## RARE EVENT SIMULATION FOR ROUGH ENERGY LANDSCAPES

Paul Dupuis
Konstantinos Spiliopoulos
Hui Wang

Lefschetz Center for Dynamical Systems
Division of Applied Mathematics
Brown University 182 George Street
Providence, RI, 02912, USA

### ABSTRACT

A rough energy landscape can be modeled by a potential function superimposed by another fast oscillating function. Modeling motion in such a rough energy landscape by a small noise stochastic differential equation with fast oscillating coefficients, we construct asymptotically optimal importance sampling schemes for the study of rare events. Standard Monte Carlo methods perform poorly for these kind of problems in the small noise limit, even without the added difficulties of the fast oscillating function. We study the situation in which the fast oscillating parameter goes to zero faster than the intensity of the noise. We identify an asymptotically optimal estimator in the sense of variance minimization using the subsolution approach. Examples and simulation results are provided.

### 1 INTRODUCTION

We study efficient importance sampling schemes for the $d$-dimensional process $X^\varepsilon \doteq \{X^\varepsilon(t), t \geq 0\}$ satisfying the first order Langevin equation

$$dX^\varepsilon(t) = -\nabla V^\varepsilon \left( X^\varepsilon(t), \frac{X^\varepsilon(t)}{\delta} \right) dt + \sqrt{\varepsilon}\sqrt{2D}dW(t), \ \ X^\varepsilon(0) = x_0. \tag{1}$$

Here $2D$ is a diffusion constant, $W(t)$ is a standard $d$-dimensional Wiener process, the two-scale potential is composed of a large-scale part and a fluctuating part in the form

$$V^\varepsilon(x, x/\delta) = \varepsilon Q(x/\delta) + V(x),$$

and $\nabla V^\varepsilon(x, x/\delta)$ denotes $\nabla[\varepsilon Q(x/\delta) + V(x)]$. It is also assumed that $\delta = \delta(\varepsilon) \downarrow 0$ in such a way that

$$\frac{\varepsilon}{\delta} \to \infty \text{ as } \varepsilon \downarrow 0. \tag{2}$$

The functions $V(x)$ and $Q(y)$ are assumed to be smooth and $Q(y)$ to be periodic with period $\lambda$ in every direction. An example of such potential is given in Figure 1.

It was suggested long ago (e.g., Lifson and Jackson 1962) that the potential surface of a protein might have a hierarchical structure with potential minima within potential minima, etc. The underlying energy landscapes of certain biomolecules can be rugged (i.e., consist of many minima separated by barriers of varying heights) due to the presence of multiple energy scales associated with the building blocks of proteins. Roughness of the energy landscapes that describe proteins has numerous effects on their folding and binding as well as on their behavior at equilibrium. Often these phenomena are described
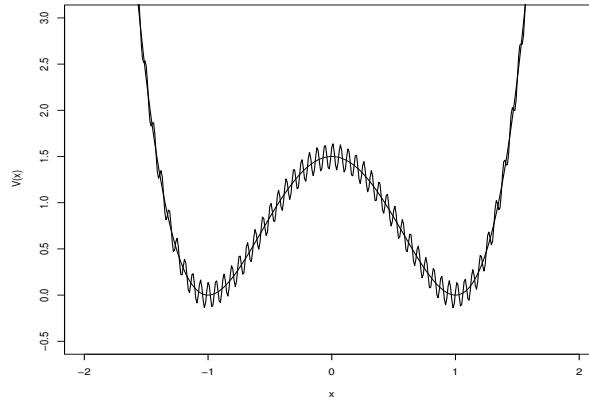
**504**

Figure 1: $V^\varepsilon(x, \frac{x}{\delta}) = \varepsilon \left( \cos(\frac{x}{\delta}) + \sin(\frac{x}{\delta}) \right) + \frac{3}{2}(x^2 - 1)^2$ and $V(x) = \frac{3}{2}(x^2 - 1)^2$ with $\varepsilon = 0.1$ and $\delta = 0.01$.

mathematically by diffusions in a rough potential made up of a smooth function superimposed by a rapidly oscillating function (see Figure 1). A representative, but by no means complete, list of references is (Ansari 2000, Bryngelson et al 1995, Hyeon and Thirumalai 2003, Mondal, Ghosh and Ray 2009, Saven, Wang and Wolynes 1994, Zwanzig 1988).

It is well known that the estimation of rare event probabilities for diffusion processes with small noise is difficult. Standard Monte Carlo sampling techniques are inadequate, in that with any fixed number of samples the relative error grows rapidly as the event of interest becomes more and more rare. Problems with fast oscillating coefficients are even more difficult, due to the additional small parameter $\delta$ and its interaction with the intensity of the noise $\varepsilon$.

The purpose of this paper is to show how importance sampling can be effectively applied to the problem of estimating rare event probabilities for diffusions in a rough energy potential. With rapidly oscillating coefficients a standard approach to model approximation and simplification is via homogenization, where one replaces the true system by a simpler approximating system with appropriate new drift and diffusion coefficients. However, in the setting of rare event estimation homogenzation must be used with some care, and in particular one cannot use importance sampling schemes that are based just on the homogenized system alone. Using large deviation results that are developed in Dupuis and Spiliopoulos 2010, we will show how to adapt the subsolution approach (Dupuis and Wang 2004, Dupuis and Wang 2007) to deal with this particular type of homogenization, and note that the methods developed can be extended to treat other forms as well.

The paper is organized as follows. In Section 2 we review importance sampling. Section 3 gives assumptions and notation, and reviews the large deviations results that will be needed. In Section 4 we recall the role of subsolutions in the context of importance sampling. In Section 5 we explain the methodology for the construction of the optimal control (equivalently, change of measure), summarize the proposed simulation scheme, and state the main results. Section 6 contains simulation results together with a discussion on the computational challenges that one faces when simulating from (1) (see Theorem 8 and the subsequent discussion).

## 2    BOUNDS ON MONTE CARLO ESTIMATION

Consider an arbitrary bounded and continuous function $h: \mathbb{R}^d \mapsto \mathbb{R}$. Suppose we are interested in estimating $\mathrm{E}_{t,x}[e^{-\frac{1}{\varepsilon}h(X^\varepsilon(T))}]$, where $\mathrm{E}_{t,x}$ denotes the expected value given $X^\varepsilon(t) = x$ and $T$ is a given final time. Consider also

$$G(t,x) = \inf_{\phi \in \mathscr{AC}([t,T];\mathbb{R}^d), \phi(t)=x} \{ S_{tT}(\phi) + h(\phi(T)) \},$$

where $\mathscr{AC}([t,T];\mathbb{R}^d)$ denotes the set of absolutely continuous functions from $[t,T]$ to $\mathbb{R}^d$ and $S_{tT}(\cdot)$ is the large deviation rate function of $\{X^\varepsilon, \varepsilon > 0\}$. The particular form of $S_{tT}(\cdot)$ was derived in Dupuis and Spiliopoulos 2010 (see also Freidlin and Sowers 1999) and will be recalled in Section 3. Then by the contraction principle

$$\lim_{\varepsilon \to 0} -\varepsilon \ln \mathrm{E}_{t,x}\left[e^{-\frac{1}{\varepsilon}h(X^\varepsilon(T))}\right] = G(t,x). \tag{3}$$

Consider any unbiased estimator, i.e., any random variable $\Gamma^\varepsilon(T)$ defined on a probability space with probability measure $\bar{\mathrm{P}}$ such that

$$\bar{\mathrm{E}}\Gamma^\varepsilon(T) = \mathrm{E}_{t,x}\left[e^{-\frac{1}{\varepsilon}h(X^\varepsilon(T))}\right],$$

where $\bar{\mathrm{E}}$ is the expectation operator associated with $\bar{\mathrm{P}}$ (we will only deal with unbiased estimators). The goal of any accelerated Monte Carlo scheme is to construct a random variable $\Gamma^\varepsilon(T)$ that can be simulated without too much difficulty, and for which the ratio of the standard deviation of the estimator to the quantity being estimated does not grow too quickly as $\varepsilon \to 0$. Because of unbiasedness, minimizing variance is equivalent to minimizing the second moment of the estimator. By Jensen's inequality

$$\bar{\mathrm{E}}(\Gamma^\varepsilon(T))^2 \geq (\bar{\mathrm{E}}\Gamma^\varepsilon(T))^2 = \left(\mathrm{E}_{t,x}\left[e^{-\frac{1}{\varepsilon}h(X^\varepsilon(T))}\right]\right)^2.$$

Taking logs in the inequality above, multiplying by $\varepsilon$ and using (3), we observe that

$$\limsup_{\varepsilon \to 0} -\varepsilon \ln \bar{\mathrm{E}}_{t,x}(\Gamma^\varepsilon(T))^2 \leq 2G(t,x).$$

Therefore, $2G(t,x)$ is the best possible rate of decay. If one can prove

$$\liminf_{\varepsilon \to 0} -\varepsilon \ln \bar{\mathrm{E}}_{t,x}(\Gamma^\varepsilon(T))^2 \geq 2G(t,x),$$

then $\Gamma^\varepsilon(T)$ achieves this best rate. Such an estimator will be called *asymptotically optimal*.

In the framework of importance sampling we consider probability measures $\bar{\mathrm{P}}$ that are absolutely continuous with respect to P (the probability measure under which the expectation operator E is defined), and use the particular form $\Gamma^\varepsilon(T) = e^{-\frac{1}{\varepsilon}h(X^\varepsilon(T))}\left[d\mathrm{P}/d\bar{\mathrm{P}}\right]$ (note that this is an unbiased estimator). Our goal is to identify a measure $\bar{\mathrm{P}}$ (or more precisely a sequence of measures indexed by $\varepsilon$) so that $\Gamma^\varepsilon(T)$ is asymptotically optimal.

## 3 PRELIMINARY RESULTS

In this section we introduce assumptions and notation, and also review the results on large deviations for (1) from Dupuis and Spiliopoulos 2010 that will be used. We work with the canonical filtered probability space $(\Omega, \mathfrak{F}, \mathrm{P})$ equipped with a filtration $\mathfrak{F}_t$ that satisfies the usual conditions, i.e., $\mathfrak{F}_t$ is right continuous and $\mathfrak{F}_0$ contains all P-negligible sets. Regarding the SDE (1) we impose the following condition:

**Condition 1** The function $Q(y)$ is $C^2(\mathbb{R}^d)$ with all derivatives continuous and periodic with respect to the $d$−dimensional torus $\mathscr{Y} = \mathbb{T}^d = [0,\lambda]^d$. The function $V(x)$ is $C^1(\mathbb{R}^d)$ with bounded and Lipschitz continuous derivative.

For notational convenience we define the operator $\cdot : \cdot$, where for two matrices $A = [a_{ij}], B = [b_{ij}]$

$$A : B \doteq \sum_{i,j} a_{ij}b_{ij}.$$

Consider the operator

$$\mathscr{L} = -\nabla Q(y) \cdot \nabla + DI : \nabla\nabla$$

equipped with periodic boundary conditions. It is easy to see that there is an explicit formula for the invariant density $\mu(y)$ corresponding to $\mathscr{L}$, which is the Gibbs distribution

$$\mu(y) = \frac{1}{L} e^{-\frac{Q(y)}{D}}, \quad L = \int_{\mathscr{Y}} e^{-\frac{Q(y)}{D}} dy. \tag{4}$$

Under Condition 1, for each $\ell \in \{1,\ldots,d\}$ there exists a unique function $\chi_\ell(y)$ that is twice differentiable and $\lambda-$periodic in every direction in $y$, and which solves

$$\mathscr{L}\chi_\ell(y) = (\nabla Q)_\ell(y), \quad \int_{\mathscr{Y}} \chi_\ell(y)\mu(dy) = 0. \tag{5}$$

For a proof see Bensoussan, Lions and Papanicolaou 1978, Theorem 3.3.4. The equation (5) is known as a *cell problem*. Let $\chi = (\chi_1, \cdots, \chi_d)$. As we shall see below, $\chi$ plays a crucial role in the design of asymptotically efficient importance sampling schemes for multiscale diffusions.

The large deviations principle for the solution of (1) together with the identification of a control that achieves the large deviations lower bound was derived in Dupuis and Spiliopoulos 2010. This control is closely related to the design of efficient importance sampling schemes (see Remark 4). We remark here that in Dupuis and Spiliopoulos 2010 the authors consider the case of general drift and diffusion coefficients, i.e., $b(x,y)$ in place of $-\nabla Q(y)$, $c(x,y)$ in place of $-\nabla V(x)$, and $\sigma(x,y)$ in place of $\sqrt{2D}$ as the diffusion coefficient. Note that the large deviations principle (without the identification of a nearly optimal control) for the particular case $b = b(y), c = c(y), \sigma = \sigma(y)$ was previously derived in Freidlin and Sowers 1999, and for $b = c = 0, \sigma = \sigma(y)$ in Baldi 1991.

**Theorem 2** Let $\{X^\varepsilon, \varepsilon > 0\}$ be the unique strong solution to (1) and suppose that $\mu(dy)$ is the Gibbs distribution (4) and $\chi(y)$ is defined by (5). Under Condition 1 $\{X^\varepsilon, \varepsilon > 0\}$ satisfies a large deviations principle with rate function

$$S_{0T}(\phi) = \begin{cases} \frac{1}{2}\int_0^T (\dot\phi(s) - r(\phi(s)))^T q^{-1}(\dot\phi(s) - r(\phi(s)))ds & \text{if } \phi \in \mathscr{AC}([0,T];\mathbb{R}^d), \phi(0) = x_0 \\ +\infty & \text{otherwise,} \end{cases}$$

where

- $r(x) = -\left[\int_{\mathscr{Y}}(I + \frac{\partial\chi}{\partial y})(y)\mu(dy)\right]\nabla V(x),$
- $q = 2D\int_{\mathscr{Y}}(I + \frac{\partial\chi}{\partial y})(y)(I + \frac{\partial\chi}{\partial y})^T(y)\mu(dy).$

When the fluctuating part is separable, i.e., $Q(y_1, y_2, \ldots, y_d) = Q_1(y_1) + Q_2(y_2) + \cdots + Q_d(y_d)$, the rate function can be made more explicit. We summarize this result in the following corollary.

**Corollary 3** Let $\{X^\varepsilon, \varepsilon > 0\}$ be the unique strong solution to (1). Assume $Q(y_1, y_2, \cdots, y_d) = Q_1(y_1) + Q_2(y_2) + \cdots + Q_d(y_d)$. Under Condition 1, $\{X^\varepsilon, \varepsilon > 0\}$ satisfies a large deviations principle with rate function

$$S_{0T}(\phi) = \begin{cases} \frac{1}{2}\int_0^1 (\dot\phi(s) - r(\phi(s)))^T q^{-1}(\dot\phi(s) - r(\phi(s)))ds & \text{if } \phi \in \mathscr{AC}([0,T];\mathbb{R}^d), \phi(0) = x_0 \\ +\infty & \text{otherwise,} \end{cases}$$

where

$$r(x) = -\Theta\nabla V(x), \quad q = 2D\Theta, \quad \Theta = \text{diag}\left[\frac{1}{Z_1\hat{Z}_1}, \cdots, \frac{1}{Z_d\hat{Z}_d}\right]$$

and for $i = 1, 2, \ldots, d$

$$Z_i = \int_{\mathbb{T}} e^{-\frac{Q_i(y_i)}{D}} dy_i, \quad \hat{Z}_i = \int_{\mathbb{T}} e^{\frac{Q_i(y_i)}{D}} dy_i.$$

Note that the effective diffusivity matrix $q$ in Corollary 3 differs from the original one. In particular, $q$ is a diagonal matrix and one can check that the elements of $q$ are always smaller than the corresponding diagonal elements of the original diffusion matrix. In the original multiscale problem there are many small energy barriers. These are not captured by the homogenized potential and hence must be accounted for in the homogenized process, and it is the trapping from these numerous local minima that is responsible for the reduction of the diffusion coefficient.

## 4 SUBSOLUTIONS TO A RELATED PDE

In this section, we recall the notion of a subsolution to a Hamilton-Jacobi-Bellman equation of the type

$$
\begin{aligned}
U_t(t,x) + \bar{H}(x, \nabla_x U(t,x)) &= 0 \\
U(T,x) &= h(x)
\end{aligned}
\tag{6}
$$

where $\bar{H}(x,p)$ is a Hamiltonian. The use of subsolutions in the context of importance sampling will become clear in Section 5.

**Definition 1** A function $\bar{U}(t,x) : [0,T] \times \mathbb{R}^d \mapsto \mathbb{R}$ is said to be a *classical subsolution* to the HJB equation (6) if

1. $\bar{U}$ is continuously differentiable,
2. $\partial_t \bar{U}(t,x) + \bar{H}(x, \nabla_x \bar{U}(t,x)) \geq 0$ for every $(t,x) \in (0,T) \times \mathbb{R}^d$.
3. $U(T,x) \leq h(x)$ for every $x \in \mathbb{R}^d$.

## 5 METHODOLOGY AND MAIN RESULTS

Define the function

$$
\Phi^\varepsilon(t,x) = \mathbb{E}_{t,x}\left[ e^{-\frac{1}{\varepsilon} h(X^\varepsilon(T))} \right],
$$

where $X^\varepsilon$ is the strong solution to the uncontrolled process (1) with $X(t) = x$ as the initial condition. Next observe that the function

$$
U^\varepsilon(t,x) = -\varepsilon \ln \Phi^\varepsilon(t,x)
$$

solves the Hamilton-Jacobi-Bellman (HJB) equation

$$
\begin{aligned}
\partial_t U^\varepsilon + H^\varepsilon(x, x/\delta, \nabla_x U^\varepsilon, \nabla_x^2 U^\varepsilon) &= 0 \\
U^\varepsilon(T,x) &= h(x),
\end{aligned}
$$

where

$$
H^\varepsilon(x,y,p,q) = \left\langle -\frac{\varepsilon}{\delta} \nabla Q(y) - \nabla V(x), p \right\rangle - D\|p\|^2 + \varepsilon D I : q.
$$

$U^\varepsilon(t,x)$ converges, as $\varepsilon \downarrow 0$, to the unique bounded, continuous viscosity solution $U(t,x)$ of the equation

$$
\begin{aligned}
U_t(t,x) + \bar{H}(x, \nabla_x U(t,x)) &= 0 \\
U(T,x) &= h(x)
\end{aligned}
\tag{7}
$$

where

$$
\bar{H}(x,p) = \langle r(x), p \rangle - \frac{1}{2} \left\| q^{1/2} p \right\|^2
$$

and where $r(x)$ and $q$ are as in Theorem 2. Note $\varepsilon/\delta \to \infty$ and so the proof is not straightforward. However, the result can be derived as in Buckdahn and Ichibara 2005 making use of the solution $\chi(y)$ to the cell problem (5) and results in section VII.8 of Fleming and Soner 2006. Define the control

$$
\bar{u}(t,x,y) = -\sqrt{2D}\left( I + \frac{\partial \chi}{\partial y}(y) \right)^T \nabla_x U(t,x).
\tag{8}
$$

Theorem 6 shows that, under appropriate regularity conditions, such a control gives rise to an asymptotically optimal estimator. Of course finding a solution to (8), even via numerical methods, is in general difficult in high dimensions. However, it is known for problems without multiscale features that it is sufficient to work with suitable subsolutions rather than solutions to the corresponding HJB equation Dupuis and Wang 2007. This is true in the present setting as well, and in fact Theorem 6 shows that it is enough to construct appropriate subsolutions to (7) (see Section 4). A classical solution to (7) is always a classical subsolution to (7) (see Definition 1), and hence there is greater flexibility in the design of schemes.

Before giving precise statements we make two observations. Although stated with reference to the solution to the HJB equation, the analogous statements apply to subsolutions and the schemes based on them as well.

**Remark 4** Assume that the function

$$G(t,x) = \inf_{\phi \in \mathscr{C}([t,T];\mathbb{R}^d), \phi(t)=x} [S_{tT}(\phi) + h(\phi(T))] \tag{9}$$

is continuous and that a minimizer of the right hand side exists for every $(t,x) \in [0,T] \times \mathbb{R}^d$. Then by Theorem II.7.2 in Fleming and Soner 2006 $G(t,x)$ is the unique viscosity solution of (7), i.e., $G = U$.

For points $(t,x) \in [0,T] \times \mathbb{R}^d$ such that there exists a unique minimizer $\bar{\phi}_{t,x}$ of the right hand side of (9), $\bar{u}$ defined by (8) can be equivalently written as

$$\bar{u}(t,x,y) = \sqrt{2D} \left( I + \frac{\partial \chi}{\partial y}(y) \right)^T q^{-1}(\dot{\bar{\phi}}_{t,x} - r(x)).$$

(In fact, this particular form of the control is useful in proving the large deviations lower bound Dupuis and Spiliopoulos 2010).

**Remark 5** Notice that the optimal control has the extra coefficient

$$\sqrt{2D} \left( I + \frac{\partial \chi}{\partial y}(y) \right)^T (q^{-1/2})^T$$

multiplying the expression that one would expect based on just the rate function, i.e., $-(q^{1/2})^T \nabla_x U(t,x)$. This reflects the fact that one cannot base an asymptotically optimal scheme on the homogenized problem alone. Instead one needs to include more detailed information on the fast oscillatory coefficients. The importance of this term will be numerically illustrated in Section 6.

From now on we write $\bar{u}$ to refer to the control defined in (8) when $U = \bar{U}$, where $\bar{U}$ is a classical subsolution. For notational convenience, we sometimes abbreviate

$$\bar{u}(s) = \bar{u}\left(s, X_{t,x}^\varepsilon(s), \frac{X_{t,x}^\varepsilon(s)}{\delta}\right).$$

Consider the family of distributions $\bar{P}$ given by the change of measure

$$\frac{d\bar{P}}{dP} = Z_{t,T} = e^{-\frac{1}{2\varepsilon} \int_t^T |\bar{u}(s)|^2 ds + \frac{1}{\sqrt{\varepsilon}} \int_t^T \bar{u}(s) dW(s)}.$$

Note that $\bar{P}$ is absolutely continuous with respect to P. By Girsanov's Theorem the process $\bar{W}(s) = W(s) - \frac{1}{\sqrt{\varepsilon}} \int_t^s \bar{u}(\rho) d\rho$ is a Brownian motion under measure $\bar{P}$ and the dynamics of $X^\varepsilon$ under $\bar{P}$ are given by

$$
\begin{aligned}
dX_{t,x}^\varepsilon(s) &= \left[ -\frac{\varepsilon}{\delta} \nabla Q\left(\frac{X_{t,x}^\varepsilon(s)}{\delta}\right) - \nabla V\left(X_{t,x}^\varepsilon(s)\right) + \sqrt{2D}\bar{u}(s) \right] ds + \sqrt{\varepsilon}\sqrt{2D} d\bar{W}(s) \\
X_{t,x}^\varepsilon(t) &= x.
\end{aligned}
$$

By the standard change of measure

$$\mathrm{E}_{t,x}\left[e^{-\frac{1}{\varepsilon}h(X^{\varepsilon}(T))}\right] = \bar{\mathrm{E}}_{t,x}\left[e^{-\frac{1}{\varepsilon}h(X^{\varepsilon}(T))}\frac{d\mathrm{P}}{d\bar{\mathrm{P}}}(X^{\varepsilon})\right]$$

Define

$$Q_h^{\varepsilon}(t,x;\bar{u}) = \bar{\mathrm{E}}_{t,x}^{\varepsilon}\left(\left[e^{-\frac{1}{\varepsilon}h(X^{\varepsilon}(T))}\frac{d\mathrm{P}}{d\bar{\mathrm{P}}}(X^{\varepsilon}(\cdot))\right]^2\right).$$

The main theorem of this paper is Theorem 6, whose proof can be found in Dupuis, Spiliopoulos and Wang 2011.

**Theorem 6** Let $h : \mathbb{R}^d \mapsto \mathbb{R}$ be an arbitrary bounded and continuous function and assume Condition 1. Consider the control $\bar{u}(t,x,y)$ defined by (8) with a classical subsolution $\bar{U}$ in place of $U$. Then

$$\liminf_{\varepsilon \to 0} -\varepsilon \ln Q_h^{\varepsilon}(t,x;\bar{u}) \geq G(t,x) + \bar{U}(t,x). \tag{10}$$

**Remark 7** Theorem 6 quantifies the speedup obtained using the subsolution approach. Standard Monte-Carlo corresponds to a lower bound as in (10) with $\bar{U}(t,x) = 0$. The optimal bound from (10) is twice the value of the maximal subsolution (which of course is twice the value of the solution), $2G(t,x)$. Hence a subsolution that corresponds to a bound between these two extremes identifies an important sampling scheme that performs better than standard Monte-Carlo. It is also important to note that one can have schemes corresponding to different subsolutions whose performance as measured by (10) is the same since one can, in principle, construct different subsolutions whose values are the same at the point $(t,x)$.

## 5.1 Simulation Scheme

In this subsection we summarize the proposed simulation scheme. Let $\bar{X}^{\varepsilon} = \bar{X}^{\varepsilon,\bar{u}}$ be the solution to the SDE

$$\begin{aligned}
d\bar{X}^{\varepsilon}(s) &= \left[-\frac{\varepsilon}{\delta}\nabla Q\left(\frac{\bar{X}^{\varepsilon}(s)}{\delta}\right) - \nabla V\left(\bar{X}^{\varepsilon}(s)\right) + \sqrt{2D}\bar{u}(s)\right]dt + \sqrt{\varepsilon}\sqrt{2D}dW(s) \tag{11} \\
\bar{X}(0) &= x_0.
\end{aligned}$$

The simulation scheme that estimates $\theta_{t,x}(\varepsilon) \doteq \mathbb{E}_{t,x}\left[e^{-\frac{1}{\varepsilon}h(X^{\varepsilon}(T))}\right]$ is as follows.

1. Consider $\bar{u}(t,x,y)$ as defined by (8) with $U = \bar{U}$, where $\bar{U}$ is a classical subsolution.
2. Construct
$$Z_j^{\bar{u}} \doteq e^{-\frac{1}{2\varepsilon}\int_t^T\left|\bar{u}\left(s,\bar{X}_j^{\varepsilon,\bar{u}}(s),\bar{X}_j^{\varepsilon,\bar{u}}(s)/\delta\right)\right|^2 ds - \frac{1}{\sqrt{\varepsilon}}\int_t^T\bar{u}\left(s,\bar{X}_j^{\varepsilon,\bar{u}}(s),\bar{X}_j^{\varepsilon,\bar{u}}(s)/\delta\right)dW_j(s)},$$

   and $(W_j, \bar{X}_j^{\varepsilon,\bar{u}})$ are the $N$ independent samples generated from (11) with control
$\bar{u}(s) = \bar{u}(s,\bar{X}_j^{\varepsilon,\bar{u}}(s),\bar{X}_j^{\varepsilon,\bar{u}}(s)/\delta)$.
3. Return the estimate
$$\hat{\theta}_{t,x}(\varepsilon) \doteq \frac{1}{N}\sum_{j=1}^{N}\left[e^{-\frac{1}{\varepsilon}h(\bar{X}_j^{\varepsilon,\bar{u}})}Z_j^{\bar{u}}\right]. \tag{12}$$

## 6  SIMULATION RESULTS

In this section we test the performance of the proposed estimator (12) by estimating $\theta(\varepsilon) \doteq \mathrm{E}_{t_0,x_0}[e^{-\frac{1}{\varepsilon}h(X^{\varepsilon}(T))}]$ for a continuous function $h$. We consider a one dimensional example and assume that the diffusion process $X^{\varepsilon}$ satisfies the first order Langevin equation (1).

Let $u_2(t,x) = -\sqrt{q}G_x(t,x)$. This is the control one would consider without fully taking into account the homogenization phenomena. In other words, this control does not have the prefactor $\sqrt{2Dq}(1+\chi_y(y))$ (see Remark 5).

We compare the proposed estimator $\hat{\theta}_1$ given by (12) with the standard Monte-Carlo estimator $\hat{\theta}_0$, and and also with the estimator $\hat{\theta}_2$ based on $u_2$, i.e.,

$$\hat{\theta}_0(\varepsilon) = \frac{1}{N}\sum_{j=1}^{N}\left[e^{-\frac{1}{\varepsilon}h(X_j^\varepsilon(1))}\right] \text{ and } \hat{\theta}_2(\varepsilon) = \frac{1}{N}\sum_{j=1}^{N}\left[e^{-\frac{1}{\varepsilon}h(\bar{X}_j^{\varepsilon,u_2}(1))}Z_j^{u_2}\right].$$

As we will see the estimator given by (12) outperforms the other two estimators significantly as $\varepsilon \downarrow 0$.

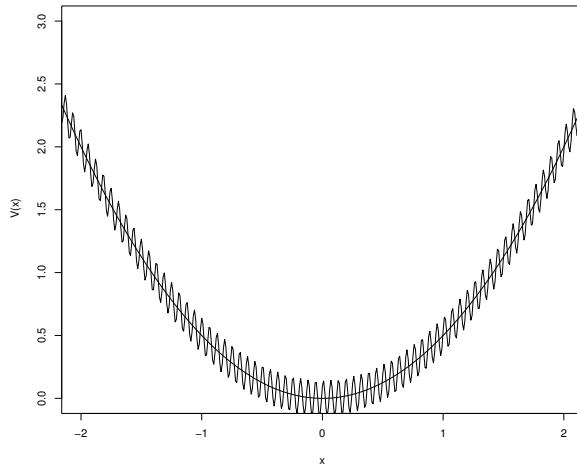For purposes of illustration, we consider the potential function that is drawn in Figure 2, i.e,



Figure 2: $V^\varepsilon(x,\frac{x}{\delta}) = \varepsilon\left(\cos(\frac{x}{\delta})+\sin(\frac{x}{\delta})\right) + \frac{1}{2}x^2$ and $V(x) = \frac{1}{2}x^2$ with $\varepsilon = 0.1$ and $\delta = 0.01$.

$$V(x) = \frac{1}{2}x^2$$
$$Q(y) = \cos(y)+\sin(y)$$

The function $h(\cdot)$ is chosen to be

$$h(x) = \begin{cases} (x-1)^2 & x \geq 0, \\ (x+1)^2 & x < 0. \end{cases}$$

The period is $\lambda = 2\pi$. Let $\kappa = 4\pi^2/L\hat{L}$. The effective drift $r(x)$ and diffusion $q$ are respectively

$$r(x) = -\kappa x, \quad q = 2D\kappa.$$

The limiting HJB equation (7) is

$$U_t(t,x) - \kappa x U_x(t,x) - \kappa D|U_x(t,x)|^2 = 0,$$
$$U(T,x) = h(x).$$

Under appropriate conditions (which are satisfied here) one can represent the unique viscosity solution to this equation by the value function

$$G(t,x) = \inf_{\phi \in \mathscr{C}([t,T];\mathbb{R}^d),\phi(t)=x}\left[S_{tT}(\phi)+h(\phi(T))\right],$$

**511**

where $S_{tT}(\cdot)$ is given by Corollary 3 with $r(x) = -\kappa x$ and $q = 2D\kappa$. One can solve this variational problem explicitly and obtain

$$U(t,x) = \frac{(e^{\kappa T} - |x|e^{\kappa t})^2}{-2De^{2\kappa t} + (1+2D)e^{2\kappa T}}.$$

One can fit this problem into the subsolution framework by defining

$$U_1(t,x) = \frac{(e^{\kappa T} - xe^{\kappa t})^2}{-2De^{2\kappa t} + (1+2D)e^{2\kappa T}}, \quad U_2(t,x) = \frac{(e^{\kappa T} + xe^{\kappa t})^2}{-2De^{2\kappa t} + (1+2D)e^{2\kappa T}}$$

and then considering the subsolution $\bar{U}(t,x) = U_1(t,x) \wedge U_2(t,x)$. In general, one needs to mollify such a $\bar{U}$ in order to produce a smooth subsolution. However, it is known (see Vanden-Eijnden and Weare 2009 for an analogous situation) that the bound on the performance is still valid if the subsolution is the minimum of two classical sense solutions with a single discontinuity interface.

The optimal control is given by $\bar{u}(t,x,y) = -\sqrt{2D}(1 + \frac{\partial \chi}{\partial y}(y))\bar{U}_x(t,x)$. Observing that

$$1 + \frac{\partial \chi}{\partial y}(y) = \frac{2\pi}{\hat{L}}e^{Q(y)/D} = \frac{2\pi}{\hat{L}}e^{(\cos(y)+\sin(y))/D},$$

we obtain the expression for the optimal control

$$u_1(t,x) \doteq \bar{u}(t,x) = \begin{cases} -\sqrt{2D}\frac{2\pi}{\hat{L}}e^{(\cos(\frac{x}{\delta})+\sin(\frac{x}{\delta}))/D}\frac{-2e^{\kappa t}(e^{\kappa T}-xe^{\kappa t})}{-2De^{2\kappa t}+(1+2D)e^{2\kappa T}} & x > 0 \\ -\sqrt{2D}\frac{2\pi}{\hat{L}}e^{(\cos(\frac{x}{\delta})+\sin(\frac{x}{\delta}))/D}\frac{2e^{\kappa t}(e^{\kappa T}+xe^{\kappa t})}{-2De^{2\kappa t}+(1+2D)e^{2\kappa T}} & x < 0 \end{cases}.$$

On the other hand the control $u_2$ is

$$u_2(t,x) = \begin{cases} -\sqrt{q}\frac{-2e^{\kappa t}(e^{\kappa T}-xe^{\kappa t})}{-2De^{2\kappa t}+(1+2D)e^{2\kappa T}} & x > 0 \\ -\sqrt{q}\frac{2e^{\kappa t}(e^{\kappa T}+xe^{\kappa t})}{-2De^{2\kappa t}+(1+2D)e^{2\kappa T}} & x < 0 \end{cases}.$$

We choose $D = 1$, initial point $(t_0, x_0) = (0, 0.05)$ and final time $T = 1$, and calculate that

$$\hat{L} = 9.83999, \quad \kappa = \frac{4\pi^2}{L\hat{L}} = 0.407728.$$

Before presenting the simulation results we comment on the simulation aspect of the problem. Let $Y_n^\varepsilon$ denote the numerical approximation to $X_t^\varepsilon$ with weak order of convergence $p$. For such a numerical scheme one has the following error bound.

**Theorem 8** If $\varepsilon, \delta$ and the discretization step $\Delta$ are such that $\Delta\varepsilon/\delta^2 \ll 1$, then for every $T > 0$ and every smooth function $f$ with compact support, there exists $h_0 > 0$ and a constant $C_0$ that is independent of $\varepsilon, \delta$ and $\Delta$ such that whenever $\Delta\varepsilon/\delta^2 < h_0$

$$\sup_{n \leq T/\Delta} |\mathbb{E}_{x_0} f(X_{t_n}^\varepsilon) - \mathbb{E}_{x_0} f(Y_n^\varepsilon)| \leq C_0 \left(\Delta\varepsilon/\delta^2\right)^p. \tag{13}$$

The proof of this theorem follows standard arguments (see for example E, Liu and Vanden-Eijnden 2005 for a proof in an analogous situation). It illustrates the computational difficulty of the problem. Consider a fixed error tolerance $\zeta$. The error bound (13) indicates that we would need a time step of order

$$\Delta = O\left(\frac{\delta^2}{\varepsilon}\zeta^{1/p}\right).$$

The corresponding computational cost per unit of time is proportional to $N = 1/\Delta$.

In the following table we summarize some simulation results. We used a predictor-corrector Euler scheme to simulate trajectories of (1) (or of the associated control SDE). By Theorem 8 we know that the error in the Euler approximation is bounded by the term $\Delta\varepsilon/\delta^2$, where $\Delta$ is the discretization step. For each choice of $\varepsilon$ and $\delta$ we choose $\Delta$ so that the aforementioned error bound is of the order of 0.001. In other words, we set $\Delta = 0.001(\delta^2/\varepsilon)$. Since $\delta^2/\varepsilon \to 0$ as $\varepsilon \to 0$ by assumption (2), one should make the discretization step $\Delta$ smaller as $\varepsilon$ gets smaller. Simulations were done using parallel computing in the C programming language. The simulation results are based on $N = 10^7$ trajectories and, as numerical experimentation indicated, the reported values for the accelerated Monte-Carlo estimator and its estimated relative error are stable. We used Mersenne Twister (Matsumoto and Nishimura 1998) for the random number generator.

The measure used to compare the different estimators is the relative error of the estimator. In order to distinguish among the different Monte-Carlo procedures, we denote the relative error for control $u = 0$ (standard Monte-Carlo), $u = u_1$ (importance sampling with the optimal control) and $u = u_2$ (importance sampling without taking into consideration the homogenization effect) by $\rho_0^\varepsilon$, $\rho_1^\varepsilon$, and $\rho_2^\varepsilon$, respectively.

Let $\mathrm{Var}(\hat{\theta}_i(\varepsilon))$ be the variance of the estimator based on the change of measure induced by the control $u_i$ for $i = 0, 1, 2$. The corresponding relative error of this estimator is defined by

$$\rho_i^\varepsilon \doteq \rho(\hat{\theta}_i(\varepsilon)) = \frac{\sqrt{\mathrm{Var}(\hat{\theta}_i(\varepsilon))}}{\theta(\varepsilon)} = \frac{1}{\sqrt{N}} \sqrt{\frac{\bar{\mathrm{E}}_{t_0,x_0}^\varepsilon \left[ e^{-\frac{2}{\varepsilon}h(X^\varepsilon(T))} \left( \frac{d\mathrm{P}}{d\bar{\mathrm{P}}_i}(X_{t_0,x_0}^\varepsilon) \right)^2 \right]}{\left( \mathrm{E}_{t_0,x_0}^\varepsilon \left[ e^{-\frac{1}{\varepsilon}h(X^\varepsilon(T))} \right] \right)^2} - 1}. \tag{14}$$

Theorem 6 then implies that for small $\varepsilon$, $\rho_1^\varepsilon$ will remain close to 0 whereas for $i = 0, 2$, $\rho_i^\varepsilon$ will get larger as $\varepsilon$ gets smaller.

Notice however that we do not have a closed form expression for $\theta(\varepsilon) = \bar{\mathrm{E}}_{t_0,x_0}^\varepsilon[\hat{\theta}_i(\varepsilon)] = \mathrm{E}_{t_0,x_0}^\varepsilon[e^{-\frac{1}{\varepsilon}h(X^\varepsilon(T))}]$, the true value of the quantity that we want to estimate or for its variance. Thus, we cannot compute (14) in practice. Instead, we compute the estimated relative error

$$\hat{\rho}_i^\varepsilon \doteq \hat{\rho}(\hat{\theta}_i(\varepsilon)) = \frac{\sqrt{\widehat{\mathrm{Var}(\hat{\theta}_i(\varepsilon))}}}{\hat{\theta}_1(\varepsilon)}.$$

The corresponding results are in Table 1.

Table 1: Comparison table.

| No. | $\varepsilon$ | $\delta$ | $\varepsilon/\delta$ | $\hat{\theta}_1(\varepsilon)$ | $\hat{\rho}_0^\varepsilon$ | $\hat{\rho}_1^\varepsilon$ | $\hat{\rho}_2^\varepsilon$ |
|---|---|---|---|---|---|---|---|
| 1 | 0.25 | 0.1 | 2.5 | $2.25e-01$ | $3.36e-4$ | $1.76e-3$ | $6.34e-4$ |
| 2 | 0.125 | 0.04 | 3.125 | $3.65e-02$ | $8.40e-4$ | $1.80e-3$ | $1.43e-3$ |
| 3 | 0.0625 | 0.015625 | 4 | $8.75e-04$ | $1.06e-2$ | $1.29e-3$ | $4.08e-3$ |
| 4 | 0.03125 | 0.007 | 4.46 | $6.87e-07$ | $4.47e-2$ | $8.00e-4$ | $3.32e-2$ |
| 5 | 0.025 | 0.004 | 6.25 | $1.61e-08$ | $6.85e-2$ | $7.55e-4$ | $3.06e-2$ |
| 6 | 0.02 | 0.002 | 10 | $1.99e-10$ | $4.09e-1$ | $3.82e-4$ | $4.97e-2$ |
| 7 | 0.015 | 0.0013 | 11.54 | $1.37e-13$ | $2.53e-1$ | $3.01e-4$ | $1.86e-1$ |

As we see from Table 1, the estimator corresponding to the change of measure induced by the control $u_1$ outperforms both the standard Monte Carlo estimator ($u = 0$) and the estimator corresponding to the change of measure induced by the control $u_2$ as $\varepsilon$ gets smaller. In particular, as $\varepsilon$ gets smaller, $\hat{\rho}_1^\varepsilon$ is considerably smaller than both $\hat{\rho}_0^\varepsilon$ and $\hat{\rho}_2^\varepsilon$. Moreover, in contrast to the estimated relative errors $\hat{\rho}_0^\varepsilon$ and $\hat{\rho}_2^\varepsilon$, $\hat{\rho}_1^\varepsilon$ decreases as $\varepsilon, \delta$ get smaller while $\varepsilon/\delta$ increases.

## ACKNOWLEDGMENTS

## REFERENCES

Ansari, A. 2000. "Mean first passage time solution of the Smoluchowski equation: Application of relaxation dynamics in myoglobin". *Journal of Chemical Physics*. 112 (5):2516-2522.

Baldi, P. 1991. "Large deviations for diffusions processes with homogenization and applications" *Annals of Probability*. 19(2): 509–524.

Bensoussan A., J.L. Lions and G. Papanicolaou. 1978. *Asymptotic Analysis for Periodic Structures*. Vol 5, Studies in Mathematics and its Applications, North-Holland Publishing Co., Amsterdam.

Buckdahn R.,and N. Ichihara. 2005. "Limit theorem for controlled backward SDEs and homogenization of Hamilton-Jacobi-Bellman equations". *Applied Mathematics and Optimization*. 51: 1-33.

Bryngelson J.D., J. N. Onuchic, N. D. Socci and P. G. Wolynes. 1995. "Funnels, pathways and the energy landscape of protein folding: A synthesis". *Proteins*. 21 (3): 167-195.

Dupuis P., and K. Spiliopoulos. 2010. "Large deviations for multiscale problems via weak convergence methods". submitted.

Dupuis P., K. Spiliopoulos and H. Wang. 2011. "Importance sampling for multiscale diffusions". preprint.

Dupuis P., and H. Wang. 2004. "Importance sampling, large deviations and differential games". *Stochastics and Stochastics Reports* 76: 481-508.

Dupuis P., and H. Wang. 2007. "Subsolutions of an Isaacs equation and efficient schemes of importance sampling". *Mathematics of Operations Research*, 32 (3): 723-757.

E W., D. Liu and E. Vanden-Eijnden. 2005. "Analysis of multiscale methods for stochastic differential equations". *Comm. Pure Appl. Math.*, 58 (11): 1544-1585.

Fleming W.H., and H.M. Soner. 2006. *Controlled Markov Processes and Viscosity Solutions*, Springer, 2nd Ed.

Freidlin M., and R, Sowers. 1999. "A comparison of homogenization and large deviations, with applications to wavefront propagation". *Stochastic Process and Their Applications*, 82 (1): 23–52.

Hyeon C., and D. Thirumalai. 2003. "Can energy landscapes roughness of proteins and RNA be measured by using mechanical unfolding experiments¿". *Proc. Natl. Acad. Sci.* USA, 100 (18): 10249-10253.

Lifson S., and J.L. Jackson. 1962. "On the self-diffusion of ions in a polyelectrolyte solution" *Journal of Chemical Physics*. 36: 2410-2414.

Mondal D., P.K. Ghosh and D.S. Ray. 2009. "Noise-induced transport in a rough racket potential", *Journal of Chemical Physics*, 130: 074703.1-074703.7.

Matsumoto M.,and T. Nishimura. 1998. "Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator". *ACM Trans. Model. Comput. Simul.*, 8(1): 3-30.

Saven J.G., J. Wang and P.G.Wolynes. 1994. "Kinetics of protein folding: The dynamics of globally connected rough energy landscapes with biases". *Journal of Chemical Physics*, 101 (12): 11037-11043.

Vanden-Eijnden E., and J. Weare. 2009. "Rare event simulation with vanishing error for small noise diffusions". preprint.

Zwanzig R. 1988. "Diffusion in a rough potential". *Proc. Natl. Acad. Sci. USA*, 85: 2029-2030.

## AUTHOR BIOGRAPHIES

**PAUL DUPUIS** is Professor in the Division of Applied Mathematics at Brown University. He received a B.Sc. degree from Brown University in Applied Mathematics, an M.Sc. degree from Northwestern University in Engineering and Applied Mathematics, and a Ph.D. from Brown University in Applied Mathematics. Email: dupuis@dam.brown.edu.

**KONSTANTINOS SPILIOPOULOS** is a Prager Assistant Professor at the Division of Applied Mathematics at Brown University. He received his undergraduate diploma in Applied Mathematics and Physical Sciences from National Technical University of Athens in Greece and Ph.D. in Mathematical Statistics from University of Maryland at College Park in USA. Email: kspiliop@dam.brown.edu.

**HUI WANG** is an Associate Professor at the Division of Applied Mathematics at Brown University. He received his undergraduate diploma in Engineering from Tsinghua University in P.R. China and Ph.D. in Statistics from Columbia University in USA. Email: huiwang@cfm.brown.edu.