

## Rare event simulation in the neighborhood of a rest point

Paul Dupuis

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

Konstantinos Spiliopoulos

Department of Mathematics and Statistics  
Boston University 111 Cummington Mall, Boston, MA, 02215, USA

### ABSTRACT

In this paper, we construct efficient importance sampling Monte Carlo schemes for finite time exit probabilities in the presence of rest points. We focus on reversible diffusion processes with small noise that have an asymptotically stable equilibrium point. The main novelty of the work is the inclusion of rest points in the domain of interest. We motivate the construction of schemes that perform well both asymptotically and non-asymptotically. We concentrate on the regime where the noise is small and the time horizon is large. Examples and simulation results are provided.

### 1 INTRODUCTION

In this paper we develop provably efficient rare event simulation schemes for the estimation of finite time exit and transition probabilities that involve escape from an equilibrium point. We consider the  $d$ -dimensional process

$$dX_t^\varepsilon = -\nabla V(X_t^\varepsilon)dt + \sqrt{\varepsilon\beta}dW_t, \quad X_0^\varepsilon = y, \quad (1)$$

where  $V(x)$  is an appropriately smooth function,  $\beta > 0$  is a fixed constant,  $0 < \varepsilon \ll 1$  and  $W_t$  is a standard  $d$ -dimensional Wiener process. For  $L > 0$ , define  $A_L = \{x \in \mathbb{R}^d : V(x) = L\}$  and let

$$\mathcal{D} = \{x \in \mathbb{R}^d : V(x) \leq L\}.$$

Let  $\tau^\varepsilon$  be the time of exit of the process  $X_t^\varepsilon$  from  $\mathcal{D}$ , i.e.,

$$\tau^\varepsilon = \left\{ t > 0 : X_t^\varepsilon \in \mathbb{R}^d \setminus \mathcal{D} \right\}.$$

We are interested in developing provably efficient importance sampling schemes for the estimation of quantities of the form

$$\theta^\varepsilon(y) = \mathbb{P}_y[X^\varepsilon \text{ hits } A_L \text{ before time } T].$$

where  $y$  is such that  $X_0^\varepsilon = y$  and  $0 \leq V(y) < L$ . If  $V$  is bounded from below then  $\mathcal{D}$  contains critical points, i.e., points  $x_k$  such that  $\nabla V(x_k) = 0$  with  $\nabla V(x) \neq 0$  for  $x \neq x_k$ , and estimating  $\theta^\varepsilon(y)$  becomes a rare event simulation problem as  $\varepsilon \downarrow 0$ . This problem is closely related to the estimation of exit times and transition probabilities for dynamical systems with multiple equilibrium states, i.e., with metastability issues.

It is known in the literature, e.g., (Freidlin and Wentzell 1984, Glasserman and Kou 1995, Glasserman and Wang 1997, Dupuis and Wang 2004, Dupuis and Wang 2007, Dupuis, Spiliopoulos and Wang 2012, Dupuis, Spiliopoulos and Zhou 2013) that estimation of such probabilities has several mathematical and

computational difficulties, in that standard Monte Carlo sampling techniques lead to exponentially large relative errors as  $\varepsilon$  gets smaller. As we will discuss, there are additional difficulties particular to the situation when rest points are in the domain of interest  $\mathcal{D}$ . Hence the development of provably efficient simulation methods becomes of interest. This is in particular relevant when one deals with rough environments, e.g., (Dupuis, Spiliopoulos and Wang 2012), where analytic formulas such as Kramers-Eyring are not available.

An underlying reason for the poor performance of standard Monte Carlo methods or for poorly constructed Monte Carlo schemes when rest points are present is that trajectories tend to spend a lot of time around stable points before a successful escape, which can be shown to lead to large variance. The situation is of course more critical when the time interval is long. In this case two exponential scalings come into play—one indexed by the strength of the noise  $1/\varepsilon$  and the other by the length of time  $T$ . The majority of the current accelerated Monte Carlo literature is focused on asymptotic optimality, which is designed to address the first exponential scaling. However, the literature does not address the problems due to rest points, and as has been documented in (Dupuis, Spiliopoulos and Zhou 2013, Dupuis and Spiliopoulos 2014, Spiliopoulos 2013), this may cause seemingly-reasonable Monte Carlo methods to perform poorly even though they may be asymptotically optimal. Hence it becomes important to construct accelerated Monte-Carlo methods that have good performance even before taking the limit as  $\varepsilon \downarrow 0$ .

Based on the theory of subsolutions to the associated Hamilton-Jacobi-Bellman equation (HJB), see (Dupuis and Wang 2004, Dupuis and Wang 2007), we develop provably-efficient importance sampling methods for estimating probabilities such as  $\theta^\varepsilon(y)$ . We focus on the case that  $\mathcal{D}$  contains a stable equilibrium point of the system. The schemes constructed are easily implementable in any dimension and have provably good performance not only asymptotically, i.e., when  $\varepsilon \downarrow 0$ , but also pre-asymptotically. Moreover, the prelimit behavior does not degrade as the time horizon gets large.

The rest of the paper is organized as follows. In Section 2 we recall the related large deviations and importance sampling theory and the role of subsolutions to the associated HJB equations. In Section 3 we present the suggested simulation scheme together with a statement on its performance. In Section 4 we present numerical examples in one and two dimensions.

## 2 REVIEW OF RELATED LARGE DEVIATIONS AND IMPORTANCE SAMPLING THEORY

In this section we state assumptions and review the related large deviations and importance sampling theory. Moreover, we also review the role of subsolutions to the associated HJB equation in the performance of importance sampling methods.

We assume that  $V \in \mathcal{C}^1(\mathbb{R}^d; \mathbb{R})$  is bounded with Lipschitz continuous derivative. When rare events dominate, the standard Monte-Carlo methods perform poorly in the small noise regime. Importance sampling is an accelerated Monte-Carlo method for variance reduction. To minimize the variance (equivalently second moment) of the estimator, we simulate the system under a different probability measure chosen so that the new unbiased estimate has a smaller (and hopefully much smaller) variance.

For example to estimate a quantity such as  $\mathbb{E}_y[e^{-\frac{1}{\varepsilon}h(X_T^\varepsilon)}]$ , one generates iid samples  $X^\varepsilon(i)$  under  $\bar{\mathbb{P}}$  and uses the importance sampling estimator

$$\frac{1}{N} \sum_{i=1}^N e^{-\frac{1}{\varepsilon}h(X_T^\varepsilon(i))} \frac{d\bar{\mathbb{P}}}{d\mathbb{P}}(X^\varepsilon(i)), \quad (2)$$

where  $\bar{\mathbb{P}}$  is any probability measure which is absolutely continuous with respect to the original probability measure  $\mathbb{P}$ . If  $\bar{\mathbb{E}}$  is the expectation under  $\bar{\mathbb{P}}$ , then the goal is to choose  $\bar{\mathbb{P}}$  such that the second moment

$$\bar{\mathbb{E}}_y \left[ e^{-\frac{1}{\varepsilon}h(X_T^\varepsilon)} \frac{d\bar{\mathbb{P}}}{d\mathbb{P}}(X^\varepsilon) \right]^2$$

is minimized. The large deviations rate function for the process  $X_t^\varepsilon$  on  $[t, T]$  is given by, see (Freidlin and Wentzell 1984),  $S_{tT}(\phi) = +\infty$  for  $\phi \notin \mathcal{AC}([t, T])$  and

$$S_{tT}(\phi) = \frac{1}{2\beta} \int_t^T |\dot{\phi}_s + \nabla V(\phi_s)|^2 ds, \text{ if } \phi \in \mathcal{AC}([t, T]). \quad (3)$$

Then, under fairly general conditions, one has the heuristic approximation (made precise by large deviation theory)

$$\mathbb{E}_y[e^{-\frac{1}{\varepsilon}h(X_T^\varepsilon)}] \approx e^{-\frac{1}{\varepsilon} \inf\{S_{0T}(\phi) + h(\phi_T) : \phi, \phi_0 = y\}}, \text{ as } \varepsilon \downarrow 0.$$

Going back to importance sampling, if for a suitable process  $u : [0, T] \rightarrow \mathbb{R}^d$  we define

$$\frac{d\bar{\mathbb{P}}}{d\mathbb{P}} = e^{-\frac{1}{2\varepsilon} \int_0^T \|u(s)\|^2 ds + \frac{1}{\sqrt{\varepsilon}} \int_0^T \langle u(s), dW_s \rangle},$$

then Girsanov's theorem implies that under  $\bar{\mathbb{P}}$  the dynamics for  $X_t^\varepsilon$  take the form

$$dX_t^\varepsilon = \left[ -\nabla V(X_t^\varepsilon) + \sqrt{\beta}u(t) \right] dt + \sqrt{\varepsilon} \sqrt{\beta} d\bar{W}_t, \quad \text{with } \bar{W}_t = W_t - \frac{1}{\sqrt{\varepsilon}} \int_0^t u(\rho) d\rho \quad (4)$$

a  $\bar{\mathbb{P}}$ -Wiener process.

So, the problem becomes how to choose the control  $u(t)$  so that the second moment

$$\mathbb{E}_y \left[ e^{-\frac{1}{\varepsilon}h(X_T^\varepsilon)} \frac{d\bar{\mathbb{P}}}{d\mathbb{P}}(X^\varepsilon) \right]^2$$

(and hence the variance) is small and perhaps nearly minimized. Once the control  $u$  has been chosen, then one uses the estimator (2), where  $X^\varepsilon(i)$  are iid samples generated from  $\bar{\mathbb{P}}$  under (4).

It turns out that the selection of a good  $\bar{\mathbb{P}}$ , equivalently a good control  $u(t)$ , is related to the HJB equation with Hamiltonian

$$\mathbb{H}(x, p) = \langle p, -\nabla V(x) \rangle - \frac{1}{2} \beta \|p\|^2.$$

In particular, let us consider the HJB equation

$$U_t(t, x) + \mathbb{H}(x, \nabla U(t, x)) = 0 \text{ for } (t, x) \in [0, T] \times \mathcal{D}, \quad (5)$$

$$U(t, x) = h(x) \text{ for } t \leq T, x \in \partial\mathcal{D}, \quad U(T, x) = \infty \text{ for } x \in \mathcal{D}. \quad (6)$$

Then under mild conditions on  $h$  it can be shown, see for example (Fleming and Soner 2006, Freidlin and Wentzell 1984), that for each  $(t, x) \in [0, T] \times \mathcal{D}$

$$G(t, x) \doteq \lim_{\varepsilon \downarrow 0} \left\{ -\varepsilon \ln \mathbb{E}_{t,x} \left[ e^{-\frac{1}{\varepsilon}h(X_T^\varepsilon)} \chi_{\{\tau^\varepsilon \leq T\}} \right] \right\} = \inf_{\phi \in \Lambda(t,x)} [S_{tT}(\phi) + h(\phi(T))],$$

where  $S_{tT}(\phi)$  is given by (3), and

$$\Lambda(t, x) = \left\{ \phi \in \mathcal{C}([t, T] : \mathbb{R}^d) : \phi(t) = x, \phi(s) \in \mathcal{D} \text{ for } s \in [t, T], \phi(T) \in \partial\mathcal{D} \right\}.$$

Here,  $G$  is the unique continuous viscosity solution of (5) and (6), see (Fleming and Soner 2006).

Let us now recall the notion of subsolutions.

**Definition 1** A function  $\bar{U}(t, x) : [0, T] \times \mathbb{R}^d \mapsto \mathbb{R}$  is a classical subsolution to the HJB equation (5) if

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

The connection between subsolutions and performance of importance sampling schemes has been established in several papers, such as (Dupuis and Wang 2007, Dupuis, Spiliopoulos and Wang 2012). To be more precise, the following theorem on asymptotic optimality is known.

**Theorem 1** [Theorem 4.1 in (Dupuis, Spiliopoulos and Wang 2012)] Let  $\{X^\varepsilon, \varepsilon > 0\}$  be the unique strong solution to (1). Consider a bounded and continuous function  $h : \mathbb{R}^d \mapsto \mathbb{R}$ . Let  $\bar{U}(t, x)$  be a subsolution according to Definition 1 and define the control  $\bar{u}(t, x) = -\sqrt{\beta} \nabla_x \bar{U}(t, x)$ . Let  $X_T^{\varepsilon, \bar{u}}$  solve (8) below with  $u(t) = \bar{u}(t, X_t^\varepsilon)$ . Then

$$G(t, x) + \bar{U}(t, x) \leq \liminf_{\varepsilon \rightarrow 0} -\varepsilon \ln \bar{\mathbb{E}}_{t, x} [e^{-\frac{1}{\varepsilon} h(X_T^\varepsilon)} d\mathbb{P}/d\bar{\mathbb{P}}]^2 \leq 2G(t, x). \quad (7)$$

The subsolution property of  $\bar{U}$  implies that for every  $(t, x) \in [0, T] \times \mathcal{D}$  we have  $\bar{U}(t, y) \leq G(t, y)$ . Hence the scheme is asymptotically optimal if  $\bar{U}(t, x) = G(t, x)$  at the starting point  $(t, x)$ . Thus, for the purposes of asymptotic optimality, one should choose subsolutions such that at the point  $(0, y)$   $0 \ll \bar{U}(0, y) \leq G(0, y)$ , and value as close to  $G(0, y)$  as possible. Note that  $h$  can be chosen so that the expectation becomes a probability of escape or transition,

Let us summarize here the proposed simulation scheme. Let  $X^{\varepsilon, u}$  be the solution to the SDE

$$\begin{aligned} dX_t^{\varepsilon, u} &= \left[ -\nabla V(X_t^{\varepsilon, u}) + \sqrt{\beta} u(t) \right] dt + \sqrt{\varepsilon \beta} dW_s \\ X_0^\varepsilon &= y \end{aligned} \quad (8)$$

The proposed simulation scheme in order to estimate  $\theta^\varepsilon(y) \doteq \mathbb{E}_y \left[ e^{-\frac{1}{\varepsilon} h(X_T^\varepsilon)} \right]$  goes as follows.

1. Consider  $u(t) = \bar{u}(t, X_t^{\varepsilon, u}) = -\sqrt{\beta} \nabla_x \bar{U}(t, X_t^{\varepsilon, u})$  with  $\bar{U}$  an appropriate subsolution as defined in Definition 1.
2. Consider the estimator

$$\hat{\theta}^\varepsilon(y) \doteq \frac{1}{N} \sum_{j=1}^N \left[ e^{-\frac{1}{\varepsilon} h(X_T^{\varepsilon, u}(j))} Z_j^u \right] \quad (9)$$

where

$$Z_j^u \doteq e^{-\frac{1}{2\varepsilon} \int_0^T \|\bar{u}(t, X_t^{\varepsilon, u}(j))\|^2 dt - \frac{1}{\sqrt{\varepsilon}} \int_0^T \langle \bar{u}(t, X_t^{\varepsilon, u}(j)), dW_t(j) \rangle}$$

and  $(W(j), X^{\varepsilon, u}(j))$  is an independent sample generated from (8) with control  $u(t) = \bar{u}(t, X_t^{\varepsilon, u}(j))$ .

We conclude this section with the following remark. We address why following large deviations optimal paths (which corresponds to using the actual solution of the HJB equation and not a suitable smooth subsolution) may not be in general a good strategy for importance sampling Monte-Carlo methods.

**Remark 2** Obtaining an accurate solution, analytical or numerical, of the HJB equation (5) in high dimensions is very difficult. However, even if available the solution would not by itself be suitable for the design of importance sampling schemes for estimating the sorts of escape probabilities we consider here. The difficulty is because the solution may have a discontinuous derivative in  $x$  precisely at the rest point, and the generalized second derivative is negative definite. The exact solution to the HJB equation attempts at each time and location to force the simulated trajectory to follow a most likely large deviations path, i.e., the velocity of an extremal  $\phi_t^*$  of the variational problem for  $G(t, x)$ . However, the noise can cause trajectories to return to a neighborhood of the origin, thereby producing large likelihood ratios. As

we shall see in the numerical examples of Section 4, using  $u(t, x) = -\sqrt{\beta} \nabla G(t, x)$  to define a change of measure when  $\nabla G$  is not sufficiently smooth is not just a theoretical or technical problem, but can lead to poor performance. Hence, it is important to view this from a global point of view and not local.

The goal of this paper is to construct a subsolution in *explicit form* that is tailored to work for escape probabilities and that does not suffer from the shortcomings just mentioned.

### 3 MAIN RESULTS

In this section, we focus on estimating escape probabilities via importance sampling. We present an explicit construction of a subsolution  $\bar{U}(t, x)$ , which is both numerically implementable (i.e., it does not rely on solving differential equations), and provably efficient both asymptotically and pre-asymptotically.

We are interested in estimating via importance sampling the probability

$$\theta^\varepsilon(y) = \mathbb{P}_y [X^\varepsilon \text{ hits } A_L \text{ before time } T].$$

where  $y$  is such that  $X_0^\varepsilon = y$  and  $0 \leq V(y) < L$ . Moreover, we assume for brevity that  $V$  has a critical point at  $x_0 = 0$  such that  $V(x_0) = 0$  and  $\nabla V(x_0) = 0$  with  $\nabla V(x) \neq 0$  for all  $x \neq x_0$  such that  $0 \leq V(x) < L$ . Furthermore, for notational convenience, let us set  $\beta = 1$ .

As it can be easily verified, a global subsolution can be defined in terms of the quasipotential defined in (Freidlin and Wentzell 1984). In the gradient case that we study, the subsolution takes the explicit form

$$F_1(x) = 2(L - V(x)).$$

If we set  $\bar{U}(t, x) = F_1(x)$  and use it to define a global change of measure, we would be essentially reversing the dynamics under the new probability measure. However, it turns out that it is not optimal to use  $u(t, x) = -\nabla F_1(x)$  as a global change of measure, and the problem is mainly due to the fact that the rest point  $x_0$  is included in the domain of interest  $\mathcal{D}$ . As it is explained in (Dupuis, Spiliopoulos and Zhou 2013), the mathematical reason for the failure of the quasipotential is that the presence of the rest points induces a competition, between  $-1/\varepsilon$  and the length of the simulation  $T$  in the prelimit, which implies that one may need to go to very small values of  $\varepsilon$  in order to observe good performance. In particular, the following bound can be derived (see (Dupuis, Spiliopoulos and Zhou 2013))

$$\bar{\mathbb{E}}_y \left[ e^{-\frac{1}{\varepsilon} h(X_T^\varepsilon)} \frac{d\mathbb{P}}{d\bar{\mathbb{P}}} \right]^2 \geq e^{-\frac{1}{\varepsilon} C_1 + C_2(T-K)}.$$

for some positive constants  $C_1, C_2, K > 0$ .

This lower bound on the second moment indicates that if  $T$  is large, one may need to go to very small values of  $\varepsilon$  before the logarithmically optimal (in  $\varepsilon$ ) asymptotic performance is observed in any practice sense. As explained previously, a more detailed analysis must be done in the neighborhood of the rest point. To illustrate the point with minimal notation, let us assume for the moment that  $V(x)$  is symmetric and quadratic with  $x \in \mathbb{R}$ . In particular, we have  $V(x) = \frac{\gamma}{2}x^2$  with  $\gamma > 0$ . In this case, we can compute  $G(t, x)$  in closed form. To be specific, we have

$$\begin{aligned} G(t, x) &= \inf_{\phi_t = x, V(\phi_T) = L} \left\{ \frac{1}{2} \int_t^T \|\dot{\phi}_s + \Gamma \phi_s\|^2 ds \right\} \\ &= \inf_{\hat{x} \in V^{-1}(L)} \gamma \frac{(\hat{x} - x e^{\gamma(t-T)})^2}{1 - e^{2\gamma(t-T)}}. \end{aligned} \tag{10}$$

The point now to realize is that  $G(t, x)$  provides a good change of measure while near the rest point, whereas the quasipotential  $F_1(x)$  provides a good change of measure away from the rest point. There are a

few more issues to deal with though. The first one is that  $G(t, x)$  is discontinuous near  $t = T$ . The second one is that we need to put them together in a smooth way that will define a global subsolution.

In order to deal with the discontinuities near  $t = T$ , we shall introduce two mollification parameters  $t^*$  and  $M$  that will be appropriately chosen as functions of  $\varepsilon$  later on. Let us also fix another parameter  $\hat{L} \in (0, L]$ . In the one-dimensional case, it is easy to solve the equation  $V(\hat{x}) = \hat{L}$  and in particular we get that  $\pm\hat{x} = \pm\sqrt{\frac{2\hat{L}}{\gamma}}$ . So, instead of using  $G(t, x)$  directly, we define

$$F_2^M(t, x; \hat{x}) = \gamma \frac{(\hat{x} - xe^{\gamma(t-T)})^2}{\frac{1}{M} + 1 - e^{2\gamma(t-T)}}$$

In order now to pass smoothly between the  $F_1(x)$  and  $F_2^M(t, x; \hat{x})$  without violating the subsolution property, we use the exponential mollification (Dupuis and Wang 2007)

$$U^\delta(t, x) = -\delta \log \left( e^{-\frac{1}{\delta}F_1(x)} + e^{-\frac{1}{\delta}[F^M(t, x; \hat{x}) + F_1(\hat{x})]} + e^{-\frac{1}{\delta}[F^M(t, x; -\hat{x}) + F_1(-\hat{x})]} \right)$$

It is easy to see that  $F_1(\hat{x}) = 2(L - \hat{L})$  and if  $\hat{L} = L$ , then  $F_1(\hat{x}) = 0$ .

Then, a provably efficient importance sampling scheme can be constructed, see (Dupuis, Spiliopoulos and Zhou 2013) for proof and more details, based on the subsolution

$$\bar{U}^\delta(t, x) = \begin{cases} F_1(x), & t > T - t^* \\ U^\delta(t, x), & t \leq T - t^* \end{cases}$$

It turns out that  $\bar{U}^\delta(t, x)$  is a global smooth subsolution which has provably good performance both pre-asymptotically and asymptotically. The precise optimality bound and its proof guide the choice of the parameters  $\delta, t^*, M$  and  $\hat{L}$ . Before presenting the theoretical result, let us elaborate a little bit more on the exponential mollification. The control used for the simulation is

$$u(t, x) = -\nabla \bar{U}^\delta(t, x) = - \left[ \rho_1(t, x; \delta) \nabla F_1(x) + \rho_2^{M,+}(t, x; \delta) \nabla F_2^M(t, x, \hat{x}) + \rho_2^{M,-}(t, x; \delta) \nabla F_2^M(t, x, -\hat{x}) \right]$$

where the weights  $\rho_i$  are

$$\rho_1(t, x; \delta) = \frac{e^{-\frac{1}{\delta}F_1(x)}}{e^{-\frac{1}{\delta}F_2^M(t, x; \hat{x})} + e^{-\frac{1}{\delta}F_2^M(t, x; -\hat{x})} + e^{-\frac{1}{\delta}F_1(x)}}$$

and

$$\rho_2^{M,\pm}(t, x; \delta) = \frac{e^{-\frac{1}{\delta}F_2^M(t, x, \pm\hat{x})}}{e^{-\frac{1}{\delta}F_2^M(t, x; \hat{x})} + e^{-\frac{1}{\delta}F_2^M(t, x; -\hat{x})} + e^{-\frac{1}{\delta}F_1(x)}}$$

Note that  $\rho_1(t, x; \delta) + \rho_2^{M,+}(t, x; \delta) + \rho_2^{M,-}(t, x; \delta) = 1$ . A further analysis of this scheme shows that the weights are such that the corresponding subsolutions dominate in their corresponding area of intended dominance. To be more precise, for  $\delta$  small enough,  $\rho_1(t, x; \delta)$  turns out to be close to zero when  $x$  is close to the rest point  $x_0 = 0$  (i.e.,  $F_2^M$  will dominate there), whereas  $\rho_2^{M,\pm}(t, x; \delta)$  turn out to be close to zero when  $x$  is away from the rest point  $x_0 = 0$  (i.e.,  $F_1$  will dominate there).

The precise theorem characterizing the performance of such a scheme is as follows (See Theorem 4.6 of (Dupuis, Spiliopoulos and Zhou 2013) for the proof).

**Theorem 3** Let us assume  $d = 1$ ,  $\delta = 2\varepsilon$ ,  $\theta \in (2, 4)$ ,  $\eta \in \left[ \varepsilon / \left( \varepsilon + \theta \frac{2\hat{L}}{\beta M} \right), 1/4 \right]$  and that  $\eta \frac{\hat{L}}{M} > 12\varepsilon$ .

Consider the scheme constructed in Section 2 with  $\bar{u}(t, x) = -\sqrt{\beta} D \bar{U}^\delta(t, x)$ . Then up to an exponentially negligible term, we have

$$-\varepsilon \log \bar{\mathbb{E}}_y [e^{-\frac{1}{\varepsilon} h(X_T^\varepsilon)} d\mathbb{P} / d\bar{\mathbb{P}}]^2 \geq 2I_1(\varepsilon, \eta, T, \hat{L}, M) 1_{\{T \geq t^*\}} + 2I_2(\varepsilon, T) 1_{\{T < t^*\}},$$

where

$$I_1(\varepsilon, \eta, T, \hat{x}, M) = (1 - \eta) \bar{U}^\delta(0, 0) + \left( \log \left[ \frac{1}{\hat{x}} \left( \hat{x} M^{-1/2} / 2 - \sqrt{4\varepsilon\beta^2/\gamma\eta} \right) \right] \wedge 0 \right) \varepsilon.$$

$$I_2(\varepsilon, T) = 2L - \gamma T \varepsilon.$$

Moreover,

$$\bar{U}^\delta(0, 0) \geq \frac{2\hat{L}/\beta}{\frac{1}{M} + 1 - e^{-2\gamma T}} + 2\beta^{-1}(L - \hat{L}) - \delta \log 3.$$

Several remarks are now relevant.

**Remark 4** (Choice of free parameters) Note that even though the bound provided by the theorem appears to be in a slightly complicated form, it does not degrade as  $T \rightarrow \infty$ . This is also reflected in the simulation data presented in Section 4. An analysis of the bound and of its proof, see (Dupuis, Spiliopoulos and Zhou 2013), reveals appropriate choices of  $\delta, M, t^*$  as functions of  $\varepsilon$ . In particular we can set  $\delta = 2\varepsilon$ ,  $t^* \geq -\frac{2}{\gamma} \log \frac{1}{M}$  and  $M = \hat{L}\varepsilon^{-2\kappa}$  with  $\kappa \in (0, 1/2)$ . It turns out that the choice of  $\hat{L}$  is not that important and one can simply pick  $\hat{L} = L/2$  or even  $\hat{L} = L$ . Then  $M = M(\varepsilon) \uparrow \infty$  as  $\varepsilon \downarrow 0$ , and uniformly in  $T$  the decay rate (i.e., lower bound of the algorithm as  $\varepsilon \downarrow 0$ ) is

$$2\beta^{-1}L + 2\beta^{-1}\hat{L} \frac{e^{-2\lambda\hat{T}}}{1 - e^{-2\lambda\hat{T}}}.$$

For the convenience of the reader and for purposes of an easy reference we present in Table 1 the suggested values for  $(\delta, \hat{L}, M, t^*)$ , given the value of the strength of the noise  $\varepsilon > 0$ .

parameter	$\delta$	$\hat{L}$	$M$	$t^*$
values	$2\varepsilon$	$O(1)$ or $\varepsilon^{2\lambda}$ with $\lambda < \kappa$	$\max\{\frac{\hat{L}}{\varepsilon^{2\kappa}}, 4\}$ with $\kappa \in (0, 1/2)$	$-\frac{2}{\gamma} \log \frac{1}{M}$

Table 1: Parameter values for the algorithm based on a given value of  $\varepsilon > 0$ .

**Remark 5** (General non-quadratic potential) In the case of nonlinear dynamics, one can approximate the potential near the rest point by its quadratic approximation and thus define  $F_2^M$ . Because of the design of the exponential mollification, the algorithm uses  $F_1(x)$  away from the neighborhood of the rest point, which is well defined independently of whether one has quadratic potential or not. Near the rest point, the algorithm uses  $F_2^M$  which is a good mollified approximation for  $G(t, x)$  when  $x$  is close to  $x_0 = 0$ . It turns out that the localization error induced by such an approximation is exponentially negligible as  $\varepsilon \downarrow 0$ , see Theorem 5.7 in (Dupuis, Spiliopoulos and Zhou 2013).

**Remark 6** (Multidimensional case) Away from the rest point mostly  $F_1(x)$  is used, which is given in closed form in the gradient case for any dimension. Near the rest point, the mollified version to the quadratic approximation to  $G(t, x)$  is used. Hence, even though the details are more elaborate, the algorithm can be extended to higher dimensions with complexity that is linear in the dimension. This program is carried out in detail in (Dupuis and Spiliopoulos 2014) and we refer the interested reader there for details.

## 4 NUMERICAL SIMULATIONS

In this section we present one and two dimensional simulation results to illustrate the points made previously. For convenience, we set  $\beta = 1$ . In all examples the number of trajectories simulated is  $N = 10^7$ . The performance measure is relative error per sample, defined by

$$\rho^\varepsilon \doteq \rho(\hat{\theta}^\varepsilon) = \frac{\sqrt{\text{Var}(\hat{\theta}^\varepsilon)}}{\theta^\varepsilon} = \frac{1}{\sqrt{N}} \sqrt{\frac{\bar{\mathbb{E}}_0 \left[ 1_{\{X^\varepsilon \text{ exits } [-1, 1] \text{ before time } T\}} \left( \frac{dP}{d\bar{P}_1}(X^\varepsilon) \right)^2 \right]}{(\mathbb{E}_0 [1_{\{X^\varepsilon \text{ exits } [-1, 1] \text{ before time } T\}}])^2}} - 1. \quad (11)$$

The estimated relative error indicates the number of samples required for a given accuracy.

We first start with a very simple one dimensional example by taking  $V(x) = \frac{1}{2}x^2$ . Assume that the initial point of the process is the equilibrium point, i.e.,  $y = x_0 = 0$ . Also let  $L = 1$ . Thus the problem of interest is to compute the probability of exit from the interval  $[-1, 1]$  by some given time  $T$ , when starting from  $y = 0$  and with some level of noise  $\varepsilon$ .

We first present in Table 2 estimated values of the probability of interest for different combinations of  $(\varepsilon, T)$ . The values have been obtained using the suggested algorithm. We note that the values range from probabilities of order  $O(10^{-1})$  to probabilities of the order of  $O(10^{-10})$ .

$\varepsilon \mid T$	1.5	2.5	5	7	10	14	18	23
0.16	$2.2e-03$	$6.6e-03$	$1.8e-02$	$2.7e-02$	$4.0e-02$	$5.7e-02$	$7.4e-02$	$9.5e-02$
0.13	$5.1e-04$	$1.6e-03$	$4.6e-03$	$6.9e-03$	$1.1e-02$	$1.5e-02$	$2.0e-02$	$2.6e-02$
0.11	$1.1e-04$	$3.9e-04$	$1.2e-03$	$1.8e-03$	$2.8e-03$	$4.1e-03$	$5.4e-03$	$7.0e-03$
0.09	$1.3e-05$	$5.2e-05$	$1.7e-04$	$2.6e-04$	$4.1e-04$	$5.9e-04$	$7.8e-04$	$1.0e-03$
0.07	$4.3e-07$	$2.2e-06$	$7.6e-06$	$1.2e-05$	$1.9e-05$	$2.8e-05$	$3.7e-05$	$4.8e-05$
0.05	$9.7e-10$	$6.9e-09$	$2.8e-08$	$4.4e-08$	$7.0e-08$	$1.1e-07$	$1.4e-07$	$1.8e-07$

Table 2: Estimated values for different pairs  $(\varepsilon, T)$ .  $\hat{L} = 1$ .

In Table 3 we present relative errors per sample when the subsolution based on just the quasipotential is used as a global change of measure, i.e., when the control is  $u(t, x) = -F'_1(x) = 2x$ . Note that for  $T$  small and  $T$  large the performance of the algorithm is bad uniformly in the size of the noise  $\varepsilon$ . Dashes indicate that no single trajectory was successful for the particular pair  $(\varepsilon, T)$ .

$\varepsilon \mid T$	0.25	0.5	1	1.5	2.5	10	14	18
0.16	253	10	2	1	1	10	48	139
0.13	748	16	3	1	1	9	48	378
0.11	1594	26	3	1	1	10	42	272
0.09	—	49	4	2	1	9	43	357
0.07	—	127	5	2	1	8	47	251
0.05	—	714	8	2	1	8	42	145

Table 3: Using the subsolution based on quasipotential throughout. Relative errors per sample for different pairs  $(\varepsilon, T)$ .

In Table 4, we present relative errors per sample, where the algorithm being used is based on using the actual solution to the HJB equation as a global change of measure. Namely, we use the change of measure that is based on the control  $u(t, x) = -G_x(t, x)$ , where  $G(t, x) = \inf_{\hat{x} \in V^{-1}(1)} \frac{(\hat{x} - xe^{(t-T)})^2}{1 - e^{2(t-T)}}$ . We notice that even though the behavior is good when  $T$  is small, when  $T$  gets large the algorithm deteriorates uniformly in the value of  $\varepsilon$ .

$\varepsilon \mid T$	0.25	0.5	1	1.5	2.5	10	14	18
0.16	3	3	1	1	1	10	55	265
0.13	4	2	1	1	1	9	51	394
0.11	3	2	2	1	1	9	44	177
0.09	3	3	2	1	1	9	43	314
0.07	3	3	2	2	1	9	67	520
0.05	4	2	2	2	1	8	50	278

Table 4: Using the subsolution based on the explicit solution to the  $\varepsilon = 0$  HJB equation. Relative errors per sample for different pairs  $(\varepsilon, T)$ .

Then, in Table 5 we investigate the numerical performance of the suggested algorithm which correctly balances out between  $F_1(x)$  and  $F_2^M(t,x)$  as described previously. It is clear that the improvement is noticeable and it is uniform across all values of  $\varepsilon$  and  $T$ .

$\varepsilon \mid T$	2.5	5	7	10	14	18	23
0.16	1.5	1.1	0.8	0.7	0.7	0.7	0.9
0.13	1.7	1.2	1.0	0.8	0.7	0.7	0.8
0.11	1.8	1.4	1.2	0.9	0.8	0.7	0.8
0.09	2.0	1.6	1.3	1.1	0.9	0.8	0.8
0.07	2.2	1.9	1.6	1.3	1.1	1.0	0.9
0.05	2.4	2.5	2.1	1.7	1.5	1.3	1.1

Table 5: Relative errors per sample for different pairs  $(\varepsilon, T)$  when exponential mollification is being used.  $M = \frac{2}{\sqrt{\varepsilon}}$  and  $\hat{L} = 1$ .

Next we consider a two-dimensional nonlinear example. We consider the potential function

$$V(x, y) = (x^2 - 1)^2 + \frac{1}{2}(3y + x^2 - 1)^2,$$

with the exit level  $L = 1/2$ . This potential has two stable points at  $(-1, 0)$  and at  $(1, 0)$ . We set the initial point of the process to be one of them, i.e.,  $y = x_0 = (-1, 0)$ . By linearizing the dynamics around the equilibrium point, we can apply the algorithm presented. The estimated probabilities of exit by time  $T$  from the level set of the potential  $L = 1/2$  are given in the Table 6.

$\varepsilon \mid T$	1	2	3	4	6	8	10	12	14
0.06	9.63e-06	2.23e-05	3.51e-05	4.79e-05	7.33e-05	9.89e-05	1.24e-04	1.51e-04	1.94e-04
0.05	3.93e-07	9.34e-07	1.47e-06	2.00e-06	3.08e-06	4.17e-06	5.27e-06	6.36e-06	7.39e-06
0.04	3.14e-09	7.57e-09	1.20e-08	1.64e-08	2.52e-08	3.41e-08	4.31e-08	5.18e-08	6.06e-08
0.03	9.38e-13	2.30e-12	3.68e-12	5.04e-12	7.79e-12	1.06e-11	1.32e-11	1.61e-11	1.88e-11
0.02	7.34e-20	1.84e-19	2.99e-19	4.13e-19	6.44e-19	8.74e-19	1.10e-19	1.32e-18	1.56e-18

Table 6: Estimated probability values for different pairs  $(\varepsilon, T)$  using the exponential mollification with  $M = \hat{L}\varepsilon^{-0.5}$  and  $\hat{L} = L$ .

Relative errors when the quasipotential is used as a global change of measure throughout are given in Table 7. Large relative errors occurs for all pairs of values  $(\varepsilon, T)$ .

$\varepsilon \mid T$	1	2	3	4	6	8	10	12	14
0.06	415	1472	249	209	164	320	354	451	388
0.05	196	590	105	127	366	181	186	155	293
0.04	309	918	157	181	339	313	246	207	109
0.03	144	614	159	1302	315	1438	126	368	132
0.02	368	216	493	338	110	639	247	798	966

Table 7: Estimated relative errors per sample for different pairs  $(\varepsilon, T)$  using quasipotential.

Lastly, we present in Table 8 relative errors per sample for different pairs of  $(\varepsilon, T)$  using the exponential mollification. It is clear that the performance has been dramatically improved. Note that some approximation error due to localization is expected. Notice also that as  $T$  gets large and  $\varepsilon$  gets small the performance of the algorithm gets better, as theory indicates, even though the performance is good even for larger values of  $\varepsilon$ , i.e., non-asymptotically.

We conclude by mentioning that the large relative error values of Tables 3, 4 and 7, do not just imply that one would need so many more times the number of trajectories to achieve the same level of accuracy given by the exponential mollification. These large values imply that the simulation data are all-together unreliable, even though those data was generated using asymptotically nearly optimal schemes.

$\varepsilon \mid T$	1	2	3	4	6	8	10	12	14
0.06	3	3	2	3	5	7	8	11	13
0.05	4	3	3	3	3	4	5	6	8
0.04	7	4	4	3	3	3	3	4	4
0.03	12	7	5	5	4	3	3	3	3
0.02	23	16	14	12	12	9	8	7	5

Table 8: Estimated relative errors per sample for different pairs  $(\varepsilon, T)$  using the exponential mollification with  $M = \hat{L}\varepsilon^{-0.5}$  and  $\hat{L} = L$ .

## ACKNOWLEDGMENTS

Research of P.D. supported in part by the National Science Foundation (DMS-1317199) and the Air Force Office of Scientific Research (FA9550-12-1-0399). K.S. was partially supported by the National Science Foundation, NSF-DMS 1312124.

## REFERENCES

- Dupuis P., Spiliopoulos K., and Wang H. 2012. “Importance sampling for multiscale diffusions”, *SIAM J. Multiscale Model. and Simul.* 10:1–27.
- Dupuis P., Spiliopoulos K., and Zhou X., 2014. “Escaping from an Attractor: Importance Sampling and Rest Points, Part I”, *Annals of Applied Probability*, accepted subject to minor revision.
- Dupuis P., and Spiliopoulos K., 2014. “Escaping from an Attractor: Importance Sampling and Rest Points, Part II”, *preprint*.
- Dupuis P., and Wang H. 2004. “Importance sampling, large deviations and differential games”. *Stochastics and Stochastics Reports* 76: 481-508.
- Dupuis P., and Wang H. 2007. “Subsolutions of an Isaacs equation and efficient schemes of importance sampling”. *Mathematics of Operations Research*, 32 (3): 723-757.
- Fleming W.H., and Soner H.M. 2006. *Controlled Markov Processes and Viscosity Solutions*, Springer, 2nd Ed.
- Freidlin M. I. and Wentzell A. D. 1984. *Random Perturbations of Dynamical Systems*, Springer-Verlag, New York, 2nd Ed.
- Glasserman P., and Kou S., 1995. “Analysis of an importance sampling estimator for tandem queues” *ACM Trans. Modeling Comp. Simulation*, 4:22–42.
- Glasserman P., and Wang Y., 1997, “Counter examples in importance sampling for large deviations probabilities” *Ann. Appl. Prob.*, 7:731–746.
- Spiliopoulos K., 2013. “Non-asymptotic performance analysis of importance sampling schemes for small noise diffusions”, *submitted*.

## 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. His email is dupuis@dam.brown.edu.

**KONSTANTINOS SPILIOPOULOS** is an Assistant Professor at the Department of Mathematics and Statistics at Boston 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. His e-mail is kspiliop@math.bu.edu.