

Large Deviations and Importance Sampling for Systems of Slow-Fast Motion

Konstantinos Spiliopoulos

Published online: 2 October 2012
© Springer Science+Business Media New York 2012

Abstract In this paper we develop the large deviations principle and a rigorous mathematical framework for asymptotically efficient importance sampling schemes for general, fully dependent systems of stochastic differential equations of slow and fast motion with small noise in the slow component. We assume periodicity with respect to the fast component. Depending on the interaction of the fast scale with the smallness of the noise, we get different behavior. We examine how one range of interaction differs from the other one both for the large deviations and for the importance sampling. We use the large deviations results to identify asymptotically optimal importance sampling schemes in each case. Standard Monte Carlo schemes perform poorly in the small noise limit. In the presence of multiscale aspects one faces additional difficulties and straightforward adaptation of importance sampling schemes for standard small noise diffusions will not produce efficient schemes. It turns out that one has to consider the so called cell problem from the homogenization theory for Hamilton-Jacobi-Bellman equations in order to guarantee asymptotic optimality. We use stochastic control arguments.

Keywords Importance sampling · Monte Carlo · Large deviations · Homogenization · Multiscale · Slow-fast motion

K. Spiliopoulos
Division of Applied Mathematics, Brown University, Providence, RI 02912, USA
e-mail: kspiliop@dam.brown.edu

K. Spiliopoulos (✉)
Department of Mathematics and Statistics, Boston University, Boston, MA 02215, USA
e-mail: kspiliop@math.bu.edu

1 Introduction

Let us consider the $m + (d - m)$ dimensional process $(X^\epsilon, Y^\epsilon) = \{(X^\epsilon(s), Y^\epsilon(s)), 0 \leq s \leq T\}$ satisfying the system of stochastic differential equations (SDE's)

$$\begin{aligned}
 dX^\epsilon(s) &= \left[\frac{\epsilon}{\delta} b(X^\epsilon(s), Y^\epsilon(s)) + c(X^\epsilon(s), Y^\epsilon(s)) \right] ds \\
 &\quad + \sqrt{\epsilon} \sigma(X^\epsilon(s), Y^\epsilon(s)) dW(s), \\
 dY^\epsilon(s) &= \frac{1}{\delta} \left[\frac{\epsilon}{\delta} f(X^\epsilon(s), Y^\epsilon(s)) + g(X^\epsilon(s), Y^\epsilon(s)) \right] ds \\
 &\quad + \frac{\sqrt{\epsilon}}{\delta} [\tau_1(X^\epsilon(s), Y^\epsilon(s)) dW(s) \\
 &\quad + \tau_2(X^\epsilon(s), Y^\epsilon(s)) dB(s)], \\
 X^\epsilon(0) &= x_0, \quad Y^\epsilon(0) = y_0
 \end{aligned}
 \tag{1.1}$$

where $\delta = \delta(\epsilon) \downarrow 0$ as $\epsilon \downarrow 0$ and $(W(s), B(s))$ is a 2κ -dimensional standard Wiener process. The functions $b(x, y), c(x, y), \sigma(x, y), f(x, y), g(x, y), \tau_1(x, y)$ and $\tau_2(x, y)$ are assumed to be sufficiently smooth (see Condition 2.1) and periodic with period λ in every direction with respect to the second variable.

One can interpret the system of (1.1) as a system of slow and fast motion with X^ϵ playing the role of the slow motion and Y^ϵ playing the role of the fast motion. The goal of this paper is to provide a large deviations analysis of (1.1) that allows to rigorously develop the importance sampling theory for estimation of functionals such as

$$\theta(\epsilon) \doteq \mathbb{E} \left[e^{-\frac{1}{\epsilon} h(X^\epsilon(T))} \mid X^\epsilon(0) = x_0, Y^\epsilon(0) = y_0 \right].
 \tag{1.2}$$

Importance sampling is a variance reduction technique in Monte Carlo simulation. As it is well known, standard Monte Carlo sampling techniques perform very poorly in that the relative errors under a fixed computational effort grow rapidly as the event becomes more and more rare. Estimating rare event probabilities in the context of slow-fast systems presents extra difficulties due to the underlying fast motion and its interaction with the intensity of the noise ϵ .

Depending on the order that ϵ, δ go to zero, we have three different regimes of interaction:

$$\lim_{\epsilon \downarrow 0} \frac{\epsilon}{\delta} = \begin{cases} \infty & \text{Regime 1,} \\ \gamma \in (0, \infty) & \text{Regime 2,} \\ 0 & \text{Regime 3.} \end{cases}
 \tag{1.3}$$

If δ goes to zero faster than ϵ (Regime 1) then homogenization occurs first, whereas if ϵ goes to zero faster than δ (Regime 3) then large deviations theory tells how quickly (1.1) converges to the averaged deterministic ODE given by setting ϵ equal to zero. If the two parameters go to zero together then one has an intermediate situation (Regime 2).

The study of rare events in the multiscale context is a difficult problem due to the presence of the underlying fast motion. The first necessary step is to develop the associated large deviations theory. Using weak convergence arguments the authors in [12] prove the large deviations principle for the special case $f = b$, $g = c$, $\tau_1 = \sigma$ and $\tau_2 = 0$. We extend the results of [12] to the current general setup. Then, using the large deviation results and stochastic control arguments we construct asymptotically optimal importance sampling schemes with rigorous bounds on performance. The construction is based on subsolutions for an associated Hamilton-Jacobi-Bellman (HJB) equation as in [13, 15]. The situation here is complicated due to the presence of the fast motion. It turns out that changes of measure that are implied by the homogenized system do not lead to efficient importance sampling schemes. The standard arguments have to be modified taking into account the solution to the related “cell problem” which is different for each regime. This is also tightly related to the homogenization theory for HJB equations. A control in full feedback form, i.e., a function of both the slow variable X^ϵ and the fast variable Y^ϵ , is used to construct dynamic importance sampling schemes with precise asymptotic performance bounds. The control involves both the solution to the appropriate homogenized HJB equation and to its corresponding cell problem.

The novelty of this work lies in developing (a) the large deviations principle and (b) a general and rigorous mathematical framework for the study of importance sampling schemes for systems of slow-fast motion as in (1.1) for all three regimes of interaction, (1.3). Multiscale stochastic control problems and related large deviations problems have been studied elsewhere as well under various assumptions and dependencies of the coefficients of the system on the slow and fast motion, see [3, 7, 12, 18, 19, 24, 25, 28, 30, 33, 34]. The papers [18, 19, 25, 33, 34] address the large deviations principle for Regime 2 for special cases of dependence of the coefficients on (x, y) . With the exception of [12, 25], they express it through a Legendre-Fenchel transform of the limit of the normalized logarithm of an exponential moment or of the first eigenvalue of an associated operator. Here we provide an explicit characterization of the action functional. Also, the large deviations arguments in the aforementioned papers do not cover the full nonlinear case that we study here and do not seem to provide insights into how to construct asymptotically efficient importance sampling schemes. Some related importance sampling results on this problem have been recently obtained in [13]. There the authors study the special case of $f = b$, $g = c$, $\tau_1 = \sigma$ and $\tau_2 = 0$ for Regime 1 only and provide simulation studies for that particular case as well. It is also demonstrated there that straightforward adaptation of importance sampling schemes for standard diffusion processes (without the multiscale aspect) will have poor results in the multiscale setting. This translates in that one needs to consider the solution to the cell problem in problems with multiple scales in order to guarantee good asymptotic performance. The treatment of the general case, that is the content of the current paper, requires additional considerations. In particular, the identification of the optimal control and of the associated subsolutions and cell problems are more involved here even for Regime 1. The case of Regimes 2 and 3 is studied in this paper for the first time. This work is closely related to the homogenization theory of HJB equations, e.g., [1, 2, 9, 17, 21, 27], see Sect. 5.

We note here that one may possibly be able to relax the periodicity assumption both for the large deviations and for the importance sampling. In particular, in the

case of Regime 1 using the results and methodology of [12, 30] and of the present paper, we can probably prove an analogous result when the fast variable takes values in \mathbb{R}^{d-m} instead of the torus. Of course, one would need to impose the appropriate recurrence conditions for the fast motion. In the case of Regime 2, the extension to the whole space with full dependence of the coefficients on (x, y) is more involved. However, it seems plausible that the methods of the current paper can be combined with those of [5, 12, 22] to weaken the periodicity assumption for Regime 2 as well. This will be addressed elsewhere.

The need to simulate rare events occurs in many application areas including telecommunication, finance, insurance and chemistry. We present some examples in Sect. 6. A model of interest in chemical physics and chemistry is the first order Langevin equation in a rough potential, e.g. [14, 26, 29, 32, 35]. This is a special case of the system (1.1) with $f = b = -\nabla Q(y)$, $g = c = -\nabla V(x)$, $\tau_1 = \sigma = \text{constant}$ and $\tau_2 = 0$ and is discussed in Sect. 6.1. Another example, discussed in Sect. 6.2, is related to short time asymptotics of a process that depends on another fast mean reverting process.

The rest of the paper is organized as follows. In Sect. 2 we introduce necessary notation and our assumptions. Section 3 is devoted to the related large deviations theory. In Sect. 4 we develop the importance sampling theory for all three possible regimes of interaction that guarantees asymptotic optimality. In Sect. 5 we discuss the connection of the importance sampling theory with the homogenization of HJB equations. We conclude with Sect. 6 where we examine how our results look like in some special cases of interest.

2 Notation and Assumptions

In this section we establish some notation and lay out our main assumptions. Let us assume a filtered probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ equipped with a filtration \mathfrak{F}_t that satisfies the usual conditions, namely, \mathfrak{F}_t is right continuous and \mathfrak{F}_0 contains all \mathbb{P} -negligible sets.

The main assumption for the coefficients of (1.1) is as follows.

Condition 2.1

- (i) *The functions $b(x, y)$, $c(x, y)$, $\sigma(x, y)$, $f(x, y)$, $g(x, y)$, $\tau_1(x, y)$ and $\tau_2(x, y)$ are bounded in both variables and periodic with period λ in the second variable in each direction. We additionally assume that they are $C^1(\mathbb{R}^{d-m})$ in y and $C^2(\mathbb{R}^m)$ in x with all partial derivatives continuous and globally bounded in x and y .*
- (ii) *The diffusion matrices $\sigma\sigma^T$ and $\tau_1\tau_1^T + \tau_2\tau_2^T$ are uniformly nondegenerate.*

Under Condition 2.1 the system (1.1) has a unique strong solution. The smoothness assumptions are stronger than necessary, but they guarantee smoothness and boundedness of the associated cell problems that will appear in the development of

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

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

Let $\mathcal{Y} = \mathbb{T}^{d-m}$ be the $(d - m)$ -dimensional torus. This is the state space of the fast motion. For the purposes of consistency with the related literature we use similar notation as in [12, 13] with the appropriate modifications in order to cover the more general set-up that we treat here.

Under Regime 1, we also impose the following condition.

Condition 2.2 Let $F \in \mathcal{C}^2(\mathcal{Y}; \mathbb{R})$ and consider the operator

$$\mathcal{L}_x^1 F(y) = f(x, y) \cdot \nabla_y F(y) + \frac{1}{2}(\tau_1 \tau_1^T + \tau_2 \tau_2^T)(x, y) : \nabla_y \nabla_y F(y)$$

equipped with periodic boundary conditions in y . Under Regime 1, we assume the centering condition (see [6]):

$$\int_{\mathcal{Y}} b(x, y) \mu(dy|x) = 0,$$

where $\mu(dy|x)$ is the unique invariant measure corresponding to the operator \mathcal{L}_x^1 .

Under Conditions 2.1 and 2.2, Theorem 3.3.4 in [6] guarantees that for each $\ell \in \{1, \dots, m\}$ there is a unique, twice differentiable, with all partial derivatives up to second order bounded, λ -periodic in each direction function $\chi_\ell(x, y)$ that satisfies the cell problem:

$$\mathcal{L}_x^1 \chi_\ell(x, y) = -b_\ell(x, y), \quad \int_{\mathcal{Y}} \chi_\ell(x, y) \mu(dy|x) = 0. \tag{2.1}$$

We write $\chi = (\chi_1, \dots, \chi_m)$.

Let us denote $\mathcal{Z} = \mathbb{R}^k$. This will be the space in which the control processes that will appear in the next sections take values.

Definition 2.3 For $(x, y, z_1, z_2) \in \mathbb{R}^m \times \mathcal{Y} \times \mathcal{Z} \times \mathcal{Z}$ and for Regime $i = 1, 2, 3$ defined in (1.3) we define the operators $\mathcal{L}_{z_1, z_2, x}^i$. For $i = 1, 2$ we let $\mathcal{D}(\mathcal{L}_{z, x}^i) = \mathcal{C}^2(\mathcal{Y})$ and for $i = 3$, $\mathcal{D}(\mathcal{L}_{z, x}^3) = \mathcal{C}^1(\mathcal{Y})$. For $F \in \mathcal{D}(\mathcal{L}_{z, x}^i)$ define

$$\begin{aligned} \mathcal{L}_x^1 F(y) &= f(x, y) \cdot \nabla_y F(y) + \frac{1}{2}(\tau_1 \tau_1^T + \tau_2 \tau_2^T)(x, y) : \nabla_y \nabla_y F(y), \\ \mathcal{L}_{z_1, z_2, x}^2 F(y) &= [\gamma f(x, y) + g(x, y) + \tau_1(x, y)z_1 + \tau_2(x, y)z_2] \cdot \nabla_y F(y) \\ &\quad + \gamma \frac{1}{2}(\tau_1 \tau_1^T + \tau_2 \tau_2^T)(x, y) : \nabla_y \nabla_y F(y), \\ \mathcal{L}_{z_1, z_2, x}^3 F(y) &= [g(x, y) + \tau_1(x, y)z_1 + \tau_2(x, y)z_2] \cdot \nabla_y F(y). \end{aligned}$$

Definition 2.4 For $(x, y, z_1, z_2) \in \mathbb{R}^m \times \mathcal{Y} \times \mathcal{Z} \times \mathcal{Z}$ and for Regime $i = 1, 2, 3$ defined in (1.3) we define the functions $\lambda_i(x, y, z_1, z_2) : \mathbb{R}^m \times \mathcal{Y} \times \mathcal{Z} \times \mathcal{Z} \rightarrow \mathbb{R}^m$ by

$$\begin{aligned} \lambda_1(x, y, z_1, z_2) &= c(x, y) + \frac{\partial \chi}{\partial y}(x, y)g(x, y) + \sigma(x, y)z_1 \\ &\quad + \frac{\partial \chi}{\partial y}(x, y)(\tau_1(x, y)z_1 + \tau_2(x, y)z_2), \\ \lambda_2(x, y, z_1, z_2) &= \gamma b(x, y) + c(x, y) + \sigma(x, y)z_1, \\ \lambda_3(x, y, z_1, z_2) &= c(x, y) + \sigma(x, y)z_1, \end{aligned}$$

where $\chi = (\chi_1, \dots, \chi_m)$ is defined by (2.1).

For a Polish space \mathcal{S} , let $\mathcal{P}(\mathcal{S})$ be the space of probability measures on \mathcal{S} . Next we recall the notion of viability as defined in [12].

Definition 2.5 A pair $(\psi, P) \in \mathcal{C}([0, T]; \mathbb{R}^m) \times \mathcal{P}(\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, T])$ will be called viable with respect to (λ, \mathcal{L}) and write $(\psi, P) \in \mathcal{V}_{(\lambda, \mathcal{L})}$, if the following hold:

- The function ψ_t is absolutely continuous.
- The measure P is square integrable in the sense that $\int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, T]} \|z\|^2 \times P(dz_1 dz_2 dy ds) < \infty$.
- For all $t \in [0, T]$

$$\psi_t = x_0 + \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, t]} \lambda(\psi_s, y, z_1, z_2)P(dz_1 dz_2 dy ds). \tag{2.2}$$

- For all $t \in [0, T]$ and for every $f \in \mathcal{D}(\mathcal{L})$

$$\int_0^t \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y}} \mathcal{L}_{z_1, z_2, \psi_s} f(y)P(dz_1 dz_2 dy ds) = 0. \tag{2.3}$$

- For all $t \in [0, T]$

$$P(\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, t]) = t. \tag{2.4}$$

Notice that (2.4) implies that the last marginal of P is Lebesgue measure, and hence P can be decomposed in the form $P(dz_1 dz_2 dy dt) = P_t(dz_1 dz_2 dy) dt$.

3 Large Deviations Principle

The authors in [12] establish the large deviations principle related to (1.1) in the special case of $f = b, g = c, \tau_1 = \sigma$ and $\tau_2 = 0$. We extend the results of [12] to the current general setup. A uniform approach to the large deviations problem for (1.1) is presented, allowing to essentially treat all three regimes with the same general strategy, even though the technical details might be different from regime to regime. Moreover, in the course of the proof of the large deviations lower bound, we need to

construct a nearly optimal control that attains the large deviations bound. As we will see in Sect. 4, this control can guide the construction of efficient importance sampling for the estimation of quantities such as (1.2).

Essentially, in each regime, the action functional is given by the infimization of a quadratic functional, where the infimum is determined by the averaging of an appropriate controlled version of the limiting slow motion with respect to the corresponding fast motion. Both the limiting slow motion and the fast motion with respect to which the averaging is being done, differ from regime to regime. This is related to the notion of viability from Definition 2.5 where the viable pairs (λ, \mathcal{L}) are obtained from Definitions 2.3 and 2.4 for each regime. What differs from the special case considered in [12] is the form of the appropriate viable pair in each case. We present this characterization below.

In preparation for stating the main large deviations results, we recall the concept of a Laplace principle.

Definition 3.1 Let $\{X^\epsilon, \epsilon > 0\}$ be a family of random variables taking values in a Polish space \mathcal{S} and let I be a rate function on \mathcal{S} . We say that $\{X^\epsilon, \epsilon > 0\}$ satisfies the Laplace principle with rate function I if for every bounded and continuous function $h : \mathcal{S} \rightarrow \mathbb{R}$

$$\lim_{\epsilon \downarrow 0} -\epsilon \ln \mathbb{E} \left[\exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] = \inf_{x \in \mathcal{S}} [I(x) + h(x)]. \tag{3.1}$$

If the level sets of the rate function (equivalently action functional) are compact, then the Laplace principle is equivalent to the corresponding large deviations principle with the same rate function (Theorems 2.2.1 and 2.2.3 in [11]).

The derivation of the large deviations and importance sampling results are based on a variational representation for functionals of Wiener process derived in [8] that allows to rewrite the prelimit left hand side of (3.1). Let $Z(\cdot)$ be a standard n -dimensional Wiener process and $F(\cdot)$ a bounded and measurable real-valued function define on the set of \mathbb{R}^n -valued continuous functions on $[0, T]$. By Theorem 3.1 in [8] we have

$$-\log \mathbb{E} \left[\exp \{ -F(Z(\cdot)) \} \right] = \inf_{u \in \mathcal{A}} \mathbb{E} \left[\frac{1}{2} \int_0^T \|u(s)\|^2 ds + F \left(Z(\cdot) + \int_0^\cdot u(s) ds \right) \right] \tag{3.2}$$

where \mathcal{A} is the set of all \mathfrak{F}_s -progressively measurable n -dimensional processes $u \doteq \{u(s), 0 \leq s \leq T\}$ satisfying

$$\mathbb{E} \int_0^T \|u(s)\|^2 ds < \infty.$$

In the present case, let $Z(\cdot) = (W(\cdot), B(\cdot))$ and $n = 2k$. Under Condition 2.1, the system has (1.1) has a unique strong solution. Therefore X^ϵ and Y^ϵ are measurable functions of $Z(\cdot) = (W(\cdot), B(\cdot))$. After setting $F(Z(\cdot)) = h(X^\epsilon(\cdot))/\epsilon$ and rescaling

the controls by $\frac{1}{\sqrt{\epsilon}}$ we get the representation

$$\begin{aligned}
 & -\epsilon \ln \mathbb{E}_{x_0, y_0} \left[\exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] \\
 & = \inf_{u \in \mathcal{A}} \mathbb{E}_{x_0, y_0} \left[\frac{1}{2} \int_0^T [\|u_1(s)\|^2 + \|u_2(s)\|^2] ds + h(\bar{X}^\epsilon) \right] \tag{3.3}
 \end{aligned}$$

where the pair $(\bar{X}^\epsilon, \bar{Y}^\epsilon)$ is the unique strong solution to

$$\begin{aligned}
 d\bar{X}^\epsilon(s) & = \left[\frac{\epsilon}{\delta} b(\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s)) + c(\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s)) + \sigma(\bar{X}_t^\epsilon, \bar{Y}_t^\epsilon) u_1(s) \right] ds \\
 & \quad + \sqrt{\epsilon} \sigma(\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s)) dW(s), \\
 d\bar{Y}^\epsilon(s) & = \frac{1}{\delta} \left[\frac{\epsilon}{\delta} f(\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s)) + g(\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s)) + \tau_1(\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s)) u_1(s) \right. \\
 & \quad \left. + \tau_2(\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s)) u_2(s) \right] ds + \frac{\sqrt{\epsilon}}{\delta} [\tau_1(\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s)) dW(s) \\
 & \quad + \tau_2(\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s)) dB(s)], \\
 \bar{X}^\epsilon(0) & = x_0, \quad \bar{Y}^\epsilon(0) = y_0.
 \end{aligned}
 \tag{3.4}$$

Therefore in order to derive the Laplace principle for $\{X^\epsilon\}$, it is enough to study the limit of the right hand side of the variational representation (3.3). The first step in doing so is to consider the weak limit of the slow motion \bar{X}^ϵ of the controlled couple (3.4). Due to the involved controls, it is convenient to introduce the following occupation measure. Let $\Delta = \Delta(\epsilon) \downarrow 0$ as $\epsilon \downarrow 0$. The role of $\Delta(\epsilon)$ is to exploit a time-scale separation. Let A_1, A_2, B, Γ be Borel sets of $\mathcal{Z}, \mathcal{Z}, \mathcal{Y}, [0, T]$ respectively. Let $u_i^\epsilon \in A_i, i = 1, 2$ and let $(\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s))$ solve (3.4) with u_i^ϵ in place of u_i . We associate with $(\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s))$ and u_i^ϵ a family of occupation measures $\mathbb{P}^{\epsilon, \Delta}$ defined by

$$\begin{aligned}
 & \mathbb{P}^{\epsilon, \Delta}(A_1 \times A_2 \times B \times \Gamma) \\
 & = \int_\Gamma \left[\frac{1}{\Delta} \int_t^{t+\Delta} 1_{A_1}(u_1^\epsilon(s)) 1_{A_2}(u_2^\epsilon(s)) 1_B(\bar{Y}_s^\epsilon \text{ mod } \lambda) ds \right] dt.
 \end{aligned}$$

We assume that $u_i^\epsilon(s) = 0$ for $i = 1, 2$ if $s > T$.

For presentation purposes, we devote Sect. 3.1 to the limiting behavior of the controlled process $\{(\bar{X}^\epsilon, \bar{Y}^\epsilon), \epsilon > 0\}$ in (3.4) as $\epsilon \downarrow 0$. This is a law of large numbers result. The large deviations result is in Sect. 3.2

3.1 Limiting Behavior of the Controlled Process (3.4)

Theorem 3.2, deals with the limiting behavior of the controlled process (3.4) under each of the three regimes, and uses the notion of a viable pair.

Theorem 3.2 Assume Condition 2.1 and under Regime 1 assume Condition 2.2. Fix the initial point $(x_0, y_0) \in \mathbb{R}^m \times \mathbb{R}^{d-m}$ and consider a family $\{u^\epsilon = (u_1^\epsilon, u_2^\epsilon), \epsilon > 0\}$ of controls in \mathcal{A} satisfying

$$\sup_{\epsilon > 0} \mathbb{E} \int_0^T [\|u_1^\epsilon(s)\|^2 + \|u_2^\epsilon(s)\|^2] ds < \infty. \tag{3.5}$$

Then the family $\{(\bar{X}^\epsilon, P^{\epsilon, \Delta}), \epsilon > 0\}$ is tight. Given the particular regime of interaction $i = 1, 2, 3$ and given any subsequence of $\{(\bar{X}^\epsilon, P^{\epsilon, \Delta}), \epsilon > 0\}$, there exists a subsubsequence that converges in distribution with limit (\bar{X}^i, P^i) . With probability 1, the limit point $(\bar{X}^i, P^i) \in \mathcal{V}_{(\lambda_i, \mathcal{L}^i)}$, according to Definition 2.5, with the pairs $(\lambda_i, \mathcal{L}^i)$ given by Definitions 2.3 and 2.4.

Proof We will only present the proof for Regime 1, since the proof for the other regimes is completely analogous. We start by proving tightness and then we identify the limit.

Tightness of $\{\bar{X}^\epsilon, \epsilon > 0\}$ follows if we establish that for every $\eta > 0$

$$\lim_{\rho \downarrow 0} \limsup_{\epsilon \downarrow 0} \mathbb{P}_{x_0, y_0} \left[\sup_{|t_1 - t_2| < \rho, 0 \leq t_1 < t_2 \leq T} |\bar{X}^\epsilon(t_1) - \bar{X}^\epsilon(t_2)| \geq \eta \right] = 0. \tag{3.6}$$

The difficulty to obtain this estimate in Regime 1 is due to the unclear behavior of the term $\frac{\epsilon}{\delta} \int_0^t b(\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s)) ds$ as $\epsilon/\delta \uparrow \infty$. We treat this term by applying Itô formula to $\chi(x, y)$, the solution to the cell problem (2.1). By doing so, we can rewrite the first component of (3.4), omitting function arguments in some places for notational convenience, as

$$\begin{aligned} d\bar{X}^\epsilon(s) &= \lambda_1(\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s), u_1(s), u_2(s)) ds \\ &+ \left(\epsilon \frac{\partial \chi}{\partial x} b + \delta \frac{\partial \chi}{\partial x} (c + \sigma u_1(s)) + \frac{\epsilon \delta}{2} \sigma \sigma^T : \frac{\partial^2 \chi}{\partial x^2} + \epsilon \sigma \tau_1^T : \frac{\partial^2 \chi}{\partial x \partial y} \right) \\ &\times (\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s)) ds \\ &+ \left(\sqrt{\epsilon} \left(\sigma + \frac{\partial \chi}{\partial y} \tau_1 \right) + \sqrt{\epsilon} \delta \frac{\partial \chi}{\partial x} \sigma \right) (\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s)) dW(s) \\ &+ \sqrt{\epsilon} \frac{\partial \chi}{\partial y} \tau_2 (\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s)) dB_s. \end{aligned}$$

Then, from this representation and the boundedness of the coefficients and of the derivatives of χ (Chap. 3, Sect. 6 of [6]), statement (3.6) follows, which then gives tightness of $\{\bar{X}^\epsilon, \epsilon > 0\}$.

Tightness of the occupation measures $\{P^{\epsilon, \Delta}, \epsilon > 0\}$ follows from the bound

$$\sup_{\epsilon \in (0, 1]} \mathbb{E}_{x_0, y_0} [g(P^{\epsilon, \Delta})] < \infty, \tag{3.7}$$

for the tightness function $g(r) = \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, T]} [\|z_1\|^2 + \|z_2\|^2] r(dz_1 dz_2 dy dt)$, $r \in \mathcal{P}(\mathcal{Z}^2 \times \mathcal{Y} \times [0, T])$. To be precise, notice that the function $g(r)$ is a tightness function since it is bounded from below, with relatively compact level sets

$G_k = \{r \in \mathcal{P}(\mathcal{Z}^2 \times \mathcal{Y} \times [0, T]) : g(r) \leq k\}$ for each $k < \infty$. Then, by Theorem A.19 in [11], it is known that tightness of $\{\mathbb{P}^{\epsilon, \Delta}, \epsilon > 0\}$ holds if (3.7) holds. However, this follows directly from Condition (3.5), due to the estimate

$$\begin{aligned} & \sup_{\epsilon \in (0, 1]} \mathbb{E}_{x_0, y_0} [g(\mathbb{P}^{\epsilon, \Delta})] \\ &= \sup_{\epsilon \in (0, 1]} \mathbb{E}_{x_0, y_0} \left[\int_{\mathcal{Z}^2 \times \mathcal{Y} \times [0, T]} [\|z_1\|^2 + \|z_2\|^2] \mathbb{P}^{\epsilon, \Delta}(dz_1 dz_2 dy dt) \right] \\ &= \sup_{\epsilon \in (0, 1]} \mathbb{E}_{x_0, y_0} \int_0^T \frac{1}{\Delta} \int_t^{t+\Delta} [\|u_1^\epsilon(s)\|^2 + \|u_2^\epsilon(s)\|^2] ds dt \\ &< \infty. \end{aligned}$$

Hence, we have established that the family $\{(\bar{X}^\epsilon, \mathbb{P}^{\epsilon, \Delta}), \epsilon > 0\}$ is tight. Next, we prove that any accumulation point will be a viable pair according to Definition 2.5. Tightness guarantees that for any subsequence of $\epsilon > 0$ there exists subsubsequence that converges, in distribution, to some limit (\bar{X}, \mathbb{P}) such that

$$(\bar{X}^\epsilon, \mathbb{P}^{\epsilon, \Delta}) \rightarrow (\bar{X}, \mathbb{P}).$$

Making use of the Skorokhod representation theorem we may assume, by the introduction of another probability space, which we omit writing in the notation, that this convergence holds with probability 1 (Theorem 1.8 [16]). Fatou’s Lemma gives us

$$\mathbb{E}_{x_0, y_0} \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, T]} [\|z_1\|^2 + \|z_2\|^2] \mathbb{P}(dz_1 dz_2 dy dt) < \infty \tag{3.8}$$

which then implies that $\int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, T]} [\|z_1\|^2 + \|z_2\|^2] \mathbb{P}(dz_1 dz_2 dy dt) < \infty$ w.p.1.

Thus, we need now to prove that the limit point $(\bar{X}, \mathbb{P}) \in \mathcal{V}_{(\lambda_1, \mathcal{L}^1)}$, according to Definition 2.5. Some of the computations here are analogous to those of the proof of Theorem 2.8 in [12]. We recall for completeness the main arguments appropriately modified to cover the more general case that is considered here.

We start with (2.2). This follows from the characterization of solutions to SDE’s via the martingale problem formulation and the averaging principle [6, 16]. We fix a collection of elements $p, q, S, t_i, \tau, F, \phi_j, \zeta$ that are defined as follows. $S, t_i, \tau \geq 0, i \leq q$ are such that $t_i \leq S \leq S + \tau \leq T$. The real valued functions F, ϕ_j are smooth and have compact support. Moreover, ζ is a real valued, bounded and continuous function with compact support on $(\mathbb{R}^m)^q \times \mathbb{R}^{pq}$.

Then, we define $\bar{A}_t^{\epsilon, \Delta}$ by

$$\bar{A}_t^{\epsilon, \Delta} F(x) = \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y}} \lambda(x, y, z_1, z_2) \nabla F(x) \mathbb{P}_t^{\epsilon, \Delta}(dz_1 dz_2 dy) \tag{3.9}$$

where

$$\mathbb{P}_t^{\epsilon, \Delta}(dz_1 dz_2 dy) = \frac{1}{\Delta} \int_t^{t+\Delta} 1_{dz_1}(u_1^\epsilon(s)) 1_{dz_2}(u_2^\epsilon(s)) 1_{dy}(\bar{Y}^\epsilon(s) \bmod \lambda) ds.$$

With these definitions at hand, (2.2) follow, if we prove that, as $\epsilon \downarrow 0$,

$$\mathbb{E}_{x_0, y_0} \left[\zeta(\bar{X}_t^\epsilon, (\mathbb{P}^{\epsilon, \Delta}, \phi_j)_{t_i}, i \leq q, j \leq p) \left[F(\bar{X}^\epsilon(S + \tau)) - F(\bar{X}^\epsilon(S)) - \int_S^{S+\tau} \bar{A}_t^{\epsilon, \Delta} F(\bar{X}^\epsilon(t)) dt \right] \right] \rightarrow 0 \tag{3.10}$$

and, in probability,

$$\int_S^{S+\tau} \bar{A}_s^{\epsilon, \Delta} F(\bar{X}^\epsilon(s)) ds - \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [S, S+\tau]} \lambda_1(\bar{X}_s, y, z_1, z_2) \times \nabla F(\bar{X}(s)) \bar{P}(dz_1 dz_2 dy ds) \rightarrow 0. \tag{3.11}$$

Then, the pair (\bar{X}, \bar{P}) solves the martingale problem associated with (2.2), which implies the latter.

So, let us first prove (3.10). Notice that under the topology of weak convergence and due to the fact that the last marginal of \mathbb{P} is Lebesgue measure w.p.1., we have that for every $t \in [0, T]$ and for every real valued, compactly supported and continuous function ϕ

$$(\mathbb{P}^{\epsilon, \Delta}, \phi)_t \rightarrow (\mathbb{P}, \phi)_t \quad \text{w.p.1.}$$

Next, we recall the solution $\chi(x, y)$ to the cell problem (2.1). Let $\psi = \{\psi_1, \dots, \psi_d\}$ be defined by $\psi_\ell(x, y) = \chi_\ell(x, y) F_{x_\ell}(x)$ for $\ell = 1, \dots, d$. Notice that $\psi_\ell(x, y)$ is periodic in every direction in y , with period λ , and satisfies

$$\mathcal{L}_x^1 \psi_\ell(x, y) = -b_\ell(x, y) F_{x_\ell}(x), \quad \int_{\mathcal{Y}} \psi_\ell(x, y) \mu(dy|x) = 0. \tag{3.12}$$

Applying Itô’s formula to $\psi(\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s))$ gives us that relation (3.12) and the boundedness of $\chi(x, y)$ and its derivatives guarantee the validity of (3.10), if

$$\int_S^{S+\tau} [\bar{A}_s^{\epsilon, \Delta} F(\bar{X}^\epsilon(s)) - \lambda_1(\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s), u_1^\epsilon(s), u_2^\epsilon(s)) \nabla F(\bar{X}^\epsilon(s))] ds \rightarrow 0, \tag{3.13}$$

as $\epsilon \downarrow 0$,

in probability. However, this is exactly the second statement of Lemma 3.2 in [12] by taking as $g(x, y, z_1, z_2) = \lambda_1(x, y, z_1, z_2) \cdot \nabla f(x)$. On the other hand, (3.11) is the first statement of Lemma 3.2 in [12], with the function g as was just specified. These give the proof of (2.2).

Next, we establish relation (2.3). Let $\mathcal{A}_{z_1, z_2, x}^\epsilon$ be the operator associated with the fast motion \bar{Y}^ϵ in (3.4) with $z_1 = u_1, z_2 = u_2$ and $x = \bar{X}^\epsilon$ fixed,

$$\mathcal{A}_{z_1, z_2, x}^\epsilon F(y) = \left[\frac{\epsilon}{\delta^2} f(x, y) + \frac{1}{\delta} [g(x, y) + \tau_1(x, y)z_1 + \tau_2(x, y)z_2] \right] \cdot \nabla_y F(y) + \frac{\epsilon}{\delta^2} \frac{1}{2} (\tau_1 \tau_1^T + \tau_2 \tau_2^T)(x, y) : \nabla_y \nabla_y F(y) \tag{3.14}$$

for functions $F \in \mathcal{C}^2(\mathcal{Y})$. Consider, $\{F_\ell : \mathcal{Y} \mapsto \mathbb{R}, \ell \in \mathbb{N}\}$ to be a smooth and dense family in $\mathcal{C}^2(\mathcal{Y})$. Then, notice that

$$M_t^\epsilon = F_\ell(Y^\epsilon(t)) - F_\ell(y_0) - \int_0^t \mathcal{A}_{u_1(s), u_2(s), \bar{X}^\epsilon(s)}^\epsilon F_\ell(\bar{Y}^\epsilon(s)) ds$$

is an \mathfrak{F}_t -martingale. Next, let

$$g(\epsilon) = \frac{\delta^2}{\epsilon}$$

and notice that $g(\epsilon)\mathcal{A}_{z_1, z_2, x}^\epsilon$ converges to $\mathcal{L}_{z_1, z_2, x}^1$ under Regime 1, as $\epsilon \downarrow 0$. Based on this observation and denoting $\mathcal{G}_{x, y, z_1, z_2} F_\ell(y) = [g(x, y) + \tau_1(x, y)z_1 + \tau_1(x, y)z_2] \cdot \nabla_y F_\ell(y)$ we write

$$\begin{aligned} &g(\epsilon)M_t^\epsilon - g(\epsilon)[F_\ell(\bar{Y}^\epsilon(t)) - F_\ell(y_0)] \\ &+ g(\epsilon)\left[\int_0^t \frac{1}{\Delta} \left[\int_s^{s+\Delta} \mathcal{A}_{u_1^\epsilon(\rho), u_2^\epsilon(\rho), \bar{X}^\epsilon(\rho)}^\epsilon F_\ell(\bar{Y}^\epsilon(\rho)) d\rho\right] ds \right. \\ &\left. - \int_0^t \mathcal{A}_{u_1^\epsilon(s), u_2^\epsilon(s), \bar{X}^\epsilon(s)}^\epsilon F_\ell(\bar{Y}^\epsilon(s)) ds\right] \\ &= -\frac{\delta}{\epsilon} \left(\int_0^t \frac{1}{\Delta} \left[\int_s^{s+\Delta} [\mathcal{G}_{\bar{X}^\epsilon(\rho), \bar{Y}^\epsilon(\rho), u_1^\epsilon(\rho), u_2^\epsilon(\rho)} F_\ell(\bar{Y}^\epsilon(\rho)) \right. \right. \\ &\left. \left. - \mathcal{G}_{\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s), u_1^\epsilon(s), u_2^\epsilon(s)} F_\ell(\bar{Y}^\epsilon(s))] d\rho\right] ds \right) \\ &- \frac{\delta}{\epsilon} \left(\int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, t]} \mathcal{G}_{\bar{X}^\epsilon(s), y, z_1, z_2} F_\ell(y) \bar{P}^{\epsilon, \Delta}(dz_1 dz_2 dy ds) \right) \\ &- \int_0^t \frac{1}{\Delta} \left[\int_s^{s+\Delta} [\mathcal{L}_{\bar{X}^\epsilon(\rho)}^1 F_\ell(\bar{Y}^\epsilon(\rho)) - \mathcal{L}_{\bar{X}^\epsilon(s)}^1 F_\ell(\bar{Y}^\epsilon(\rho))] d\rho\right] ds \\ &- \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, t]} \mathcal{L}_{\bar{X}^\epsilon(s)}^1 F_\ell(y) P^{\epsilon, \Delta}(dz_1 dz_2 dy dt). \tag{3.15} \end{aligned}$$

Let us now analyze the different terms in (3.15). In particular

(i) We have that $g(\epsilon)M_t^\epsilon \downarrow 0$ as $\epsilon \downarrow 0$ in probability. Indeed, we can rewrite

$$\begin{aligned} M_t^\epsilon &= \frac{\sqrt{\epsilon}}{\delta} \left[\int_0^t \nabla_y F_\ell(\bar{Y}^\epsilon(s)) \cdot \tau_1(\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s)) dW(s) \right. \\ &\left. + \int_0^t \nabla_y F_\ell(\bar{Y}^\epsilon(s)) \cdot \tau_2(\bar{X}^\epsilon(s), \bar{Y}^\epsilon(s)) dB(s) \right], \end{aligned}$$

which allows us to obtain that $\mathbb{E}_{x_0, y_0}[M_T^\epsilon]^2 \leq C_0 \frac{1}{g(\epsilon)}$, and so $g(\epsilon)M_t^\epsilon \downarrow 0$ follows from $g(\epsilon) \downarrow 0$.

(ii) Since F is bounded, we have that $g(\epsilon)[F_\ell(\bar{Y}^\epsilon(t)) - F_\ell(y_0)]$ converges to zero uniformly.

(iii) By Condition 2.1 and since $\Delta \downarrow 0, \delta/\epsilon \downarrow 0$, the term

$$g(\epsilon) \left[\int_0^t \frac{1}{\Delta} \left[\int_s^{s+\Delta} \mathcal{A}_{u_1^\epsilon(\rho), u_2^\epsilon(\rho), \bar{X}^\epsilon(\rho)}^\epsilon F_\ell(\bar{Y}^\epsilon(\rho)) d\rho \right] ds - \int_0^t \mathcal{A}_{u_1^\epsilon(s), u_2^\epsilon(s), \bar{X}^\epsilon(s)}^\epsilon F_\ell(\bar{Y}^\epsilon(s)) ds \right]$$

converges to zero in probability.

(iv) Tightness of $\{\bar{X}^\epsilon, \epsilon > 0\}$ and Conditions 2.1, (3.5), imply that the first and the third term in the right hand side of (3.15) converge to zero in probability as $\delta/\epsilon \downarrow 0$.

(v) Uniform integrability of $P^{\epsilon, \Delta}$ and the fact that $\delta/\epsilon \downarrow 0$ imply that the second term on the right hand side of (3.15) converges to zero in probability.

Therefore, we finally obtain that

$$\int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, T]} \mathcal{L}_{z_1, z_2, \bar{X}_t}^i F_\ell(y) P^{\epsilon, \Delta}(dz_1 dz_2 dy dt) \rightarrow 0, \quad \text{in probability.}$$

Then, this implies (2.3), by continuity in $t \in [0, T]$ and density of $\{F_\ell : \mathcal{Y} \mapsto \mathbb{R}, \ell \in \mathbb{N}\}$.

It remains to prove (2.4). It is clear, that the analogous property holds at the pre-limit level. Moreover, since $P(\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times \{t\}) = 0$ and the map $t \rightarrow P(\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, t])$ is continuous, one can deal with null sets. Thus, (2.4) follows.

The proof for Regimes 2 and 3 is completely analogous with the only exception that for the purposes of the proof (2.3), for Regime 2 we define $g(\epsilon) = \epsilon$ and for Regime 3, $g(\epsilon) = \delta$. This completes the proof of the theorem. \square

3.2 Large Deviations for $\{X^\epsilon, \epsilon > 0\}$

In this subsection we present the main large deviations result. The main difference from the case considered in [12] is the identification of the correct viable pair with respect to which the large deviations principle is expressed to. The proper viable pair in each regime is indicated by Theorem 3.2.

Theorem 3.3 *Let $\{(X^\epsilon, Y^\epsilon), \epsilon > 0\}$ be the unique strong solution to (1.1) and consider Regime $i = 1, 2, 3$. Assume Condition 2.1 and under Regime 1 assume Condition 2.2. Define*

$$S^i(\phi) = \inf_{(\phi, P) \in \mathcal{V}_{(\lambda_i, \mathcal{L}^i)}} \left[\frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, T]} [\|z_1\|^2 + \|z_2\|^2] P(dz_1 dz_2 dy dt) \right], \tag{3.16}$$

with the convention that the infimum over the empty set is ∞ . The pairs $(\lambda_i, \mathcal{L}^i)$ are given in Definitions 2.3 and 2.4. Then, we have

(i) The level sets of S^i are compact. In particular, for each $s < \infty$, the set

$$\Phi_s^i = \{ \phi \in \mathcal{C}([0, T]; \mathbb{R}^m) : S^i(\phi) \leq s \}$$

is a compact subset of $\mathcal{C}([0, T]; \mathbb{R}^m)$.

(ii) For every bounded and continuous function h mapping $\mathcal{C}([0, T]; \mathbb{R}^m)$ into \mathbb{R}

$$\liminf_{\epsilon \downarrow 0} -\epsilon \ln \mathbb{E}_{x_0, y_0} \left[\exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] \geq \inf_{\phi \in \mathcal{C}([0, T]; \mathbb{R}^m)} [S^i(\phi) + h(\phi)].$$

(iii) In the case of Regime 3 assume either that we are in dimension 1 i.e., $m = 1$, $d = 2$, or that $g(x, y) = g(y)$ and $\tau_i(x, y) = \tau_i(y)$, $i = 1, 2$ for the general multidimensional case. Then for every bounded and continuous function h mapping $\mathcal{C}([0, T]; \mathbb{R}^m)$ into \mathbb{R}

$$\limsup_{\epsilon \downarrow 0} -\epsilon \ln \mathbb{E}_{x_0, y_0} \left[\exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] \leq \inf_{\phi \in \mathcal{C}([0, T]; \mathbb{R}^m)} [S^i(\phi) + h(\phi)].$$

In other words, under the imposed assumptions, $\{X^\epsilon, \epsilon > 0\}$ satisfies the large deviations principle with action functional S^i .

For the sake of presentation, the proof of Theorem 3.3 is deferred to the end of this section. In the case of Regime 1 we can get an explicit characterization of the rate function.

Theorem 3.4 Let $\{(X^\epsilon, Y^\epsilon), \epsilon > 0\}$ be the unique strong solution to (1.1) and consider Regime 1. Under Conditions 2.1 and 2.2, $\{X^\epsilon, \epsilon > 0\}$ satisfies a large deviations principle with rate function

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

Proof It follows by putting Lemma 3.6 and Theorem 3.7 below together. □

Notice that the coefficients $r(x)$ and $q(x)$ that enter into the action functional for Regime 1 are those obtained if we had first taken to (1.1) $\delta \downarrow 0$ with ϵ fixed and then consider the large deviations for the homogenized system. Indeed if $\epsilon = 1$, then $X^{\epsilon, \delta} = X^{1, \delta}$ can be shown to converge weakly in the space of continuous functions in $\mathcal{C}([0, T]; \mathbb{R}^m)$, as $\delta \downarrow 0$, to the solution of an SDE with drift coefficient $r(x)$ and diffusion coefficient $q^{1/2}(x)$. This can be derived via standard homogenization theory [6, 31]. The action functional for a small noise diffusion with drift coefficient $r(x)$ and diffusion coefficient $\sqrt{\epsilon}q^{1/2}(x)$ is the one given by Theorem 3.4. This is in accordance to intuition since under Regime 1, δ goes to zero faster, so homogenization should occur first as it indeed does.

Remark 3.5 Notice that if we set $f = b, g = c, \sigma = \tau_1$ and $\tau_2 = 0$ in the statements of Theorems 3.2–3.4, then one recovers the results of [12].

The reader may wonder, why we have imposed further structural restrictions for part (iii) of the theorem for Regime 3. This is because we were not able to prove some smoothness requirements of the constructed nearly optimal controls in the prelimit level with respect to x in the general multidimensional case when the coefficients g , τ_1 and τ_2 depend on x . Similar issues arise in the case considered in [12] and are discussed in detail there. However, observing the viable pairs that characterize the large deviations principle for Regimes 2 and 3, $(\lambda_2, \mathcal{L}^2)$ and $(\lambda_3, \mathcal{L}^3)$ respectively, we notice that Regime 3 can be thought of as a limiting case of Regime 2 with $\gamma = 0$. So, one is led to conjecture that the extra assumptions for Regime 3 are not necessary, even though we currently do not have a proof for this.

Next, we proceed with the proof of Theorem 3.3. The proof is based on two intermediate results, which we present now. Lemma 3.6 allows us to equivalently rewrite the action functional S^i .

Let $\mathcal{AC}([0, T]; \mathbb{R}^m)$ be the set of absolutely continuous functions from $[0, T]$ to \mathbb{R}^m .

Lemma 3.6 *Consider the set-up of Theorem 3.3. Then, for $i = 1, 2, 3$ we have*

$$S^i(\phi) = \begin{cases} \int_0^T L_i(\phi(s), \dot{\phi}(s)) ds & \text{if } \phi \in \mathcal{AC}([0, T]; \mathbb{R}^m) \text{ and } \phi(0) = x_0 \\ +\infty & \text{otherwise,} \end{cases}$$

where

$$L_i(x, \beta) = \inf_{(v, \mu) \in \mathcal{A}_{x, \beta}^i} \left\{ \frac{1}{2} \int_{\mathcal{Y}} \|v(y)\|^2 \mu(dy) \right\} \tag{3.17}$$

with

$$\begin{aligned} \mathcal{A}_{x, \beta}^i = & \left\{ v(\cdot) = (v_1(\cdot), v_2(\cdot)) : \mathcal{Y} \mapsto \mathbb{R}^{2\kappa}, \mu \in \mathcal{P}(\mathcal{Y}) : (v, \mu) \right. \\ & \text{satisfy } \int_{\mathcal{Y}} \mathcal{L}_{v_1(y), v_2(y), x}^i F(y) \mu(dy) = 0 \\ & \text{for all } F \in \mathcal{D}(\mathcal{L}_{v_1, v_2, x}^i), \int_{\mathcal{Y}} \|v(y)\|^2 \mu(dy) < \infty \\ & \left. \text{and } \beta = \int_{\mathcal{Y}} \lambda_i(x, y, v_1(y), v_2(y)) \mu(dy) \right\}. \end{aligned}$$

Proof The proof of this lemma follows easily by an appropriate rewriting of the corresponding expressions. First, notice that (3.16) can be written in terms of a local rate function

$$S^i(\phi) = \int_0^T L_i^r(\phi(s), \dot{\phi}(s)) ds$$

(if ϕ is absolutely continuous). This follows from the definition of a viable pair by setting

$$L_i^r(x, \beta) = \inf_{P \in \mathcal{A}_{x,\beta}^{i,r}} \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y}} \frac{1}{2} [\|z_1\|^2 + \|z_2\|^2] P(dz_1 dz_2 dy), \tag{3.18}$$

where

$$\begin{aligned} \mathcal{A}_{x,\beta}^{i,r} = \left\{ P \in \mathcal{P}(\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y}) : \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y}} \mathcal{L}_{z_1, z_2, x}^i F(y) P(dz_1 dz_2 dy) = 0 \right. \\ \text{for all } F \in \mathcal{D}(\mathcal{L}_{z_1, z_2, x}^i) \\ \left. \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y}} [\|z_1\|^2 + \|z_2\|^2] P(dz_1 dz_2 dy) < \infty \text{ and} \right. \\ \left. \beta = \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y}} \lambda_i(x, y, z_1, z_2) P(dz_1 dz_2 dy) \right\}. \end{aligned}$$

Second, we note that $P \in \mathcal{P}(\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y})$ can be decomposed into marginals as follows

$$P(dz_1 dz_2 dy) = \eta(dz_1 dz_2 | y) \mu(dy).$$

This, the convexity of the cost on (z_1, z_2) and the affine dependence of λ_i on (z_1, z_2) imply that the relaxed control formulation (3.18) and the ordinary control formulation (3.17) are equivalent by taking

$$v_i(y) = \int_{\mathcal{Z} \times \mathcal{Z}} z_i \eta(dz_1 dz_2 | y).$$

Hence, the statement of the lemma holds. □

Now, we use the representations in Lemma 3.6 to obtain the controls needed in the proof of part (iii) of Theorem 3.3. In the case of Regime 1 we can be even more specific and obtain a closed form expression for the variational problem associated to the local rate function $L_1(x, \beta)$ appearing in Lemma 3.6. The derivation of the closed form expression is based on identifying an optimal control that is then used to prove Theorem 3.3. The proof of this statement is based on a straightforward Lagrange multiplier type of analysis of the variational problem (3.17) for $i = 1$ and thus omitted (see also Theorem 5.2 in [12] for an analogous situation). In the case of Regime 2 and Regime 3, we can obtain that there is pair (\bar{v}, μ) that attains the infimum in (3.17). We collect these statements in the following theorem. Its proof is omitted, since it follows analogously to the corresponding proofs of Theorems 5.2, 6.2 for Regimes 1 and 2 respectively and from Sect. 7 for Regime 3, of [12].

Theorem 3.7 *Assume Condition 2.1 and in the case of Regime 1 assume Condition 2.2. The infimization problem (3.17) for $i = 1$ has the explicit solution*

$$L_1(x, \beta) = \frac{1}{2} (\beta - r(x))^T q^{-1}(x) (\beta - r(x)),$$

where

- $r(x) = \int_{\mathcal{Y}} (c(x, y) + \frac{\partial \chi}{\partial y}(x, y)g(x, y)) \mu(dy|x),$
- $q(x) = \int_{\mathcal{Y}} [(\sigma + \frac{\partial \chi}{\partial y} \tau_1)(\sigma + \frac{\partial \chi}{\partial y} \tau_1)^T + (\frac{\partial \chi}{\partial y} \tau_2)(\frac{\partial \chi}{\partial y} \tau_2)^T](x, y) \mu(dy|x),$

and where $\mu(dy|x)$ is the unique invariant measure corresponding to the operator \mathcal{L}_x^1 and $\chi(x, y)$ is defined by (2.1). The control

$$\begin{aligned} \bar{u}(y) &= (\bar{u}_{1,\beta}(x, y), \bar{u}_{2,\beta}(x, y)) \\ &= \left(\left(\sigma + \frac{\partial \chi}{\partial y} \tau_1 \right)^T (x, y) q^{-1}(x) (\beta - r(x)), \right. \\ &\quad \left. \left(\frac{\partial \chi}{\partial y} \tau_2 \right)^T (x, y) q^{-1}(x) (\beta - r(x)) \right) \end{aligned}$$

attains the infimum in (3.17).

In the case of Regime 2 and in the one dimensional setting of Regime 3, there is a pair $(\bar{u}, \bar{\mu})$ that achieves the infimum in (3.17) such that $\bar{u} = \bar{u}_\beta(x, y)$ is, for each fixed $\beta \in \mathbb{R}^d$, continuous in x , Lipschitz continuous in y and measurable in (x, y, β) . Moreover, $\bar{\mu}(dy) = \bar{\mu}_{\bar{u}}^i(dy|x)$ is the unique invariant measure corresponding to the operator $\mathcal{L}_{\bar{u}_\beta(x,y),x}^i, i = 2, 3$, and it is weakly continuous as a function of x . In the x -independent multidimensional case of Regime 3, there is a $P \in \mathcal{A}_\beta^{3,r}$ that achieves the infimum in (3.18) or equivalently in (3.17).

Now we have the necessary tools to prove Theorem 3.3.

Proof of Theorem 3.3 Part (i). As in Lemma 4.1 and Lemma 4.3 of [12], one can establish that Φ_s^i is precompact and closed, respectively. These two statements, then give compactness of Φ_s^i .

Part (ii). We have the following chain of inequalities:

$$\begin{aligned} &\liminf_{\epsilon \downarrow 0} \left(-\epsilon \ln \mathbb{E}_{x_0, y_0} \left[\exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] \right) \\ &\geq \liminf_{\epsilon \downarrow 0} \left(\mathbb{E}_{x_0, y_0} \left[\frac{1}{2} \int_0^T [\|u_1^\epsilon(t)\|^2 + \|u_2^\epsilon(t)\|^2] dt + h(\bar{X}^\epsilon) \right] - \epsilon \right) \\ &\geq \liminf_{\epsilon \downarrow 0} \left(\mathbb{E}_{x_0, y_0} \left[\frac{1}{2} \int_0^T \frac{1}{\Delta} \int_t^{t+\Delta} [\|u_1^\epsilon(s)\|^2 + \|u_2^\epsilon(s)\|^2] ds dt + h(\bar{X}^\epsilon) \right] \right) \\ &= \liminf_{\epsilon \downarrow 0} \left(\mathbb{E}_{x_0, y_0} \left[\frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y} \times [0, T]} [\|z_1\|^2 + \|z_2\|^2] P^{\epsilon, \Delta}(dz_1 dz_2 dy dt) + h(\bar{X}^\epsilon) \right] \right) \\ &\geq \mathbb{E}_{x_0, y_0} \left[\frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y} \times [0, T]} [\|z_1\|^2 + \|z_2\|^2] \bar{P}(dz_1 dz_2 dy dt) + h(\bar{X}) \right] \\ &\geq \inf_{(\phi, P) \in \mathcal{V}} \left\{ \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Z} \times \mathcal{Y} \times [0, T]} [\|z_1\|^2 + \|z_2\|^2] P(dz_1 dz_2 dy dt) + h(\phi) \right\} \\ &= \inf_{\phi \in \mathcal{C}([0, T]; \mathbb{R}^m)} [S^i(\phi) + h(\phi)]. \end{aligned}$$

The first line follows by the representation formula (3.3). The fourth line by Theorem 3.2 and Fatou’s Lemma. This establishes the lower bound.

Part (iii). In each regime we follow the same general steps. What differs from regime to regime, is the form of the viable pair $(\lambda_i, \mathcal{L}^i)$ in the definition of the action functional $S^i(\cdot)$. To prove the Laplace principle upper bound we must show that for all bounded, continuous functions h mapping $\mathcal{C}([0, T]; \mathbb{R}^m)$ into \mathbb{R}

$$\limsup_{\epsilon \downarrow 0} -\epsilon \ln \mathbb{E}_{x_0, y_0} \left[\exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] \leq \inf_{\phi \in \mathcal{C}([0, T]; \mathbb{R}^m)} [S^i(\phi) + h(\phi)].$$

By the variational representation formula (3.3), it is enough to prove that

$$\begin{aligned} & \limsup_{\epsilon \downarrow 0} \inf_{u=(u_1, u_2) \in \mathcal{A}} \mathbb{E}_{x_0, y_0} \left[\frac{1}{2} \int_0^T [\|u_1(s)\|^2 + \|u_2(s)\|^2] ds + h(\bar{X}^\epsilon) \right] \\ & \leq \inf_{\phi \in \mathcal{C}([0, T]; \mathbb{R}^m)} [S^i(\phi) + h(\phi)]. \end{aligned} \tag{3.19}$$

In each regime, we consider for the limiting variational problem in the Laplace principle a nearly optimal control pair (ψ, P) . In particular, let $\eta > 0$ be given and consider $\psi \in \mathcal{C}([0, T]; \mathbb{R}^m)$ with $\psi_0 = x_0$ such that

$$S^i(\psi) + h(\psi) \leq \inf_{\phi \in \mathcal{C}([0, T]; \mathbb{R}^m)} [S^i(\phi) + h(\phi)] + \eta < \infty.$$

Let us consider first Regimes 1 and 2 and Regime 3 in dimension 1. By Lemma 3.6, the local rate function of S^i is $L_i(x, \beta)$ given by (3.17). By Theorem 3.7, it is clear that $L_1(x, \beta)$ is continuous and finite at each $(x, \beta) \in \mathbb{R}^{2m}$. As in Theorem 6.3 in [12] the same is true for Regime 2 and as in Theorem 7.2 of [12] for Regime 3 in dimension 1. Thus, a standard mollification argument, allows us to further assume that $\dot{\psi}$ is piecewise constant (see for example Lemmas 6.5.3 and 6.5.5 in Sect. 6.5 of [11]). Then, by Theorem 3.7, we obtain that with $\beta = \dot{\psi}_t$, there is $\bar{u}_{\dot{\psi}_t}(x, y) = (\bar{u}_{1, \dot{\psi}_t}(x, y), \bar{u}_{2, \dot{\psi}_t}(x, y))$ that is bounded, continuous in x and Lipschitz continuous in y , and piecewise constant in t such that

$$\bar{u}_{\dot{\psi}_t}(x, \cdot) \in \operatorname{argmin}_v \left\{ \frac{1}{2} \int_Y \|v(y)\|^2 \mu(dy) : (v, \mu) \in \mathcal{A}_{x, \dot{\psi}_t}^i \right\}. \tag{3.20}$$

Let us denote by $\bar{\mu}_{\bar{u}}(dy)$ the unique invariant measure corresponding to the operator $\mathcal{L}_{\bar{u}, x}^i$ for this particular control \bar{u} (which exists due to Theorem 3.7). The feedback type of control used to prove the upper bound is then

$$\bar{u}^\epsilon(t) = \bar{u}_{\dot{\psi}_t}(\bar{X}^\epsilon(t), \bar{Y}^\epsilon(t)) = (\bar{u}_{1, \dot{\psi}_t}(\bar{X}^\epsilon(t), \bar{Y}^\epsilon(t)), \bar{u}_{2, \dot{\psi}_t}(\bar{X}^\epsilon(t), \bar{Y}^\epsilon(t))).$$

By Condition 2.1 and since \bar{u} is continuous in x and y , (3.4) has a strong solution with $(u_1(t), u_2(t)) = \bar{u}^\epsilon(t)$.

Then, standard averaging theory (e.g., Sect. 6, Chap. 3 of [6]) and the fact that $\bar{\mu}_{\bar{u}_{\dot{\psi}_t}(x, \cdot)}(\cdot)$ is weakly continuous in x (Theorem 3.7) and piecewise continuous in t

we have that $\bar{X}^\epsilon \xrightarrow{\mathcal{D}} \bar{X}$, where

$$\bar{X}_t = x_0 + \int_0^t \int_{\mathcal{Y}} \lambda_i(\bar{X}_s, y, \bar{u}_{\psi_s}(\bar{X}_s, y)) \bar{\mu}_{\bar{u}_{\psi_s}(\bar{X}_s, \cdot)}(dy) ds.$$

Then for ψ such that $\psi_0 = x_0$, (3.20) and the definition of $\mathcal{A}_{x, \psi_t}^i$ gives us that

$$\bar{X}_t = x_0 + \int_0^t \dot{\psi}_s ds = \psi_t \quad \text{for any } t \in [0, T], \text{ w.p.1.}$$

Therefore, by the representation formula (3.3) and (3.20) we finally obtain that

$$\begin{aligned} & \limsup_{\epsilon \downarrow 0} \left[-\epsilon \ln \mathbb{E}_{x_0, y_0} \left[\exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] \right] \\ &= \limsup_{\epsilon \downarrow 0} \inf_{(u_1, u_2)} \mathbb{E}_{x_0, y_0} \left[\frac{1}{2} \int_0^T [\|u_1(t)\|^2 + \|u_2(t)\|^2] dt + h(\bar{X}^\epsilon) \right] \\ &\leq \limsup_{\epsilon \downarrow 0} \mathbb{E}_{x_0, y_0} \left[\frac{1}{2} \int_0^T \|\bar{u}^\epsilon(t)\|^2 dt + h(\bar{X}^\epsilon) \right] \\ &= \mathbb{E}_{x_0, y_0} [S^i(\bar{X}) + h(\bar{X})] \\ &\leq \inf_{\phi \in \mathcal{C}([0, T]; \mathbb{R}^m)} [S^i(\phi) + h(\phi)] + \eta. \end{aligned}$$

Since η is arbitrary, we are done. The proof of part (iii) for Regime 3 follows as in Sect. 7 of [12], using the fact that there is (by Theorem 3.7 below) a $P \in \mathcal{A}_\beta^{3, T}$ that achieves the infimum in (3.18). \square

4 Importance Sampling

The purpose of this section is to utilize the large deviations results of Section 3 in order to obtain asymptotically efficient importance sampling schemes for quantities like (1.2). Simulation problems involving rare events unavoidably have a number of mathematical and computational challenges. As it is well known, standard Monte Carlo sampling techniques perform very poorly in that the relative errors under a fixed computational effort grow rapidly as the event becomes more rare. Rare event estimation problems for systems of fast and slow motion present extra difficulties due to the underlying fast motion and its interaction with the intensity of the noise ϵ . In particular, one needs to take into account the solution to the appropriate cell problem associated with the homogenization theory of HJB equations in order to guarantee asymptotic optimality. Related simulation results are provided in [13, 14] for the special case $f(x, y) = b(x, y) = -\nabla Q(y)$, $g(x, y) = c(x, y) = -\nabla V(x)$, $\sigma(x, y) = \tau_1(x, y) = \text{constant}$ and $\tau_2(x, y) = 0$.

We start by reviewing general things about importance sampling adjusting the discussion to our setting of interest. Consider a bounded continuous function h :

$\mathbb{R}^m \mapsto \mathbb{R}$ and suppose that one is interested in estimating

$$\theta(\epsilon) \doteq \mathbb{E}\left[e^{-\frac{1}{\epsilon}h(X^\epsilon(T))} \mid X^\epsilon(t_0) = x_0, Y^\epsilon(t_0) = y_0\right]$$

by Monte Carlo, where the pair of slow and fast motion (X^ϵ, Y^ϵ) has initial point $X^\epsilon(t_0) = x_0, Y^\epsilon(t_0) = y_0$. For Regime $i = 1, 2, 3$, let

$$G_i(t_0, x_0) \doteq \inf_{\phi \in \mathcal{C}([t_0, T]; \mathbb{R}^m), \phi(t_0) = x_0} \left[S_{t_0 T}^i(\phi) + h(\phi(T)) \right]. \tag{4.1}$$

As we shall see below, under regularity conditions, the function $G_i(t, x)$ satisfies a PDE of HJB type. Now, depending on the regime of interaction, the contraction principle implies

$$\lim_{\epsilon \rightarrow 0} -\epsilon \log \theta(\epsilon) = G_i(t_0, x_0). \tag{4.2}$$

Notice that the limit is independent of the initial point y_0 of the fast motion Y^ϵ . This is due to the averaging that takes place, as we shall also see later on in the rigorous proofs.

Let $\Gamma^\epsilon(t_0, x_0, y_0)$ be any unbiased estimator of $\theta(\epsilon)$ that is defined on some probability space with probability measure $\bar{\mathbb{P}}$. With $\bar{\mathbb{E}}$ denoting the expectation operator associated with $\bar{\mathbb{P}}$ we have that $\Gamma^\epsilon(t_0, x_0, y_0)$ is a random variable such that

$$\bar{\mathbb{E}}\Gamma^\epsilon(t_0, x_0, y_0) = \theta(\epsilon).$$

In Monte Carlo simulation, one generates a number of independent copies of $\Gamma^\epsilon(t, x, y)$ and the estimate is the sample mean. The specific number of samples required depends on the desired accuracy, which is measured by the variance of the sample mean. Because of unbiasedness, minimizing the variance is equivalent to minimizing the second moment. Jensen’s inequality implies

$$\bar{\mathbb{E}}(\Gamma^\epsilon(t_0, x_0, y_0))^2 \geq (\bar{\mathbb{E}}\Gamma^\epsilon(t_0, x_0, y_0))^2 = \theta(\epsilon)^2.$$

This and (4.2) say that

$$\limsup_{\epsilon \rightarrow 0} -\epsilon \log \bar{\mathbb{E}}(\Gamma^\epsilon(t_0, x_0, y_0))^2 \leq 2G_i(t_0, x_0).$$

Hence, $2G_i(t_0, x_0)$ is the best possible rate of decay of the second moment. If

$$\liminf_{\epsilon \rightarrow 0} -\epsilon \log \bar{\mathbb{E}}(\Gamma^\epsilon(t_0, x_0, y_0))^2 \geq 2G_i(t_0, x_0),$$

then $\Gamma^\epsilon(t_0, x_0, y_0)$ achieves this best decay rate, and is said to be *asymptotically optimal*.

It is important to note here that asymptotic optimality is not the only practical concern. Rare events associated with multiscale problems are rather complicated and many times is it very difficult to construct asymptotically optimal schemes. One way to circumvent this difficulty is by constructing appropriate sub-optimal schemes with

precise bounds on asymptotic performance. This is the content of Theorems 4.6, 4.8 and 4.10 for Regime $i = 1, 2, 3$ respectively.

Fix the Regime $i = 1, 2, 3$ and assume that we are given a control $\bar{u}(s, x, y; i)$ that is sufficiently smooth and bounded. Let us recall the 2κ -dimensional Wiener process $Z(\cdot) = (W(\cdot), B(\cdot))$. Consider the family of probability measures \bar{P}^ϵ defined by the change of measure

$$\begin{aligned} \frac{d\bar{P}^\epsilon}{dP} &= \exp \left\{ -\frac{1}{2\epsilon} \int_{t_0}^T \|\bar{u}(s, X^\epsilon(s), Y^\epsilon(s); i)\|^2 ds \right. \\ &\quad \left. + \frac{1}{\sqrt{\epsilon}} \int_{t_0}^T \langle \bar{u}(s, X^\epsilon(s), Y^\epsilon(s); i), dZ(s) \rangle \right\}. \end{aligned}$$

By Girsanov’s Theorem

$$\bar{Z}(s) = Z(s) - \frac{1}{\sqrt{\epsilon}} \int_{t_0}^s \bar{u}(\rho, X^\epsilon(\rho), Y^\epsilon(\rho); i) d\rho, \quad t_0 \leq s \leq T$$

is a Wiener process on $[t_0, T]$ under the probability measure \bar{P}^ϵ , and (X^ϵ, Y^ϵ) satisfies $X^\epsilon(t_0) = x_0, Y^\epsilon(t_0) = y_0$ and for $s \in (t, T]$ it is the unique strong solution of (3.4) with $\bar{Z}(\cdot) = (\bar{W}(\cdot), \bar{B}(\cdot))$ in place of $Z(\cdot) = (W(\cdot), B(\cdot))$ and $\bar{u}(s, x, y; i) = (\bar{u}_1(s, x, y; i), \bar{u}_2(s, x, y; i))$ in place of $u(s) = (u_1(s), u_2(s))$.

Letting

$$\Gamma^\epsilon(t_0, x_0, y_0) = \exp \left\{ -\frac{1}{\epsilon} h(X^\epsilon(T)) \right\} \frac{dP}{d\bar{P}^\epsilon}(X^\epsilon, Y^\epsilon),$$

it follows easily that under $\bar{P}^\epsilon, \Gamma^\epsilon(t_0, x_0, y_0)$ is an unbiased estimator for $\theta(\epsilon)$. The performance of this estimator is characterized by the decay rate of its second moment

$$Q^\epsilon(t_0, x_0, y_0; \bar{u}) \doteq \bar{E} \left[\exp \left\{ -\frac{2}{\epsilon} h(X^\epsilon(T)) \right\} \left(\frac{dP}{d\bar{P}^\epsilon}(X^\epsilon, Y^\epsilon) \right)^2 \right]. \quad (4.3)$$

We construct asymptotically efficient importance sampling schemes by choosing the control \bar{u} in (4.3) such that the behavior of the second moment $Q^\epsilon(t_0, x_0, y_0; \bar{u})$ is controlled. Two are the main ingredients in the construction of \bar{u} :

- (i) The gradient of a subsolution to the PDE that the function $G_i(t, x)$ defined in (4.1) satisfies. Under appropriate regularity conditions $G_i(t, x)$ satisfies a PDE of Hamilton-Jacobi-Bellman (HJB) type.
- (ii) The solution to the associated cell problem or in other words the so-called corrector from the homogenization theory of HJB equations.

Depending on the regime of interaction the HJB equation and the corresponding cell problem take a different form. These will be made precise in Sects. 4.1–4.3.

As mentioned before, we work with appropriate subsolutions to the associated HJB equation. Thus, let us now recall the notion of a subsolution to an HJB equation of the form

$$G_s(s, x) + \bar{H}(x, \nabla_x G(s, x)) = 0, \quad G(T, x) = h(x). \quad (4.4)$$

Definition 4.1 A function $\bar{U}(s, x) : [0, T] \times \mathbb{R}^m \mapsto \mathbb{R}$ is a classical subsolution to the HJB (4.4) if

- (i) \bar{U} is continuously differentiable,
- (ii) $\bar{U}_s(s, x) + \bar{H}(x, \nabla_x \bar{U}(s, x)) \geq 0$ for every $(s, x) \in (0, T) \times \mathbb{R}^m$,
- (iii) $\bar{U}(T, x) \leq h(x)$ for $x \in \mathbb{R}^m$.

We will impose stronger regularity conditions on the subsolutions to be considered than those of Definition 4.1. This is convenient for the purposes of illustrations since then the feedback control is uniformly bounded and thus several technical problems are avoided. However, we mention that the uniform bounds that will be assumed in Condition 4.2 can be replaced by milder conditions with the expense of working harder to establish the results.

Condition 4.2 \bar{U} has continuous derivatives up to order 1 in t and order 2 in x , and the first and second derivatives in x are uniformly bounded.

Roughly speaking, our main result is as follows.

Theorem 4.3 Consider a bounded and continuous function $h : \mathbb{R}^m \mapsto \mathbb{R}$ and assume Conditions 2.1 and under Regime 1 assume Condition 2.2. Let $\{(X^\epsilon(s), Y^\epsilon(s)), \epsilon > 0\}$ be the solution to (1.1) for $s \in [t_0, T]$ with initial point (x_0, y_0) at time t_0 . Under Regime $i = 1, 2, 3$ let $\bar{u}(s, x, y; i)$ be an appropriately defined and smooth control in terms of a subsolution $\bar{U}_i(s, x)$ to the HJB satisfied by $G_i(s, x)$ and the corrector from the corresponding cell problem. Then

$$\liminf_{\epsilon \rightarrow 0} -\epsilon \ln Q^\epsilon(t_0, x_0, y_0; \bar{u}(\cdot; i)) \geq G_i(t_0, x_0) + \bar{U}_i(t_0, x_0). \tag{4.5}$$

Once we have established a Theorem like 4.3, we can make a claim for estimating probabilities of the form $P_{t,x_0,y_0}[X^\epsilon(T) \in A]$ as well. The claim of the following proposition is not readily covered by Theorem 4.3, since the function h is neither bounded nor continuous. However, by an approximating argument analogous to [15] the claim can be established. We omit the details of the proof and only present the statement.

Proposition 4.4 Assume Conditions 2.1 and 4.2 and under Regime 1 assume Condition 2.2. Let $\{(X^\epsilon, Y^\epsilon), \epsilon > 0\}$ be the solution to (1.1) with initial point (t_0, x_0, y_0) . Under Regime i , let $A \subset \mathbb{R}^m$ be a regular set with respect to the action functional S^i and the initial point (t_0, x_0, y_0) , i.e., the infimum of S^i over the closure \bar{A} is the same as the infimum over the interior A° . Let

$$h(x) = \begin{cases} 0 & \text{if } x \in A \\ +\infty & \text{if } x \notin A. \end{cases}$$

Let $\bar{u}(s, x, y; i)$ be an appropriately defined and smooth control as in Theorem 4.3. Then (4.5) holds.

Notice that the lower asymptotic bound of Theorem 4.3 and Proposition 4.4 is independent of the initial point y_0 of the fast component Y^ϵ . This is due to averaging.

Remark 4.5 Since \bar{U}_i is a subsolution, we get that $\bar{U}_i(s, x) \leq G_i(s, x)$ everywhere. By (4.5) this implies that the scheme is asymptotically optimal if $\bar{U}_i(t_0, x_0) = G_i(t_0, x_0)$ at the starting point (t_0, x_0) . Standard Monte Carlo corresponds to choosing the subsolution $\bar{U}_i = 0$. Hence, any subsolution with value at the origin (t_0, x_0) such that

$$0 \ll \bar{U}_i(t_0, x_0) \leq G_i(t_0, x_0)$$

will have better asymptotic performance than that of standard Monte Carlo.

In the next subsections we present how one can choose the controls $\bar{u}(s, x, y; i)$ in terms of a subsolution \bar{U} and its corresponding cell problem such that the bound mentioned in Theorem 4.3 is attained. The situation is subtle here due to the multiscale aspect of the problem.

4.1 Importance Sampling for Regime 1

In this subsection we construct asymptotically efficient importance sampling schemes for Regime 1. In Regime 1, the form of the Hamiltonian $\bar{H}(x, p)$ in (4.4) is naturally suggested by the calculus of variation problem (4.1) and the explicit formula of the rate function $S_{iT}^1(\phi)$ in Theorem 3.4:

$$\bar{H}(x, p) = \langle r(x), p \rangle - \frac{1}{2} \langle p, q(x)p \rangle. \tag{4.6}$$

In fact, under mild conditions G_1 from (4.1) is the unique viscosity solution to (4.4) with $\bar{H}(x, p)$ defined by (4.6).

We have the following Theorem.

Theorem 4.6 *Let $\{(X^\epsilon(s), Y^\epsilon(s)), \epsilon > 0\}$ be the solution to (1.1) for $s \in [t_0, T]$ with initial point (x_0, y_0) at time t_0 . Consider a bounded and continuous function $h : \mathbb{R}^m \mapsto \mathbb{R}$ and assume Conditions 2.1, 2.2 and 4.2. Let $\bar{U}_1(s, x)$ be a subsolution to the associated HJB equation. Define the feedback control $\bar{u}(s, x, y; 1) = (\bar{u}_1(s, x, y; 1), \bar{u}_2(s, x, y; 1))$ by*

$$\bar{u}(s, x, y; 1) = \left(-\left(\sigma + \frac{\partial \chi}{\partial y} \tau_1 \right)^T (x, y) \nabla_x \bar{U}_1(s, x), -\left(\frac{\partial \chi}{\partial y} \tau_2 \right)^T (x, y) \nabla_x \bar{U}_1(s, x) \right).$$

Then the conclusion of Theorem 4.3 holds, i.e.

$$\liminf_{\epsilon \rightarrow 0} -\epsilon \ln Q^\epsilon(t_0, x_0, y_0; \bar{u}(\cdot; 1)) \geq G_1(t_0, x_0) + \bar{U}_1(t_0, x_0).$$

Before proceeding with the proof, we notice that the feedback control (4.6) is essentially implied by the solution to the variational problem associated with the local rate function in the definition of the action functional for Regime 1, Theorem 3.7.

Proof Note that under the given conditions $\bar{u}(s, x, y; 1)$ is Lipschitz continuous in (x, y) , continuous in (t, x, y) , and uniformly bounded. For notational convenience, we omit the subscript 1 from G_1 and \bar{u}_1 and we write (t, x, y) in place of (t_0, x_0, y_0) .

Boundedness of h and \bar{u} imply by the representation formula (3.3) and by the Lemma 4.3 of [13] that

$$\begin{aligned}
 & -\epsilon \log Q^\epsilon(t, x, y; \bar{u}) \\
 &= \inf_{v \in \mathcal{A}} \mathbb{E} \left[\frac{1}{2} \int_t^T \|v(s)\|^2 ds - \int_t^T \|\bar{u}(s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s))\|^2 ds + 2h(\hat{X}^\epsilon(T)) \right],
 \end{aligned}
 \tag{4.7}$$

where $v(s) = (v_1(s), v_2(s))$, $\bar{u}(s, x, y; 1) = (\bar{u}_1(s, x, y; 1), \bar{u}_2(s, x, y; 1))$ and (\hat{X}, \hat{Y}) satisfying

$$\begin{aligned}
 d\hat{X}^\epsilon(s) &= \left[\frac{\epsilon}{\delta} b(\hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) + \bar{c}(\hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) + \sigma(\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon) v_1(s) \right] ds \\
 &\quad + \sqrt{\epsilon} \sigma(\hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) dW(s), \\
 d\hat{Y}^\epsilon(s) &= \frac{1}{\delta} \left[\frac{\epsilon}{\delta} f(\hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) + \bar{g}(\hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) + \tau_1(\hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) v_1(s) \right. \\
 &\quad \left. + \tau_2(\hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) v_2(s) \right] ds + \frac{\sqrt{\epsilon}}{\delta} [\tau_1(\hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) dW(s) \\
 &\quad + \tau_2(\hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) dB(s)],
 \end{aligned}
 \tag{4.8}$$

$$\hat{X}^\epsilon(0) = x_0, \quad \hat{Y}^\epsilon(0) = y_0$$

with

$$\begin{aligned}
 \bar{c}(s, x, y) &= c(x, y) - \sigma(x, y) \bar{u}_1(s, x, y; 1) \\
 \bar{g}(s, x, y) &= g(x, y) - \tau_1(x, y) \bar{u}_1(s, x, y; 1) - \tau_2(x, y) \bar{u}_2(s, x, y; 1).
 \end{aligned}
 \tag{4.9}$$

The next step is to take the limit infimum in the representation (4.7). The right hand side of (4.7) can be bounded by below in the limit $\epsilon \downarrow 0$ using statement (ii) of Theorem 3.3 with two differences. The first difference is that the functions c, g in the definition of the first component of appropriate viable pair $(\lambda_1, \mathcal{L}^1)$, see Definition 2.4, are replaced with \bar{c}, \bar{g} . Using Theorem 3.7, the local rate function takes the form

$$L_1(x, \beta) = \frac{1}{2} (\beta - \bar{r}(s, x))^T q^{-1}(x) (\beta - \bar{r}(s, x))$$

where $\bar{r}(s, x) = r(x) - \int_{\mathcal{Y}} (\sigma \bar{u}_1 + \frac{\partial \chi}{\partial y} \tau_1 \bar{u}_1 + \frac{\partial \chi}{\partial y} \tau_2 \bar{u}_2)(s, x, y) \mu(dy|x)$. Here $\mu(dy|x)$ is the invariant measure defined in Condition 2.2. This takes care of the limit of first term on the right hand side of (4.7). The second difference is the presence of the additional integral term $-\int_t^T \|\bar{u}(s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s))\|^2 ds$. Using classical averaging arguments, see [6], appropriately modified to treat controlled processes, as in Lemma 3.2

in [12], this term can be replaced in the limit as $\epsilon \downarrow 0$ by its averaged version with respect to $\mu(dy|x)$. This takes care of the limit of second term on the right hand side of (4.7). Putting these together, we have

$$\begin{aligned} & \liminf_{\epsilon \rightarrow 0} -\epsilon \log Q^\epsilon(t, x; \bar{u}) \\ & \geq \inf_{\phi, \phi(t)=x} \left[\int_t^T L_1(\phi(s), \dot{\phi}(s)) ds - \int_t^T \int_Y [\|\bar{u}_1(s, \phi(s), y)\|^2 \right. \\ & \quad \left. + \|\bar{u}_2(s, \phi(s), y)\|^2] \mu(dy|\phi(s)) ds + 2h(\phi(T)) \right]. \end{aligned} \tag{4.10}$$

By recalling the formula for $\bar{u} = (\bar{u}_1, \bar{u}_2)$ we have for $\phi \in \mathcal{AC}([t, T]; \mathbb{R}^m)$

$$\begin{aligned} & \int_t^T \int_Y [\|\bar{u}_1(s, \phi(s), y)\|^2 + \|\bar{u}_2(s, \phi(s), y)\|^2] \mu(dy|\phi(s)) ds \\ & = \int_t^T \langle \nabla_x \bar{U}(s, \phi(s)), q(x) \nabla_x \bar{U}(s, \phi(s)) \rangle ds. \end{aligned}$$

Thus, we have

$$\begin{aligned} & \liminf_{\epsilon \rightarrow 0} -\epsilon \log Q^\epsilon(t, x; \bar{u}) \\ & \geq \inf_{\phi, \phi(t)=x} \left[\frac{1}{2} \int_t^T \|\dot{\phi}(s) - r(\phi(s))\| \right. \\ & \quad - \int_Y \left(\sigma \bar{u}_1 + \frac{\partial \chi}{\partial y} \tau_1 \bar{u}_1 + \frac{\partial \chi}{\partial y} \tau_2 \bar{u}_2 \right) (s, \phi(s), y) \mu(dy|\phi(s)) \Big\|_{q^{-1}(\phi(s))}^2 ds \\ & \quad \left. - \int_t^T \int_Y [\|\bar{u}_1(s, \phi(s), y)\|^2 + \|\bar{u}_2(s, \phi(s), y)\|^2] \mu(dy|\phi(s)) ds + 2h(\phi(T)) \right] \\ & = \inf_{\phi \in \mathcal{AC}([t, T]; \mathbb{R}^m), \phi(t)=x} \left[\int_t^T \left[\frac{1}{2} \|\dot{\phi}(s) - r(\phi(s))\|_{q^{-1}(\phi(s))}^2 \right. \right. \\ & \quad \left. \left. - \langle \dot{\phi}(s) - r(\phi(s)), \nabla_x \bar{U}(s, \phi(s)) \rangle \right] ds \right. \\ & \quad \left. - \frac{1}{2} \int_t^T \langle \nabla_x \bar{U}(s, \phi(s)), q(x) \nabla_x \bar{U}(s, \phi(s)) \rangle ds + 2h(\phi(T)) \right] \\ & = \inf_{\phi \in \mathcal{AC}([t, T]; \mathbb{R}^m), \phi(t)=x} \left[S_{tT}^1(\phi) + 2h(\phi(T)) \right. \\ & \quad \left. - \int_t^T \langle \dot{\phi}(s) - r(\phi(s)), \nabla_x \bar{U}(s, \phi(s)) \rangle \right. \\ & \quad \left. + \frac{1}{2} \langle \nabla_x \bar{U}(s, \phi(s)), q(x) \nabla_x \bar{U}(s, \phi(s)) \rangle ds \right]. \end{aligned} \tag{4.11}$$

In the first equality we have used the definition of $\bar{u} = (\bar{u}_1, \bar{u}_2)$ whereas in the second equality we used the definition of the action functional by Theorem 3.4.

Given an arbitrary $\phi \in \mathcal{AC}([t, T]; \mathbb{R}^m)$ with $\phi(t) = x$, the subsolution property implies that

$$\begin{aligned} & -\langle \dot{\phi}(s) - r(\phi(s)), \nabla_x \bar{U}(s, \phi(s)) \rangle - \frac{1}{2} \langle \nabla_x \bar{U}(s, \phi(s)), q(\phi(s)) \nabla_x \bar{U}(s, \phi(s)) \rangle \\ & - \partial_t \bar{U}(s, \phi(s)) - \langle \nabla_x \bar{U}(s, \phi(s)), \dot{\phi}(s) \rangle \\ & = -\frac{d}{ds} \bar{U}(s, \phi(s)). \end{aligned}$$

Let us now integrate both sides on $[t, T]$. Using the terminal condition $\bar{U}(T, x) \leq h(x)$, we have

$$\begin{aligned} & -\int_t^T (\langle \dot{\phi}(s) - r(\phi(s)), \nabla_x \bar{U}(s, \phi(s)) \rangle \\ & + \frac{1}{2} \langle \nabla_x \bar{U}(s, \phi(s)), q(\phi(s)) \nabla_x \bar{U}(s, \phi(s)) \rangle) ds \geq -h(\phi(T)) + \bar{U}(t, x). \end{aligned}$$

Thus, the right hand side of (4.11) is bounded from below by

$$\inf_{\phi \in \mathcal{AC}([t, T]; \mathbb{R}^m), \phi(t)=x} [S_{tT}^1(\phi) + h(\phi(T))] + \bar{U}(t, x).$$

Thus, since by definition $G(t, x) = \inf_{\phi \in \mathcal{AC}([t, T]; \mathbb{R}^m), \phi(t)=x} [S_{tT}^1(\phi) + h(\phi(T))]$ we can conclude that

$$\liminf_{\epsilon \rightarrow 0} -\epsilon \log Q^\epsilon(t, x; \bar{u}) \geq G(t, x) + \bar{U}(t, x).$$

This concludes the proof. □

4.2 Importance Sampling for Regime 2

Let us now study the construction of efficient importance samplings for Regime 2. The situation here is more subtle than it is for Regime 1. This is also seen from the large deviations principle, Theorem 3.6. The key difference between the LDP for Regimes 1 and 2 is that $\mathcal{L}_{z_1, z_2, x}^2$ depends on (z_1, z_2) , while \mathcal{L}_x^1 did not. This means that relations between the elements of a viable pair are more complex, and in particular that the joint distribution of the control (z_1, z_2) and fast variable y is important. Thus, in contrast to Regime 1 where the action functional can be written down explicitly, in the case of Regime 2 the formula of the action functional is in terms of value function to a variational problem.

This implicit characterization partially carries over to the importance sampling. The optimal control is again in terms of a corresponding cell problem as it was for Regime 1 (recall the cell problem (2.1) for Regime 1). The difference here is that the cell problem is defined implicitly rather than explicitly. As we discuss in Sect. 5, this is related to the homogenization theory of HJB equations.

In what follows, the subscript γ is to emphasize the dependence on γ (see (1.3)). Define

$$\begin{aligned}
 H_\gamma(x, y, p, q, P, Q, R) &= \inf_{u_1, u_2 \in \mathcal{Z}} \left[\frac{1}{2} \sigma \sigma^T : P + \gamma \frac{1}{2} (\tau_1 \tau_1^T + \tau_2 \tau_2^T) : Q \right. \\
 &\quad + \gamma \tau_1 \sigma^T : R + \langle \gamma b + c + \sigma u_1, p \rangle \\
 &\quad \left. + \langle \gamma f + g + \tau_1 u_1 + \tau_2 u_2, q \rangle + \frac{1}{2} \|u_1\|^2 + \frac{1}{2} \|u_2\|^2 \right] \\
 &= \frac{1}{2} \sigma \sigma^T : P + \gamma \frac{1}{2} (\tau_1 \tau_1^T + \tau_2 \tau_2^T) : Q \\
 &\quad + \gamma \tau_1 \sigma^T : R + \langle \gamma b + c, p \rangle \\
 &\quad + \langle \gamma f + g, q \rangle - \frac{1}{2} \|\sigma^T p + \tau_1^T q\|^2 - \frac{1}{2} \|\tau_2^T q\|^2.
 \end{aligned}
 \tag{4.12}$$

The infimum in (4.12) is attained for

$$u_1 = -\sigma^T(x, y)p - \tau_1^T(x, y)q \quad \text{and} \quad u_2 = -\tau_2^T(x, y)q.
 \tag{4.13}$$

The control $u = (u_1, u_2)$ motivates the asymptotically optimal change of measure in Theorem 4.8. Let us now define the associated HJB equation of interest together with the associated cell problem. We start with the cell problem. For each fixed (x, p) consider the unique value $\bar{H}_\gamma(x, p)$ such that there is a periodic solution ξ to the cell problem

$$H_\gamma(x, y, p, \nabla_y \xi_\gamma, 0, \nabla_y^2 \xi_\gamma, 0) = \bar{H}_\gamma(x, p).
 \tag{4.14}$$

The unknown in (4.14) is the pair $(\xi_\gamma, \bar{H}_\gamma)$. As it can be obtained by Theorem II.2 in [2], ξ_γ is the unique (up to an additive constant) periodic solution to (4.14) such that $\xi_\gamma \in C^2(\mathbb{R}^{d-m})$. Moreover, $\bar{H}_\gamma(x, p)$ is continuous in x and concave in p (see Propositions 11 and 12 in [1]).

Consider then the HJB equation (4.4) with $\bar{H}(x, p)$ replaced by $\bar{H}_\gamma(x, p)$. Under the standing assumptions, this HJB equation has a unique viscosity solution which we denote by $G_2(s, x)$. Actually, under mild conditions the value function of the variational problem (4.1) is this unique viscosity solution. This can be derived as in [1] and it will be recalled in Sect. 5.

In accordance to what we did for Regime 1, we consider a classical subsolution to that HJB equation, which we denote by $\bar{U}_2(s, x)$, where the Hamiltonian is $\bar{H}_\gamma(x, p)$. Notice that ξ_γ depends on the triple (x, y, p) with (x, p) seen as parameters. In the computations p will be substituted by the gradient of the subsolution $\nabla_x \bar{U}_2(s, x)$. So, in principle ξ_γ and $\bar{U}_2(s, x)$ are coupled. This coupling is in line with the coupling that appears in large deviations, see Theorem 3.6.

Similarly to what we did for Regime 1, we impose stronger regularity conditions. This is done to ease exposition. In particular, the following condition guarantees the feedback control used in importance sampling is uniformly bounded and that we can apply Itô formula directly without approximations. Thus, a number of technicalities are circumvented.

Condition 4.7 \bar{U}_2 has continuous derivatives up to order 1 in t and order 2 in x , and the first and second derivatives in x are uniformly bounded. Similarly, ξ_γ is twice continuous differentiable in (x, y, p) , periodic with respect to y and all of the mixed derivatives up to order 2 are bounded.

The following verification theorem is the analogous of Theorem 4.6 for Regime 2.

Theorem 4.8 Let $\{(X^\epsilon(s), Y^\epsilon(s)), \epsilon > 0\}$ be the solution to (1.1) for $s \in [t_0, T]$ with initial point (x_0, y_0) at time t_0 . Assume that we are considering Regime 2. Consider a bounded and continuous function $h : \mathbb{R}^m \mapsto \mathbb{R}$ and assume Conditions 2.1. Let $\xi_\gamma(x, y, p)$ be the unique (up to a constant) periodic solution to the cell problem (4.14) and $\bar{U}_2(s, x)$ be a classical subsolution according to Definition 4.1 and assume Condition 4.7. Define the control $\bar{u}(s, x, y; 2) = (\bar{u}_1(s, x, y; 2), \bar{u}_2(s, x, y; 2))$ by

$$\begin{aligned} \bar{u}(s, x, y; 2) = & (-\sigma^T(x, y)\nabla_x \bar{U}_2(s, x) - \tau_1^T(x, y)\nabla_y \xi_\gamma(x, y, \nabla_x \bar{U}_2(s, x)), \\ & -\tau_2^T(x, y)\nabla_y \xi_\gamma(x, y, \nabla_x \bar{U}_2(s, x))) \end{aligned}$$

Then the conclusion of Theorem 4.3 holds, i.e.

$$\liminf_{\epsilon \rightarrow 0} -\epsilon \ln Q^\epsilon(t_0, x_0, y_0; \bar{u}(\cdot; 2)) \geq G_2(t_0, x_0) + \bar{U}_2(t_0, x_0).$$

Proof For notational convenience, we omit the subscripts 2 and γ from $G_2, \bar{U}_2, \xi_\gamma$ and \bar{H}_γ . Also we write $\xi(s, x, y)$ in place of $\xi(x, y, \nabla_x \bar{U}(s, x))$ and (t, x, y) in place of (t_0, x_0, y_0) for the initial point.

The first step is to write, as in Regime 1, that

$$\begin{aligned} & -\epsilon \log Q^\epsilon(t, x, y; \bar{u}(\cdot; 2)) \\ & = \inf_{v \in \mathcal{A}} \mathbb{E} \left[\frac{1}{2} \int_t^T \|v(s)\|^2 ds - \int_t^T \|\bar{u}(s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s); 2)\|^2 ds + 2h(\hat{X}^\epsilon(T)) \right], \end{aligned} \tag{4.15}$$

where (\hat{X}, \hat{Y}) satisfies (4.8) with $v(s) = (v_1(s), v_2(s))$ and $\bar{u}(s, x, y; 2) = (\bar{u}_1(s, x, y; 2), \bar{u}_2(s, x, y; 2))$.

The next step is to rewrite the right hand side of (4.15). Recall the definition of the operator $\mathcal{L}_{z,x}^2$ from Definition 2.3 with $z = (z_1, z_2)$. Denote by $\mathcal{L}_{z,x}^{\epsilon/\delta, 2}$ the operator $\mathcal{L}_{z,x}^2$ with $\frac{\epsilon}{\delta}$ in place of γ . We will write $\mathcal{L}_{0,x}^{\epsilon/\delta, 2}$ to denote the operator with the control variable $z = 0$.

Apply Itô formula to $\xi(s, x, y)$. After some term rearrangement, we get

$$\begin{aligned} & - \int_t^T \mathcal{L}_{0, \hat{X}^\epsilon(s)}^{\epsilon/\delta, 2} \xi(s, \hat{X}^\epsilon(s), \bar{Y}^\epsilon(s)) ds \\ & = \int_t^T (\nabla_y \xi, \tau_1(v_1 - \bar{u}_1) + \tau_2(v_2 - \bar{u}_2))(s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) ds + R_1(\epsilon, v) \end{aligned} \tag{4.16}$$

where the random variable $R_1(\epsilon, v)$ is

$$\begin{aligned}
 R_1(\epsilon, v) = & \delta \left[\xi(x, x/\delta, t) - \xi(T, \hat{X}^\epsilon(T), \hat{Y}^\epsilon(T)) + \int_t^T \partial_t \xi(s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) ds \right] \\
 & + \epsilon \int_t^T \tau_1^T \sigma : \nabla_x \nabla_y \xi(s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) ds \\
 & + \delta \int_t^T \left[\left\langle \frac{\epsilon}{\delta} b + c + \sigma(v_1(s) - \bar{u}_1), \nabla_x \xi \right\rangle (s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) \right. \\
 & \left. + \frac{\epsilon}{2} \sigma \sigma^T : \nabla_{xx} \xi(s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) \right] ds \\
 & + \sqrt{\epsilon} \delta \int_t^T \langle \nabla_x \xi, \sigma dW(s) \rangle (s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) + \sqrt{\epsilon} \int_t^T \langle \nabla_y \xi, \tau_1 dW(s) \\
 & + \tau_2 dB(s) \rangle (s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)).
 \end{aligned}$$

Under our assumptions, the random variable $R_1(\epsilon, v)$ converges in L^2 to zero as $\epsilon, \delta \downarrow 0$ uniformly in $v \in \mathcal{A}$.

Next, we apply Itô formula to $\bar{U}(t, x)$. Omitting some function arguments for notational convenience and using the subsolution property for \bar{U} , we get

$$\begin{aligned}
 h(\hat{X}^\epsilon(T)) \geq & \bar{U}(t, x) + \int_t^T \left[-\bar{H}(\hat{X}^\epsilon(s), \nabla_x \bar{U}) \right. \\
 & \left. + \left\langle \nabla_x \bar{U}, \frac{\epsilon}{\delta} b + c + \sigma(v_1(s) - \bar{u}_1) \right\rangle \right] (s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) ds \\
 & + \frac{\epsilon}{2} \int_t^T \sigma \sigma^T (\hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) : \nabla_x \nabla_x \bar{U}(s, \hat{X}^\epsilon(s)) ds \\
 & + \sqrt{\epsilon} \int_t^T \langle \nabla_x \bar{U}(s, \hat{X}^\epsilon(s)), \sigma(\hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) dW(s) \rangle. \tag{4.17}
 \end{aligned}$$

Recalling the definition of \bar{H} by (4.14) and adding and subtracting the term $\int_t^T \mathcal{L}_{0, \hat{X}^\epsilon(s)}^{\epsilon/\delta, 2} \xi(s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) ds$, relation (4.17) becomes, after using (4.16),

$$\begin{aligned}
 h(\hat{X}^\epsilon(T)) - \bar{U}(t, x) \geq & (\epsilon/\delta - \gamma) \int_t^T \langle \nabla_x \bar{U}(s, \hat{X}^\epsilon(s)), b(\hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) \rangle ds \\
 & + \int_t^T \langle \nabla_x \bar{U}(s, \hat{X}^\epsilon(s)), \sigma(v_1(s) - \bar{u}_1)(s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) \rangle ds \\
 & + \int_t^T \langle \nabla_y \xi, \tau_1(v_1(s) - \bar{u}_1) + \tau_2(v_2(s) - \bar{u}_2) \rangle ds \\
 & + \frac{1}{2} \int_t^T \|\bar{u}_1(s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s); 2)\|^2 ds
 \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{2} \int_t^T \|\bar{u}_2(s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s); 2)\|^2 ds \\
 & + R_1(\epsilon, v) + R_2(\epsilon, v)
 \end{aligned} \tag{4.18}$$

where $R_1(\epsilon, v)$ was defined before and $R_2(\epsilon, v)$ is as follows

$$\begin{aligned}
 R_2(\epsilon, v) = & \frac{\epsilon}{2} \int_t^T \sigma \sigma^T : \nabla_x \nabla_x \bar{U}(\hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) ds \\
 & + \sqrt{\epsilon} \int_t^T \langle \nabla_x \bar{U}(s, \hat{X}^\epsilon(s)), \sigma(\hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) dW(s) \rangle \\
 & + \int_t^T [\mathcal{L}_{0, \hat{X}^\epsilon(s)}^{\epsilon/\delta, 2} \xi(s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) - \mathcal{L}_{0, \hat{X}^\epsilon(s)}^2 \xi(s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s))] ds.
 \end{aligned}$$

Under our assumptions, the random variable $R_2(\epsilon, v)$ converges in L^2 to zero as $\epsilon, \delta \downarrow 0$ uniformly in $v \in \mathcal{A}$. Recalling the definitions of the controls \bar{u}_1, \bar{u}_2 we get

$$\begin{aligned}
 h(\hat{X}^\epsilon(T)) - \bar{U}(t, x) \geq & (\epsilon/\delta - \gamma) \int_t^T \langle \nabla_x \bar{U}(s, \hat{X}^\epsilon(s)), b(\hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) \rangle ds \\
 & - \int_t^T \langle \bar{u}_1, v_1(s) - \bar{u}_1 \rangle (s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) ds \\
 & - \int_t^T \langle \bar{u}_2, v_2(s) - \bar{u}_2 \rangle (s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) ds \\
 & + \frac{1}{2} \int_t^T \|\bar{u}_1(s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s); 2)\|^2 ds \\
 & + \frac{1}{2} \int_t^T \|\bar{u}_2(s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s); 2)\|^2 ds \\
 & + R_1(\epsilon, v) + R_2(\epsilon, v).
 \end{aligned}$$

Writing for notational convenience $\bar{u}_i(s) = \bar{u}_i(s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s); 2)$ for $i = 1, 2$, we get after some term rearrangement

$$\begin{aligned}
 & - \int_t^T [\|\bar{u}_1(s)\|^2 + \|\bar{u}_2(s)\|^2] ds \\
 & \geq \bar{U}(t, x) - h(\hat{X}^\epsilon(T)) + \frac{1}{2} \int_t^T [\|\bar{u}_1(s)\|^2 + \|\bar{u}_2(s)\|^2] ds \\
 & \quad - \int_t^T \langle \bar{u}_1(s), v_1(s) \rangle ds - \int_t^T \langle \bar{u}_2(s), v_2(s) \rangle ds + R(\epsilon, v)
 \end{aligned} \tag{4.19}$$

where $R(\epsilon, v) = R_1(\epsilon, v) + R_2(\epsilon, v) + (\epsilon/\delta - \gamma) \int_t^T \langle \nabla_x \bar{U}(s, \hat{X}^\epsilon(s)), b(\hat{X}^\epsilon(s), \hat{Y}^\epsilon(s)) \rangle ds$. Since $\frac{\epsilon}{\delta} \rightarrow \gamma$ and $R_1(\epsilon, v), R_2(\epsilon, v)$ converge in L^2 to zero as $\epsilon, \delta \downarrow 0$, Condition 4.7 implies that $R(\epsilon, v)$ converges in L^2 to zero uniformly in $v \in \mathcal{A}$.

Inserting (4.19) into (4.15) gives us

$$-\epsilon \ln Q^\epsilon(t, x; \bar{u}) \geq \inf_{v \in \mathcal{A}} E_{t,x,y}^\epsilon \left[\frac{1}{2} \int_t^T \|v(s) - \bar{u}(s, \hat{X}^\epsilon(s), \hat{Y}^\epsilon(s))\|^2 ds + h(\bar{X}^{\epsilon, v-\bar{u}}(T)) + \bar{U}(t, x) + R(\epsilon, v) \right].$$

Set $\bar{v}(s) = v(s) - \bar{u}(s, \hat{X}(s), \hat{Y}(s))$. Since $\bar{v} \in \mathcal{A}$, the representation formula (3.3) implies that

$$E \left[\frac{1}{2} \int_t^T \|\bar{v}(s)\|^2 ds + h(\hat{X}(T)) \right] \geq -\epsilon \log E \exp \left\{ -\frac{1}{\epsilon} h(X^\epsilon(T)) \right\}.$$

Recalling that $R(\epsilon, v)$ converges in L^2 to zero uniformly in $v \in \mathcal{A}$ as $\epsilon, \delta \downarrow 0$ and using statement (ii) of Theorem 3.3 we get

$$\begin{aligned} \liminf_{\epsilon \rightarrow 0} -\epsilon \log Q^\epsilon(t, x; \bar{u}) &\geq \liminf_{\epsilon \rightarrow 0} \inf_{\bar{v} \in \mathcal{A}} E \left[\frac{1}{2} \int_t^T \|\bar{v}(s)\|^2 ds + h(\hat{X}(T)) + R(\epsilon, v) \right] \\ &\quad + \bar{U}(t, x) \\ &\geq \liminf_{\epsilon \rightarrow 0} -\epsilon \log E \exp \left\{ -\frac{1}{\epsilon} h(X^\epsilon(T)) \right\} + \bar{U}(t, x) \\ &= G(t, x) + \bar{U}(t, x). \end{aligned} \tag{4.20}$$

This concludes the proof of the theorem. □

We conclude this subsection with the following remark. This remark relaxes the requirement of a solution pair $(\xi_\gamma(x, y, p), \bar{H}_\gamma(x, p))$ to the cell problem (4.14) to a subsolution pair. This can be useful in problems where solving the cell problem is difficult even numerically.

Remark 4.9 In the proof of the theorem, the definition of the cell problem (4.14) was only used in (4.18). However, it is easy to see that the inequality in (4.18) would be true if instead of the solution pair to the cell problem, a subsolution pair was used, i.e. a pair $(\xi_\gamma(x, y, p), \bar{H}_\gamma(x, p))$ such that $H_\gamma(x, y, p, \nabla_y \xi_\gamma, 0, \nabla_y^2 \xi_\gamma, 0) \geq \bar{H}_\gamma(x, p)$ for all $y \in \mathcal{Y}$ and $(x, p) \in \mathbb{R}^m \times \mathbb{R}^m$. So, one can seek for subsolution pairs $(\xi_\gamma(x, y, p), \bar{H}_\gamma(x, p))$ to (4.14) such that $\xi_\gamma(x, y, p)$ is periodic in y and $\bar{H}_\gamma(x, p)$ is concave in p .

4.3 Importance Sampling for Regime 3

Finally, we study the construction of efficient importance samplings for Regime 3. The procedure here is similar to that of Regime 2. This is to be expected, since Regime 3 is a limiting case of Regime 2 obtained by setting $\gamma = 0$. Therefore, we shall only present the result omitting the proof, which follows as the proof of Theorem 4.8 for Regime 2. The statement for the existence and regularity of a pair $(\xi_0(x, y, p), \bar{H}_0(x, p))$ satisfying (4.14) with $\gamma = 0$ is given in Sect. 5.

Theorem 4.10 Let $\{(X^\epsilon(s), Y^\epsilon(s)), \epsilon > 0\}$ be the solution to (1.1) for $s \in [t_0, T]$ with initial point (x_0, y_0) at time t_0 . Consider a bounded and continuous function $h : \mathbb{R}^m \mapsto \mathbb{R}$ and assume Conditions 2.1. Let $(\xi_0(x, y, p), \bar{H}_0(x, p))$ be a pair satisfying the cell problem (4.14) with $\gamma = 0$ and $\bar{U}_3(s, x)$ be a classical subsolution according to Definition 4.1 with Hamiltonian $\bar{H}_0(x, p)$ and assume Condition 4.7 with $\gamma = 0$. Define the control $\bar{u}(s, x, y; 3) = (\bar{u}_1(s, x, y; 3), \bar{u}_2(s, x, y; 3))$ by

$$\bar{u}(s, x, y; 3) = \left(-\sigma^T(x, y) \nabla_x \bar{U}_3(s, x) - \tau_1^T(x, y) \nabla_y \xi_0(x, y, \nabla_x \bar{U}_3(s, x)), \right. \\ \left. -\tau_2^T(x, y) \nabla_y \xi_0(x, y, \nabla_x \bar{U}_3(s, x)) \right)$$

Then the conclusion of Theorem 4.3 holds, i.e.

$$\liminf_{\epsilon \rightarrow 0} -\epsilon \ln Q^\epsilon(t_0, x_0, y_0; \bar{u}(\cdot; 3)) \geq G_3(t_0, x_0) + \bar{U}_3(t_0, x_0).$$

Notice here that even though in the statement for the large deviations for Regime 3 (Theorem 3.3), we require that $g(x, y) = g(y)$ and $\tau_i(x, y) = \tau_i(y)$, in the statement of the related importance sampling lower bound we do not require that assumption. The reason is that in the proof of the importance sampling bound only the Laplace principle lower bound is used (compare with (4.20)) and that holds with the x -dependence as well; see the second statement of Theorem 3.3.

5 Connection with Homogenization of Hamilton-Jacobi-Bellman Equations

It is evident from the calculations in Sect. 4 that there is an implied relation of importance sampling for multiscale problems and homogenization of a related class of HJB equations. In this section we aim to make this connection clear. We only outline the results that are relevant to the importance sampling results. We refer the interested reader to the literature of homogenization for Hamilton-Jacobi-Bellman equations for more detailed discussions, e.g. [1, 2, 9, 17, 21, 27].

Let us define the function

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

where (X^ϵ, Y^ϵ) is the strong solution to the uncontrolled process (1.1) with initial point $(X^\epsilon(t), Y^\epsilon(t)) = (x, y)$. A straightforward computation shows that the function

$$G^\epsilon(t, x, y) = -\epsilon \ln \theta^\epsilon(t, x, y)$$

solves the Hamilton-Jacobi-Bellman equation

$$\partial_t G^\epsilon + H_{\epsilon/\delta} \left(x, y, \nabla_x G^\epsilon, \frac{\nabla_y G^\epsilon}{\delta}, \epsilon \nabla_x^2 G^\epsilon, \frac{\nabla_y^2 G^\epsilon}{\delta}, \nabla_x \nabla_y G^\epsilon \right) = 0 \\ G^\epsilon(T, x, y) = h(x)$$

where the Hamiltonian $H_{\epsilon/\delta}$ is defined as in (4.12) with ϵ/δ in place of γ . Under Conditions 2.1 and 2.2 we have the following.

- In the case of *Regime 1*, we have that $G^\epsilon(t, x, y)$ converges uniformly in compact subsets of $[0, T] \times \mathbb{R}^m \times \mathbb{R}^{d-m}$ to the unique bounded and continuous viscosity solution of (4.4) with effective Hamiltonian given by (4.6). We refer the reader to [1, 9] for details.
- In the case of *Regime 2* the effective equation has again the form (4.4) but the effective Hamiltonian is given by the unique constant \bar{H}_γ such that the periodic cell problem (4.14) has a unique (up to an additive constant) periodic solution $\xi_\gamma \in C^2(\mathbb{R}^{d-m})$ (see Theorem II.2 in [2]). Under our assumptions, the effective Hamiltonian $\bar{H}_\gamma(x, p)$ is continuous in x and concave in p (see Propositions 11 and 12 in [1]).
- In the case of *Regime 3* the effective equation has again the form (4.4) but the effective Hamiltonian is given by the unique constant \bar{H}_0 such that the periodic cell problem (4.14) with $\gamma = 0$ has a Lipschitz continuous periodic solution ξ_0 (see [1, 4]). Again, under our assumptions, the effective Hamiltonian $\bar{H}_0(x, p)$ is continuous in x and concave in p (see Propositions 3 in [1]).

In regards to how these general results apply to importance sampling for multiple scale problems, we have the following remark.

Remark 5.1 In the context of importance sampling, we observe two things:

- (i) the subsolutions that we are considering are subsolutions to the corresponding limiting HJB equations, and
- (ii) the cell problem arising in homogenization of HJB equations enters in the formulation of the importance sampling scheme in each regime.

These imply that in Monte Carlo simulation for multiscale problems both the local information described by the corresponding cell problem and the homogenized information that is described by the solution to the HJB equation, enter the asymptotically optimal change of measure. As it is demonstrated in the numerical simulations presented in [13], neglecting the local information and basing the simulation only on the homogenized information can lead to estimators that perform poorly in the small noise regime.

6 Examples

In this section we present some simple examples from the existing literature to illustrate how our calculations look like. We consider two examples. The first one is the first order Langevin equation. As we said in the introduction this model can be used to model rough energy landscapes motivated by applications in chemistry; see also [14, 26, 29, 32, 35]. This model was extensively discussed in [13, 14] and the theory was also demonstrated by simulation results. We recall the formulas here for completeness for this particularly important example. The second example is related to short time asymptotics for processes that depend on another fast mean reverting process. Models of this nature appear in mathematical finance in the context of fast mean reverting stochastic volatility models, e.g., [18]. Assuming that we want to estimate

$$\theta(\epsilon) = \mathbb{E}\left[e^{-\frac{1}{\epsilon}h(X^\epsilon(T))} \mid X^\epsilon(0) = x_0, Y^\epsilon(0) = y_0\right]$$

for a given function $h(x)$ and a given corresponding subsolution \bar{U} , we also provide the control that attains the desired bounds in Theorems 4.6, 4.8 and 4.10.

6.1 The First Order Langevin Equation

We consider the first order Langevin equation

$$dX^\epsilon(s) = \left[-\frac{\epsilon}{\delta} \nabla Q \left(\frac{X^\epsilon(s)}{\delta} \right) - \nabla V(X^\epsilon(s)) \right] dt + \sqrt{\epsilon} \sqrt{2D} dW(s), \quad (6.1)$$

$$X^\epsilon(0) = x_0.$$

To connect to the notation of the general model (1.1), this corresponds to

$$f(x, y) = b(x, y) = -\nabla Q(y), \quad g(x, y) = c(x, y) = -\nabla V(x),$$

$$\tau_1(x, y) = \sigma(x, y) = \sqrt{2D}, \quad \tau_2(x, y) = 0.$$

Let us consider the case of Regime 1. The invariant distribution associated to the operator \mathcal{L}^1 is the Gibbs distribution (independent of x)

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

Moreover, Condition 2.2 is trivially satisfied. In dimension 1, an easy computation shows that the action functional takes the following explicit form

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

where

$$r(x) = -\frac{\lambda^2 V'(x)}{L \hat{L}}, \quad q = \frac{2D\lambda^2}{L \hat{L}}$$

and

$$L = \int_{\mathcal{Y}} e^{-\frac{Q(y)}{D}} dy, \quad \hat{L} = \int_{\mathcal{Y}} e^{\frac{Q(y)}{D}} dy.$$

In addition, we can also compute the optimal change of measure in regards to the importance sampling problem. Given a classical subsolution \bar{U} , the importance sampling control that appears in Theorem 4.6 takes the form

$$\bar{u}(s, x, y; 1) = \left(-\frac{\sqrt{2D}\lambda}{\hat{L}} e^{\frac{Q(y)}{D}} \partial_x \bar{U}(s, x), 0 \right).$$

The choice of the subsolution \bar{U} according to Definition 4.1 depends on the terminal cost of interest $h(x)$. See also [13, 14] for some particular examples with specific choices of subsolutions $\bar{U}(s, x)$.

6.2 Short Time Asymptotics and Fast Mean Reversion

Next we consider a particular system of slow-fast motion, where the fast motion is a fast mean reverting process. The slow motion appears due to the interest in short time asymptotics. In particular, let us consider the system in 1 + 1 dimension

$$\begin{aligned} dX(s) &= h(Y(s)) ds + \sigma(Y(s)) dW(s), \\ dY(s) &= \frac{1}{\delta^2}(m - Y(s)) ds + \frac{1}{\delta}[\rho dW(s) + \sqrt{1 - \rho^2} dB(s)] \end{aligned} \tag{6.2}$$

where $0 < \delta \ll 1$ is the fast mean reversion parameter, $m \in \mathbb{R}$ and $\rho \in [-1, 1]$ is the correlation between the noise of the X and Y process. Assume that we are interested in short time asymptotics. Then it is convenient to change time $s \mapsto \epsilon s$ with $0 < \epsilon \ll 1$. Writing the system under the new timescale, we obtain $\{(X^\epsilon(s), Y^\epsilon(s)), s \in [0, T]\}$ as the unique strong solution to:

$$\begin{aligned} dX^\epsilon(s) &= \epsilon h(Y^\epsilon(s)) ds + \sqrt{\epsilon} \sigma(Y^\epsilon(s)) dW(s), \\ dY^\epsilon(s) &= \frac{\epsilon}{\delta^2}(m - Y^\epsilon(s)) ds + \frac{\sqrt{\epsilon}}{\delta}[\rho dW(s) + \sqrt{1 - \rho^2} dB(s)]. \end{aligned} \tag{6.3}$$

Both components (X, Y) take values in \mathbb{R} . We supplement the system with initial condition $(X^\epsilon(0), Y^\epsilon(0)) = (x_0, y_0)$. To connect to the notation of the general model (1.1), this corresponds to

$$b(x, y) = 0, \quad c^\epsilon(x, y) = \epsilon h(y), \quad \sigma(x, y) = \sigma(y),$$

$$f(x, y) = m - y, \quad g(x, y) = 0, \quad \tau_1(x, y) = \rho, \quad \tau_2(x, y) = \sqrt{1 - \rho^2}.$$

Of course, this system violates the periodicity assumption. However due to the mean reverting feature of the fast motion, the conclusions hold in this case as well.

In the next subsections we see the form of the large deviations action functional and of the control that defines the asymptotically optimal change of measure for all three regimes.

6.2.1 The Case of Regime 1

A simple computation shows that the only possible solution to cell problem (2.1) is the zero solution (this is because $b = 0$). Also, it is easy to see that the invariant measure corresponding to the operator \mathcal{L}^1 is independent of x and can be explicitly computed, taking the form

$$\mu(dy) = \frac{1}{\sqrt{\pi}} e^{-(y-m)^2} dy.$$

Then this implies that the formula for the action functional (Theorem 3.4) becomes

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

where $q = \int_{\mathcal{Y}} \sigma^2(y) \mu(dy)$. Given a classical subsolution \bar{U} , the importance sampling control that appears in Theorem 4.6 takes the form

$$\bar{u}(s, x, y; 1) = (-\sigma(y)\partial_x \bar{U}(s, x), 0).$$

As in the previous example, the choice of the subsolution \bar{U} according to Definition 4.1 depends on the terminal cost of interest $h(x)$.

6.2.2 The Case of Regime 2

The situation here is more complicated because the infimization problem that appears in the definition of the local rate function, Theorem 3.3, does not necessarily have a closed form solution as it had for Regime 1. However, due to the one-dimensionality aspect of the problem we can still do some algebraic computations. A simple algebra shows that the formula for the action functional (Theorem 3.3) becomes

$$S_{0T}(\phi) = \begin{cases} \frac{1}{2} \int_0^T L_2(\phi_s, \dot{\phi}_s) ds & \text{if } \phi \in \mathcal{AC}([0, T]; \mathbb{R}) \text{ and } \phi(0) = x_0 \\ +\infty & \text{otherwise,} \end{cases}$$

where

$$L_2(x, \beta) = \inf_{v \in \mathcal{A}_{x,\beta}^2} \left\{ \frac{1}{2} \int_{\mathcal{Y}} |v(y)|^2 \mu_v(dy) \right\}.$$

with

$$\mu_v(dy) = \frac{1}{L} e^{\int_1^y [2\gamma(m-z) + 2\rho v(z)] dz} dy, \quad L = \int_{\mathcal{Y}} e^{\int_1^y [2\gamma(m-z) + 2\rho v(z)] dz} dy$$

$$\mathcal{A}_{x,\beta}^2 = \left\{ v(\cdot) : \mathcal{Y} \mapsto \mathbb{R}, \beta = \int_{\mathcal{Y}} \sigma(y)v(y)\mu_v(dy) \right\}.$$

Notice that the invariant measure $\mu(dy)$ and the control v decouple when $\rho = 0$. In the general case $\rho \in [-1, 1]$, the equation for the related cell problem (4.14) takes the form

$$\frac{\gamma}{2} \xi''_{\gamma}(y) + [\gamma(m - y) - \sigma(y)\rho p] \xi'_{\gamma}(y) - \frac{1}{2} (\xi'_{\gamma}(y))^2 - \frac{1}{2} \sigma^2(y) p^2 = \bar{H}_{\gamma}(p).$$

There is a unique pair $(\xi_{\gamma}(y), \bar{H}_{\gamma}(p))$ satisfying this equation such that $\xi_{\gamma}(y) \in W^1_{loc}$, see [2, 22]. Notice that for this model, the solution $(\xi_{\gamma}(y), \bar{H}_{\gamma}(p))$ to the cell problem is independent of the slow motion x . Obtaining closed form solutions to such equations is difficult in principle, especially because we are interested in pairs $(\xi_{\gamma}(y), \bar{H}_{\gamma}(p))$. Numerical methods such as the ones developed in [10, 20] will be useful here. Notice also that by Remark 4.9 appropriate subsolution pairs suffice. We plan to return to these issues in detail in a future work.

Given sufficient smoothness such that Theorem 4.8 is applicable and a classical subsolution \bar{U} (depending on the choice of the terminal cost $h(x)$), the importance sampling control that appears in Theorem 4.8 takes the form

$$\begin{aligned} \bar{u}(s, x, y; 2) = & (-\sigma(y)\partial_x \bar{U}(s, x) - \rho \partial_y \xi_\gamma(y, \partial_x \bar{U}(s, x)), \\ & -\sqrt{1 - \rho^2} \partial_y \xi_\gamma(y, \partial_x \bar{U}(s, x))). \end{aligned}$$

6.2.3 The Case of Regime 3

It turns out that we can make some explicit computations here. To simplify things we will assume for brevity that $\rho = 1$. With these assumptions we get that $\tau_1(x, y) = 1$ and $\tau_2(x, y) = 0$. Assume that $\sigma \in L^1(\mathcal{Y})$ and that $\int_{\mathcal{Y}} \sigma(y) dy \neq 0$. A straightforward computation shows that the local rate function takes the form

$$L_3(x, \beta) = \inf_v \left\{ \frac{1}{2} \int_{\mathcal{Y}} |v(y)|^2 \frac{\beta}{v(y) \int_{\mathcal{Y}} \sigma(y) dy} dy : \int_{\mathcal{Y}} \frac{\beta}{v(y) \int_{\mathcal{Y}} \sigma(y) dy} dy = 1 \right\}.$$

This problem can be solved explicitly yielding

$$S_{0T}(\phi) = \begin{cases} \frac{1}{2} \int_0^T |\dot{\phi}_s|^2 (\int_{\mathcal{Y}} \sigma(y) dy)^{-2} ds & \text{if } \phi \in \mathcal{AC}([0, T]; \mathbb{R}) \text{ and } \phi(0) = x_0 \\ +\infty & \text{otherwise.} \end{cases}$$

The equation for the related cell problem (4.14) with $\gamma = 0$ takes the particular simple form

$$-\sigma(y)p\xi'_0(y) - \frac{1}{2}\sigma^2(y)p^2 - \frac{1}{2}(\xi'_0(y))^2 = \bar{H}_0(p).$$

This has the form of first order Bellman equation with quadratic Hamiltonian. Such equations have been studied in the literature and our assumptions guarantee that there are pairs (ξ_0, \bar{H}_0) such that ξ_0 is a continuous viscosity solution when $\bar{H}_0 \geq \bar{H}_0^*$ where \bar{H}_0^* is a critical value. We refer the interested reader to [23] for an extensive discussion on this.

If $\sigma(y)$ is periodic in y , say with period $\lambda = 1$, then $\mathcal{Y} = \mathbb{T} = [0, 1]$ and we look for a periodic solution $\xi_0(y)$. It turns out that the Bellman equation can then be solved explicitly yielding

$$\begin{aligned} \xi_0(y, p) = & p \left(y \int_0^1 \sigma(w) dw - \int_0^y \sigma(w) dw \right), \\ \bar{H}_0(p) = & -\frac{1}{2}p^2 \left(\int_0^1 \sigma(y) dy \right)^2. \end{aligned}$$

Thus, indeed $(\xi_0(y, p), \bar{H}_0(p))$ satisfy the assumptions of Theorem 4.10. Given a classical subsolution \bar{U} (depending on the choice of the terminal cost $h(x)$), the importance sampling control that appears in Theorem 4.10 takes the particularly simple

form

$$\begin{aligned}\bar{u}(s, x, y; 3) &= (-\sigma(y)\partial_x\bar{U}(s, x) - \partial_y\xi_0(y, \partial_x\bar{U}(s, x)), 0) \\ &= \left(-\left[\int_0^1 \sigma(w) dw\right]\partial_x\bar{U}(s, x), 0\right).\end{aligned}$$

7 Conclusions

In this paper we have developed the large deviations theory and a rigorous mathematical framework for the importance sampling theory for systems of slow-fast motion like (1.1). All the possible cases of interaction of fast motion and intensity of the noise are considered. The asymptotic performance of the proposed schemes are in terms of appropriate subsolutions to related HJB equations and in terms of appropriate “cell problems”. Straightforward adaptation of importance sampling schemes from standard diffusions without multiscale features lead to poor results in the multiscale setting. We have shown how the problem can be dealt with in the general multidimensional setting for fully dependent systems of slow-fast motion, when the fast motion is periodic.

Acknowledgements The author would like to acknowledge support by the Department of Energy (DE-SC0002413) during his stay at Brown University and by a start-up fund by Boston University during completion and revision of this paper.

References

1. Alvarez, O., Bardi, M.: Viscosity solutions methods for singular perturbations in deterministic and stochastic control. *SIAM J. Control Optim.* **40**(4), 1159–1188 (2001)
2. Arisawa, M., Lions, P.-L.: On ergodic stochastic control. *Commun. Partial Differ. Equ.* **23**, 2187–2217 (1998)
3. Baldi, P.: Large deviations for diffusions processes with homogenization and applications. *Ann. Probab.* **19**(2), 509–524 (1991)
4. Bardi, M., Capuzzo Dolcetta, I.: *Optimal Control and Viscosity Solutions of Hamilton Jacobi Bellman Equations*. Birkäuser, Boston (1997)
5. Bensoussan, A., Frehse, J.: On Bellman equations of ergodic control in \mathbb{R}^n . *J. Reine Angew. Math.* **429**, 125–160 (1992)
6. Bensoussan, A., Lions, J.L., Papanicolaou, G.: *Asymptotic Analysis for Periodic Structures*. Studies in Mathematics and its Applications, vol. 5. North-Holland, Amsterdam (1978)
7. Borkar, V., Gaitsgory, V.: Averaging of singularly perturbed controlled stochastic differential equations. *Appl. Math. Optim.* **56**(2), 169–209 (2007)
8. Boué, M., Dupuis, P.: A variational representation for certain functionals of Brownian motion. *Ann. Probab.* **26**(4), 1641–1659 (1998)
9. Buckdahn, R., Ichihara, N.: Limit theorem for controlled backward SDEs and homogenization of Hamilton-Jacobi-Bellman equations. *Appl. Math. Optim.* **51**, 1–33 (2005)
10. Camilli, F., Marchi, C.: Rates of convergence in periodic homogenization of fully nonlinear uniformly elliptic PDEs. *Nonlinearity* **22**, 1481–1498 (2009)
11. Dupuis, P., Ellis, R.S.: *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley, New York (1997)
12. Dupuis, P., Spiliopoulos, K.: Large deviations for multiscale problems via weak convergence methods. *Stoch. Process. Appl.* **122**, 1947–1987 (2012)

13. Dupuis, P., Spiliopoulos, K., Wang, H.: Importance sampling for multiscale diffusions. *SIAM J. Multiscale Model. Simul.* **12**(1), 1–27 (2012)
14. Dupuis, P., Spiliopoulos, K., Wang, H.: Rare event simulation in rough energy landscapes. In 2011 Winter Simulation Conference
15. Dupuis, P., Wang, H.: Subsolutions of an Isaacs equation and efficient schemes for importance sampling. *Math. Oper. Res.* **32**(3), 723–757 (2007)
16. Eithier, S.N., Kurtz, T.G.: *Markov Processes: Characterization and Convergence*. Wiley, New York (1986)
17. Evans, L.: Periodic homogenization of certain fully nonlinear partial differential equations. *Proc. R. Soc. Edinb., Sect. A* **120**, 245–265 (1992)
18. Feng, J., Fouque, J.-P., Kumar, R.: Small-time asymptotics for fast mean-reverting stochastic volatility models. *Ann. Appl. Probab.* **22**(4), 1541–1575 (2012)
19. Freidlin, M., Sowers, R.: A comparison of homogenization and large deviations, with applications to wavefront propagation. *Stoch. Process. Appl.* **82**(1), 23–52 (1999)
20. Gomes D, D., Oberman, A.: Computing the effective Hamiltonian using a variational approach. *SIAM J. Control Optim.* **43**, 792–812 (2004)
21. Horie, K., Ishii, H.: Simultaneous effects of homogenization and vanishing viscosity in fully nonlinear elliptic equations. *Funkc. Ekvacioj* **46**(1), 63–88 (2003)
22. Kaise, H., Sheu, S.J.: On the structure of solutions of ergodic type Bellman equation related to risk-sensitive control. *Ann. Probab.* **34**(1), 284–320 (2006)
23. Kaise, H., Sheu, S.J.: Ergodic type Bellman equations of first order with quadratic Hamiltonian. *Appl. Math. Optim.* **59**(1), 37–73 (2009)
24. Kushner, H.J.: *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*. Birkhäuser, Boston (1990)
25. Kushner, H.J.: Large deviations for two-time-scale diffusions with delays. *Appl. Math. Optim.* **62**(3), 295–322 (2009)
26. Lifson, S., Jackson, J.L.: On the self-diffusion of ions in a polyelectrolyte solution. *J. Chem. Phys.* **36**, 2410–2414 (1962)
27. Lions, P.-L., Souganidis, P.E.: Homogenization of degenerate second-order PDE in periodic and almost periodic environments and applications. *Ann. Inst. Henri Poincaré, Anal. Non Linéaire* **22**(5), 667–677 (2005)
28. Lipster, R.: Large deviations for two scaled diffusions. *Probab. Theory Relat. Fields* **106**(1), 71–104 (1996)
29. Mondal, D., Ghosh, P.K., Ray, D.S.: Noise-induced transport in a rough ratchet potential. *J. Chem. Phys.* **130**, 074703 (2009)
30. Pardoux, E., Veretennikov, A.Yu.: On Poisson equation and diffusion approximation 2. *Ann. Probab.* **31**(3), 1166–1192 (2003)
31. Pavliotis, G.A., Stuart, A.M.: *Multiscale Methods: Averaging and Homogenization*. Springer, Berlin (2007)
32. Saven, J.G., Wang, J., Wolynes, P.G.: Kinetics of protein folding: the dynamics of globally connected rough energy landscapes with biases. *J. Chem. Phys.* **101**(12), 11037–11043 (1994)
33. Veretennikov, A.Yu.: On large deviations in the averaging principle for SDEs with a “full dependence”, correction, [arXiv:math/0502098v1](https://arxiv.org/abs/math/0502098v1) [math.PR] (2005). Initial article in *Ann. Probab.* **27**(1), 284–296 (1999)
34. Veretennikov, A.Yu.: On large deviations for SDEs with small diffusion and averaging. *Stoch. Process. Appl.* **89**(1), 69–79 (2000)
35. Zwanzig, R.: Diffusion in a rough potential. *Proc. Natl. Acad. Sci. USA* **85**, 2029–2030 (1988)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.