Importance sampling in path space for diffusion processes

Carsten Hartmann · Christof Schütte · Marcus Weber · Wei Zhang

Received: date / Accepted: date

Abstract Importance sampling is a widely used technique to reduce the variance of a Monte Carlo estimator by an appropriate change of measure. In this work, we study importance sampling in the framework of diffusion process and consider a change of measure which is realized by adding a control force to the original dynamics. For certain exponential type expectation, the corresponding control force of the optimal change of measure leads to a zero-variance estimator and is related to the solution of a Hamilton-Jacobi-Bellmann equation. We prove that for a certain class of multiscale diffusions, the control force obtained from the limiting dynamics is asymptotically optimal, and we provide an error bound for the importance sampling estimators under such suboptimal controls. We also discuss two other situations in which one can approximate the optimal control force by solving simplified dynamics. We demonstrate our approximation strategy with several numerical examples and discuss its application to large-scale systems, e.g.

Keywords Importance sampling · Hamilton-Jacobi-Bellmann equation · Monte Carlo method · change of measure · rare events · diffusion process.

1 Introduction

Monte Carlo (MC) methods are powerful tools to solve high-dimensional problems that are not amenable to grid-based numerical schemes [32]. Despite their quite long history since the invention of the computer, the development of the MC method and applications thereof are a field of active research. Variants of the standard Monte Carlo method include Metropolis MC [21][6], Hybrid MC [13][39], Sequential MC [33][12], to mention just a few.

A key issue for many MC methods is variance reduction to improve the convergence of the corresponding MC estimators. Although all unbiased MC estimators share the same $O(N^{-\frac{1}{2}})$
decay of their variances with the sample size $N$, the prefactor matters a lot for the performance of the MC method. Therefore variance reduction techniques (see, e.g., [2]) seek to decrease the constant prefactor and thus to increase the accuracy and efficiency of the estimators.

In this paper, we focus on the importance sampling method for variance reduction. The basic idea is to generate samples from an alternative probability distribution (rather than sampling from the original probability distribution), so that the “important” regions in state space are more frequently sampled. To give an example, consider a real-valued random variable $X$ on some probability space $(\Omega, \mathcal{A}, P)$ and the calculation of a probability

$$P(X \in B) = \mathbb{E}(\chi_B(X))$$

of an event $\{\omega \in \Omega : X(\omega) \in B\}$ that is rare. When the set $B$ is rarely hit by the random variable $X$, it may be a good idea to draw samples from another probability distribution, say, $Q$ so that the event $\{X \in B\}$ has larger probability under $Q$. An unbiased estimator of $P(X \in B)$ can then be based on the appropriately reweighted expectation under $Q$, i.e.,

$$\mathbb{E}(\chi_B(X)) = \mathbb{E}_Q(\chi_B(X)\Psi),$$

with $\Psi(\omega) = (dP/dQ)(\omega)$ being the Radon-Nikodym derivative of $P$ with respect to $Q$. The difficulty now lies in a clever choice of $Q$, because not every probability measure $Q$ that puts more weight on the “important” region $B$ leads to a variance reduction of the corresponding estimator. Especially in cases when the two probability distributions are too different from each other so that the Radon-Nikodym derivative $\Psi$ (or likelihood ratio) becomes almost degenerate, the variance typically grows and one is better off with the plain vanilla MC estimator that is based on drawing samples from the original distribution $P$. Importance sampling thus deals with clever choices of $Q$ that enhance the sampling of events like $\{X \in B\}$ while mimicking the behaviour of the original distribution in the relevant regions (e.g., in the tails). Often such a choice can be based on large deviations asymptotics that provides estimates for the probability of the event $\{X \in B\}$ as a function of a smallness parameter; see, e.g., [5, 19, 3, 15, 14, 43].

Here we focus on the path sampling problem for diffusion processes. Specifically, given a diffusion process $(X_t)_{t \geq 0}$ governed by a stochastic differential equation (SDE), our aim is to compute the expectation of some functional of $X_t$ with respect to the underlying probability measure $P$ generated by the Brownian motion. In this setting, we want to apply importance sampling and draw samples (i.e. trajectories) from a modified SDE to which a control force has been added that drives the dynamics to the important state space regions. The control force generates a new probability measure on the space of continuous functions (i.e. on the space of trajectories $(X_t)_{t \geq 0}$), and estimating the expectation of the path functional with respect to the original probability measure by sampling from the controlled SDE is possible if the trajectories are reweighted according to the Cameron-Martin-Girsanov formula [34]. We confine ourselves to certain exponential path functionals which will be explicitly given below. For this type of path functionals, an optimal change of measure exists that admits importance sampling estimator with zero variance. Furthermore, the path sampling problem admits a dual formulation in terms of a stochastic optimal control problem, in which case finding the optimal change of measure is equivalent to solving the Hamilton-Jacobi-Bellmann (HJB) equation associated with the stochastic control problem. See the recent article [44] for a discussion of the underlying optimal control problem in the context of cross-entropy minimization. We would also like to point out that this
Suboptimal importance sampling. While generally it is impractical to find the optimal control force by solving an optimal control problem, especially for high-dimensional systems or systems with vastly different time scales, there is some hope that it is possible to find computable approximations to the optimal control problem that yield importance sampling estimators that are sufficiently accurate in that they have small variance. The main purpose of this paper is to show that such approximations naturally pop up in the context of diffusions with singular perturbations and to analyze this situation. To this end, we first study the estimator’s variance when a general control force is applied and demonstrate that the closer the control force is to the optimal one, the smaller the variance is. This indicates that designing good, but suboptimal estimators can be based on approximating the underlying HJB equation. For the specific case of multiscale diffusions with slow and fast degrees of freedom, we study the time-scale separation limit in the original HJB equation. It turns out that the limiting equation corresponds to an HJB equation that is associated with a limiting dynamics, which can be obtained by disregarding the control. Our main result is that, under certain assumptions, the control force constructed from the limiting equation is asymptotically optimal in the time-scale separation limit. From a numerical point of view, we also discuss two other situations, in which the approximate control forces can be obtained by solving a simpler equation (cf. [40]).

Here we want to emphasize the necessity of the importance sampling step, even though the solution to the limiting HJB equation provides—by duality—the asymptotically exact expectation value that we want to compute. Firstly, it is difficult to verify whether the expectation value that could be obtained from the original HJB equation is close to its asymptotic limit. Secondly, even if it is, it is difficult to quantify the error between the true expectation value and its asymptotic limit by solving the asymptotic equation alone. Moreover, the importance sampling step provides us reliable results and has good performance even when the system is not in the limiting regime (see numerical results and discussions in Section 4). Last but not least, we emphasize that the asymptotic equations can be solved on a much coarser grid and thus the computation of the control forces is numerically relatively cheap.

Related work. Our work is inspired by the related works [14,43]. In [14], the authors considered multiscale diffusions with a specific form, and the corresponding control force were found by studying viscosity subsolutions of HJB equation for the rate function of an underlying large deviations principle in the limit of vanishing noise. Also see [40,42,41] for continuation studies of various types of problems by large deviation approach. On the other hand, in the closely related approach [43], the authors propose to compute the approximate control force by solving a deterministic optimal control problem that is associated with the small noise limit of the diffusion and gives estimators that are asymptotically efficient, however, multiscale aspects are not an issue there. In [43], the authors also discuss the various notions of asymptotic optimality. Compared to the aforementioned works, our idea based on suboptimal approximations is simpler in that it does neither require to be in the large deviations regime or to solve for viscosity (sub)solutions of the underlying HJB equation in the full-dimensional space. As a consequence it is easier to implement numerically, especially in high dimensions; cf. [44]. We remark that the
asymptotic optimality proof by itself may have independent value for the study of singularly perturbed Fokker-Planck equations.

**Organization of the article.** This paper is organized as follows. In Section 2 we briefly introduce the importance sampling method in the diffusion framework and discuss the variance of Monte Carlo estimators corresponding to a general control force. Section 3 states the key assumptions and our main result: a bound of the relative error based on suboptimal controls for the case of multiscale diffusions; the result is proved in Section 5 but we provide some heuristic arguments based on formal asymptotic expansions already in Section 3; possible generalizations of the approach for the case of small noise diffusions are discussed in Subsection 3.3. Section 4 records various numerical examples that demonstrate the performance of the different approximation strategies. Appendix A and B contain technical results that are used in the proof.

## 2 Importance sampling of diffusions

We consider the conditional expectation

\[ I = \mathbb{E}\left( \exp \left( -\beta \int_t^T h(z_s) \, ds \right) \mid z_t = z \right) \quad (2.1) \]

on fixed time interval \([t, T]\), where \(\beta > 0\), \(h : \mathbb{R}^n \to \mathbb{R}^+\), and \(z_s \in \mathbb{R}^n\) satisfies the dynamics

\[ \begin{align*}
    dz_s &= b(z_s) \, ds + \beta^{-1/2} \sigma(z_s) \, dw_s, \quad t \leq s \leq T \\
    z_t &= z
\end{align*} \quad (2.2) \]

with \(b : \mathbb{R}^n \to \mathbb{R}^n\), \(\sigma : \mathbb{R}^n \to M_n\) (set of \(n \times n\) real matrices), \(w_s\) is a standard \(n\)-dimensional Wiener process. It is known that the SDE (2.2) induces a probability measure \(P\) over the path ensembles \(z_s, t \leq s \leq T\) starting from \(z\). To apply the importance sampling method, we introduce

\[ dw_s = \beta^{1/2} u_s \, ds + dw_s, \quad (2.3) \]

where \(u_s\) will be referred to as the control force. Then it follows from Girsanov theorem that \(\tilde{w}_s\) is a standard Wiener process under probability measure \(\tilde{P}\), where the Radon-Nikodym derivative is

\[ \frac{d\tilde{P}}{dP} = Z_t = \exp \left( -\beta^{1/2} \int_t^T u_s \, dw_s - \frac{\beta}{2} \int_t^T |u_s|^2 \, ds \right). \quad (2.4) \]

In the following, we will omit the conditioning on the initial value at time \(t\). Let \(\tilde{E}\) denote the expectation under \(\tilde{P}\), then we have

\[ I = \mathbb{E}\left( \exp \left( -\beta \int_t^T h(z_s) \, ds \right) \right) = \tilde{E}\left( \exp \left( -\beta \int_t^T h(z_s^u) \, ds \right) Z_t^{-1} \right), \quad (2.5) \]

with variance

\[ \text{Var}_\beta I = \tilde{E}\left[ \exp \left( -2 \beta \int_t^T h(z_s^u) \, ds \right) (Z_t)^{-2} \right] - I^2. \quad (2.6) \]
Moreover, under $\bar{P}$, we have
\[
dz^u_s = b(z^u_s) ds - \sigma(z^u_s) u_s ds + \beta^{-1/2} \sigma(z^u_s) d\bar{w}_s, \quad t \leq s \leq T
\]
\[
z^u_t = z.
\]
(2.7)

Now consider the calculation of (2.5) by a Monte Carlo sampling in path space, and suppose that $N$ independent trajectories $\{z^{u,i}_s, t \leq s \leq T\}$ of (2.7) have been generated where $i = 1, 2, \cdots, N$. An unbiased estimator of (2.1) is now given by
\[
I_N = \frac{1}{N} \sum_{i=1}^{N} \left( \exp \left( - \beta \int_t^T h(z^{u,i}_s) ds \right) \left( Z^{u,i}_t \right)^{-1} \right),
\]
whose variance is
\[
\text{Var}_u I_N = \frac{\text{Var}_u I}{N} = \frac{1}{N} \left[ \mathbb{E} \left( \exp \left( - 2\beta \int_t^T h(z^{u,i}_s) ds \right) (Z^{u,i}_t)^{-2} \right) - I^2 \right].
\]
(2.8)

Notice that $Z_t = 1$ when $u_s \equiv 0$, and we recover the standard Monte Carlo method. In order to quantify the efficiency of the Monte Carlo method, we introduce the relative relative error
\[
\rho_u(I) = \sqrt{\text{Var}_u I / I^2}.
\]
(2.9)

The advantage of introducing the control force $u_s$ is that we may choose $u_s$ to decrease the variance of the estimator (2.8). From (2.6) and (2.9) we conclude that minimizing the variance of the new estimator and hence the relative error is equivalent to choosing $u_s$ such that
\[
\mathbb{E} \left( \exp \left( - 2\beta \int_t^T h(z^{u,i}_s) ds \right) (Z^{u,i}_t)^{-2} \right) \approx I^2.
\]
(2.10)

2.1 Dual optimal control problem and variance estimate.

To proceed, we make use of the following duality relation [7,11]:
\[
\log \mathbb{E} \left( \exp \left( - \beta \int_t^T h(z_s) ds \right) \right) = -\beta \inf_{u_s} \mathbb{E} \left\{ \int_t^T h(z^{u}_s) ds + \frac{1}{2} \int_t^T |u_s|^2 ds \right\}.
\]
(2.12)

We call the unique control $\hat{u}_s$, at which the infimum on the right-hand side (RHS) of (2.12) is attained the optimal control force. Accordingly we define $\hat{w}_s, \hat{Z}_t, \hat{P}$ to be the respective quantities (2.3) and (2.4) with $u_s$ replaced by $\hat{u}_s$, and we call $\hat{z}_s = \hat{z}^u_s$ the solution of (2.7) with control force $\hat{u}_s$. Using Jensen’s inequality one can show that (2.12) implies
\[
\exp \left( - \beta \int_t^T h(\hat{z}_s) ds \right) \hat{Z}_t^{-1} = I, \quad \hat{P} - a.s.
\]
(2.13)

Combining the last equation with (2.9) it follows that the change of measure induced by $\hat{u}_s$ is optimal in the sense that the variance of the importance sampling estimator (2.8) vanishes.
It is helpful to note that the RHS of (2.12) has an interpretation as the value function of a stochastic control problem:

$$U(t, z) = \inf_{u_s} \tilde{E} \left( \int_t^T h(z_s^u) \, ds + \frac{1}{2} \int_t^T |u_s|^2 \, ds \right) \bigg| z_t = z. \quad (2.14)$$

From dynamic programming principle, we know $U(t, z)$ satisfies the following Hamilton-Jacobi-Bellman or dynamic programming equation:

$$\frac{\partial U}{\partial t} + \min_{c \in \mathbb{R}^n} \left\{ c + \frac{1}{2} \sigma \cdot \nabla U + \frac{1}{2} \sigma \sigma^T \cdot \nabla^2 U \right\} = 0, \quad (2.15)$$

which implies that the optimal control force $\hat{u}_s$ is of feedback form and satisfies

$$\hat{u}_s = \sigma^T(\hat{z}_s) \nabla U(s, \hat{z}_s). \quad (2.16)$$

Now we estimate (2.11) and thus the variance (2.6) for a general $u_s$. To this end we suppose that the probability measures $\bar{P}$ and $\hat{P}$ are mutually equivalent. Then, using (2.13), we can conclude that

$$\exp \left( - \beta \int_t^T h(\hat{z}_s) \, ds \right) \tilde{Z}_t^{-1} = I, \quad \hat{P} - a.s. \quad (2.17)$$

and therefore

$$\tilde{E} \left( \exp \left( - 2 \beta \int_t^T h(z_s^u) ds \right) (\tilde{Z}_t)^{-2} \right) = \tilde{E} \left( \exp \left( - 2 \beta \int_t^T h(z_s^u) ds \right) (\tilde{Z}_t)^{-2} \left( \frac{\tilde{Z}_t}{\tilde{Z}_t} \right)^2 \right)$$

$$= I^2 \tilde{E} \left( \left( \frac{\tilde{Z}_t}{\tilde{Z}_t} \right)^2 \right), \quad (2.18)$$

where by Girsanov’s formula (2.3), we have

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

In order to simplify (2.18), we follow [14] and introduce another control force $\tilde{u}_s$ and change the measure again. Specifically, we choose $\tilde{u}_s = 2\hat{u}_s - u_s$ and define $\tilde{w}_s, \tilde{P}, \tilde{Z}_t$ as in (2.23)–(2.24), with $u_s$ being replaced by $\tilde{u}_s$. If we now let $\tilde{E}$ denote the expectation with respect to $\tilde{P}$ then, using equations (2.13) and (2.19), we obtain

$$\tilde{E} \left( \left( \frac{\tilde{Z}_t}{\tilde{Z}_t} \right)^2 \right) = \tilde{E} \left( \left( \frac{\tilde{Z}_t}{\tilde{Z}_t} \right)^2 \tilde{Z}_t^{-1} Z_t \right) = \tilde{E} \left( \exp \left( \beta \int_t^T |\tilde{u}_s - u_s|^2 ds \right) \right). \quad (2.20)$$

Hence the last equations suggests that in order to reduce the variance, we should choose a control $u$ that is uniformly close to the optimal control $\hat{u}$. In Section 3 we shall discuss several situations when such a uniformly close, suboptimal control force can be constructed based on simplifications of the original dynamics.

**Relative sampling error.** Before closing this section, we define the relative error $\rho_s(I)$ for the sample variance

$$\text{Var}_s I = \frac{1}{N} \sum_{i=1}^N \left( \exp \left( - \beta \int_t^T h(z_s^{u,i}) ds \right) (Z_s^{u,i})^{-1} Z_t \right)^2 - I_N \right)^2, \quad (2.21)$$
as
\[ \rho_s(I) = \frac{\sqrt{\text{Var}_s I}}{I_N}, \] (3.22)
which will be used in the numerical examples in Section 4. Note that both \( \text{Var}_s I \) and \( \rho_s(I) \) depend on the control \( u \), but the dependence is omitted to simplify the notations.

3 Importance sampling of multiscale diffusions

Our main result in this paper concerns multiscale dynamics. Specifically, we consider the case when the state variable \( z \in \mathbb{R}^n \) can be split into a slow variable \( x \in \mathbb{R}^k \) and a fast variable \( y \in \mathbb{R}^l \), i.e. \( z = (x, y) \), \( k + l = n \), and we assume that (3.22) is of the form
\[
\begin{align*}
\frac{dx}{ds} &= f(x, y)ds + \beta^{-1/2} \alpha_1(x, y)dw_s^1, \\
\frac{dy}{ds} &= \frac{1}{\sqrt{\epsilon}} g(x, y)ds + \beta^{-1/2} \frac{1}{\sqrt{\epsilon}} \alpha_2(x, y)dw_s^2,
\end{align*}
\] (3.1)
where \( f : \mathbb{R}^n \rightarrow \mathbb{R}^k \), \( g : \mathbb{R}^n \rightarrow \mathbb{R}^l \) are smooth vector fields, \( \alpha_1 : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times k} \), \( \alpha_2 : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times l} \) are smooth noise coefficients and \( w^1_s \in \mathbb{R}^k \), \( w^2_s \in \mathbb{R}^l \) are independent Wiener processes. The parameter \( \epsilon \ll 1 \) describes the time-scale separation.

Let \( x \in \mathbb{R}^k \) be given and suppose that the fast subsystem
\[
\frac{dy}{ds} = \frac{1}{\sqrt{\epsilon}} g(x, y)ds + \beta^{-1/2} \frac{1}{\sqrt{\epsilon}} \alpha_2(x, y)dw_s^2, \quad y_t = y
\] (3.2)
is ergodic with unique invariant measure whose density is \( \rho_s(y) \) with respect to Lebesgue measure. (Note that \( \rho_s \) should be distinguished from the relative error in (2.10) or (2.22).) Then it is well known that when \( \epsilon \rightarrow 0 \), under some mild conditions on the coefficients, the slow component of (3.1) converges in probability to the reduced dynamics (10-26,36)
\[
\begin{align*}
\frac{d\tilde{x}}{ds} &= \tilde{f}(\tilde{x})ds + \beta^{-1/2} \tilde{\alpha}(\tilde{x})dw_s, \quad t \leq s \leq T \\
\tilde{x}_t &= x.
\end{align*}
\] (3.3)
where for every \( x \in \mathbb{R}^k \), we have
\[
\tilde{f}(x) = \int_{\mathbb{R}^l} f(x, y)\rho_x(y) \, dy, \quad \tilde{\alpha}(x)\tilde{\alpha}(x)^T = \int_{\mathbb{R}^l} \alpha_1(x, y)\alpha_1(x, y)^T \rho_x(y) \, dy.
\] (3.4)
Further define
\[
\tilde{h}(x) = \int_{\mathbb{R}^l} h(x, y)\rho_x(y) \, dy
\] (3.5)
and consider the averaged value function
\[
U_0(t, x) = \inf_u \mathbb{E} \left\{ \int_t^T \tilde{h}(\tilde{x}^u_s) \, ds + \frac{1}{2} \int_t^T |u_s|^2 \, ds \right\},
\] (3.6)
where \( \tilde{x}^u_t \in \mathbb{R}^k \) is the solution of
\[
\begin{align*}
\frac{d\tilde{x}^u}{ds} &= \tilde{f}(\tilde{x}^u)ds - \tilde{\alpha}(\tilde{x}^u)u_s \, ds + \beta^{-1/2} \tilde{\alpha}(\tilde{x}^u)dw_s, \quad t \leq s \leq T \\
\tilde{x}^u_t &= x.
\end{align*}
\] (3.7)
The idea of using suboptimal controls for importance sampling of multiscale systems such as (3.1) is to use the solution of the limiting control problem (3.6)-(3.7) to construct an asymptotically optimal control of the form

\[ \hat{u}_s^0 = \left( \alpha_1^T(x_s^0, y_s^0) \nabla_x U_0(x_s^0), 0 \right). \]  

(3.8)

for the full system. Comparing (3.8) to the optimal control force (2.16), this means that we construct the control for the slow variable by using the optimal control of the reduced dynamics (3.3) and leave the fast variable uncontrolled.

**Remark 1** Another variant of a suboptimal control would be

\[ \hat{u}_s^0 = \left( \tilde{\alpha}_1^T(x_s^0) \nabla_x U_0(x_s^0), 0 \right), \]  

(3.9)

where the \( x \)-component is the optimal control of the averaged system (3.6)--(3.7). The advantage of using (3.9) rather than (3.8) is that the fast variables do not need to be explicitly known or observable to control the system. In the following we will assume that \( \alpha_1 \) is independent of \( y \), in which case (3.8) and (3.9) coincide (see Assumption 3).

### 3.1 Main result

Our main assumptions are as follows (cf. [30]).

**Assumption 1** \( f, g, h, \alpha_1, \alpha_2 \) are \( C^2 \) functions, with derivatives that are uniformly bounded by a constant \( C \). \( \alpha_1, \alpha_2 \) and \( h \) are bounded. Furthermore, there exists constant \( C_1, C_2 \), such that

\[ \xi^T \alpha_1(x, y) \alpha_1(x, y)^T \xi \geq C_1 |\xi|^2, \quad \zeta^T \alpha_2(x, y) \alpha_2(x, y)^T \zeta \geq C_2 |\zeta|^2, \]

\( \forall \xi, x \in \mathbb{R}^k, \zeta, y \in \mathbb{R}^l \).

**Assumption 2** \( \exists \lambda > 0 \), such that \( \forall x \in \mathbb{R}^k, y_1, y_2 \in \mathbb{R}^l \), we have

\[ \langle g(x, y_1) - g(x, y_2), y_1 - y_2 \rangle + \frac{3}{\beta} \| \alpha_2(x, y_1) - \alpha_2(x, y_2) \|^2 \leq -\lambda \| y_1 - y_2 \|^2, \]  

(3.10)

where \( \| \cdot \| \) denotes the Frobenius norm.

**Assumption 3** \( \alpha_1 \) and \( h \) do not depend on \( y \).

**Remark 2** 1. Assumption 1 implies the coefficients are Lipschitz functions. In particular, it holds that \( |f(x, y)| \leq C(1 + |x| + |y|) \forall x \in \mathbb{R}^k, y \in \mathbb{R}^l \) (similarly for the other coefficients).

2. For \( \tilde{f} \) as given by (3.4), Lemma B.4 in Appendix B implies that \( \tilde{f} \) is Lipschitz continuous. Unlike [30], we do not assume that \( f \) is bounded.

3. Assumption 2 guarantees that the fast dynamics are sufficiently mixing, even when \( x \) is slowly moving. As we study the asymptotic solution of (3.1) as \( \epsilon \to 0 \) at fixed noise intensity, the inverse temperature \( \beta \) can be absorbed into the coefficients \( \alpha_1, \alpha_2 \) and \( h \). In Section 5 we will therefore assume \( \beta = 1 \), in which case Assumption 2 implies that

\[ \langle \nabla_y g \xi, \xi \rangle + 3 \| \nabla_y \alpha_2 \xi \|^2 \leq -\lambda |\xi|^2, \quad \forall y, \xi \in \mathbb{R}^l, x \in \mathbb{R}^k. \]  

(3.11)
Combining this with Assumption 1 we have
\[
\langle g(x, y), y \rangle + \frac{3}{2} \| \alpha_2(x, y) \|^2 \\
\leq (g(x, y) - g(x, 0), y) + \langle g(x, 0), y \rangle + 3 \| \alpha_2(x, y) - \alpha_2(x, 0) \|^2 + 3 \| \alpha_2(x, 0) \|^2 \\
\leq - \frac{\lambda}{2} |y|^2 + C(|x|^2 + 1) \quad \forall x \in \mathbb{R}^k, y \in \mathbb{R}^l
\]

The constant 3 in (3.11) is not optimal, but it will simplify matters later on.

Now we are ready to state our main result, whose proof will be given in Section 5.

**Theorem 3.1** Let Assumptions 1–3 hold, and consider the importance sampling method for computing (2.1) with the dynamics (3.1) and control \( \hat{u}_0 \) as given by (3.8). Then, for \( \epsilon \ll 1 \), the relative error (2.10) of the importance sampling estimator satisfies
\[
\rho_{\hat{u}_0}(I) \leq C \epsilon^\frac{1}{8},
\]
where constant \( C > 0 \) is independent of \( \epsilon \).

### 3.2 Formal expansion by asymptotic analysis

The proof of the theorem in Section 5 is relatively long and technical, which is why we shall give a formal derivation of (3.8). The idea is to identify the suboptimal control \( \hat{u}_0 \) as the leading term of the optimal control using formal asymptotic expansions [4, 36]. To this end, let \( U^\epsilon \) denote the solution of (2.15), for which we seek an asymptotic expansion in powers of \( \epsilon \). Further let \( \phi^\epsilon(t, x, y) = \exp(-\beta U^\epsilon) \). From the dual relation (2.12), we know that \( \phi^\epsilon \) is the expectation (2.1) we want to compute. By the Feynman-Kac formula, we have
\[
\frac{\partial \phi^\epsilon}{\partial t} + L \phi^\epsilon - \beta h \phi^\epsilon = 0, \quad 0 \leq t \leq T \\
\phi^\epsilon(T, x, y) = 1
\]

(3.12)

where \( L = \epsilon^{-1} L_0 + L_1 \) is the infinitesimal generator of (3.1), with
\[
L_0 = g \cdot \nabla_y + \frac{1}{2\beta} \alpha_2 \alpha_2^T : \nabla_y^2 \\
L_1 = f \cdot \nabla_x + \frac{1}{2\beta} \alpha_1 \alpha_1^T : \nabla_x^2.
\]

(3.13)

Now consider the expansion \( \phi^\epsilon = \phi_0 + \epsilon \phi_1 + \ldots \) of \( \phi^\epsilon \) in powers of \( \epsilon \). Plugging it into (3.12) and comparing different powers of \( \epsilon \), we obtain to lowest order:
\[
\frac{\partial \phi_0}{\partial t} + L_0 \phi_1 + L_1 \phi_0 - \beta h \phi_0 = 0, \\
L_0 \phi_0 = 0.
\]

(3.14)

(3.15)

By the assumption that the fast dynamics (3.2) are ergodic for every \( x \in \mathbb{R}^k \) with unique invariant density \( \rho_x(y) \), it follows that \( \rho_x(y) > 0 \) is the unique solution to the linear equation \( L_0^* \rho_x = 0 \) with \( \| \rho_x \|_1 = 1 \). Here \( L_0^* \) is the formal adjoint of \( L_0 \) with respect to the standard scalar
product in the space $L^2$. Hence we can conclude from (3.15) that $\phi_0 = \phi_0(t, x)$ is independent of $y$, and integrating both sides of (3.14) against $\rho_x(y)$, we obtain a closed equation for $\phi_0$:

$$\frac{\partial \phi_0}{\partial t} + \tilde{L} \phi_0 - \beta \tilde{h} \phi_0 = 0$$

(3.16)

with

$$\tilde{L} = \tilde{f}(x) \cdot \nabla_x + \frac{\tilde{\alpha}(x) \tilde{\alpha}(x)^T}{2\beta} : \nabla_x^2,$$

(3.17)

and $\tilde{h}, \tilde{f}, \tilde{\alpha}$ as given by (3.4) and (3.5).

Notice that $\tilde{L}$ is the infinitesimal generator of the averaged dynamics (3.3). Again by the Feynman-Kac formula, the solution to (3.16) is recognized as the conditional expectation

$$\phi_0(t, x) = E\left( \exp \left( -\beta \int_t^T \tilde{h}(\tilde{x}_s) \, ds \right) \bigg| x_t = x \right)$$

(3.18)

of the averaged path functional over all realizations of the averaged dynamics (3.3) starting at $\tilde{x}_t = x$. Recalling that $U^\epsilon = -\beta^{-1} \log \phi^\epsilon$, it follows that $U^\epsilon$ has the expansion

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

(3.19)

Hence $U_0 = -\beta^{-1} \log \phi_0$. Combining (3.18) and the dual relation (2.12), we conclude that $U_0$ satisfies (3.6). Finding the corresponding expression for the optimal control is now straightforward: Setting $\hat{u}_s = (\hat{u}_{1,s}, \hat{u}_{2,s})$ the relation (2.16) between the optimal feedback and the value function yields

$$\hat{u}_{1,s} = \alpha_1^T \nabla_x U_0 + O(\epsilon) = -\beta^{-1} \frac{\alpha_1^T \nabla_x \phi_0}{\phi_0} + O(\epsilon),$$

$$\hat{u}_{2,s} = \frac{\alpha_2^T}{\sqrt{\epsilon}} \nabla_y U^\epsilon = O(\sqrt{\epsilon}).$$

(3.20)

where all functions are evaluated at $(s, x_{1,s})$ or $(s, x_{2,s}, y_{2,s})$.

The last equation shows that (3.8) appears to be the leading term of the optimal control force as $\epsilon \to 0$. Reiterating the argument given in Section 2, we expect (3.8) to be a reasonably good approximation of the exact control force that gives rise to sufficiently accurate importance sampling estimators of (2.1) in the asymptotic regime $\epsilon \ll 1$.

As for the corresponding numerical algorithm, our little derivation suggest that one possible strategy for finding good control forces for importance sampling is to first compute $U_0$ from (3.6) or (3.18), which corresponds to a low-dimensional stochastic optimal control problem, and then to construct the control force as in (3.8) to perform importance sampling. The numerical strategy will be discussed in Section 4 along with some details regarding the numerical implementation.

Remark 3 Another variant of the above are homogenization problems that appear in connection with diffusions that exhibit more than two time scales [36]. Although a rigorous treatment of multiscale diffusions with three or more time scales is beyond the scope of this work, we stress that the formal asymptotic argument carries over directly. As an example, we study a three-scale system in Subsection 4.2 using the formal framework and comparing it with results obtained in the related work [14].
3.3 Possible generalizations

In this subsection, we briefly discuss the case when $\beta \gg 1$, which corresponds to zero-temperature limit of (2.2). Instead of (2.15), we consider the equation

$$\frac{\partial \tilde{U}}{\partial t} + \min_{c \in \mathbb{R}^n} \left\{ h + \frac{1}{2} |c|^2 + (b - \sigma c) \cdot \nabla \tilde{U} \right\} = 0$$

(3.21)

and approximate $\hat{u}$ in (2.16) by

$$\hat{u}_s^0 = \sigma^T \nabla \tilde{U}(s, \eta_u^s),$$

(3.22)

with

$$\frac{d\eta_u^s}{ds} = b(\eta_u^s) - \sigma u_s, \quad \eta_u^0 = z.$$  

(3.23)

It readily follows from the dynamic programming principle that $\tilde{U}$ solving (3.21) is the value function of the following deterministic optimal control problem:

$$\tilde{U}(t, z) = \inf_{u} \left\{ \int_t^T h(\varphi(s)) ds + \frac{1}{2} \int_t^T |u_s|^2 ds \right\}$$

(3.24)

where the minimization is over all bounded measurable controls $u$ and subject to (3.23).

Replacing a stochastic control problem by the corresponding deterministic control problems may sound weird, for (3.21) is typically not very well behaved, in that the solutions of (3.21) may fail to be differentiable or unique, even though the coefficients are smooth and (2.15) has a classical solution (i.e., a solution of class $C^{1,2}([0, T] \times \mathbb{R}^n)$). Nevertheless it may be beneficial to employ the relation (3.22) with $\tilde{U}$ being a viscosity solution of (3.21): 

$$\tilde{U}(t, z) = \inf_{\varphi \in \mathcal{AC}(t, T)} \left\{ \int_t^T h(\varphi(s)) ds + \frac{1}{2} \int_t^T \|\varphi'(s) - b(\varphi(s))\|_{\sigma}^2 ds : \varphi(t) = z \right\},$$

(3.25)

where $\mathcal{AC}(t, T)$ is the space of absolutely continuous functions $\varphi: [t, T] \to \mathbb{R}^n$ and $\|v\|_{\sigma} = \sqrt{v^T (\sigma \sigma^T)^{-1} v}$ denotes the Euclidean norm of a vector $v \in \mathbb{R}^n$ with respect to the scalar product that is weighted by the inverse of the noise covariance $\sigma(\cdot) \sigma(\cdot)^T > 0$. It turns out that (3.21) has a unique viscosity solution under fairly mild assumptions, which, more importantly, can be computed by solving a finite dimensional optimization problem using a discretization of $\varphi$, rather than solving a partial differential equation like (3.21). This is exactly the idea pursued in [43], and readers are referred to this work for examples and details on the numerical implementation.
Multiscale diffusions in the zero-temperature limit. Another generalization concerns multiscale diffusions like (3.1) at low temperature, i.e. \( \epsilon \ll 1 \) and \( \beta \gg 1 \). From the dynamic programming principle and (3.6), we know that the leading term of \( U_0 \) satisfies

\[
\frac{\partial U_0}{\partial t} + \min_{c \in \mathbb{R}^k} \left\{ \tilde{h} + \frac{1}{2}|c|^2 + (\tilde{f} - \tilde{\alpha}c) \cdot \nabla U_0 + \frac{\tilde{\alpha} \tilde{\alpha}^T}{2\beta} : \nabla^2 U_0 \right\} = 0,
\]

(3.26)

where \( \tilde{f}, \tilde{h}, \tilde{\alpha} \) are given in (3.4). Pushing the idea of the small noise asymptotics a bit further, we conjecture that good approximations of the optimal control may be obtained by replacing (3.26) by its deterministic counterpart (cf. [40,42]):

\[
\frac{\partial \tilde{U}_0}{\partial t} + \min_{c \in \mathbb{R}^k} \left\{ \tilde{h} + \frac{1}{2}|c|^2 + (\tilde{f} - \tilde{\alpha}c) \cdot \nabla \tilde{U}_0 \right\} = 0,
\]

(3.27)

\( \tilde{U}_0(T, x) = 0 \).

In accordance with the previous considerations, the solutions (3.27) are associated with the low-dimensional deterministic optimal control problem:

\[
\tilde{U}_0(t, x) = \inf_u \left\{ \int_t^T \tilde{h}(\eta^u_s) \, ds + \frac{1}{2} \int_t^T |u_s|^2 \, ds \right\}
\]

(3.28)

where the minimization over all bounded measurable controls \( u \) is subject to

\[
\frac{d\eta^u_s}{ds} = \tilde{f}(\eta^u_s) - \tilde{\alpha}u_s, \quad \eta^u_t = x.
\]

(3.29)

As before, an approximation of the optimal control force \( \tilde{u} \) may be obtained from

\[
\tilde{u}^0_s = (\alpha^T_1 \nabla_x \tilde{U}_0(s, \eta^0_s), 0).
\]

(3.30)

4 Numerical examples

In this section, we present two numerical examples and discuss possible strategies for finding good suboptimal control forces as outlined in Section 3. We wish to keep the presentation as lucid as possible, which is why we confine ourselves to relatively simple, but illustrative one- and two-dimensional toy systems. We emphasize that our examples may violate Assumptions 1-3, and hence go beyond Theorem 3.1 in Subsection 3.1. As a consequence they indicate that our assumptions may be too tight and that generalizations are possible.

4.1 Two-dimensional diffusion with stiff potential

We consider the potential \( V : \mathbb{R}^2 \rightarrow \mathbb{R} \) with \( (x, y) \mapsto V_1(x) + V_2(x, y) \) and

\[
V_1(x) = \frac{1}{2}(x^2 - 1)^2, \quad V_2(x, y) = \frac{1}{2}(x - y)^2
\]

(4.1)
and the two-dimensional SDE
\[
\begin{align*}
\frac{dx_s}{ds} &= -\frac{\partial V(x_s, y_s)}{\partial x} ds + \beta^{-1/2} dw^1_s \\
\frac{dy_s}{ds} &= -\frac{1}{\epsilon} \frac{\partial V(x_s, y_s)}{\partial y} ds + \beta^{-1/2} \frac{1}{\sqrt{\epsilon}} dw^2_s
\end{align*}
\] (4.2)

A similar slow-fast system has been studied in [28]. We seek to calculate the expectation
\[
I = \mathbb{E}\left( \exp\left( -\beta \int_0^T h(x_s, y_s) ds \right) \bigg| x_0 = -1, y_0 = 0 \right),
\] (4.3)

where \( h(x, y) = (x - 1)^2 \). For every \( x \in \mathbb{R} \), the invariant measure of the fast dynamics \( y_s \) in (4.2) has the density
\[
\rho_x(y) \propto e^{-\beta(x-y)^2}.
\] (4.4)

The reduced slow dynamics then is a one-dimensional diffusion in a double well potential:
\[
\frac{d\tilde{x}_s}{ds} = -V'_1(\tilde{x}_s) ds + \beta^{-1/2} dw_s.
\] (4.5)

Before we proceed, it is worthy to briefly illustrate the difficulties to compute (4.3) by the standard Monte Carlo method, especially when \( \beta \) is large. On one hand, in path space, the exponential integrand in (4.3) is peaked around trajectories which spend a large portion of time at the minimum of \( h \), that is located at \( x = 1 \). On the other hand, in order to get close to \( x = 1 \), trajectories starting from \( x_0 = -1 \) need to cross the energy barrier \( \Delta V_1 \) of \( V_1 \) (see Fig. 1(a)). The probability of these barrier-crossing trajectories is exponentially small in \( -\beta \Delta V_1 \) when \( \beta \) is large. Combining these two facts, we can see that the rare barrier crossing events play an important role when computing (4.3). Standard Monte Carlo will be inefficient in such a situation due to insufficient sampling of the rare events (cf. the discussion in Section 1).

**Computation of a suboptimal estimator based on an averaged equation.** Now let us consider the method outlined in Subsection 3.1. Recalling (3.16), the averaged conditional expectation \( \phi_0 \) solves the linear backward evolution equation
\[
\begin{align*}
\frac{\partial \phi_0}{\partial t} + \tilde{L} \phi_0 - \beta h \phi_0 &= 0 \\
\phi_0(T, x) &= 1,
\end{align*}
\] (4.6)

with
\[
\tilde{L} = -V'_1 \frac{\partial}{\partial x} + \frac{1}{2\beta} \frac{\partial^2}{\partial x^2}, \quad \tilde{h}(x) = h(x) = (x - 1)^2.
\] (4.7)

The equation for \( \phi_0 \) is one-dimensional (in space), and can be solved by standard methods: For instance, using Rothe’s method, we can first discretize (4.6) in time, which yields
\[
\left( \frac{1}{\Delta t} - \tilde{L} \right) \phi_{0}^j = \left( \frac{1}{\Delta t} - \beta \tilde{h} \right) \phi_{0}^{j+1}, \quad j = 0, 1, \cdots, m - 1
\] (4.8)

where \( \phi_{0}^j \) denotes the approximation of \( \phi_0 \) at time \( t_j = j \Delta t \), \( j = 0, 1, \cdots, m \) with time step \( \Delta t = T/m \). Equation (4.8) is then further discretized in space using the structure-preserving finite volume method described in [29].
Even though Markov chain discretizations of control problems that are based on grid-based discretizations such as the finite difference method are a popular means to solve stochastic control problems (see [27]), we do not advocate any particular scheme here. Our choice is merely conventional and works fine when the reduced dynamics is low-dimensional. For an efficient, meshless discretizations in case of higher dimensional reduced dynamics, see, e.g., [44, 38].

In accordance with (3.1), a candidate for the suboptimal control is now given by

\[
\hat{u}_s^0 = \left( -\beta^{-1} \frac{\partial \phi_0(s, x^u_s)}{\phi_0(s, x^u_s)}, 0 \right).
\]

Plugging the last expression into (4.2) then yields the controlled dynamics

\[
\begin{align*}
    dx^u_s &= -\frac{\partial V(x^u_s, y^u_s)}{\partial x} ds + \beta^{-1} \frac{\partial \phi_0(s, x^u_s)}{\phi_0(s, x^u_s)} ds + \beta^{-1/2} d\bar{\tilde{w}}_s^1 + \\
    dy^u_s &= -\frac{1}{\epsilon} \frac{\partial V(x^u_s, y^u_s)}{\partial y} ds + \beta^{-1/2} \frac{1}{\sqrt{\epsilon}} d\bar{\tilde{w}}_s^2,
\end{align*}
\]

which will be employed to sample (4.3) using the reweighted estimator (2.8).

**Performance of the suboptimal estimator for varying \(\beta\).** Table 1 shows some numerical results of the Monte Carlo method with the above importance sampling strategy, which should be compared to Table 2 that shows the result of standard Monte Carlo. For the weighted and unweighted estimates, the sample size was set to \(N = 10^4\) trajectories of length \(T = 1\) with time step \(\Delta t \leq 10^{-7}\) that is chosen small enough to remove any possible discretization bias. The suboptimal control was computed by a finite difference method (4.8) on a grid of size \(n_x\). For comparison, we have computed a reference Monte-Carlo solution (“exact mean value”) based on \(N = 10^5\) independent realizations that is displayed in Table 1 in the column with label “I”. In order to quantify the efficacy of the control, we monitor the barrier crossing events \(x_s \geq 0\) for some \(0 < s \leq T\) and call \(R_c\) the ratio of trajectories that cross the barrier among all the trajectories.

Looking at the results shown Tables 1 and 2 more closely, we observe that, for \(\beta = 1\), both methods give acceptable mean values, while the sample variance is smaller when importance
Importance sampling in path space for diffusion processes 15

sampling is used. In the case of standard Monte Carlo, only about 0.2% of trajectories cross the barrier when \( \beta = 5 \), which becomes worse when \( \beta \) is further increased to \( \beta = 10 \), at which no barrier-crossing trajectories are sampled. This observation is in accordance with the fact that the probability of barrier crossing events decays like \( \exp(-\beta \Delta V_I) \) as \( \beta \to \infty \). It therefore does not come as a surprise that the mean values for \( \beta = 10 \) are far off the exact mean value, i.e. the standard Monte Carlo method doesn’t work in this case. In contrast to this, the barrier-crossing trajectories under importance sampling make up about 70% ~ 80% of the trajectories, showing that important events are well sampled; mean values remain close to the correct value when several runs are carried out.

**Performance of the suboptimal estimator for varying \( \epsilon \).** Another aspect we want to address is that performance of the suboptimal importance sampling estimators even for moderate values of \( \epsilon \), i.e. away from the asymptotic regime of the multiple time scale dynamics. Table 1 shows that the exact mean value may vary drastically when \( \epsilon \) varies from \( \epsilon = 0.1 \) to \( \epsilon = 0.001 \), which entails that it is not possible to approximate the exact mean value simply by taking the logarithm of the limiting value function or by Monte Carlo sampling of the averaged dynamics; especially when \( \beta \) is large, resampling of the multiscale dynamics is necessary in order to get a reliable estimate of the correct mean value. Comparing the relative errors \( \rho_s(I) \) in Table 1 and Table 2 we observe again that the importance sampling estimator is much more efficient than standard Monte Carlo for all values of \( \epsilon \in \{0.1, 0.01, 0.001\} \). In order to get a idea of how the control acts in order to enhance sampling of the rare barrier crossing event, Figure 2 shows the control force as a function of \( x \) and \( t \) for various values of \( \beta \). We clearly observe that, for all values of \( \beta \), the control acts against the energy barrier (blue region) where the effect is more pronounced for smaller values of \( t \). We moreover observe that the force converges to a stationary profile as \( \beta = +\infty \) (see below).

Table 1: Numerical results for suboptimal importance sampling with \( T = 1.0 \). Columns \( I \) and \( I_N \) are the mean values computed with \( N = 10^5 \) (“exact”) and \( N = 10^4 \). The columns \( \text{Var}_sI, \rho_s(I) \) display the sample variance and the relative error as defined in (2.21), (2.22). The column \( R_c \) shows the ratio of the trajectories that have crossed the potential barrier.

| \( \beta \) | \( \epsilon \) | \( n_c \) | \( \Delta t \) | \( I \) | \( I_N \) | \( \text{Var}_sI \) | \( \rho_s(I) \) | \( R_c \) |
|---|---|---|---|---|---|---|---|---|
| 1.0 | 0.1 | 2000 | \( 1.0 \times 10^{-7} \) | 5.36 \( \times 10^{-2} \) | 5.38 \( \times 10^{-2} \) | 2.9 \( \times 10^{-4} \) | 0.32 | 6.9 \( \times 10^{-1} \) |
| | 0.01 | | \( 1.0 \times 10^{-8} \) | 4.88 \( \times 10^{-2} \) | 4.88 \( \times 10^{-2} \) | 3.1 \( \times 10^{-5} \) | 0.11 | 6.7 \( \times 10^{-1} \) |
| 5.0 | 0.1 | 5000 | \( 1.0 \times 10^{-7} \) | 2.56 \( \times 10^{-2} \) | 2.70 \( \times 10^{-2} \) | 2.70 \( \times 10^{-4} \) | 0.10 | 6.7 \( \times 10^{-1} \) |
| | 0.01 | | \( 1.0 \times 10^{-8} \) | 2.37 \( \times 10^{-2} \) | 2.48 \( \times 10^{-2} \) | 2.70 \( \times 10^{-5} \) | 0.11 | 6.7 \( \times 10^{-1} \) |
| 10.0 | 0.1 | 8000 | \( 6.0 \times 10^{-7} \) | 1.65 \( \times 10^{-13} \) | 1.70 \( \times 10^{-13} \) | 1.80 \( \times 10^{-12} \) | 0.12 | 8.9 \( \times 10^{-1} \) |
| | 0.01 | | \( 5.0 \times 10^{-8} \) | 3.76 \( \times 10^{-14} \) | 3.80 \( \times 10^{-14} \) | 3.90 \( \times 10^{-13} \) | 0.13 | 8.9 \( \times 10^{-1} \) |

**Performance of the suboptimal viscosity solution estimator.** We also carry out the importance sampling method as discussed in Subsection 3.3, based on the averaged zero temperature dynamics where the limits are taken in the order \( \epsilon \to 0 \) and then \( \beta \to \infty \). Combining
One advantage of the viscosity solution approach is that the deterministic control problem can be
solved “on the fly” for any given $x^u_t = x$, e.g. using the geometric minimum action method [43]. Here we did not minimize the control functional on the fly, but precomputed the control forces in a parallelized fashion by minimization on a coarse grid of size $2000 \times 1000$ for $(x, t)$. The resulting control force $\tilde{u}$ is depicted in the rightmost panel of Figure 2 as a function of $x$ and $t$, and the numerical results of the corresponding importance sampling method are shown in Table 3. By comparison with Table 1 we see that the accuracy and the efficiency is in the range of the averaged stochastic control problem, where the agreement is best when $\beta = 5$ or 10. However, more computation effort is required in order to solve (4.11) rather than (4.10).

Table 2: Numerical results for standard Monte Carlo. The labels are the same as in Table 1.

| $\beta$ | $\epsilon$ | $\Delta t$ | $I_N$ | $\text{Var}_s I$ | $\rho_s(I)$ | $R_c$ |
|--------|-----------|-----------|-----|-----------------|------------|-----|
| 1.0    | 0.1       | $1.0 \times 10^{-7}$ | $5.31 \times 10^{-2}$ | $6.71 \times 10^{-3}$ | 1.54 | $2.6 \times 10^{-1}$ |
|        | 0.01      | $1.0 \times 10^{-8}$ | $4.76 \times 10^{-2}$ | $5.5 \times 10^{-3}$ | 1.56 | $2.6 \times 10^{-1}$ |
|        | 0.001     | $1.0 \times 10^{-8}$ | $4.86 \times 10^{-2}$ | $5.7 \times 10^{-3}$ | 1.55 | $2.6 \times 10^{-1}$ |
| 5.0    | 0.1       | $1.0 \times 10^{-7}$ | $6.21 \times 10^{-7}$ | $2.6 \times 10^{-11}$ | 19.46 | $2.4 \times 10^{-3}$ |
|        | 0.01      | $1.0 \times 10^{-7}$ | $2.90 \times 10^{-7}$ | $2.8 \times 10^{-10}$ | 57.70 | $1.7 \times 10^{-3}$ |
|        | 0.001     | $1.0 \times 10^{-8}$ | $2.80 \times 10^{-7}$ | $2.1 \times 10^{-10}$ | 51.75 | $1.7 \times 10^{-3}$ |
| 10.0   | 0.1       | $1.0 \times 10^{-7}$ | $8.14 \times 10^{-15}$ | $6.6 \times 10^{-26}$ | 31.56 | 0 |
|        | 0.01      | $1.0 \times 10^{-8}$ | $1.31 \times 10^{-15}$ | $2.7 \times 10^{-27}$ | 39.66 | 0 |
|        | 0.001     | $1.0 \times 10^{-8}$ | $6.20 \times 10^{-16}$ | $6.0 \times 10^{-28}$ | 39.51 | 0 |

Table 3: Numerical results for suboptimal importance sampling with $T = 1.0$, based on the
deterministic solution of the deterministic control problem (4.11) – (4.12).

| $\beta$ | $\epsilon$ | $\Delta t$ | $I_N$ | $\text{Var}_s I$ | $\rho_s(I)$ | $R_c$ |
|--------|-----------|-----------|-----|-----------------|------------|-----|
| 1.0    | 0.1       | $1.0 \times 10^{-7}$ | $5.41 \times 10^{-2}$ | $1.4 \times 10^{-3}$ | 0.69 | $8.3 \times 10^{-1}$ |
|        | 0.01      | $1.0 \times 10^{-8}$ | $4.84 \times 10^{-2}$ | $7.1 \times 10^{-4}$ | 0.55 | $8.0 \times 10^{-1}$ |
|        | 0.001     | $1.0 \times 10^{-8}$ | $4.83 \times 10^{-2}$ | $6.5 \times 10^{-4}$ | 0.53 | $8.1 \times 10^{-1}$ |
| 5.0    | 0.1       | $1.0 \times 10^{-7}$ | $3.25 \times 10^{-7}$ | $2.1 \times 10^{-13}$ | 1.41 | $8.5 \times 10^{-1}$ |
|        | 0.01      | $1.0 \times 10^{-7}$ | $1.92 \times 10^{-7}$ | $1.7 \times 10^{-14}$ | 0.68 | $8.3 \times 10^{-1}$ |
|        | 0.001     | $1.0 \times 10^{-8}$ | $1.70 \times 10^{-7}$ | $8.0 \times 10^{-15}$ | 0.53 | $8.7 \times 10^{-1}$ |
| 10.0   | 0.1       | $1.0 \times 10^{-7}$ | $1.54 \times 10^{-13}$ | $1.5 \times 10^{-25}$ | 2.51 | $9.0 \times 10^{-1}$ |
|        | 0.01      | $1.0 \times 10^{-8}$ | $3.82 \times 10^{-14}$ | $7.9 \times 10^{-28}$ | 0.74 | $9.0 \times 10^{-1}$ |
|        | 0.001     | $1.0 \times 10^{-8}$ | $3.23 \times 10^{-14}$ | $1.6 \times 10^{-28}$ | 0.39 | $9.0 \times 10^{-1}$ |
4.2 Motion in a multiscale potential

The following example is inspired by the works [14,36]. We consider the one-dimensional SDE

$$dx_s = -\nabla V(x_s)ds + \beta^{-1/2}dw_s,$$

with a periodically perturbed potential $V^\epsilon(x) = V(x) + p(x/\epsilon)$, where $p(x)$ is a $\gamma$-periodic function for some $\gamma > 0$. When $\epsilon \ll \gamma$, the potential function $V^\epsilon$ is the sum of a smooth potential function $V$ and a highly oscillatory potential $p(x/\epsilon)$. The aim is to compute the conditional expectation

$$\phi(t, x) = \mathbb{E}\left( \exp\left( -\beta \int_t^T h(x_s, x_s/\epsilon) ds \right) \bigg| x_t = x \right).$$

Setting $U = -\beta^{-1} \log \phi$ and using the duality relation (2.12), we know that $U$ is related to a stochastic optimal control problem with value function

$$U(t, x) = \inf_u \mathbb{E} \left( \int_t^T \left( h(x^u_s, x^u_s/\epsilon) + \frac{1}{2} |u_s|^2 \right) ds \bigg| x^u_t = x \right),$$

and the dynamics

$$dx^u_s = -\frac{1}{\epsilon} p'(x^u_s/\epsilon)ds - V'(x^u_s)ds - u_sds + \beta^{-1/2}dw_s.$$  

Fig. 2: The first three figures show the limiting control forces computed from (4.9) for different $\beta$ as functions of $x$ and $t$. The rightmost figure ($\beta = +\infty$) shows the control force in (1.13), which is computed by solving the averaged deterministic optimal control problem (4.11)–(4.12). 

$$-2.8 \, -2.4 \, -2.0 \, -1.6 \, -1.2 \, -0.8 \, -0.4 \, 0.0$$

$$-3 \, -2 \, -1 \, 0 \, 1 \, 2 \, 3$$

$$\beta = 1.0 \quad \beta = 5.0 \quad \beta = 10.0 \quad \beta = +\infty$$
To connect the latter with the homogenization problem mentioned in Remark 3.2 of Subsection 3.1, we introduce the auxiliary variable \( y = x/\epsilon \), by which (4.14) turns out to be equivalent to the following singularly perturbed system of equations

\[
\begin{align*}
\frac{dx_s}{ds} &= -\frac{1}{\epsilon}p'(y_s)ds - V'(x_s)ds + \beta^{-1/2}dw_s,
\frac{dy_s}{ds} &= -\frac{1}{\epsilon^2}p'(y_s)ds - \frac{1}{\epsilon}V'(x_s)ds + \frac{\beta^{-1/2}}{\epsilon}dw_s.
\end{align*}
\] (4.16)

Now consider the expectation

\[ \tilde{\phi}(t, x, y) = \mathbb{E}\left( \exp \left( -\beta \int_t^T h(x_s, y_s) \, ds \mid x_t = x, y_t = y \right) \right), \]

and the associated optimal control problem

\[ \tilde{U}(t, x, y) = \inf_u \mathbb{E} \left( \int_t^T \left( h(x_s^u, y_s^u) + \frac{1}{2} |u_s|^2 \right) \, ds \mid x_t^u = x, y_t^u = y \right), \]

with the dynamics

\[
\begin{align*}
\frac{dx_s^u}{ds} &= -\frac{1}{\epsilon}p'(y_s^u)ds - V'(x_s^u)ds - u_sds + \beta^{-1/2}dw_s,
\frac{dy_s^u}{ds} &= -\frac{1}{\epsilon^2}p'(y_s^u)ds - \frac{1}{\epsilon}V'(x_s^u)ds - \frac{1}{\epsilon}u_sds + \frac{\beta^{-1/2}}{\epsilon}dw_s.
\end{align*}
\]

Notice that the same noise and the same control are applied to both variables. Thus \( y_s = x_s/\epsilon \) for all \( t \leq s \leq T \), and the dual relation \( \tilde{U}(t, x, y) = -\beta^{-1} \log \tilde{\phi}(t, x, y) \) holds with the identification \( U(t, x) = \tilde{U}(t, x, x/\epsilon) \). Following the procedure of Subsection 3.1, we aim at computing the leading term of \( \tilde{U}(t, x, y) \) which satisfies a limiting control problem with value function

\[ U_0(t, x) = \inf_u \mathbb{E} \left( \int_t^T \left( \tilde{h}(\bar{x}_s^u) + \frac{1}{2} |u_s|^2 \right) \, ds \mid \bar{x}_t^u = x \right) \] (4.17)

under the homogenized dynamics

\[ d\tilde{x}_s^u = -K V'(\bar{x}_s^u)ds - \sqrt{K} u_sds + \sqrt{K} \beta^{-1/2}dw_s, \] (4.18)

where

\[ K = \int (I + \Phi'(y))(I + \Phi'(y))^T \rho(y)dy, \] (4.19)

\( \tilde{h} \) is given in (3.17). \( \rho(y) = \rho_s(y) \), the conditional invariant density of fast dynamics \( y_s \), and \( \Phi(y) \) solves the cell problem \( \mathcal{L}_0 \Phi = p' \), with \( \mathcal{L}_0 \) being the infinitesimal generator of the fast dynamics \( y_s \) when the slow variable \( x \) is held fixed (see [37][35][36] for the derivation).

**Suboptimal estimator based on a homogenized equation.** In our particular case, the limiting equation (4.18)–(4.19) can be explicitly computed [29]. Letting \( \phi_1 = \phi_0 + \epsilon \phi_1 + o(\epsilon) \), formal asymptotic expansion of (4.16) yields that \( \phi_0 \) is independent of \( y \). The second perturbation terms \( \phi_1(t, x, y) \) is \( \gamma \)-periodic in \( y \) and satisfies the Poisson equation

\[ \mathcal{L}_1 \phi_0 = -\mathcal{L}_1 \phi_0 = p' \frac{\partial \phi_0}{\partial x}. \]
where $\mathcal{L}_0$, $\mathcal{L}_1$ are second and first order differential operators given by $\mathcal{L}_0 = -p' \partial_y + \frac{1}{2} \beta \partial_y^2$ and $\mathcal{L}_1 = -p' \partial_x - V' \partial_y$. Here, $\mathcal{L}_0$ denotes the infinitesimal generator of the fast dynamics $y$, that has a unique invariant measure with a density $\rho(y) \propto \exp(-2\beta y)$ that is independent of $x$. The corresponding cell problem $\mathcal{L}_0 \phi(y) = p'$ can be solved analytically:

$$\phi(y) = -y + \frac{\gamma}{\int_0^y e^{2\beta y} \, dz} \int_0^y e^{2\beta y} \, dz,$$

(4.20)

The first order term is then given by $\phi_1(t, x, y) = \phi(y) \nabla_x \phi_0$. Calling

$$L = \int_0^\gamma e^{2\beta y} \, dz, \quad \tilde{L} = \int_0^\gamma e^{-2\beta y} \, dz$$

then, by (4.19), we have

$$K = \frac{\gamma^2}{LL}.$$

Then, from the duality relation (2.12), the optimal feedback control (2.16), and the perturbation expansion $\tilde{\phi}(t, x, x/\epsilon) = \phi_0(t, x) + \epsilon \phi_1(t, x, x/\epsilon) + O(\epsilon^2)$, it holds that

$$\tilde{u}_s = -\beta^{-1} \partial_x \tilde{\phi}(s, x^0_s, x^0_s/\epsilon) \phi(s, x^0_s, x^0_s/\epsilon)$$

$$= -\beta^{-1} \frac{\partial_x \phi_0(s, x^0_s) + \partial_y \phi_1(s, x^0_s) + O(\epsilon)}{\phi_0(s, x^0_s) + O(\epsilon)} + O(\epsilon)$$

$$= -\beta^{-1} \gamma e^{2\beta y} \frac{\partial_x \phi_0(s, x^0_s)}{L} + O(\epsilon).$$

(4.21)

Table 4: Numerical results for standard Monte Carlo and importance sampling (IS) based on the homogenized dynamics. Here $K = 0.407728$ when $\beta = 5.0$ and $K = 0.055302$ when $\beta = 10.0$. The column labelled “I” shows the “exact” values based on a sample of size $N = 10^5$.

| $\beta$ | $\epsilon$ | $n_x$ | $\Delta t$ | $I$ | IS | Standard |
|---------|-------------|-------|------------|-----|-----|----------|
|         |             |       |            | $I_N$ | $\text{Var}_I$ | $\rho_s(I)$ | $I_N$ | $\text{Var}_I$ | $\rho_s(I)$ |
| 5.0     | 0.05        | 5000  | $1 \times 10^{-7}$ | 0.496 | 0.497 | 2.4 x 10^{-3} | 9.8 x 10^{-2} | 0.496 | 3.0 x 10^{-2} | 0.04 |
|         | 0.02        | 5000  | $1 \times 10^{-7}$ | 0.426 | 0.426 | 4.3 x 10^{-3} | 4.9 x 10^{-2} | 0.427 | 3.4 x 10^{-2} | 0.43 |
|         | 0.01        | 5000  | $1 \times 10^{-7}$ | 0.436 | 0.436 | 1.0 x 10^{-4} | 2.3 x 10^{-2} | 0.435 | 3.3 x 10^{-2} | 0.42 |
|         | 0.008       | 5000  | $5 \times 10^{-8}$ | 0.438 | 0.438 | 6.4 x 10^{-5} | 1.8 x 10^{-2} | 0.438 | 3.3 x 10^{-2} | 0.41 |
|         | 0.005       | 5000  | $5 \times 10^{-8}$ | 0.442 | 0.441 | 3.0 x 10^{-5} | 1.2 x 10^{-2} | 0.443 | 3.4 x 10^{-2} | 0.42 |
|         | 0.002       | 5000  | $5 \times 10^{-8}$ | 0.446 | 0.446 | 4.0 x 10^{-5} | 1.4 x 10^{-2} | 0.448 | 3.3 x 10^{-2} | 0.41 |
| 10.0    | 0.05        | 2000  | $1 \times 10^{-6}$ | 0.198 | 0.198 | 1.3 x 10^{-3} | 1.8 x 10^{-1} | 0.198 | 5.8 x 10^{-3} | 0.38 |
|         | 0.02        | 2000  | $5 \times 10^{-6}$ | 0.104 | 0.104 | 3.0 x 10^{-4} | 1.7 x 10^{-1} | 0.104 | 4.4 x 10^{-3} | 0.64 |
|         | 0.01        | 2000  | $5 \times 10^{-6}$ | 0.109 | 0.109 | 5.8 x 10^{-5} | 7.0 x 10^{-2} | 0.109 | 2.3 x 10^{-3} | 0.44 |
|         | 0.008       | 2000  | $5 \times 10^{-7}$ | 0.111 | 0.111 | 3.0 x 10^{-5} | 4.9 x 10^{-2} | 0.111 | 2.0 x 10^{-3} | 0.40 |
|         | 0.006       | 2000  | $5 \times 10^{-7}$ | 0.114 | 0.114 | 1.7 x 10^{-5} | 3.6 x 10^{-2} | 0.114 | 1.6 x 10^{-3} | 0.35 |
|         | 0.004       | 2000  | $1 \times 10^{-7}$ | 0.117 | 0.117 | 7.7 x 10^{-6} | 2.4 x 10^{-2} | 0.117 | 1.8 x 10^{-3} | 0.36 |
|         | 0.002       | 2000  | $5 \times 10^{-8}$ | 0.120 | 0.120 | 4.1 x 10^{-6} | 1.7 x 10^{-2} | 0.120 | 1.7 x 10^{-3} | 0.34 |
Performance of the suboptimal estimator for varying $\epsilon$ and $\beta$. For the numerical test, we set the coefficients and simulation parameters equal to

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

Thus $p(x)$ is periodic with period $\gamma = 2\pi$. We compute $\phi_0$ and hence $\hat{u}_0$ as in the previous example in Section 4.1 with the homogenized generator given by

$$
\mathcal{L} = -KV'(x) \frac{\partial}{\partial x} + \frac{K}{2\beta} \frac{\partial^2}{\partial x^2}.
$$

(4.22)

The corresponding importance sampling estimator will be based on the control force $\hat{u}_0$ as defined in (4.21), so that instead of running the dynamics (4.14) we generate realizations of

$$
dx_s = -V'(x_s)ds - \frac{1}{\epsilon} p'(x_s/\epsilon)ds - \hat{u}_0^0 ds + \beta^{-1/2}d\bar{w}_s, \quad 0 \leq s \leq T.
$$

(4.23)

The numerical simulation results for $\beta = 5.0$ and 10.0 with different values of $\epsilon$ are shown in Table 4. Here, again, $n_x$ is the mesh size used to compute $\phi_0$, $N = 10^4$ is the number of independent realizations used in standard Monte Carlo and importance sampling, and the column “$I$” contains the “exact mean value” based on $N = 10^5$ independent realizations of (4.14). As before, the discretization time step $\Delta t$ is chosen small enough to avoid discretization bias. The results are in agreement with the averaging case that was extensively discussed in Section 3.1. By inspecting Table 4, we can see that, for all values of $\epsilon$ or $\beta$, importance sampling reduces the variance comparing to standard Monte Carlo. Furthermore, holding $\beta$ fixed, we clearly observe that sample variance of the suboptimal importance sampling estimator decreases as $\epsilon$ goes to zero, which agrees with the results presented in [20] for the closely related optimal control problem on an indefinite time-horizon.

For illustration, Figure 3 shows the non-oscillatory (homogenized) part of the control force, which drives the dynamics closer towards $x = 0$ where $h$ attains its minimum.

5 Proof of the main result

In this section, we prove our main result, Theorem 3.1 in Section 3.1. Since $\beta$ is fixed, it can be absorbed into the noise coefficients $\alpha_1$ and $\alpha_2$, $h$, and thus we can assume $\beta = 1$ without loss of generality. Our analysis is based on the solution $\phi$ of the linear backward evolution equation (3.12) and the solution $\phi_0$ of (3.10) where, by the Feynman-Kac formula, both $\phi$ and $\phi_0$ can be expressed in terms of conditional expectations like (5.1).

Idea of the proof. Under Assumption 1 it is well known that both $\phi$ and $\phi_0$ are $C^4$ function [10,18,17] and that, using the probabilistic representation (3.18), their derivatives have explicit representation formulas in terms of conditional expectations:

$$
\partial_{x_i} \phi = -\mathbb{E} \mathbb{E}^{u,y} \left[ e^{-\int_t^T h(x_s)ds} \int_t^T \nabla_x h(x_s) \cdot x_{s+i}ds \right], \quad 1 \leq i \leq k
$$

$$
\partial_{y_i} \phi = -\mathbb{E} \mathbb{E}^{u,y} \left[ e^{-\int_t^T h(x_s)ds} \int_t^T \nabla_x h(x_s) \cdot x_{s+i}ds \right], \quad 1 \leq i \leq l
$$

(5.1)

$$
\partial_{x_i} \phi_0 = -\mathbb{E} \left[ e^{-\int_t^T h(x_s)ds} \int_t^T \nabla_x h(x_s) \cdot \tilde{x}_{s+i}ds \right], \quad 1 \leq i \leq k
$$
Importance sampling in path space for diffusion processes

To this end, we follow [30] and define a partition of the interval $[0, T]$ by $[0, \Delta], [\Delta, 2\Delta], \ldots$.

(That is, the derivatives can be pulled inside the expectation, see [13,8,9].) Here, we have used that, by Assumption 3 the running cost $h$ depends only on $x$, and that $x_s, y_s$ and $\tilde{x}_s$ satisfy (3.1) and (3.3). Moreover, we have employed the shorthand $\mathbf{E}^{x,y}$ to denote the expectation conditioned on $x_t = x, y_t = y$ and similarly for $\mathbf{E}^x$.

The partial derivatives $x_{s,x_i} \in \mathbb{R}^k, y_{s,x_i} \in \mathbb{R}^l$ in (5.1) satisfy

$$
\begin{align}
    dx_{s,x_i} &= (\nabla_x f x_{s,x_i} + \nabla_y f y_{s,x_i}) ds + (\nabla_x \alpha_1 x_{s,x_i} + \nabla_y \alpha_1 y_{s,x_i}) dw_s^1 \\
    dy_{s,x_i} &= \frac{1}{\epsilon} (\nabla_x g x_{s,x_i} + \nabla_y g y_{s,x_i}) ds + \frac{1}{\sqrt{\epsilon}} (\nabla_x \alpha_2 x_{s,x_i} + \nabla_y \alpha_2 y_{s,x_i}) dw_s^2 \\
\end{align}
$$

with $x_{t,x_i} = \delta_{ij}, 1 \leq j \leq k, y_{t,x_i} = 0 \in \mathbb{R}^l$. Analogously, $x_{s,y_i} \in \mathbb{R}^k$ and $y_{s,y_i} \in \mathbb{R}^l$ satisfy

$$
\begin{align}
    dx_{s,y_i} &= (\nabla_x f x_{s,y_i} + \nabla_y f y_{s,y_i}) ds + (\nabla_x \alpha_1 x_{s,y_i} + \nabla_y \alpha_1 y_{s,y_i}) dw_s^1 \\
    dy_{s,y_i} &= \frac{1}{\epsilon} (\nabla_x g x_{s,y_i} + \nabla_y g y_{s,y_i}) ds + \frac{1}{\sqrt{\epsilon}} (\nabla_x \alpha_2 x_{s,y_i} + \nabla_y \alpha_2 y_{s,y_i}) dw_s^2 \\
\end{align}
$$

with $x_{t,y_i} = 0 \in \mathbb{R}^k, y_{t,y_i} = \delta_{ij} \in \mathbb{R}^l, 1 \leq j \leq l$ (Notice that the above dynamics also hold when $\alpha_1$ depends on both $x, y$, so terms involving $\nabla_y \alpha_1$ are kept there). The above formulas (5.1) - (5.3) will allow us to compare the dynamics $x_s, y_s, \tilde{x}_s$, the controlled dynamics and the resulting importance sampling estimators on the basis of pathwise estimates. For simplicity, we consider the dynamics on $[0, T]$ that entails similar estimates for the case $s \in [t, T]$. We therefore suppose that the initial values of $x_s, \tilde{x}_s$ are $x_0$ and the initial value of $y_s$ is $y_0$.

To prove Theorem 5.1 we will adapt some estimates used in [30]; cf. also [13,31,23,18].

Fig. 3: Homogenized feedback control for $\beta = 5, 10$ as a function of $x$ and $t$. 

$\begin{align}
\beta &= 5.0 \\
\beta &= 10.0
\end{align}$

$\begin{array}{cc}
0.0 & 0.2 & 0.4 & 0.6 & 0.8 & 1.0 \\
-2.0 & -1.2 & -0.4 & 0.4 & 1.2 & 2.0
\end{array}$
\[(M - 1)\Delta, M\Delta\] with \(\Delta = T/M, \ M > 0,\) and consider the auxiliary process

\[
d\hat{x}_s = f(x_{j\Delta}, \hat{y}_s)ds + \alpha_1(x_s)dw^1_s
\]
\[
d\hat{y}_s = \frac{1}{\epsilon}g(x_{j\Delta}, \hat{y}_s)ds + \frac{1}{\sqrt{\epsilon}}\alpha_2(x_{j\Delta}, \hat{y}_s)dw^2_s
\]

(5.4)

for \(s \in [j\Delta, (j + 1)\Delta), 0 \leq j \leq (M - 1),\) with the continuity condition

\[
\hat{x}_{(j+1)\Delta} = \lim_{s \to (j+1)\Delta^-} \hat{x}_s, \quad \hat{y}_{(j+1)\Delta} = \lim_{s \to (j+1)\Delta^-} \hat{y}_s.
\]

and initial conditions \(\hat{x}_0 = x_0, \hat{y}_0 = y_0.\) Without loss of generality, we can suppose that \(\Delta \leq 1.\)

This auxiliary process will serve as a bridge between (3.1) and (3.3). In contrast to [30] and owed to the fact that we consider controlled dynamics, estimates for 4th-order moments as well as for the processes (5.2) and (5.3) will be needed to prove the theorem.

Before entering the details of the estimates, we first summarize our main technical results, the proofs of which will be given in the following subsections.

For the derivative processes satisfying (5.2) and (5.3), we have (see Theorem 5.6 and Lemma 5.4 below):

**Theorem 5.1** Let Assumptions 1–3 hold. Then \(\exists C > 0,\) independent of \(\epsilon, x_0\) and \(y_0,\) such that

\[
\max_{0 \leq s \leq T} \mathbb{E}|x_s, x_i|^2 \leq C, \quad \max_{0 \leq s \leq T} \mathbb{E}|y_s, x_i|^2 \leq C, \quad 1 \leq i \leq k.
\]

\[
\max_{0 \leq s \leq T} \mathbb{E}|x_s, y_i|^2 \leq C\epsilon^2, \quad \mathbb{E}|y_t, y_i|^2 \leq e^{-\lambda t} + C\epsilon^2, \quad t \in [0, T], \quad 1 \leq l.
\]

For the approximation results, we have (see Theorem 5.7 and Theorem 5.8 below):

**Theorem 5.2** Let Assumptions 1–3 hold. Then \(\exists C > 0,\) independent of \(\epsilon\) (possibly depending on \(x_0\) and \(y_0\)), such that

\[
\mathbb{E}|x_s - \tilde{x}_s|^4 \leq C\epsilon^2, \quad s \in [0, T].
\]

**Theorem 5.3** Let Assumptions 1–3 hold. Then \(\exists C > 0,\) independent of \(\epsilon\) (possibly depending on \(x_0\) and \(y_0\)), such that

\[
\mathbb{E}|x_s, x_i - \tilde{x}_s, x_i|^2 \leq C\epsilon^2, \quad s \in [0, T].
\]

From these results that will be proved in the remainder of this Section we then obtain:

**Theorem 5.4** Let Assumptions 1–3 hold. Then \(\exists C > 0,\) independent of \(\epsilon\) (possibly depending on \(x\) and \(y\)), such that

1.

\[
|\nabla_y \phi' | \leq C\epsilon, \quad |\nabla_x \phi' - \nabla_x \phi_0| \leq C\epsilon^2
\]

(5.5)
2. For $U^c = -\log \phi^c$, $U_0 = -\log \phi_0$, we have

$$|\nabla_y U^c| \leq C\epsilon, \quad |\nabla_x U^c - \nabla_x U_0| \leq C\epsilon^{\frac{1}{2}}$$

(5.6)

**Proof** We use (5.1). For $\nabla_y \phi^c$, using Assumption 1 and Theorem 5.1 we have

$$|\partial_y \phi^c| \leq \mathbb{E} \left( e^{-\int_0^T h(x_s) ds} \int_0^T |\nabla_x h(x_s)| |x_{s,y_i}| ds \right)$$

$$\leq C \mathbb{E} \int_0^T |x_{s,y_i}| ds \leq C \int_0^T \left( \mathbb{E} |x_{s,y_i}|^2 \right)^{\frac{1}{2}} ds \leq C$$

To compare $\nabla_x \phi^c$ with $\nabla_x \phi_0$, we use that

$$|\partial_x \phi^c - \partial_x \phi_0|$$

$$\leq \left| \mathbb{E} \left[ e^{-\int_0^T h(x_s) ds} \left( \int_0^T (\nabla_x h(x_s) \cdot x_{s,x_i} - \nabla_x h(\tilde{x}_s) \cdot \tilde{x}_{s,x_i}) ds \right) \right] \right|$$

$$+ \left| \mathbb{E} \left[ \left( e^{-\int_0^T h(x_s) ds} - e^{-\int_0^T h(\tilde{x}) ds} \right) \left( \int_0^T \nabla_x h(\tilde{x}_s) \cdot \tilde{x}_{s,x_i} ds \right) \right] \right|$$

$$= I_1 + I_2$$

For $I_1$, using Assumption 1, Theorem 5.2 and Theorem 5.3, it follows that

$$I_1 \leq C \mathbb{E} \left[ \int_0^T \left( |x_s - \tilde{x}_s| |x_{s,x_i}| + |x_{s,x_i} - \tilde{x}_{s,x_i}| \right) ds \right]$$

$$\leq C \int_0^T \left( \mathbb{E} |x_s - \tilde{x}_s|^2 \right)^{\frac{1}{2}} \left( \mathbb{E} |x_{s,x_i}|^2 \right)^{\frac{1}{2}} + \left( \mathbb{E} |x_{s,x_i} - \tilde{x}_{s,x_i}|^2 \right)^{\frac{1}{2}} ds \leq C\epsilon^{\frac{1}{2}}$$

For $I_2$, we have

$$I_2 \leq \left[ \mathbb{E} \left( e^{-\int_0^T h(x_s) ds} - e^{-\int_0^T h(\tilde{x}) ds} \right)^2 \right]^{\frac{1}{2}} \left[ \mathbb{E} \left( \int_0^T \nabla_x h(\tilde{x}_s) \cdot \tilde{x}_{s,x_i} ds \right)^2 \right]$$

$$\leq C \left\{ \mathbb{E} \left[ \int_0^T e^{-\int_0^T (1-r) h(x_s) + r h(\tilde{x}) ds} \left( \int_0^T |h(\tilde{x}_s) - h(x_s)| dr \right) \right] \right\}^{\frac{1}{2}} \left[ \mathbb{E} \int_0^T |\tilde{x}_{s,x_i}|^2 ds \right]^{\frac{1}{2}}$$

$$\leq C \left( \int_0^T \mathbb{E} |\tilde{x}_s - x_s|^2 ds \right)^{\frac{1}{2}} \leq C\epsilon^{\frac{1}{2}}$$

which then entails the estimates for $\phi^c$. The estimates for $U^c$, $U_0$ follows from the fact that $\phi^c \to \phi^0$ uniformly on any bounded subset of $[0,T] \times \mathbb{R}^k \times \mathbb{R}^l$ with rate $\epsilon$ (see [30,30,16]), and the fact that $e^{-C(T-t)} \leq \phi^c \leq 1$ is uniformly bounded for all $\epsilon > 0$.

Recall that, in Section 2 and Subsection 3.1, $\tilde{u}$ is the optimal control as given by (2.10) and that the averaged force $\tilde{u}^0$ as defined in (3.8) is a candidate for the suboptimal control which is used for estimating (2.1) with nearly optimal variance. Theorem (3.1) that is entailed by the above results expresses this fact, and we restate it for the readers’ convenience:

**Theorem 5.5** Let Assumptions 1 and 2 hold, and consider the importance sampling method for computing (2.1) under the dynamics (3.1). When the control $\tilde{u}^0$ as given in (3.8) is used to perform the importance sampling, the relative error (2.17) of the Monte Carlo estimator satisfies

$$\rho_{\tilde{u}^0}(I) \leq C\epsilon^{\frac{1}{2}}$$

for $\epsilon \ll 1$ where $C > 0$ is a constant independent of $\epsilon$. 
Then Theorem 5.4 implies that $\hat{\phi}$ converges uniformly to $\bar{\phi}^0$ on any bounded subset of $[0, T] \times \mathbb{R}^k \times \mathbb{R}^l$. Furthermore, we know both of them are uniformly bounded on $[0, T] \times \mathbb{R}^k \times \mathbb{R}^l$ from the boundness of $\bar{\phi}', \alpha_1, \alpha_2$ and formula (5.1).

Now call $\tilde{x}_s^u, \tilde{y}_s^u$ the controlled dynamics (2.7) corresponding to the control $\tilde{u}_s = 2\bar{u}_s - \hat{u}_s^0$. Further let $R > 0$. For $y \in \mathbb{R}^l$, we define $\chi_R(y) = 1$, if $|y| \leq R$, and $\chi_R(y) = 0$ otherwise. Similarly, for $x \in \mathbb{R}^k, y \in \mathbb{R}^l$, we define $\chi_R(x, y) = 1$, if $|x| \leq R$ and $|y| \leq R$, otherwise $\chi_R(x, y) = 0$. Then we can recast (2.20) as

$$
\tilde{E} \left[ \exp \left( \int_t^T |\tilde{u}_s - \hat{u}_s^0|^2 \chi_R(\bar{x}_s^u, \bar{y}_s^u)ds + \int_t^T |\hat{u}_s - \hat{u}_s^0|^2 (1 - \chi_R(\bar{x}_s^u, \bar{y}_s^u))ds \right) \right] 
\leq e^{C_\eta(T-t)\epsilon^2} \tilde{E} \left[ \exp \left( \int_t^T |\tilde{u}_s - \hat{u}_s^0|^2 (1 - \chi_R(\bar{x}_s^u, \bar{y}_s^u))ds \right) \right] 
\leq e^{C_\eta(T-t)\epsilon^2} \tilde{E} \left[ \exp \left( C \int_t^T (1 - \chi_R(\bar{x}_s^u, \bar{y}_s^u))ds \right) \right] 
\leq e^{C_\eta(T-t)\epsilon^2} \left[ e^{C_\eta} + e^{CT} \tilde{P} \left( \int_t^T (1 - \chi_R(\bar{x}_s^u, \bar{y}_s^u))ds \geq \delta \right) \right]
$$

where $C_R$ is a constant that depends on $R$. Since $\tilde{u}_s$ is uniformly bounded, it readily follows that Lemma 5.2 and Lemma 5.3 in Subsection 5.2 hold for the dynamics $\tilde{x}_s^u, \tilde{y}_s^u$ (see pp. 26 and 26).

Therefore, the above quantity is bounded by

$$
e^{C_\eta(T-t)\epsilon^2} \left[ e^{C_\eta} + e^{CT} \frac{CT(1 + |x|^4 + |y|^4)}{\delta R^4} \right]
$$

Now for any $\delta > 0$ we can choose $R > 0$ such that

$$
\tilde{E} \left( \exp \left( \int_t^T |\tilde{u}_s - \hat{u}_s^0|^2 ds \right) \right) \leq 2e^{CT} \epsilon \delta R^4
$$

where $C > 0$ is independent of $\epsilon$. Combing this with (2.6) and (2.10), we conclude that

$$
\rho_\bar{\phi}(t) \leq C_\epsilon \delta R^4
$$

whenever $\epsilon$ is sufficiently small.

5.1 Estimates for $x_{s,y_i}$ and $y_{s,y_i}$

We consider $x_{s,y_i}$ and $y_{s,y_i}$ first, since the arguments are simpler and largely unrelated to the rest of the proof. In the following and throughout this section, we denote by $C$ a generic constant that is independent of $\epsilon$ and which value may change from line to line.

**Lemma 5.1** Under Assumptions 7, 8 there exists $C > 0$, independent of $\epsilon$, $x_0$ and $y_0$, such that

$$
\max_{0 \leq s \leq T} \mathbb{E}|x_{s,y_i}|^2 \leq C \epsilon, \quad \mathbb{E}|y_{s,y_i}|^2 \leq e^{-\frac{4}{\epsilon^2}} + C \epsilon, \quad t \in [0, T], \quad 1 \leq i \leq l.
$$

(5.7)
Proof Applying Ito’s formula to $|x_{s,y_i}|^2$ and $|y_{s,y_i}|^2$, equation (5.3) yields

$$d\mathbb{E}[x_{s,y_i}]^2 = 2\mathbb{E}(\nabla_x f x_{s,y_i}, x_{s,y_i})ds + 2\mathbb{E}(\nabla_y f y_{s,y_i}, x_{s,y_i})ds + \mathbb{E}\|\nabla_x \alpha_1 x_{s,y_i} + \nabla_y \alpha_1 y_{s,y_i}\|^2ds$$

$$d\mathbb{E}[y_{s,y_i}]^2 = \frac{2}{\epsilon}\mathbb{E}(\nabla_y g y_{s,y_i}, y_{s,y_i})ds + \frac{2}{\epsilon}\mathbb{E}(\nabla_y g y_{s,y_i}, y_{s,y_i})ds + \frac{1}{\epsilon}\mathbb{E}\|\nabla_x \alpha_2 x_{s,y_i} + \nabla_y \alpha_2 y_{s,y_i}\|^2ds$$

(5.8)

Then, by using the Cauchy-Schwarz inequality, Lipschitz continuity of the SDE coefficients (Assumption 2) and Remark 2 on page 9, it follows that

$$\frac{d\mathbb{E}[x_{s,y_i}]}{ds} \leq C(\mathbb{E}[x_{s,y_i}]^2 + \mathbb{E}[y_{s,y_i}]^2)$$

$$\frac{d\mathbb{E}[y_{s,y_i}]}{ds} \leq -\frac{\lambda}{\epsilon}\mathbb{E}[y_{s,y_i}]^2 + \frac{C}{\epsilon}\mathbb{E}[x_{s,y_i}]^2$$

(5.9)

with $\mathbb{E}[x_{0,y_i}]^2 = 0$, $\mathbb{E}[y_{0,y_i}]^2 = 1$. The conclusion then follows from Claim [A.1] in Appendix A.

The last result can be improved if we additionally impose Assumption 3 and if we treat the (transient) initial layer at $t = 0$ more carefully.

**Theorem 5.6** Let Assumptions 4, 5 hold. Then $\exists C > 0$, independent of $t$, $x_0$ and $y_0$, such that

$$\max_{0 \leq s \leq T} \mathbb{E}[x_{s,y_i}]^2 \leq C\epsilon^2$$

$$\mathbb{E}[y_{s,y_i}]^2 \leq e^{-\frac{\lambda t}{\epsilon}} + C\epsilon^2$$

$t \in [0,T]$, $1 \leq i \leq l$.

Proof Set $t_1 = -\frac{2 \ln \lambda}{\epsilon}$ and introduce the function $\eta: [0,T] \rightarrow [0,1]$ by

$$\eta(t) = \begin{cases} 
1 - \frac{t}{t_1}, & 0 \leq t \leq t_1 \\
0, & t_1 < t \leq T 
\end{cases}$$

(5.10)

Then

$$\mathbb{E}(\nabla_y f y_{s,y_i}, x_{s,y_i}) \leq C \left( \epsilon^{-\eta(s)} \mathbb{E}[x_{s,y_i}]^2 + \epsilon^{\eta(s)} \mathbb{E}[y_{s,y_i}]^2 \right)$$

$$\mathbb{E}(\nabla_y g x_{s,y_i}, y_{s,y_i}) \leq C \left( \frac{\mathbb{E}[x_{s,y_i}]^2}{2} + \frac{\mathbb{E}[y_{s,y_i}]^2}{2} \right),$$

from which, using a similar argument as in (5.9), we thus obtain

$$\frac{d\mathbb{E}[x_{s,y_i}]}{ds} \leq C(1 + \epsilon^{-\eta(s)})\mathbb{E}[x_{s,y_i}]^2 + C\epsilon^{\eta(s)}\mathbb{E}[y_{s,y_i}]^2$$

$$\frac{d\mathbb{E}[y_{s,y_i}]}{ds} \leq -\frac{\lambda}{\epsilon}\mathbb{E}[y_{s,y_i}]^2 + \frac{C}{\epsilon}\mathbb{E}[x_{s,y_i}]^2$$

with $\mathbb{E}[x_{0,y_i}]^2 = 0$, $\mathbb{E}[y_{0,y_i}]^2 = 1$. The conclusion follows from Claim [A.2] in Appendix A.
5.2 Stability estimates

We start with some basic facts related to the stability of the dynamics (3.1), (3.3), (5.2) and (5.4). Bear in mind that $\beta = 1$ throughout this section. Then, for $x_s, y_s$ satisfying (3.1), we have:

**Lemma 5.2** Under Assumption 1, 2, there exists $C > 0$, independent of $\epsilon, x_0$ and $y_0$, such that

$$
\max_{0 \leq s \leq T} E|x_s|^4 \leq C(|x_0|^4 + |y_0|^4 + 1), \quad \max_{0 \leq s \leq T} E|y_s|^4 \leq C(|y_0|^4 + |x_0|^4 + 1). \tag{5.11}
$$

**Proof** Ito’s formula implies that

$$
\frac{dE|x_s|^4}{ds} \leq 4E\left(|x_s|^2(f(x_s, y_s), x_s)\right) + 6E\left(|x_s|^2\alpha_1(x_s, y_s)\right),
$$

$$
\frac{dE|y_s|^4}{ds} \leq \frac{4}{\epsilon}E\left(|y_s|^2(g(x_s, y_s), y_s)\right) + \frac{6}{\epsilon}E\left(|y_s|^2\alpha_2(x_s, y_s)\right).
$$

By Assumption 1 $f$ is Lipschitz and $\alpha_1$ is bounded. Hence we obtain (cf. Remark 2)

$$
\frac{dE|x_s|^4}{ds} \leq C\left(E|x_s|^4 + E|y_s|^4 + 1\right),
$$

$$
\frac{dE|y_s|^4}{ds} \leq -\frac{\lambda}{\epsilon}E|y_s|^4 + \frac{C}{\epsilon}\left(E|x_s|^4 + 1\right).
$$

An argument similar to the one in Claim A.1 of Appendix A provides us with the desired estimates.

**Remark 4** Reiterating the last step, it follows that the solutions of (5.4) and (5.3) satisfy

$$
\max_{0 \leq s \leq T} E|x_s|^4 \leq C(|x_0|^4 + |y_0|^4 + 1), \quad \max_{0 \leq s \leq T} E|y_s|^4 \leq C(|y_0|^4 + |x_0|^4 + 1). \tag{5.12}
$$

and

$$
\max_{0 \leq s \leq T} E|\tilde{x}_s|^4 \leq C(|x_0|^4 + 1), \tag{5.13}
$$

since $\tilde{f}$ is Lipschitz as well (Remark 2).

The above results entail estimates for the supremum of the solution $x_s$ of the averaged SDE (3.1), as well as for the occupation times of $y_s$ on arbitrary finite time intervals:

**Lemma 5.3** Letting Assumptions 1, 2 hold, there exists $C > 0$, independent of $\epsilon, x_0$ and $y_0$, such that

$$
E\left(\sup_{0 \leq s \leq T} |x_s|^4\right) \leq C(T^4 + 1)(1 + |x_0|^4 + |y_0|^4)
$$

Moreover, for all $\delta, R > 0$, it holds

$$
P\left(\int_0^T (1 - \chi_R(y_s))ds \geq \delta\right) \leq \frac{CT(1 + |x_0|^4 + |y_0|^4)}{\delta R^4},
$$

$$
P\left(\int_0^T (1 - \chi_R(x_s, y_s))ds \geq \delta\right) \leq \frac{CT(1 + |x_0|^4 + |y_0|^4)}{\delta R^4}$$

where $\chi_R$ denotes the characteristic function of $[-R, R]$. Theorem 5.1 will now follow from the next result, which is a consequence of Lemma 5.3.
The first integral in the last equation can be bounded using Lemma 5.2, whereas the second one.

Taking first the supremum and then the expected value in both sides, we find

\[ \mathbb{E}(\sup_{0 \leq s \leq T} |x_s|^4) \leq C \left( |x_0|^4 + T^4 \mathbb{E} \left( \int_0^T (|x_s|^4 + |y_s|^4 + 1) \, ds \right) \right). \]

The first integral in the last equation can be bounded using Lemma 5.2 whereas the second one is bounded by the maximal martingale inequality [25]. Hence

\[ \mathbb{E}(\sup_{0 \leq s \leq T} |x_s|^4) \leq C(T^4 + 1)(|x_0|^4 + |y_0|^4 + 1) + C \left( \mathbb{E} \left( \int_0^T |\alpha_1(x_s, y_s)|^2 \, ds \right)^2 \right). \]

and the boundness of \( \alpha_1 \) entails

\[ \mathbb{E}(\sup_{0 \leq s \leq T} |x_s|^4) \leq C(T^4 + 1)(1 + |x_0|^4 + |y_0|^4). \]

As for the second part of the assertion, notice that for all \( \delta > 0 \) and \( R > 0 \) it holds:

\[ R^4 \mathbb{E} \left( \int_0^T (1 - \chi_R(y_s)) \, ds \right) \leq \mathbb{E} \left( \int_0^T |y_s|^4 (1 - \chi_R(y_s)) \, ds \right) \leq \mathbb{E} \left( \int_0^T |y_s|^4 \, ds \right) \leq CT(1 + |x_0|^4 + |y_0|^4) \]

Thus, by Chebyshev’s inequality,

\[ \mathbb{P} \left( \int_0^T (1 - \chi_R(y_s)) \, ds \geq \delta \right) \leq \frac{CT(1 + |x_0|^4 + |y_0|^4)}{\delta R^4} \]

The second inequality follows in the same fashion.

We proceed our analysis by inspecting the SDE \(5.2\) for \( x_{s,x_i}, y_{s,x_i} \), for which we seek the analogue of the inequality \(5.3\). In this case the initial values satisfy \( \mathbb{E}|x_{0,x_i}|^2 = 1, \mathbb{E}|y_{0,x_i}|^2 = 0 \) and by similar argument as in the proof of Lemma 5.1 we find:

**Lemma 5.4** Under Assumptions 7.13 there exists \( C > 0 \), independent of \( \epsilon, x_0 \) and \( y_0 \), such that

\[ \max_{0 \leq s \leq T} \mathbb{E}|x_{s,x_i}|^2 \leq C, \quad \max_{0 \leq s \leq T} \mathbb{E}|y_{s,x_i}|^2 \leq C, \quad 1 \leq i \leq k. \quad (5.14) \]

Upper bounds on 4th moments can be obtained in the same manner:

**Lemma 5.5** Under Assumptions 7.13 there exists \( C > 0 \), independent of \( \epsilon, x_0 \) and \( y_0 \), such that

\[ \max_{0 \leq s \leq T} \mathbb{E}|x_{s,x_i}|^4 \leq C, \quad \max_{0 \leq s \leq T} \mathbb{E}|y_{s,x_i}|^4 \leq C, \quad 1 \leq i \leq k. \quad (5.15) \]
Proof By Ito’s formula,
\[ dE|x_{s,x_i}|^4 = 4E \left( |x_{s,x_i}|^2 (\nabla_x f x_{s,x_i} + \nabla_y f y_{s,x_i}, x_{s,x_i}) \right) ds + 2E \left( |x_{s,x_i}|^2 (\nabla_x \alpha_1 x_{s,x_i} + \nabla_y \alpha_1 y_{s,x_i}) || y_{s,x_i} \right) ds \]
\[ + 4E \left( |\nabla_x \alpha_1 x_{s,x_i} + \nabla_y \alpha_1 y_{s,x_i}, x_{s,x_i}|^2 \right) ds \]
\[ dE|y_{s,x_i}|^4 = \frac{4}{\epsilon} E \left( |y_{s,x_i}|^2 (\nabla_x g x_{s,x_i} + \nabla_y g y_{s,x_i}, y_{s,x_i}) \right) ds + \frac{2}{\epsilon} E \left( |y_{s,x_i}|^2 (\nabla_x \alpha_2 x_{s,x_i} + \nabla_y \alpha_2 y_{s,x_i}) || y_{s,x_i} \right) ds \]
\[ + \frac{4}{\epsilon} E \left( |\nabla_x \alpha_2 x_{s,x_i} + \nabla_y \alpha_2 y_{s,x_i}, y_{s,x_i}|^2 \right) ds \]
(5.16)

Assumption \([1]\) (cf. Remark \([2]\)) and the Cauchy-Schwarz inequality now readily imply that
\[ \frac{dE|x_{s,x_i}|^4}{ds} \leq C (E|x_{s,x_i}|^4 + E|y_{s,x_i}|^4) \]
\[ \frac{dE|y_{s,x_i}|^4}{ds} \leq -\frac{2\lambda}{\epsilon} E|y_{s,x_i}|^4 + \frac{C}{\epsilon} E|x_{s,x_i}|^4, \]
with \(E|y_{0,x_i}|^4 = 0, E|x_{0,x_i}|^4 = 1\). The assertion then follows by the same argument as the in the proof of Claim \([A.1]\) in Appendix \([A]\).

For the discretized processes, we have the following bounds.

Lemma 5.6 Let \(\Delta \leq 1, s \in [j\Delta, (j + 1)\Delta], 0 \leq j \leq M - 1\). Further let Assumptions \([1]\) hold. Then, for all \(p \geq 1\), we have
\[ E|x_{s,x_i} - x_{j\Delta,x_i}|^{2p} \leq C' (s - j\Delta)^p \leq C\Delta^p, \]
(5.17)

with constants \(C', C\) that are independent of \(\epsilon, x_0, y_0\). The same inequality holds if \(x_{s,x_i}\) is replaced by the process \(\tilde{x}_{s,x_i}\) and \(\tilde{x}_{s,x_i}\) (and the corresponding discrete counterpart).

For \(x_s\), it holds
\[ E|x_s - x_{j\Delta}|^4 \leq C(s - j\Delta)^2, \]
(5.18)

with a constant \(C\) that may depend on \(x_0\) and \(y_0\). The same bound is satisfied by the processes \(\tilde{x}_s, \tilde{x}_s\) (and their discretizations).

Proof The first part of the Lemma follows from Lemma \([5.1]\) and the moment inequalities for martingales, since coefficients in \([5.2]\) are bounded.

As for the second part, related to the processes \(x_s, \tilde{x}_s, \tilde{y}_s\), using that \(f\) is Lipschitz and \(\alpha_1\) is bounded (Assumption \([1]\)) and Lemma \([5.2]\) we can conclude that
\[ E|x_s - x_{j\Delta}|^4 \leq C \left( \int_{j\Delta}^s (1 + |x_r| + |y_r|) dr \right)^4 \leq C \left( \int_{j\Delta}^s \alpha_1(x_r, y_r) dw_r^y \right)^4 \]
\[ \leq C \left| x_0 \right|^4 + \left| y_0 \right|^4 + 1(s - j\Delta)^4 + C(s - j\Delta)^2 \leq C(s - j\Delta)^2 \]

where, in the last inequality, we have taken advantage of the fact that \(\Delta \leq 1\).
5.3 Approximation by the auxiliary process

In this subsection, we will study the approximations of the original dynamics (3.1) by the auxiliary discrete process (5.4) and the reduced dynamics (3.3). To this end, we recall Hölder and Young inequalities: Given two random variables $X, Y$, and $\frac{1}{p} + \frac{1}{q} = 1, p, q > 0$, it holds that

$$
\mathbb{E}|XY| \leq (\mathbb{E}|X|^p)^{\frac{1}{p}} (\mathbb{E}|Y|^q)^{\frac{1}{q}} \leq \frac{\mathbb{E}|X|^p}{p} + \frac{\mathbb{E}|Y|^q}{q}.
$$

**Lemma 5.7** Suppose that Assumptions $[1,2]$ are met. Then there exists $C > 0$, independent of $\epsilon$ (possibly depending on $x_0, y_0$), such that

$$
\mathbb{E}|y_s - \hat{y}_s|^4 \leq C \Delta^2, \quad \mathbb{E}|x_s - \hat{x}_s|^4 \leq C \Delta^2 \quad (5.19)
$$

**Proof** Let $j = \lfloor \frac{s}{\Delta} \rfloor$. By Itô’s formula, using Assumptions $[1,2]$ we have

\[
\frac{d\mathbb{E}|y_s - \hat{y}_s|^4}{ds} = 4\mathbb{E}\left([y_s - \hat{y}_s]^2|y_s - \hat{y}_s, g(x_s, y_s) - g(x_{j\Delta}, \hat{y}_s)\right) + \frac{4}{\epsilon}\mathbb{E}\left(||\alpha_2(x_s, y_s) - \alpha_2(x_{j\Delta}, \hat{y}_s)||^2\right)
\]

\[
\leq 4\mathbb{E}\left([y_s - \hat{y}_s]^2|y_s - \hat{y}_s, g(x_s, y_s) - g(x_{j\Delta}, \hat{y}_s)\right) + \frac{6}{\epsilon}\mathbb{E}\left(||\alpha_2(x_s, y_s) - \alpha_2(x_{j\Delta}, \hat{y}_s)||^2\right)
\]

\[
\leq 4\mathbb{E}\left([y_s - \hat{y}_s]^2|y_s - \hat{y}_s, g(x_s, y_s) - g(x_{j\Delta}, \hat{y}_s)\right) + \frac{12C}{\epsilon}\mathbb{E}\left(|y_s - \hat{y}_s|^3|x_s - x_{j\Delta}|\right) + \frac{4C}{\epsilon}\mathbb{E}\left(|y_s - \hat{y}_s|^3|x_s - x_{j\Delta}|\right)
\]

\[
\leq -2\lambda\mathbb{E}|y_s - \hat{y}_s|^4 + \frac{C}{\epsilon}\mathbb{E}|x_s - x_{j\Delta}|^4
\]

\[
\leq -2\lambda\mathbb{E}|y_s - \hat{y}_s|^4 + \frac{C}{\epsilon}, \Delta^2
\]

which, by Gronwall’s Lemma, yields the first inequality. The second inequality follows from

\[
\frac{d\mathbb{E}|x_s - \hat{x}_s|^4}{ds} = 4\mathbb{E}\left(|\dot{x}_s - x_s|^2|f(x_{j\Delta}, \hat{y}_s) - f(x_s, y_s), \dot{x}_s - x_s|\right)
\]

\[
\leq C\mathbb{E}\left(|\dot{x}_s - x_s|^3(|x_{j\Delta} - x_s| + |\hat{y}_s - y_s|\right)
\]

\[
\leq C\left(\mathbb{E}|x_s - x_{j\Delta}|^4 + \mathbb{E}|x_{j\Delta} - x_s|^4 + \mathbb{E}|\hat{y}_s - y_s|^4\right)
\]

\[
\leq C\mathbb{E}|x_s - x_{j\Delta}|^4 + C\Delta^2
\]

The following auxiliary result will be use below.

**Claim 5.1** Let $F(x) = |x|^2x$. Then $|F(x) - F(y)| \leq \frac{1}{2}(|x|^2 + |y|^2)|x - y|$. 
\[ |F(x) - F(y)| = \left| \int_0^1 \frac{d}{dt} F((1-t)x + ty) dt \right| \]
\[ = \int_0^1 \left[ 2((1-t)x + ty, y - x)((1-t)x + ty) + ((1-t)x + ty)^2(y - x) \right] dt \]
\[ \leq 3 \int_0^1 ((1-t)x + ty)^2|y - x| dt \leq \frac{3}{2}(|x|^2 + |y|^2)|x - y| \]

As the next step, we show that the averaged process \( \tilde{x}_s \) can be approximated by the time-discrete process as well.

**Lemma 5.8** Under Assumptions 1–3, there exists \( C > 0 \), independent of \( \epsilon \) (possibly depending on \( x_0 \) and \( y_0 \)), such that

\[
\max_{0 \leq s \leq T} \mathbb{E}|\hat{x}_s - \tilde{x}_s|^4 \leq C \left( \frac{\epsilon}{\lambda \Delta} + \Delta \right) e^{C(1 + \frac{1}{\epsilon \lambda \Delta}) T} \tag{5.20}
\]

Especially, for \( \Delta = \epsilon^{1/2} \), we have \( \max_{0 \leq s \leq T} \mathbb{E}|\hat{x}_s - \tilde{x}_s|^4 \leq Ce^T \).

**Proof** Applying Ito’s formula to \( |\hat{x}_s - \tilde{x}_s|^4 \), integrating both sides and taking expectations, we obtain

\[
\mathbb{E}|\tilde{x}_s - \hat{x}_s|^4 \leq 4 \int_0^s \mathbb{E}\left(|\hat{x}_r - \tilde{x}_r|^2|\hat{x}_r - \hat{x}_r, f(x|\frac{x}{\Delta}, \tilde{y}_r) - \tilde{f}(\tilde{x}_r)\right) dr + 6 \int_0^s \mathbb{E}\left(|\hat{x}_r - \tilde{x}_r|^2|\alpha_1(x_r) - \alpha_1(\tilde{x}_r)|^2\right) dr
\]
\[= 4 \int_0^s \mathbb{E}\left(F(\hat{x}_r|\Delta) - \tilde{x}_r|\Delta, f(x|\frac{x}{\Delta}, \tilde{y}_r) - \tilde{f}(\tilde{x}_r)\right) dr
\]
\[+ 4 \int_0^s \mathbb{E}\left(F(\hat{x}_r - \tilde{x}_r) - F(\hat{x}_r|\Delta) - \tilde{x}_r|\Delta, f(x|\frac{x}{\Delta}, \tilde{y}_r) - \tilde{f}(\tilde{x}_r)\right) dr
\]
\[+ 4 \int_0^s \mathbb{E}\left(F(\hat{x}_r - \tilde{x}_r), \tilde{f}(\tilde{x}_r) - \tilde{f}(\tilde{x}_r)\right) dr
\]
\[+ 6 \int_0^s \mathbb{E}\left(|\hat{x}_r - \tilde{x}_r|^2|\alpha_1(x_r) - \alpha_1(\tilde{x}_r)|^2\right) dr
\]
\[= I_1 + I_2 + I_3 + I_4
\]
We estimate the 4 terms in the sum separately.
\[
\begin{align*}
|I_1| & \leq 4 \sum_{j=0}^{[s/\lambda]} \int_{j\Delta}^{(j+1)\Delta} \mathbb{E}\left( |\tilde{x}_{j\Delta} - \tilde{x}_{j\Delta}|^4 \mathbb{E}_j f(x_{j\Delta}, \tilde{y}_r) - \tilde{f}(x_{j\Delta}) | \right) dr \\
& \leq C \sum_{j=0}^{[s/\lambda]} \int_{j\Delta}^{(j+1)\Delta} \mathbb{E}\left( |\tilde{x}_{j\Delta} - \tilde{x}_{j\Delta}|^3 (|x_{j\Delta}| + |\tilde{y}_r| + 1) e^{-\frac{\lambda j\Delta}{r}} \right) dr \\
& \leq \frac{C}{\lambda} \mathbb{E}\left[ \left( \sum_{j=0}^{[s/\lambda]} |\tilde{x}_{j\Delta} - \tilde{x}_{j\Delta}|^1 \left( \sum_{j=0}^{[s/\lambda]} (|x_{j\Delta}| + |\tilde{y}_r| + 1)^{1/2} \right) \right)^2 \right] \\
& \leq \frac{C}{\lambda} \mathbb{E}\left[ \left( \sum_{j=0}^{[s/\lambda]} |\tilde{x}_{j\Delta} - \tilde{x}_{j\Delta}|^4 + \mathbb{E} \sum_{j=0}^{[s/\lambda]} (|x_{j\Delta}| + |\tilde{y}_r| + 1)^4 \right) \right] \\
& \leq \frac{C}{\lambda} \mathbb{E}\left[ \int_0^s |\tilde{x}_r - \tilde{y}_r|^4 dr + \frac{C}{\lambda} \mathbb{E}\int_0^s |\tilde{x}_r - \tilde{y}_r|^2 dr \leq \frac{C}{\lambda} \mathbb{E}\left[ \int_0^s |\tilde{x}_r - \tilde{y}_r|^4 dr + \frac{C}{\lambda} \mathbb{E}\int_0^s |\tilde{x}_r - \tilde{y}_r|^2 dr \right] \\
& \leq \frac{C}{\lambda} \mathbb{E}\left[ \int_0^s |\tilde{x}_r - \tilde{y}_r|^4 dr + C s \Delta + \Delta^2 \right]
\end{align*}
\]

In the first inequality, \( \mathbb{E}_{j\Delta} \) denotes the expectation conditional \( \tilde{y}_r \) at time \( s = j\Delta \), and Lemma \[B.3\] in Appendix \[B\] is used to derive the second inequality. Therefore, by Lemma \[5.2\] and Remark \[4\]

\[
|I_1| \leq \frac{C}{\lambda} \mathbb{E}\int_0^s |\tilde{x}_r - \tilde{y}_r|^4 dr + \frac{C s \Delta}{\lambda} \Delta.
\]

For \( I_2 \), using Claim \[5.1\] the fact that \( f, \tilde{f} \) are Lipschitz, and Lemmas \[5.2\] and \[5.6\] we conclude that
\[
\begin{align*}
|I_2| & \leq C \mathbb{E}\int_0^s |\tilde{x}_r - \tilde{y}_r|^2 + |\tilde{x}_r - \tilde{x}_r|^2 \\
& \times (\tilde{x}_r - \tilde{y}_r) \left( 1 + |\tilde{x}_{\Delta}| + |y_r| \right) dr \\
& \leq C \mathbb{E}\int_0^s |\tilde{x}_r - \tilde{y}_r|^2 + |\tilde{x}_r - \tilde{x}_r|^2 \\
& \times (\tilde{x}_r - \tilde{x}_r) \left( 1 + |\tilde{x}_{\Delta}| + |y_r| \right) dr \\
& \leq C \mathbb{E}\int_0^s |\tilde{x}_r - \tilde{y}_r|^4 dr + C \mathbb{E}\int_0^s |\tilde{x}_r - \tilde{y}_r|^2 (1 + |\tilde{x}_{\Delta}| + |y_r|)^2 dr \\
& \leq C \int_0^s |\tilde{x}_r - \tilde{y}_r|^2 dr \\
& \leq C \int_0^s \left( \mathbb{E}(\tilde{x}_r - \tilde{y}_r)^4 \right)^{1/2} (1 + |\tilde{x}_{\Delta}| + |y_r|)^{1/2} dr \\
& \leq \frac{C}{\lambda} \mathbb{E}\int_0^s |\tilde{x}_r - \tilde{y}_r|^4 dr + C s (\Delta + \Delta^2)
\end{align*}
\]
For $I_3$, since $\tilde{f}$ is Lipschitz, we have

$$|I_3| \leq C E \int_0^s |\tilde{x}_r - \tilde{x}_r|^3 |x_{1/4} | \Delta - \tilde{x}_r | dr$$

$$\leq C E \int_0^s |\tilde{x}_r - \tilde{x}_r|^4 dr + C E \int_0^s |\tilde{x}_r - \tilde{x}_r|^3 |x_{1/4} | |x_r| dr + C E \int_0^s |\tilde{x}_r - \tilde{x}_r|^3 |x_r - \tilde{x}_r| dr$$

$$\leq C E \int_0^s |\tilde{x}_r - \tilde{x}_r|^4 dr + C s \Delta^2$$

Finally, since $\alpha_1$ is Lipschitz, we obtain the following bound for $I_4$:

$$|I_4| \leq C E \int_0^s |\tilde{x}_r - \tilde{x}_r|^2 |x_r - \tilde{x}_r|^2 dr$$

$$\leq C E \int_0^s |\tilde{x}_r - \tilde{x}_r|^4 dr + C E \int_0^s |\tilde{x}_r - \tilde{x}_r|^2 |x_r - \tilde{x}_r|^2 dr$$

$$\leq C E \int_0^s |\tilde{x}_r - \tilde{x}_r|^4 dr + C s \Delta^2.$$

We thus have proved the bound (assuming $\Delta \leq 1$)

$$E[\tilde{x}_s - \tilde{x}_s]^4 \leq C \left(1 + \frac{e}{\lambda \Delta}\right) E \int_0^s |\tilde{x}_r - \tilde{x}_r|^4 dr + C s \left(\frac{e}{\lambda \Delta} + \Delta\right)$$

(5.21)

and Gronwall’s Lemma yields the assertion with

$$E[|\tilde{x}_s - \tilde{x}_s|^4] \leq C \left(\frac{e}{\lambda \Delta} + \Delta\right) e^{C(1 + \frac{e}{\Delta})^s}$$

(5.22)

Summarizing Lemma 5.7 and Lemma 5.8, we have proved the following estimate for the 4th moments of the process (see [30] for stronger result on the 2nd moments):

**Theorem 5.7** Suppose that Assumption 1–3 hold. Then there exists $C > 0$, independent of $\epsilon$ (possibly depending on $x_0$ and $y_0$), such that

$$E[|x_s - \tilde{x}_s|^4] \leq C e^{\frac{s}{\Delta}}, \quad s \in [0,T].$$

As the next step, we consider the derivative of the time-discrete auxiliary process (5.23)

$$d\tilde{x}_{s,x_i} = (\nabla_x f x_j, x_i) + \frac{1}{\epsilon} \nabla_x g x_j, x_i) ds + \frac{1}{\sqrt{\epsilon}} \nabla_x \alpha_2 \tilde{x}_s x_{j, x_i} dw_s$$

where $j = \lceil \frac{s}{\Delta} \rceil$. The following lemma shows that (5.23) is an approximation of (5.2).

**Lemma 5.9** Under Assumptions 1–3 there exists $C > 0$, independent of $\epsilon$ (possibly depending on $x_0$ and $y_0$), such that

$$E[\tilde{y}_{s,x_i - \tilde{y}_{s,x_i}}^2] \leq C \Delta, \quad E[|x_{s,x_i - \tilde{x}_{s,x_i}}|^2] \leq C \Delta$$

(5.24)
Proof Let $j = \lfloor \frac{i}{\epsilon} \rfloor$. Using Itô’s formula, it follows that
\[
\frac{dE(y_{s,x_i} - \hat{y}_{s,x_i})^2}{ds} = 2E(\nabla_x g(x_s, y_s)x_{s,x_i} - \nabla_x g(x_{j,\Delta} \hat{y}_s)x_{j,\Delta,x_i}, y_{s,x_i} - \hat{y}_{s,x_i})
\]
\[
+ 2E(\nabla_y g(x_s, y_s)y_{s,x_i} - \nabla_y g(x_{j,\Delta} \hat{y}_s)\hat{y}_{s,x_i}, y_{s,x_i} - \hat{y}_{s,x_i})
\]
\[
+ \frac{1}{\epsilon} E\left( \left\| \nabla_x \alpha_2(x_s, y_s)x_{s,x_i} + \nabla_y \alpha_2(x_s, y_s)y_{s,x_i} - \nabla_x \alpha_2(x_{j,\Delta} \hat{y}_s)x_{j,\Delta,x_i} - \nabla_y \alpha_2(x_{j,\Delta} \hat{y}_s)\hat{y}_{s,x_i} \right\|^2 \right)
\]

The first term can be bounded by
\[
E(\nabla_x g(x_s, y_s)x_{s,x_i} - \nabla_x g(x_{j,\Delta} \hat{y}_s)x_{j,\Delta,x_i}, y_{s,x_i} - \hat{y}_{s,x_i})
\]
\[
= E\left( \left\| \nabla_x g(x_s, y_s) - \nabla_x g(x_{j,\Delta} \hat{y}_s) \right\| y_{s,x_i} + \nabla_y g(x_{j,\Delta} \hat{y}_s)(y_{s,x_i} - \hat{y}_{s,x_i}), y_{s,x_i} - \hat{y}_{s,x_i} \right)\]
\[
\leq C\left[ (E|\nabla_x x_i|^4)^{1/2} (E|x_s - x_{j,\Delta}|^4 + E|y_s - \hat{y}_s|^4)^{1/2} + \frac{\lambda}{4} E|y_{s,x_i} - \hat{y}_{s,x_i}|^2 \right]
\]

Inspecting the second term, we find
\[
E(\nabla_y g(x_s, y_s)y_{s,x_i} - \nabla_y g(x_{j,\Delta} \hat{y}_s)\hat{y}_{s,x_i}, y_{s,x_i} - \hat{y}_{s,x_i})
\]
\[
= E\left( \left\| \nabla_y g(x_s, y_s) - \nabla_y g(x_{j,\Delta} \hat{y}_s) \right\| y_{s,x_i} + \nabla_x g(x_{j,\Delta} \hat{y}_s)(y_{s,x_i} - \hat{y}_{s,x_i}), y_{s,x_i} - \hat{y}_{s,x_i} \right)\]
\[
\leq C\left[ (E|\nabla_y y_s|^4)^{1/2} (E|x_s - x_{j,\Delta}|^4 + E|y_s - \hat{y}_s|^4)^{1/2} + \frac{\lambda}{4} E|y_{s,x_i} - \hat{y}_{s,x_i}|^2 \right]
\]

For the third term, we have
\[
E\left( \left\| \nabla_x \alpha_2(x_s, y_s)x_{s,x_i} + \nabla_y \alpha_2(x_s, y_s)y_{s,x_i} - \nabla_x \alpha_2(x_{j,\Delta} \hat{y}_s)x_{j,\Delta,x_i} - \nabla_y \alpha_2(x_{j,\Delta} \hat{y}_s)\hat{y}_{s,x_i} \right\|^2 \right)
\]
\[
\leq 2E\left( \left\| \nabla_x \alpha_2(x_s, y_s)x_{s,x_i} - \nabla_x \alpha_2(x_{j,\Delta} \hat{y}_s)x_{j,\Delta,x_i} \right\|^2 \right)
\]
\[
+ 2E\left( \left\| \nabla_y \alpha_2(x_s, y_s)y_{s,x_i} - \nabla_y \alpha_2(x_{j,\Delta} \hat{y}_s)\hat{y}_{s,x_i} \right\|^2 \right)
\]
\[
\leq C\left[ (E|\nabla_x x_i|^4)^{1/2} + (E|\nabla_y y_s|^4)^{1/2} \right] \left( E|x_s - x_{j,\Delta}|^4 + E|y_s - \hat{y}_s|^4 \right)^{1/2} + E|x_{s,x_i} - x_{j,\Delta,x_i}|^2
\]
\[
+ 4E\left( \left\| \nabla_y \alpha_2(x_{j,\Delta} \hat{y}_s)(y_{s,x_i} - \hat{y}_{s,x_i}) \right\|^2 \right)
\]

Now by Lemma 5.5, Lemma 5.6, Lemma 5.7 and Assumption 2 we conclude that
\[
\frac{dE|y_{s,x_i} - \hat{y}_{s,x_i}|^2}{ds} \leq -\frac{\lambda}{\epsilon} E|y_{s,x_i} - \hat{y}_{s,x_i}|^2 + \frac{C\Delta}{\epsilon}
\]

and the first part of the assertion follows from Gronwall’s Lemma. In the same way, we can conclude that
\[
\frac{dE|x_{s,x_i} - \hat{x}_{s,x_i}|^2}{ds}
\]
\[
\leq -\frac{\lambda}{\epsilon} E|x_{s,x_i} - \hat{x}_{s,x_i}|^2 + C\Delta
\]

which, by Gronwall’s Lemma, implies the second part of the assertion.
We continue our journey by comparing $\tilde{x}_{s,x_i}$ with $\tilde{x}_{s,x_i}$, where
\[
d\tilde{x}_{s,x_i} = \nabla \tilde{f}(\tilde{x}_s)\tilde{x}_{s,x_i} ds + \nabla \alpha_1(\tilde{x}_s)\tilde{x}_{s,x_i} dw_s^1, \tag{5.25}
\]
Recalling (3.4) we can write
\[
\tilde{f}(\tilde{x}_s) = \mathbb{E}^{\xi_t}(f(\tilde{x}_s, \xi_t^s)),
\]
\[
\nabla \tilde{f}(\tilde{x}_s)\tilde{x}_{s,x_i} = \mathbb{E}^{\xi_t}(\nabla_x f(\tilde{x}_s, \xi_t^s) + \nabla_y f(\tilde{x}_s, \xi_t^s)\xi_t^s)\tilde{x}_{s,x_i},
\]
where $\xi_t^s$ is the stationary process defined in Appendix B and $\mathbb{E}^{\xi_t}$ denotes the expectation with respect to the stationary process.

**Lemma 5.10** Let $\Delta = \epsilon^{1/2}$, and let Assumptions 1–3 be satisfied. Then there exists $C > 0$, independent of $\epsilon$ (possibly depending on $x_0, y_0$), such that
\[
\max_{0 \leq t \leq T} \mathbb{E}|\tilde{x}_{s,x_i} - \tilde{x}_{s,x_i}|^2 \leq C \epsilon^\Delta.
\]

**Proof** Let $j = \lfloor \frac{\epsilon}{\Delta} \rfloor$. By Ito's formula, we have
\[
\mathbb{E}|\tilde{x}_{s,x_i} - \tilde{x}_{s,x_i}|^2
\]
\[
= 2 \int_0^s \mathbb{E}(\nabla_x f(x_{j,t}, \tilde{y}_r) x_{j,t,x_i} + \nabla_y f(x_{j,t}, \tilde{y}_r) \tilde{y}_{r,x_i} - \nabla_x \tilde{f}(\tilde{x}_r)\tilde{x}_{r,x_i} - \tilde{x}_{r,x_i}) dr
\]
\[
+ \int_0^s \mathbb{E}\|\nabla_x \alpha_1(x_r) x_{r,x_i} - \nabla_x \alpha_1(\tilde{x}_r)\tilde{x}_{r,x_i}\|^2 dr
\]
\[
= 2 \int_0^s \mathbb{E}(\nabla_x f(x_{j,t}, \tilde{y}_r) x_{j,t,x_i} - \mathbb{E}^{\xi_t}(\nabla_x f(\tilde{x}_r, \xi_t^s))\tilde{x}_{r,x_i} - \tilde{x}_{r,x_i}) dr
\]
\[
+ 2 \int_0^s \mathbb{E}(\nabla_y f(x_{j,t}, \tilde{y}_r) \tilde{y}_{r,x_i} - \mathbb{E}^{\xi_t}(\nabla_y f(\tilde{x}_r, \xi_t^s)\xi_t^s)\tilde{x}_{r,x_i} - \tilde{x}_{r,x_i}) dr
\]
\[
+ \int_0^s \mathbb{E}\|\nabla_x \alpha_1(x_r) x_{r,x_i} - \nabla_x \alpha_1(\tilde{x}_r)\tilde{x}_{r,x_i}\|^2 dr
\]
\[
= I_1 + I_2 + I_3.
\]

Using the notations in Appendix B we can identify $\tilde{y}_r$ with $\xi_{j,t,r}^{x_i}$ and $\tilde{y}_{r,x_i}$ with $\xi_{j,t,r,x_i}^{x_i}$. Then, the term $I_1$ on the right hand side can be recast as
\[
\int_0^s \mathbb{E}(\nabla_x f(x_{j,t}, \xi_{j,t}^{x_i}) x_{j,t,x_i} - \mathbb{E}^{\xi_t}(\nabla_x f(\tilde{x}_r, \xi_t^s))\tilde{x}_{r,x_i} - \tilde{x}_{r,x_i}) dr
\]
\[
= \int_0^s \mathbb{E}(\nabla_x f(x_{j,t}, \xi_{j,t}^{x_i}) x_{j,t,x_i} - \mathbb{E}^{\xi_t}(\nabla_x f(x_{j,t}, \xi_{j,t}^{x_i}))) x_{j,t,x_i} - \tilde{x}_{r,x_i}) dr
\]
\[
+ \int_0^s \mathbb{E}(\mathbb{E}^{\xi_t}(\nabla_x f(\tilde{x}_r, \xi_t^s))) \tilde{x}_{r,x_i} - \tilde{x}_{r,x_i}) dr
\]
\[
+ \int_0^s \mathbb{E}(\mathbb{E}^{\xi_t}(\nabla_x f(\tilde{x}_r, \xi_t^s))\tilde{x}_{r,x_i} - \tilde{x}_{r,x_i}) dr
\]
\[
= I_{1,1} + I_{1,2} + I_{1,3} + I_{1,4}
\]
For $I_{1,1}$, using Lemma B.3 in Appendix B and Lemma 5.6, we have

$$|I_{1,1}| \leq \left| \int_0^s E\left( \nabla_x f(x_{j\Delta}^\Delta, \xi_{j\Delta}^\Delta) x_{j\Delta, x_i} - E^{\xi_{t\Delta}}(\nabla_x f(x_{j\Delta, x_i}, \xi_{j\Delta, x_i}) x_{j\Delta, x_i}, \tilde{x}_{j\Delta, x_i} - \tilde{x}_{j\Delta, x_i})dr \right) + \int_0^s E\left( \nabla_x f(x_{j\Delta}^\Delta, \xi_{j\Delta}^\Delta) x_{j\Delta, x_i} - E^{\xi_{t\Delta}}(\nabla_x f(x_{j\Delta, x_i}, \xi_{j\Delta, x_i}) x_{j\Delta, x_i}, \tilde{x}_{r, x_i} - \tilde{x}_{j\Delta, x_i})dr \right) + \int_0^s E\left( \nabla_x f(x_{j\Delta}^\Delta, \xi_{j\Delta}^\Delta) x_{j\Delta, x_i} - E^{\xi_{t\Delta}}(\nabla_x f(x_{j\Delta, x_i}, \xi_{j\Delta, x_i}) x_{j\Delta, x_i}, \tilde{x}_{r, x_i} - \tilde{x}_{j\Delta, x_i})dr \right) \right|$$

$$\leq C \sum_{j=0}^{\lfloor s/\Delta \rfloor} \int_0^{(j+1)\Delta} E\left( (1 + |x_{j\Delta}| + |y_{j\Delta}|)|x_{j\Delta, x_i}| |\tilde{x}_{j\Delta, x_i} - \tilde{x}_{j\Delta, x_i}| \right) e^{-\frac{\lambda |r| \Delta}{2}} dr + C s \Delta^\frac{3}{2}$$

$$\leq C e \sum_{j=0}^{\lfloor s/\Delta \rfloor} E|\tilde{x}_{j\Delta, x_i} - \tilde{x}_{j\Delta, x_i}|^2 + Cs(\Delta^\frac{3}{2} + \epsilon) \leq C e \sum_{j=0}^{\lfloor s/\Delta \rfloor} E|\tilde{x}_{r, x_i} - \tilde{x}_{r, x_i}|^2 dr + Cs(\Delta^\frac{3}{2} + \epsilon)$$

For $I_{1,2}$, since $f$ is Lipschitz, it follows that

$$|I_{1,2}| \leq C \int_0^s E|\tilde{x}_{r, x_i} - \tilde{x}_{r, x_i}|^2 dr$$

For $I_{1,3}$, Lemma B.4 implies that

$$\left| E^{\xi_{t\Delta}}(\nabla_x f(\tilde{x}_r, \xi_{t\Delta}^r)) - E^{\xi_{t\Delta}}(\nabla_x f(\tilde{x}_{r}, \xi_{t\Delta}^r)) \right| \leq CE^{\xi_{t\Delta}}(|\tilde{x}_r - \tilde{x}_r| + |\xi_{t\Delta}^r - \xi_{t\Delta}^r|) \leq C|\tilde{x}_r - \tilde{x}_r|$$

and therefore

$$|I_{1,3}| \leq C \int_0^s E(|\tilde{x}_r - \tilde{x}_r| |\tilde{x}_{r, x_i} - \tilde{x}_{r, x_i}|) dr$$

$$\leq C \int_0^s E|\tilde{x}_{r, x_i} - \tilde{x}_{r, x_i}|^2 dr + C \int_0^s (E|\tilde{x}_r - \tilde{x}_r|^4)^{\frac{1}{2}} (E|\tilde{x}_{r, x_i}|^4)^{\frac{1}{2}} dr$$

$$\leq C \int_0^s E|\tilde{x}_{r, x_i} - \tilde{x}_{r, x_i}|^2 dr + C \int_0^s (E|\tilde{x}_r - \tilde{x}_r|^4)^{\frac{1}{2}} dr$$

The remaining term $I_{1,4}$ can be estimated in pretty much the same way as $I_{1,2}$ and $I_{1,3}$:

$$|I_{1,4}| \leq C \int_0^s E\left( |x_{j\Delta, x_i} - \tilde{x}_{r, x_i}| |\tilde{x}_{r, x_i} - \tilde{x}_{r, x_i}| \right) dr + C \int_0^s E\left( |x_{j\Delta} - \tilde{x}_r| |\tilde{x}_{r, x_i} - \tilde{x}_{r, x_i}| \right) dr$$

$$\leq C \int_0^s E|\tilde{x}_{r, x_i} - \tilde{x}_{r, x_i}|^2 dr + C \int_0^s E|\tilde{x}_{r, x_i} - \tilde{x}_{r, x_i}|^2 dr + C \int_0^s (E|\tilde{x}_r - \tilde{x}_r|^4)^{\frac{1}{2}} dr$$

$$\leq C \int_0^s E|\tilde{x}_{r, x_i} - \tilde{x}_{r, x_i}|^2 dr + C s \Delta,$$

where the last inequality follows from Lemmas 5.6, 5.7, and 5.9.
We proceed with $I_2$, which can be bound similarly to $I_1$, noting that
\[
\int_0^s \mathbb{E} \left\langle \nabla_y f(x_{j,\Delta}, \hat{y}_t) \theta_{r,x_i} - \mathbb{E} \xi \left( \nabla_y f(x_{j,\Delta}, \xi_{t,x_i}) \xi_{t,x_i} \right) \hat{x}_{r,x_i} - \hat{x}_{r,x_i} \right\rangle dr \\
= \int_0^s \mathbb{E} \left( \nabla_y f(x_{j,\Delta}, \xi_{t,x_i}) \xi_{t,x_i} x_{j,\Delta,x} - \mathbb{E} \xi \left( \nabla_y f(x_{j,\Delta}, \xi_{t,x_i}) \xi_{t,x_i} \right) x_{j,\Delta,x} \right) \hat{x}_{r,x_i} - \hat{x}_{r,x_i} dr \\
+ \int_0^s \mathbb{E} \left( \nabla_y f(x_{j,\Delta}, \xi_{t,x_i}) \xi_{t,x_i} \theta_{r,x_i} \right) \hat{x}_{r,x_i} - \hat{x}_{r,x_i} dr \\
+ \int_0^s \mathbb{E} \left( \nabla_y f(x_{j,\Delta}, \xi_{t,x_i}) \xi_{t,x_i} \right) \hat{x}_{r,x_i} - \hat{x}_{r,x_i} dr \\
= I_{2,1} + I_{2,2} + I_{2,3} + I_{2,4} \quad (5.26)
\]

For $I_3$, Lemma 5.4 Lemma 5.10 and the assumption that $\alpha_1$ is Lipschitz entail
\[
|I_3| \leq \int_0^s \mathbb{E} || \nabla_x \alpha_1(x_r) - \nabla_x \alpha_1(\bar{x}_r) ||^2 dr + 3 \int_0^s \mathbb{E} || \nabla_x \alpha_1(\bar{x}_r) (x_{r,x_i} - \bar{x}_{r,x_i}) ||^2 dr \\
+ 3 \int_0^s \mathbb{E} || \nabla_x \alpha_1(\bar{x}_r) (\hat{x}_{r,x_i} - \bar{x}_{r,x_i}) ||^2 dr \\
\leq C \int_0^s \mathbb{E} (|x_r| - \bar{x}_r|^2 |x_{r,x_i}|)^2 dr + C \int_0^s \mathbb{E} |x_{r,x_i} - \bar{x}_{r,x_i}|^2 dr + C \int_0^s \mathbb{E} |\hat{x}_{r,x_i} - \bar{x}_{r,x_i}|^2 dr \\
\leq C \int_0^s \mathbb{E} |x_{r,x_i} - \bar{x}_{r,x_i}|^2 dr + C \int_0^s (\mathbb{E} |\hat{x}_r - \bar{x}_r|)^{\frac{3}{2}} dr + Cs \Delta \\
\]

Upon combining the bounds for $I_1$, $I_2$ and $I_3$, we conclude that
\[
\mathbb{E} |\hat{x}_{r,x_i} - \bar{x}_{r,x_i}|^2 \leq C(1 + \frac{K}{\Delta}) \int_0^s \mathbb{E} |\hat{x}_{r,x_i} - \bar{x}_{r,x_i}|^2 dr \\
+ C \int_0^s (\mathbb{E} |\hat{x}_r - \bar{x}_r|^\frac{3}{2})^\frac{3}{2} dr + Cs(\Delta + \Delta^\frac{3}{2} + \epsilon) .
\]

Now letting $\Delta = \epsilon^{\frac{3}{2}}$ and using Lemma 5.8 it follows that
\[
\mathbb{E} |\hat{x}_{r,x_i} - \bar{x}_{r,x_i}|^2 \leq C \int_0^s \mathbb{E} |\hat{x}_{r,x_i} - \bar{x}_{r,x_i}|^2 dr + C \epsilon^{\frac{3}{2}}
\]
and applying Gronwall’s Lemma yields the conclusion.

Combining Lemma 5.9 and Lemma 5.10 we have proved:

**Theorem 5.8** Suppose that Assumptions 4.13 hold. Then there exists $C > 0$, independent of $\epsilon$ (possibly depending on $x_0$ and $y_0$), such that
\[
\mathbb{E} |x_{r,x_i} - \bar{x}_{r,x_i}|^2 \leq Cs^{\frac{3}{2}} , \quad s \in [0, T] .
\]
6 Conclusions

Importance sampling is a widely used variance reduction technique for the design of efficient Monte Carlo estimators. A crucial point in order to achieve substantial variance reduction, is a clever (and careful) change of measure. In the diffusion process setting, this change of measure can be realized by adding a control force to the original system, where the optimal control that leads to a zero-variance estimator is related to a Hamilton-Jacobi-Bellman (HJB) equation that may not be easily solvable numerically, e.g. when the state space is high-dimensional.

Our starting point is that, although it may not be possible to compute the optimal control, it is possible to approximate it in such a way that and the resulting estimators remain efficient. Based on the dynamic programming principle of stochastic control, we study approximations to the solution of the HJB equation in three different situations that give rise to efficient importance sampling estimators: time-scale separation between degrees of freedom, small temperature (i.e. low noise), and a combination of the two former. For multiscale diffusions with time-scale separation between the degrees of freedom and exponential type expectations, the asymptotic optimality of the approximation based on a low-dimensional averaged (or homogenized) equation has been proved. Our result implies that it is possible to design efficient importance sampling strategies that are close-to-optimal, in that they yield estimators with vanishing variance and a small relative error, based on low-dimensional averaged equations.

Even though the situations studied in this paper are prototypical for some multiscale system that arise in molecular physics, climate modelling, material science etc., it remains an open question how to treat systems that have multiscale features, but in which the scale separation is not strong and which therefore cannot be treated by asymptotic methods. Another aspect concerns systems that are so large that even a reduced (e.g. averaged or homogenized) equation cannot be discretized by any grid-based method. We leave these questions for future work and refer to [11,20] for some recent algorithmic and methodologic developments in this regard.

Acknowledgements

The authors acknowledge financial support by the DFG Research Center MATHEON, the Einstein Center for Mathematics ECMATH and the DFG Sonderforschungsbereich 1114.

A Two useful inequalities

Claim A.1 Consider a system of linear equations on $t \in [0, T]$ satisfying

$$
\dot{x}_1(t) \leq a_{11} x_1(t) + a_{12} x_2(t) \\
\dot{x}_2(t) \leq \frac{a_{21}}{\epsilon} x_1(t) - \frac{a_{22}}{\epsilon} x_2(t)
$$

with $x_1(0) = 0, x_2(0) = 1$, $a_{ij} > 0$, $1 \leq i, j \leq 2$. Further assume that $x_1(t) \geq 0$ for all $t \in [0, T]$. Then there is constant $C > 0$ depending on $a_{ij}$ and $T$, such that

$$
\max_{0 \leq s \leq T} x_1(s) \leq C \epsilon, \quad x_2(t) \leq e^{-\frac{a_{22} t}{\epsilon}} + C \epsilon, \quad t \in [0, T]. \quad (A.1)
$$
Proof Applying Gronwall’s inequality to the equation of $x_2$, we have

$$x_2(t) \leq e^{-\frac{a_2 t^2}{2}} + \int_0^t e^{-\frac{a_2 (t-s)}{2}} \frac{a_2 t}{\varepsilon} x_1(s) ds \leq e^{-\frac{a_2 t^2}{2}} + \frac{a_2 t}{\varepsilon} \max_{0 \leq s \leq t} x_1(s). \quad (A.2)$$

Applying Gronwall’s inequality to $x_1$ and using (A.2), we find

$$x_1(t) \leq a_1 \int_0^t e^{\alpha_1 (t-s)} \left[ e^{-\frac{a_2 s}{2}} + \frac{a_2}{\varepsilon} \max_{0 \leq r \leq s} x_1(r) \right] ds. \quad (A.3)$$

Since the right hand side in the last inequality is monotonically increasing, it follows that

$$\max_{0 \leq s \leq t} x_1(s) \leq a_1 \int_0^t e^{\alpha_1 (t-s)} \left[ e^{-\frac{a_2 s}{2}} + \frac{a_2}{\varepsilon} \max_{0 \leq r \leq s} x_1(r) \right] ds \leq \frac{a_1}{\alpha_2} e^{\alpha_1 t} + \frac{a_1 a_2}{\varepsilon a_2} \int_0^t e^{\alpha_1 (t-s)} \max_{0 \leq r \leq s} x_1(r) ds. \quad (A.4)$$

The assertion then follows by using Gronwall’s inequality in integral form again, using (A.2)

For $0 < \varepsilon < 1$, we set $t_1 = -\frac{a_3}{\lambda} \max x_1(0) > 0$ and introduce the function $\eta: [0, \bar{T}] \rightarrow [0, 1]$ by

$$\eta(t) = \begin{cases} 1 & 0 \leq t \leq t_1 \\ 0 & t_1 < t \leq \bar{T}. \end{cases} \quad (A.5)$$

Claim A.2 Consider the following system of linear equations on $t \in [0, \bar{T}]$:

$$\dot{x}_1(t) \leq a_1 (1 - \eta(t)) x_1(t) + a_2 \eta(t) x_2(t)$$

$$\dot{x}_2(t) \leq \frac{a_3 x_1(t)}{\lambda} - \frac{\lambda x_2(t)}{\varepsilon},$$

where $\eta$ is given in (A.5), $a_i \geq 0, 1 \leq i \leq 3$, and $x_1(0) = 0, x_2(0) = 1$. Further assume that $x_1(t) \geq 0$ on $t \in [0, \bar{T}]$. Then there is a constant $C > 0$ independent of $\varepsilon$, such that

$$\max_{0 \leq s \leq \bar{T}} x_1(s) \leq C \varepsilon^2, \quad x_2(t) \leq e^{-\frac{a_3}{\lambda} t} + C \varepsilon^2, \quad t \in [0, \bar{T}]. \quad (A.6)$$

Proof As in Claim A.1 we have

$$x_2(t) \leq e^{-\frac{a_3}{\lambda} t} + \frac{a_3}{\lambda} \max_{0 \leq s \leq t} x_1(s). \quad (A.7)$$

$$\max_{0 \leq s \leq t} x_1(s) \leq a_2 \int_0^t e^{a_1 t_1 (1+b+\eta(r))dr} \eta(s) e^{-\frac{a_3}{\lambda} t} + \frac{a_3}{\lambda} \max_{0 \leq r \leq s} x_1(r) ds. \quad (A.8)$$

Then, for $t < t_1$,

$$\max_{0 \leq s \leq t} x_1(s) \leq C \varepsilon^2 + \frac{a_3 a_4}{\lambda} \int_0^t e^{a_1 t_1 (1+b+\eta(r))dr} \eta(s) \max_{0 \leq r \leq s} x_1(r) ds. \quad (A.9)$$

Using (A.7) and Gronwall’s Lemma again, we conclude that

$$\max_{0 \leq s \leq t_1} x_1(s) \leq C \varepsilon^2, \quad x_2(t) \leq e^{-\frac{a_3}{\lambda} t} + C \varepsilon^2, \quad t \leq t_1. \quad (A.10)$$

Repeating the above argument for $t \in [t_1, \bar{T}]$, noticing that $x_1(t_1) \leq C \varepsilon^2, x_2(t_1) \leq C \varepsilon^2, \eta(t) \equiv 0, t \in [t_1, \bar{T}]$, it follows that

$$\max_{t_1 \leq s \leq \bar{T}} x_1(s) \leq C \varepsilon^2, \quad x_2(t) \leq C \varepsilon^2, \quad t \in [t_1, \bar{T}]. \quad (A.11)$$

The proof is completed by combining (A.10) and (A.11).
**B Properties of the stationary process**

For fixed $x \in \mathbb{R}^k$ and $\tau \in \mathbb{R}$, we introduce the process

$$d\xi^\varepsilon_{\tau,s} = \frac{1}{\varepsilon} g(x, \xi^\varepsilon_{\tau,s})ds + \frac{1}{\sqrt{\varepsilon}} \omega_2(x, \xi^\varepsilon_{\tau,s})dw_s, \quad s \geq \tau, \quad \xi^\varepsilon_{\tau,\tau} = y \tag{B.1}$$

where $w_s$ is a standard Wiener process in $\mathbb{R}^l$. In this subsection, we summarize some properties related to the above process that we called the fast subsystem in Section 3. See also [20] for additional results.

**Lemma B.1** Under Assumptions [20] there exists a constant $C > 0$, independent of $\varepsilon, x, y$, such that:

1. $E|\xi^\varepsilon_{\tau},s|^4 \leq e^{-\frac{\lambda(1-C)}{2\varepsilon}}|y|^4 + C(|x|^4 + 1)$.
2. For $\tau_1 \leq \tau_2$, it holds

$$E|\xi^\varepsilon_{\tau_2,s} - \xi^\varepsilon_{\tau_1,s}|^4 \leq C(1 + |x|^4 + |y|^4)e^{-\frac{4(1-C)}{2\varepsilon}}.$$  

3. For $x, x' \in \mathbb{R}^k$ and $\tau_1 \leq \tau_2$,

$$E|\xi^\varepsilon_{\tau_2,s} - \xi^\varepsilon_{\tau_1,s}|^4 \leq e^{-\frac{2(1-C)}{\varepsilon}}(|y|^4 + |x|^4 + 1) + C|x'-x|^4$$

**Proof** 1. By Ito’s formula, we have

$$\frac{dE|\xi^\varepsilon_{\tau},s|^4}{ds} = \frac{1}{\varepsilon} E\left[|\xi^\varepsilon_{\tau},s|^2(4g(x, \xi^\varepsilon_{\tau},s), \xi^\varepsilon_{\tau},s) + 2\|\omega_2(x, \xi^\varepsilon_{\tau},s)\|^2 + 4\|\omega_2(x, \xi^\varepsilon_{\tau},s)\|^2|\xi^\varepsilon_{\tau},s|^2\right]$$

which together with Remark 2 implies that

$$\frac{dE|\xi^\varepsilon_{\tau},s|^4}{ds} \leq -\frac{2\lambda}{\varepsilon} E|\xi^\varepsilon_{\tau},s|^4 + C E\left[|\xi^\varepsilon_{\tau},s|^2(|x|^2 + 1)\right]$$

$$\leq -\frac{\lambda}{\varepsilon} E|\xi^\varepsilon_{\tau},s|^4 + C \left(|x|^4 + 1\right).$$

The first statement then follows from Gronwall’s Lemma.

2. As for the second statement, simply notice that

$$\frac{dE|\xi^\varepsilon_{\tau_2,s} - \xi^\varepsilon_{\tau_1,s}|^4}{ds} \leq -\frac{4\lambda}{\varepsilon} E|\xi^\varepsilon_{\tau_2,s} - \xi^\varepsilon_{\tau_1,s}|^4.$$

Therefore,

$$E|\xi^\varepsilon_{\tau_2,s} - \xi^\varepsilon_{\tau_1,s}|^4 \leq e^{-\frac{4\lambda(1-C)}{2\varepsilon}} E|\xi^\varepsilon_{\tau_1,s} - y|^4 \leq C(1 + |x|^4 + |y|^4)e^{-\frac{4\lambda(1-C)}{2\varepsilon}}.$$  

3. Finally, the third statement is a consequence of Ito’s formula and the Lipschitz property of $g$ and $\omega_2$:

$$\frac{dE|\xi^\varepsilon_{\tau_2,s} - \xi^\varepsilon_{\tau_1,s}|^4}{ds} \leq -\frac{2\lambda}{\varepsilon} E|\xi^\varepsilon_{\tau_2,s} - \xi^\varepsilon_{\tau_1,s}|^4 + C \left(|x' - x|^4\right).$$

Gronwall’s inequality then yields the assertion.

Now consider the derived process

$$d\xi^\varepsilon_{\tau,s,x_i} = \frac{1}{\varepsilon}\left(D_x g(x, \xi^\varepsilon_{\tau,s,x}), \nabla_y g(x, \xi^\varepsilon_{\tau,s,x})\xi^\varepsilon_{\tau,s,x_i}\right)ds + \frac{1}{\sqrt{\varepsilon}}\left(D_x \omega_2(x, \xi^\varepsilon_{\tau,s,x}) + \nabla_y \omega_2(x, \xi^\varepsilon_{\tau,s,x})\xi^\varepsilon_{\tau,s,x_i}\right)dw_s,$$

with $s \geq \tau, \xi^\varepsilon_{\tau,\tau} = 0, 1 \leq i \leq k$. The following is true.

**Lemma B.2** Under Assumptions [20] there exists a constant $C > 0$, independent of $\varepsilon, x, y$, such that

1. For $x \in \mathbb{R}^k, s \geq \tau, E|\xi^\varepsilon_{\tau,s,x_i}|^4 \leq C$.
2. For $\tau_1 \leq \tau_2, x_i \in \mathbb{R}^k$,

$$E|\xi^\varepsilon_{\tau_2,s,x_i} - \xi^\varepsilon_{\tau_1,s,x_i}|^2 \leq C(1 + |x|^2 + |y|^2)e^{-\frac{\lambda(1-C)}{2\varepsilon}}.$$
3. For $\tau_1 \leq \tau_2$, $x, x' \in \mathbb{R}^d$, 
\[
\mathbb{E}(|\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'}|^2) \leq C e^{-\frac{\lambda(\tau_2 - \tau_1)}{\epsilon}} \left[1 + \frac{\lambda - \tau_2}{\epsilon} (1 + |x|^2 + |y|^2)\right] + C|x - x'|^2
\]

Proof 1. Using Ito’s formula, Assumption 1 (Lipschitz continuity) as well as Assumption 2 we see that
\[
\begin{align*}
\frac{d\mathbb{E}(|\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'}|^2)}{d\tau} & \leq \frac{2}{\epsilon} \mathbb{E}\left[(D_x \xi(x, \xi_{\tau_2, x, x_1}^{x'})) \mathbb{E}(|\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'}|^2) \right] + \mathbb{E}\left((D_x \xi(x, \xi_{\tau_2, x, x_1}^{x'})) \mathbb{E}(|\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'}|^2) \right) + \mathbb{E}\left(\mathbb{E}(|\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'}|^2) \right) \\
& \leq -\frac{2}{\epsilon} \mathbb{E}(|\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'}|^2) + C \mathbb{E}(|\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'}|^2)
\end{align*}
\]
and therefore $\mathbb{E}(|\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'}|^2) \leq C$ by Gronwall’s inequality.

2. Now consider $\xi_{\tau_1, x, x_1}^{x'}, \xi_{\tau_2, x, x_1}^{x'}$ with $\tau_1 \leq \tau_2$:
\[
\frac{d\mathbb{E}(|\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'}|^2)}{d\tau} = \frac{2}{\epsilon} \mathbb{E}((D_x \xi(x, \xi_{\tau_2, x, x_1}^{x'})) \mathbb{E}(|\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'}|^2) \right) + \mathbb{E}\left(\mathbb{E}(|\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'}|^2) \right) \\
\leq -\frac{2}{\epsilon} \mathbb{E}(|\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'}|^2) + C \mathbb{E}(|\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'}|^2)
\]
Then Gronwall’s Lemma entails
\[
\mathbb{E}(|\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'}|^2) \leq C(1 + |x|^2 + |y|^2)e^{-\frac{\lambda(\tau_2 - \tau_1)}{\epsilon}}
\]

3. Consider $\xi_{\tau_1, x, x_1}^{x'}, \xi_{\tau_2, x, x_1}^{x'}$ with $\tau_1 \leq \tau_2$. Now,
\[
\frac{d\mathbb{E}(|\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'}|^2)}{d\tau} \leq -\frac{\lambda}{\epsilon} \mathbb{E}(\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'}) (|x' - x|^2 + \mathbb{E}(|\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'})^2 + \mathbb{E}(\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'})) \frac{1}{\epsilon}
\]
and thus
\[
\mathbb{E}(|\xi_{\tau_2, x, x_1}^{x'} - \xi_{\tau_1, x, x_1}^{x'}|^2) \leq Ce^{-\frac{\lambda(\tau_2 - \tau_1)}{\epsilon}} \left[1 + \frac{\lambda - \tau_2}{\epsilon} (1 + |x|^2 + |y|^2)\right] + C|x' - x|^2
\]

The above results allow us to define the stationary process $\xi_x^\epsilon = \xi_x^{x',0}$ with $\xi_x^{x',0} \sim \rho_x(y)$ where $\rho_x$ is the stationary probability density with respect to Lebesgue measure, and also the derived process $\xi_x^{x',i}$ for $1 \leq i \leq k$, satisfying that $\forall f \in C_b^1(\mathbb{R}^d)$ and $f(x) = E(f(x, \xi_x^\epsilon)) = \int f(x) \rho_x(y) dy$, it holds
\[
D_x f(x) = E(D_x f(x, \xi_x^\epsilon)) + \nabla_y f(x, \xi_x^\epsilon) \xi_x^{x',i}_{x, x_1}.
\]

The derived process $\xi_x^{x',i}$ has the following properties:

**Lemma B.3** Under Assumptions 4 and 5 there is constant $C > 0$, independent of $\epsilon$, such that $\forall f \in C_b^1(\mathbb{R}^d)$:
Importance sampling in path space for diffusion processes

1. \[ |E[f(\xi_{0,t}^x)] - \int_{\mathbb{R}} f(y)p_x(y)dy| \leq \sup |f'| \left( |x| + |y| + 1 \right)e^{-\lambda t/2} \] (B.3)

2. \[ \left| E\left(f(\xi_{0,t}^x)\xi_{0,t}^{x',y}ight) - E\left(f(\xi_{0,t}^x)\xi_{0,t}^{x,y}ight) \right| \leq C \left( \sup |f| + \sup |f'| \right) \left( 1 + |x| + |y| \right)e^{-\lambda t/2} \] (B.4)

Proof: We only prove the second inequality, as the first one follows by a similar fashion. Using Lemma [B.1] and Lemma [B.2] we readily conclude that

\[ \left| E\left(f(\xi_{0,t}^x)\xi_{0,t}^{x',y}ight) - E\left(f(\xi_{0,t}^x)^{\xi_{0,t}^{x,y}}\right) \right| \leq C \left( \sup |f| + \sup |f'| \right) \left( 1 + |x| + |y| \right)e^{-\lambda t/2} \]

An analogous property for the stationary process \( \xi_{0,y}^x \) is the following:

Lemma B.4 Under Assumption [1] and [2], there exists constant \( C > 0 \), independent of \( x, x' \), such that

1. For \( x \in \mathbb{R} \), \( E[|\xi_{0,y}^x|] \leq C. \)
2. For \( x, x' \in \mathbb{R} \), \( E[|\xi_{0,y}^x - \xi_{0,y}^{x'}|] \leq C|x - x'|. \)
3. For \( x, x' \in \mathbb{R} \), \( E[|\xi_{0,y}^x - \xi_{0,y}^{x'}|^2] \leq C|x - x'|^2. \)

References

1. A. Aries and N. Krylov, Controlled Diffusion Processes, Stochastic Modelling and Applied Probability, Springer, 2008.
2. S. Asmussen and P. W. Glynn, Stochastic Simulation, Springer, 2007.
3. S. Asmussen and D. P. Kroese, Improved algorithms for rare event simulation with heavy tails, Adv. Appl. Prob., 38 (2006), pp. 545–558.
4. A. Bensoussan, J.-L. Lions, and G. Papanicolaou, Asymptotic analysis for periodic structures, Studies in mathematics and its applications, North-Holland, 1978.
5. J. Blanchet and P. Glynn, Efficient rare-event simulation for the maximum of heavy-tailed random walks, Ann. Appl. Probab., 18 (2008), pp. 1351–1378.
6. S. P. Brooks, Markov chain Monte Carlo method and its application, J. R. Stat. Soc. Series D (The Statistician), 47 (1998), pp. 69–100.
7. A. Budhiraja and P. Dupuis, A variational representation for positive functionals of infinite dimensional Brownian motion, Probab. Math. Statist., 20 (2000), pp. 39–61.
8. S. Cerrai, Second Order PDE’s in Finite and Infinite Dimension: A Probabilistic Approach, no. Nr. 1762 in Lecture Notes in Mathematics, Springer, 2001.
9. G. Da Prato and J. Zabczyk, Ergodicity for Infinite Dimensional Systems, Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology, Cambridge University Press, 1996.
10. Second Order Partial Differential Equations in Hilbert Spaces, London Mathematical Society Lecture Note Series, Cambridge University Press, 2002.
11. P. Dai Pra, L. Meneghini, and W. J. Runggaldier, Connections between stochastic control and dynamic games, Math. Control Signals Systems, 9 (1996), pp. 303–326.
12. A. Doucet, N. De Freitas, and N. Gordon, eds., Sequential Monte Carlo methods in practice, Springer, 2001.
13. S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Hybrid Monte Carlo, Phys. Lett. B, 195 (1987), pp. 216–222.
14. P. Dupuis, K. Spiliopoulos, and H. Wang, Importance sampling for multiscale diffusions, Multiscale Model. Simul., 10 (2012), pp. 1–27.
15. P. Dupuis and H. Wang, Importance sampling, large deviations, and differential games, Stochastics and Stochastic Rep., 76 (2004), pp. 481–508.
16. M. Freidlin and A. Wentzell, *Random Perturbations of Dynamical Systems*, vol. 260 of Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg, 2012.
17. A. Friedman, *Partial differential equations of parabolic type*, Prentice-Hall, 1964.
18. D. Givon, *Strong convergence rate for two-time-scale jump-diffusion stochastic differential systems*, Multi-scale Model. Simul., 6 (2007), pp. 577–594.
19. P. Glasserman, P. Heidelberger, and P. Shahabuddin, *Asymptotically optimal importance sampling and stratification for pricing path-dependent options*, Math. Finance, 9 (1999), pp. 117–152.
20. C. Hartmann, J. Latorre, G. Pavliotis, and W. Zhang, *Optimal control of multiscale systems using reduced-order models*, J. Computational Dynamics, 1 (2014), pp. 279–306.
21. W. K. Hastings, *Monte Carlo sampling methods using markov chains and their applications*, Biometrika, 57 (1970), pp. 97–109.
22. L. O. Hedges, R. L. Jack, J. P. Garrahan, and D. Chandler, *Dynamic order-disorder in atomistic models of structural glass formers*, Science, 323 (2009), pp. 1309–1313.
23. C. Huang and D. Liu, *Strong convergence and speed up of nested stochastic simulation algorithm*, Commun. Comput. Phys., 15 (2014), pp. 1207–1236.
24. R. L. Jack and P. Sollich, *Effective interactions and large deviations in stochastic processes*, arXiv:1501.01154, (2015).
25. I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, Springer, 2 ed., 1991.
26. R. Khasminskii, *Principle of averaging for parabolic and elliptic differential equations and for markov processes with small diffusion*, Theory Probab. Appl., 8 (1963), pp. 1–21.
27. H. J. Kushner and P. G. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer, New York, 2001.
28. J. C. Latorre, C. Hartmann, and C. Schütte, *Free energy computation by controlled Langevin dynamics*, Procedia Comput. Sci., 1 (2010), pp. 1597 – 1606.
29. J. C. Latorre, P. Metzner, C. Hartmann, and C. Schütte, *A structure-preserving numerical discretization of reversible diffusions*, Commun. Math. Sci., 9 (2010), pp. 1051–1072.
30. D. Liu, *Strong convergence of principle of averaging for multiscale stochastic dynamical systems*, Commun. Math. Sci., 8 (2010), pp. 999–1020.
31. G. Pavliotis and A. Stuart, *Multiscale Methods: Averaging and Homogenization*, Springer, 2008.
32. G. A. Pavliotis and A. M. Stuart, *Parameter estimation for multiscale diffusions*, J. Stat. Phys., 127 (2007), pp. 741–781.
33. M. Sarich, R. Banisch, C. Hartmann, and C. Schütte, *Markov state models for rare events in molecular dynamics*, Entropy, 16 (2014), pp. 258–286.
34. C. Schütte, A. Fischer, W. Huisinga, and P. Deufhard, *A direct approach to conformational dynamics based on Hybrid Monte Carlo*, J. Comput. Phys., 151 (1999), pp. 146 – 168.
35. K. Spiliopoulos, *Large deviations and importance sampling for systems of slow-fast motion*, Appl. Math. Optim., 67 (2013), pp. 123–161.
36. K. Spiliopoulos, *Rare event simulation for multiscale diffusions in random environments*, arXiv:1410.0386, (2014).
37. K. Spiliopoulos, P. Dupuis, and X. Zhou, *Escaping from an attractor: Importance sampling and rest points, part I*, Ann. Appl. Probab., 1303.0450v1 (2014).
38. E. Vanden-Eijnden and J. Weare, *Rare event simulation of small noise diffusions*, Comm. Pure Appl. Math., 65 (2012), pp. 1770–1803.
39. W. Zhang, H. Wang, C. Hartmann, M. Weber, and C. Schütte, *Applications of the cross-entropy method to importance sampling and optimal control of diffusions*, J. Sci. Comput., 36 (2014), pp. A2654–A2672.