Noname manuscript No.
(will be inserted by the editor)

Importance sampling in path space for diffusion processes with slow-fast variables

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 the 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 focus on certain diffusions with both slow and fast variables, and the main result is that we obtain an upper bound of the relative error for the importance sampling estimators with control obtained from the limiting dynamics. We demonstrate our approximation strategy with a simple numerical example.

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 [33]. Despite their quite long history since the invention of the computer, the development of MC method and applications thereof are a field of active research. Variants of the standard Monte Carlo method include Metropolis MC [24,7], Hybrid MC [13,39], Sequential MC [34,12], to mention just a few.

A key issue for many MC methods is variance reduction in order to improve the convergence of the corresponding MC estimators. Although all unbiased MC estimators share the same $\mathcal{O}(N^{-\frac{1}{2}})$ 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., [1,33]) seek to decrease the constant prefactor and thus to increase the accuracy and efficiency of the estimators.

C. Hartmann, W. Zhang
Institute of Mathematics, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany
E-mail: carsten.hartmann@fu-berlin.de, wei.zhang@fu-berlin.de

C. Schütte, M. Weber
Zuse Institute Berlin, Takustrasse 7, 14195 Berlin, Germany
E-mail: schuette@zib.de, weber@zib.de

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 $(\Omega, \mathcal{F}, \mathbf{P})$ and the calculation of a probability

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

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

$$\mathbf{E}(\chi_B(X)) = \mathbf{E}_{\mathbf{Q}}(\chi_B(X)\Psi) \ ,$$

with $\Psi(\omega) = (\mathrm{d}\mathbf{P}/\mathrm{d}\mathbf{Q})(\omega)$ being the Radon-Nikodym derivative of $\mathbf{P}$ with respect to $\mathbf{Q}$. The difficulty now lies in a clever choice of $\mathbf{Q}$, because not every probability measure $\mathbf{Q}$ 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 Ψ (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 $\mathbf{P}$. Importance sampling thus deals with clever choices of $\mathbf{Q}$ that enhance the sampling of events like $\{X \in B\}$ while mimicking the behaviour of the original distribution in the relevant regions. Often such a choice can be based on large deviation asymptotics that provides estimates for the probability of the event $\{X \in B\}$ as a function of a smallness parameter; see, e.g., [5, 22, 2, 16, 15, 44].

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 path functional of X_t with respect to the underlying probability measure $\mathbf{P}$ 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 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 [36]. We confine ourselves to certain exponential path functionals which will be explicitly given below. For this type of path functionals, the 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.

While in general it is impractical to find the exact optimal control force by solving an optimal control problem, there is some hope to find computable approximations to the optimal control that yield importance sampling estimators which are sufficiently accurate in that they have small variance. A general theoretical framework has been established by Dupuis and Wang in [17, 16], where they connected the subsolutions of HJB equation and the rate of variance decay for the corresponding importance sampling estimators. This theoretical framework has been further applied by Dupuis, Spiliopoulos and Wang in a series of papers [14,15,40,42] to study systems of quite general forms and several adaptive importance sampling schemes were suggested based on large deviation analysis. In many cases, these importance sampling schemes were shown to be asymptotically optimal in logarithmic sense. Also see discussions in [44,41]. More closely related to our present work, dynamics involving two parameters δ, ϵ and with slow-fast variables were studied in [40]. The author there performed a systematic analysis for dynamics within different regimes according to the asymptotics of ratio $\frac{\epsilon}{\delta}$ as $\epsilon \to 0$, where $\delta = \delta(\epsilon)$. Importance sampling for systems in the regime when $\frac{\epsilon}{\delta} \to +\infty$ with random environment was studied in [42]. Also, in [44] the authors proposed a numerical way to compute control which leads to importance sampling estimator with vanishing relative error for diffusion processes in the small noise limit. On the other hand, while it is crucial to study importance sampling in the small noise limit when $\epsilon \to 0$, some recent work [43,41] considered the performance of importance sampling estimators when ϵ is small but fixed (pre-asymptotic), especially when systems' metastability is involved [43].

Inspired by these previous studies, in the present work we consider importance sampling problem for diffusions with two different time scales. See dynamics (3.1) in Section 3. Instead of studying importance sampling estimators associated with general subsolutions of the HJB equation as in [16,14,15,40,42], we consider a specific control which can be constructed from the low-dimensional limiting dynamics. The main contribution of the present work is Theorem 3.1 in Section 3. It states that, under certain assumptions, the importance sampling estimator associated to this specific control is asymptotically optimal in the time scale separation limit and an upper bound on the relative error of the corresponding estimator is obtained. To the best of our knowledge, this is the first result where the dependence of the relative error of the importance sampling estimator on the time-scale separation parameter is explicitly given. As a secondary contribution, since the proof is based on a careful study of the multiscale process and the limiting process, several convergence results related to the original process and the limiting process are obtained as a by-product. See Theorem 5.2-5.4 in Section 5.

Before concluding the introduction, we compare our results with the previous work in more details and discuss some limitations. First of all, the dynamics (3.1) considered in the present work is less general than the dynamics considered in [40,42]. Specifically, dynamics (3.1) is a special case of [40,42] corresponding to coefficients $b = g = \tau_1 = 0$ there. Secondly, instead of considering asymptotic regime for both $\epsilon, \delta \to 0$ as in [15,40,42], here we only consider the time-scale separation limit and assume the other parameter β in (3.1), which is related to system's temperature, is fixed (although could be large). Roughly speaking, this is equivalent to the case when $\delta \to 0$ with fixed ϵ in [40,42]. Accordingly, the constant in Theorem 3.1 also depends on β. Thirdly, we assume Lipschitz conditions on system's coefficients, which may be restrictive in many applications. Generalizing the theoretical results to non-Lipschitz case is possible but not trivial and will be considered in future work. See [9] for related studies on reaction-diffusion equations.

Nevertheless, dynamics (3.1) is an interesting mathematical model which exhibits both slow and fast time scales and belongs to the "averaging case" in the literatures [3,37] and our results are of different type comparing to the above mentioned literatures. In applications, especially in climate sciences and molecular dynamics [4,35,38], systems may have a few degrees of freedom which evolves on a large time scale and exhibits *metastability* feature, while the other degrees of freedom are rapidly evolving. In this situation, due to the existence of systems' metastability, standard Monte Carlo sampling may become inefficient with a large variance even for moderate temperature β (also see [43]). We expect our results will be instructive for studying importance sampling in this situation. Also see Section 4 for more discussions on a simple illustrative numerical example.

Organization of the article. This paper is organized as follows. In Section 2, we briefly introduce the importance sampling method in the diffusion setting and discuss the variance of Monte Carlo estimators corresponding to a general control force. Section 3 states the assumptions and our main result: an upper bound of the relative error for the importance sampling estimator based on suboptimal controls for the multiscale diffusions; the result is proved in Section 5, but we provide some heuristic arguments based on formal asymptotic expansions already in Section 3. Section 4 shows a simple numerical example that demonstrate the performance of the importance sampling method. Appendix A and B contain technical results that are used in the proof.

2 Importance sampling of diffusions

We consider the conditional expectation

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

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

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

with $b : \mathbb{R}^n \to \mathbb{R}^n$, $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times m}$, w_s is a standard m-dimensional Wiener process. An expectation similar to (2.1) may arise either as an object to study importance sampling method [15,40,42,44], or due to its connection to certain optimal control problem [6,18]. In recent years, it has also been exploited by physicists to study phase transitions [27,25].

In the following of this section, we will introduce the importance sampling method to compute quantify (2.1). To simplify matters, we assume all the coefficients are smooth and the controls satisfy the Novikov condition such that the Girsanov theorem can be applied [36]. Specific assumptions and the concrete form of dynamics will be given in Section 3.

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

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

where $u_s \in \mathbb{R}^m$ will be referred to as the *control force*. Then it follows from Girsanov theorem [36] that $\bar{w}_s$ is a standard m-dimensional Wiener process under probability measure $\bar{\mathbf{P}}$, where the Radon-Nikodym derivative is

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

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

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

with variance

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

Moreover, under $\bar{\mathbf{P}}$, we have

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

Now consider the calculation of (2.5) by a Monte Carlo sampling in path space, and suppose that N independent trajectories $\{z^{u,i}_s, t \le s \le T\}$ of (2.7) have been generated where $i = 1, 2, \cdots, N$. An unbiased estimator of (2.1) is now given by

$$I_N = \frac{1}{N}\sum_{i=1}^N \Big[\exp\Big(-\beta\int_t^T h(z^{u,i}_s)\,ds\Big)(Z^{u,i}_t)^{-1}\Big], \tag{2.8}$$

whose variance is

$$\mathrm{Var}_u I_N = \frac{\mathrm{Var}_u I}{N} = \frac{1}{N}\Big[\bar{\mathbf{E}}\Big(\exp\Big(-2\beta\int_t^T h(z^u_s)\,ds\Big)(Z_t)^{-2}\Big) - I^2\Big]. \tag{2.9}$$

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 error* [16,44]

$$\mathrm{RE}_u(I) = \frac{\sqrt{\mathrm{Var}_u I}}{I}\,. \tag{2.10}$$

The advantage of introducing the control force u_s is that we may choose u_s to reduce the relative error of the estimator (2.8). From (2.6) and (2.9), we can see that minimizing the relative error of the new estimator is equivalent to choosing u_s such that

$$\frac{1}{I^2}\bar{\mathbf{E}}\Big[\exp\Big(-2\beta\int_t^T h(z^u_s)\,ds\Big)(Z_t)^{-2}\Big] \tag{2.11}$$

is as close as possible to 1.

2.1 Dual optimal control problem and estimate of relative error

To proceed, we make use of the following duality relation [6]:

$$\log\mathbf{E}\Big[\exp\Big(-\beta\int_t^T h(z_s)\,ds\Big)\Big] = -\beta\inf_{u_s}\bar{\mathbf{E}}\Big\{\int_t^T h(z^u_s)\,ds + \frac{1}{2}\int_t^T |u_s|^2 ds\Big\}\,, \tag{2.12}$$

where the infimum is over all processes u_s which are progressively measurable with respect to the augmented filtration generated by the Brownian motion. See [6] for more discussions. It is known that there is a feedback control $\hat{u}_s$ such that the infimum on the right-hand side (RHS) of

(2.12) is attained (see Theorem 3.1 in [18]). We will call $\hat{u}_s$ the *optimal control force*. Accordingly we define $\hat{w}_s, \hat{Z}_t, \hat{\mathbf{P}}$ to be the respective quantities in (2.3) and (2.4) with u_s replaced by $\hat{u}_s$, and we denote $\hat{z}_s = \hat{z}_s^{\hat{u}}$ the solution of (2.7) with control force $\hat{u}_s$. Using Jensen's inequality one can show that (2.12) implies

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

Combining the above equality 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:

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

From dynamic programming principle [18], we know $U(t,z)$ satisfies the following *Hamilton-Jacobi-Bellman* or *dynamic programming* equation:

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

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

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

Now we estimate (2.11) and thus the relative error (2.10) for a general control u_s. To this end we suppose that the probability measures $\bar{\mathbf{P}}$ and $\hat{\mathbf{P}}$ are mutually equivalent. Then, using (2.13), we can conclude that

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

and therefore

$$\begin{aligned} \frac{1}{I^2}\bar{\mathbf{E}}\Big[\exp\Big(-2\beta\int_t^T h(z_s^u)ds\Big)(Z_t)^{-2}\Big] =& \frac{1}{I^2}\bar{\mathbf{E}}\Big[\exp\Big(-2\beta\int_t^T h(\hat{z}_s)ds\Big)(\hat{Z}_t)^{-2}\Big(\frac{\hat{Z}_t}{Z_t}\Big)^2\Big] \\ =& \bar{\mathbf{E}}\Big[\Big(\frac{\hat{Z}_t}{Z_t}\Big)^2\Big], \end{aligned} \tag{2.18}$$

where by Girsanov's formula (2.4), we have

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

In order to simplify (2.18), we follow [15] and introduce another control force $\tilde{\bar{u}}_s$ and change the measure again. Specifically, we choose $\tilde{\bar{u}}_s = 2\hat{u}_s - u_s$ and define $\tilde{\bar{w}}_t, \tilde{\bar{\mathbf{P}}}, \tilde{\bar{Z}}_t$ as in (2.3)–(2.4), with u_s being replaced by $\tilde{\bar{u}}_s$. If we now let $\tilde{\bar{\mathbf{E}}}$ denote the expectation with respect to $\tilde{\bar{\mathbf{P}}}$ then, using equations (2.18) and (2.19), we obtain

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

Roughly speaking, the last equation indicates that the relative error (2.10) of the importance sampling estimator associated to a general control u depends on the difference between control u and the optimal control $\hat{u}$. This relation will be further used in Section 5 to prove the upper bound for the relative error of importance sampling estimator.

3 Importance sampling of multiscale diffusions

Our main result in this paper concerns dynamics with two time scales. 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 (2.2) is of the form

$$\begin{aligned} dx_s &= f(x_s, y_s)ds + \beta^{-1/2}\alpha_1(x_s, y_s)dw_s^1 \\ dy_s &= \frac{1}{\epsilon}g(x_s, y_s)ds + \beta^{-1/2}\frac{1}{\sqrt{\epsilon}}\alpha_2(x_s, y_s)dw_s^2 \end{aligned} \tag{3.1}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}^k$, $g\colon \mathbb{R}^n \to \mathbb{R}^l$ are smooth vector fields, $\alpha^1\colon \mathbb{R}^n \to \mathbb{R}^{k\times m_1}$, $\alpha^2\colon \mathbb{R}^n \to \mathbb{R}^{l\times m_2}$ are smooth noise coefficients and $w_s^1 \in \mathbb{R}^{m_1}$, $w_s^2 \in \mathbb{R}^{m_2}$ are independent Wiener processes with $m_1, m_2 > 0$. The parameter $\epsilon \ll 1$ describes the time-scale separation between processes x_s and y_s.

Let $x \in \mathbb{R}^k$ be given and suppose that the fast subsystem

$$dy_s = \frac{1}{\epsilon}g(x, y_s)ds + \beta^{-1/2}\frac{1}{\sqrt{\epsilon}}\alpha_2(x, y_s)dw_s^2, \qquad y_t = y \in \mathbb{R}^l\,, \tag{3.2}$$

is ergodic with unique invariant measure whose density is $\rho_x(y)$ with respect to Lebesgue measure (see Appendix B for more details). Then it is well known that when $\epsilon \to 0$, under some mild conditions on the coefficients, the slow component of (3.1) converges in probability to the averaged dynamics [19,29,37,32]

$$\begin{aligned} d\widetilde{x}_s &= \widetilde{f}(\widetilde{x}_s)ds + \beta^{-1/2}\widetilde{\alpha}(\widetilde{x}_s)dw_s, \quad t \le s \le T \\ \widetilde{x}_t &= x\,, \end{aligned} \tag{3.3}$$

where for every $x \in \mathbb{R}^k$, we have

$$\widetilde{f}(x) = \int_{\mathbb{R}^l} f(x, y)\rho_x(y)\,dy, \qquad \widetilde{\alpha}(x)\widetilde{\alpha}(x)^T = \int_{\mathbb{R}^l} \alpha_1(x, y)\alpha_1(x, y)^T\rho_x(y)\,dy. \tag{3.4}$$

Further define

$$\widetilde{h}(x) = \int_{\mathbb{R}^l} h(x, y)\rho_x(y)\,dy \tag{3.5}$$

and consider the averaged value function

$$U_0(t, x) = \inf_u \bar{\mathbf{E}}\Big\{\int_t^T \widetilde{h}(\widetilde{x}_s^u)\,ds + \frac{1}{2}\int_t^T |u_s|^2 ds\Big\}, \tag{3.6}$$

where $\widetilde{x}_s^u \in \mathbb{R}^k$ is the solution of

$$\begin{aligned} d\widetilde{x}_s^u &= \widetilde{f}(\widetilde{x}_s^u)ds - \widetilde{\alpha}(\widetilde{x}_s^u)u_s ds + \beta^{-1/2}\widetilde{\alpha}(\widetilde{x}_s^u)dw_s, \quad t \le s \le T \\ \widetilde{x}_t^u &= x\,. \end{aligned} \tag{3.7}$$

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

$$\hat{u}_s^0 = \left(\alpha_1^T(x_s^u, y_s^u)\nabla_x U_0(x_s^u), 0\right) , \tag{3.8}$$

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 averaged value function U_0 in (3.6) and leave the fast variable uncontrolled. Notice that control (3.8) has also been suggested in [40] for more general dynamics with a general subsolution of the HJB equation.

Remark 1 Another variant of a suboptimal control would be

$$\hat{u}_s^0 = \left(\widetilde{\alpha}^T(x_s^u)\nabla_x U_0(x_s^u), 0\right) , \tag{3.9}$$

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 in order to control the system. In the following we will assume that α_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.

Assumption 1 *$f, g, h, \alpha_1, \alpha_2$ are C^2 functions, with derivatives that are uniformly bounded by a constant $C > 0$. α_1, α_2 and h are bounded. Furthermore, there exist constants $C_1 > 0$, such that*

$$\zeta^T \alpha_2(x, y)\alpha_2(x, y)^T \zeta \geq C_1|\zeta|^2 ,$$

$x \in \mathbb{R}^k, \zeta, y \in \mathbb{R}^l$.

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

$$\langle g(x, y_1) - g(x, y_2), y_1 - y_2\rangle + \frac{3}{\beta}\|\alpha_2(x, y_1) - \alpha_2(x, y_2)\|^2 \leq -\lambda|y_1 - y_2|^2, \tag{3.10}$$

where $\|\cdot\|$ denotes the Frobenius norm.

Assumption 3 *α_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 $\widetilde{f}$ as given by (3.4), Lemma B.4 in Appendix B implies that $\widetilde{f}$ is Lipschitz continuous. Unlike [32], we do not assume that f is bounded.
3. Assumption 2 guarantees that the fast dynamics are quickly mixing. As we study the asymptotic solution of (3.1) as $\epsilon \to 0$ at fixed noise intensity, the inverse temperature β can be absorbed into the coefficients α_1, α_2 and h. In Section 5, we will therefore assume $\beta = 1$, in which case Assumption 2 implies that

$$\langle \nabla_y g\, \xi, \xi\rangle + 3\|\nabla_y \alpha_2\, \xi\|^2 \leq -\lambda|\xi|^2, \quad \forall y, \xi \in \mathbb{R}^l, x \in \mathbb{R}^k , \tag{3.11}$$

where $\nabla_y \alpha_2 \xi$ is an $l \times m_2$ matrix with components

$$\left(\nabla_y \alpha_2 \xi\right)_{ij} = \sum_{r=1}^{l} \frac{\partial (\alpha_2)_{ij}}{\partial y_r} \xi_r \,, \qquad 1 \le i \le l \,, \quad 1 \le j \le m_2 \,. \tag{3.12}$$

Combining this with Assumption 1, we have

$$\begin{aligned}
&\langle g(x,y), y\rangle + \frac{3}{2}\|\alpha_2(x,y)\|^2 \\
\le&\langle g(x,y) - g(x,0), y\rangle + \langle g(x,0), y\rangle + 3\|\alpha_2(x,y) - \alpha_2(x,0)\|^2 + 3\|\alpha_2(x,0)\|^2 \\
\le& -\frac{\lambda}{2}|y|^2 + C(|x|^2+1) \qquad \forall x \in \mathbb{R}^k, y \in \mathbb{R}^l \,.
\end{aligned} \tag{3.13}$$

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 *Suppose Assumptions 1–3 hold, and consider the importance sampling method for computing (2.1) with 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*

$$RE_{\hat{u}^0}(I) \le C\epsilon^{\frac{1}{8}} \,,$$

where constant $C > 0$ is independent of ϵ.

3.2 Formal expansion by asymptotic analysis

The proof of Theorem 3.1 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 [3,37]. To this end, let U^ϵ denote the solution of (2.15), for which we seek an asymptotic expansion in powers of ϵ. Further let $\phi^\epsilon(t,x,y) = \exp(-\beta U^\epsilon)$. From the dual relation (2.12), we know that ϕ^ϵ is the expectation (2.1) we want to compute. By the Feynman-Kac formula, we have

$$\begin{aligned}
&\frac{\partial \phi^\epsilon}{\partial t} + \mathcal{L}\phi^\epsilon - \beta h \phi^\epsilon = 0 \,, \quad 0 \le t \le T \\
&\phi^\epsilon(T,x,y) = 1
\end{aligned} \tag{3.14}$$

where $\mathcal{L} = \epsilon^{-1}\mathcal{L}_0 + \mathcal{L}_1$ is the infinitesimal generator of process (3.1), with

$$\begin{aligned}
\mathcal{L}_0 &= g \cdot \nabla_y + \frac{1}{2\beta}\alpha_2\alpha_2^T : \nabla_y^2 \\
\mathcal{L}_1 &= f \cdot \nabla_x + \frac{1}{2\beta}\alpha_1\alpha_1^T : \nabla_x^2 \,.
\end{aligned} \tag{3.15}$$

Now consider the expansion $\phi^\epsilon = \phi_0 + \epsilon\phi_1 + \ldots$ of ϕ^ϵ in powers of ϵ. Plugging it into (3.14) and comparing different powers of ϵ, we obtain to lowest order:

$$\frac{\partial \phi_0}{\partial t} + \mathcal{L}_0\phi_1 + \mathcal{L}_1\phi_0 - \beta h \phi_0 = 0, \tag{3.16}$$

$$\mathcal{L}_0\phi_0 = 0, \tag{3.17}$$

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 $\mathcal{L}_0^* \rho_x = 0$ with $\int_{\mathbb{R}^l} \rho_x(y) dy = 1$. Here $\mathcal{L}_0^*$ is the adjoint operator of $\mathcal{L}_0$ with respect to the standard scalar product in the space $L^2(\mathbb{R}^l)$. Hence we can conclude from (3.17) that $\phi_0 = \phi_0(t, x)$ is independent of y. Integrating both sides of (3.16) against $\rho_x(y)$, we obtain a closed equation for ϕ_0:

$$\frac{\partial \phi_0}{\partial t} + \widetilde{\mathcal{L}} \phi_0 - \beta \widetilde{h} \phi_0 = 0 \tag{3.18}$$

with

$$\widetilde{\mathcal{L}} = \widetilde{f}(x) \cdot \nabla_x + \frac{\widetilde{\alpha}(x) \widetilde{\alpha}(x)^T}{2\beta} : \nabla_x^2 \,, \tag{3.19}$$

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

Notice that $\widetilde{\mathcal{L}}$ is the infinitesimal generator of the averaged dynamics (3.3). Again by the Feynman-Kac formula, the solution to (3.18) is recognized as the conditional expectation

$$\phi_0(t, x) = \mathbf{E}\Big[\exp\Big(-\beta \int_t^T \widetilde{h}(\widetilde{x}_s) \, ds \Big) \;\Big|\; x_t = x \Big] \tag{3.20}$$

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

$$U^\epsilon = -\beta^{-1} \log(\phi_0 + \epsilon \phi_1 + o(\epsilon)) = -\beta^{-1} \log \phi_0 - \beta^{-1} \frac{\phi_1}{\phi_0} \epsilon + o(\epsilon). \tag{3.21}$$

Combining (3.21) with (3.20) and the dual relation (2.12), we conclude that U_0 in (3.6) satisfies $U_0 = -\beta^{-1} \log \phi_0$ and is the leading term of U^ϵ in expansion (3.21). Finding the corresponding expression for the optimal control is now straightforward: Setting $\hat{u}_s = (\hat{u}_{s,1}, \hat{u}_{s,2})$, the relation (2.16) between the optimal feedback control and the value function yields

$$\begin{aligned} \hat{u}_{s,1} &= \alpha_1^T \nabla_x U_0 + \mathcal{O}(\epsilon) = -\beta^{-1} \frac{\alpha_1^T \nabla_x \phi_0}{\phi_0} + \mathcal{O}(\epsilon), \\ \hat{u}_{s,2} &= \frac{\alpha_2^T}{\sqrt{\epsilon}} \nabla_y U^\epsilon = \mathcal{O}(\epsilon^{\frac{1}{2}}) \,, \end{aligned} \tag{3.22}$$

where all functions are evaluated at $(s, x_s^{\hat{u}}, y_s^{\hat{u}})$.

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 derivations suggest that one possible strategy for finding good control forces for importance sampling is to first compute U_0 from (3.6) or (3.20), 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 dynamics (3.1) is the homogenization problems where systems exhibit more than two time scales [37]. 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. See [15, 40, 42] for large deviations and importance sampling studies of related dynamics.

4 Numerical example

In this section, we study a numerical example and discuss some algorithmic issues for solving suboptimal control force (3.8) proposed in Section 3. The dynamics we considered here is described by the two-dimensional SDE

$$
\begin{aligned}
dx_s &= -\frac{\partial V(x_s, y_s)}{\partial x} ds + \beta^{-1/2} dw_s^1 \\
dy_s &= -\frac{1}{\epsilon}\frac{\partial V(x_s, y_s)}{\partial y} ds + \beta^{-1/2}\frac{1}{\sqrt{\epsilon}} dw_s^2 \,,
\end{aligned}
\tag{4.1}
$$

where $(x_s, y_s) \in \mathbb{R}^2$, $w_s = (w_s^1, w_s^2)$ is a two-dimensional Wiener process, $\beta, \epsilon > 0$ and potential $V(x, y) = V_1(x) + V_2(x, y)$ with

$$
\begin{aligned}
V_1(x) =& \frac{1}{2}\big(1 - \eta(x) - \eta(-x)\big) \cos\Big(\frac{4\pi x}{5}\Big) + 3\eta(x)(x-1)^2 + 3\eta(-x)(x+1)^2, \\
V_2(x, y) =& \frac{1}{2}(x-y)^2 \,.
\end{aligned}
\tag{4.2}
$$

In the above, function $\eta(x) = e^{-\frac{1}{x}}$ if $x > 0$, and $\eta(x) = 0$ elsewise. Potential $V_1(x)$ is a smooth function and contains two "wells" centered around $x = -1$ and $x = 1$. As in (2.1), we aim at computing the expectation

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

where

$$
h(x) = \eta\Big(\frac{x+2}{w}\Big)\eta\Big(\frac{4-x}{w}\Big)(x-1)^2 + 10\Big[2 - \eta\Big(\frac{x+2}{w}\Big) - \eta\Big(\frac{4-x}{w}\Big)\Big] \,, \tag{4.4}
$$

with parameter $w = 0.02$. The profiles of functions η, V_1 and h are shown in Figure 1. Notice that function η is introduced in (4.2) and (4.4) in order that Assumption 1-3 of Theorem 3.1 in Section 3 are satisfied. More discussions on these assumptions can be found in the section of Introduction and Conclusions.

For dynamics (4.1), using the specific form of potential V, we can obtain that the invariant measure of the fast dynamics y_s for each fixed $x \in \mathbb{R}$ has the density

$$
\rho_x(y) \propto e^{-\beta(x-y)^2} \tag{4.5}
$$

with respect to the Lebesgue measure. Also following the discussions in Section 3, especially (3.3) and (3.4), we know the averaged dynamics is a one-dimensional diffusion in a double well potential :

$$
d\widetilde{x}_s = -V_1'(\widetilde{x}_s) ds + \beta^{-1/2} dw_s \,, \tag{4.6}
$$

where potential V_1 is given in (4.2) and w_s is a one-dimensional Wiener process.

Before we proceed, it is helpful to briefly illustrate the difficulty to compute (4.3) with the standard Monte Carlo method, which is mainly due to system's metastability when β is moderate or relatively large. On one hand, in the path space, the exponential integrand in (4.3) is peaked around trajectories which spend a large portion of time at the minimum of h, which is located around $x = 1$ (Figure 1(c)). On the other hand, in order to get close to state $x = 1$, trajectories starting from $x_0 = -1$ need to cross the energy barrier $\Delta V_1(\approx V_1(0) - V_1(-1))$ of V_1 (Figure 1(b)). The probability of these barrier-crossing trajectories is roughly of order $\exp(-\beta\Delta V_1)$ when $\beta\Delta V_1$ is large. Combining these facts, we can conclude that the rare barrier crossing events play an important role when computing (4.3). And standard Monte Carlo method will be inefficient in such a situation due to insufficient sampling of these rare events (cf. the discussion in Section 1).

Computation of the suboptimal estimator based on the averaged equation. Now let us consider the method outlined in Subsection 3.1. Recalling (3.18), the averaged conditional expectation ϕ_0 solves the linear backward evolution equation

$$\begin{aligned} &\frac{\partial \phi_0}{\partial t} + \widetilde{\mathcal{L}}\phi_0 - \beta\widetilde{h}\phi_0 = 0 \\ &\phi_0(T,x) = 1, \end{aligned} \tag{4.7}$$

with

$$\widetilde{\mathcal{L}} = -V_1'\frac{\partial}{\partial x} + \frac{1}{2\beta}\frac{\partial^2}{\partial^2 x}, \quad \widetilde{h}(x) = h(x)\,. \tag{4.8}$$

The equation for ϕ_0 is one-dimensional (in space), and can be solved by standard method: For instance, using Rothe's method, we can first discretize (4.7) in time, which yields

$$\Big(\frac{1}{\Delta t} - \widetilde{\mathcal{L}}\Big)\phi_0^j = \Big(\frac{1}{\Delta t} - \beta h\Big)\phi_0^{j+1}\,, \quad j = 0, 1, \cdots, m-1 \tag{4.9}$$

where ϕ_0^j denotes the approximation of ϕ_0 at time $t_j = j\Delta t$, $j = 0, 1, \cdots, m$ with time step size $\Delta t = T/m$. Equation (4.9) is then further discretized in space using the structure-preserving finite volume method described in [31]. Starting from $\phi_0^m \equiv 1$, we can obtain all ϕ_0^j for $j = m-1, m-2, \cdots, 1$ by solving (4.9) backwardly.

After obtaining ϕ_0, we can compute the feedback control force (3.8) as

$$\hat{u}_s^0 = \left(-\beta^{-1}\frac{\partial_x\phi_0(s, x_s^u)}{\phi_0(s, x_s^u)}, 0\right), \tag{4.10}$$

when system's state is at (x_s^u, y_s^u) at time s. Plugging the last expression into (4.1) then yields the controlled dynamics (also see (2.7))

$$\begin{aligned} dx_s^u &= -\frac{\partial V(x_s^u, y_s^u)}{\partial x}ds + \beta^{-1}\frac{\partial_x\phi_0(s, x_s^u)}{\phi_0(s, x_s^u)}ds + \beta^{-1/2}dw_s^1 \\ dy_s^u &= -\frac{1}{\epsilon}\frac{\partial V(x_s^u, y_s^u)}{\partial y}ds + \beta^{-1/2}\frac{1}{\sqrt{\epsilon}}dw_s^2, \end{aligned} \tag{4.11}$$

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

Numerical results. Now we turn to the numerical results. Table 1 shows the numerical results of the Monte Carlo method with the above importance sampling strategy, i.e. (4.11), which should be compared to Table 2 that shows the result of standard Monte Carlo method. For both 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 discretization bias. The control (4.10) was obtained by computing ϕ_0 from (4.9) on a grid of size n_x. For comparison, we have computed a reference importance sampling Monte-Carlo solution ("exact"

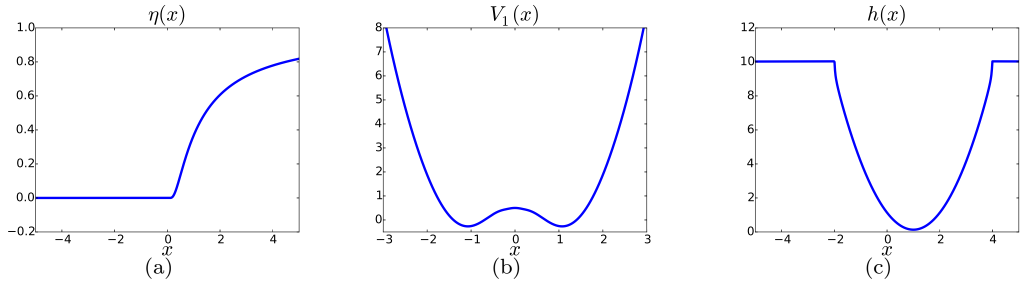

Fig. 1: (a) Function $\eta(x)$ used to define potential V_1. (b) Double well potential $V_1(x)$. (c) Function h in (4.3).

mean value) based on $N = 10^5$ independent realizations that is displayed in Table 1 in the column with label "I". The performance of the Monte Carlo methods can be evaluated based on the variance (2.6) and the relative error (2.10). In our numerical study, they are estimated from the sampled trajectories as

$$\begin{aligned} \mathrm{Var}_u I =& \frac{1}{N}\sum_{i=1}^{N}\Big[\Big(\exp\Big(-\beta\int_0^T h(x_s^{u,i})\,ds\Big)(Z_t^{u,i})^{-1}\Big) - I_N\Big]^2, \\ \mathrm{RE}_u(I) =& \frac{\sqrt{\mathrm{Var}_u I}}{I_N}, \end{aligned} \tag{4.12}$$

where $x_s^{u,i}$ is the i-th trajectories, $1 \le i \le N$, I_N is the estimator (2.8) of I, and u denotes the control force. See Section 2 for details. Furthermore, in order to illustrate the actual effect of the control force, we monitor the barrier crossing events with $x_s \ge 0$ for some $0 < s \le T = 1$ and let R_c record the ratio of trajectories which cross the barrier among all the trajectories.

In Table 1, for different values of β, we can see that the relative error of the importance sampling estimator becomes smaller as ϵ decreases from 0.1 to 0.001. This indicates that the importance sampling estimator performs better and better when ϵ deceases and therefore is accordance with the conclusion of Theorem 3.1 in Section 3.

It is also worth making a comparison of both the importance sampling estimator and the standard Monte Carlo estimator. For the importance sampling estimator (Table 1), we observed that both the mean values and the variances, estimated with $N = 10^4$ trajectories, are stable after we ran several times and are close to the results estimated with $N = 10^5$ trajectories, which we take as the "exact" mean value. For the standard Monte Carlo method (Table 2), at $\beta = 1$, while it gives acceptable mean values, the sample variances (and the relative errors) are larger comparing to the importance sampling estimator. For $\beta = 5$, 8, however, the results with standard Monte Carlo method become far away from the "exact" mean values and are unstable when we ran several times. These results indicate that the standard Monte Carlo method is inefficient in this situation.

The above results can be better understood if we consider the barrier-crossing events occurred during time $[0, 1]$. These events are related to the metastability of the system and become rare for $\beta = 5$ and $\beta = 8$. In the "R_c" column of Table 2, we see that very few trajectories can cross the energy barrier when $\beta = 5$, and it becomes even rarer when β is further increased to $\beta = 8$, at which no barrier-crossing trajectories are sampled with $N = 10^4$ trajectories. This observation reveals the fact that crossing the energy barrier is a rare event (in the uncontrolled system) due to system's metastability at moderate temperature. And it also explains why the estimations of the mean values are largely underestimated by the standard Monte Carlo method (compare Table 1 and Table 2). On the other hand, as shown in "R_c" column of Table 1, the barrier-crossing events are much better sampled by the importance sampling estimator. Also Figure 2 shows the control force (4.10) as a function of x and time s for various values of β. We clearly observe that the control acts against the energy barrier (blue region) and assists the slow variable x_s of the system to transit from $x = -1$ to $x = 1$.

We conclude this section with the following remark on some further numerical issues.

Remark 4 1. It is necessary to solve the averaged equation (3.6) for U_0, or equivalently (3.18) for ϕ_0, in order to compute control (3.8). Solving ϕ_0 from (3.18) may be relatively easier because the equation is linear. Furthermore, since equation (3.18) doesn't involve small parameter ϵ any more, it can be solved on a coarser grid and the numerical computation is not expensive.
2. In our example, the probability density $\rho_x(y)$ can be solved analytically and used to obtain averaged dynamics (3.3) or (4.6). More generally, coefficients (3.4) of the averaged dynamics (3.3) could be numerically computed from the time integration of the fast subsystem (3.2). See Chapter 10-11 of [37] and also [45] for more details.
3. In principle, the method described above for solving linear PDE (4.7) is computationally applicable when the dimension k of system's slow variables x is smaller or equal to 3. Alternative approaches need to be studied for systems with more slow variables. See Conclusions for further discussions.

Table 1: Numerical results for importance sampling Monte Carlo method with $T = 1.0$. Columns I and I_N are the mean values computed with $N = 10^5$ ("exact") and $N = 10^4$, respectively. Columns $\mathrm{Var}_u I, \mathrm{RE}_u(I)$ display the variance and the relative error defined in (2.6) and (2.10) estimated from trajectories as in (4.12). Column R_c shows the ratio of the trajectories that have crossed the potential barrier.

β	ϵ	n_x	Δt	I	I_N	$\mathrm{Var}_u I$	$\mathrm{RE}_u(I)$	R_c
1.0	0.1	2000	1.0×10^{-7}	3.52×10^{-2}	3.54×10^{-2}	1.5×10^{-4}	0.35	6.5×10^{-1}
	0.01		1.0×10^{-7}	3.12×10^{-2}	3.12×10^{-2}	1.5×10^{-5}	0.12	6.3×10^{-1}
	0.001		1.0×10^{-8}	3.09×10^{-2}	3.09×10^{-2}	1.5×10^{-6}	0.04	6.2×10^{-1}
5.0	0.1	5000	1.0×10^{-7}	3.82×10^{-8}	3.81×10^{-8}	3.5×10^{-15}	1.55	8.1×10^{-1}
	0.01		1.0×10^{-7}	1.60×10^{-8}	1.62×10^{-8}	4.9×10^{-17}	0.43	7.6×10^{-1}
	0.001		1.0×10^{-8}	1.47×10^{-8}	1.47×10^{-8}	3.7×10^{-18}	0.13	7.6×10^{-1}
8.0	0.1	8000	1.0×10^{-7}	1.59×10^{-12}	1.47×10^{-12}	1.1×10^{-23}	2.26	8.9×10^{-1}
	0.01		5.0×10^{-8}	3.68×10^{-13}	3.68×10^{-13}	4.9×10^{-26}	0.60	8.7×10^{-1}
	0.001		1.0×10^{-8}	3.18×10^{-13}	3.18×10^{-13}	3.2×10^{-27}	0.18	8.7×10^{-1}

5 Proof of the main result

In this section, we prove our main result, Theorem 3.1 in Section 3.1. Since parameter β is fixed, it can be absorbed into coefficients α_1 and α_2, h, and we can assume $\beta = 1$ without loss of

Table 2: Numerical results for standard Monte Carlo method ($u = 0$). The labels have the same meaning as in Table 1.

β	ϵ	Δt	I_N	$\mathrm{Var}_u I$	$\mathrm{RE}_u(I)$	R_c
1.0	0.1	1.0×10^{-7}	3.58×10^{-2}	4.3×10^{-3}	1.83	1.9×10^{-1}
	0.01	1.0×10^{-7}	3.27×10^{-2}	3.9×10^{-3}	1.91	1.8×10^{-1}
	0.001	1.0×10^{-8}	3.14×10^{-2}	3.4×10^{-3}	1.86	1.8×10^{-1}
5.0	0.1	1.0×10^{-7}	2.27×10^{-8}	6.3×10^{-13}	34.97	3.0×10^{-4}
	0.01	1.0×10^{-7}	2.98×10^{-9}	6.4×10^{-16}	8.49	0
	0.001	1.0×10^{-8}	3.61×10^{-9}	6.8×10^{-15}	22.84	1.0×10^{-4}
8.0	0.1	1.0×10^{-7}	3.68×10^{-14}	1.1×10^{-24}	28.50	0
	0.01	5.0×10^{-8}	1.87×10^{-14}	3.8×10^{-25}	32.96	0
	0.001	1.0×10^{-8}	2.01×10^{-14}	4.4×10^{-25}	33.00	0

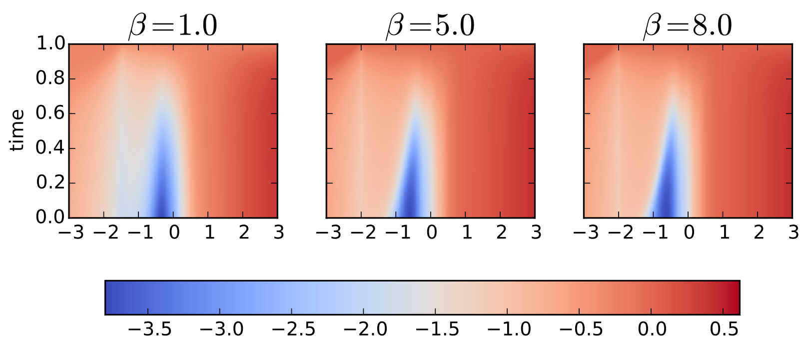

Fig. 2: x-component of control force $\hat{u}^0_s$ defined in (4.10) for different β as function of x and time s.

generality. Also recall that $\|\cdot\|$ denotes the Frobenius norm of matrices and $|\cdot|$ is the Euclidean norm of vectors or the absolute value of a scalar.

Our analysis is based on the solution ϕ^ϵ of the linear backward evolution equation (3.14) and the solution ϕ_0 of (3.18) where, by Feynman-Kac formula, both ϕ^ϵ and ϕ_0 can be expressed in terms of conditional expectations like (3.20).

Idea of the proof. Under Assumption 1, it is well known that both ϕ^ϵ and ϕ_0 are C^1 functions [11,8,20] and that, using the probabilistic representation (3.20), their derivatives have explicit representation formulas in terms of conditional expectations :

$$
\begin{aligned}
\partial_{x_i}\phi^\epsilon &= -\,\mathbf{E}^{x,y}\Big[e^{-\int_t^T h(x_s)ds}\int_t^T \nabla_x h(x_s)\cdot x_{s,x_i}\,ds\Big], \quad 1\le i\le k\\
\partial_{y_i}\phi^\epsilon &= -\,\mathbf{E}^{x,y}\Big[e^{-\int_t^T h(x_s)ds}\int_t^T \nabla_x h(x_s)\cdot x_{s,y_i}\,ds\Big], \quad 1\le i\le l\\
\partial_{x_i}\phi_0 &= -\,\mathbf{E}^{x}\Big[e^{-\int_t^T h(\widetilde{x}_s)ds}\int_t^T \nabla_x h(\widetilde{x}_s)\cdot \widetilde{x}_{s,x_i}\,ds\Big], \quad 1\le i\le k\,.
\end{aligned}
\tag{5.1}
$$

That is, the derivatives can be put inside the expectation, see Section 1.3 of [8] and Section 2.7-2.8 of [30]. Here, we have used Assumption 3 that the running cost h depends only on x, and that dynamics x_s, y_s and $\widetilde{x}_s$ 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 $\mathbf{E}^x$.

Processes $x_{s,x_i}\in\mathbb{R}^k, y_{s,x_i}\in\mathbb{R}^l$ in (5.1) describe the partial derivatives of processes x_s and y_s with respect to the initial conditions and satisfy dynamics

$$
\begin{aligned}
dx_{s,x_i} &= (\nabla_x f\, x_{s,x_i} + \nabla_y f\, y_{s,x_i})ds + (\nabla_x\alpha_1\, x_{s,x_i} + \nabla_y\alpha_1\, y_{s,x_i})dw_s^1\\
dy_{s,x_i} &= \frac{1}{\epsilon}(\nabla_x g\, x_{s,x_i} + \nabla_y g\, y_{s,x_i})ds + \frac{1}{\sqrt{\epsilon}}(\nabla_x\alpha_2\, x_{s,x_i} + \nabla_y\alpha_2\, y_{s,x_i})dw_s^2
\end{aligned}
\quad 1\le i\le k
\tag{5.2}
$$

with $x^j_{t,x_i} = \delta_{ij}, 1\le j\le k$, $y_{t,x_i} = 0\in\mathbb{R}^l$. In the above $\nabla_x\alpha_1 x_{s,x_i}$ denotes the $k\times m_1$ matrix whose components are

$$
(\nabla_x\alpha_1 x_{s,x_i})_{j_1 j_2} = \sum_{r=1}^k \frac{\partial(\alpha_1)_{j_1 j_2}}{\partial x_r}x^r_{s,x_i}\,, \qquad 1\le j_1\le k\,,\quad 1\le j_2\le m_1\,,
\tag{5.3}
$$

and the meanings of other terms are similar. Analogously, processes $x_{s,y_i}\in\mathbb{R}^k$ and $y_{s,y_i}\in\mathbb{R}^l$ satisfy

$$
\begin{aligned}
dx_{s,y_i} &= (\nabla_x f\, x_{s,y_i} + \nabla_y f\, y_{s,y_i})ds + (\nabla_x\alpha_1\, x_{s,y_i} + \nabla_y\alpha_1\, y_{s,y_i})dw_s^1\\
dy_{s,y_i} &= \frac{1}{\epsilon}(\nabla_x g\, x_{s,y_i} + \nabla_y g\, y_{s,y_i})ds + \frac{1}{\sqrt{\epsilon}}(\nabla_x\alpha_2\, x_{s,y_i} + \nabla_y\alpha_2\, y_{s,y_i})dw_s^2
\end{aligned}
\quad 1\le i\le l
\tag{5.4}
$$

with $x_{t,y_i} = 0\in\mathbb{R}^k$, $y^j_{t,y_i} = \delta_{ij}\in\mathbb{R}^l, 1\le j\le l$ (Notice that the above dynamics also hold when coefficient α_1 depends on both x, y, so terms involving $\nabla_y\alpha_1$ are kept there). The above formulas (5.1)–(5.4) will allow us to compare the dynamics x_s, y_s, $\widetilde{x}_s$, the controlled dynamics and the resulting importance sampling estimators. 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 $x_s, \widetilde{x}_s$ are $x_0\in\mathbb{R}^k$ and the initial value of y_s is $y_0\in\mathbb{R}^l$. The notation $\mathbf{E}$ below will always refer to expectation conditioned on these initial values.

To prove Theorem 3.1, we will adapt some estimates used in [32]. Also see [10,8,26,21] for similar techniques. To this end, we follow [32] and define a partition of the interval $[0,T]$ by $[0,\Delta]$, $[\Delta, 2\Delta]$, $\cdots$, $[(M-1)\Delta, M\Delta]$ with $\Delta = T/M$, $M>0$, and consider the auxiliary process

$$
\begin{aligned}
d\hat{x}_s &= f(x_{j\Delta}, \hat{y}_s)ds + \alpha_1(x_s)dw_s^1\\
d\hat{y}_s &= \frac{1}{\epsilon}g(x_{j\Delta}, \hat{y}_s)ds + \frac{1}{\sqrt{\epsilon}}\alpha_2(x_{j\Delta}, \hat{y}_s)dw_s^2
\end{aligned}
\tag{5.5}
$$

for $s \in [j\Delta, (j+1)\Delta)$, $0 \le j \le (M-1)$, 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 $\Delta \le 1$. This auxiliary process will serve as a bridge between (3.1) and (3.3). In contrast to [32] and owed to the fact that we consider controlled dynamics, estimates for 4th-order moments as well as for the processes (5.2) and (5.4) will be needed in order to prove the theorem.

Before entering the details of the various 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 (5.4), 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 ϵ, x_0 and y_0, such that*

$$\begin{aligned} &\max_{0\le s\le T} \mathbf{E}|x_{s,x_i}|^2 \le C, \qquad \max_{0\le s\le T} \mathbf{E}|y_{s,x_i}|^2 \le C, \qquad 1 \le i \le k.\\ &\max_{0\le s\le T} \mathbf{E}|x_{s,y_i}|^2 \le C\epsilon^2, \qquad \mathbf{E}|y_{t,y_i}|^2 \le e^{-\frac{\lambda t}{\epsilon}} + C\epsilon^2, \quad t\in[0,T]\,, \quad 1 \le i \le l. \end{aligned}$$

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 ϵ and can be chosen uniformly for x_0, y_0 which are contained in some bounded domain of $\mathbb{R}^k \times \mathbb{R}^l$, such that*

$$\max_{0\le s\le T} \mathbf{E}|x_s - \widetilde{x}_s|^4 \le C\epsilon^{\frac{1}{2}}\,.$$

Theorem 5.3 *Let Assumptions 1–3 hold. Then $\exists C > 0$, independent of ϵ and can be chosen uniformly for x_0, y_0 which are contained in some bounded domain of $\mathbb{R}^k \times \mathbb{R}^l$, such that*

$$\max_{0\le s\le T} \mathbf{E}|x_{s,x_i} - \widetilde{x}_{s,x_i}|^2 \le C\epsilon^{\frac{1}{4}}\,.$$

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 ϵ and can be chosen uniformly for x, y which are contained in some bounded domain of $\mathbb{R}^k \times \mathbb{R}^l$, such that*

1. *$|\nabla_y \phi^\epsilon| \le C\epsilon$, $\quad |\nabla_x \phi^\epsilon - \nabla_x \phi_0| \le C\epsilon^{\frac{1}{8}}$.*
2. *For $U^\epsilon = -\log\phi^\epsilon$, $U_0 = -\log\phi_0$, we have*

$$|\nabla_y U^\epsilon| \le C\epsilon, \qquad |\nabla_x U^\epsilon - \nabla_x U_0| \le C\epsilon^{\frac{1}{8}}\,. \tag{5.6}$$

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

$$\begin{aligned}|\partial_{y_i}\phi^\epsilon| \leq & \mathbf{E}\Big(e^{-\int_t^T h(x_s)ds}\int_t^T |\nabla_x h(x_s)||x_{s,y_i}|ds\Big)\\ \leq & C\mathbf{E}\int_t^T |x_{s,y_i}|ds \leq C\int_t^T (\mathbf{E}|x_{s,y_i}|^2)^{\frac{1}{2}}ds \leq C\epsilon\,.\end{aligned}$$

To compare $\nabla_x\phi^\epsilon$ with $\nabla_x\phi_0$, we compute that

$$\begin{aligned}&|\partial_{x_i}\phi^\epsilon - \partial_{x_i}\phi_0|\\ \leq&\Big|\mathbf{E}\Big[e^{-\int_t^T h(x_s)ds}\Big(\int_t^T \big(\nabla_x h(x_s)\cdot x_{s,x_i} - \nabla_x h(\widetilde{x}_s)\cdot \widetilde{x}_{s,x_i}\big)\,ds\Big)\Big]\Big|\\ &+\Big|\mathbf{E}\Big[\Big(e^{-\int_t^T h(x_s)ds} - e^{-\int_t^T h(\widetilde{x}_s)ds}\Big)\Big(\int_t^T \nabla_x h(\widetilde{x}_s)\cdot\widetilde{x}_{s,x_i}\,ds\Big)\Big]\Big|\\ =&I_1 + I_2\,.\end{aligned}$$

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

$$\begin{aligned}I_1 \leq&\Big|\mathbf{E}\Big(\int_t^T \big(\nabla_x h(x_s)\cdot x_{s,x_i} - \nabla_x h(\widetilde{x}_s)\cdot\widetilde{x}_{s,x_i}\big)\,ds\Big)\Big|\\ =&\Big|\mathbf{E}\Big(\int_t^T \big[\big(\nabla_x h(x_s) - \nabla_x h(\widetilde{x}_s)\big)\cdot x_{s,x_i} + \nabla_x h(\widetilde{x}_s)\cdot(x_{s,x_i} - \widetilde{x}_{s,x_i})\big]\,ds\Big)\Big|\\ \leq& C\mathbf{E}\Big[\int_t^T \Big(|x_s - \widetilde{x}_s||x_{s,x_i}| + |x_{s,x_i} - \widetilde{x}_{s,x_i}|\Big)\,ds\Big]\\ \leq& C\int_t^T \Big[\big(\mathbf{E}|x_s-\widetilde{x}_s|^2\big)^{\frac{1}{2}}\big(\mathbf{E}|x_{s,x_i}|^2\big)^{\frac{1}{2}} + \big(\mathbf{E}|x_{s,x_i} - \widetilde{x}_{s,x_i}|^2\big)^{\frac{1}{2}}\Big]\,ds \leq C\epsilon^{\frac{1}{8}}\,.\end{aligned}$$

For I_2, we have

$$\begin{aligned}I_2 \leq&\Big[\mathbf{E}\Big(e^{-\int_t^T h(x_s)ds} - e^{-\int_t^T h(\widetilde{x}_s)ds}\Big)^2\Big]^{\frac{1}{2}}\Big[\mathbf{E}\Big(\int_t^T \nabla_x h(\widetilde{x}_s)\cdot\widetilde{x}_{s,x_i}\,ds\Big)^2\Big]^{\frac{1}{2}}\\ \leq& C\Big\{\mathbf{E}\Big[\int_0^1 e^{-\int_t^T (1-r)h(x_s)+rh(\widetilde{x}_s)ds}\Big(\int_t^T |h(\widetilde{x}_s) - h(x_s)|ds\Big)dr\Big]^2\Big\}^{\frac{1}{2}}\Big(\mathbf{E}\int_t^T |\widetilde{x}_{s,x_i}|^2\,ds\Big)^{\frac{1}{2}}\\ \leq& C\Big(\mathbf{E}\int_t^T |\widetilde{x}_s - x_s|^2 ds\Big)^{\frac{1}{2}} \leq C\epsilon^{\frac{1}{8}}\,,\end{aligned}$$

which then entails the estimates for the derivatives of ϕ^ϵ. Meanwhile, using a similar argument,

$$\begin{aligned}|\phi^\epsilon - \phi_0| =&\Big|\mathbf{E}\Big(e^{-\int_t^T h(x_s)ds} - e^{-\int_t^T h(\widetilde{x}_s)ds}\Big)\Big|\\ \leq&\mathbf{E}\Big[\int_0^1 e^{-\int_t^T (1-r)h(x_s)+rh(\widetilde{x}_s)ds}\Big(\int_t^T |h(\widetilde{x}_s) - h(x_s)|ds\Big)dr\Big]\\ \leq& C\mathbf{E}\Big(\int_t^T |h(\widetilde{x}_s) - h(x_s)|ds\Big)\\ \leq& C\int_t^T \big(\mathbf{E}|\widetilde{x}_s - x_s|^4\big)^{\frac{1}{4}}ds \leq C\epsilon^{\frac{1}{8}}\,.\end{aligned}$$

Since h is bounded by Assumption 1, we have $e^{-C(T-t)} \leq \phi^\epsilon \leq e^{C(T-t)}$ is uniformly bounded (and bounded away from zero) for all $\epsilon > 0$. The conclusion concerning $|\nabla_y U^\epsilon|$ and $|\nabla_x U^\epsilon - \nabla_x U_0|$ follows directly from the above estimates. □

Recall that, in Section 2 and Subsection 3.1, $\hat{u}$ is the optimal control as given by (2.16) and that the control $\hat{u}^0$ 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 1–3 hold, and consider the importance sampling method for computing (2.1) under the dynamics (3.1). When the control $\hat{u}^0$ as given in (3.8) is used to perform the importance sampling, the relative error (2.10) of the Monte Carlo estimator satisfies*

$$RE_{\hat{u}^0}(I) \le C\epsilon^{\frac{1}{8}}$$

for $\epsilon \ll 1$ where $C>0$ is a constant independent of ϵ.

Proof In the following we will regard the optimal control $\hat{u}$ and control $\hat{u}^0$ as functions of t, x and y. Using (2.16) and (3.8), we see that Theorem 5.4 implies that $|\hat{u}_s - \hat{u}^0_s| \le C\epsilon^{\frac{1}{8}}$ uniformly on $[0,T]\times D$ where D is any bounded domain of $\mathbb{R}^k \times \mathbb{R}^l$ and constant C depends on domain D. Furthermore, both of them are uniformly bounded on $[0,T]\times\mathbb{R}^k\times\mathbb{R}^l$ from the boundedness of $\phi^\epsilon, \alpha_1, \alpha_2$ and formula (5.1).

Now call $\tilde{\tilde{x}}^u_s, \tilde{\tilde{y}}^u_s$ the controlled dynamics of (3.1) corresponding to the control $\tilde{\tilde{u}}_s = 2\hat{u}_s - \hat{u}^0_s$. Specifically, using (2.16) and (3.8) again, we have (for $\beta = 1$ and assume Assumption 3)

$$\begin{aligned} d\tilde{\tilde{x}}^u_s &= f(\tilde{\tilde{x}}^u_s, \tilde{\tilde{y}}^u_s)ds - \alpha_1(\tilde{\tilde{x}}^u_s)\alpha_1^T(\tilde{\tilde{x}}^u_s)(2\nabla_x U^\epsilon(\tilde{\tilde{x}}^u_s, \tilde{\tilde{y}}^u_s) - \nabla_x U_0(\tilde{\tilde{x}}^u_s)) + \alpha_1(\tilde{\tilde{x}}^u_s)dw^1_s \\ d\tilde{\tilde{y}}^u_s &= \frac{1}{\epsilon}g(\tilde{\tilde{x}}^u_s, \tilde{\tilde{y}}^u_s)ds - \frac{2}{\epsilon}\alpha_2(\tilde{\tilde{x}}^u_s, \tilde{\tilde{y}}^u_s)\alpha_2^T(\tilde{\tilde{x}}^u_s, \tilde{\tilde{y}}^u_s)\nabla_y U^\epsilon(\tilde{\tilde{x}}^u_s, \tilde{\tilde{y}}^u_s) + \frac{1}{\sqrt{\epsilon}}\alpha_2(\tilde{\tilde{x}}^u_s, \tilde{\tilde{y}}^u_s)dw^2_s\,, \end{aligned} \tag{5.7}$$

and control $\tilde{\tilde{u}}_s$ is bounded on $[0,T]\times\mathbb{R}^k\times\mathbb{R}^l$ uniformly for ϵ. This especially implies that Lemma 5.2 and Lemma 5.3 in Subsection 5.2 also hold for dynamics $\tilde{\tilde{x}}^u_s, \tilde{\tilde{y}}^u_s$ (see Remark 6).

Let $R>0$ and for $y\in\mathbb{R}^l$, we define $\chi_R(y)=1$, if $|y|\le R$, 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 both $|x|,|y|\le R$, and otherwise $\chi_R(x,y)=0$. Then applying the uniform approximation $|\hat{u}_s-\hat{u}^0_s|\le C_R\epsilon^{\frac{1}{8}}$ on bounded domain defined by $\chi_R(x,y)$ and the boundedness of both controls, we can recast (2.20) as

$$\begin{aligned} &\tilde{\tilde{\mathbf{E}}}\Big[\exp\Big(\int_t^T |\hat{u}_s-\hat{u}^0_s|^2\chi_R(\tilde{\tilde{x}}^u_s,\tilde{\tilde{y}}^u_s)ds + \int_t^T |\hat{u}_s-\hat{u}^0_s|^2\big(1-\chi_R(\tilde{\tilde{x}}^u_s,\tilde{\tilde{y}}^u_s)\big)ds\Big)\Big] \\ \le& e^{C_R(T-t)\epsilon^{\frac{1}{4}}}\,\tilde{\tilde{\mathbf{E}}}\Big[\exp\Big(\int_t^T |\hat{u}_s-\hat{u}^0_s|^2(1-\chi_R(\tilde{\tilde{x}}^u_s,\tilde{\tilde{y}}^u_s))ds\Big)\Big] \\ \le& e^{C_R(T-t)\epsilon^{\frac{1}{4}}}\,\tilde{\tilde{\mathbf{E}}}\Big[\exp\Big(C\int_t^T (1-\chi_R(\tilde{\tilde{x}}^u_s,\tilde{\tilde{y}}^u_s))ds\Big)\Big] \\ \le& e^{C_R(T-t)\epsilon^{\frac{1}{4}}}\Big[e^{C\delta} + e^{CT}\mathbf{P}\Big(\int_t^T (1-\chi_R(\tilde{\tilde{x}}^u_s,\tilde{\tilde{y}}^u_s))ds \ge \delta\Big)\Big] \end{aligned} \tag{5.8}$$

where $\delta>0$ and C_R is a constant that depends on $R>0$. In the last inequality we have split the expectation according to event $\{\int_t^T(1-\chi_R(\tilde{\tilde{x}}^u_s,\tilde{\tilde{y}}^u_s))ds\ge\delta\}$ and its complement. Therefore, applying the conclusion of Lemma 5.3 to processes $\tilde{\tilde{x}}^u_s,\tilde{\tilde{y}}^u_s$, we can bound the above quantity (5.8) by

$$e^{C_R(T-t)\epsilon^{\frac{1}{4}}}\big[e^{C\delta} + e^{CT}\frac{CT(1+|x|^4+|y|^4)}{\delta R^4}\big]\,.$$

Now we can first choose a small δ and then a large R such that

$$\tilde{\bar{\mathbf{E}}}\Big[\exp\Big(\int_t^T |\hat{u}_s - \hat{u}^0_s|^2 ds\Big)\Big] \le 2e^{C(T-t)\epsilon^{\frac{1}{4}}}$$

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

$$\mathrm{RE}_{\hat{u}^0}(I) \le C\epsilon^{\frac{1}{8}}$$

whenever ϵ is sufficiently small. □

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

We first consider processes x_{s,y_i} and y_{s,y_i} in (5.4), 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 ϵ and whose value may change from line to line.

Lemma 5.1 *Under Assumptions 1–2, there exists $C > 0$, independent of ϵ, x_0 and y_0, such that*

$$\max_{0\le s\le T} \mathbf{E}|x_{s,y_i}|^2 \le C\epsilon, \quad \mathbf{E}|y_{t,y_i}|^2 \le e^{-\frac{\lambda t}{\epsilon}} + C\epsilon, \quad t \in [0,T], \quad 1 \le i \le l. \tag{5.9}$$

Proof Recall the notation in (5.3) and apply Ito's formula to $|x_{s,y_i}|^2$ and $|y_{s,y_i}|^2$. After taking expectation, equation (5.4) yields

$$\begin{aligned} d\mathbf{E}|x_{s,y_i}|^2 &= 2\mathbf{E}\langle \nabla_x f\, x_{s,y_i}, x_{s,y_i}\rangle ds + 2\mathbf{E}\langle \nabla_y f\, y_{s,y_i}, x_{s,y_i}\rangle ds + \mathbf{E}\|\nabla_x \alpha_1\, x_{s,y_i} + \nabla_y \alpha_1\, y_{s,y_i}\|^2 ds \\ d\mathbf{E}|y_{s,y_i}|^2 &= \frac{2}{\epsilon}\mathbf{E}\langle \nabla_x g\, x_{s,y_i}, y_{s,y_i}\rangle ds + \frac{2}{\epsilon}\mathbf{E}\langle \nabla_y g\, y_{s,y_i}, y_{s,y_i}\rangle ds + \frac{1}{\epsilon}\mathbf{E}\|\nabla_x \alpha_2\, x_{s,y_i} + \nabla_y \alpha_2\, y_{s,y_i}\|^2 ds\,, \end{aligned} \tag{5.10}$$

where $\|\cdot\|$ denotes the Frobenius norm of a given matrix. Then, using Cauchy-Schwarz inequality, Lipschitz continuity of the coefficients (Assumption 1) and inequality (3.11) in Remark 2, it follows that

$$\begin{aligned} \frac{d\mathbf{E}|x_{s,y_i}|^2}{ds} &\le C\big(\mathbf{E}|x_{s,y_i}|^2 + \mathbf{E}|y_{s,y_i}|^2\big) \\ \frac{d\mathbf{E}|y_{s,y_i}|^2}{ds} &\le -\frac{\lambda}{\epsilon}\mathbf{E}|y_{s,y_i}|^2 + \frac{C}{\epsilon}\mathbf{E}|x_{s,y_i}|^2 \end{aligned} \tag{5.11}$$

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

The above result can be improved if we additionally impose Assumption 3 and if we treat the initial layer near $t = 0$ more carefully.

Theorem 5.6 *Let Assumptions 1–3 hold. Then $\exists C > 0$, independent of ϵ, x_0 and y_0, such that*

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

Proof Applying Ito's formula in the same way as in Lemma 5.1 and noticing that now coefficient α_1 is independent of y, we can obtain

$$\begin{aligned}
d\mathbf{E}|x_{s,y_i}|^2 &= 2\mathbf{E}\langle \nabla_x f\, x_{s,y_i}, x_{s,y_i}\rangle ds + 2\mathbf{E}\langle \nabla_y f\, y_{s,y_i}, x_{s,y_i}\rangle ds + \mathbf{E}\|\nabla_x \alpha_1\, x_{s,y_i}\|^2 ds\\
d\mathbf{E}|y_{s,y_i}|^2 &= \frac{2}{\epsilon}\mathbf{E}\langle \nabla_x g\, x_{s,y_i}, y_{s,y_i}\rangle ds + \frac{2}{\epsilon}\mathbf{E}\langle \nabla_y g\, y_{s,y_i}, y_{s,y_i}\rangle ds + \frac{1}{\epsilon}\mathbf{E}\|\nabla_x \alpha_2\, x_{s,y_i} + \nabla_y \alpha_2\, y_{s,y_i}\|^2 ds\,.
\end{aligned} \tag{5.12}$$

Now set $t_1 = -\frac{2\epsilon \ln \epsilon}{\lambda}$ and introduce the function $\eta \colon [0,T] \to [0,1]$ by

$$\eta(t) = \begin{cases} 1 - \frac{t}{t_1} & 0 \le t \le t_1 \\ 0 & t_1 < t \le T \end{cases} \tag{5.13}$$

Then using Cauchy-Schwarz inequality and Lipschitz condition in Assumption 1, we have

$$\begin{aligned}
\mathbf{E}\langle \nabla_y f\, y_{s,y_i}, x_{s,y_i}\rangle &\le C\Big(\epsilon^{-\eta(s)}\frac{\mathbf{E}|x_{s,y_i}|^2}{2} + \epsilon^{\eta(s)}\frac{\mathbf{E}|y_{s,y_i}|^2}{2}\Big)\\
\mathbf{E}\langle \nabla_y g\, x_{s,y_i}, y_{s,y_i}\rangle &\le \frac{C^2}{\lambda}\frac{\mathbf{E}|x_{s,y_i}|^2}{2} + \lambda\frac{\mathbf{E}|y_{s,y_i}|^2}{2}\,.
\end{aligned}$$

Substituting them into (5.12) and apply inequality (3.11) in Remark 2, we can obtain

$$\begin{aligned}
\frac{d\mathbf{E}|x_{s,y_i}|^2}{ds} &\le C(1+\epsilon^{-\eta(s)})\mathbf{E}|x_{s,y_i}|^2 + C\epsilon^{\eta(s)}\mathbf{E}|y_{s,y_i}|^2\\
\frac{d\mathbf{E}|y_{s,y_i}|^2}{ds} &\le -\frac{\lambda}{\epsilon}\mathbf{E}|y_{s,y_i}|^2 + \frac{C}{\epsilon}\mathbf{E}|x_{s,y_i}|^2
\end{aligned}$$

with $\mathbf{E}|x_{0,y_i}|^2 = 0$, $\mathbf{E}|y_{0,y_i}|^2 = 1$. 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 (5.5). Bear in mind that $\beta = 1$ throughout this section. For processes x_s, y_s satisfying (3.1), we have:

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

$$\max_{0\le s\le T}\mathbf{E}|x_s|^4 \le C\big(|x_0|^4 + |y_0|^4 + 1\big), \qquad \max_{0\le s\le T}\mathbf{E}|y_s|^4 \le C\big(|y_0|^4 + |x_0|^4 + 1\big). \tag{5.14}$$

Proof Applying Ito's formula to $|x_s|^4$ and taking expectation, we can obtain

$$\begin{aligned}
\frac{d\mathbf{E}|x_s|^4}{ds} =& 4\mathbf{E}\Big(|x_s|^2\langle f(x_s,y_s), x_s\rangle\Big) + 2\mathbf{E}\Big(|x_s|^2\|\alpha_1(x_s,y_s)\|^2\Big) + 4\mathbf{E}\Big(|\alpha_1^T(x_s,y_s)x_s|^2\Big)\\
\le& 4\mathbf{E}\Big(|x_s|^2\langle f(x_s,y_s), x_s\rangle\Big) + 6\mathbf{E}\Big(|x_s|^2\|\alpha_1(x_s,y_s)\|^2\Big)\,,
\end{aligned}$$

and similarly for $|y_s|^4$,

$$\frac{d\mathbf{E}|y_s|^4}{ds} \le \frac{4}{\epsilon}\mathbf{E}\Big(|y_s|^2\langle g(x_s,y_s), y_s\rangle\Big) + \frac{6}{\epsilon}\mathbf{E}\Big(|y_s|^2\|\alpha_2(x_s,y_s)\|^2\Big).$$

By Assumption 1, f is Lipschitz and α_1 is bounded. We also know from Remark 2 that $|f(x_s,y_s)| \leq C(1+|x_s|+|y_s|)$ and inequality (3.13) holds. Together with Young's inequality, we obtain

$$\begin{aligned}\frac{d\mathbf{E}|x_s|^4}{ds} \leq & C\Big(\mathbf{E}|x_s|^4 + \mathbf{E}|y_s|^4 + 1\Big)\\ \frac{d\mathbf{E}|y_s|^4}{ds} \leq & -\frac{\lambda}{\epsilon}\mathbf{E}|y_s|^4 + \frac{C}{\epsilon}\Big(\mathbf{E}|x_s|^4 + 1\Big)\,.\end{aligned}$$

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

Remark 5 Reiterating the above argument, we can prove that the solutions of (5.5) and (3.3) satisfy

$$\max_{0\leq s\leq T}\mathbf{E}|\hat{x}_s|^4 \leq C\big(|x_0|^4+|y_0|^4+1\big), \qquad \max_{0\leq s\leq T}\mathbf{E}|\hat{y}_s|^4 \leq C\big(|y_0|^4+|x_0|^4+1\big)\,, \tag{5.15}$$

and

$$\max_{0\leq s\leq T}\mathbf{E}|\widetilde{x}_s|^4 \leq C\big(|x_0|^4+1\big)\,, \tag{5.16}$$

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

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

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

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

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

$$\begin{aligned}\mathbf{P}\Big(\int_0^T \big(1-\chi_R(y_s)\big)ds \geq \delta\Big) \leq & \frac{C\big(1+|x_0|^4+|y_0|^4\big)}{\delta R^4}\,,\\ \mathbf{P}\Big(\int_0^T \big(1-\chi_R(x_s,y_s)\big)ds \geq \delta\Big) \leq & \frac{C\big(1+|x_0|^4+|y_0|^4\big)}{\delta R^4}\,.\end{aligned}$$

Proof The proof is standard. Since f is Lipschitz, using Hölder's inequality, we have

$$\begin{aligned}|x_s|^4 \leq & C\Big(|x_0|^4 + \Big|\int_0^s f(x_r,y_r)dr\Big|^4 + \Big|\int_0^s \alpha_1(x_r,y_r)dw_r^1\Big|^4\Big)\\ \leq & C\Big(|x_0|^4 + s^3\int_0^s |f(x_r,y_r)|^4 dr + \Big|\int_0^s \alpha_1(x_r,y_r)dw_r^1\Big|^4\Big)\\ \leq & C\Big(|x_0|^4 + T^3\int_0^T (|x_r|^4+|y_r|^4+1)dr + \Big|\int_0^s \alpha_1(x_r,y_r)dw_r^1\Big|^4\Big)\,.\end{aligned}$$

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

$$\mathbf{E}(\sup_{0\leq s\leq T}|x_s|^4) \leq C\Big[|x_0|^4 + T^3\mathbf{E}\int_0^T (|x_r|^4+|y_r|^4+1)dr + \mathbf{E}\Big(\sup_{0\leq s\leq T}\Big(\int_0^s \alpha_1(x_r,y_r)dw_r^1\Big)^4\Big)\Big]\,.$$

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 [28]. Hence

$$\mathbf{E}(\sup_{0\le s\le T}|x_s|^4) \le C(|x_0|^4+|y_0|^4+1)+C\Big(\mathbf{E}\int_0^T|\alpha_1(x_r,y_r)|^2dr\Big)^2$$

and the boundedness of α_1 entails

$$\mathbf{E}(\sup_{0\le s\le T}|x_s|^4) \le C\big(1+|x_0|^4+|y_0|^4\big)\,.$$

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

$$\begin{aligned}R^4\mathbf{E}\Big[\int_0^T\big(1-\chi_R(y_s)\big)ds\Big] &\le \mathbf{E}\Big[\int_0^T|y_s|^4(1-\chi_R(y_s))ds\Big]\\ &\le \mathbf{E}\Big(\int_0^T|y_s|^4ds\Big)\le C\big(1+|x_0|^4+|y_0|^4\big)\,.\end{aligned}$$

Thus, by Chebyshev's inequality,

$$\mathbf{P}\Big(\int_0^T\big(1-\chi_R(y_s)\big)ds\ge\delta\Big)\le\frac{C\big(1+|x_0|^4+|y_0|^4\big)}{\delta R^4}\,.$$

The second inequality follows in the same fashion. □

Remark 6 Based on the result of Theorem 5.4, we could prove that the same conclusions of Lemma 5.2 and Lemma 5.3 also hold for processes (5.7). See discussions in the proof of Theorem 5.5.

We proceed our analysis by inspecting SDE (5.2) for processes x_{s,x_i}, y_{s,x_i}, for which we seek the analogue of the inequality (5.11). In this case the initial values satisfy $\mathbf{E}|x_{0,x_i}|^2=1$, $\mathbf{E}|y_{0,x_i}|^2=0$ and by similar argument as in the proof of Lemma 5.1, we find:

Lemma 5.4 *Under Assumptions 1–2, there exists $C>0$, independent of ϵ, x_0 and y_0, such that*

$$\max_{0\le s\le T}\mathbf{E}|x_{s,x_i}|^2\le C,\quad \max_{0\le s\le T}\mathbf{E}|y_{s,x_i}|^2\le C,\qquad 1\le i\le k. \tag{5.17}$$

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

Lemma 5.5 *Under Assumptions 1–2, there exists $C>0$, independent of ϵ, x_0 and y_0, such that*

$$\max_{0\le s\le T}\mathbf{E}|x_{s,x_i}|^4\le C,\quad \max_{0\le s\le T}\mathbf{E}|y_{s,x_i}|^4\le C,\quad 1\le i\le k. \tag{5.18}$$

Proof The proof is similar to Lemma 5.2. Using Ito's formula, we obtain

$$\begin{aligned}d\mathbf{E}|x_{s,x_i}|^4 =&4\mathbf{E}\Big(|x_{s,x_i}|^2\langle\nabla_x f\,x_{s,x_i}+\nabla_y f\,y_{s,x_i},x_{s,x_i}\rangle\Big)ds+2\mathbf{E}\Big(|x_{s,x_i}|^2\|\nabla_x\alpha_1\,x_{s,x_i}+\nabla_y\alpha_1\,y_{s,x_i}\|^2\Big)ds\\ &+4\mathbf{E}\Big(|(\nabla_x\alpha_1\,x_{s,x_i}+\nabla_y\alpha_1\,y_{s,x_i})^Tx_{s,x_i}|^2\Big)ds\\ \le&4\mathbf{E}\Big(|x_{s,x_i}|^2\langle\nabla_x f\,x_{s,x_i}+\nabla_y f\,y_{s,x_i},x_{s,x_i}\rangle\Big)ds+6\mathbf{E}\Big(|x_{s,x_i}|^2\|\nabla_x\alpha_1\,x_{s,x_i}+\nabla_y\alpha_1\,y_{s,x_i}\|^2\Big)ds\\ d\mathbf{E}|y_{s,x_i}|^4 =&\frac{4}{\epsilon}\mathbf{E}\Big(|y_{s,x_i}|^2\langle\nabla_x g\,x_{s,x_i}+\nabla_y g\,y_{s,x_i},y_{s,x_i}\rangle\Big)ds+\frac{2}{\epsilon}\mathbf{E}\Big(|y_{s,x_i}|^2\|\nabla_x\alpha_2\,x_{s,x_i}+\nabla_y\alpha_2\,y_{s,x_i}\|^2\Big)ds\\ &+\frac{4}{\epsilon}\mathbf{E}\Big(|(\nabla_x\alpha_2\,x_{s,x_i}+\nabla_y\alpha_2\,y_{s,x_i})^Ty_{s,x_i}|^2\Big)ds\\ \le&\frac{4}{\epsilon}\mathbf{E}\Big(|y_{s,x_i}|^2\langle\nabla_x g\,x_{s,x_i}+\nabla_y g\,y_{s,x_i},y_{s,x_i}\rangle\Big)ds+\frac{6}{\epsilon}\mathbf{E}\Big(|y_{s,x_i}|^2\|\nabla_x\alpha_2\,x_{s,x_i}+\nabla_y\alpha_2\,y_{s,x_i}\|^2\Big)ds\,.\end{aligned} \tag{5.19}$$

Lipschitz conditions on the coefficients in Assumption 1, Assumption 2, especially inequality (3.11) in Remark 2 as well as Young's inequality now readily imply that

$$\begin{aligned}\frac{d\mathbf{E}|x_{s,x_i}|^4}{ds} &\leq C\big(\mathbf{E}|x_{s,x_i}|^4 + \mathbf{E}|y_{s,x_i}|^4\big)\\ \frac{d\mathbf{E}|y_{s,x_i}|^4}{ds} &\leq -\frac{2\lambda}{\epsilon}\mathbf{E}|y_{s,x_i}|^4 + \frac{C}{\epsilon}\mathbf{E}|x_{s,x_i}|^4\,,\end{aligned}$$

with $\mathbf{E}|y_{0,x_i}|^4 = 0$, $\mathbf{E}|x_{0,x_i}|^4 = 1$. The assertion then follows by the same argument as in the proof of Claim A.1 in Appendix A. □

We also have the following simple bounds for processes x_s and x_{s,x_i}.

Lemma 5.6 *Let $\Delta \leq 1$, $s \in [j\Delta, (j+1)\Delta)$, $0 \leq j \leq M-1$. Further let Assumptions 1–2 hold.*

1. *For process x_s satisfying (3.1), it holds*

$$\mathbf{E}|x_s - x_{j\Delta}|^4 \leq C(s-j\Delta)^2, \tag{5.20}$$

where constant $C > 0$ is independent of ϵ, Δ and can be chosen uniformly for x_0 and y_0 which are contained in some bounded domain of $\mathbb{R}^k \times \mathbb{R}^l$. The same bound is satisfied by processes $\widetilde{x}_s, \hat{x}_s$.

2. *For process x_{s,x_i} in (5.2), we have*

$$\mathbf{E}|x_{s,x_i} - x_{j\Delta,x_i}|^4 \leq C(s-j\Delta)^2 \leq C\Delta^2, \tag{5.21}$$

with constant $C > 0$ that is independent of ϵ, x_0, y_0. The same inequality holds if x_{s,x_i} is replaced by processes $\hat{x}_{s,x_i}$ and $\widetilde{x}_{s,x_i}$.

Proof For the first part of the conclusion, using that function f is Lipschitz and therefore $|f(x_r, y_r)| \leq C(1+|x_r|+|y_r|)$ (Remark 2), α_1 is bounded (Assumption 1), as well as Lemma 5.2, we can conclude that

$$\begin{aligned}\mathbf{E}|x_s - x_{j\Delta}|^4 =&\mathbf{E}\Big[\int_{j\Delta}^s f(x_r,y_r)dr + \int_{j\Delta}^s \alpha_1(x_r,y_r)dw_r^1\Big]^4\\ \leq& C\mathbf{E}\Big[\int_{j\Delta}^s \big(1+|x_r|+|y_r|\big)dr\Big]^4 + C\mathbf{E}\Big[\int_{j\Delta}^s \alpha_1(x_r,y_r)dw_r^1\Big]^4\\ \leq& C\big(|x_0|^4+|y_0|^4+1\big)(s-j\Delta)^4 + C(s-j\Delta)^2\\ \leq& C(s-j\Delta)^2\,,\end{aligned}$$

where, in the last inequality, we have used the fact that $\Delta \leq 1$. It is clear that a common constant C can be chosen for x_0, y_0 which are contained in some bounded domain.

The second part of the conclusion can be obtained in a similar way by using Lipschitz condition of coefficients as well as Lemma 5.5. □

5.3 Approximation by the auxiliary process

In this subsection, we study the approximations of the original dynamics (3.1) by the auxiliary discrete process (5.5) and the averaged dynamics (3.3). To this end, we recall Hölder and Young's inequalities : Given two random variables X, Y, and $p, q > 0$ with $\frac{1}{p} + \frac{1}{q} = 1$, it holds that

$$\mathbf{E}|XY| \leq \big(\mathbf{E}|X|^p\big)^{\frac{1}{p}} \big(\mathbf{E}|Y|^q\big)^{\frac{1}{q}} \leq \frac{\mathbf{E}|X|^p}{p} + \frac{\mathbf{E}|Y|^q}{q} \,. \tag{5.22}$$

First of all, we have

Lemma 5.7 *Suppose that Assumptions 1–3 are met. For processes x_s, y_s satisfying (3.1) and the auxiliary processes $\hat{x}_s$, $\hat{y}_s$ defined in (5.5), we have*

$$\max_{0\leq s\leq T} \mathbf{E}|y_s - \hat{y}_s|^4 \leq C\Delta^2 \,, \qquad \max_{0\leq s\leq T} \mathbf{E}|x_s - \hat{x}_s|^4 \leq C\Delta^2 \,, \tag{5.23}$$

where constant $C > 0$ is independent of ϵ, Δ and can be chosen uniformly for x_0, y_0 which are contained on some bounded domain of $\mathbb{R}^k \times \mathbb{R}^l$.

Proof Let $j = \lfloor \frac{s}{\Delta} \rfloor$, which is the largest integer smaller or equal to $\frac{s}{\Delta}$. Applying Ito's formula and using the Lipschitz condition on coefficients g, α_2 in Assumptions 1, the inequality in Assumption 2, the conclusion of Lemma 5.6, as well as inequality (5.22), we can estimate

$$\begin{aligned}
&\frac{d\mathbf{E}|y_s - \hat{y}_s|^4}{ds}\\
=&\frac{4}{\epsilon}\mathbf{E}\Big(|y_s - \hat{y}_s|^2 \langle y_s - \hat{y}_s, g(x_s, y_s) - g(x_{j\Delta}, \hat{y}_s)\rangle\Big) + \frac{2}{\epsilon}\mathbf{E}\Big(|y_s - \hat{y}_s|^2 \|\alpha_2(x_s, y_s) - \alpha_2(x_{j\Delta}, \hat{y}_s)\|^2\Big)\\
&+ \frac{4}{\epsilon}\mathbf{E}\Big(\Big|\big(\alpha_2(x_s, y_s) - \alpha_2(x_{j\Delta}, \hat{y}_s)\big)^T (y_s - \hat{y}_s)\Big|^2\Big)\\
\leq&\frac{4}{\epsilon}\mathbf{E}\Big(|y_s - \hat{y}_s|^2 \langle y_s - \hat{y}_s, g(x_s, y_s) - g(x_{j\Delta}, \hat{y}_s)\rangle\Big) + \frac{6}{\epsilon}\mathbf{E}\Big(|y_s - \hat{y}_s|^2 \|\alpha_2(x_s, y_s) - \alpha_2(x_{j\Delta}, \hat{y}_s)\|^2\Big)\\
\leq&\frac{4}{\epsilon}\mathbf{E}\Big[|y_s - \hat{y}_s|^2 \Big(\langle y_s - \hat{y}_s, g(x_s, y_s) - g(x_s, \hat{y}_s)\rangle + 3\|\alpha_2(x_s, y_s) - \alpha_2(x_s, \hat{y}_s)\|^2\Big)\Big]\\
&+ \frac{4}{\epsilon}\mathbf{E}\Big[|y_s - \hat{y}_s|^2 \Big(\langle y_s - \hat{y}_s, g(x_s, \hat{y}_s) - g(x_{j\Delta}, \hat{y}_s)\rangle + 3\|\alpha_2(x_s, \hat{y}_s) - \alpha_2(x_{j\Delta}, \hat{y}_s)\|^2\Big)\Big]\\
\leq& -\frac{4\lambda}{\epsilon}\mathbf{E}|y_s - \hat{y}_s|^4 + \frac{C}{\epsilon}\mathbf{E}\Big(|y_s - \hat{y}_s|^3 |x_s - x_{j\Delta}|\Big) + \frac{C}{\epsilon}\mathbf{E}\Big(|y_s - \hat{y}_s|^2 |x_s - x_{j\Delta}|^2\Big)\\
\leq& -\frac{2\lambda}{\epsilon}\mathbf{E}|y_s - \hat{y}_s|^4 + \frac{C}{\epsilon}\mathbf{E}|x_s - x_{j\Delta}|^4\\
\leq& -\frac{2\lambda}{\epsilon}\mathbf{E}|y_s - \hat{y}_s|^4 + \frac{C}{\epsilon}\Delta^2
\end{aligned}$$

which, by Gronwall's inequality, yields the first inequality. For the second inequality, applying Ito's formula, taking Assumption 1, Lemma 5.6 and the above estimate into account, we obtain

$$\begin{aligned}
\frac{d\mathbf{E}|\hat{x}_s - x_s|^4}{ds} =& 4\mathbf{E}\Big(|\hat{x}_s - x_s|^2 \langle f(x_{j\Delta}, \hat{y}_s) - f(x_s, y_s), \hat{x}_s - x_s\rangle\Big)\\
\leq& C\mathbf{E}\Big[|\hat{x}_s - x_s|^3 \Big(|x_{j\Delta} - x_s| + |\hat{y}_s - y_s|\Big)\Big]\\
\leq& C\Big(\mathbf{E}|\hat{x}_s - x_s|^4 + \mathbf{E}|x_{j\Delta} - x_s|^4 + \mathbf{E}|\hat{y}_s - y_s|^4\Big)\\
\leq& C\mathbf{E}|\hat{x}_s - x_s|^4 + C\Delta^2 \,,
\end{aligned}$$

and the conclusion follows again by applying Gronwall's inequality. □

The following elementary estimate will be useful.

Claim 5.1 *Define $F(x)=|x|^2x$, $\forall x\in\mathbb{R}^m$. We have $|F(x)-F(y)|\le\frac{3}{2}(|x|^2+|y|^2)|x-y|$.*

Proof We have

$$
\begin{aligned}
&|F(x)-F(y)|\\
=&\Big|\int_0^1 \frac{d}{dt}F\big((1-t)x+ty\big)dt\Big|\\
=&\Big|\int_0^1 \Big[2\langle(1-t)x+ty,y-x\rangle\big((1-t)x+ty\big)+|(1-t)x+ty|^2(y-x)\Big]dt\Big|\\
\le&3\int_0^1 |(1-t)x+ty|^2|y-x|dt\le\frac{3}{2}\big(|x|^2+|y|^2\big)|x-y|\,.
\end{aligned}
$$

□

As the next step, we show that the averaged process $\widetilde{x}_s$ in (3.3) can be approximated by the time-discrete process (5.5) as well.

Lemma 5.8 *Under Assumptions 1–3, we have*

$$
\max_{0\le s\le T}\mathbf{E}|\hat{x}_s-\widetilde{x}_s|^4\le C\left(\frac{\epsilon}{\lambda\Delta}+\Delta\right)e^{C\left(1+\frac{\epsilon}{\lambda\Delta}\right)T}\,. \tag{5.24}
$$

where constant $C>0$ is independent of ϵ,Δ and can be chosen uniformly for x_0, y_0 which are contained in some bounded domain of $\mathbb{R}^k\times\mathbb{R}^l$. Especially, for $\Delta=\epsilon^{\frac{1}{2}}$, we have $\max\limits_{0\le s\le T}\mathbf{E}|\hat{x}_s-\widetilde{x}_s|^4\le C\epsilon^{\frac{1}{2}}$.

Proof We apply Ito's formula to $|\hat{x}_s-\widetilde{x}_s|^4$ and take expectations similarly as before. Using the function F defined in Claim 5.1, we can estimate

$$
\begin{aligned}
&\mathbf{E}|\hat{x}_s-\widetilde{x}_s|^4\\
\le&4\int_0^s\mathbf{E}\Big(|\hat{x}_r-\widetilde{x}_r|^2\big\langle\hat{x}_r-\widetilde{x}_r,f(x_{\lfloor\frac{r}{\Delta}\rfloor\Delta},\hat{y}_r)-\widetilde{f}(\widetilde{x}_r)\big\rangle\Big)dr+6\int_0^s\mathbf{E}\Big(|\hat{x}_r-\widetilde{x}_r|^2|\alpha_1(x_r)-\alpha_1(\widetilde{x}_r)|^2\Big)dr\\
=&4\int_0^s\mathbf{E}\Big(\big\langle F(\hat{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta}-\widetilde{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta}),f(x_{\lfloor\frac{r}{\Delta}\rfloor\Delta},\hat{y}_r)-\widetilde{f}(x_{\lfloor\frac{r}{\Delta}\rfloor\Delta})\big\rangle\Big)dr\\
&+4\int_0^s\mathbf{E}\Big(\big\langle F(\hat{x}_r-\widetilde{x}_r)-F(\hat{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta}-\widetilde{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta}),f(x_{\lfloor\frac{r}{\Delta}\rfloor\Delta},\hat{y}_r)-\widetilde{f}(x_{\lfloor\frac{r}{\Delta}\rfloor\Delta})\big\rangle\Big)dr\\
&+4\int_0^s\mathbf{E}\Big(\big\langle F(\hat{x}_r-\widetilde{x}_r),\widetilde{f}(x_{\lfloor\frac{r}{\Delta}\rfloor\Delta})-\widetilde{f}(\widetilde{x}_r)\big\rangle\Big)dr\\
&+6\int_0^s\mathbf{E}\Big(|\hat{x}_r-\widetilde{x}_r|^2|\alpha_1(x_r)-\alpha_1(\widetilde{x}_r)|^2\Big)dr\\
=&I_1+I_2+I_3+I_4\,.
\end{aligned}
$$

We estimate the above four terms in the sum separately. For I_1, we have

$$
\begin{aligned}
|I_1| \leq & 4 \sum_{j=0}^{\lfloor s/\Delta \rfloor} \int_{j\Delta}^{[(j+1)\Delta]\wedge s} \mathbf{E}\Big(|\hat{x}_{j\Delta} - \widetilde{x}_{j\Delta}|^3 |\mathbf{E}_{j\Delta} f(x_{j\Delta}, \hat{y}_r) - \widetilde{f}(x_{j\Delta})|\Big) dr \\
\leq & C \sum_{j=0}^{\lfloor s/\Delta \rfloor} \int_{j\Delta}^{[(j+1)\Delta]\wedge s} \mathbf{E}\Big(|\hat{x}_{j\Delta} - \widetilde{x}_{j\Delta}|^3 (|x_{j\Delta}| + |\hat{y}_{j\Delta}| + 1)\Big) e^{-\frac{\lambda(r-j\Delta)}{\epsilon}} dr \\
\leq & \frac{\epsilon C}{\lambda} \mathbf{E}\Big[\Big(\sum_{j=0}^{\lfloor s/\Delta \rfloor} |\hat{x}_{j\Delta} - \widetilde{x}_{j\Delta}|^4\Big)^{\frac{3}{4}} \Big(\sum_{j=0}^{\lfloor s/\Delta \rfloor} \big(|x_{j\Delta}| + |\hat{y}_{j\Delta}| + 1\big)^4\Big)^{\frac{1}{4}}\Big] \\
\leq & \frac{\epsilon C}{\lambda} \Big(\mathbf{E} \sum_{j=0}^{\lfloor s/\Delta \rfloor} |\hat{x}_{j\Delta} - \widetilde{x}_{j\Delta}|^4\Big)^{\frac{3}{4}} \Big(\mathbf{E} \sum_{j=0}^{\lfloor s/\Delta \rfloor} (|x_{j\Delta}| + |\hat{y}_{j\Delta}| + 1)^4\Big)^{\frac{1}{4}} \\
\leq & \frac{\epsilon C}{\lambda\Delta} \Big(\mathbf{E} \sum_{j=0}^{\lfloor s/\Delta \rfloor} |\hat{x}_{j\Delta} - \widetilde{x}_{j\Delta}|^4 \Delta + \mathbf{E} \sum_{j=0}^{\lfloor s/\Delta \rfloor} (|x_{j\Delta}| + |\hat{y}_{j\Delta}| + 1)^4 \Delta\Big) \\
\leq & \frac{\epsilon C}{\lambda\Delta} \mathbf{E} \int_0^s |\hat{x}_r - \widetilde{x}_r|^4 dr + \frac{\epsilon C}{\lambda\Delta} \mathbf{E} \int_0^s \Big| |\hat{x}_{\lfloor \frac{r}{\Delta} \rfloor \Delta} - \widetilde{x}_{\lfloor \frac{r}{\Delta} \rfloor \Delta}|^4 - |\hat{x}_r - \widetilde{x}_r|^4 \Big| dr \\
& + \frac{\epsilon C}{\lambda\Delta} \mathbf{E} \sum_{j=0}^{\lfloor s/\Delta \rfloor} \Big(|x_{j\Delta}| + |\hat{y}_{j\Delta}| + 1\Big)^4 \Delta \,,
\end{aligned}
$$

where in the first inequality above, $\mathbf{E}_{j\Delta}$ denotes the expectation conditioned on $\hat{y}_s$ at time $s = j\Delta$. We have used Lemma B.3 in Appendix B to derive the second inequality. Hölder inequality and Young's inequality (5.22) were also used. Therefore, by Lemma 5.2 and Remark 5, the last inequality implies

$$
|I_1| \leq \frac{\epsilon C}{\lambda\Delta} \mathbf{E} \int_0^s |\hat{x}_r - \widetilde{x}_r|^4 dr + \frac{C s \epsilon}{\lambda\Delta}.
$$

For I_2, since functions $f, \widetilde{f}$ are Lipschitz, we have

$$
\begin{aligned}
|f(x_{\lfloor \frac{r}{\Delta} \rfloor \Delta}, \hat{y}_r)| \leq & C\big(1 + |x_{\lfloor \frac{r}{\Delta} \rfloor \Delta}| + |\hat{y}_r|\big) \,, \\
|\widetilde{f}(x_{\lfloor \frac{r}{\Delta} \rfloor \Delta})| \leq & C\big(1 + |x_{\lfloor \frac{r}{\Delta} \rfloor \Delta}|\big) \,.
\end{aligned}
$$

Then using Claim 5.1, Lemma 5.2 and Lemma 5.6, as well as Hölder and Young's inequalities (5.22), we can estimate

$$
\begin{aligned}
|I_2| \leq & C\mathbf{E}\int_0^s \Big(|\hat{x}_r-\widetilde{x}_r|^2+|\hat{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta}-\widetilde{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta}|^2\Big)\\
&\times\Big|(\hat{x}_r-\hat{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta})-(\widetilde{x}_r-\widetilde{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta})\Big|\Big(1+|x_{\lfloor\frac{r}{\Delta}\rfloor\Delta}|+|\hat{y}_r|\Big)dr\\
\leq & C\mathbf{E}\int_0^s \Big(|\hat{x}_r-\widetilde{x}_r|^2+|(\hat{x}_r-\hat{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta})-(\widetilde{x}_r-\widetilde{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta})|^2\Big)\\
&\times\Big|(\hat{x}_r-\hat{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta})-(\widetilde{x}_r-\widetilde{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta})\Big|\Big(1+|x_{\lfloor\frac{r}{\Delta}\rfloor\Delta}|+|\hat{y}_r|\Big)dr\\
\leq & C\mathbf{E}\int_0^s |\hat{x}_r-\widetilde{x}_r|^4dr+C\mathbf{E}\int_0^s\Big|(\hat{x}_r-\hat{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta})-(\widetilde{x}_r-\widetilde{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta})\Big|^3\Big(1+|x_{\lfloor\frac{r}{\Delta}\rfloor\Delta}|+|\hat{y}_r|\Big)dr\\
&+C\mathbf{E}\int_0^s\Big|(\hat{x}_r-\hat{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta})-(\widetilde{x}_r-\widetilde{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta})\Big|^2\Big(1+|x_{\lfloor\frac{r}{\Delta}\rfloor\Delta}|+|\hat{y}_r|\Big)^2dr\\
\leq & C\mathbf{E}\int_0^s |\hat{x}_r-\widetilde{x}_r|^4dr\\
&+C\int_0^s\Big[\mathbf{E}\big|(\hat{x}_r-\hat{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta})-(\widetilde{x}_r-\widetilde{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta})\big|^4\Big]^{\frac{3}{4}}\Big[\mathbf{E}\big(1+|x_{\lfloor\frac{r}{\Delta}\rfloor\Delta}|+|\hat{y}_r|\big)^4\Big]^{\frac{1}{4}}dr\\
&+C\int_0^s\Big[\mathbf{E}\big|(\hat{x}_r-\hat{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta})-(\widetilde{x}_r-\widetilde{x}_{\lfloor\frac{r}{\Delta}\rfloor\Delta})\big|^4\Big]^{\frac{1}{2}}\Big[\mathbf{E}\big(1+|x_{\lfloor\frac{r}{\Delta}\rfloor\Delta}|+|\hat{y}_r|\big)^4\Big]^{\frac{1}{2}}dr\\
\leq & C\mathbf{E}\int_0^s |\hat{x}_r-\widetilde{x}_r|^4dr+Cs(\Delta+\Delta^{\frac{3}{2}})\,.
\end{aligned}
$$

For I_3, since function $\widetilde{f}$ is Lipschitz, we have

$$
\begin{aligned}
|I_3| \leq & C\mathbf{E}\int_0^s |\hat{x}_r-\widetilde{x}_r|^3|x_{\lfloor\frac{r}{\Delta}\rfloor\Delta}-\widetilde{x}_r|dr\\
= & C\mathbf{E}\int_0^s |\hat{x}_r-\widetilde{x}_r|^3\Big|(x_{\lfloor\frac{r}{\Delta}\rfloor\Delta}-x_r)+(x_r-\hat{x}_r)+(\hat{x}_r-\widetilde{x}_r)\Big|dr\\
\leq & C\mathbf{E}\int_0^s |\hat{x}_r-\widetilde{x}_r|^4dr+C\mathbf{E}\int_0^s |\hat{x}_r-\widetilde{x}_r|^3|x_{\lfloor\frac{r}{\Delta}\rfloor\Delta}-x_r|dr+C\mathbf{E}\int_0^s |\hat{x}_r-\widetilde{x}_r|^3|x_r-\hat{x}_r|dr\\
\leq & C\mathbf{E}\int_0^s |\hat{x}_r-\widetilde{x}_r|^4dr+C\mathbf{E}\int_0^s |x_{\lfloor\frac{r}{\Delta}\rfloor\Delta}-x_r|^4dr+C\mathbf{E}\int_0^s |x_r-\hat{x}_r|^4dr\\
\leq & C\mathbf{E}\int_0^s |\hat{x}_r-\widetilde{x}_r|^4dr+Cs\Delta^2\,,
\end{aligned}
$$

where Lemma 5.6, Lemma 5.7 and Young's inequality have been used.

Finally, using that coefficient α_1 is Lipschitz and Lemma 5.7, we obtain the following bound for I_4:

$$
\begin{aligned}
|I_4| \leq & C\mathbf{E}\int_0^s |\hat{x}_r-\widetilde{x}_r|^2|x_r-\widetilde{x}_r|^2dr\\
= & C\mathbf{E}\int_0^s |\hat{x}_r-\widetilde{x}_r|^2|(x_r-\hat{x}_r)+(\hat{x}_r-\widetilde{x}_r)|^2dr\\
\leq & C\mathbf{E}\int_0^s |\hat{x}_r-\widetilde{x}_r|^4dr+C\mathbf{E}\int_0^s |\hat{x}_r-\widetilde{x}_r|^2|x_r-\hat{x}_r|^2dr\\
\leq & C\mathbf{E}\int_0^s |\hat{x}_r-\widetilde{x}_r|^4dr+C\mathbf{E}\int_0^s |x_r-\hat{x}_r|^4dr\\
\leq & C\mathbf{E}\int_0^s |\hat{x}_r-\widetilde{x}_r|^4dr+Cs\Delta^2\,.
\end{aligned}
$$

Combining the above estimates, we have obtained the bound (assuming $\Delta \le 1$)

$$\mathbf{E}|\hat{x}_s - \widetilde{x}_s|^4 \le C\left(1+\frac{\epsilon}{\lambda\Delta}\right)\mathbf{E}\int_0^s |\hat{x}_r - \widetilde{x}_r|^4 dr + Cs\left(\frac{\epsilon}{\lambda\Delta} + \Delta\right), \tag{5.25}$$

and Gronwall's inequality yields the assertion

$$\mathbf{E}|\hat{x}_s - \widetilde{x}_s|^4 \le C\left(\frac{\epsilon}{\lambda\Delta} + \Delta\right) e^{C\left(1+\frac{\epsilon}{\lambda\Delta}\right)s}. \tag{5.26}$$

□

Summarizing Lemma 5.7 and Lemma 5.8, we have proved the following estimate for the 4th moments of processes x_s and $\widetilde{x}_s$ (see [32] for stronger result about the 2nd moments):

Theorem 5.7 *Suppose that Assumption 1–3 hold. Then there exists $C > 0$, independent of ϵ and can be chosen uniformly for x_0, y_0 which are contained in some bounded domain of $\mathbb{R}^k \times \mathbb{R}^l$, such that*

$$\max_{0\le s\le T} \mathbf{E}|x_s - \widetilde{x}_s|^4 \le C\epsilon^{\frac{1}{2}}.$$

As the next step, we consider derivatives of the auxiliary processes (5.5)

$$\begin{aligned} d\hat{x}_{s,x_i} &= \big(\nabla_x f\, x_{j\Delta,x_i} + \nabla_y f\, \hat{y}_{s,x_i}\big)ds + \big(\nabla_x \alpha_1\, x_{s,x_i}\big)dw_s^1 \\ d\hat{y}_{s,x_i} &= \frac{1}{\epsilon}\big(\nabla_x g\, x_{j\Delta,x_i} + \nabla_y g\, \hat{y}_{s,x_i}\big)ds + \frac{1}{\sqrt{\epsilon}}\big(\nabla_x \alpha_2\, x_{j\Delta,x_i} + \nabla_y \alpha_2\, \hat{y}_{s,x_i}\big)dw_s^2, \end{aligned} \quad 1 \le i \le k \tag{5.27}$$

where $j = \lfloor \frac{s}{\Delta} \rfloor$ and we have assumed that Assumption 3 holds. The following lemma shows that (5.27) is an approximation of (5.2).

Lemma 5.9 *Under Assumptions 1–3, there exists $C > 0$, independent of ϵ, Δ and can be chosen uniformly for x_0, y_0 which are contained in some bounded domain of $\mathbb{R}^k \times \mathbb{R}^l$, such that*

$$\mathbf{E}|y_{s,x_i} - \hat{y}_{s,x_i}|^2 \le C\Delta, \qquad \mathbf{E}|x_{s,x_i} - \hat{x}_{s,x_i}|^2 \le C\Delta. \tag{5.28}$$

Proof Let $j = \lfloor \frac{s}{\Delta} \rfloor$. Applying Ito's formula to $|y_{s,x_i} - \hat{y}_{s,x_i}|^2$ and taking expectation, we obtain

$$\begin{aligned} &\frac{d\mathbf{E}|y_{s,x_i} - \hat{y}_{s,x_i}|^2}{ds} \\ =&\frac{2}{\epsilon}\mathbf{E}\big\langle \nabla_x g(x_s, y_s)x_{s,x_i} - \nabla_x g(x_{j\Delta}, \hat{y}_s)x_{j\Delta,x_i}, y_{s,x_i} - \hat{y}_{s,x_i}\big\rangle \\ &+ \frac{2}{\epsilon}\mathbf{E}\big\langle \nabla_y g(x_s, y_s)y_{s,x_i} - \nabla_y g(x_{j\Delta}, \hat{y}_s)\hat{y}_{s,x_i}, y_{s,x_i} - \hat{y}_{s,x_i}\big\rangle \\ &+ \frac{1}{\epsilon}\mathbf{E}\Big(\big\|\nabla_x \alpha_2(x_s, y_s)\, x_{s,x_i} + \nabla_y \alpha_2(x_s, y_s)\, y_{s,x_i} - \nabla_x \alpha_2(x_{j\Delta}, \hat{y}_s)\, x_{j\Delta,x_i} - \nabla_y \alpha_2(x_{j\Delta}, \hat{y}_s)\hat{y}_{s,x_i}\big\|^2\Big). \end{aligned}$$

We estimate each terms using Hölder and Young's inequality (5.22). For the first term,

$$\begin{aligned} &\mathbf{E}\big\langle \nabla_x g(x_s, y_s)x_{s,x_i} - \nabla_x g(x_{j\Delta}, \hat{y}_s)x_{j\Delta,x_i}, y_{s,x_i} - \hat{y}_{s,x_i}\big\rangle \\ =&\mathbf{E}\Big\langle \big(\nabla_x g(x_s, y_s) - \nabla_x g(x_{j\Delta}, \hat{y}_s)\big)x_{s,x_i} + \nabla_x g(x_{j\Delta}, \hat{y}_s)\big(x_{s,x_i} - x_{j\Delta,x_i}\big), y_{s,x_i} - \hat{y}_{s,x_i}\Big\rangle \\ \le&\frac{4}{\lambda}\mathbf{E}\big|\big(\nabla_x g(x_s, y_s) - \nabla_x g(x_{j\Delta}, \hat{y}_s)\big)x_{s,x_i}\big|^2 + \frac{4}{\lambda}\mathbf{E}\big|\nabla_x g(x_{j\Delta}, \hat{y}_s)\big(x_{s,x_i} - x_{j\Delta,x_i}\big)\big|^2 + \frac{\lambda}{4}\mathbf{E}|y_{s,x_i} - \hat{y}_{s,x_i}|^2 \\ \le&C\Big[\big(\mathbf{E}|x_{s,x_i}|^4\big)^{1/2}\big(\mathbf{E}|x_s - x_{j\Delta}|^4 + \mathbf{E}|y_s - \hat{y}_s|^4\big)^{1/2} + \mathbf{E}|x_{s,x_i} - x_{j\Delta,x_i}|^2\Big] + \frac{\lambda}{4}\mathbf{E}|y_{s,x_i} - \hat{y}_{s,x_i}|^2. \end{aligned}$$

For the second term, in a similar way, we find

$$\begin{aligned}
&\mathbf{E}\big\langle \nabla_y g(x_s,y_s) y_{s,x_i} - \nabla_y g(x_{j\Delta},\hat{y}_s)\hat{y}_{s,x_i}, y_{s,x_i}-\hat{y}_{s,x_i}\big\rangle \\
=&\mathbf{E}\big\langle \big(\nabla_y g(x_s,y_s) - \nabla_y g(x_{j\Delta},\hat{y}_s)\big) y_{s,x_i} + \nabla_y g(x_{j\Delta},\hat{y}_s)(y_{s,x_i}-\hat{y}_{s,x_i}), y_{s,x_i}-\hat{y}_{s,x_i}\big\rangle \\
\leq & C\Big[\big(\mathbf{E}|y_{s,x_i}|^4\big)^{1/2}\big(\mathbf{E}|x_s - x_{j\Delta}|^4 + \mathbf{E}|y_s-\hat{y}_s|^4\big)^{1/2}\Big] + \frac{\lambda}{4}\mathbf{E}|y_{s,x_i}-\hat{y}_{s,x_i}|^2 \\
& + \mathbf{E}\big\langle \nabla_y g(x_{j\Delta},\hat{y}_s)(y_{s,x_i}-\hat{y}_{s,x_i}), y_{s,x_i}-\hat{y}_{s,x_i}\big\rangle\,.
\end{aligned}$$

For the third term,

$$\begin{aligned}
&\mathbf{E}\Big(\big\|(\nabla_x \alpha_2(x_s,y_s)\, x_{s,x_i} + \nabla_y\alpha_2(x_s,y_s)\, y_{s,x_i} - \nabla_x\alpha_2(x_{j\Delta},\hat{y}_s)\, x_{j\Delta,x_i} - \nabla_y\alpha_2(x_{j\Delta},\hat{y}_s)\,\hat{y}_{s,x_i})\big\|^2\Big) \\
\leq& 4\mathbf{E}\Big(\big\|\big(\nabla_x\alpha_2(x_s,y_s) - \nabla_x\alpha_2(x_{j\Delta},\hat{y}_s)\big)\, x_{s,x_i}\big\|^2\Big) + 4\mathbf{E}\Big(\big\|\nabla_x\alpha_2(x_{j\Delta},\hat{y}_s)\big(x_{s,x_i}-x_{j\Delta,x_i}\big)\big\|^2\Big) \\
& + 4\mathbf{E}\Big(\big\|\big(\nabla_y\alpha_2(x_s,y_s) - \nabla_y\alpha_2(x_{j\Delta},\hat{y}_s)\big)\, y_{s,x_i}\big\|^2\Big) + 4\mathbf{E}\Big(\big\|\nabla_y\alpha_2(x_{j\Delta},\hat{y}_s)\big(y_{s,x_i}-\hat{y}_{s,x_i}\big)\big\|^2\Big) \\
\leq& C\mathbf{E}\Big|\big(|x_s - x_{j\Delta}| + |y_s-\hat{y}_s|\big)\, x_{s,x_i}\Big|^2 + C\mathbf{E}\big|x_{s,x_i} - x_{j\Delta,x_i}\big|^2 \\
& + C\mathbf{E}\Big|\big(|x_s - x_{j\Delta}| + |y_s-\hat{y}_s|\big)\, y_{s,x_i}\Big|^2 + 4\mathbf{E}\Big(\big\|\nabla_y\alpha_2(x_{j\Delta},\hat{y}_s)\big(y_{s,x_i}-\hat{y}_{s,x_i}\big)\big\|^2\Big) \\
\leq& C\Big[\Big(\big(\mathbf{E}|y_{s,x_i}|^4\big)^{1/2} + \big(\mathbf{E}|x_{s,x_i}|^4\big)^{1/2}\Big)\Big(\mathbf{E}|x_s - x_{j\Delta}|^4 + \mathbf{E}|y_s-\hat{y}_s|^4\Big)^{1/2} + \mathbf{E}|x_{s,x_i} - x_{j\Delta,x_i}|^2\Big] \\
& + 4\mathbf{E}\|\nabla_y\alpha_2(x_{j\Delta},\hat{y}_s)\,(y_{s,x_i}-\hat{y}_{s,x_i})\|^2\,.
\end{aligned}$$

Now combining the above estimates and applying Lemma 5.5, Lemma 5.6, Lemma 5.7 as well as inequality (3.11) in Assumption 2, we conclude that

$$\frac{d\mathbf{E}|y_{s,x_i}-\hat{y}_{s,x_i}|^2}{ds} \leq -\frac{\lambda}{\epsilon}\mathbf{E}|y_{s,x_i}-\hat{y}_{s,x_i}|^2 + \frac{C\Delta}{\epsilon}\,,$$

and the first part of the assertion follows from Gronwall's inequality. In the same way, we can compute that

$$\begin{aligned}
&\frac{d\mathbf{E}|x_{s,x_i}-\hat{x}_{s,x_i}|^2}{ds} \\
=&2\mathbf{E}\big\langle \nabla_x f(x_s,y_s)x_{s,x_i} - \nabla_x f(x_{j\Delta},\hat{y}_s)x_{k\Delta,x_i}, x_{s,x_i}-\hat{x}_{s,x_i}\big\rangle \\
& + 2\mathbf{E}\big\langle \nabla_y f(x_s,y_s)y_{s,x_i} - \nabla_y f(x_{j\Delta},\hat{y}_s)\hat{y}_{s,x_i}, x_{s,x_i}-\hat{x}_{s,x_i}\big\rangle \\
=&2\mathbf{E}\big\langle \big(\nabla_x f(x_s,y_s) - \nabla_x f(x_{j\Delta},\hat{y}_s)\big)x_{s,x_i}, x_{s,x_i}-\hat{x}_{s,x_i}\big\rangle + 2\mathbf{E}\big\langle \nabla_x f(x_{j\Delta},\hat{y}_s)\big(x_{s,x_i}-x_{k\Delta,x_i}\big), x_{s,x_i}-\hat{x}_{s,x_i}\big\rangle \\
& + 2\mathbf{E}\big\langle \big(\nabla_y f(x_s,y_s) - \nabla_y f(x_{j\Delta},\hat{y}_s)\big)y_{s,x_i}, x_{s,x_i}-\hat{x}_{s,x_i}\big\rangle \\
& + 2\mathbf{E}\big\langle \nabla_y f(x_{j\Delta},\hat{y}_s)\big(y_{s,x_i}-\hat{y}_{s,x_i}\big), x_{s,x_i}-\hat{x}_{s,x_i}\big\rangle \\
\leq&\mathbf{E}\big|\big(\nabla_x f(x_s,y_s) - \nabla_x f(x_{j\Delta},\hat{y}_s)\big)x_{s,x_i}\big|^2 + \mathbf{E}|x_{s,x_i}-\hat{x}_{s,x_i}|^2 + C\mathbf{E}\big|\big\langle x_{s,x_i}-x_{k\Delta,x_i}, x_{s,x_i}-\hat{x}_{s,x_i}\big\rangle\big| \\
& + \mathbf{E}\big|\big(\nabla_y f(x_s,y_s) - \nabla_y f(x_{j\Delta},\hat{y}_s)\big)y_{s,x_i}\big|^2 + \mathbf{E}|x_{s,x_i}-\hat{x}_{s,x_i}\big|^2 \\
& + C\mathbf{E}\big|\big\langle y_{s,x_i}-\hat{y}_{s,x_i}, x_{s,x_i}-\hat{x}_{s,x_i}\big\rangle\big| \\
\leq& C\Big[\mathbf{E}\big|\big(|x_s-x_{j\Delta}| + |y_s-\hat{y}_s|\big)x_{s,x_i}\big|^2 + \mathbf{E}\big|\big(|x_s-x_{j\Delta}| + |y_s-\hat{y}_s|\big)y_{s,x_i}\big|^2 + \mathbf{E}|x_{s,x_i}-x_{j\Delta,x_i}|^2 \\
& + \mathbf{E}|y_{s,x_i}-\hat{y}_{s,x_i}|^2 + \mathbf{E}|x_{s,x_i}-\hat{x}_{s,x_i}|^2\Big] \\
\leq& C\Big[\big((\mathbf{E}|y_{s,x_i}|^4)^{1/2} + (\mathbf{E}|x_{s,x_i}|^4)^{1/2}\big)\big(\mathbf{E}|x_s-x_{j\Delta}|^4 + \mathbf{E}|y_s-\hat{y}_s|^4\big)^{1/2} + \mathbf{E}|x_{s,x_i}-x_{j\Delta,x_i}|^2 + \mathbf{E}|y_{s,x_i}-\hat{y}_{s,x_i}|^2\Big] \\
& + C\mathbf{E}|x_{s,x_i}-\hat{x}_{s,x_i}|^2 \\
\leq& C\Delta + C\mathbf{E}|x_{s,x_i}-\hat{x}_{s,x_i}|^2\,,
\end{aligned}$$

where Lemma 5.5, Lemma 5.6, Lemma 5.7, as well as the first part of conclusion have been used to obtain the last inequality. Now Gronwall's inequality implies the second part of the assertion. □

We continue our study by comparing processes $\hat{x}_{s,x_i}$ with $\widetilde{x}_{s,x_i}$, where

$$d\widetilde{x}_{s,x_i} = \nabla\widetilde{f}(\widetilde{x}_s)\widetilde{x}_{s,x_i}ds + \nabla\alpha_1(\widetilde{x}_s)\widetilde{x}_{s,x_i}dw^1_s\,. \tag{5.29}$$

Recalling (3.4), we can write

$$\begin{aligned} \widetilde{f}(\widetilde{x}_s) &= \mathbf{E}^\xi\left[f(\widetilde{x}_s, \xi_t^{\widetilde{x}_s})\right],\\ \nabla\widetilde{f}(\widetilde{x}_s)\widetilde{x}_{s,x_i} &= \mathbf{E}^\xi\left[\nabla_x f(\widetilde{x}_s, \xi_t^{\widetilde{x}_s}) + \nabla_y f(\widetilde{x}_s, \xi_t^{\widetilde{x}_s})\xi_{t,x}^{\widetilde{x}_s}\right]\widetilde{x}_{s,x_i}, \end{aligned} \tag{5.30}$$

where ξ_t^x is the stationary process defined in Appendix B, $\xi_{t,x}^x$ is the derivative process of ξ_t^x with respect to x, and $\mathbf{E}^\xi$ denotes the expectation with respect to the stationary process. We have

Lemma 5.10 *Let $\Delta = \epsilon^{\frac{1}{2}}$ and Assumptions 1–3 be satisfied. Then there exists $C > 0$, independent of ϵ and can be chosen uniformly for x_0, y_0 which are contained in some bounded domain of $\mathbb{R}^k \times \mathbb{R}^l$, such that*

$$\max_{0\le s\le T}\mathbf{E}|\hat{x}_{s,x_i} - \widetilde{x}_{s,x_i}|^2 \le C\epsilon^{\frac{1}{4}}.$$

Proof Let $j = \lfloor \frac{r}{\Delta}\rfloor$. By Ito's formula and equality (5.30), we have

$$\begin{aligned} &\mathbf{E}|\hat{x}_{s,x_i} - \widetilde{x}_{s,x_i}|^2\\ =&2\int_0^s \mathbf{E}\big\langle \nabla_x f(x_{j\Delta}, \hat{y}_r)\, x_{j\Delta,x_i} + \nabla_y f(x_{j\Delta}, \hat{y}_r)\,\hat{y}_{r,x_i} - \nabla_x\widetilde{f}(\widetilde{x}_r)\widetilde{x}_{r,x_i}, \hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}\big\rangle\, dr\\ &+ \int_0^s \mathbf{E}\big\|\nabla_x\alpha_1(x_r)\, x_{r,x_i} - \nabla_x\alpha_1(\widetilde{x}_r)\widetilde{x}_{r,x_i}\big\|^2\, dr\\ =&2\int_0^s \mathbf{E}\big\langle \nabla_x f(x_{j\Delta}, \hat{y}_r)\, x_{j\Delta,x_i} - \mathbf{E}^\xi\big(\nabla_x f(\widetilde{x}_r, \xi_t^{\widetilde{x}_r})\big)\widetilde{x}_{r,x_i}, \hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}\big\rangle dr\\ &+ 2\int_0^s \mathbf{E}\big\langle \nabla_y f(x_{j\Delta}, \hat{y}_r)\,\hat{y}_{r,x_i} - \mathbf{E}^\xi\big(\nabla_y f(\widetilde{x}_r, \xi_t^{\widetilde{x}_r})\xi_{t,x}^{\widetilde{x}_r}\big)\widetilde{x}_{r,x_i}, \hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}\big\rangle dr\\ &+ \int_0^s \mathbf{E}\big\|\nabla\alpha_1(x_r)\, x_{r,x_i} - \nabla\alpha_1(\widetilde{x}_r)\widetilde{x}_{r,x_i}\big\|^2 dr\\ =&I_1 + I_2 + I_3\,. \end{aligned}$$

Using the notations in Appendix B, we can identify process $\hat{y}_r$ with $\xi^{x_{j\Delta}}_{j\Delta,r}$ and process $\hat{y}_{r,x_i}$ with $\xi^{x_{j\Delta}}_{j\Delta,r,x}x_{j\Delta,x_i}$. Then, the term I_1 on the right hand side above can be recasted as

$$\begin{aligned} &\int_0^s \mathbf{E}\big\langle \nabla_x f(x_{j\Delta}, \hat{y}_r)\, x_{j\Delta,x_i} - \mathbf{E}^\xi\big(\nabla_x f(\widetilde{x}_r, \xi_t^{\widetilde{x}_r})\big)\widetilde{x}_{r,x_i}, \hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}\big\rangle\, dr\\ =&\int_0^s \mathbf{E}\big\langle \nabla_x f(x_{j\Delta}, \xi^{x_{j\Delta}}_{j\Delta,r})x_{j\Delta,x_i} - \mathbf{E}^\xi\big(\nabla_x f(x_{j\Delta}, \xi_t^{x_{j\Delta}})\big)x_{j\Delta,x_i}, \hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}\big\rangle dr\\ &+ \int_0^s \mathbf{E}\big\langle \mathbf{E}^\xi\big(\nabla_x f(\hat{x}_r, \xi_t^{\hat{x}_r})\big)(\hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}), \hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}\big\rangle dr\\ &+ \int_0^s \mathbf{E}\big\langle \big[\mathbf{E}^\xi\big(\nabla_x f(\hat{x}_r, \xi_t^{\hat{x}_r})\big) - \mathbf{E}^\xi\big(\nabla_x f(\widetilde{x}_r, \xi_t^{\widetilde{x}_r})\big)\big]\widetilde{x}_{r,x_i}, \hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}\big\rangle dr\\ &+ \int_0^s \mathbf{E}\big\langle \mathbf{E}^\xi\big(\nabla_x f(x_{j\Delta}, \xi_t^{x_{j\Delta}})\big)x_{j\Delta,x_i} - \mathbf{E}^\xi\big(\nabla_x f(\hat{x}_r, \xi_t^{\hat{x}_r})\big)\hat{x}_{r,x_i}, \hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}\big\rangle dr\\ =&I_{1,1} + I_{1,2} + I_{1,3} + I_{1,4}\,. \end{aligned}$$

For $I_{1,1}$, using Lemma B.3 in Appendix B and Lemma 5.6, we have

$$\begin{aligned}
|I_{1,1}| \leq& \Big|\int_0^s \mathbf{E}\Big\langle \nabla_x f(x_{j\Delta},\xi^{x_{j\Delta}}_{j\Delta,r})x_{j\Delta,x_i} - \mathbf{E}^\xi\big(\nabla_x f(x_{j\Delta},\xi^{x_{j\Delta}}_t)\big)x_{j\Delta,x_i}, \hat{x}_{j\Delta,x_i} - \widetilde{x}_{j\Delta,x_i}\Big\rangle dr\Big|\\
&+ \Big|\int_0^s \mathbf{E}\Big\langle \nabla_x f(x_{j\Delta},\xi^{x_{j\Delta}}_{j\Delta,r})x_{j\Delta,x_i} - \mathbf{E}^\xi\big(\nabla_x f(x_{j\Delta},\xi^{x_{j\Delta}}_t)\big)x_{j\Delta,x_i}, \hat{x}_{r,x_i} - \hat{x}_{j\Delta,x_i}\Big\rangle dr\Big|\\
&+ \Big|\int_0^s \mathbf{E}\Big\langle \nabla_x f(x_{j\Delta},\xi^{x_{j\Delta}}_{j\Delta,r})x_{j\Delta,x_i} - \mathbf{E}^\xi\big(\nabla_x f(x_{j\Delta},\xi^{x_{j\Delta}}_t)\big)x_{j\Delta,x_i}, \widetilde{x}_{r,x_i} - \widetilde{x}_{j\Delta,x_i}\Big\rangle dr\Big|\\
\leq& C\sum_{j=0}^{\lfloor s/\Delta\rfloor}\int_{j\Delta}^{[(j+1)\Delta]\wedge s} \mathbf{E}\Big(\big(1+|x_{j\Delta}|+|\hat{y}_{j\Delta}|\big)\big|x_{j\Delta,x_i}\big|\big|\hat{x}_{j\Delta,x_i}-\widetilde{x}_{j\Delta,x_i}\big|\Big)e^{-\frac{\lambda(r-j\Delta)}{\epsilon}}\,dr + Cs\Delta^{\frac{1}{2}}\\
\leq& \frac{C\epsilon}{\lambda}\sum_{j=0}^{\lfloor s/\Delta\rfloor}\Big[\mathbf{E}\big(1+|x_{j\Delta}|+|\hat{y}_{j\Delta}|\big)^4 + \mathbf{E}\big|x_{j\Delta,x_i}\big|^4 + \mathbf{E}\big|\hat{x}_{j\Delta,x_i}-\widetilde{x}_{j\Delta,x_i}\big|^2\Big] + Cs\Delta^{\frac{1}{2}}\\
\leq& \frac{C\epsilon}{\lambda}\sum_{j=0}^{\lfloor s/\Delta\rfloor}\mathbf{E}|\hat{x}_{j\Delta,x_i}-\widetilde{x}_{j\Delta,x_i}|^2 + Cs(\Delta^{\frac{1}{2}}+\epsilon) \leq \frac{C\epsilon}{\lambda\Delta}\int_0^s \mathbf{E}|\hat{x}_{r,x_i}-\widetilde{x}_{r,x_i}|^2 dr + Cs(\Delta^{\frac{1}{2}}+\epsilon)\,,
\end{aligned}$$

where 4th order estimates in Lemma 5.2, Lemma 5.5, as well as Remark 5 are used in the last two inequalities. For $I_{1,2}$, since function f is Lipschitz, it follows that

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

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

$$\Big|\mathbf{E}^\xi\big(\nabla_x f(\hat{x}_r,\xi^{\hat{x}_r}_t)\big) - \mathbf{E}^\xi\big(\nabla_x f(\widetilde{x}_r,\xi^{\widetilde{x}_r}_t)\big)\Big| \leq C\mathbf{E}^\xi\big(|\hat{x}_r-\widetilde{x}_r| + |\xi^{\hat{x}_r}_t - \xi^{\widetilde{x}_r}_t|\big) \leq C|\hat{x}_r-\widetilde{x}_r|\,,$$

and therefore using inequality (5.22),

$$\begin{aligned}
|I_{1,3}| \leq& C\int_0^s \mathbf{E}\big(|\hat{x}_r-\widetilde{x}_r|\,|\widetilde{x}_{r,x_i}|\,|\hat{x}_{r,x_i}-\widetilde{x}_{r,x_i}|\big)\,dr\\
\leq& C\int_0^s \mathbf{E}|\hat{x}_{r,x_i}-\widetilde{x}_{r,x_i}|^2 dr + C\int_0^s \big(\mathbf{E}|\hat{x}_r-\widetilde{x}_r|^4\big)^{\frac{1}{2}}\big(\mathbf{E}|\widetilde{x}_{r,x_i}|^4\big)^{\frac{1}{2}}dr\\
\leq& C\int_0^s \mathbf{E}|\hat{x}_{r,x_i}-\widetilde{x}_{r,x_i}|^2 dr + C\int_0^s \big(\mathbf{E}|\hat{x}_r-\widetilde{x}_r|^4\big)^{\frac{1}{2}}dr\,.
\end{aligned}$$

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

$$\begin{aligned}
|I_{1,4}| \leq& C\int_0^s \mathbf{E}\Big(|x_{j\Delta,x_i}-\hat{x}_{r,x_i}||\hat{x}_{r,x_i}-\widetilde{x}_{r,x_i}|\Big)dr + C\int_0^s \mathbf{E}\Big(|x_{j\Delta}-\hat{x}_r|\,|\hat{x}_{r,x_i}|\,|\hat{x}_{r,x_i}-\widetilde{x}_{r,x_i}|\Big)dr\\
\leq& C\int_0^s \mathbf{E}|\hat{x}_{r,x_i}-\widetilde{x}_{r,x_i}|^2 dr + C\int_0^s \mathbf{E}|x_{j\Delta,x_i}-\hat{x}_{r,x_i}|^2 dr + C\int_0^s \mathbf{E}\Big(|x_{j\Delta}-\hat{x}_r|^2\,|\hat{x}_{r,x_i}|^2\Big)dr\\
\leq& C\int_0^s \mathbf{E}|\hat{x}_{r,x_i}-\widetilde{x}_{r,x_i}|^2 dr + C\int_0^s \mathbf{E}|x_{j\Delta,x_i}-\hat{x}_{r,x_i}|^2 dr + C\int_0^s \big(\mathbf{E}|x_{j\Delta}-\hat{x}_r|^4\big)^{\frac{1}{2}}dr\\
\leq& C\int_0^s \mathbf{E}|\hat{x}_{r,x_i}-\widetilde{x}_{r,x_i}|^2 dr + C\int_0^s \mathbf{E}|x_{j\Delta,x_i}-x_{r,x_i}|^2 dr + C\int_0^s \mathbf{E}|x_{r,x_i}-\hat{x}_{r,x_i}|^2 dr\\
&+ C\int_0^s \big(\mathbf{E}|x_{j\Delta}-x_r|^4\big)^{\frac{1}{2}}dr + C\int_0^s \big(\mathbf{E}|x_r-\hat{x}_r|^4\big)^{\frac{1}{2}}dr\\
\leq& C\int_0^s \mathbf{E}|\hat{x}_{r,x_i}-\widetilde{x}_{r,x_i}|^2 dr + Cs\Delta\,,
\end{aligned}$$

where the last inequality follows from Lemma 5.6, Lemma 5.7 and Lemma 5.9.

We proceed with I_2. Similarly as I_1, we have

$$\begin{aligned}
&\int_0^s \mathbf{E}\langle \nabla_y f(x_{j\Delta}, \hat{y}_r)\, \hat{y}_{r,x_i} - \mathbf{E}^\xi\big(\nabla_y f(\widetilde{x}_r, \xi_t^{\widetilde{x}_r})\xi_{t,x}^{\widetilde{x}_r}\big)\widetilde{x}_{r,x_i}, \hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}\rangle dr\\
=&\int_0^s \mathbf{E}\langle \nabla_y f(x_{j\Delta}, \xi_{j\Delta,r}^{x_{j\Delta}})\xi_{j\Delta,r,x}^{x_{j\Delta}} x_{j\Delta,x_i} - \mathbf{E}^\xi\big(\nabla_y f(x_{j\Delta}, \xi_t^{x_{j\Delta}})\xi_{t,x}^{x_{j\Delta}}\big)x_{j\Delta,x_i}, \hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}\rangle dr\\
&+\int_0^s \mathbf{E}\langle \mathbf{E}^\xi\big(\nabla_y f(\hat{x}_r, \xi_t^{\hat{x}_r})\xi_{t,x}^{\hat{x}_r}\big)(\hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}), \hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}\rangle dr\\
&+\int_0^s \mathbf{E}\langle \big[\mathbf{E}^\xi\big(\nabla_y f(\hat{x}_r, \xi_t^{\hat{x}_r})\xi_{t,x}^{\hat{x}_r}\big) - \mathbf{E}^\xi\big(\nabla_y f(\widetilde{x}_r, \xi_t^{\widetilde{x}_r})\xi_{t,x}^{\widetilde{x}_r}\big)\big]\widetilde{x}_{r,x_i}, \hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}\rangle dr\\
&+\int_0^s \mathbf{E}\langle \mathbf{E}^\xi\big(\nabla_y f(x_{j\Delta}, \xi_t^{x_{j\Delta}})\xi_{t,x}^{x_{j\Delta}}\big)x_{j\Delta,x_i} - \mathbf{E}^\xi\big(\nabla_y f(\hat{x}_r, \xi_t^{\hat{x}_r})\xi_{t,x}^{\hat{x}_r}\big)\hat{x}_{r,x_i}, \hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}\rangle dr\\
=&I_{2,1} + I_{2,2} + I_{2,3} + I_{2,4}\,.
\end{aligned} \tag{5.31}$$

Using Lemma B.3 and Lemma B.4, we can estimate the above four terms similarly as terms $I_{1,1}$ to $I_{1,4}$, and obtain

$$\begin{aligned}
I_{2,1} \le& \frac{C\epsilon}{\lambda\Delta}\int_0^s \mathbf{E}|\hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}|^2 dr + Cs(\Delta^{\frac{1}{2}} + \epsilon)\,,\\
I_{2,2} \le& C\int_0^s \mathbf{E}|\hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}|^2 dr\,,\\
I_{2,3} \le& C\int_0^s \mathbf{E}|\hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}|^2 dr + C\int_0^s \big(\mathbf{E}|\hat{x}_r - \widetilde{x}_r|^4\big)^{\frac{1}{2}} dr\,,\\
I_{2,4} \le& C\int_0^s \mathbf{E}|\hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}|^2 dr + Cs\Delta\,.
\end{aligned}$$

For I_3, Lemma 5.5, Lemma 5.7 and the assumption that α_1 is Lipschitz entail

$$\begin{aligned}
|I_3| \le& 3\int_0^s \mathbf{E}\|(\nabla_x\alpha_1(x_r) - \nabla_x\alpha_1(\widetilde{x}_r))x_{r,x_i}\|^2 dr + 3\int_0^s \mathbf{E}\|\nabla_x\alpha_1(\widetilde{x}_r)(x_{r,x_i} - \hat{x}_{r,x_i})\|^2 dr\\
&+ 3\int_0^s \mathbf{E}\|\nabla_x\alpha_1(\widetilde{x}_r)(\hat{x}_{r,x_i} - \widetilde{x}_{r,x_i})\|^2 dr\\
\le& C\int_0^s \mathbf{E}\big(|x_r - \widetilde{x}_r|^2|x_{r,x_i}|^2\big) dr + C\int_0^s \mathbf{E}|x_{r,x_i} - \hat{x}_{r,x_i}|^2 dr + C\int_0^s \mathbf{E}|\hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}|^2 dr\\
\le& C\int_0^s \mathbf{E}|\hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}|^2 dr + C\int_0^s \big(\mathbf{E}|x_r - \widetilde{x}_r|^4\big)^{\frac{1}{2}}\big(\mathbf{E}|x_{r,x_i}|^4\big)^{\frac{1}{2}} dr + Cs\Delta\\
\le& C\int_0^s \mathbf{E}|\hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}|^2 dr + C\int_0^s \big(\mathbf{E}|\hat{x}_r - \widetilde{x}_r|^4\big)^{\frac{1}{2}} dr + Cs\Delta\,.
\end{aligned}$$

Upon combining the bounds for I_1, I_2 and I_3, we conclude that

$$\begin{aligned}
\mathbf{E}|\hat{x}_{s,x_i} - \widetilde{x}_{s,x_i}|^2 \le& C(1 + \frac{\epsilon}{\lambda\Delta})\int_0^s \mathbf{E}|\hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}|^2 dr\\
&+ C\int_0^s \big(\mathbf{E}|\hat{x}_r - \widetilde{x}_r|^4\big)^{\frac{1}{2}} dr + Cs(\Delta + \Delta^{\frac{1}{2}} + \epsilon)\,.
\end{aligned}$$

Now letting $\Delta = \epsilon^{\frac{1}{2}}$ and using Lemma 5.8, it follows that

$$\mathbf{E}|\hat{x}_{s,x_i} - \widetilde{x}_{s,x_i}|^2 \le C\int_0^s \mathbf{E}|\hat{x}_{r,x_i} - \widetilde{x}_{r,x_i}|^2 dr + Cs\epsilon^{\frac{1}{4}}$$

and applying Gronwall's inequality yields the conclusion. □

Combining Lemma 5.9 and Lemma 5.10, we have proved:

Theorem 5.8 *Suppose that Assumptions 1–3 hold. Then there exists $C > 0$, independent of ϵ and can be chosen uniformly for x_0, y_0 which are contained in some bounded domain of $\mathbb{R}^k \times \mathbb{R}^l$, such that*

$$\mathbf{E}|x_{s,x_i} - \widetilde{x}_{s,x_i}|^2 \leq C\epsilon^{\frac{1}{4}}\,, \quad s \in [0,T]\,.$$

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, 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 the resulting estimators remain efficient. In the case of exponential type expectations and for multiscale diffusions with both slow and fast variables, the asymptotic optimality of the approximation based on a low-dimensional averaged equation has been proved and an upper bound for the relative error of the importance sampling estimator is obtained. We expect our results to be helpful for the design of importance sampling methods as well as for the study of multiscale diffusion processes.

There are many possible extensions related to the current work. For the theoretical aspects, our main result concerns the time scale separation limit ($\epsilon \to 0$) for diffusion with slow and fast variables and assumes the temperature β is fixed. As a result, the constant in Theorem 3.1 may depend on β. It is interesting to consider asymptotics for both parameters ϵ, β together. Generalizing our results to dynamics with non-Lipschitz coefficients as well as to more general types of dynamics is also important. For the numerical aspects, realistic systems in climate science, molecular dynamics may be high-dimensional and even the averaged equation cannot be easily discretized and solved by usual grid-based methods. In more general situations, it may be impossible to separate systems' states into slow and fast ones with an explicit time scale separation parameter. We leave these questions for future work and refer to [46, 23] for some recent algorithmic and methodological 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-CRC 1114 "Scaling Cascades in Complex Systems". Special thanks also go to anonymous referees whose valuable comments and criticism have helped to improve this paper.

A Two useful inequalities

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

$$\begin{aligned}\dot{x}_1(t) &\le a_{11}\,x_1(t) + a_{12}\,x_2(t)\\ \dot{x}_2(t) &\le \frac{a_{21}}{\epsilon} x_1(t) - \frac{a_{22}}{\epsilon} x_2(t)\end{aligned}$$

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

$$\max_{0\le s\le T} x_1(s) \le C\epsilon, \qquad x_2(t) \le e^{-\frac{a_{22}t}{\epsilon}} + C\epsilon\,, \quad t\in[0,T]. \tag{A.1}$$

Proof Applying Gronwall's inequality to the equation of x_2, we have

$$\begin{aligned}x_2(t) &\le e^{-\frac{a_{22}t}{\epsilon}} + \int_0^t e^{-\frac{a_{22}}{\epsilon}(t-s)} \frac{a_{21}}{\epsilon} x_1(s)ds\\ &\le e^{-\frac{a_{22}t}{\epsilon}} + \frac{a_{21}}{a_{22}} \max_{0\le s\le t} x_1(s)\,.\end{aligned} \tag{A.2}$$

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

$$x_1(t) \le a_{12} \int_0^t e^{a_{11}(t-s)} \Big[e^{-\frac{a_{22}s}{\epsilon}} + \frac{a_{21}}{a_{22}} \max_{0\le r\le s} x_1(r)\Big] ds\,. \tag{A.3}$$

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

$$\begin{aligned}\max_{0\le s\le t} x_1(s) &\le a_{12} \int_0^t e^{a_{11}(t-s)} \Big[e^{-\frac{a_{22}s}{\epsilon}} + \frac{a_{21}}{a_{22}} \max_{0\le r\le s} x_1(r)\Big] ds\\ &\le \frac{a_{12}}{a_{22}} e^{a_{11}T}\epsilon + \frac{a_{12}a_{21}}{a_{22}} \int_0^t e^{a_{11}(t-s)} \max_{0\le r\le s} x_1(r) ds\,.\end{aligned} \tag{A.4}$$

The first part of the assertion then follows by using Gronwall's inequality in integral form to $\max\limits_{0\le s\le t} x_1(s)$, while the second part is obtained using (A.2). □

For $0 < \epsilon < 1$, we set $t_1 = -\frac{2\epsilon\ln\epsilon}{\lambda} > 0$ and introduce the function η: $[0,T] \to [0,1]$ by

$$\eta(t) = \begin{cases} 1 - \frac{t}{t_1} & 0 \le t \le t_1\\ 0 & t_1 < t \le T\,.\end{cases} \tag{A.5}$$

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

$$\begin{aligned}\dot{x}_1(t) &\le a_1(1 + \epsilon^{-\eta(t)})x_1(t) + a_2\epsilon^{\eta(t)}x_2(t)\\ \dot{x}_2(t) &\le \frac{a_3x_1(t)}{\epsilon} - \frac{\lambda x_2(t)}{\epsilon}\,,\end{aligned}$$

where η is given in (A.5), $a_i \ge 0, 1 \le i \le 3$, and $x_1(0) = 0, x_2(0) = 1$. Further assume that $x_1(t) \ge 0$ on $t \in [0,T]$. Then there is a constant $C > 0$ independent of ϵ, such that

$$\max_{0\le s\le T} x_1(s) \le C\epsilon^2, \qquad x_2(t) \le e^{-\frac{\lambda t}{\epsilon}} + C\epsilon^2\,, \quad t\in[0,T]\,. \tag{A.6}$$

Proof As in Claim A.1, we can obtain

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

$$\max_{0\le s\le t} x_1(s) \le a_2 \int_0^t e^{a_1\int_s^t (1+\epsilon^{-\eta(r)})dr} \epsilon^{\eta(s)} \big[e^{-\frac{\lambda s}{\epsilon}} + \frac{a_3}{\lambda} \max_{0\le r\le s} x_1(r)\big] ds\,. \tag{A.8}$$

Then, for $t < t_1$, the second inequality above implies

$$\max_{0\le s\le t} x_1(s) \le C\epsilon^2 + \frac{a_2a_3}{\lambda} \int_0^t e^{a_1\int_s^t (1+\epsilon^{-\eta(r)})dr} \epsilon^{\eta(s)} \max_{0\le r\le s} x_1(r) ds\,. \tag{A.9}$$

Using (A.7) and Gronwall's inequality again, we conclude that

$$\max_{0\le s\le t_1} x_1(s) \le C\epsilon^2, \quad x_2(t) \le e^{-\frac{\lambda t}{\epsilon}} + C\epsilon^2, \qquad t \le t_1\,. \tag{A.10}$$

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

$$\max_{t_1\le s\le T} x_1(s) \le C\epsilon^2, \quad x_2(t) \le C\epsilon^2, \qquad t \in [t_1, T]\,. \tag{A.11}$$

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

$$d\xi^x_{\tau,s} = \frac{1}{\epsilon} g(x, \xi^x_{\tau,s})ds + \frac{1}{\sqrt{\epsilon}}\alpha_2(x, \xi^x_{\tau,s})dw_s\,, \quad s \ge \tau\,, \quad \xi^x_{\tau,\tau} = y \tag{B.1}$$

where w_s is a standard Wiener process in $\mathbb{R}^{m_1}$. In the following, we summarize some properties related to the above process that we called *the fast subsystem* in Section 3. See also [32, 10] for additional results.

Lemma B.1 *Under Assumptions 1–2, there exists a constant $C > 0$, independent of ϵ, x, y, such that:*

1. $\mathbf{E}|\xi^x_{\tau,s}|^4 \le e^{-\frac{\lambda(s-\tau)}{\epsilon}}|y|^4 + C\big(|x|^4 + 1\big)$.
2. *For $\tau_1 \le \tau_2$, it holds*
$$\mathbf{E}|\xi^x_{\tau_2,s} - \xi^x_{\tau_1,s}|^4 \le C\big(1 + |x|^4 + |y|^4\big)\, e^{-\frac{4\lambda(s-\tau_2)}{\epsilon}}\,.$$
3. *For $x, x' \in \mathbb{R}^k$ and $\tau_1 \le \tau_2$,*
$$\mathbf{E}|\xi^{x'}_{\tau_2,s} - \xi^x_{\tau_1,s}|^4 \le e^{-\frac{2\lambda(s-\tau_2)}{\epsilon}}\big(|y|^4 + |x|^4 + 1\big) + C|x' - x|^4\,.$$

Proof 1. By Ito's formula, we have

$$\begin{aligned}\frac{d\mathbf{E}|\xi^x_{\tau,s}|^4}{ds} =& \frac{1}{\epsilon}\mathbf{E}\Big[|\xi^x_{\tau,s}|^2\big(4\langle g(x,\xi^x_{\tau,s}), \xi^x_{\tau,s}\rangle + 2\|\alpha_2(x,\xi^x_{\tau,s})\|^2\big) + 4|\alpha_2^T(x,\xi^x_{\tau,s})\xi^x_{\tau,s}|^2\Big]\\ \le& \frac{1}{\epsilon}\mathbf{E}\Big[|\xi^x_{\tau,s}|^2\big(4\langle g(x,\xi^x_{\tau,s}), \xi^x_{\tau,s}\rangle + 6\|\alpha_2(x,\xi^x_{\tau,s})\|^2\big)\Big]\,.\end{aligned}$$

Applying inequality (3.13) in Remark 2 and inequality (5.22), we obtain

$$\begin{aligned}\frac{d\mathbf{E}|\xi^x_{\tau,s}|^4}{ds} \le& -\frac{2\lambda}{\epsilon}\mathbf{E}|\xi^x_{\tau,s}|^4 + \frac{C}{\epsilon}\mathbf{E}\Big[|\xi^x_{\tau,s}|^2(|x|^2+1)\Big]\\ \le& -\frac{\lambda}{\epsilon}\mathbf{E}|\xi^x_{\tau,s}|^4 + \frac{C}{\epsilon}\Big(|x|^4+1\Big)\,,\end{aligned}$$

and the first statement follows from Gronwall's inequality.

2. For the second statement, using Ito's formula and Assumption 2, it follows

$$\begin{aligned}&\frac{d\mathbf{E}|\xi^x_{\tau_2,s} - \xi^x_{\tau_1,s}|^4}{ds}\\ =&\frac{1}{\epsilon}\mathbf{E}\Big[|\xi^x_{\tau_2,s} - \xi^x_{\tau_1,s}|^2\big(4\langle g(x,\xi^x_{\tau_2,s}) - g(x,\xi^x_{\tau_1,s}), \xi^x_{\tau_2,s} - \xi^x_{\tau_1,s}\rangle\\ &+ 2\|\alpha_2(x,\xi^x_{\tau_2,s}) - \alpha_2(x,\xi^x_{\tau_1,s})\|^2\big) + 4\Big|\big(\alpha_2(x,\xi^x_{\tau_2,s}) - \alpha_2(x,\xi^x_{\tau_1,s})\big)^T\big(\xi^x_{\tau_2,s} - \xi^x_{\tau_1,s}\big)\Big|^2\Big]\\ \le&\frac{1}{\epsilon}\mathbf{E}\Big[|\xi^x_{\tau_2,s} - \xi^x_{\tau_1,s}|^2\big(4\langle g(x,\xi^x_{\tau_2,s}) - g(x,\xi^x_{\tau_1,s}), \xi^x_{\tau_2,s} - \xi^x_{\tau_1,s}\rangle + 6\|\alpha_2(x,\xi^x_{\tau_2,s}) - \alpha_2(x,\xi^x_{\tau_1,s})\|^2\big)\Big]\\ \le& -\frac{4\lambda}{\epsilon}\mathbf{E}|\xi^x_{\tau_2,s} - \xi^x_{\tau_1,s}|^4\,.\end{aligned}$$

Therefore, integrating and using the first statement above, we obtain

$$\mathbf{E}|\xi^x_{\tau_2,s} - \xi^x_{\tau_1,s}|^4 \le e^{-\frac{4\lambda(s-\tau_2)}{\epsilon}}\mathbf{E}|\xi^x_{\tau_1,\tau_2} - y|^4 \le C\big(1 + |x|^4 + |y|^4\big)e^{-\frac{4\lambda(s-\tau_2)}{\epsilon}}\,.$$

3. For the third statement, in a similar way, applying Ito's formula, using Assumption 2, as well as Lipschitz property of functions g and α_2, we have

$$\begin{aligned}
&\frac{d\mathbf{E}|\xi^{x'}_{\tau_2,s}-\xi^{x}_{\tau_1,s}|^4}{ds}\\
=&\frac{1}{\epsilon}\mathbf{E}\Big[|\xi^{x'}_{\tau_2,s}-\xi^{x}_{\tau_1,s}|^2\big(4\langle g(x',\xi^{x'}_{\tau_2,s})-g(x,\xi^{x}_{\tau_1,s}),\xi^{x'}_{\tau_2,s}-\xi^{x}_{\tau_1,s}\rangle\\
&+2\|\alpha_2(x',\xi^{x'}_{\tau_2,s})-\alpha_2(x,\xi^{x}_{\tau_1,s})\|^2\big)+4\big|\big(\alpha_2(x',\xi^{x'}_{\tau_2,s})-\alpha_2(x,\xi^{x}_{\tau_1,s})\big)^T\big(\xi^{x'}_{\tau_2,s}-\xi^{x}_{\tau_1,s}\big)\big|^2\Big]\\
\le&\frac{1}{\epsilon}\mathbf{E}\Big[|\xi^{x'}_{\tau_2,s}-\xi^{x}_{\tau_1,s}|^2\big(4\langle g(x',\xi^{x'}_{\tau_2,s})-g(x,\xi^{x}_{\tau_1,s}),\xi^{x'}_{\tau_2,s}-\xi^{x}_{\tau_1,s}\rangle+6\|\alpha_2(x',\xi^{x'}_{\tau_2,s})-\alpha_2(x,\xi^{x}_{\tau_1,s})\|^2\big)\Big]\\
\le&\frac{1}{\epsilon}\mathbf{E}\Big[|\xi^{x'}_{\tau_2,s}-\xi^{x}_{\tau_1,s}|^2\big(4\langle g(x',\xi^{x'}_{\tau_2,s})-g(x',\xi^{x}_{\tau_1,s}),\xi^{x'}_{\tau_2,s}-\xi^{x}_{\tau_1,s}\rangle+12\|\alpha_2(x',\xi^{x'}_{\tau_2,s})-\alpha_2(x',\xi^{x}_{\tau_1,s})\|^2\big)\Big]\\
&+\frac{1}{\epsilon}\mathbf{E}\Big[|\xi^{x'}_{\tau_2,s}-\xi^{x}_{\tau_1,s}|^2\big(4\langle g(x',\xi^{x}_{\tau_1,s})-g(x,\xi^{x}_{\tau_1,s}),\xi^{x'}_{\tau_2,s}-\xi^{x}_{\tau_1,s}\rangle+12\|\alpha_2(x',\xi^{x}_{\tau_1,s})-\alpha_2(x,\xi^{x}_{\tau_1,s})\|^2\big)\Big]\\
\le&-\frac{4\lambda}{\epsilon}\mathbf{E}|\xi^{x'}_{\tau_2,s}-\xi^{x}_{\tau_1,s}|^4+\frac{C}{\epsilon}\mathbf{E}\big(|\xi^{x'}_{\tau_2,s}-\xi^{x}_{\tau_1,s}|^3|x'-x|\big)+\frac{C}{\epsilon}\mathbf{E}\big(|\xi^{x'}_{\tau_2,s}-\xi^{x}_{\tau_1,s}|^2|x'-x|^2\big)\\
\le&-\frac{2\lambda}{\epsilon}\mathbf{E}|\xi^{x'}_{\tau_2,s}-\xi^{x}_{\tau_1,s}|^4+\frac{C}{\epsilon}|x'-x|^4\,,
\end{aligned}$$

where inequality (5.22) is used to obtain the last inequality. Gronwall's inequality together with the first statement above then yield the assertion.

□

Now consider the derivative process

$$d\xi^x_{\tau,s,x_i}=\frac{1}{\epsilon}\Big(D_{x_i}g(x,\xi^x_{\tau,s})+\nabla_y g(x,\xi^x_{\tau,s})\xi^x_{\tau,s,x_i}\Big)ds+\frac{1}{\sqrt{\epsilon}}\Big(D_{x_i}\alpha_2(x,\xi^x_{\tau,s})+\nabla_y\alpha_2(x,\xi^x_{\tau,s})\xi^x_{\tau,s,x_i}\Big)dw_s\,,$$

with $s\ge\tau\,,\xi^x_{\tau,\tau,x_i}=0$, $1\le i\le k$. We summarize its properties in the following result.

Lemma B.2 *Under Assumptions 1–2, there exists a constant $C>0$, independent of ϵ,x,y, such that*

1. *For $x\in\mathbb{R}^k$, $s\ge\tau$, $\mathbf{E}|\xi^x_{\tau,s,x_i}|^4\le C$.*
2. *For $\tau_1\le\tau_2$, $x\in\mathbb{R}^k$,*

$$\mathbf{E}|\xi^x_{\tau_2,s,x_i}-\xi^x_{\tau_1,s,x_i}|^2\le C\big(1+|x|^2+|y|^2\big)e^{-\frac{\lambda(s-\tau_2)}{\epsilon}}\,.$$

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

$$\mathbf{E}|\xi^{x'}_{\tau_2,s,x_i}-\xi^x_{\tau_1,s,x_i}|^2\le Ce^{-\frac{\lambda(s-\tau_2)}{\epsilon}}\Big[1+\frac{s-\tau_2}{\epsilon}\big(1+|x|^2+|y|^2\big)\Big]+C|x-x'|^2\,.$$

Proof 1. Using Ito's formula, Assumption 1 (Lipschitz continuity of functions g and α_2), inequality (3.11) in Remark 2, as well as inequality (5.22), we see that

$$\begin{aligned}
&\frac{d\mathbf{E}|\xi^x_{\tau,s,x_i}|^4}{ds}\\
\le&\frac{1}{\epsilon}\mathbf{E}\Big[|\xi^x_{\tau,s,x_i}|^2\Big(4\langle D_{x_i}g(x,\xi^x_{\tau,s})+\nabla_y g(x,\xi^x_{\tau,s})\xi^x_{\tau,s,x_i},\xi^x_{\tau,s,x_i}\rangle+6\|D_{x_i}\alpha_2(x,\xi^x_{\tau,s})+\nabla_y\alpha_2(x,\xi^x_{\tau,s})\xi^x_{\tau,s,x_i}\|^2\Big)\Big]\\
\le&\frac{1}{\epsilon}\mathbf{E}\Big[|\xi^x_{\tau,s,x_i}|^2\Big(C|\xi^x_{\tau,s,x_i}|+4\langle\nabla_y g(x,\xi^x_{\tau,s})\xi^x_{\tau,s,x_i},\xi^x_{\tau,s,x_i}\rangle+C+12\|\nabla_y\alpha_2(x,\xi^x_{\tau,s})\xi^x_{\tau,s,x_i}\|^2\Big)\Big]\\
\le&-\frac{2\lambda}{\epsilon}\mathbf{E}|\xi^x_{\tau,s,x_i}|^4+\frac{C}{\epsilon}
\end{aligned}$$

and therefore $\mathbf{E}|\xi^x_{\tau,s,x_i}|^4\le C$ by Gronwall's inequality.

2. Now consider $\xi^x_{\tau_1,s,x_i}, \xi^x_{\tau_2,s,x_i}$ with $\tau_1 \le \tau_2$. Using Lipschitz condition of functions g, α_2, inequality (3.11) in Remark 2, as well as inequality (5.22), it follows

$$
\begin{aligned}
&\frac{d\mathbf{E}|\xi^x_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}|^2}{ds} \\
=&\frac{2}{\epsilon}\mathbf{E}\langle D_{x_i} g(x,\xi^x_{\tau_2,s}) - D_{x_i} g(x,\xi^x_{\tau_1,s}) + \nabla_y g(x,\xi^x_{\tau_2,s})\xi^x_{\tau_2,s,x_i} - \nabla_y g(x,\xi^x_{\tau_1,s})\xi^x_{\tau_1,s,x_i}, \xi^x_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}\rangle \\
&+ \frac{1}{\epsilon}\mathbf{E}\| D_{x_i}\alpha_2(x,\xi^x_{\tau_2,s}) - D_{x_i}\alpha_2(x,\xi^x_{\tau_1,s}) + \nabla_y\alpha_2(x,\xi^x_{\tau_2,s})\xi^x_{\tau_2,s,x_i} - \nabla_y\alpha_2(x,\xi^x_{\tau_1,s})\xi^x_{\tau_1,s,x_i}\|^2 \\
\le&\frac{C}{\epsilon}\mathbf{E}\Big(|\xi^x_{\tau_2,s} - \xi^x_{\tau_1,s}||\xi^x_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}|\Big) + \frac{2}{\epsilon}\mathbf{E}\langle\big(\nabla_y g(x,\xi^x_{\tau_2,s}) - \nabla_y g(x,\xi^x_{\tau_1,s})\big)\xi^x_{\tau_1,s,x_i}, \xi^x_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}\rangle \\
&+ \frac{2}{\epsilon}\mathbf{E}\langle\nabla_y g(x,\xi^x_{\tau_2,s})(\xi^x_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}), \xi^x_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}\rangle + \frac{C}{\epsilon}\mathbf{E}|\xi^x_{\tau_2,s} - \xi^x_{\tau_1,s}|^2 \\
&+ \frac{3}{\epsilon}\mathbf{E}\|\big(\nabla_y\alpha_2(x,\xi^x_{\tau_2,s}) - \nabla_y\alpha_2(x,\xi^x_{\tau_1,s})\big)\xi^x_{\tau_1,s,x_i}\|^2 + \frac{3}{\epsilon}\mathbf{E}\|\nabla_y\alpha_2(x,\xi^x_{\tau_2,s})(\xi^x_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i})\|^2 \\
\le& -\frac{\lambda}{\epsilon}\mathbf{E}|\xi^x_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}|^2 + \frac{C}{\epsilon}\big(\mathbf{E}|\xi^x_{\tau_2,s} - \xi^x_{\tau_1,s}|^4\big)^{\frac{1}{2}}\big(\mathbf{E}|\xi^x_{\tau_1,s,x_i}|^4\big)^{\frac{1}{2}} + \frac{C}{\epsilon}\mathbf{E}|\xi^x_{\tau_2,s} - \xi^x_{\tau_1,s}|^2 \\
\le& -\frac{\lambda}{\epsilon}\mathbf{E}|\xi^x_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}|^2 + \frac{C}{\epsilon}(1 + |x|^2 + |y|^2)e^{-\frac{2\lambda(s-\tau_2)}{\epsilon}},
\end{aligned}
$$

where the first assertion above and Lemma B.1 have been used in the last inequality. Then Gronwall's inequality entails

$$
\mathbf{E}|\xi^x_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}|^2 \le C\big(1 + |x|^2 + |y|^2\big)e^{-\frac{\lambda(s-\tau_2)}{\epsilon}}.
$$

3. Consider $\xi^x_{\tau_1,s,x_i}, \xi^{x'}_{\tau_2,s,x_i}$ with $\tau_1 \le \tau_2$. In a similar way, we have

$$
\begin{aligned}
&\frac{d\mathbf{E}|\xi^{x'}_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}|^2}{ds} \\
=&\frac{2}{\epsilon}\mathbf{E}\langle D_{x_i} g(x',\xi^{x'}_{\tau_2,s}) - D_{x_i} g(x,\xi^x_{\tau_1,s}) + \nabla_y g(x',\xi^{x'}_{\tau_2,s})\xi^{x'}_{\tau_2,s,x_i} - \nabla_y g(x,\xi^x_{\tau_1,s})\xi^x_{\tau_1,s,x_i}, \xi^{x'}_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}\rangle \\
&+ \frac{1}{\epsilon}\mathbf{E}\| D_{x_i}\alpha_2(x',\xi^{x'}_{\tau_2,s}) - D_{x_i}\alpha_2(x,\xi^x_{\tau_1,s}) + \nabla_y\alpha_2(x',\xi^{x'}_{\tau_2,s})\xi^{x'}_{\tau_2,s,x_i} - \nabla_y\alpha_2(x,\xi^x_{\tau_1,s})\xi^x_{\tau_1,s,x_i}\|^2 \\
\le&\frac{2}{\epsilon}\mathbf{E}\langle D_{x_i} g(x',\xi^{x'}_{\tau_2,s}) - D_{x_i} g(x',\xi^x_{\tau_1,s}) + \nabla_y g(x',\xi^{x'}_{\tau_2,s})(\xi^{x'}_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}), \xi^{x'}_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}\rangle \\
&+ \frac{2}{\epsilon}\mathbf{E}\langle D_{x_i} g(x',\xi^x_{\tau_1,s}) - D_{x_i} g(x,\xi^x_{\tau_1,s}) + \big(\nabla_y g(x',\xi^{x'}_{\tau_2,s}) - \nabla_y g(x,\xi^x_{\tau_1,s})\big)\xi^x_{\tau_1,s,x_i}, \xi^{x'}_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}\rangle \\
&+ \frac{3}{\epsilon}\mathbf{E}\| D_{x_i}\alpha_2(x',\xi^{x'}_{\tau_2,s}) - D_{x_i}\alpha_2(x,\xi^x_{\tau_1,s})\|^2 + \frac{3}{\epsilon}\mathbf{E}\|\nabla_y\alpha_2(x',\xi^{x'}_{\tau_2,s})(\xi^{x'}_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i})\|^2 \\
&+ \frac{3}{\epsilon}\mathbf{E}\|\big(\nabla_y\alpha_2(x',\xi^{x'}_{\tau_2,s}) - \nabla_y\alpha_2(x,\xi^x_{\tau_1,s})\big)\xi^x_{\tau_1,s,x_i}\|^2 \\
\le& -\frac{2\lambda}{\epsilon}\mathbf{E}|\xi^{x'}_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}|^2 + \frac{C}{\epsilon}\mathbf{E}\big(|\xi^{x'}_{\tau_2,s} - \xi^x_{\tau_1,s}||\xi^{x'}_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}|\big) + \frac{C}{\epsilon}\mathbf{E}\big(|x'-x||\xi^{x'}_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}|\big) \\
&+ \frac{C}{\epsilon}\mathbf{E}\big[\big(|x'-x| + |\xi^{x'}_{\tau_2,s} - \xi^x_{\tau_1,s}|\big)|\xi^x_{\tau_1,s,x_i}||\xi^{x'}_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}|\big] + \frac{C}{\epsilon}|x-x'|^2 + \frac{C}{\epsilon}\mathbf{E}|\xi^{x'}_{\tau_2,s} - \xi^x_{\tau_1,s}|^2 \\
&+ \frac{C}{\epsilon}\mathbf{E}\big[\big(|x'-x| + |\xi^{x'}_{\tau_2,s} - \xi^x_{\tau_1,s}|\big)|\xi^x_{\tau_1,s,x_i}|^2\big] \\
\le& -\frac{\lambda}{\epsilon}\mathbf{E}|\xi^{x'}_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}|^2 + \frac{C}{\epsilon}\Big(|x'-x|^2 + \mathbf{E}|\xi^{x'}_{\tau_2,s} - \xi^x_{\tau_1,s}|^2 + \big(\mathbf{E}|\xi^{x'}_{\tau_2,s} - \xi^x_{\tau_1,s}|^4\big)^{\frac{1}{2}}\Big) \\
\le& -\frac{\lambda}{\epsilon}\mathbf{E}|\xi^{x'}_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}|^2 + \frac{C}{\epsilon}\Big[(1 + |x|^2 + |y|^2)e^{-\frac{\lambda(s-\tau_2)}{\epsilon}} + |x'-x|^2\Big],
\end{aligned}
$$

and thus

$$
\mathbf{E}|\xi^{x'}_{\tau_2,s,x_i} - \xi^x_{\tau_1,s,x_i}|^2 \le Ce^{-\frac{\lambda(s-\tau_2)}{\epsilon}}\big[1 + \frac{s-\tau_2}{\epsilon}(1 + |x|^2 + |y|^2)\big] + C|x'-x|^2.
$$

□

The above results allow us to define the stationary process $\xi^x_s = \xi^x_{-\infty,s}$ with $\xi^x_s \sim \rho_x(y)\,dy$ where ρ_x is the stationary probability density with respect to Lebesgue measure, and also the derivative process ξ^x_{s,x_i} for $1 \le i \le k$, satisfying that $\forall f \in C^1_b(\mathbb{R}^k \times \mathbb{R}^l)$ and $\widetilde{f}(x) = \mathbf{E}(f(x,\xi^x_s)) = \int_{\mathbb{R}^l} f(x,y)\rho_x(y)dy$, it holds

$$
D_{x_i}\widetilde{f}(x) = \mathbf{E}\big(D_{x_i} f(x,\xi^x_s) + \nabla_y f(x,\xi^x_s)\xi^x_{s,x_i}\big). \tag{B.2}
$$

Processes ξ^x_s and ξ^x_{s,x_i} have the following properties:

Lemma B.3 *Under Assumptions 1 and 2, there is constant $C > 0$, independent of ϵ, x and y, such that $\forall f \in C_b^1(\mathbb{R}^l)$:*

1.

$$\Big|\mathbf{E}f(\xi^x_{0,s}) - \int_{\mathbb{R}^l} f(y)\rho_x(y)dy\Big| \le \sup|f'|\Big(|x|+|y|+1\Big)e^{-\frac{\lambda s}{\epsilon}}\,. \tag{B.3}$$

2.

$$\Big|\mathbf{E}\Big(f(\xi^x_{0,s})\xi^x_{0,s,x_i}\Big) - \mathbf{E}\Big(f(\xi^x_s)\xi^x_{s,x_i}\Big)\Big| \le C\Big(\sup|f| + \sup|f'|\Big)\Big(1+|x|+|y|\Big)e^{-\frac{\lambda s}{2\epsilon}}\,. \tag{B.4}$$

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

$$\begin{aligned}
&\Big|\mathbf{E}\big(f(\xi^x_{0,s})\xi^x_{0,s,x_i}\big) - \mathbf{E}\big(f(\xi^x_s)\xi^x_{s,x_i}\big)\Big| \\
\le&\Big|\mathbf{E}\big[f(\xi^x_s)(\xi^x_{0,s,x_i} - \xi^x_{s,x_i})\big]\Big| + \Big|\mathbf{E}\big[(f(\xi^x_{0,s}) - f(\xi^x_s))\xi^x_{0,s,x_i}\big]\Big| \\
\le&C\big(\sup|f| + \sup|f'|\big)\big(1+|x|+|y|\big)e^{-\frac{\lambda s}{2\epsilon}}
\end{aligned}$$

□

An analogous property for the stationary process ξ^x_s is the following:

Lemma B.4 *Under Assumption 1 and 2, there exists constant $C > 0$, independent of x, x', such that*

1. *For $x \in \mathbb{R}^k$, $\mathbf{E}|\xi^x_{s,x_i}|^4 \le C$.*
2. *For $x, x' \in \mathbb{R}^k$, $\mathbf{E}|\xi^{x'}_s - \xi^x_s|^4 \le C|x-x'|^4$.*
3. *For $x, x' \in \mathbb{R}^k$, $\mathbf{E}|\xi^{x'}_{s,x_i} - \xi^x_{s,x_i}|^2 \le C|x-x'|^2$.*

Proof The conclusions follow directly by letting $\tau_1, \tau_2 \to -\infty$ in Lemma B.1 and Lemma B.2.

References

1. S. Asmussen and P. W. Glynn, *Stochastic Simulation: Algorithms and Analysis*, Springer, 2007.
2. S. Asmussen and D. P. Kroese, *Improved algorithms for rare event simulation with heavy tails*, Adv. Appl. Prob., 38 (2006), pp. 545–558.
3. A. Bensoussan, J.-L. Lions, and G. Papanicolaou, *Asymptotic analysis for periodic structures*, Studies in mathematics and its applications, North-Holland, 1978.
4. N. BERGLUND and B. GENTZ, *Metastability in simple climate models: Pathwise analysis of slowly driven langevin equations*, Stoch. and Dyn., 02 (2002), pp. 327–356.
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. M. Boué and P. Dupuis, *A variational representation for certain functionals of Brownian motion*, Ann. Probab., 26 (1998), pp. 1641–1659.
7. S. P. Brooks, *Markov chain Monte Carlo method and its application*, J. R. Stat. Soc. Series D (The Statistician), 47 (1998), pp. 69–100.
8. 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. ———, *Averaging principle for systems of reaction-diffusion equations with polynomial nonlinearities perturbed by multiplicative noise*, Siam J. Math. Anal., 43 (2011), pp. 2482–2518.
10. G. Da Prato and J. Zabczyk, *Ergodicity for Infinite Dimensional Systems*, Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology, Cambridge University Press, 1996.
11. ———, *Second Order Partial Differential Equations in Hilbert Spaces*, London Mathematical Society Lecture Note Series, Cambridge University Press, 2002.
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, *Rare event simulation for rough energy landscapes*, in Simulation Conference (WSC), Proceedings of the 2011 Winter, Dec 2011, pp. 504–515.
15. P. Dupuis, K. Spiliopoulos, and H. Wang, *Importance sampling for multiscale diffusions*, Multiscale Model. Simul., 10 (2012), pp. 1–27.
16. P. Dupuis and H. Wang, *Importance sampling, large deviations, and differential games*, Stochastics and Stochastic Rep., 76 (2004), pp. 481–508.
17. P. Dupuis and H. Wang, *Subsolutions of an isaacs equation and efficient schemes for importance sampling*, Mathematics of Operations Research, 32 (2007), pp. 723–757.
18. W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, Springer, 2006.
19. M. Freidlin and A. Wentzell, *Random Perturbations of Dynamical Systems*, vol. 260 of Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg, 2012.
20. A. Friedman, *Partial differential equations of parabolic type*, Prentice-Hall, 1964.
21. D. Givon, *Strong convergence rate for two-time-scale jump-diffusion stochastic differential systems*, Multiscale Model. Simul., 6 (2007), pp. 577–594.
22. 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.
23. C. Hartmann, J. Latorre, G. Pavliotis, and W. Zhang, *Optimal control of multiscale systems using reduced-order models*, J. Comput. Dyn., 1 (2014), pp. 279–306.
24. W. K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika, 57 (1970), pp. 97–109.
25. 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.
26. C. Huang and D. Liu, *Strong convergence and speed up of nested stochastic simulation algorithm*, Commun. Comput. Phys., 15 (2014), pp. 1207–1236.
27. R. L. Jack and P. Sollich, *Effective interactions and large deviations in stochastic processes*, Eur. Phys. J. Special Topics, 224 (2015), pp. 2351–2367.
28. I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, Springer, 2 ed., 1991.
29. 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.
30. N. Krylov, *Controlled Diffusion Processes*, Stochastic Modelling and Applied Probability, Springer, 1980.
31. 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.
32. D. Liu, *Strong convergence of principle of averaging for multiscale stochastic dynamical systems*, Commun. Math. Sci., 8 (2010), pp. 999–1020.
33. J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer, 2nd ed., 2008.
34. J. S. Liu and R. Chen, *Sequential Monte Carlo methods for dynamic systems*, J. Amer. Statist. Assoc., 93 (1998), pp. 1032–1044.
35. A. Majda, C. Franzke, and B. Khouider, *An applied mathematics perspective on stochastic modelling for climate*, Philos. Trans. A Math. Phys. Eng. Sci., 366 (2008), pp. 2429–2455.
36. B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, Springer, 6th ed., 2010.
37. G. Pavliotis and A. Stuart, *Multiscale Methods: Averaging and Homogenization*, Springer, 2008.
38. J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, *Markov models of molecular kinetics: Generation and validation*, J. Chem. Phys., 134 (2011).
39. C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard, *A direct approach to conformational dynamics based on hybrid Monte Carlo*, J. Comput. Phys., 151 (1999), pp. 146 – 168.
40. K. Spiliopoulos, *Large deviations and importance sampling for systems of slow-fast motion*, Appl. Math. Optim., 67 (2013), pp. 123–161.
41. ———, *Nonasymptotic performance analysis of importance sampling schemes for small noise diffusions*, J. Appl. Probab., 52 (2015), pp. 797–810.
42. ———, *Rare event simulation for multiscale diffusions in random environments*, Multiscale Model. Simul., 13 (2015), pp. 1290–1311.
43. K. Spiliopoulos, P. Dupuis, and X. Zhou, *Escaping from an attractor: Importance sampling and rest points, part I*, Ann. Appl. Probab., 25 (2015), pp. 2909–2958.

44. E. Vanden-Eijnden and J. Weare, *Rare event simulation of small noise diffusions*, Comm. Pure Appl. Math., 65 (2012), pp. 1770–1803.
45. E. Weinan, D. Liu, and E. Vanden-Eijnden, *Analysis of multiscale methods for stochastic differential equations*, Comm. Pure Appl. Math., 58 (2005), pp. 1544–1585.
46. 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*, SIAM J. Sci. Comput., 36 (2014), pp. A2654–A2672.