Analysis of Multiscale Methods for Stochastic Differential Equations

WEINAN E
Princeton University

DI LIU
Courant Institute

AND

ERIC VANDEN-EIJNDEN
Courant Institute

Abstract

We analyze a class of numerical schemes proposed in [26] for stochastic differential equations with multiple time scales. Both advective and diffusive time scales are considered. Weak as well as strong convergence theorems are proven. Most of our results are optimal. They in turn allow us to provide a thorough discussion on the efficiency as well as optimal strategy for the method. © 2005 Wiley Periodicals, Inc.

1 Introduction

Multiscale modeling and computation have received a great deal of interest in recent years (for a review, see [8]). Yet there are relatively few analytical results available that help to assess the performance and provide guidance for designing these methods. The main purpose of the present paper is to provide a thorough analysis of a recently proposed numerical technique [26] (see also [11]) for stochastic differential equations with multiple time scales.

Consider the following generic example for $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$:

$$\begin{cases} \dot{X}_t^\varepsilon = f\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big), & X_0^\varepsilon = x, \\ \dot{Y}_t^\varepsilon = \dfrac{1}{\varepsilon} g\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big), & Y_0^\varepsilon = y. \end{cases} \tag{1.1}$$

Here $f(\cdot) \in \mathbb{R}^n$ and $g(\cdot) \in \mathbb{R}^m$ are $O(1)$ functions (possibly random) in ε, and ε is a small parameter representing the ratio of the time scales in the system. We have assumed that the phase space can be decomposed into slow degrees of freedom x and fast ones y. Systems of this type arise from molecular dynamics, material sciences, atmospheric and ocean sciences, etc. Standard computational schemes may fail due to the separation between the $O(\varepsilon)$ time scale that must be dealt with and the $O(1)$ and $O(\varepsilon^{-1})$ time scales that are of actual interest. On the analytical

Communications on Pure and Applied Mathematics, Vol. LVIII, 1544–1585 (2005)

side [12, 17, 24] (see also [3, 4, 20]), the following is known about (1.1). On the $O(1)$ time scale (advective time scale), if the dynamics for Y_t^ε with $X_t^\varepsilon = x$ fixed has an invariant probability measure $\mu_x^\varepsilon(dy)$ and the following limit exists:

$$\bar{f}(x) = \lim_{\varepsilon \to 0} \int_{\mathbb{R}^m} f(x, y, \varepsilon) \mu_x^\varepsilon(dy), \tag{1.2}$$

then in the limit of $\varepsilon \to 0$, X_t^ε converges to the solution of

$$\dot{\bar{X}}_t = \bar{f}(\bar{X}_t), \quad \bar{X}_0 = x. \tag{1.3}$$

On the longer $O(\varepsilon^{-1})$ time scale (diffusive time scale), fluctuations become important and additional terms must be included in (1.3). Under appropriate assumptions on f and g, for small ε the dynamics for X_t^ε can be approximated by the stochastic differential equation

$$\dot{\bar{X}}_t^\varepsilon = \bar{f}\big(X_t^\varepsilon\big) + \varepsilon \bar{b}\big(\bar{X}_t^\varepsilon\big) + \sqrt{\varepsilon}\bar{\sigma}\big(\bar{X}_t^\varepsilon\big)\dot{W}_t, \quad \bar{X}_0^\varepsilon = x, \tag{1.4}$$

where W_t is a Wiener process and the coefficients $\bar{b}$ and $\bar{\sigma}$ are expressed in terms of limits of expectations similar to (1.2); see Section 3. The importance of including the new terms proportional to $\bar{b}$ and $\bar{\sigma}$ is especially clear when $\bar{f}$ is either 0 or $O(\varepsilon)$ due, for example, to some symmetry, and the evolution of the slow variable arises only on the $O(\varepsilon^{-1})$ time scale.

It is often the case that the dynamics of the fast variables Y_t^ε is too complicated for the coefficients $\bar{f}$, $\bar{b}$, and $\bar{\sigma}$ to be computed analytically. The basic idea in [26] is to approximate $\bar{f}$, $\bar{b}$, and $\bar{\sigma}$ numerically by solving the original fine scale problem on time intervals of an intermediate scale, and use that data to evolve the slow variables with macroscopic time steps. Several related techniques have been proposed [7, 13, 16]. For kinetic Monte Carlo schemes involving disparate rates, Novotny et al. proposed in [16] a technique called projective dynamics, which reduces the Markov chain onto a smaller state space involving only the slow processes. A similar idea, also named projective dynamics, was proposed in [13] for dissipative deterministic ODEs with separated time scales. The method in [13] can also be viewed as a special case of a general class of methods called Chebyshev methods for stiff ODEs [19].

Of particular relevance to the present work is the framework of "heterogeneous multiscale" methods proposed in [7] (HMM for short), since it provides a very natural setting for the method proposed in [26]. At the same time, it also gives a general principle for the analysis of this kind of method. In the present setting, the general theorem proven in [7] states that if the macrosolver is stable, then the numerical error consists of two parts: a part due to the error in the macrosolver and a new part due to the approximation of the macroscale data (here the $\bar{\sigma}$, $\bar{a}$, and $\bar{b}$) by the microsolver. In general, the second part consists of the error in the microsolver, the relaxation error, and the sampling error. This general principle has been used for the analysis of several classes of multiscale methods (see in particular

[6, 9, 11, 23]). It is also the strategy that we will follow in this paper. Deterministic analogues of the algorithm were analyzed in [6, 10].

We will study equations like (1.1) both on the advective time scale (Section 2) and the diffusive time scale (Section 3). After presenting the multiscale numerical schemes (Sections 2.1, 3.1, and 3.5), we prove convergence theorems for these schemes (Sections 2.2, 2.3, and 3.2), and use these results (Sections 2.4 and 3.3) to determine the optimal set of numerical parameters to be used at a given error tolerance. We also illustrate the schemes and test our theorems on numerical examples (Sections 2.5 and 3.4).

Before ending this introduction, let us note that a simple trick for dealing with the multiscaled nature of the problem above is to increase the parameter ε to an optimal value according to a given error tolerance. This idea is indeed used in the artificial compressibility method for computing nearly incompressible flows [2] and the Car-Parrinello method [1], and has proven to be very successful. The multiscale scheme is much more efficient than direct solutions of the microscale model with the original $\varepsilon \ll 1$. More interestingly, our results show that if used correctly, the multiscale scheme is at least as efficient (on the advective time scale) or much more efficient (on the diffusive time scale) than a direct scheme even if an optimal value of ε is used in the microscale model to minimize the cost. In addition, the multiscale scheme can be applied even in situations when explicitly increasing the value of ε in the original equations can be difficult.

2 Advective Time Scale

Consider the following dynamics:

$$\begin{cases} \dot{X}_t^\varepsilon = a\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big), & X_0^\varepsilon = x, \\ \dot{Y}_t^\varepsilon = \dfrac{1}{\varepsilon} b\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big) + \dfrac{1}{\sqrt{\varepsilon}} \sigma\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big)\dot{W}_t, & Y_0^\varepsilon = y, \end{cases} \tag{2.1}$$

where $a \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and $\sigma \in \mathbb{R}^m \times \mathbb{R}^d$ are deterministic functions and W_t is a standard d-dimensional Wiener process. Define $\mathbb{C}_b^\infty$ to be the space of smooth functions with bounded derivatives of any order. We assume the following:

Assumption 2.1. The coefficients a, b, and σ, viewed as functions of (x, y, ε), are in $\mathbb{C}_b^\infty$; a and σ are bounded.

Assumption 2.2. There exists an $\alpha > 0$ such that $\forall (x, y, \varepsilon)$,

$$|\sigma^T(x, y, \varepsilon) y|^2 \geq \alpha |y|^2.$$

Assumption 2.3. There exists a $\beta > 0$ such that $\forall (x, y_1, y_2, \varepsilon)$,

$$\langle (y_1 - y_2), (b(x, y_1, \varepsilon) - b(x, y_2, \varepsilon))\rangle + \|\sigma(x, y_1, \varepsilon) - \sigma(x, y_2, \varepsilon)\|^2 \leq - \beta |y_1 - y_2|^2,$$

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

Assumption 2.1 can be weakened but is used here for simplicity of presentation. Assumption 2.2 means that the diffusion is nondegenerate for the y-process. Assumption 2.3 is a dissipative condition. Under these assumptions, one can show that for each (x, ε), the dynamics

$$\dot{Y}_t^{x,\varepsilon} = \frac{1}{\varepsilon} b\big(x, Y_t^{x,\varepsilon}, \varepsilon\big) + \frac{1}{\sqrt{\varepsilon}} \sigma\big(x, Y_t^{x,\varepsilon}, \varepsilon\big) \dot{W}_t, \quad Y_0^{x,\varepsilon} = y, \tag{2.2}$$

is exponentially mixing with a unique invariant probability measure $\mu_x^\varepsilon(dy)$. Define

$$\bar{a}(x) = \lim_{\varepsilon \to 0} \int_{\mathbb{R}^m} a(x, y, \varepsilon) \mu_x^\varepsilon(dy). \tag{2.3}$$

It is proven later (with error estimates) that under Assumptions 2.1, 2.2, and 2.3, X_t^ε converges strongly as $\varepsilon \to 0$ to the solution $\bar{X}_t$ of the following dynamics:

$$\dot{\bar{X}}_t = \bar{a}(\bar{X}_t), \quad \bar{X}_0 = x. \tag{2.4}$$

2.1 The Numerical Scheme

Usually of interest is the behavior of the slow variable X_t^ε, whose leading-order term for small ε is $\bar{X}_t$. But the coefficient $\bar{a}$ in the effective equation (2.4) for $\bar{X}_t$ is given via an expectation with respect to measure $\mu_x^\varepsilon(dy)$ that is usually difficult or impossible to obtain analytically, especially when the dimension m is large.

The basic idea proposed in [26] is to solve (2.4) with a macrosolver in which $\bar{a}$ is estimated by solving the microscale problem (2.2). This leads to multiscale schemes whose structure is explained next. (For simplicity we restrict ourselves to explicit solvers. Extension to implicit solvers is straightforward, but it tends to make the algorithm and implementation more involved.)

At each macrotime step n, having the numerical solution X_n, we need to estimate $\bar{a}(X_n)$ in order to move to step $n + 1$. Since $\bar{X}_t$ is deterministic, as a macrosolver we may use any stable explicit ODE solver such as a forward Euler, a Runge-Kutta, or a linear multistep method. For instance, in the simplest case when the forward Euler is selected as the macrosolver, we have

$$X_{n+1} = X_n + \tilde{a}_n \Delta t, \tag{2.5}$$

where Δt is the macrotime-step size and $\tilde{a}_n$ is the approximation of $\bar{a}(X_n)$ that we obtain in a two-step procedure:

(1) We solve (2.2) using a microsolver for stochastic ODEs and denote the solution by $\{Y_{n,m}\}$ where m labels the microtime steps. Multiple independent replicas can be created, in which case we denote the solutions by $\{Y_{n,m,j}\}$ where j is the replica number.

(2) We then define an approximation of $\bar{a}(X_n)$ by the following time and ensemble average:

$$\tilde{a}_n = \frac{1}{MN}\sum_{j=1}^{M}\sum_{m=n_T}^{n_T+N-1} a(X_n, Y_{n,m,j}, \varepsilon),$$

where M is the number of replicas, N is the number of steps in the time averaging, and n_T is the number of steps we skip to eliminate transients.

For the microsolver, denoting by ℓ its weak order of accuracy, for each realization we may use the first-order scheme ($\ell = 1$) [15]

$$\begin{aligned} Y^i_{n,m+1} = Y^i_{n,m} &+ \frac{1}{\sqrt{\varepsilon}}\sum_j \sigma^{ij}(X_n, Y_{n,m}, \varepsilon)\xi^j_{m+1}\sqrt{\delta t} \\ &+ \frac{1}{\varepsilon} b^i(X_n, Y_{n,m}, \varepsilon)\delta t + \frac{1}{\varepsilon}\sum_{jk} A^{ijk}(X_n, Y_{n,m}, \varepsilon) s^{kj}_{m+1}\delta t \end{aligned} \tag{2.6}$$

or the second-order scheme ($\ell = 2$)

$$\begin{aligned} Y^i_{n,m+1} = Y^i_{n,m} &+ \frac{1}{\sqrt{\varepsilon}}\sum_j \sigma^{ij}(X_n, Y_{n,m}, \varepsilon)\xi^j_{m+1}\sqrt{\delta t} \\ &+ \frac{1}{\varepsilon} b^i(X_n, Y_{n,m}, \varepsilon)\delta t + \frac{1}{\varepsilon}\sum_{jk} A^{ijk}(X_n, Y_{n,m}, \varepsilon) s^{kj}_{m+1}\delta t \\ &+ \frac{1}{2\varepsilon^{3/2}}\sum_j B^{ij}(X_n, Y_{n,m}, \varepsilon)\xi^j_{m+1}\delta t^{3/2} \\ &+ \frac{1}{2\varepsilon^2} C^i(X_n, Y_{n,m}, \varepsilon)\delta t^2. \end{aligned} \tag{2.7}$$

For the initial condition, we take $Y_{0,0} = 0$ and

$$Y_{n,0} = Y_{n-1,n_T+N-1}; \tag{2.8}$$

i.e., the initial values for the microvariables at macrotime step n are chosen to be their final values from macrotime step $n-1$.

In (2.6) and (2.7) δt is the microtime step size (note that it only appears in terms of the ratio $\delta t/\varepsilon =: \Delta\tau$), and the coefficients are defined as

$$\begin{cases} A^{ijk} = \sum_l (\partial^l \sigma^{ij})\sigma^{lk}, \\ B^{ij} = \sum_l (\sigma^{lj}\partial^l b^i + b^l \partial^l \sigma^{ij}) + \frac{1}{2}\sum_{kl} g^{kl}\partial^k\partial^l\sigma^{ij}, \\ C^i = \sum_j b^j\partial^j b^i + \frac{1}{2}\sum_{jk} g^{jk}\partial^j\partial^k b^i, \end{cases}$$

where $g = \sigma\sigma^T$ and the derivatives are taken with respect to y. The random variables $\{\xi_m^j\}$ are i.i.d. Gaussian with mean 0 and variance 1, and

$$s_m^{kj} = \begin{cases} \frac{1}{2}\xi_m^k \xi_m^j + z_m^{kj}, & k < j, \\ \frac{1}{2}\xi_m^k \xi_m^j - z_m^{jk}, & k > j, \\ \frac{1}{2}\left((\xi_m^j)^2 - 1\right), & k = j, \end{cases}$$

where $\{z_m^{kj}\}$ are i.i.d. with $\mathbb{P}\{z_m^{kj} = \frac{1}{2}\} = \mathbb{P}\{z_m^{kj} = -\frac{1}{2}\} = \frac{1}{2}$.

2.2 Strong Convergence Theorem

In this section, we give the rate of strong convergence for the scheme described above under Assumptions 2.1, 2.2, and 2.3 given at the beginning of Section 2. Throughout the remainder of the paper, we will denote by C a generic positive constant that may change its value from line to line.

THEOREM 2.4 *Assume that the macrosolver is stable and of k^{th}-order accuracy for* (2.4) *and that Δt and $\delta t/\varepsilon$ are small enough. Then for any $T_0 > 0$, there exists a constant $C > 0$ independent of $(\varepsilon, \Delta t, \delta t, n_T, M, N)$ such that*

$$\begin{aligned} &\sup_{n \le T_0/\Delta t} \mathbb{E}|X_{t_n}^\varepsilon - X_n| \le \\ &\quad C\Bigg(\sqrt{\varepsilon} + \Delta t^k + (\delta t/\varepsilon)^\ell \\ &\qquad + \frac{e^{-\frac{1}{2}\beta n_T(\delta t/\varepsilon)}}{\sqrt{N(\delta t/\varepsilon)+1}}\left(R + \sqrt{R}\right) + \frac{\sqrt{\Delta t}}{\sqrt{M(N(\delta t/\varepsilon)+1)}}\Bigg), \end{aligned} \tag{2.9}$$

where $t_n = n\Delta t$ and

$$R = \frac{\Delta t}{1 - e^{-\frac{1}{2}\beta(n_T+N-1)(\delta t/\varepsilon)}}. \tag{2.10}$$

The error estimate on $|X_{t_n}^\varepsilon - X_n|$ in (2.9) can be divided into three parts:

(1) $|X_t^\varepsilon - \bar{X}_t|$, where $\bar{X}_t$ is the solution of the effective equation (2.4),
(2) $|\bar{X}_{t_n} - \bar{X}_n|$, where $\bar{X}_n$ is the approximation of $\bar{X}_{t_n}$ given by the selected macrosolver assuming that $\bar{a}(x)$ is known, and
(3) $|\bar{X}_n - X_n|$.

The first part is a principle of averaging estimates for stochastic ODEs, and we will show in Lemma 2.5 that

$$\sup_{0 \le t \le T_0} \mathbb{E}|X_t^\varepsilon - \bar{X}_t| \le C\sqrt{\varepsilon}.$$

This part gives rise to the first term in (2.9).

The second part is a standard ODE estimate and based on the smoothness of $\bar{a}$ given by Lemma A.4 in the appendix; we have

$$\sup_{n\le T_0/\Delta t}\left|\bar{X}_{t_n}-\bar{X}_n\right|\le C\Delta t^k. \tag{2.11}$$

This is the second term in (2.9).

The third part accounts for the error caused by using $\tilde{a}_n$ instead of $\bar{a}(X_n)$ in the macrosolver. This part gives rise to a term of order $O(\varepsilon)$ that is dominated by $C\sqrt{\varepsilon}$ and to the remaining three terms in (2.9), which will be estimated later using Lemma 2.6. In this part of the error, the term $(\delta t/\varepsilon)^\ell$ is due to the microtime discretization that induces a difference between the invariant measures of the continuous and discrete dynamical systems. The term

$$\frac{e^{-\frac{1}{2}\beta n_T(\delta t/\varepsilon)}}{\sqrt{N(\delta t/\varepsilon)+1}}\bigl(R+\sqrt{R}\bigr)$$

accounts for the errors caused by relaxation of the fast variables. As will become clear from the proof, the factor R appears due to the way in (2.8) that we initialize the fast variables at each macrotime step. Different initializations will lead to similar error estimates with different values of R. For example, if we use $Y_{n,0}=0$, then $R=1$. As we will see in Section 2.4, the factor R is not essential for the multiscale scheme to be more efficient than a direct scheme, but the presence of this factor permits us to achieve even bigger efficiency gains. Finally, the term

$$\frac{\sqrt{\Delta t}}{\sqrt{M(N(\delta t/\varepsilon)+1)}}$$

accounts for the sampling errors when the fast variable reaches local equilibrium (via a central-limit-theorem type of estimate).

Before proceeding with the proof of Theorem 2.4, we point out a property of (2.9) that may seem somewhat surprising at first sight, namely, that the HMM scheme converges as $\Delta t\to 0$, $\delta t\to 0$, on any sequence such that $R\to 0$, even if one takes one realization only, $M=1$, and makes only one microtime step per macrotime step, $n_T=1$, $N=1$ (in this case $R=\Delta t/(\delta t/\varepsilon)$ plus higher-order terms). Indeed, in this case, (2.9) reduces to

$$\sup_{n\le T_0/\Delta t}\mathbb{E}\left|X^\varepsilon_{t_n}-X_n\right|\le C\bigl(\sqrt{\varepsilon}+\sqrt{\Delta t/(\delta t/\varepsilon)}+(\delta t/\varepsilon)^\ell\bigr). \tag{2.12}$$

While the set of parameters leading to (2.12) may not be optimal (see the discussion in Section 2.5), (2.12) is clearly a nice property of the multiscale scheme since, at fixed Δt and δt, the smaller M, n_T, and N, the more efficient the scheme is. The ultimate reason that the multiscale scheme converges even when $n_T=N=M=1$ has to be found in the proof of Theorem 2.4, but it is worthwhile to give an intuitive explanation for this fact.

The parameter n_T can be small and even equal to 1 because we reinitialize the fast variables at macrotime step n by their final value at macrotime step $n-1$

(see (2.8)). Therefore they already sample $\mu^{\varepsilon}_{X_{n-1}}(dy)$ initially when one lets them evolve to sample $\mu^{\varepsilon}_{X_n}(dy)$. Since $X_n - X_{n-1} = O(\Delta t)$, these two measures become closer and closer as $\Delta t \to 0$, and relaxation requires fewer and fewer microtime steps. This gives convergence even when $n_T = 1$ provided that $\Delta t/(\delta t/\varepsilon) \to 0$. (Note that if we do not use (2.8) to reinitialize the fast variables, the multiscale scheme still converges with $M = N = 1$ but it requires that $n_T(\delta t/\varepsilon) \to \infty$.)

On the other hand, the reason that the multiscale scheme converges even when $N = M = 1$ is best explained through a simple example. Consider the two-dimensional system

$$\begin{cases} \dot{X}^{\varepsilon}_t = -Y^{\varepsilon}_t, & X_0 = x, \\ \dot{Y}^{\varepsilon}_t = -\dfrac{1}{\varepsilon}(Y^{\varepsilon}_t - X^{\varepsilon}_t) + \dfrac{1}{\sqrt{\varepsilon}}\dot{W}_t, & Y_0 = y, \end{cases} \tag{2.13}$$

which leads to the following equation for X^{ε}_t in the limit as $\varepsilon \to 0$:

$$\dot{X}_t = -X_t, \quad X_0 = x. \tag{2.14}$$

To focus on sampling errors rather than relaxation errors, let us build a forward Euler multiscale scheme where, at each macrotime step, we draw only one realization of Y^{ε}_t out of the conditional measure

$$\mu_{X_n}(dy) = \frac{e^{-(y-X_n)^2}}{\sqrt{\pi}}\, dy \tag{2.15}$$

where X_n is the current value of the slow variable in the scheme. Extracting the mean explicitly, this amounts to using

$$X_{n+1} = X_n - X_n\Delta t + \frac{1}{\sqrt{2}}\xi_n \Delta t, \quad X_0 = x, \tag{2.16}$$

where the ξ_n's are i.i.d. Gaussian random variables with mean 0 and variance 1. Note that ξ_n is multiplied by Δt, not $\sqrt{\Delta t}$ as in a standard SDE. This is not a misprint and, in fact, is the reason that the noise term in (2.16) induces an error that disappears as $\Delta t \to 0$. To see this explicitly, note that the solution of (2.16) is

$$X_n = x(1-\Delta t)^n + \frac{\Delta t}{\sqrt{2}}\sum_{j=1}^{n-1}(1-\Delta t)^j \xi_{n-j}. \tag{2.17}$$

The first term on the right-hand side is what would have been provided by a forward Euler scheme for the limiting equation in (2.14). Therefore, a strong estimate accounting for the error introduced by sampling is

$$\mathbb{E}\big|X_n - x(1-\Delta t)^n\big|^2 = \frac{\Delta t^2}{2}\sum_{j=1}^{n-1}(1-\Delta t)^{2j} = O(\Delta t) \tag{2.18}$$

for $n = O(\Delta t^{-1})$. Even though the scheme above makes an $O(1)$ error in sampling at each macrotime step, it converges as $\Delta t \to 0$ because the fast variable is sampled over and over again before the slow variable has changed significantly. This leads to an effective number of realizations of the order $O(\Delta t^{-1})$ and shows the importance of assessing the quality of the estimator as integrated in the multiscale scheme rather than as a tool to evaluate the conditional expectation of the fast variables at each macrotime step.

LEMMA 2.5 *For any $T_0 > 0$, there exists a constant $C > 0$ independent of ε such that*

$$\sup_{0\le t\le T_0} \mathbb{E}\big|X_t^\varepsilon - \bar{X}_t\big| \le C\sqrt{\varepsilon}. \tag{2.19}$$

PROOF: Because $a(x, y, \varepsilon)$ is bounded, the slow process $\{X_t^\varepsilon\}_{0\le t\le T_0}$ is also bounded on the finite time interval $[0, T_0]$. Partitioning $[0, T_0]$ into subintervals of the same length $\Delta = \sqrt{\varepsilon}$ and denoting by $\lfloor z\rfloor$ the largest integer less than or equal to z, we construct the following auxiliary processes $(\tilde{X}_t^\varepsilon, \tilde{Y}_t^\varepsilon)$ such that for $t \in [k\Delta, (k+1)\Delta)$,

$$\begin{cases} \dot{\tilde{X}}_t^\varepsilon = a\big(X_{k\Delta}^\varepsilon, \tilde{Y}_t^\varepsilon, \varepsilon\big), & \tilde{X}_0^\varepsilon = x, \\ \dot{\tilde{Y}}_t^\varepsilon = \dfrac{1}{\varepsilon} b\big(X_{k\Delta}^\varepsilon, \tilde{Y}_t^\varepsilon, \varepsilon\big) + \dfrac{1}{\sqrt{\varepsilon}}\sigma\big(X_{k\Delta}^\varepsilon, \tilde{Y}_t^\varepsilon, \varepsilon\big)\dot{W}_t, & \tilde{Y}_{k\Delta}^\varepsilon = Y_{k\Delta}^\varepsilon. \end{cases}$$

A direct computation with Itô's formula gives for $t \in [k\Delta, (k+1)\Delta)$,

$$\begin{aligned} d\,\mathbb{E}\big|Y_t^\varepsilon - \tilde{Y}_t^\varepsilon\big|^2 &= \frac{2}{\varepsilon}\mathbb{E}(Y_t^\varepsilon - \tilde{Y}_t^\varepsilon)\cdot\big(b(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon) - b\big(X_{k\Delta}^\varepsilon, \tilde{Y}_t^\varepsilon, \varepsilon\big)\big)dt \\ &\quad + \frac{1}{\varepsilon}\mathbb{E}\big|\sigma\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big) - \sigma\big(X_{k\Delta}^\varepsilon, \tilde{Y}_t^\varepsilon, \varepsilon\big)\big|^2\,dt, \end{aligned} \tag{2.20}$$

where Y_t^ε solves (2.1). Using Assumptions 2.1, 2.2, and 2.3, we have

$$\begin{aligned} &\big(Y_t^\varepsilon - \tilde{Y}_t^\varepsilon\big)\cdot\big(b\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big) - b\big(X_{k\Delta}^\varepsilon, \tilde{Y}_t^\varepsilon, \varepsilon\big)\big) \\ &\quad + \frac{1}{2}\big|\sigma\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big) - \sigma\big(X_{k\Delta}^\varepsilon, \tilde{Y}_t^\varepsilon, \varepsilon\big)\big|^2 \\ &\quad \le \big(Y_t^\varepsilon - \tilde{Y}_t^\varepsilon\big)\cdot\big(b\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big) - b\big(X_t^\varepsilon, \tilde{Y}_t^\varepsilon, \varepsilon\big)\big) \\ &\qquad + \big(Y_t^\varepsilon - \tilde{Y}_t^\varepsilon\big)\cdot\big(b\big(X_t^\varepsilon, \tilde{Y}_t^\varepsilon, \varepsilon\big) - b\big(X_{k\Delta}^\varepsilon, \tilde{Y}_t^\varepsilon, \varepsilon\big)\big) \\ &\qquad + \big|\sigma\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big) - \sigma\big(X_t^\varepsilon, \tilde{Y}_t^\varepsilon, \varepsilon\big)\big|^2 \\ &\qquad + \big|\sigma\big(X_t^\varepsilon, \tilde{Y}_t^\varepsilon, \varepsilon\big) - \sigma\big(X_{k\Delta}^\varepsilon, \tilde{Y}_t^\varepsilon, \varepsilon\big)\big|^2 \\ &\quad \le -\beta\big|Y_t^\varepsilon - \tilde{Y}_t^\varepsilon\big|^2 + C\big(\big|Y_t^\varepsilon - \tilde{Y}_t^\varepsilon\big|\,\big|X_t^\varepsilon - X_{k\Delta}^\varepsilon\big| + \big|X_t^\varepsilon - X_{k\Delta}^\varepsilon\big|^2\big). \end{aligned}$$

Noting that for any $\beta > 0$, we have

$$C\big|Y_t^\varepsilon - \tilde{Y}_t^\varepsilon\big|\,\big|X_t^\varepsilon - X_{k\Delta}^\varepsilon\big| \le \frac{1}{2}\beta\big|Y_t^\varepsilon - \tilde{Y}_t^\varepsilon\big|^2 + \frac{C^2}{2\beta}\big|X_t^\varepsilon - X_{k\Delta}^\varepsilon\big|^2,$$

which can be written as

$$\begin{aligned}(2.21)\qquad &\big(Y_t^\varepsilon - \tilde{Y}_t^\varepsilon\big) \cdot \big(b\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big) - b\big(X_{k\Delta}^\varepsilon, \tilde{Y}_t^\varepsilon, \varepsilon\big)\big) \\ &\quad + \frac{1}{2}\big|\sigma\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big) - \sigma\big(X_{k\Delta}^\varepsilon, \tilde{Y}_t^\varepsilon, \varepsilon\big)\big|^2 \\ &\quad \le -\frac{1}{2}\beta\big|Y_t^\varepsilon - \tilde{Y}_t^\varepsilon\big|^2 + C\big|X_t^\varepsilon - X_{k\Delta}^\varepsilon\big|^2.\end{aligned}$$

By the boundedness of a, for $t \in [k\Delta, (k+1)\Delta)$,

$$(2.22)\qquad \mathbb{E}\big|X_t^\varepsilon - X_{k\Delta}^\varepsilon\big|^2 \le C\Delta^2.$$

Combining (2.20), (2.21), and (2.22), it follows that

$$d\,\mathbb{E}\big|Y_t^\varepsilon - \tilde{Y}_t^\varepsilon\big|^2 \le -\frac{\beta}{\varepsilon}\big|Y_t^\varepsilon - \tilde{Y}_t^\varepsilon\big|^2\,dt + \frac{C}{\varepsilon}\Delta^2\,dt.$$

Since $\mathbb{E}|Y_{k\Delta}^\varepsilon - \tilde{Y}_{k\Delta}^\varepsilon|^2 = 0$ by construction, the Gronwall inequality then implies that

$$\mathbb{E}\big|Y_t^\varepsilon - \tilde{Y}_t^\varepsilon\big|^2 \le C\Delta^2.$$

This is true for each $t \in [k\Delta, (k+1)\Delta)$, and hence for $0 \le t \le T_0$. Therefore we get

$$\begin{aligned}&\mathbb{E}\big|X_t^\varepsilon - \tilde{X}_t^\varepsilon\big|^2 \\ &\quad = \mathbb{E}\bigg|\int_0^t \big(a\big(X_s^\varepsilon, Y_s^\varepsilon, \varepsilon\big) - a\big(\tilde{X}_s^\varepsilon, \tilde{Y}_s^\varepsilon, \varepsilon\big)\big)ds\bigg|^2 \\ &\quad \le C\mathbb{E}\int_0^t \big(\big|X_s^\varepsilon - X_{\lfloor s/\Delta\rfloor\Delta}^\varepsilon\big|^2 + \big|\tilde{X}_s^\varepsilon - X_{\lfloor s/\Delta\rfloor\Delta}^\varepsilon\big|^2 + \big|Y_s^\varepsilon - \tilde{Y}_s^\varepsilon\big|^2\big)ds \\ &\quad \le C\varepsilon.\end{aligned}$$

This implies that

$$(2.23)\qquad \mathbb{E}\big|X_t^\varepsilon - \tilde{X}_t^\varepsilon\big| \le C\sqrt{\varepsilon}.$$

On the other hand, based on the smoothness of functions a and

$$\hat{a}(x, \varepsilon) := \int_{\mathbb{R}^m} a(x, y, \varepsilon)\mu_x^\varepsilon(dy),$$

and the exponential mixing property established in Lemma A.4 and Proposition A.2 given in the appendix, we have

$$
\begin{aligned}
&\mathbb{E}\left|\int_{k\Delta}^{(k+1)\Delta}\big(a\big(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_t,\varepsilon\big)-\bar{a}\big(X^\varepsilon_{k\Delta}\big)\big)dt\right|\\
&\quad\le \mathbb{E}\left|\int_{k\Delta}^{(k+1)\Delta}\big(a\big(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_t,\varepsilon\big)-\hat{a}(X^\varepsilon_{k\Delta},\varepsilon)\big)dt\right|\\
&\qquad+\left|\hat{a}\big(X^\varepsilon_{k\Delta},\varepsilon\big)-\bar{a}\big(X^\varepsilon_{k\Delta}\big)\right|\Delta \qquad\qquad (2.24)\\
&\quad\le C\varepsilon\mathbb{E}\big(\big|\tilde{Y}^\varepsilon_{k\Delta}\big|^2+1\big)\\
&\quad\le C\varepsilon,
\end{aligned}
$$

where the last step uses $\mathbb{E}(|\tilde{Y}^\varepsilon_{k\Delta}|^2+1)\le C$, which follows from the energy estimate (A.3) established in the appendix. By the smoothness of $\bar{a}(x)=\hat{a}(x,0)$, we have

$$
\begin{aligned}
\mathbb{E}\big|\tilde{X}^\varepsilon_t-\bar{X}_t\big| &= \mathbb{E}\left|\int_0^t\big(a(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta},\tilde{Y}^\varepsilon_s,\varepsilon)-\bar{a}(\bar{X}_s)\big)ds\right|\\
&\le \mathbb{E}\left|\int_0^t\big(a\big(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta},\tilde{Y}^\varepsilon_s,\varepsilon\big)-\bar{a}\big(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta}\big)\big)ds\right|\\
&\quad+\mathbb{E}\left|\int_0^t\big(\bar{a}\big(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta}\big)-\bar{a}\big(X^\varepsilon_s\big)\big)\,ds\right|\\
&\quad+\mathbb{E}\left|\int_0^t\big(\bar{a}\big(X^\varepsilon_s\big)-\bar{a}\big(\bar{X}_s\big)\big)\,ds\right|\\
&\le \mathbb{E}\left|\int_0^t a\big(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta},\tilde{Y}^\varepsilon_s\big)ds-\int_0^t\bar{a}\big(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta}\big)ds\right|\\
&\quad+C\Delta+C\int_0^t\mathbb{E}\big|X^\varepsilon_s-\bar{X}_s\big|ds.
\end{aligned}
$$

Breaking the first integral at the right hand-side into $\lfloor t/\Delta\rfloor$ pieces and using (2.24), we conclude that

$$
\mathbb{E}\big|\tilde{X}^\varepsilon_t-\bar{X}_t\big|\le C\left(\sqrt{\varepsilon}+\int_0^t\mathbb{E}\big|X^\varepsilon_s-\bar{X}_s\big|ds\right). \qquad (2.25)
$$

Since

$$
\mathbb{E}\big|X^\varepsilon_t-\bar{X}_t\big|\le\mathbb{E}\big|X^\varepsilon_t-\tilde{X}^\varepsilon_t\big|+\mathbb{E}\big|\tilde{X}^\varepsilon_t-\bar{X}_t\big|,
$$

combining (2.23) and (2.25) and using the Gronwall inequality, we arrive at (2.19). □

Denoting by $\mathbb{E}_{X_n}$ the conditional expectation with respect to X_n, we have the following:

LEMMA 2.6 *Under the assumptions in Theorem* 2.4, *for each* $T_0 > 0$, *there exists an independent constant* $C > 0$ *such that* $\forall n \in [0, T_0/\Delta t]$,

$$\begin{aligned}&\mathbb{E}|\mathbb{E}_{X_n}\tilde{a}_n - \bar{a}(X_n)|^2 \le \\ &\quad C\left(\varepsilon^2 + (\delta t/\varepsilon)^{2\ell} + \frac{e^{-\beta n_T(\delta t/\varepsilon)}}{N(\delta t/\varepsilon)+1}\left(e^{-\beta(n-1)N_m(\delta t/\varepsilon)} + R^2\right)\right),\end{aligned} \tag{2.26}$$

and

$$\begin{aligned}&\mathbb{E}|\tilde{a}_n - \bar{a}(X_n)|^2 \\ &\quad \le C\left(\varepsilon^2 + (\delta t/\varepsilon)^{2\ell} + \frac{e^{-\beta n_T(\delta t/\varepsilon)}}{N(\delta t/\varepsilon)+1}\left(e^{-\beta(n-1)N_m(\delta t/\varepsilon)} + R^2\right)\right) \\ &\qquad + C\frac{1}{M(N(\delta t/\varepsilon)+1)},\end{aligned} \tag{2.27}$$

where $N_m = n_T + N - 1$ *is the total number of microtime steps per macrotime step and realization.*

PROOF: Notice first that since $a(x, y, \varepsilon)$ is bounded, $\{X_n\}_{n\le T_0/\Delta t}$ is in a compact set. Let

$$\hat{a}(x,\varepsilon) = \int_{\mathbb{R}^m} a(x, y, \varepsilon)\mu^\varepsilon(dy).$$

Using the smoothness of $\hat{a}(x,\varepsilon)$ established in Lemma A.4, it follows that $\bar{a}(x) = \hat{a}(x, 0)$. Since

$$\mathbb{E}|\mathbb{E}_{X_n}\tilde{a}_n - \bar{a}(X_n)|^2 \le 2\mathbb{E}|\mathbb{E}_{X_n}\tilde{a}_n - \hat{a}(X_n,\varepsilon)|^2 + 2\mathbb{E}|\hat{a}(X_n,\varepsilon) - \bar{a}(X_n)|^2,$$

and $\mathbb{E}|\hat{a}(X_n,\varepsilon) - \bar{a}(X_n)|^2 \le C\varepsilon^2$ by Lemma A.4, this term gives the factor $C\varepsilon^2$ in (2.26), and it suffices to estimate $\mathbb{E}|\mathbb{E}_{X_n}\tilde{a}_n - \hat{a}(X_n,\varepsilon)|^2$ to derive the remaining terms. A similar argument gives the $C\varepsilon^2$ term in (2.27) and links the remaining terms to the estimate of $\mathbb{E}|\tilde{a}_n - \hat{a}(X_n,\varepsilon)|^2$.

We first compute

$$\begin{aligned}&\mathbb{E}|\mathbb{E}_{X_n}\tilde{a}_n - \hat{a}(X_n,\varepsilon)|^2 \\ &\quad = \frac{1}{M^2N^2}\mathbb{E}\left|\sum_{m,j}\mathbb{E}_{X_n}a(X_n, Y_{n,m,j},\varepsilon) - \hat{a}(X_n,\varepsilon)\right|^2 \\ &\quad \le \frac{1}{MN}\mathbb{E}\sum_{m,j}\left|\mathbb{E}_{X_n}a(X_n, Y_{n,m,j},\varepsilon) - \hat{a}(X_n,\varepsilon)\right|^2.\end{aligned}$$

By Lemma A.3 in the appendix, if $\delta t/\varepsilon$ is small enough, for each n and j, $Y_{n,m,j}$ is exponentially mixing with unique invariant probability measure $\mu_{X_n}^{\delta t,\varepsilon}$, and there exists a random variable $\zeta^{X_n,\delta t,\varepsilon}$ with distribution $\mu_{X_n}^{\delta t,\varepsilon}$ that is independent of the driving Wiener processes. Denote by $\zeta_{n,m}$ the solution provided by the microsolver with the initial condition $\zeta^{X_n,\delta t,\varepsilon}$. Then, by construction, the distribution of $\zeta_{n,m}$ is

$\mu_{X_n}^{\delta t,\varepsilon}$ for all $m > 0$. We have (recall that $N_m = n_T + N - 1$ is the total number of microtime steps per macrotime step and realization),

$$\begin{aligned}
&\big(\mathbb{E}\big|\mathbb{E}_{X_n} a(X_n, Y_{n,m,j}, \varepsilon) - \hat{a}(X_n, \varepsilon)\big|^2\big)^{1/2} \\
&\quad \le \big(\mathbb{E}\big|\mathbb{E}_{X_n}\big(a(X_n, Y_{n,m,j}, \varepsilon) - a(X_n, \zeta_{n,nN_m+m}, \varepsilon)\big)\big|^2\big)^{1/2} \\
&\qquad + \big(\mathbb{E}\big|\mathbb{E}_{X_n} a(X_n, \zeta_{n,nN_m+m}, \varepsilon) - \hat{a}(X_n, \varepsilon)\big|^2\big)^{1/2}.
\end{aligned}$$

The smoothness of a guarantees that

$$\mathbb{E}\big|\mathbb{E}_{X_n}\big(a(X_n, Y_{n,m,j}, \varepsilon) - a(X_n, \zeta_{n,nN_m+m}, \varepsilon)\big)\big|^2 \le C\mathbb{E}|Y_{n,m,j} - \zeta_{n,nN_m+m}|^2,$$

while (A.14) implies that

$$\mathbb{E}\big|\mathbb{E}_{X_n} a(X_n, \zeta_{n,nN_m+m}, \varepsilon) - \hat{a}(X_n, \varepsilon)\big|^2 \le C(\delta t/\varepsilon)^{2\ell}.$$

Therefore

$$\begin{aligned}
(2.28)\quad &\big(\mathbb{E}\big|\mathbb{E}_{X_n} a(X_n, Y_{n,m,j}, \varepsilon) - \hat{a}(X_n, \varepsilon)\big|^2\big)^{1/2} \le \\
&\qquad C\big(\big(\mathbb{E}|Y_{n,m,j} - \zeta_{n,nN_m+m}|^2\big)^{1/2} + (\delta t/\varepsilon)^{\ell}\big).
\end{aligned}$$

The exponential mixing property established in Lemma A.3 implies that

$$\begin{aligned}
&\big(\mathbb{E}|Y_{n,m,j} - \zeta_{n,nN_m+m}|^2\big)^{1/2} \\
&\quad \le e^{-\frac{1}{2}\beta m(\delta t/\varepsilon)}\big(\mathbb{E}|Y_{n-1,N_m,j} - \zeta_{n,nN_m}|^2\big)^{1/2} \\
&\quad \le e^{-\frac{1}{2}\beta m(\delta t/\varepsilon)}\big(\mathbb{E}|Y_{n-1,N_m,j} - \zeta_{n-1,nN_m}|^2\big)^{1/2} \\
&\qquad + e^{-\frac{1}{2}\beta m(\delta t/\varepsilon)}\big(\mathbb{E}|\zeta_{n,nN_m} - \zeta_{n-1,nN_m}|^2\big)^{1/2} \\
&\quad \le e^{-\frac{1}{2}\beta m(\delta t/\varepsilon)}\big((\mathbb{E}|Y_{n-1,N_m,j} - \zeta_{n-1,nN_m})|^2\big)^{1/2} + C\Delta t\big),
\end{aligned}$$

where the last inequality follows since, by Assumptions 2.1 and 2.3,

$$\mathbb{E}|\zeta_{n,nN_m} - \zeta_{n-1,nN_m}|^2 \le C\mathbb{E}|X_n - X_{n-1}|^2 \le C\Delta t^2.$$

Repeating the above argument at each macrotime step from $n-1$ to $n = 0$, we have

$$\begin{aligned}
&\big(\mathbb{E}|Y_{n-1,N_m,j} - \zeta_{n-1,nN_m}|^2\big)^{1/2} \\
&\quad \le e^{-\frac{1}{2}\beta N_m(\delta t/\varepsilon)}\big((\mathbb{E}|Y_{n-2,N_m,j} - \zeta_{n-2,(n-1)N_m}|^2)^{1/2} + C\Delta t\big) \\
&\quad \le C(e^{-\frac{1}{2}\beta(n-1)N_m(\delta t/\varepsilon)} + R).
\end{aligned}$$

Inserting these results in (2.28), we arrive at

$$\begin{aligned}
&\mathbb{E}\big|\mathbb{E}_{X_n} a(X_n, Y_{n,m,j}, \varepsilon) - \hat{a}(X_n, \varepsilon)\big|^2 \\
&\qquad \le C\big(e^{-\beta m(\delta t/\varepsilon)}(e^{-\beta(n-1)N_m(\delta t/\varepsilon)} + R^2) + (\delta t/\varepsilon)^{2\ell}\big).
\end{aligned}$$

The inequality also holds for $n = 0$ with an appropriate choice of C. Summing over $m \in [n_T, n_T + N - 1]$ and $j \in [1, M]$, we obtain

$$\begin{aligned}\mathbb{E}\left|\mathbb{E}_{X_n}\tilde{a}_n - \hat{a}(X_n,\varepsilon)\right|^2 \\ \le C\left(e^{-\beta n_T(\delta t/\varepsilon)}\big(e^{-\beta(n-1)N_m(\delta t/\varepsilon)} + R^2\big)\frac{1-e^{-\beta N(\delta t/\varepsilon)}}{N(1-e^{-\beta(\delta t/\varepsilon)})} + (\delta t/\varepsilon)^{2\ell}\right).\end{aligned}$$

Assuming $\delta t/\varepsilon \in (0,1)$, we have

$$\frac{1-e^{-\beta N(\delta t/\varepsilon)}}{N(1-e^{-\beta(\delta t/\varepsilon)})} \le C\frac{1-e^{-\beta N(\delta t/\varepsilon)}}{N(\delta t/\varepsilon)} \le \frac{C'}{N(\delta t/\varepsilon)+1},$$

and the last two terms in (2.26) follow. Next we compute

$$\begin{aligned}\mathbb{E}\big|\tilde{a}_n - \hat{a}(X_n,\varepsilon)\big|^2 = \frac{1}{M^2N^2}\sum_{j,m,k,l}\mathbb{E}\big(a(X_n, Y_{n,m,j},\varepsilon) - \hat{a}(X_n,\varepsilon)\big) \\ \cdot\big(a(X_n, Y_{n,l,k},\varepsilon) - \hat{a}(X_n,\varepsilon)\big).\end{aligned}$$

By the same analysis as above and by independence between $Y_{n,m,j}$ and $Y_{n,m,k}$ for $j \ne k$ for given $\{X_{n'}\}_{n'\le n}$ and $\{Y_{n',\cdot,\cdot}\}_{n'<n}$, we have for $j \ne k$ in the sum above

$$\begin{aligned}&\left|\sum_{j\ne k}\sum_{m,l}\mathbb{E}\big(a(X_n, Y_{n,m,j},\varepsilon) - \hat{a}(X_n,\varepsilon)\big)\cdot\big(a(X_n, Y_{n,l,k},\varepsilon) - \hat{a}(X_n,\varepsilon)\big)\right| \\ &\quad\le \sum_{j\ne k}\sum_{m,l}\mathbb{E}\big|\mathbb{E}_n\big(a(X_n, Y_{n,m,j},\varepsilon) - \hat{a}(X_n,\varepsilon)\big)\cdot\mathbb{E}_n\big(a(X_n, Y_{n,l,k},\varepsilon) - \hat{a}(X_n,\varepsilon)\big)\big| \\ &\quad\le CM^2N^2\left(\frac{e^{-\beta n_T(\delta t/\varepsilon)}}{N(\delta t/\varepsilon)+1}\big(e^{-\beta(n-1)N_m(\delta t/\varepsilon)} + R^2\big) + (\delta t/\varepsilon)^{2\ell}\right),\end{aligned}$$

where E_n denotes the expectation conditioned on $\{X_{n'}\}_{n'\le n}$ and $\{Y_{n',\cdot,\cdot}\}_{n'<n}$. When $j = k$ we have

$$\begin{aligned}&\big|\mathbb{E}\big(a(X_n, Y_{n,m,j},\varepsilon) - \hat{a}(X_n,\varepsilon)\big)\cdot\big(a(X_n, Y_{n,l,j},\varepsilon) - \hat{a}(X_n,\varepsilon)\big)\big| \le \\ &\quad\big|\mathbb{E}\big(\mathbb{E}_{X_n}\big(a(X_n, Y_{n,m,j},\varepsilon) - \hat{a}(X_n,\varepsilon)\big)\cdot\mathbb{E}_{n,m,j}\big(a(X_n, Y_{n,l,j},\varepsilon) - \hat{a}(X_n,\varepsilon)\big)\big)\big|,\end{aligned}$$

when $m \le l$ and similarly when $m > l$. Here $\mathbb{E}_{n,m,j}$ denotes the conditional expectation with respect to $Y_{n,m,j}$.

By the energy estimate (A.3) and the same analysis as above using exponential mixing, we deduce

$$\mathbb{E}\big|\mathbb{E}_{X_n}\big(a(X_n, Y_{n,m,j},\varepsilon) - \hat{a}(X_n,\varepsilon)\big)\big|^2 \le C\left(e^{-\beta m(\delta t/\varepsilon)} + (\delta t/\varepsilon)^{2\ell}\right)$$

and

$$\mathbb{E}\big|\mathbb{E}_{n,m,j}\big(a(X_n, Y_{n,l,j},\varepsilon) - \hat{a}(X_n,\varepsilon)\big)\big|^2 \le C\left(e^{-\beta(l-m)(\delta t/\varepsilon)} + (\delta t/\varepsilon)^{2\ell}\right).$$

Summing over $j \in [1, M]$ and $m, l \in [n_T, n_T + N - 1]$, this leads to

$$\left| \sum_j \sum_{m,l} \mathbb{E}\big(a(X_n, Y_{n,m,j}, \varepsilon) - \hat{a}(X_n, \varepsilon)\big) \cdot \big(a(X_n, Y_{n,l,j}, \varepsilon) - \hat{a}(X_n, \varepsilon)\big) \right| \leq CMN^2 \left(\frac{1}{N(\delta t/\varepsilon) + 1} + (\delta t/\varepsilon)^{2\ell} \right),$$

and the last two terms in (2.27) follow. □

PROOF OF THEOREM 2.4: We will prove (2.9) for the case when the macrosolver is the forward Euler method. The extension to general stable macrosolvers mentioned in Section 2.1 is straightforward. By the boundedness of a, the solutions of all the equations for the slow processes X are all in a compact set. Letting $e_n = \bar{X}_n - X_n$, we have

$$\begin{aligned} e_{n+1} &= e_n + \Delta t\big(\bar{a}(\bar{X}_n) - \tilde{a}_n\big) \\ &= e_n + \Delta t\big(\bar{a}(\bar{X}_n) - \bar{a}(X_n)\big) + \Delta t\big(\bar{a}(X_n) - \tilde{a}_n\big) \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}e_{n+1}^2 &= \mathbb{E}\big(e_n + \Delta t\big(\bar{a}(\bar{X}_n) - \bar{a}(X_n)\big)\big)^2 + \Delta t^2 \mathbb{E}\big(\bar{a}(X_n) - \tilde{a}_n\big)^2 \\ &\quad + 2\Delta t \mathbb{E}\big(e_n + \Delta t\big(\bar{a}(\bar{X}_n) - \bar{a}(X_n)\big)\big) \cdot (\bar{a}(X_n) - \tilde{a}_n) \\ &=: I_1 + I_2 + I_3. \end{aligned}$$

For I_1, by the smoothness of $\bar{a}$, we have

$$I_1 \leq (1 + C\Delta t)e_n^2.$$

For I_2, letting $N_m = n_T + N - 1$, by (2.27) we have

$$\begin{aligned} I_2 &\leq C\Delta t^2 \left(\frac{e^{-\beta n_T (\delta t/\varepsilon)}}{N(\delta t/\varepsilon) + 1} \big(e^{-\beta(n-1)N_m(\delta t/\varepsilon)} + R^2\big) + (\delta t/\varepsilon)^{2\ell} + \varepsilon^2 \right) \\ &\quad + C \frac{\Delta t^2}{M(N(\delta t/\varepsilon) + 1)}. \end{aligned}$$

For I_3, by (2.26), we have

$$\begin{aligned} |I_3| &\leq 2\Delta t \mathbb{E}\big|e_n + \Delta t\big(\bar{a}(\bar{X}_n) - \bar{a}(X_n)\big)\big| \, \big|\mathbb{E}_{X_n}\tilde{a}_n - \bar{a}(X_n)\big| \\ &\leq \Delta t \mathbb{E}\big|e_n + \Delta t\big(\bar{a}(\bar{X}_n) - \bar{a}(X_n)\big)\big|^2 + C\Delta t \mathbb{E}\big|\mathbb{E}_{X_n}\tilde{a}_n - \bar{a}(X_n)\big|^2 \\ &\leq \Delta t(1 + C\Delta t)\mathbb{E}e_n^2 \\ &\quad + C\Delta t \left(\frac{e^{-\beta n_T(\delta t/\varepsilon)}}{N(\delta t/\varepsilon) + 1} \big(e^{-\beta(n-1)N_m(\delta t/\varepsilon)} + R^2\big) + (\delta t/\varepsilon)^{2\ell} + \varepsilon^2 \right). \end{aligned}$$

It follows that

$$\begin{aligned}\mathbb{E}e_{n+1}^2 &\le (1+C\Delta t)\mathbb{E}e_n^2 \\ &\quad + C\Delta t\left(\frac{e^{-\beta n_T(\delta t/\varepsilon)}}{N(\delta t/\varepsilon)+1}\left(e^{-\beta(n-1)N_m(\delta t/\varepsilon)}+R^2\right)+(\delta t/\varepsilon)^{2\ell}+\varepsilon^2\right) \\ &\quad + C\frac{\Delta t^2}{M(N(\delta t/\varepsilon)+1)}.\end{aligned}$$

Therefore, using the fact that $1-x^2 \ge 1-x$ for $x \in (0,1)$, we have

$$\begin{aligned}\mathbb{E}e_n^2 &\le C\frac{e^{-\beta n_T(\delta t/\varepsilon)}}{N(\delta t/\varepsilon)+1}(R^2+R) \\ &\quad + C\left((\delta t/\varepsilon)^{2\ell}+\varepsilon^2+\frac{\Delta t}{M(N(\delta t/\varepsilon)+1)}\right),\end{aligned}$$

which, together with (2.19), completes the proof of Theorem 2.4. □

2.3 Weak Convergence Theorem

Next we give the rate of weak convergence for the multiscale scheme under Assumptions 2.1, 2.2, and 2.3.

THEOREM 2.7 *Assume that the assumptions in Theorem* 2.4 *hold. Then for any* $f \in \mathbb{C}_0^\infty$ *and* $T_0 > 0$, *there exists a constant* $C > 0$ *independent of* $(\varepsilon, \Delta t, \delta t, n_T, M, N)$ *such that*

$$\begin{aligned}&\sup_{n\le T_0/\Delta t}\left|\mathbb{E}f\left(X_{t_n}^\varepsilon\right)-\mathbb{E}f(X_n)\right| \\ &\quad \le C\left(\varepsilon+\Delta t^k+(\delta t/\varepsilon)^\ell+\frac{e^{-\frac{1}{2}\beta n_T(\delta t/\varepsilon)}}{\sqrt{N(\delta t/\varepsilon)+1}}(R+R^2)\right. \\ &\qquad \left.+\frac{\Delta t}{M(N(\delta t/\varepsilon)+1)}\right).\end{aligned} \tag{2.29}$$

As before, we split the estimate of $|\mathbb{E}f(X_{t_n}^\varepsilon)-\mathbb{E}f(X_n)|$ into three parts:

(1) $|\mathbb{E}f(X_t^\varepsilon)-f(\bar{X}_t)|$,
(2) $|f(\bar{X}_{t_n})-f(\bar{X}_n)|$, and
(3) $|f(\bar{X}_n)-\mathbb{E}f(X_n)|$.

The first part accounts for the error caused by replacing X_t^ε by the solution $\bar{X}_t$ of the asymptotic equation in (2.4). It has been proven (e.g., in [12, 24]) that

$$\sup_{0\le t\le T_0}\left|\mathbb{E}f\left(X_t^\varepsilon\right)-f\left(\bar{X}_t\right)\right| \le C\varepsilon. \tag{2.30}$$

This is the first term in (2.29), and we give a formal derivation of this result by perturbation analysis at the end of this section. The second part accounts for the error caused by the macrosolver assuming that the coefficient $\bar{a}$ in (2.4) is known exactly. A standard ODE estimate using the smoothness of $\bar{a}$ (see Lemma A.4 in the appendix) gives

$$\sup_{n\le T_0/\Delta t}\left|f\left(\bar{X}_{t_n}\right)-f\left(\bar{X}_n\right)\right|\le C\Delta t^k. \tag{2.31}$$

This is the second term in (2.9). The third part, the HMM error, accounts for the error introduced by using $\tilde{a}_n$ instead of $\bar{a}(X_n)$ in the macrosolver. This part gives rise to a term of order $O(\varepsilon)$ that can be absorbed in $C\varepsilon$ and to the last three terms in (2.29). These terms account for discretization error in the microscheme, as well as relaxation and sampling errors, and are estimated in Lemma 2.8.

LEMMA 2.8 *For any $f\in\mathbb{C}_0^\infty$ and $T_0>0$, there exists a constant $C>0$ such that*

$$\begin{aligned}\left|f\left(\bar{X}_n\right)-\mathbb{E}f\left(X_n\right)\right| &\\ \le C\Bigg(\varepsilon+(\delta t/\varepsilon)^\ell+\frac{e^{-\frac{1}{2}\beta n_T(\delta t/\varepsilon)}}{\sqrt{N(\delta t/\varepsilon)+1}}(R+R^2) &\\ +\frac{\Delta t}{M(N(\delta t/\varepsilon)+1)}\Bigg).\end{aligned} \tag{2.32}$$

PROOF: Again for simplicity we will discuss only the case when the macrosolver is the forward Euler method. To estimate $|f(\bar{X}_n)-\mathbb{E}f(X_n)|$, we define an auxiliary function $u(k,x)$ for $k\le n$ as follows:

$$u(n,x)=f(x),\quad u(k,x)=u(k+1,x+\Delta t\bar{a}(x)).$$

Then we have $u(0,x)=f(\bar{X}_n)$. By the boundedness of a, the solutions of the equations for X are uniformly bounded in a compact set K. By the smoothness of $\bar{a}$, it is easy to show that

$$\sup_{k,x}\left\{\left|\partial_x u(k,x)\right|+\left|\partial_x^2 u(k,x)\right|\right\}$$

is uniformly bounded on K for different Δt. Hence we have

$$\begin{aligned}&\left|\mathbb{E}\big(u(k+1,X_{k+1})-u(k,X_k)\big)\right|\\ &\quad=\left|\mathbb{E}\big(u(k+1,X_k+\Delta t\tilde{a}_k)-u(k+1,X_k+\Delta t\bar{a}(X_k))\big)\right|\\ &\quad\le\Delta t\left|\mathbb{E}\partial_x u(k+1,X_k)\cdot(\mathbb{E}_{X_k}\tilde{a}_k-\bar{a}(X_k))\right|\\ &\qquad+\frac{1}{2}\Delta t^2\sup_{y\in K}\left|\partial_x^2 u(k+1,y)\right|\mathbb{E}|\tilde{a}_k-\bar{a}(X_k)|^2\\ &\quad\le C\big(\Delta t\mathbb{E}|\mathbb{E}_{X_k}\tilde{a}_k-\bar{a}(X_k)|+\Delta t^2\mathbb{E}|\tilde{a}_k-\bar{a}(X_k)|^2\big).\end{aligned}$$

So, by Lemma 2.6, we have

$$
\begin{aligned}
\left|\mathbb{E}f\left(X_n\right) - f\left(\bar{X}_n\right)\right| &= \left|\mathbb{E}u(n, X_n) - u(0, x)\right| \\
&= \left|\sum_{0\le k\le n-1} \mathbb{E}(u(k+1, X_{k+1}) - u(k, X_k))\right| \\
&\le C\left(\varepsilon + \sum_{0\le k\le n-1} \Delta t \frac{e^{-\frac{1}{2}\beta n_T(\delta t/\varepsilon)}}{\sqrt{N\delta t + 1}} \cdot e^{-\frac{1}{2}\beta k N_m(\delta t/\varepsilon)}\right) \\
&\quad + C\Delta t\left(\frac{e^{-\frac{1}{2}\beta n_T(\delta t/\varepsilon)}}{\sqrt{N(\delta t/\varepsilon)+1}} R + \Delta t \frac{e^{-\beta n_T(\delta t/\varepsilon)}}{N(\delta t/\varepsilon)+1} R^2\right) \\
&\quad + C\Delta t\left((\delta t/\varepsilon)^{\ell} + \frac{\Delta t}{M(N(\delta t/\varepsilon)+1)}\right),
\end{aligned}
$$

and we are done. □

Finally, we give a formal argument for (2.30) using perturbation analysis [24] . It is known that $u^\varepsilon(t, x, y) = \mathbb{E}\{f(X_t^\varepsilon)\}$ satisfies the following backward Fokker-Planck equation:

$$
\frac{\partial u^\varepsilon}{\partial t} = \left(\frac{1}{\varepsilon} L_1 + L_2 + \varepsilon L_3\right) u^\varepsilon, \quad u(0) = f. \tag{2.33}
$$

Here

$$
L_1 = b(x, y, 0)\partial_y + \frac{1}{2}\sigma\sigma^T(x, y, 0)\partial_y^2, \quad L_2 = a(x, y, 0)\partial_x + L_2^y,
$$

where L_2^y is a differential operator in y only, and εL_3 contains higher-order terms in ε. It is known [14] that under Assumptions 2.1, 2.2, and 2.3, for each x, the process associated with L_1 has $\mu_x(dy) \equiv \mu_x^{\varepsilon=0}(dy)$ as a unique invariant measure, and this measure has a density, $\mu_x(dy) = p_x(y)dy$.

Define P by

$$
Pf(x) = \int_{\mathbb{R}^m} f(x, y) p_x(y) dy.
$$

Notice that P is the projection onto the null space of L_1. Let u^ε be formally represented by a power series

$$
u^\varepsilon = u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \dots .
$$

Inserting into (2.33) and equating coefficients of equal powers of ε, we get

$$
L_1 u_0 = 0, \quad \frac{\partial u_0}{\partial t} = L_1 u_1 + L_2 u_0, \quad \dots .
$$

Suppose $Pu_0(0) = u_0(0)$. Then $Pu_0 = u_0$ for all $t > 0$ and acting with P on both sides of the second equation, we obtain the following transport equation for u_0:

$$\frac{\partial u_0}{\partial t} = PL_2 u_0 = Pa(x, y, 0)\partial_x u_0 = \bar{a}(x)\partial_x u_0, \quad u_0(0) = f.$$

u_1 is given by

$$u_1 = L_1^{-1}(PL_2 - L_2)u_0.$$

Now we have

$$\begin{aligned}
&\left(\frac{\partial}{\partial t} - \frac{1}{\varepsilon}L_1 - L_2 - \varepsilon L_3\right)(u^\varepsilon - u_0 - \varepsilon u_1) \\
&\quad = \left(\frac{1}{\varepsilon}L_1 + L_2 + \varepsilon L_3 - \frac{\partial}{\partial t}\right)(u_0 + \varepsilon u_1) \\
&\quad = \varepsilon\left(L_2 + \varepsilon L_3 - \frac{\partial}{\partial t}\right)u_1 + \varepsilon L_3 u_0 \\
&\quad = O(\varepsilon).
\end{aligned}$$

This means that on finite time intervals, as long as u_0 and u_1 are bounded, we have

$$u^\varepsilon - u_0 = O(\varepsilon).$$

The boundedness of u_0 and u_1 is implied by the smoothness of $\bar{a}$ and the exponential mixing.

Remark 2.9. The derivation above implies that even if Assumptions 2.1 and 2.3 are not satisfied as in the numerical example in the next section, the weak convergence still holds provided u_0 and u_1 are bounded on finite time intervals.

2.4 Efficiency and Consistency Analysis

A measure of the cost of the multiscale scheme described in Section 2.1 is the number of microtime steps per unit of time,

$$\text{cost} = \frac{MN_m}{\Delta t} = \frac{M(n_T + N - 1)}{\Delta t}.$$

Suppose that one wishes to compute with an error tolerance λ and ε is such that $\sqrt{\varepsilon} \ll \lambda$ if strong convergence is required, or $\varepsilon \ll \lambda$ if weak convergence is enough. Then the multiscale scheme is applicable, and the best numerical strategy is to choose the parameters in the scheme so that in estimate (2.9) or (2.29) each term but the first is of order λ. Suppose that we take one realization only, $M = 1$, and evaluate expectations via time average. Then the optimal parameters are

$$\begin{aligned}
&\Delta t = O(\lambda^{1/k}), && \delta t/\varepsilon = O(\lambda^{1/\ell}), \\
&n_T = O(\lambda^{-1/\ell}), && N = O(\lambda^{-\alpha+1/k-1/\ell}),
\end{aligned} \tag{2.34}$$

with $\alpha = 2$ for strong convergence and $\alpha = 1$ for weak convergence. The corresponding cost is

$$\text{cost} = \frac{n_T + N - 1}{\Delta t} = O(\lambda^{-\alpha-1/\ell}). \tag{2.35}$$

Similarly, if one takes $N = 1$ and evaluates expectations via ensemble average only, we arrive at

$$\begin{aligned} \Delta t &= O(\lambda^{1/k}), & \delta t/\varepsilon &= O(\lambda^{1/\ell}), \\ n_T &= O(\lambda^{-1/\ell}\log\lambda^{-1}), & M &= O\big(\lambda^{-\alpha+1/k}\big), \end{aligned} \tag{2.36}$$

and the cost is

$$\text{cost} = \frac{M n_T}{\Delta t} = O(\lambda^{-\alpha-1/\ell}\log\lambda^{-1}). \tag{2.37}$$

This indicates that (2.35) is in fact optimal over both N and M, though the additional cost of using ensemble instead of time averaging is only marginal and proportional to $O(\log\lambda^{-1})$. Notice that the cost decreases as the order of the microscheme increases, but there is no gain in utilizing a higher-order macroscheme since (2.35) and (2.37) are independent of k. The reason is quite simple. A higher-order macroscheme in principle allows one to use a larger macrotime step, but this increases the sampling errors unless a larger N or M is used as well. The two effects balance each other exactly and the cost remains the same. Therefore, the only way to drive the cost down would be to use a higher-order estimator in conjunction with higher-order micro- and macroschemes. This can be done in the nonrandom case (see, e.g., [6, 10]), but it is difficult to construct higher-order estimators when the fast process is governed by an SDE.

The costs in (2.35) and (2.37) indicate that, for small ε, the multiscale scheme is cheaper than a direct scheme for (2.1). Denote by X_n^ε the numerical approximation provided by a scheme of weak order ℓ (same as in the microsolver used in the multiscale scheme) applied directly to (2.1) and assume that the corresponding strong order of the scheme is $\ell/2$. Then the following error estimates hold:

$$\sup_{n \le T_0/\delta t} \mathbb{E}\big|X_{t_n}^\varepsilon - X_n^\varepsilon\big| \le C(\delta t/\varepsilon)^{\ell/2}, \tag{2.38}$$

$$\sup_{n \le T_0/\delta t} \big|\mathbb{E} f\big(X_{t_n}^\varepsilon\big) - \mathbb{E} f\big(X_n^\varepsilon\big)\big| \le C(\delta t/\varepsilon)^{\ell}. \tag{2.39}$$

Thus, at error tolerance λ, a time step of order $\delta t/\varepsilon = O(\lambda^{\alpha/\ell})$ must be used, leading to

$$\text{cost} = 1/\delta t = O(\varepsilon^{-1}\lambda^{-\alpha/\ell}),$$

where, as before, $\alpha = 2$ for strong convergence and $\alpha = 1$ for weak convergence. This is much higher than the cost of the multiscale scheme when $\varepsilon \ll \lambda^\alpha$.

A much tougher test for the multiscale scheme is to compare it with a direct scheme not for (2.1) but rather for an equation like (2.1) where an optimal ε, say,

$\varepsilon' > \varepsilon$, is used. Denote by $X_n^{\varepsilon'}$ the numerical approximation for this equation provided by a microscheme of weak order ℓ (same as in the microsolver used in the multiscale scheme) and strong order $\ell/2$. Then analysis similar to the ones presented in Sections 2.2 and 2.3 gives the following error estimates for the scheme:

$$\sup_{n \le T_0/\delta t'} \mathbb{E}\big|X_{t_n}^{\varepsilon} - X_n^{\varepsilon'}\big| \le C\big(\sqrt{\varepsilon'} + (\delta t'/\varepsilon')^{\ell/2}\big), \tag{2.40}$$

$$\sup_{n \le T_0/\delta t'} \big|\mathbb{E} f\big(X_{t_n}^{\varepsilon}\big) - \mathbb{E} f\big(X_n^{\varepsilon'}\big)\big| \le C\big(\varepsilon' + (\delta t'/\varepsilon')^{\ell}\big). \tag{2.41}$$

Given the error tolerance λ, the biggest ε' one may take is therefore $\varepsilon' = \lambda^{\alpha}$. Then

$$\delta t' = O(\lambda^{\alpha(1+1/\ell)}), \tag{2.42}$$

and the cost is

$$\text{cost} = 1/\delta t' = O(\lambda^{-\alpha(1+1/\ell)}). \tag{2.43}$$

For weak convergence ($\alpha = 1$), this cost is identical to (2.35), but for strong convergence ($\alpha = 2$), it is higher by a factor of order $O(\lambda^{-1/\ell})$; i.e., the multiscale scheme is more efficient than a direct calculation with an optimally increased ε. In essence, this is because the multiscale scheme only requires weak convergence of the fast process even when strong convergence of the slow process is sought, whereas the direct scheme with optimized ε leads to either weak or strong convergence of both processes by construction.

It is interesting to corroborate the analysis of the efficiency of the multiscale scheme with the analysis of its consistency. The multiscale scheme is consistent with the limiting equation in (2.4) if $\Delta t/(n_T + N - 1)\delta t \to 0$ as $\Delta t \to 0$, $\delta t \to 0$. But this scaling will not lead to a gain in efficiency in general. On the other hand, because of the way in (2.8) by which we initialize the fast process at each macrotime step, it is easy to see that the multiscale scheme is consistent with (compared with (2.1))

$$\begin{cases} \dot{X}_t = \dfrac{1}{M}\displaystyle\sum_{j=1}^{M} a(X_t, Y_t^j, \varepsilon), & X_0 = x, \\ \dot{Y}_t^j = \dfrac{1}{\varepsilon'} b\big(X_t, Y_t^j, \varepsilon\big) + \dfrac{1}{\sqrt{\varepsilon'}}\sigma\big(X_t, Y_t^j, \varepsilon\big)\dot{W}_t^j, & Y_0^j = 0, \end{cases} \tag{2.44}$$

as $\Delta t \to 0$, $\delta t \to 0$, with $\Delta t/((n_T + N - 1)(\delta t/\varepsilon)) \to \varepsilon'$ (note that it does not matter what happens with n_T, N, and M in this limit, and we may just as well keep these parameters fixed).

This scaling may lead to a gain in efficiency. In particular, using the parameters leading to (2.35) in the weak convergence case, we have

$$\varepsilon' = \frac{\Delta t}{(n_t + N - 1)(\delta t/\varepsilon)} = O(\lambda), \tag{2.45}$$

which is precisely the optimal value of ε we deduced before. In other words, the multiscale scheme can also be thought of as a seamless way to compute with a system where the value of ε has been optimized in terms of the error tolerance. This is a rather remarkable property of the multiscale scheme.

2.5 Numerical Example

Consider the following example with $(X, Y) \in \mathbb{R}^2$:

$$\begin{cases} \dot{X}_t = -Y_t - Y_t^3 + \cos(\pi t) + \sin(\sqrt{2}\pi t), & X_0 = x, \\ \dot{Y}_t = -\dfrac{1}{\varepsilon}\big(Y_t + Y_t^3 - X_t\big) + \dfrac{1}{\sqrt{\varepsilon}}\dot{W}_t, & Y_0 = y. \end{cases} \tag{2.46}$$

From (2.3) the effective equation for X_t is

$$\dot{\bar{X}}_t = -\bar{X}_t + \cos(\pi t) + \sin(\sqrt{2}\pi t), \quad \bar{X}_0 = x. \tag{2.47}$$

Since the error caused by the principle of averaging is independent of the computational parameters $(\Delta t, \delta t, M, N, n_T)$ and the analysis for the error caused by the macrosolver is standard, we only analyze the difference between the numerical solution X_n given by the multiscale scheme and the numerical solution $\bar{X}_n$ given by the macrosolver for (2.47) (in other words, we analyze the error caused by using $\tilde{a}_n$ instead of $\bar{a}(X_n)$). We focus on strong error estimates. Recall that according to Theorem 2.4, in the case when

$$N_m(\delta t/\varepsilon) = (n_T + N - 1)(\delta t/\varepsilon) > 1,$$

we have $R = \Delta t$ plus higher-order terms, and

$$\sup_{n \le T_0/\Delta t} \mathbb{E}\big|X_n - \bar{X}_n\big| \le C\left((\delta t/\varepsilon)^\ell + \frac{e^{-\frac{1}{2}\beta n_T(\delta t/\varepsilon)}\sqrt{\Delta t}}{\sqrt{N(\delta t/\varepsilon)+1}} + \frac{\sqrt{\Delta t}}{\sqrt{M(N(\delta t/\varepsilon)+1)}}\right). \tag{2.48}$$

Here Δt is a fixed parameter. Suppose that we want to bound the error by $O(2^{-p})$ for $p = 0, 1, \dots,$ and assume that $M = 1$. Then, proceeding as in Section 2.4, we deduce that the optimal choice is to take

$$\delta t/\varepsilon = O(2^{-p/\ell}), \quad n_T = O(1), \quad N = O(2^{p(2+1/\ell)}),$$

which leads to a cost scaled as

$$\text{cost} = \frac{n_T + N - 1}{\Delta t} = O(2^{p(2+1/\ell)}).$$

In the numerical calculations, we took

$$(T_0, \Delta t, \delta t/\varepsilon, n_T, M, N) = (6,\ 0.01,\ 0.01 \times 2^{-p/\ell},\ 100,\ 1,\ 10 \times 2^{p(2+1/\ell)}),$$

and we computed the following error estimate for one realization of the solution

$$E_p^\ell = \frac{\Delta t}{T_0} \sum_{n \le \lfloor T_0/\Delta t \rfloor} \big|\tilde{X}_n - \bar{X}_n\big|.$$

A comparison between $\bar{X}_n$ and the solution X_n provided by the multiscale scheme for $\ell = 1$ and $p = 2$ is shown in Figure 2.1. The magnitudes of the errors for

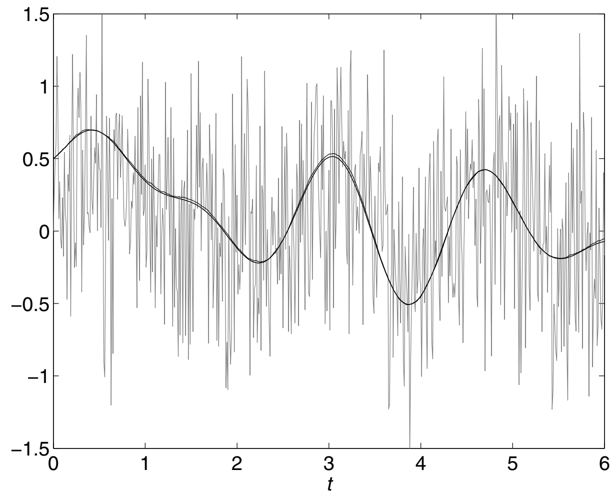

FIGURE 2.1. The comparison between $\bar{X}_n$ and X_n produced by the multiscale scheme with $\ell = 1$, $p = 4$ (black curves). Also shown is the fast process Y_{n,n_T+N} used in the microsolver (gray curve).

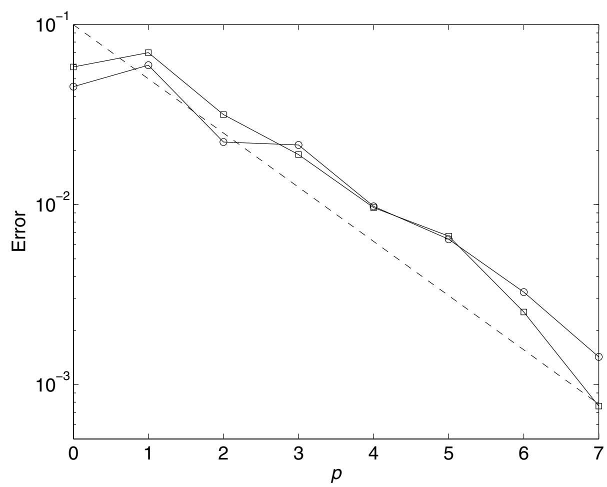

FIGURE 2.2. The error $E_p^\ell = (\Delta t/T_0) \sum_{n \le \lfloor T_0/\Delta t \rfloor} |\tilde{X}_n - \bar{X}_n|$ in the function of p when $\ell = 1$ (circles) and $\ell = 2$ (squares). The dashed line is 0.1×2^{-p}, consistent with the predicted error estimate $E_p^\ell = O(2^{-p})$.

various p and $\ell = 1, 2$ are listed in Table 2.1 and shown in Figure 2.2. As predicted, we observe $E_p^\ell = O(2^{-p})$.

	$p=0$	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=6$	$p=7$
$\ell=1$	.058	.070	.032	.019	.0096	.0067	.0025	.00076
$\ell=2$	.045	.059	.022	.021	.0098	.0064	.0033	.0014

TABLE 2.1. The computed values for the error E_p^ℓ.

3 Diffusive Time Scale

Consider the following equation for $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$:

$$\begin{cases} \dot{X}_t^\varepsilon = \dfrac{1}{\varepsilon} a\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big), & X_0^\varepsilon = x, \\ \dot{Y}_t^\varepsilon = \dfrac{1}{\varepsilon^2} b\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big) + \dfrac{1}{\varepsilon}\sigma\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big)\dot{W}_t, & Y_0^\varepsilon = y. \end{cases} \tag{3.1}$$

We assume that the coefficients satisfy Assumptions 2.1, 2.2, and 2.3. This guarantees that the process $Y_s^{x,\varepsilon}$, the solution of the second equation in (3.1) with fixed $X_t^\varepsilon = x$ and rescaled time $s = t/\varepsilon^2$, is exponentially mixing with unique invariant probability measure $\mu_x^\varepsilon(\cdot)$. In addition, we assume the following:

Assumption 3.1. The coefficients b and σ are of the form

$$\begin{cases} b(x, y, \varepsilon) = b_0(y) + \varepsilon b_1(x, y, \varepsilon), \\ \sigma(x, y, \varepsilon) = \sigma_0(y) + \varepsilon\sigma_1(x, y, \varepsilon). \end{cases}$$

Notice that by this assumption, $Y_s^{x,\varepsilon=0}$ and $\mu = \mu_x^{\varepsilon=0}$ are independent of x.

We assume the following centering condition:

Assumption 3.2.

$$\forall (x, \varepsilon) : \int_{\mathbb{R}^m} a(x, y, \varepsilon)\mu(dy) = 0. \tag{3.2}$$

These two assumptions make the multiscale scheme simpler and facilitate the analysis of its convergence properties but are not essential and will be relaxed in Section 3.5.

To give the effective dynamics for X_t^ε when ε is small, it will be convenient to define for each (x, ε) the following auxiliary processes (Y_t^1, Y_t^2):

$$\begin{cases} \dot{Y}_t^1 = \dfrac{1}{\varepsilon^2} b_0\big(Y_t^1\big) + \dfrac{1}{\varepsilon}\sigma_0\big(Y_t^1\big)\dot{W}_t, & Y_0^1 = y_1, \\ \dot{Y}_t^2 = \dfrac{1}{\varepsilon^2}\partial b_0\big(Y_t^1\big)Y_t^2 + \dfrac{1}{\varepsilon}\partial\sigma_0\big(Y_t^1\big)Y_t^2\dot{W}_t & \\ \qquad + \dfrac{1}{\varepsilon^2} b_1\big(x, Y_t^1, \varepsilon\big) + \dfrac{1}{\varepsilon}\sigma_1\big(x, Y_t^1, \varepsilon\big)\dot{W}_t, & Y_0^2 = y_2. \end{cases} \tag{3.3}$$

Assumptions 2.2 and 2.3 imply that process Y_t^1 is exponentially mixing with unique invariant probability measure $\mu(dy_1)$ defined as above. It is proven in Appendix B that for each (x, ε), the process (Y_t^1, Y_t^2) defined by (3.3) is exponentially mixing with unique invariant probability measure $\nu_x^\varepsilon(dy_1, dy_2)$.

It will be shown later that X_t^ε converges to the solution of the following stochastic differential equation:

$$\dot{\bar{X}}_t = \bar{a}\big(\bar{X}_t\big) + \bar{\sigma}\big(\bar{X}_t\big)\dot{W}_t, \quad \bar{X}_0 = x, \tag{3.4}$$

where W_t is an n-dimensional Wiener process and

$$\begin{cases} \bar{a}(x) = \lim\limits_{\varepsilon\to 0}\displaystyle\int_{\mathbb{R}^m\times\mathbb{R}^m} \nu_x^\varepsilon(dy_1, dy_2)\partial_y a(x, y_1, \varepsilon)y_2 \\ \qquad + \lim\limits_{\varepsilon\to 0}\displaystyle\int_{\mathbb{R}^m} \mu(dy_1)\int_0^\infty \mathbb{E}_{y_1}\,\partial_x a(x, Y^1_{\varepsilon^2 s}, \varepsilon)a(x, y_1, \varepsilon)ds, \\ \bar{\sigma}(x)\bar{\sigma}^T(x) = \\ \qquad 2\lim\limits_{\varepsilon\to 0}\displaystyle\int_{\mathbb{R}^m} \mu(dy_1)a(x, y_1, \varepsilon)\otimes\int_0^\infty \mathbb{E}_{y_1} a\big(x, Y^1_{\varepsilon^2 s}, \varepsilon\big)ds. \end{cases} \tag{3.5}$$

We will assume $\bar{\sigma}$ is well-defined and belongs to $\mathbb{C}_b^\infty$.

3.1 The Numerical Scheme

The scheme for (3.1) consists of a macrosolver for (3.4), a microsolver for (Y_t^1, Y_t^2), and an estimator for $\bar{a}(\cdot)$ and $\bar{\sigma}(\cdot)$. For the macrosolver, we may use any stable explicit solver, such as (in the simplest case) the forward Euler method:

$$X_{n+1} = X_n + \tilde{a}_n\Delta t + \tilde{\sigma}_n\tilde{\xi}_{n+1}\sqrt{\Delta t}, \tag{3.6}$$

where $\{\tilde{\xi}_n\}$ are i.i.d. Gaussian with mean 0, variance 1, and independent of the ones used in the microscheme, and $\tilde{a}_n$ and $\tilde{\sigma}_n$ are the approximations of $\bar{a}(X_n)$ and $\bar{\sigma}(X_n)$ provided by the estimator. For the microsolver for (3.3), we may use a first- or second-order scheme, similar to (2.6) or (2.7) (note that the microtime step will now appear only as the ratio $\delta t/\varepsilon^2$ in these schemes). In order to estimate the expectation in (3.5) (see (3.8) below), we integrate these equations over n_T+N+N' microtime steps, and we reinitialize the fast variables as in (2.8) at each macrotime step; i.e., we take

$$Y^1_{n,0} = Y^1_{n-1,n_T+N+N'-1}, \quad Y^2_{n,0} = Y^2_{n-1,n_T+N+N'-1}. \tag{3.7}$$

For the estimator, we use the following time and ensemble average:

$$\begin{cases} \tilde{a}_n = \dfrac{1}{MN}\displaystyle\sum_{j=1}^{M}\sum_{m=n_T}^{n_T+N-1} \partial_y a\big(X_n, Y^1_{n,m,j}, \varepsilon\big) Y^2_{n,m,j} \\ \qquad + \dfrac{(\delta t/\varepsilon^2)}{MN}\displaystyle\sum_{j=1}^{M}\sum_{m=n_T}^{n_T+N-1}\sum_{m'=0}^{N'} \partial_x a\big(X_n, Y^1_{n,m+m',j}, \varepsilon\big) a\big(X_n, Y^1_{n,m,j}, \varepsilon\big), \\ \tilde{B}_n = \dfrac{2(\delta t/\varepsilon^2)}{MN}\displaystyle\sum_{j=1}^{M}\sum_{m=n_T}^{n_T+N-1}\sum_{m'=0}^{N'} a\big(X_n, Y^1_{n,m,j}, \varepsilon\big) \otimes a\big(X_n, Y^1_{n,m+m',j}, \varepsilon\big). \end{cases} \tag{3.8}$$

$\tilde{\sigma}_n$ is obtained by Cholesky decomposition of $\tilde{B}_n$ so that $\tilde{\sigma}_n \tilde{\sigma}_n^T = \tilde{B}_n$. Here n_T is the number of microtime steps that we skip to eliminate transients, N is the number of microtime steps that we use for time averaging, and N' is the number of microtime steps we use to estimate the integrals over s in (3.5). M is the number of realizations of the fast auxiliary processes (Y^1_t, Y^2_t).

3.2 Convergence of the Scheme

THEOREM 3.3 *Suppose Δt and $\delta t/\varepsilon^2$ are sufficiently small. Then for any $f \in \mathbb{C}_0^\infty$ and $T_0 > 0$, there exists a constant C independent of the parameters $(\varepsilon, \Delta t, \delta t, n_T, M, N, N')$ such that*

$$\begin{aligned} &\sup_{n \le T_0/\delta t} \big|\mathbb{E} f\big(X^\varepsilon_{t_n}\big) - \mathbb{E} f(X_n)\big| \\ &\quad \le C\big(\varepsilon + \Delta t + (\delta t/\varepsilon^2)^\ell + e^{-\frac{1}{2}\beta N'(\delta t/\varepsilon^2)}\big) \\ &\qquad + C\bigg(\frac{e^{-\frac{1}{2}\beta n_T(\delta t/\varepsilon^2)}}{\sqrt{N(\delta t/\varepsilon^2)+1}}\bar{R} + \frac{\Delta t}{M(N(\delta t/\varepsilon^2)+1)}\bigg), \end{aligned} \tag{3.9}$$

where

$$\bar{R} = \frac{\sqrt{\Delta t}}{1 - e^{-\frac{1}{2}\beta(n_T+N+N'-1)(\delta t/\varepsilon^2)}}.$$

We divide the estimate of $|\mathbb{E} f(X^\varepsilon_t) - f(X_n)|$ into two parts:

(1) $|\mathbb{E} f(X^\varepsilon_t) - \mathbb{E} f(\bar{X}_t)|$ and
(2) $|\mathbb{E} f(\bar{X}_{t_n}) - \mathbb{E} f(X_n)|$.

For the first part, it is known [12, 24] that

$$\sup_{0 \le t \le T_0} \big|\mathbb{E} f\big(X^\varepsilon_t\big) - \mathbb{E} f\big(\bar{X}_t\big)\big| \le C\varepsilon,$$

which gives rise to the first term in (3.9). We give a formal derivation of this result at the end of this section. Now we estimate the second part.

Using Lemmas B.2 and B.3 and repeating the analysis of Lemma 2.6, we can show that for each $T_0 > 0$, there exists an independent constant C such that $\forall n \in [0, T_0/\Delta t]$,

$$\begin{aligned}
&\mathbb{E}|\mathbb{E}_{X_n}\tilde{a}_n - \bar{a}(X_n)|^2 + \mathbb{E}\big\|\mathbb{E}_{X_n}\tilde{B}_n - \bar{\sigma}(X_n)\bar{\sigma}^T(X_n)\big\|^2 \\
&\le C\big(\varepsilon^2 + (\delta t/\varepsilon^2)^{2\ell} + e^{-\beta N'(\delta t/\varepsilon^2)}\big) \\
&\quad + C\frac{e^{-\beta n_T(\delta t/\varepsilon^2)}}{N(\delta t/\varepsilon^2)+1}\big(e^{-\beta n(n_T+N+N'-1)(\delta t/\varepsilon^2)} + \bar{R}^2\big),
\end{aligned} \tag{3.10}$$

and

$$\begin{aligned}
&\mathbb{E}|\tilde{a}_n - \bar{a}(X_n)|^2 + \mathbb{E}\|\tilde{B}_n - \bar{\sigma}(X_n)\bar{\sigma}^T(X_n)\|^2 \\
&\quad \le C\big(\varepsilon^2 + (\delta t/\varepsilon^2)^{2\ell} + e^{-\beta N'(\delta t/\varepsilon^2)}\big) \\
&\qquad + C\frac{e^{-\beta n_T(\delta t/\varepsilon^2)}}{N(\delta t/\varepsilon^2)+1}\big(e^{-\beta n(n_T+N+N'-1)(\delta t/\varepsilon^2)} + \bar{R}^2\big) \\
&\qquad + C\frac{1}{M(N(\delta t/\varepsilon^2)+1)}.
\end{aligned} \tag{3.11}$$

Based on these estimates, we have the following:

LEMMA 3.4 *For any $f \in \mathbb{C}_0^\infty$ and $T_0 > 0$, there exists an independent constant C such that*

$$\begin{aligned}
&\sup_{n\le T_0/\delta t}\big|\mathbb{E}f\big(\bar{X}_{t_n}\big) - \mathbb{E}f(X_n)\big| \\
&\quad \le C\big(\varepsilon + \Delta t + (\delta t/\varepsilon^2)^{\ell} + e^{-\frac{1}{2}\beta N'(\delta t/\varepsilon^2)}\big) \\
&\qquad + C\bigg(\frac{e^{-\frac{1}{2}\beta n_T(\delta t/\varepsilon^2)}}{\sqrt{N(\delta t/\varepsilon^2)+1}}\bar{R} + \frac{\Delta t}{\sqrt{M(N(\delta t/\varepsilon^2)+1)}}\bigg).
\end{aligned} \tag{3.12}$$

PROOF: Define the function $u(k, x)$ for $k \le n$ similarly as in the proof of Lemma 2.8:

$$u(n, x) = f(x), \quad u(k, x) = \mathbb{E}\big(u(k+1, x + \Delta t\bar{a}(x) + \sqrt{\Delta t}\,\bar{\sigma}(x)\xi_n\big).$$

It is easy to show by the smoothness of $\bar{a}$, $\bar{\sigma}$, and f and the compactness of f that $u(k, x)$ is a smooth function of x with uniformly bounded derivatives for all k. By Taylor expansion we then have

$$\begin{aligned}
&\big|\mathbb{E}\big(u(k+1, X_{k+1}) - u(k, X_k)\big)\big| \\
&\quad = \big|\mathbb{E}\big(u\big(k+1, X_k + \Delta t\tilde{a}_k + \sqrt{\Delta t}\tilde{\sigma}_k\tilde{\xi}_k\big) \\
&\qquad - u\big(k+1, X_k + \Delta t\bar{a}(X_k) + \sqrt{\Delta t}\,\bar{\sigma}(X_k)\xi_k\big)\big|
\end{aligned}$$

$$
\begin{aligned}
&\le \Delta t \mathbb{E}|\partial_x u(k+1, X_k)|\,|\mathbb{E}_{X_k}\tilde{a}_k - \bar{a}(X_k)| \\
&\quad + \frac{1}{2}\Delta t^2 \mathbb{E}\big|\partial_x^2 u(k+1, X_k)\big|\,\big|\mathbb{E}_{X_k}\tilde{a}_k^2 - \bar{a}^2(X_k)\big| \\
&\quad + \frac{1}{2} C\Delta t \mathbb{E}\big|\partial_x^2 u(k+1, X_k)\big|\,\big|\mathbb{E}_{X_k}\tilde{B}_k - \bar{\sigma}(X_k)\bar{\sigma}^T(X_k)\big| \\
&\quad + \frac{1}{6}\mathbb{E}\big|\partial_x^3 u(k+1, X_k)\big|\,\big|\mathbb{E}_{X_k}\big(\Delta t\big(\tilde{a}_k - \bar{a}(X_k)\big) + \sqrt{\Delta t}\big(\tilde{\sigma}_k\tilde{\xi}_k - \bar{\sigma}(X_k)\xi_k\big)\big)^3\big| \\
&\quad + \frac{1}{12}\mathbb{E}\big|\partial_x^4 u(k+1, y_k)\big|\,\mathbb{E}\big|\Delta t\big(\tilde{a}_k - \bar{a}(X_k)\big) + \sqrt{\Delta t}\big(\tilde{\sigma}_k\tilde{\xi}_k - \bar{\sigma}(X_k)\xi_k\big)\big|^4 \\
&\le C\Delta t\big(\mathbb{E}\big|\mathbb{E}_{X_k}\tilde{a}_k - \bar{a}(X_k)\big|^2 + \mathbb{E}\big|\mathbb{E}_{X_k}\tilde{B}_k - \bar{\sigma}(X_k)\bar{\sigma}^T(X_k)\big|^2\big)^{1/2} \\
&\quad + C\Delta t^2\big(\mathbb{E}|\tilde{a}_k - \bar{a}(X_k)|^2 + \mathbb{E}\big|\tilde{B}_k - \bar{\sigma}(X_k)\bar{\sigma}^T(X_k)\big|^2\big)^{1/2} + C\Delta t^2,
\end{aligned}
$$

where

$$
y_k = X_k + \theta_k\big(\Delta t\tilde{a}_k + \sqrt{\Delta t}\tilde{\sigma}_k\tilde{\xi}_k - \Delta t\bar{a}(X_k) - \sqrt{\Delta t}\bar{\sigma}(X_k)\xi_k\big)
$$

for some $\theta_k \in [0, 1]$. Hence, using (3.10) and (3.11), we deduce

$$
\begin{aligned}
\big|\mathbb{E}f(X_n) - f\big(\bar{X}_n\big)\big| &= |\mathbb{E}u(n, X_n) - u(0, x)| \\
&= \Bigg|\sum_{0\le k\le n-1}\mathbb{E}(u(k+1, X_{k+1}) - u(k, X_k))\Bigg| \\
&\le C\big(\varepsilon + \Delta t + (\delta t/\varepsilon^2)^\ell + e^{-\frac{1}{2}\beta N'(\delta t/\varepsilon^2)}\big) \\
&\quad + C\Bigg(\frac{e^{-\frac{1}{2}\beta n_T(\delta t/\varepsilon^2)}}{\sqrt{N(\delta t/\varepsilon^2)+1}}\bar{R} + \frac{\Delta t}{\sqrt{M(N(\delta t/\varepsilon^2)+1)}}\Bigg).
\end{aligned}
$$

Since

$$
\big|\mathbb{E}\big(f\big(\bar{X}_{t_n}\big) - f\big(\bar{X}_n\big)\big)\big| \le C\Delta t,
$$

(3.12) follows. □

Now we give a formal derivation for the convergence of the solution X_t^ε of (3.1) to the solution $\bar{X}_t$ of (3.4) by perturbation analysis. Letting $Z_t^\varepsilon = (Y_t^\varepsilon - Y_t^1)/\varepsilon$, (3.1) can be written as the following enlarged system:

$$
\begin{cases}
\dot{X}_t^\varepsilon = \dfrac{1}{\varepsilon}a\big(X_t^\varepsilon, Y_t^1, 0\big) + \partial_y a\big(X_t^\varepsilon, Y_t^1, 0\big)Z_t^\varepsilon + \partial_\varepsilon a\big(X_t^\varepsilon, Y_t^1, 0\big) + \varepsilon E \\
\dot{Y}_t^1 = \dfrac{1}{\varepsilon^2}b_0\big(Y_t^1\big) + \dfrac{1}{\varepsilon}\sigma_0\big(Y_t^1\big)\dot{W}_t, \\
\dot{Z}_t^\varepsilon = \dfrac{1}{\varepsilon^2}\partial b_0\big(Y_t^1\big)Z_t^\varepsilon + \dfrac{1}{\varepsilon^2}b_1\big(X_t^\varepsilon, Y_t^1, 0\big) + \dfrac{1}{\varepsilon}F \\
\qquad + \dfrac{1}{\varepsilon}\partial\sigma_0\big(Y_t^{1,\varepsilon}\big)Z_t^\varepsilon\dot{W}_t + \dfrac{1}{\varepsilon}\sigma_1\big(X_t^\varepsilon, Y_t^1, 0\big)\dot{W}_t + G\dot{W}_t,
\end{cases}
\tag{3.13}
$$

with $X_0^\varepsilon = x$, $Y_0^1 = y$, and $Z_0^\varepsilon = 0$. Furthermore, $E(\cdot)$, $F(\cdot)$, and $G(\cdot)$ are appropriate functions of $(x, y_1, y_2, \varepsilon)$ whose actual values are not important for the limiting equation. The generator of this enlarged system can be written as

$$L^\varepsilon = \frac{1}{\varepsilon^2} L_1 + \frac{1}{\varepsilon} L_2 + L_3 + \varepsilon L_4,$$

where

$$\begin{cases} L_1 = b_0(y_1)\dfrac{\partial}{\partial y_1} + (\partial b_0(y_1)z + b_1(x, y_1, 0))\dfrac{\partial}{\partial z} + \dfrac{1}{2} A A^\mathsf{T} \dfrac{\partial^2}{\partial y^2} \\ \qquad \big(A = \mathrm{diag}(\sigma_0(y_1),\ \partial\sigma_0(y_1)z + \sigma_1(x, y_1, 0)),\ y = (y_1, z)\big), \\ L_2 = a(x, y_1, 0)\dfrac{\partial}{\partial x} + L_2^z, \\ L_3 = \big(\partial_y a(x, y_1, 0)z + \partial_\varepsilon a(x, y_1, 0)\big)\dfrac{\partial}{\partial x} + \dfrac{1}{2} G G^\mathsf{T} \dfrac{\partial^2}{\partial z^2}, \\ L_4 = E\dfrac{\partial}{\partial x}, \end{cases}$$

and L_2^z is a differential operator in z.

Notice that L_1/ε^2 is the infinitesimal generator of process (Y_t^1, Y_t^2) defined by (3.3) with b_1 and σ_1 evaluated at $\varepsilon = 0$. By Lemma B.1 given in the appendix, L_1 generates an exponentially mixing process, i.e., for $f \in \mathbb{C}_b^\infty$,

$$\left| e^{L_1 t} f - P_x f \right| \le B\big(|y_1|^2 + |z|^2 + 1\big) e^{-\beta t};$$

here

$$P_x f = \int_{\mathbb{R}^m \times \mathbb{R}^m} f(x, y_1, z)\nu_x(dy_1, dz).$$

where $\nu_x(dy_1, dy_2) = \nu_x^{\varepsilon=0}(dy_1, dy_2)$ is the invariant measure for (3.3) with b_1 and σ_1 evaluated at $\varepsilon = 0$. Assumption 3.2 implies that

$$PL_2P = 0. \tag{3.14}$$

The equation $u^\varepsilon(t, x, y_1, y_2) = \mathbb{E}_{x, y_1, y_2} f(X_t^\varepsilon)$ satisfies the following equation:

$$\frac{\partial u^\varepsilon}{\partial t} = L^\varepsilon u^\varepsilon, \quad u^\varepsilon(0) = f. \tag{3.15}$$

Represent u^ε in the power series

$$u^\varepsilon = u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \cdots .$$

Inserting this into (3.15) and equating coefficients of different powers of ε, we have

$$L_1 u_0 = 0, \quad L_1 u_1 = -L_2 u_0, \quad L_1 u_2 = -L_2 u_1 - L_3 u_0 + \frac{\partial u_0}{\partial t}, \quad \ldots .$$

Suppose that $u_0(0) = Pu_0(0)$. Projecting onto P, by the solvability condition (3.14), we have

$$\frac{\partial u_0}{\partial t} = \big(PL_3P - PL_2L_1^{-1}L_2P\big)u_0 = \bar{L}u_0, \quad u_0(0) = f, \tag{3.16}$$

and

$$u_1 = -L_1^{-1}L_2u_0, \quad u_2 = -L_1^{-1}\big(L_3 - L_2L_1^{-1}L_2 - \bar{L}\big)u_0.$$

By definition and uniqueness of the invariant measures, for any bounded function f,

$$\int_{\mathbb{R}^m} f(y_1)\mu(dy_1) = \int_{\mathbb{R}^m\times\mathbb{R}^m} f(y_1)\nu_x^\varepsilon(dy_1, dz).$$

Assumption 3.2 implies that

$$\int_{\mathbb{R}^m} \partial_\varepsilon a(x, y_1, 0)\mu(dy_1) = \int_{\mathbb{R}^m\times\mathbb{R}^m} \partial_\varepsilon a(x, y_1, 0)\nu_x^\varepsilon(dy_1, dz) = 0. \tag{3.17}$$

A direct computation with (3.17) and Lemma B.3 shows that

$$\bar{L} = \bar{a}\frac{\partial}{\partial x} + \frac{1}{2}\bar{\sigma}\bar{\sigma}^T\frac{\partial^2}{\partial x^2}.$$

By the same “bootstrap” argument as in Section 2.3, on finite intervals, as long as u_0, u_1, and u_2 have bounded solutions, we have

$$u^\varepsilon - u_0 = O(\varepsilon).$$

The boundedness of u_0, u_1, and u_2 are guaranteed by the smoothness of the coefficients and exponential mixing.

Remark 3.5. Weak convergence to the effective dynamics is implied by the above analysis even if the smoothness assumptions on the coefficients are not satisfied, as in the numerical example below.

3.3 Efficiency and Consistency Analysis

We proceed as in Section 2.4. At fixed error tolerance λ, assuming that $\lambda \gg \varepsilon$, we will see that the multiscale scheme is then appropriate. Due to the fact that the sampling error is dominated by the macrotime discretization error in (3.9), the optimal choice of parameters is

$$\Delta t = O(\lambda), \qquad \delta t/\varepsilon^2 = O(\lambda^{1/\ell}), \tag{3.18}$$

$$M = N = 1, \qquad n_T = N' = O(\lambda^{1/\ell}\log\lambda^{-1}). \tag{3.19}$$

This leads to

$$\text{cost} = \frac{M(n_T + 1 + N')}{\Delta t} = O(\lambda^{-1-1/\ell}\log\lambda^{-1}). \tag{3.20}$$

In comparison, a direct scheme for (3.1) with weak order ℓ (same as in the microsolver used in the multiscale scheme) leads to an error estimate as

$$\sup_{n\le T_0/\delta t} \left|\mathbb{E}f\left(X^\varepsilon_{t_n}\right) - \mathbb{E}f\left(X^\varepsilon_n\right)\right| \le C(\delta t/\varepsilon^2)^\ell, \tag{3.21}$$

where X^ε_n is the numerical approximation provided by the direct scheme. At error tolerance λ, a time step $\delta t = O(\varepsilon^2\lambda^{1/\ell})$, and the cost is $1/\delta t = O(\varepsilon^{-2}\lambda^{-1/\ell})$. This is much more expensive than the multiscale scheme when $\varepsilon \ll \lambda$.

As in Section 2.4, we can compare the cost of the multiscale scheme to that of a direct scheme for (3.1) where ε is chosen optimally as a function of the error tolerance. The error estimate for such a direct scheme when ε is increased to the value ε' is

$$\sup_{n\le T_0/\delta t} \left|\mathbb{E}f\left(X^\varepsilon_{t_n}\right) - \mathbb{E}f\left(X^{\varepsilon'}_n\right)\right| \le C(\varepsilon' + (\delta t/\varepsilon'^2)^\ell), \tag{3.22}$$

Thus as optimal parameters we should take $\varepsilon' = \lambda$, a time step of order

$$\delta t = O(\lambda^{2+1/\ell}), \tag{3.23}$$

and the cost is

$$\text{cost} = 1/\delta t = O\left(\lambda^{-2-1/\ell}\right). \tag{3.24}$$

This cost is still higher by a factor of order $O(\lambda^{-1})$ than the one in (3.20) of the multiscale scheme.

3.4 Numerical Example

Consider the following equation:

$$\begin{cases} \dot X^\varepsilon_t = -\dfrac{2}{\varepsilon} Y^\varepsilon_t Z^\varepsilon_t, & X^\varepsilon_0 = x, \\ \dot Y^\varepsilon_t = -\dfrac{1}{\varepsilon^2} Y^\varepsilon_t + \dfrac{1}{\varepsilon} X^\varepsilon_t Z^\varepsilon_t + \dfrac{1}{\varepsilon}\dot W^1_t, & Y^\varepsilon_0 = y, \\ \dot Z^\varepsilon_t = -\dfrac{2}{\varepsilon^2} Z^\varepsilon_t + \dfrac{1}{\varepsilon} X^\varepsilon_t Y^\varepsilon_t + \dfrac{1}{\varepsilon}\dot W^2_t, & Z^\varepsilon_0 = z. \end{cases}$$

The fast time scale processes are given by the following dynamics:

$$\begin{cases} \dot Y^1_t = -\dfrac{1}{\varepsilon^2} Y^1_t + \dfrac{1}{\varepsilon}\dot W^1_t, & Y^1_0 = y_1, \\ \dot Y^2_t = -\dfrac{1}{\varepsilon^2} Y^2_t + \dfrac{1}{\varepsilon^2} x Z^1_t, & Y^2_0 = y_2, \\ \dot Z^1_t = -\dfrac{2}{\varepsilon^2} Z^1_t + \dfrac{1}{\varepsilon}\dot W^2_t, & Z^1_0 = z_1, \\ Z^2_t = -\dfrac{2}{\varepsilon^2} Z^2_t + \dfrac{1}{\varepsilon^2} x Y^1_t, & Z^2_0 = z_2. \end{cases}$$

The coefficients of the effective dynamics are

$$\begin{cases} \bar{a}(x) = \int_{\mathbb{R}^4} (-2y_1z_2 - 2z_1y_2)\mu_x(dy_1, dz_1, dy_2, dz_2) = -\frac{1}{2}x, \\ \bar{\sigma}^2(x) = 2\int_{\mathbb{R}^2} \mu(dy_1, dz_1)(2y_1z_1)\int_0^\infty \mathbb{E}(2Y^1_{\varepsilon^2 s} Z^1_{\varepsilon^2 s})ds = \frac{1}{3}; \end{cases}$$

i.e., the effective equation is

$$\dot{\bar{X}}_t = -\frac{1}{2}\bar{X}_t + \frac{1}{\sqrt{3}}\dot{W}_t. \tag{3.25}$$

Since the error caused by the principle of averaging and macrotime discretization is standard, we only compute the error caused by using $(\tilde{a}, \tilde{\sigma})$ instead of $(\bar{a}, \bar{\sigma})$ in the scheme. In the case when

$$(n_T + N - 1)\delta t > 1,$$

we have $\bar{R} \approx \sqrt{\Delta t}$. Relation (3.11) and its proof imply that for fixed Δt,

$$\begin{aligned} &\sup_{n \le T_0/\delta t} \mathbb{E}|\tilde{a}_n - \bar{a}(X_n)| + \mathbb{E}\big\|\tilde{B}_n - \bar{\sigma}(X_n)\bar{\sigma}^T(X_n)\big\| \\ &\quad \le C\big((\delta t/\varepsilon^2)^\ell + e^{-\beta N'(\delta t/\varepsilon^2)/2}\big) \\ &\qquad + C\left(\frac{e^{-\beta n_T(\delta t/\varepsilon^2)/2}}{\sqrt{N(\delta t/\varepsilon^2) + 1}} + \frac{1}{\sqrt{M(N(\delta t/\varepsilon^2) + 1)}}\right). \end{aligned} \tag{3.26}$$

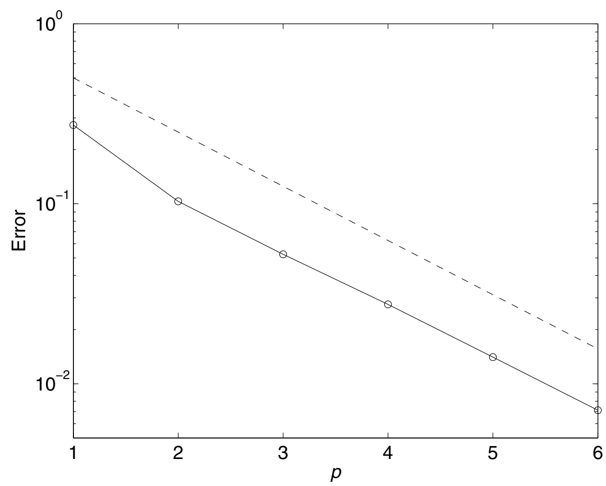

FIGURE 3.1. The error $E_p^\ell = (\Delta t/T_0)\sum_{n \le \lfloor T_0/\Delta t \rfloor} |\tilde{a}_n + \frac{1}{2}X_n| + |\tilde{\sigma}_n - \frac{1}{\sqrt{3}}|$ in function of p when $\ell = 1$ (circles). Also shown is the predicted error estimate 2^{-p} (dashed line).

Suppose, assuming $M = 1$, we want to bound the error by 2^{-p} for $p = 0, 1, \ldots$. Then with the same analysis as before, the optimal choices for the parameters can be given as

$$\delta t/\varepsilon^2 = O(2^{-p/\ell}), \quad n_T = O(1), \quad N = O(2^{p(2+1/\ell)}), \quad N' = O(2^{p/\ell} p),$$

which leads to a cost scaled as

$$\text{cost} = \frac{M(n_T + N + N')}{\Delta t} = O(2^{p(2+1/\ell)}).$$

In the numerical experiments, we took

$$(T_0, \Delta t, \delta t/\varepsilon^2, N_T, M, N, N') = (1, .001, 2^{-p/\ell}, 16, 1, 10 \times 2^{p(2+1/\ell)}, 2^{p/\ell} p), \tag{3.27}$$

and computed for one realization of the solution the following error between the $(-\frac{1}{2}X_n, \frac{1}{\sqrt{3}})$ of $(\tilde{a}_n, \tilde{\sigma}_n)$:

$$E_p^\ell = \frac{\Delta t}{T_0} \sum_{n \le \lfloor T_0/\Delta t \rfloor} \left| \tilde{a}_n + \frac{1}{2} X_n \right| + \left| \tilde{\sigma}_n - \frac{1}{\sqrt{3}} \right|.$$

We choose the microsolver to be the first-order scheme (2.6). The magnitudes of the above error are listed in Table 3.1 and shown in Figure 3.1.

	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 6$
$\ell = 1$	.274	.103	.052	.028	.014	.0071

TABLE 3.1. The computed values for the error E_p^ℓ.

3.5 Generalizations

In this section, we want to discuss two more general cases of equation (3.1). The first is when the centering assumption, Assumption 3.2, is not satisfied. In this case the effective dynamics for small ε can be expressed in the following form [17]:

$$\dot{\bar{X}}_t = \bar{a}(\bar{X}_t, \varepsilon) + \bar{\sigma}(\bar{X}_t)\dot{W}_t, \tag{3.28}$$

where

$$\begin{cases} \bar{a}(x,\varepsilon) = \dfrac{1}{\varepsilon}\displaystyle\int_{\mathbb{R}^m} \mu(dy_1)a(x,y_1,\varepsilon) \\ \qquad + \displaystyle\int_{\mathbb{R}^m\times\mathbb{R}^m} \nu_x^\varepsilon(dy_1,dy_2)\partial_y a(x,y_1,\varepsilon)y_2 \\ \qquad + \displaystyle\int_{\mathbb{R}^m} \mu(dy_1)\int_0^\infty \Big(\mathbb{E}_{y_1}\partial_x a\big(x,Y^1_{\varepsilon^2 s},\varepsilon\big) \\ \qquad\qquad - \displaystyle\int_{\mathbb{R}^m} \mu(dy_1)\partial_x a(x,y_1,\varepsilon)\Big)a(x,y_1,\varepsilon)\,ds \\ \bar{\sigma}(x)\bar{\sigma}^T(x) = \displaystyle\lim_{\varepsilon\to 0} 2\int_{\mathbb{R}^m} \mu(dy_1)a(x,y_1,\varepsilon) \\ \qquad \otimes \displaystyle\int_0^\infty \Big(\mathbb{E}_{y_1} a\big(x,Y^1_{\varepsilon^2 s},\varepsilon\big) - \int_{\mathbb{R}^m}\mu(dy_1)a(x,y_1,\varepsilon)\Big)ds. \end{cases}$$

Notice that the above formula is no more complicated than (3.5). So the same scheme as before can be used with minor modifications.

The second case of interest is when the principal component of the fast dynamics depends on the slow dynamics. In other words,

$$b_0 = b_0(x,y), \quad \sigma_0 = \sigma_0(x,y). \tag{3.29}$$

Just for simplicity, we assume (3.2). In this case the effective dynamics has the following form:

$$\dot{\bar{X}}_t = \bar{a}\big(\bar{X}_t\big) + \bar{\sigma}\big(\bar{X}_t\big)\dot{W}_t, \tag{3.30}$$

with

$$\begin{cases} \bar{a}(x) = \displaystyle\lim_{\varepsilon\to 0}\int_{\mathbb{R}^m\times\mathbb{R}^m} \nu_x^\varepsilon(dy_1,dy_2)\partial_y a(x,y_1,\varepsilon)y_2 \\ \qquad + \displaystyle\int_{\mathbb{R}^m} \mu_x(dy_1)\int_0^\infty \mathbb{E}_{y_1}\Big(\partial_x a\big(x,Y^1_{\varepsilon^2 s},\varepsilon\big) \\ \qquad\qquad + \partial_y a\big(x,Y^1_{\varepsilon^2 s},\varepsilon\big)U_{\varepsilon^2 s}\Big)a(x,y_1,\varepsilon)\,ds \\ \bar{\sigma}(x)\bar{\sigma}^T(x) = \displaystyle\lim_{\varepsilon\to 0} 2\int_{\mathbb{R}^m} \mu_x(dy_1)a(x,y_1,\varepsilon)\otimes\int_0^\infty \mathbb{E}_{y_1}a(x,Y^1_{\varepsilon^2 s},\varepsilon)\,ds, \end{cases}$$

where $U_t = \partial_x Y_t^1 \in \mathbb{R}^m \times \mathbb{R}^n$ is the process satisfying the following dynamics:

$$\begin{aligned} \dot{U}_t &= \frac{1}{\varepsilon^2}\partial_x b_0\big(x,Y_t^1\big) + \frac{1}{\varepsilon^2}\partial_y b_0\big(x,Y_t^1\big)U_t \\ &\quad + \frac{1}{\varepsilon}\partial_x\sigma_0\big(x,Y_t^1\big)\dot{W}_t + \frac{1}{\varepsilon}\partial_y\sigma_0\big(x,Y_t^1\big)U_t\,\dot{W}_t. \end{aligned}$$

Provided the stability condition such that the above integrals exist, the multiscale scheme can also be applied to this case.

Appendix A: Limiting Properties—Advective Time Scale

Here we give some limiting properties of the auxiliary process $Y_t^{x,\varepsilon}$ defined by (2.2) and its time discretization. We assume that Assumptions 2.1, 2.2, and 2.3 hold.

Taking $y_1 = y$ and $y_2 = 0$ in Assumption 2.3 and using Assumption 2.1, we deduce that for some positive constant C,

$$y \cdot b(x, y, \varepsilon) \leq -\frac{\beta}{2}|y|^2 + C(|x|^2 + \varepsilon^2 + 1). \tag{A.1}$$

Using this inequality and Itô's formula, it is easy to check that for any $p \geq 1$, $V(y) = |y|^{2p}$ is a Lyapunov function for $Y_t^{x,\varepsilon}$ in the sense that

$$LV(y) \leq -\frac{\beta}{\varepsilon}V(y) + \frac{1}{\varepsilon}H(x, \varepsilon), \tag{A.2}$$

where L is the infinitesimal generator of $Y_t^{x,\varepsilon}$ and $H(x, \varepsilon)$ is a positive smooth function. This implies that

$$\limsup_{t\to\infty} \mathbb{E}V\left(Y_t^{x,\varepsilon}\right) \leq H(x, \varepsilon). \tag{A.3}$$

By theorem 6.1 in [21] (see also [18]), $Y_t^{x,\varepsilon}$ is exponentially mixing with unique invariant probability measure $\mu_x^\varepsilon(\cdot)$ in the following sense: For each (x, ε) and $p \in \mathbb{N}$, there exist positive constants B and κ such that for any function $f : \mathbb{R}^m \to \mathbb{R}$ with $|f(y)| \leq |y|^{2p} + 1$,

$$\left|\mathbb{E}f\left(Y_t^{x,\varepsilon}\right) - \int_{\mathbb{R}^m} \mu_x^\varepsilon(dy) f(y)\right| \leq B(|y|^{2p} + 1)e^{-\kappa t/\varepsilon}, \tag{A.4}$$

where $y = Y_{t=0}^{x,\varepsilon}$.

For each (x, ε), we can construct an independent random variable $\zeta^{x,\varepsilon}$ whose law is $\mu_x^\varepsilon(\cdot)$, i.e., $\mathcal{L}(\zeta^{x,\varepsilon}) = \mu_x^\varepsilon(\cdot)$. Let $\zeta_t^{x,\varepsilon}$ be the solution of (2.2) with initial condition $\zeta_{t=0}^{x,\varepsilon} = \zeta^{x,\varepsilon}$. Then

$$\mathcal{L}\left(\zeta_t^{x,\varepsilon}\right) = \mu_x^\varepsilon(\cdot).$$

Relation (A.4) implies that

$$\mathbb{E}|\zeta^{x,\varepsilon}|^2 = \lim_{t\to\infty} \left|Y_t^{x,\varepsilon}\right|^2 \leq C(|x|^2 + \varepsilon^2 + 1). \tag{A.5}$$

The following lemma gives the exponentially mixing property of process $Y_t^{x,\varepsilon}$ towards $\zeta_t^{x,\varepsilon}$.

LEMMA A.1 *For any (x, ε),*

$$\mathbb{E}\left|Y_t^{x,\varepsilon} - \zeta_t^{x,\varepsilon}\right|^2 \leq \mathbb{E}\left|y - \zeta^{x,\varepsilon}\right|^2 e^{-2\beta t/\varepsilon}. \tag{A.6}$$

where $y = Y_{t=0}^{x,\varepsilon}$.

PROOF: Using Itô's formula and Assumption 2.3, we deduce

$$d\mathbb{E}\big|Y_t^{x,\varepsilon} - \zeta_t^{x,\varepsilon}\big|^2 \le -\frac{2\beta}{\varepsilon}\mathbb{E}\big|Y_t^{x,\varepsilon} - \zeta_t^{x,\varepsilon}\big|^2\,dt. \tag{A.7}$$

Hence (A.6) follows. □

From (A.5) and (A.6), it follows that

$$\mathbb{E}\big|Y_t^{x,\varepsilon} - \zeta_t^{x,\varepsilon}\big|^2 \le C\mathbb{E}(|y|^2 + 1)e^{-2\beta t/\varepsilon} \tag{A.8}$$

uniformly in time as long as (x, ε) is in a compact set.

PROPOSITION A.2 *Suppose* (x, ε) *is in a compact set*; *then there exists a constant* $C > 0$ *such that for any function* f *with Lipschitz constant less than* 1 *and* $t \in [0, \infty)$,

$$\mathbb{E}\left|\frac{1}{T}\int_t^{t+T} f\big(x, Y_s^{x,\varepsilon}, \varepsilon\big)ds - \int f(x, y, \varepsilon)\mu_x^\varepsilon(dy)\right| \le C\frac{(|y|^2 + 1)e^{-\beta t/\varepsilon}}{T}. \tag{A.9}$$

PROOF: Since $\mathcal{L}(\zeta_t^{x,\varepsilon}) = \mu_x^\varepsilon(\cdot)$, we have

$$\mathbb{E}f\big(x, Y_t^{x,\varepsilon}, \varepsilon\big) - \int f(x, y, \varepsilon)\mu_x^\varepsilon(dy) = \mathbb{E}f\big(x, Y_t^{x,\varepsilon}, \varepsilon\big) - \mathbb{E}f\big(x, \zeta_t^{x,\varepsilon}, \varepsilon\big).$$

By Lemma A.1 and the assumption on f, we have

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{T}\int_t^{t+T} f\big(x, Y_s^{x,\varepsilon}, \varepsilon\big)ds - \int f(x, y, \varepsilon)\mu_x^\varepsilon(dy)\right| \\
&\quad\le \frac{1}{T}\int_0^T \mathbb{E}\big|f\big(x, Y_s^{x,\varepsilon}, \varepsilon\big) - f\big(x, \zeta_s^{x,\varepsilon}, \varepsilon\big)\big|ds \\
&\quad\le C(|y|^2 + 1)\frac{e^{-\beta t/\varepsilon}}{T}.
\end{aligned}$$

□

Similar ergodic properties hold at the discrete level. Suppose $Y_n^{x,\varepsilon}$ is the solution of the microsolver (2.6) or (2.7) with parameter (x, ε) and microtime step δt. By the smoothness assumption, Assumption 2.1, for each $x \in \mathbb{R}^n$, $p \in \mathbb{N}$, and δt small enough, there exists $\lambda < 1$ such that

$$\mathbb{E}\big|Y_{n+1}^{x,\varepsilon}\big|^{2p} \le \lambda\big|Y_n^{x,\varepsilon}\big|^{2p} + F(x, \varepsilon), \tag{A.10}$$

where F is a smooth function. The results in [21] imply that under Assumptions 2.2 and 2.3, for each (x, ε) and δt small enough, $Y_n^{x,\varepsilon}$ is ergodic with unique invariant probability measure $\mu_x^{\delta t,\varepsilon}$. By the same analysis as in the proof of Lemma A.1, the following can be shown:

LEMMA A.3 *Suppose (x,ε) is in a compact set. For δt small enough, there exists a family of random variables $\{\zeta^{x,\delta t,\varepsilon}\}$ independent of the Wiener process in* (2.6) *or* (2.7) *with measure $\mu_x^{\delta t,\varepsilon}$ such that*

$$\mathbb{E}\left|Y_n^{x,\varepsilon}-\zeta_n^{x,\delta t,\varepsilon}\right|^2 \le \mathbb{E}|y-\zeta^{x,\delta t,\varepsilon}|^2 e^{-\beta n\delta t/\varepsilon}, \tag{A.11}$$

where $Y_n^{x,\varepsilon}$ (respectively, $\zeta_n^{x,\delta t,\varepsilon}$) is the solution of the microsolver (2.6) *or* (2.7) *with initial condition y (respectively, $\zeta^{x,\delta t,\varepsilon}$).*

For any smooth function $f=(x,y,\varepsilon)$ with polynomial growth in y, we define

$$\hat f(x,\varepsilon)=\int_{\mathbb{R}^m} f(x,y,\varepsilon)\mu_x^{\varepsilon}(dy) \tag{A.12}$$

and

$$\hat f^{\delta t}(x,\varepsilon)=\int_{\mathbb{R}^m} f(x,y,\varepsilon)\mu_x^{\delta t,\varepsilon}(dy). \tag{A.13}$$

Using energy estimate (A.3) and following the proof of theorem 3.3 in [25], it can be shown that under Assumptions 2.1, 2.2, and 2.3, if (x,ε) is in a compact set, for δt small enough, we have

$$\left|\hat f(x,\varepsilon)-\hat f^{\delta t}(x,\varepsilon)\right| \le C(\delta t/\varepsilon)^{\ell}. \tag{A.14}$$

This gives us an estimate on the error induced by the discretization on computing expectations.

The following lemma gives the property we need for

$$\hat a(x,\varepsilon)=\int_{\mathbb{R}^m} a(x,y,\varepsilon)\mu_x^{\varepsilon}(dy).$$

LEMMA A.4 *The function $\hat a(x,\varepsilon)$ is smooth.*

PROOF: For simplicity, we only discuss the case when $\mathbb{R}^n=\mathbb{R}^m=\mathbb{R}$. The proof for higher dimensions is similar. Suppose (x,ε) is in a compact set. Let $u(t,x,y,\varepsilon)=\mathbb{E}\,a(x,Y_t^{x,\varepsilon},\varepsilon)$ be the solution of the backward Fokker-Planck equation

$$\frac{\partial u}{\partial t}=Lu,\quad u(0,x,y,\varepsilon)=a(x,y,\varepsilon). \tag{A.15}$$

A straightforward generalization of Mikulyavichyus result [22] shows that $u(t,x,y,\varepsilon)$ is infinitely differentiable with respect to (x,y,ε). It is also proven in [25] that under Assumptions 2.1, 2.2, and 2.3, for any (x,ε) and $n\in\mathbb{N}$, there exists $s_n\in\mathbb{N}$, $B_n\in[0,\infty)$, and $\beta_n>0$ such that

$$\left|\partial_y^n u(t,x,y,\varepsilon)\right| \le B_n(|y|^{s_n}+1)e^{-\beta_n t/\varepsilon}. \tag{A.16}$$

Taking derivatives with respect to x on both sides of equation (A.15), we have

$$\frac{\partial}{\partial t}\partial_x u=\partial_x a(x,y,\varepsilon)\partial_y u+\partial_x\sigma^2(x,y,\varepsilon)\partial_y^2 u+L\partial_x u. \tag{A.17}$$

Using Duhamel's principle, $\partial_x u$ can be formally expressed in terms of u as

$$\partial_x u = e^{Lt}\partial_x a(x, y, \varepsilon) + \int_0^t e^{L(t-s)}\big(\partial_x a(x, \cdot, \varepsilon)\partial_y u + \partial_x \sigma^2(x, \cdot, \varepsilon)\partial_y^2 u\big)ds. \tag{A.18}$$

By the uniform mixing (A.8), the following limit exists:

$$\partial_x \hat{a}(x, \varepsilon) = \partial_x \lim_{t\to\infty} u(t, x, y, \varepsilon) = \lim_{t\to\infty} \partial_x u(t, x, y, \varepsilon).$$

Differentiating (A.17) with respect to y and using (A.16), we deduce that the semigroup generating $\partial_y u$ has a positive exponential decay. By the same type of analysis, we deduce that

$$|\partial_{xy} u(t, x, y, \varepsilon)| \le B(|y|^2 + 1)\exp(-\eta t),$$

where B and η are positive constants. Hence,

$$\partial_{xy}\bar{a}(x, \varepsilon) = \partial_{xy} \lim_{t\to\infty} u(t, x, y, \varepsilon) = \lim_{t\to\infty} \partial_{xy} u(t, x, y, \varepsilon) = 0.$$

The lemma follows by repeating the same analysis to higher-order derivatives in x and ε. □

Appendix B: Limiting Properties—Diffusive Time Scale

Now we want to give the exponential mixing property of the process (Y_t^1, Y_t^2) given by (3.3). We will assume that the coefficients are smooth and Assumption 3.2 holds.

LEMMA B.1 *For each (x, ε), the process (Y_t^1, Y_t^2) is exponentially mixing with unique invariant measure $\nu_x^\varepsilon(dy_1, dy_2)$. Furthermore, there exist stationary processes (ζ_t^1, ζ_t^2) with $\mathcal{L}(\zeta_t^1, \zeta_t^2) = \nu_x^\varepsilon$ such that for some $B < \infty$ and all $t \in [0, \infty)$,*

$$\mathbb{E}\big|Y_t^1 - \zeta_t^1\big|^2 + \big|Y_t^2 - \zeta_t^2\big|^2 \le B(|y_1| + |y_2| + 1)e^{-2\beta t/\varepsilon^2}. \tag{B.1}$$

PROOF: By (A.4), Y_t^1 is exponentially mixing. Now we want to show that (Y_t^1, Y_t^2) is also exponentially mixing. For any $\tau > 0$, denote by (V_t^τ, Y_t^τ) the unique solution of the equation

$$\begin{cases} \dot{V}_t = \dfrac{1}{\varepsilon^2} b_0(V_t) + \dfrac{1}{\varepsilon}\sigma_0(V_t)\dot{W}_t, \\ \dot{Y}_t = \dfrac{1}{\varepsilon^2}\partial b_0(V_t)Y_t + \dfrac{1}{\varepsilon}\partial\sigma_0(V_t)Y_t\dot{W}_t, + \dfrac{1}{\varepsilon^2} b_1(x, V_t, \varepsilon) + \dfrac{1}{\varepsilon}\sigma_1(x, V_t, \varepsilon)\dot{W}_t, \\ Y_{-\tau} = y_2, \quad V_{-\tau} = y_1. \end{cases}$$

It is easily seen that the distributions of (V_0^τ, Y_0^τ) and (Y_τ^1, Y_τ^2) coincide, i.e.,

$$\mathcal{L}\big(V_0^\tau, Y_0^\tau\big) = \mathcal{L}\big(Y_\tau^1, Y_\tau^2\big).$$

For ergodicity, it is sufficient [5] to prove that there exist random variables ξ and ζ such that the following convergence is exponential:

$$\lim_{\tau\to\infty} \mathbb{E}\left|V_0^\tau - \xi\right|^2 + \mathbb{E}\left|Y_0^\tau - \zeta\right|^2 = 0, \quad \forall (y_1, y_2).$$

Then (Y_t^1, Y_t^2) is ergodic with unique invariant measure $\nu_x^\varepsilon = \mathcal{L}(\xi, \zeta)$. The existence of ξ is guaranteed by the exponential mixing of Y_t^1.

Now we give the existence of ζ. Taking $\varepsilon = 0$ and $|y_2 - y_1| \to 0$ in Assumption 2.3, we have for any (y, y'),

$$\langle y, \partial b_0(y')y\rangle + \|\partial\sigma_0(y')y\|^2 \le -\beta|y|^2. \tag{B.2}$$

By the smoothness assumption, Assumption 2.1, (B.2), and Itô's lemma, we have for some constant C and all $t \ge -\tau$,

$$d\mathbb{E}\left|Y_t^\tau\right|^2 \le -\frac{\beta}{\varepsilon^2}\mathbb{E}\left|Y_t^\tau\right|^2 dt + \frac{C}{\varepsilon^2}(|y_1|^2 + 1)dt.$$

This means that for $\forall \tau > 0$ and $t \in [-\tau, \infty)$,

$$\mathbb{E}\left|Y_t^\tau\right|^2 \le e^{-\beta(t+\tau)/\varepsilon^2}|y_2|^2 + C(|y_1|^2 + 1). \tag{B.3}$$

Let $\gamma > \tau$ and $Z_t = Y_t^\tau - Y_t^\gamma$, $t \ge -\gamma$; then Z_t is the solution of the following equation:

$$\dot{Z}_t = \frac{1}{\varepsilon^2}\partial b(V_t)Z_t + \frac{1}{\varepsilon}\partial\sigma(V_t)Z_t\dot{W}_t, \quad Z(-\tau) = y_2 - Y_{-\tau}^\gamma.$$

Direct computation with Itô's formula and (B.3) shows that

$$\mathbb{E}|Z_0|^2 = \mathbb{E}\left|Y_0^\tau - Y_0^\gamma\right|^2 \le C(|y_1|^2 + |y_2|^2 + 1)e^{-2\beta\tau/\varepsilon^2}.$$

This implies that there exists a random variable ζ such that

$$\mathbb{E}\left|Y_0^\tau - \zeta\right|^2 \le C_0(|y_1|^2 + |y_2|^2 + 1)e^{-2\beta\tau/\varepsilon^2}.$$

The same analysis will show that ζ is independent of initial value (y_1, y_2). Taking (ζ_t^1, ζ_t^2) to be the solution of (3.3) with independent initial distribution ν_x^ε, by the same analysis above using Assumptions 2.3, we have (B.1). □

By a similar argument above we can prove the following lemma for the time discretization of (Y_n^1, Y_n^2) by the scheme (2.6) and (2.7).

LEMMA B.2 *For microstep δ small enough and each (x, ε), (Y_n^1, Y_n^2) is exponentially mixing with unique invariant measure $\nu_{x,\varepsilon}^\delta(dy_1, dy_2)$. Furthermore, there exist stationary processes (ζ_n^1, ζ_n^2) with distribution $\nu_{x,\varepsilon}^{\delta t}$ such that for some $B < \infty$,*

$$\mathbb{E}\left|Y_n^1 - \zeta_n^1\right|^2 + \left|Y_n^2 - \zeta_n^2\right|^2 \le B(|y_1| + |y_2| + 1)e^{-\beta n\delta t/\varepsilon^2}. \tag{B.4}$$

By the independence of Y_t^1 on (x, ε) and the linear form of the equation for Y_t^2, using the same analysis as in [25], we also have for any $x \in \mathbb{R}^n$ and any function of polynomial growth,

$$\left| \int_{\mathbb{R}^m \times \mathbb{R}^m} \nu_x^\varepsilon(dy_1, dy_2) f(x, y_1, y_2, \varepsilon) - \int_{\mathbb{R}^m \times \mathbb{R}^m} \nu_{x,\varepsilon}^{\delta t}(dy_1, dy_2) f(x, y_1, y_2, \varepsilon) \right| \le B\delta t^\ell, \tag{B.5}$$

where B is an independent constant.

Define

$$\begin{cases} \tilde{a}(x, \varepsilon) = \int_{\mathbb{R}^m \times \mathbb{R}^m} \nu_x^\varepsilon(dy_1, dy_2) \partial_y a(x, y_1, \varepsilon) y_2 \\ \qquad\qquad + \int_{\mathbb{R}^m} \mu(dy_1) \int_0^\infty \mathbb{E}_{y_1} \partial_x a\big(x, Y_{\varepsilon^2 s}^1, \varepsilon\big) a(x, y_1, \varepsilon) ds, \\ \tilde{\sigma}(x, \varepsilon)\tilde{\sigma}^T(x, \varepsilon) = 2 \int_{\mathbb{R}^m} \mu(dy_1) a(x, y_1, \varepsilon) \\ \qquad\qquad \otimes \int_0^\infty \mathbb{E}_{y_1} a\big(x, Y_{\varepsilon^2 s}^1, \varepsilon\big) ds. \end{cases} \tag{B.6}$$

LEMMA B.3 *The functions $\tilde{a}(x, \varepsilon)$ and $\tilde{\sigma}(x, \varepsilon)\tilde{\sigma}^T(x, \varepsilon)$ are smooth functions of (x, ε) with derivatives of polynomial growth.*

PROOF: To repeat the proof of Lemma A.4, we only need to show that the following functions are smooth with derivatives of polynomial growth,

$$\int_0^\infty \mathbb{E}_{y_1} a\big(x, Y_{\varepsilon^2 s}^1, \varepsilon\big) ds, \qquad \int_0^\infty \mathbb{E}_{y_1} \partial_x a\big(x, Y_{\varepsilon^2 s}^1, \varepsilon\big) ds.$$

Based on the centering condition, the exponential mixing, and the independence of $Y_{\varepsilon^2 s}^1$ with respect to x, using the same analysis for the proof of Lemma A.4, we have that functions $\mathbb{E}_{y_1} a(x, Y_{\varepsilon^2 s}^1, \varepsilon)$ and $\mathbb{E}_{y_1} \partial_x a(x, Y_{\varepsilon^2 s}^1, \varepsilon)$ and their arbitrary derivatives with respect to y_1 decay exponentially to 0 and hence are integrable on an infinite time interval. Hence interchanging the order of limits, we have the following limit:

$$\partial_{y_1} \int_0^\infty \mathbb{E}_{y_1} a\big(x, Y_{\varepsilon^2 s}^1, \varepsilon\big) dt = \int_0^\infty \partial_{y_1} \mathbb{E}_{y_1} a\big(x, Y_{\varepsilon^2 s}^1, \varepsilon\big) dt$$

and

$$\partial_{y_1} \int_0^\infty \mathbb{E}_{y_1} \partial_x a\big(x, Y_{\varepsilon^2 s}^1, \varepsilon\big) dt = \int_0^\infty \partial_{y_1} \mathbb{E}_{y_1} \partial_x a\big(x, Y_{\varepsilon^2 s}^1, \varepsilon\big) dt.$$

The same differentiability holds for higher-order derivatives for (x, ε) and the result follows. □

Acknowledgment. We are grateful to Ibrahim Fatkullin and Jonathan Mattingly for stimulating discussions and insightful comments. The work of E is partially supported by NSF Grant DMS01-30107. Liu is partially supported by NSF Grant DMS97-29992 and by ONR Grant N00014-01-1-0674. Vanden-Eijnden is partially supported by NSF Grants DMS02-09959 and DMS02-39625.

Bibliography

[1] Car, R.; Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **55** (1985), no. 22, 2471–2474.

[2] Chorin, A. J. *Computational fluid mechanics.* Selected papers. Academic Press, Boston, 1989.

[3] Chorin, A. J.; Kast, A. P.; Kupferman, R. Optimal prediction of underresolved dynamics. *Proc. Natl. Acad. Sci. USA* **95** (1998), no. 8, 4094–4098.

[4] Chorin, A. J.; Kast, A. P.; Kupferman, R. Unresolved computation and optimal predictions. *Comm. Pure Appl. Math.* **52** (1999), no. 10, 1231–1254.

[5] Da Prato, G.; Zabczyk, J. *Stochastic equations in infinite dimensions*. Encyclopedia of Mathematics and Its Applications, 44. Cambridge University Press, Cambridge, 1992.

[6] E, W. Analysis of the heterogeneous multiscale method for ordinary differential equations. *Commun. Math. Sci.* **1** (2003), no. 3, 423–436.

[7] E, W.; Engquist, B. The heterogeneous multiscale methods. *Commun. Math. Sci.* **1** (2003), no. 1, 87–132.

[8] E, W.; Engquist, B. Multiscale modeling and computation. *Notices Amer. Math. Soc.* **50** (2003), no. 9, 1062–1070.

[9] E, W.; Ming, P. B.; Zhang, P. W. Analysis of the heterogeneous multiscale method for elliptic homogenization problems. *J. Amer. Math. Soc.* **18**, no. 1, 121–156.

[10] Engquist, B.; Tsai, R. The heterogeneous multiscale method for stiff ODEs. *Math. Comp.*, in press.

[11] Fatkullin, I.; Vanden-Eijnden, E. A computational strategy for multiscale systems with applications to Lorenz 96 model. *J. Comput. Phys.* **200** (2004), no. 2, 605–638.

[12] Freidlin, M. I.; Wentzell, A. D. *Random perturbations of dynamical systems*. Second edition. Grundlehren der Mathematischen Wissenschaften, 260. Springer-Verlag, New York, 1998.

[13] Gear, C. W.; Kevrekidis, I. G. Projective methods for stiff differential equations: problems with gaps in their eigenvalue spectrum. *SIAM J. Sci. Comput.* **24** (2003), no. 4, 1091–1106.

[14] Has′minskiĭ, R. Z. *Stochastic stability of differential equations.* Monographs and Textbooks on Mechanics of Solids and Fluids: Mechanics and Analysis, 7. Sijthoff & Noordhoff, Alphen aan den Rijn–Germantown, Md., 1980.

[15] Kloeden, P. E.; Platen, E. *Numerical solution of stochastic differential equations*. Applications of Mathematics (New York), 23. Springer-Verlag, Berlin, 1992.

[16] Kolesik, M.; Novotny, M. A.; Rikvold, P. A. Projection method for statics and dynamics of lattice spin systems. *Phys. Rev. Lett.* **80** (1998), no. 15, 3384–3387.

[17] Kurtz, T. G. A limit theorem for perturbed operator semigroups with applications to random evolutions. *J. Functional Analysis* **12** (1973), 55–67.

[18] Lamba, H.; Mattingly, J. C.: Stuart, A. M. An adaptive Euler-Maruyama scheme for SDEs: Part I, convergence, Part II, stability. Preprint, 2004.

[19] Lebedev, V. I.; Finogenov, S. I. Explicit methods of second order for the solution of stiff systems of ordinary differential equations. *Zh. Vychisl. Mat. Mat. Fiz.* **16** (1976), 895–910.

[20] Majda, A. J.; Timofeyev, I.; Vanden-Eijnden, E. A mathematical framework for stochastic climate models. *Comm. Pure Appl. Math.* **54** (2001), no. 8, 891–974.

[21] Meyn, S. P.; Tweedie, R. L. Stability of Markovian processes. I. Criteria for discrete-time chains. *Adv. in Appl. Probab.* 24 (1992), no. 3, 542–574. II. Continuous-time processes and sampled chains. *Adv. in Appl. Probab.* 25 (1993), no. 3, 487–517. III. Foster-Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.* 25 (1993), no. 3, 518–548.

[22] Mikulyavichyus, R. Some properties of solutions of stochastic differential equations. *Litovsk. Mat. Sb.* **23** (1983), no. 4, 18–31.

[23] Ming, P. B.; Zhang, P. W. Analysis of the heterogeneous multiscale method for parabolic homogenization problems. Submitted to *Math. Comp.*

[24] Papanicolaou, G. C. Introduction to the asymptotic analysis of stochastic equations. *Modern modeling of continuum phenomena (Ninth Summer Sem. Appl. Math., Rensselaer Polytech. Inst., Troy, N.Y., 1975)*, 109–147. Lectures in Applied Mathematics, Vol. 16. American Mathematical Society, Providence, R.I., 1977.

[25] Talay, D. Second order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stoch. Stoch. Rep.* **29** (1990), 13–36.

[26] Vanden-Eijnden, E. Numerical techniques for multi-scale dynamical systems with stochastic effects. *Commun. Math. Sci.* **1** (2003), no. 2, 385–391.

WEINAN E
Department of Mathematics
and PACM
Princeton University
Washington Road
Princeton, NJ 08544
E-mail: weinan@princeton.edu

DI LIU
Courant Institute
251 Mercer Street
New York, NY 10012
E-mail: diliu@cims.nyu.edu

ERIC VANDEN-EIJNDEN
Courant Institute
251 Mercer Street
New York, NY 10012
E-mail: eve2@cims.nyu.edu

Received May 2004.