MULTISCALE MODEL. SIMUL.
Vol. 8, No. 3, pp. 944–964

ANALYSIS OF MULTISCALE METHODS FOR STOCHASTIC DYNAMICAL SYSTEMS WITH MULTIPLE TIME SCALES*

DI LIU†

Abstract. We prove the convergence of a class of numerical schemes for stochastic dynamical systems with well-separated time scales. A sharp error estimate for the schemes is provided. The optimality of the convergence rate is illustrated through examples.

Key words. stochastic differential equations, time scale separation, multiscale methods

AMS subject classifications. 60H10, 34E13, 65C20

DOI. 10.1137/090750664

1. Introduction.

Consider the following stochastic dynamical system with a time scale separation measured by $\varepsilon \ll 1$:

$$\begin{cases} \dot{X}_t^\varepsilon = a\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big) + \sigma\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big)\dot{W}_t, & X_0^\varepsilon = x, \\ \dot{Y}_t^\varepsilon = \dfrac{1}{\varepsilon}B\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big) + \dfrac{1}{\sqrt{\varepsilon}}C\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big)\dot{W}_t, & Y_0^\varepsilon = y, \end{cases} \tag{1.1}$$

where $X_t^\varepsilon \in \mathbb{R}^n$ and $Y_t^\varepsilon \in \mathbb{R}^m$, and W_t is a standard d-dimensional Wiener process. Notice that $a(\cdot) \in \mathbb{R}^n$ and $B(\cdot) \in \mathbb{R}^m$ are vector valued functions, while $\sigma(\cdot) \in \mathbb{R}^n \times \mathbb{R}^d$ and $C(\cdot) \in \mathbb{R}^m \times \mathbb{R}^d$ are matrix valued functions. Under appropriate assumptions on $B(\cdot)$ and $C(\cdot)$, the dynamics for Y_t^ε with $X_t^\varepsilon = x$ fixed is ergodic with a unique invariant measure $\mu_x^\varepsilon(dy)$. In this case, it has been proved [5, 8] that the effective dynamics for X_s^ε in the limit of $\varepsilon \to 0$, in both the weak and the strong sense, is an SDE in the following form:

$$\dot{\bar{X}}_t = \bar{a}\big(\bar{X}_t\big) + \bar{\sigma}\big(\bar{X}_t\big)\dot{W}_t, \qquad \bar{X}_0 = x, \tag{1.2}$$

where

$$\begin{aligned} \bar{a}(x) &= \lim_{\varepsilon\to 0}\int a(x,y,\varepsilon)\mu_x^\varepsilon(dy), \\ \bar{\sigma}(x)\bar{\sigma}^T(x) &= \lim_{\varepsilon\to 0}\int \sigma(x,y,\varepsilon)\sigma^T(x,y,\varepsilon)\mu_x^\varepsilon(dy). \end{aligned} \tag{1.3}$$

In [14], a multiscale integration scheme was proposed to deal with systems in the form of (1.1) by solving the effective dynamics (1.2). Fitting into the framework of heterogeneous multiscale methods (HMM) [1, 2, 15], the scheme consists of a macro solver to evolve (1.2) and a micro solver to simulate the fast dynamics in (1.1) with fixed slow variables. An estimator is chosen to approximate the coefficients $\bar{a}(\cdot)$ and $\bar{\sigma}(\cdot)$ on the fly at each macro time step, using data obtained from the fast simulations with the micro solver. Without having to resolve all the details of the fast processes

*Received by the editors February 23, 2009; accepted for publication (in revised form) January 29, 2010; published electronically April 14, 2010. This work was supported by the NSF under grant DMS-0609315.
http://www.siam.org/journals/mms/8-3/75066.html

†Department of Mathematics, Michigan State University, East Lansing, MI 48824 (richardl@math.msu.edu).

on the $O(\varepsilon)$ fine time scale, the method is able to overcome the numerical stiffness induced by the time scale separation. To fully justify the strategy, a thorough analysis of convergence and efficiency of the scheme is needed. In [3], the convergence of the HMM scheme for (1.1) is proved assuming that the slow dynamics is deterministic, i.e., $\sigma = 0$. For systems with the slow diffusion depending only on the slow variables such that $\sigma = \sigma(x)$, the convergence of the scheme is established in [4]. In this paper, we are going to study the more general case of fully coupled systems where the slow diffusion depends on both slow and fast variables. We will provide the convergence rate for the HMM scheme in both strong and weak senses. We will also illustrate the optimality of results with simple examples and numerical experiments. It will be shown that the error estimate here is much sharper compared with those in similar works [4].

Throughout the remainder of this paper, we will denote by $\mathbb{C}$ a generic positive constant which may change its value from line to line. In chains of inequalities, we will adopt $\mathbb{C}$, $\mathbb{C}'$, $\mathbb{C}''$, ... or $\mathbb{C}_1$, $\mathbb{C}_2$, $\mathbb{C}_3$, ... to avoid confusion.

2. The multiscale scheme. For a system in the form of (1.1), what is usually of interest is the behavior of the slow variable X_t^ϵ, whose leading order term for small ε is $\bar{X}_t$, as described by (1.2). The difficulty of dealing with $\bar{X}_t$ lies in the fact that the coefficients $\bar{a}(\cdot)$ and $\bar{\sigma}(\cdot)$ are given via expectations with respect to measure $\mu_x^\varepsilon(dy)$, which is usually difficult or impossible to obtain analytically, especially when the dimension m is large. The basic idea of HMM is to solve $\bar{X}_t$ with $\bar{a}(\cdot)$ and $\bar{\sigma}(\cdot)$ being estimated on the fly using a time-ensemble average of the original slow coefficients $a(\cdot)$ and $\sigma(\cdot)$ with respect to numerical solutions of the fast processes, which gives a three-step procedure:

1. *Macro solver.* At each time step n, having the numerical solution X_n, in order to move to step $n+1$, we evolve (1.2) with an explicit stable SDE solver using approximate coefficients $\tilde{a}$ and $\tilde{\sigma}$. For instance, in the simplest case when the Euler–Maruyama scheme is selected, we have

$$X_{n+1} = X_n + \tilde{a}_n \Delta t + \tilde{\sigma}_n \Delta B_{n+1}, \tag{2.1}$$

   where Δt is the macro time step size and ΔB_n's are independent $N(0, \Delta t)$ Gaussian random variables.

2. *Micro solver.* To obtain $\tilde{a}$ and $\tilde{\sigma}$ used in the macro solver, we solve the following fast scale problem:

$$\dot{Y}_t = \frac{1}{\varepsilon} B\big(x, Y_t, \varepsilon\big) + \frac{1}{\sqrt{\varepsilon}} C\big(x, Y_t, \varepsilon\big) \dot{W}_t, \tag{2.2}$$

   with fixed parameter $x = X_n$, and we denote the solution by $\{Y_{n,m}\}$, where m labels the micro time 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.

3. *Estimator.* $\tilde{a}$ and $\tilde{\sigma}$ can be given by the following time and ensemble average:

$$\tilde{a}_n = \frac{1}{MN} \sum_{j=1}^{M} \sum_{m=n_r+1}^{n_r+N} a\big(X_n, Y_{n,m,j}, \varepsilon\big) \tag{2.3}$$

   and

$$\tilde{\sigma}_n \tilde{\sigma}_n^T = \frac{1}{MN} \sum_{j=1}^{M} \sum_{m=n_r+1}^{n_r+N} \sigma_n\big(X_n, Y_{n,m,j}, \varepsilon\big) \sigma_n^T\big(X_n, Y_{n,m,j}, \varepsilon\big), \tag{2.4}$$

where M is the number of replicas, N is the number of steps in the time averaging, and n_r is the number of steps we skip to eliminate transients. $\tilde{\sigma}_n$ is then obtained by a Cholesky decomposition.

For the macro solver, any stable explicit SDE solver such as the Euler–Maruyama, a Taylor, or a multistep method can be adopted. For simplicity, we consider only explicit solvers here. Extension to implicit solvers is straightforward, but it tends to make the implementation and analysis more involved. For the micro solver, denoting by ℓ the weak order of accuracy, we can use the following Euler–Maruyama scheme, which is first order accurate ($\ell = 1$),

$$Y^i_{n,m+1} = Y^i_{n,m} + \frac{1}{\varepsilon} B^i\big(X_n, Y_{n,m}, \varepsilon\big)\delta t + \frac{1}{\sqrt{\varepsilon}} \sum_j C^{ij}\big(X_n, Y_{n,m}, \varepsilon\big) W^j_{m+1}\sqrt{\delta t}, \tag{2.5}$$

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

$$\begin{aligned} Y^i_{n,m+1} = Y^i_{n,m} &+ \frac{1}{\sqrt{\varepsilon}} \sum_j C^{ij}\big(X_n, Y_{n,m}, \varepsilon\big) W^j_{m+1}\sqrt{\delta t} \\ &+ \frac{1}{\varepsilon} B^i\big(X_n, Y_{n,m}, \varepsilon\big)\delta t + \frac{1}{\varepsilon}\sum_{jk} A^{ijk}\big(X_n, Y_{n,m}, \varepsilon\big) s^{kj}_{m+1}\delta t \\ &+ \frac{1}{2\varepsilon^{3/2}} \sum_j D^{ij}\big(X_n, Y_{n,m}, \varepsilon\big) W^j_{m+1}\delta t^{3/2} \\ &+ \frac{1}{2\varepsilon^2} E^i\big(X_n, Y_{n,m}, \varepsilon\big)\delta t^2. \end{aligned} \tag{2.6}$$

In (2.5) and (2.6), δt is the micro time step size (note that it appears only in terms of the ratio $\delta t/\varepsilon =: \mathit{\Delta}\tau$), and the coefficients are defined as

$$\begin{aligned} A^{ijk} &= \sum_l \big(\partial^l C^{ij}\big) C^{lk}, \\ D^{ij} &= \sum_l \big(C^{lj}\partial^l B^i + B^l \partial^l C^{ij}\big) + \frac{1}{2}\sum_{kl} G^{kl}\partial^k\partial^l C^{ij}, \\ E^i &= \sum_j B^j \partial^j B^i + \frac{1}{2}\sum_{jk} G^{jk}\partial^j\partial^k B^i, \end{aligned}$$

where $G = CC^T$ and the derivatives are taken with respect to y. The random variables $\{W^j_m\}$ are independent and identically distributed (i.i.d.) Gaussian with mean zero and variance one, and

$$s^{kj}_m = \begin{cases} \frac{1}{2} W^k_m W^j_m + \xi^{kj}_m, & k < j, \\ \frac{1}{2} W^k_m W^j_m - \xi^{jk}_m, & k > j, \\ \frac{1}{2}\left((W^j_m)^2 - 1\right), & k = j, \end{cases} \tag{2.7}$$

where $\{\xi^{kj}_m\}$ are i.i.d. with $\mathbb{P}\{\xi^{kj}_m = \frac{1}{2}\} = \mathbb{P}\{\xi^{kj}_m = -\frac{1}{2}\} = \frac{1}{2}$. At each macro time step, the fast process is reinitialized such that

$$Y_{n+1,0,j} = Y_{n,n_r+N,j}; \tag{2.8}$$

i.e., the initial values for the micro variables at macro time step $n+1$ are chosen to be their final values from macro time step n.

3. Convergence of the scheme. In this section, we provide the rate of convergence for the scheme described in section 2. We will first discuss the convergence in the strong sense and then give the convergence in the weak sense. The optimality of the estimated convergence rate will also be addressed. Define $\mathbb{C}_b^\infty$ to be the space of smooth functions with bounded derivatives of any order. We make the following assumptions.

Assumption 3.1. The coefficients $a(\cdot)$, $\sigma(\cdot)$, $B(\cdot)$, and $C(\cdot)$, viewed as functions of (x, y, ε), are in $\mathbb{C}_b^\infty$. Moreover, $a(\cdot)$ and $\sigma(\cdot)$ are bounded.

Assumption 3.2. There exists a constant $\alpha > 0$ such that $\forall\ (x, y, \varepsilon)$

$$y^T C(x, y, \varepsilon) C^T(x, y, \varepsilon) y \geq \alpha |y|^2. \tag{3.1}$$

Assumption 3.3. There exists a constant $\beta > 0$ such that $\forall\ (x, y_1, y_2, \varepsilon)$

$$\begin{aligned}&\big\langle (y_1 - y_2),\, B(x, y_1, \varepsilon) - B(x, y_2, \varepsilon)\big\rangle + \big\|C(x, y_1, \varepsilon) - C(x, y_2, \varepsilon)\big\|_F^2\\ &\quad\leq -\beta |y_1 - y_2|^2,\end{aligned} \tag{3.2}$$

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

Assumption 3.4. There exists a constant γ such that $\forall x$

$$\big\|\big(\bar\sigma(x)\bar\sigma^T(x)\big)^{-1}\big\|_2 \leq \gamma \quad \text{and} \quad \bar\sigma_{ii}(x) \geq 0, \tag{3.3}$$

where $\|\cdot\|_2$ denotes the l_2-norm.

Assumption 3.1 can be weakened but is used here for simplicity of presentation. Assumption 3.2 means that the diffusion is nondegenerate for the y-process, which guarantees a smooth transition probability density. Assumption 3.3 is a dissipative condition for the stability of the fast process. Under these assumptions, one can show that for each (x, ε) the fast dynamics (2.2) is exponentially mixing with a unique invariant measure. Assumption 3.4 is for the well-posedness of the Cholesky decomposition.

3.1. Strong convergence rate. Base on Assumptions 3.1–3.4, we can prove the following theorem on the strong convergence rate of the HMM scheme.

THEOREM 3.5. *Assume that the macro solver is kth order accurate in the strong sense and the micro solver is ℓth order accurate in the weak sense. Suppose Δt and $(\delta t/\varepsilon)$ are small enough. Then, for any $T_0 > 0$, there exists a constant $\mathbb{C} > 0$ independent of $(\epsilon, \Delta t, \delta t, n_r, M, N)$ such that*

$$\begin{aligned}&\sup_{n \leq T_0/\Delta t} \Big(\mathbb{E}\big|\bar X_{t_n} - X_n\big|^2\Big)^{1/2}\\ &\quad\leq \mathbb{C}\left(\Delta t^k + \varepsilon + \left(\frac{\delta t}{\varepsilon}\right)^\ell + \frac{e^{-\beta(n_r+1)(\delta t/\varepsilon)}}{N(\delta t/\varepsilon)+1}\sqrt{R} + \frac{1}{\sqrt{M(N(\delta t/\varepsilon)+1)}}\right),\end{aligned} \tag{3.4}$$

where

$$R = \frac{\Delta t}{1 - e^{-2\beta(n_r+N)(\delta t/\varepsilon)}}. \tag{3.5}$$

To understand the above error estimate, we should first of all notice that the error estimate on $|\bar X_{t_n} - X_n|$ in (3.4) can be divided into two parts:

E1. $|\bar X_{t_n} - \bar X_n|$, where $\bar X_n$ is the approximation of $\bar X_{t_n}$ given by the selected macro solver assuming that $\bar a(\cdot)$ and $\bar\sigma(\cdot)$ are known exactly.

E2. $|\bar{X}_n - X_n|$, where the error is induced by using $\tilde{a}$ and $\tilde{\sigma}$, instead of $\bar{a}(\cdot)$ and $\bar{\sigma}(\cdot)$, to advance effective dynamics (1.2).

The estimate of E1 is a standard numerical SDE error estimate. Based on the smoothness of $\bar{a}(\cdot)$ and $\bar{\sigma}(\cdot)$ given in the appendix, we have

$$\mathbb{E}\big|\bar{X}_{t_n} - \bar{X}_n\big| \leq \mathbb{C}\Delta t^k, \tag{3.6}$$

which is the first term on the right-hand side of (3.4). Error E2 can be further divided into four parts:

1. The second term of $O(\varepsilon)$ in (3.4) arises from the dependence of $a(\cdot)$ and $\sigma(\cdot)$ on ε. This error does not exist if $a(\cdot)$ and $\sigma(\cdot)$ are independent of ε such that $a = a(x, y)$ and $\sigma = \sigma(x, y)$.
2. The term $(\delta t/\varepsilon)^\ell$ is due to the micro time discretization which induces a difference between the invariant measures of the continuous and discretized processes.
3. The term
$$\frac{e^{-\beta(n_r+1)(\delta t/\varepsilon)}}{N(\delta t/\varepsilon) + 1}\sqrt{R}$$
accounts for the errors caused by the time scale for the relaxation of the fast variables. The exponential term and its denominator are from the geometric convergence of the distributions of the fast process to its quasi-equilibrium distribution. When $(n_r + N)(\delta t/\varepsilon) \gg 1$, the fast process will be fully relaxed to the equilibrium at each macro time step. In this case, we can have
$$\sqrt{R} \approx \sqrt{\Delta t},$$
which is consistent with the fact that moving to the next macro time step will generate only an $O(\Delta W)$ deviation in the equilibrium distributions of the fast processes.
4. Finally, the central limit theorem type of estimate
$$\frac{1}{\sqrt{M(N(\delta t/\varepsilon) + 1)}}$$
gives the sampling errors when the fast variable reaches its quasi equilibrium.

The optimality of the convergence rate (3.4) can be illustrated through the following example:

$$\begin{cases} \dot{X}_t^\varepsilon = \left(Y_t^\varepsilon - \dfrac{1}{2}\right)\dot{B}_t, & X_0^\varepsilon = x, \\ \dot{Y}_t^\varepsilon = -\dfrac{1}{\varepsilon}Y_t^\varepsilon + \dfrac{1}{\sqrt{\varepsilon}}\dot{W}_t, & Y_0^\varepsilon = y. \end{cases} \tag{3.7}$$

The effective dynamics prescribed by (1.2) has the solution

$$\dot{\bar{X}}_t = \frac{\sqrt{3}}{2}\dot{B}_t, \qquad \bar{X}_0 = x. \tag{3.8}$$

For simplicity, we choose the Euler–Maruyama scheme as both macro and micro solvers for the HMM scheme. Since the effective dynamics (3.8) has a constant coefficient, the Euler solver for (3.8) is exact and the macro time step error E1 does not exist. Also, since the coefficient of the slow dynamics does not depend on ε, the $O(\varepsilon)$ error also does not show up. These errors can be easily introduced if we change the slow coefficient from $Y_t^\varepsilon - \frac{1}{2}$ to $Y_t^\varepsilon + X_t^\varepsilon$ or $Y_t^\varepsilon + \varepsilon$, so the effective slow dynamics will depend on x or ε. As can be shown by little modification of the computation below, these changes will not affect other errors. For simplicity, we choose the example in its current form to avoid these terms and focus on the relaxation and sampling errors.

The micro solver for the fast process therefore has the following form if we let $\Delta\tau = \delta t/\varepsilon$:

$$\begin{cases} Y_{n,m+1} = (1-\Delta\tau)Y_{n,m} + W_{n,m+1}\sqrt{\Delta\tau}, \\ Y_{n,0} = Y_{n-1,n_r+N}, \quad Y_{0,0} = y. \end{cases}$$

The solution of the above equation is a Gaussian process with mean and covariance

$$\begin{aligned} \mathbb{E}Y_{n,m} &= (1-\Delta\tau)^{n(n_r+N)+m} y, \\ Cov(Y_{n,m}, Y_{n.l}) &= \frac{(1-\Delta\tau)^{|m-l|} - \big(1-\Delta\tau\big)^{2n(n_r+N)+m+l}}{2(1-\Delta\tau/2)}. \end{aligned}$$

The HMM scheme for the effective dynamics has the form

$$X_{n+1} = X_n + \left(\frac{1}{MN}\sum_{m=n_r+1}^{n_r+N}\sum_{j=1}^{M}\left(Y_{n,m,j} - \frac{1}{2}\right)^2\right)^{1/2} \Delta B_{n+1}, \qquad X_0 = x.$$

Itô isometry suggests that

$$\begin{aligned} &\mathbb{E}\big|\bar{X}_{n\Delta t} - X_n\big|^2 \\ &\quad = \Delta t \sum_{k=0}^{n-1} \mathbb{E}\left(\left(\frac{1}{MN}\sum_{m=n_r+1}^{n_r+N}\sum_{j=1}^{M}\left(Y_{k,m,j} - \frac{1}{2}\right)^2\right)^{1/2} - \frac{\sqrt{3}}{2}\right)^2. \end{aligned} \tag{3.9}$$

To verify (3.4), we calculate the error of the scheme for two asymptotic regimes of the numerical parameters, where either the relaxation error or the sampling error will dominate.

1. We first consider the situation $M \to \infty$, which means the equilibrium distribution of the fast processes is perfectly sampled. In this case, there is no sampling error in (3.4). By the law of large numbers, we have, with probability one,

$$\begin{aligned} &\frac{1}{M}\sum_{j=1}^{M}\left(Y_{k,n_r+m,j} - \frac{1}{2}\right)^2 \\ &\quad \longrightarrow \frac{1 - \big(1-\Delta\tau\big)^{2k(n_r+N)+2m}}{2(1-\Delta\tau/2)} + \left(\big(1-\Delta\tau\big)^{k(n_r+N)+m} y - \frac{1}{2}\right)^2. \end{aligned}$$

By Taylor expansion, $(1-\Delta\tau) = e^{-\Delta\tau} + O(\Delta\tau^2)$. Therefore we can have

$$\lim_{M\to\infty} \frac{1}{N}\sum_m \frac{1}{M}\sum_{j=1}^{M}\left(Y_{k,n_r+m,j}-\frac{1}{2}\right)^2$$
$$= O\left(\frac{3}{4}+\Delta\tau+\frac{e^{-(k(n_r+N)+n_r+1)\Delta\tau}(1-e^{-2N\Delta\tau})}{N\Delta\tau}\right)$$
$$= O\left(\frac{3}{4}+\Delta\tau+\frac{e^{-(k(n_r+N)+n_r+1)\Delta\tau}}{N\Delta\tau+1}\right)$$

and

$$\lim_{M\to\infty}\mathbb{E}\left|\bar{X}_{n\Delta t}-X_n\right|^2 = O\left(\Delta\tau^2+\frac{e^{-2(n_r+1)\Delta\tau}}{(N\Delta\tau+1)^2}\frac{\Delta t}{1-e^{-2(n_r+N)\Delta\tau}}\right).$$

The exponential decay rate is consistent with estimate (3.4) since system (3.7) satisfies Assumption 3.3 with $\beta=1$.

2. We next let $n_r\to\infty$, which means the fast processes are fully relaxed to the equilibrium. Estimate (3.4) implies that the error consists only of the micro time step error and the sampling error. We can compute the following variance:

$$\lim_{n_r\to\infty}\mathbb{E}\left(\frac{1}{N}\sum_m\left(Y_{n,m,j}-\frac{1}{2}\right)^2-\mathbb{E}\left(Y_{n,m,j}-\frac{1}{2}\right)^2\right)^2$$
$$= O\left(\frac{1}{N^2}\sum_{m,l}\frac{(1-\Delta\tau)^{|l-m|}}{2(1-\Delta\tau/2)}\right)$$
$$= O\left(\frac{1}{N\Delta\tau+1}\right).$$

By the central limit theorem, when $M\gg 1$,

$$\frac{1}{\sqrt{M}}\sum_j\left(\frac{1}{N}\sum_m\left(Y_{k,m,j}-\frac{1}{2}\right)^2-\mathbb{E}\left(Y_{k,m,j}-\frac{1}{2}\right)^2\right)$$

can be approximated by a Gaussian random variable with mean zero and the above variance. Again by Taylor expansion, we have

$$\lim_{n_r\to\infty}\mathbb{E}\left|\bar{X}_{n\Delta t}-X_n\right|^2$$
$$=\lim_{n_r\to\infty}\Delta t$$
$$\sum_k\mathbb{E}\left(\left(\frac{1}{N}\sum_{m,j}\mathbb{E}\left(Y_{n,m,j}-\frac{1}{2}\right)^2\right)^{1/2}+O\left(\frac{1}{\sqrt{M(N\Delta\tau+1)}}\right)-\frac{\sqrt{3}}{2}\right)^2$$
$$=O\left(\Delta\tau^2+\frac{1}{M(N\Delta\tau+1)}\right),$$

which is consistent with estimate (3.4).

Results similar to Theorem 3.5 have been given for the case when $\sigma=\sigma(x)$ in [4]. Under the assumption that $a(\cdot)$ and $b(\cdot)$ are independent of ε, it is proved in [4] that when $M = 1$ and $n_r = 0$

$$\mathbb{E}\left|\bar{X}_{t_n} - X_n\right|^2 \le \mathbb{C}\left(\Delta t + \left(\frac{\delta t}{\varepsilon}\right)^{1/2} + \frac{\ln^2\bigl(N(\delta t/\varepsilon)\bigr) + 1)}{N(\delta t/\varepsilon)} + \frac{1}{N}\right), \tag{3.10}$$

while (3.4) suggests that the error should be

$$\begin{aligned}\mathbb{E}\left|\bar{X}_{t_n} - X_n\right|^2 &\le \mathbb{C}\left(\Delta t^{2k} + \left(\frac{\delta t}{\varepsilon}\right)^{2\ell} + \frac{\Delta t}{(N(\delta t/\varepsilon)+1)^2} + \frac{1}{N(\delta t/\varepsilon)+1}\right)\\ &\le \mathbb{C}'\left(\Delta t^{2k} + \left(\frac{\delta t}{\varepsilon}\right)^{2\ell} + \frac{1}{N(\delta t/\varepsilon)+1}\right).\end{aligned} \tag{3.11}$$

Both (3.10) and (3.11) imply that the numerical scheme gives accurate results only in the following regime of the parameters:

$$\Delta t,\ (\delta t/\varepsilon) \ll 1, \quad N(\delta t/\varepsilon) \gg 1, \tag{3.12}$$

which reduces (3.10) to

$$\mathbb{E}\left|\bar{X}_{t_n} - X_n\right|^2 \le \mathbb{C}\left(\Delta t + \left(\frac{\delta t}{\varepsilon}\right)^{1/2} + \frac{\ln^2\bigl(N(\delta t/\varepsilon)\bigr)}{N(\delta t/\varepsilon)}\right). \tag{3.13}$$

It can be seen from (3.13) and (3.11) that the error estimate we provide here is sharper than that in [4]. The current estimate gives the right order of accuracy for the time discretization errors. It also takes into account the relaxation and sampling errors in a more accurate and consistent way.

3.2. Proof of strong convergence. We first want to establish some axillary estimates for the estimator $\tilde{a}$.

PROPOSITION 3.6. *Under the assumptions in Theorem* 3.5, *for each* $T_0 > 0$, *there exists an independent constant* $\mathbb{C} > 0$ *such that* $\forall\ n \in [0, T_0/\Delta t]$

$$\begin{aligned}&\mathbb{E}\left|\mathbb{E}_{X_n}\tilde{a}_n - \bar{a}\bigl(X_n\bigr)\right|^2\\ &\quad\le \mathbb{C}\left(\varepsilon^2 + \left(\frac{\delta t}{\varepsilon}\right)^{2\ell} + \frac{e^{-2\beta(n_r+1)(\delta t/\varepsilon)}}{(N(\delta t/\varepsilon)+1)^2}\left(e^{-2\beta n(n_r+N)(\delta t/\varepsilon)} + R\right)\right)\end{aligned} \tag{3.14}$$

and

$$\begin{aligned}&\mathbb{E}\left|\tilde{a}_n - \bar{a}\bigl(X_n\bigr)\right|^2\\ &\quad\le \mathbb{C}\left(\varepsilon^2 + \left(\frac{\delta t}{\varepsilon}\right)^{2\ell} + \frac{e^{-2\beta(n_r+1)(\delta t/\varepsilon)}}{(N(\delta t/\varepsilon)+1)^2}\left(e^{-2\beta n(n_r+N)(\delta t/\varepsilon)} + R\right)\right)\\ &\qquad + \frac{\mathbb{C}}{M(N(\delta t/\varepsilon)+1)}.\end{aligned} \tag{3.15}$$

Proof. Let $\bar{a}(x,\varepsilon) = \int a(x,y,\varepsilon)\mu^\varepsilon(dy)$. Using the smoothness of $\bar{a}(x,\varepsilon)$ established in Lemma A.6, we have $\bar{a}(x) = \bar{a}(x,0)$ and

$$\begin{aligned}\mathbb{E}\left|\mathbb{E}_{X_n}\tilde{a}_n - \bar{a}\bigl(X_n\bigr)\right|^2 &\le 2\mathbb{E}\left|\mathbb{E}_{X_n}\tilde{a}_n - \bar{a}\bigl(X_n,\varepsilon\bigr)\right|^2 + 2\mathbb{E}\left|\bar{a}\bigl(X_n,\varepsilon\bigr) - \bar{a}\bigl(X_n\bigr)\right|^2\\ &\le 2\mathbb{E}\left|\mathbb{E}_{X_n}\tilde{a}_n - \bar{a}\bigl(X_n,\varepsilon\bigr)\right|^2 + \mathbb{C}\varepsilon^2,\end{aligned}$$

which gives the first term on the right-hand sides of (3.14) and (3.15). Letting $Y_{n,m} = Y_{n,m,1}$ and making use of independence, we can compute for the weak estimate (3.14)

$$\mathbb{E}\Big|\mathbb{E}_{X_n}\tilde{a}_n - \bar{a}\big(X_n,\varepsilon\big)\Big|^2 = \frac{1}{M^2N^2}\mathbb{E}\Big|\sum_{m,j}\mathbb{E}_{X_n}a\big(X_n,Y_{n,m,j},\varepsilon\big) - \bar{a}\big(X_n,\varepsilon\big)\Big|^2$$
$$= \frac{1}{N^2}\mathbb{E}\Big|\sum_{m}\mathbb{E}_{X_n}a\big(X_n,Y_{n,m},\varepsilon\big) - \bar{a}\big(X_n,\varepsilon\big)\Big|^2.$$

It is explained in the appendix that when $\Delta\tau = (\delta t/\varepsilon)$ is sufficiently small, for each n, $Y_{n,m}$ is exponentially mixing with unique invariant probability measure μ_n. Lemma A.2 also implies that we can have a process $Z_{n,m}$ over the whole time domain $m \in (-\infty,\infty)$ satisfying the same equation as $Y_{n,m}$ with the stationary distribution μ_n. Now we have

$$\mathbb{E}_{X_n}a\big(X_n,Y_{n,m},\varepsilon\big) - \bar{a}\big(X_n,\varepsilon\big) = \mathbb{E}_{X_n}a\big(X_n,Y_{n,m},\varepsilon\big) - \mathbb{E}_{X_n}a\big(X_n,Z_{n,m},\varepsilon\big)$$
$$+ \mathbb{E}_{X_n}a\big(X_n,Z_{n,m},\varepsilon\big) - \bar{a}\big(X_n,\varepsilon\big).$$

Inequality (A.14) implies that

$$\mathbb{E}\Big|\mathbb{E}_{X_n}a\big(X_n,Z_{n,m},\varepsilon\big) - \bar{a}\big(X_n,\varepsilon\big)\Big|^2 \le \mathbb{C}(\Delta\tau)^{2\ell}.$$

Therefore

$$\begin{aligned}\mathbb{E}\Big|\mathbb{E}_{X_n}a\big(X_n,Y_{n,m},\varepsilon\big) - \bar{a}\big(X_n,\varepsilon\big)\Big|^2 \\ \le \mathbb{C}\Big(\Big|\mathbb{E}_{X_n}a(X_n,Y_{n,m},\varepsilon) - a(X_n,Z_{n,m},\varepsilon)\Big|^2 + \Delta\tau^{2\ell}\Big).\end{aligned} \tag{3.16}$$

Let $V_{n,m}(y) = \nabla_y Y_{n,m}(y)$, where ∇_y denotes derivatives with respect to initial value y. Then we have by (A.12), for some $\xi_n \in \big[Y_{n,0},Z_{n,0}\big]$,

$$\begin{aligned}&\mathbb{E}\Big|\mathbb{E}_{X_n}\Big(a\big(X_n,Y_{n,m},\varepsilon\big) - a\big(X_n,Z_{n,m},\varepsilon\big)\Big)\Big|^2 \\ &= \mathbb{E}\Big|Y_{n,0} - Z_{n,0}\Big|^2\Big|\mathbb{E}_{X_n}\nabla_y a(X_n,Y_{n,m}(\xi_n),\varepsilon)V_{n,m}(\xi_n)\Big|^2 \\ &\le \mathbb{C}e^{-2m\beta(\delta t/\varepsilon)}\mathbb{E}\Big|Y_{n,0} - Z_{n,0}\Big|^2.\end{aligned}$$

Remember we set the initial condition of the fast simulation $Y_{n,0} = Y_{n-1,n_r+N}$. By the same argument as above and (A.12)–(A.13), we can have

$$\begin{aligned}&\mathbb{E}\Big|Y_{n,0} - Z_{n,0}\Big|^2 \\ &= \mathbb{E}\Big|Y_{n-1,n_r+N} - Z_{n,0}\Big|^2 \\ &\le \mathbb{E}\Big|Y_{n-1,n_r+N} - Z_{n-1,n_r+N}\Big|^2 \\ &\quad + 2\Big|\mathbb{E}\big(Y_{n-1,n_r+N} - Z_{n-1,n_r+N}\big)\Big|\Big|\mathbb{E}_{X_{n-1}}\nabla_x Z_{n-1,n_r+N}|_{x=X_{n-1}}(X_n - X_{n-1})\Big| \\ &\quad + \mathbb{E}\Big|\mathbb{E}_{X_{n-1}}\nabla_x^2 Z_{n-1,n_r+N}|_{x=\eta_{n-1}}(X_n - X_{n-1})^2\Big| \\ &\le e^{-2\beta(n_r+N)(\delta t/\varepsilon)}\mathbb{E}\Big|Y_{n-1,0} - Z_{n-1,0}\Big|^2 \\ &\quad + \mathbb{C}\Delta t\end{aligned}$$

for some $\eta_{n-1} \in [X_{n-1}, X_n]$. Repeating the above argument from macro time step $n-1$ till 0, using (A.3), we have

$$\begin{aligned}\mathbb{E}\big|Y_{n,0} - Z_{n,0}\big|^2 &\leq e^{-2\beta n(n_r+N)\Delta\tau}\Big(\mathbb{E}\big|Y_{0,0} - Z_{0,0}\big|^2 + \mathbb{C}\Delta t\Big)\\ &\leq \mathbb{C}'\Big(e^{-2\beta n(n_r+N)\Delta\tau} + R\Big).\end{aligned}$$

Inserting these results in (3.16), we arrive at

$$\begin{aligned}&\mathbb{E}\Big|\mathbb{E}_{X_n} a\big(X_n, Y_{n,m}, \varepsilon\big) - \bar{a}\big(X_n, \varepsilon\big)\Big|^2\\ &\qquad\leq \mathbb{C}\Big(e^{-2\beta m\Delta\tau}\Big(e^{-2\beta n(n_r+N)\Delta\tau} + R\Big) + \Delta\tau^{2\ell}\Big).\end{aligned}$$

Summing over $m \in [n_r+1,\, n_r+N]$, we obtain

$$\begin{aligned}&\mathbb{E}\Big|\mathbb{E}_{X_n}\tilde{a}_n - \bar{a}\big(X_n, \varepsilon\big)\Big|^2\\ &\quad\leq \mathbb{C}\frac{1}{N^2}\left|\sum_{m=n_r+1}^{n_r+N} e^{-\beta m\Delta\tau}\Big(e^{-\beta n(n_r+N)\Delta\tau} + \sqrt{R}\Big) + \Delta\tau^{\ell}\right|^2\\ &\quad\leq \mathbb{C}'\left(e^{-2\beta(n_r+1)\Delta\tau}\frac{\big(1-e^{-\beta N\Delta\tau}\big)^2}{N^2\big(1-e^{-\beta\Delta\tau}\big)^2}\Big(e^{-2\beta n(n_r+N)\Delta\tau} + R\Big) + \Delta\tau^{2\ell}\right)\\ &\quad\leq \mathbb{C}''\left(\frac{e^{-2\beta(n_r+1)\Delta\tau}}{(N\Delta\tau+1)^2}\Big(e^{-2\beta n(n_r+N)\Delta\tau} + R\Big) + \Delta\tau^{2\ell}\right),\end{aligned}$$

and the last two terms on the right-hand side of (3.14) follow.

Next we compute

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

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

$$\begin{aligned}&\left|\sum_{m,l}\sum_{j\neq k}\mathbb{E}\Big(a\big(X_n, Y_{n,m,j}, \varepsilon\big) - \bar{a}\big(X_n, \varepsilon\big)\Big)\cdot\Big(a\big(X_n, Y_{n,l,k}, \varepsilon\big) - \bar{a}\big(X_n, \varepsilon\big)\Big)\right|\\ &\quad\leq \sum_{m,l}\sum_{j\neq k}\mathbb{E}\Big|\mathbb{E}_n\Big(a\big(X_n, Y_{n,m,j}, \varepsilon\big) - \bar{a}\big(X_n, \varepsilon\big)\Big)\\ &\qquad\qquad\cdot\mathbb{E}_n\Big(a\big(X_n, Y_{n,l,k}, \varepsilon\big) - \bar{a}\big(X_n, \varepsilon\big)\Big)\Big|\\ &\quad\leq \mathbb{C}M^2N^2\left(\frac{e^{-2\beta(n_r+1)\Delta\tau}}{(N\Delta\tau+1)^2}\Big(e^{-2\beta n(n_r+N)\Delta\tau} + R\Big) + \Delta\tau^{2\ell}\right),\end{aligned}$$

where E_n denotes the expectation conditioned on $\{X_{n'}\}_{n'\leq n}$, $\{Y_{n',\cdot,\cdot}\}_{n'<n}$. When

$j = k$, we have

$$\left|\mathbb{E}\Big(a\big(X_n, Y_{n,m,j}, \varepsilon\big) - \bar{a}\big(X_n, \varepsilon\big)\Big) \cdot \Big(a\big(X_n, Y_{n,l,j}, \varepsilon\big) - \bar{a}\big(X_n, \varepsilon\big)\Big)\right|$$
$$\leq \mathbb{E}\left(\mathbb{E}_{X_n}\left|a\big(X_n, Y_{n,m,j}, \varepsilon\big) - \bar{a}\big(X_n, \varepsilon\big)\right| \cdot \left|\mathbb{E}_{n,m,j} a\big(X_n, Y_{n,l,j}, \varepsilon\big) - \bar{a}\big(X_n, \varepsilon\big)\right|\right)$$

when $m \leq l$ and similarly when $m > l$. Here $\mathbb{E}_{n,m,j}$ denote the conditional expectation with respect to $Y_{n,m,j}$. The same analysis for (3.16) gives

$$\mathbb{E}\left|\mathbb{E}_{n,m,j} a\big(X_n, Y_{n,l,j}, \varepsilon\big) - \bar{a}\big(X_n, \varepsilon\big)\right|^2 \leq \mathbb{C}\Big(e^{-2\beta(l-m)\Delta\tau} + \Delta\tau^{2\ell}\Big).$$

Summing over $m, l \in [n_r + 1,\, n_r + N]$, this leads to

$$\left|\sum_j \sum_{m,l} \mathbb{E}\Big(a\big(X_n, Y_{n,m,j}, \varepsilon\big) - \bar{a}\big(X_n, \varepsilon\big)\Big) \cdot \Big(a\big(X_n, Y_{n,l,j}, \varepsilon\big) - \bar{a}\big(X_n, \varepsilon\big)\Big)\right|$$
$$\leq \mathbb{C} M N^2 \left(\frac{1}{N\Delta\tau + 1} + \Delta\tau^{2\ell}\right),$$

which gives the last term in (3.15). $\square$

We can also have similar results for $\tilde{\sigma}$.

PROPOSITION 3.7. *Under the assumptions in Theorem* 3.5, *for each* $T_0 > 0$, *there exists an independent constant* $\mathbb{C} > 0$ *such that* $\forall\ n \in [0, T_0/\Delta t]$

$$\begin{aligned} &\mathbb{E}\left\|\mathbb{E}_{X_n}\tilde{\sigma}_n\tilde{\sigma}_n^T - \bar{\sigma}\big(X_n\big)\bar{\sigma}^T\big(X_n\big)\right\|_F^2 \\ &\quad\leq \mathbb{C}\left(\varepsilon^2 + \left(\frac{\delta t}{\varepsilon}\right)^{2\ell} + \frac{e^{-2\beta(n_r+1)(\delta t/\varepsilon)}}{\big(N(\delta t/\varepsilon)+1\big)^2}\Big(e^{-2\beta n(n_r+N)(\delta t/\varepsilon)} + R\Big)\right) \end{aligned} \tag{3.17}$$

and

$$\begin{aligned} &\max\left\{\mathbb{E}\left\|\tilde{\sigma}_n\tilde{\sigma}_n^T - \bar{\sigma}\bar{\sigma}^T\big(X_n\big)\right\|_F^2,\quad \mathbb{E}\left\|\tilde{\sigma}_n - \bar{\sigma}\big(X_n\big)\right\|_F^2\right\} \\ &\quad\leq \mathbb{C}\left(\varepsilon^2 + \left(\frac{\delta t}{\varepsilon}\right)^{2\ell} + \frac{e^{-2\beta(n_r+1)(\delta t/\varepsilon)}}{\big(N(\delta t/\varepsilon)+1\big)^2}\Big(e^{-2\beta n(n_r+N)(\delta t/\varepsilon)} + R\Big)\right) \\ &\qquad + \frac{\mathbb{C}}{M(N(\delta t/\varepsilon)+1)}. \end{aligned} \tag{3.18}$$

Proof. Inequality (3.17) and the first inequality in (3.18) can be obtained by the same proof as for Proposition 3.6. For the second inequality, we know from [12] that when

$$\left\|\big(\bar{\sigma}\bar{\sigma}^T\big(X_n\big)\big)^{-1}\right\|_2 \left\|\tilde{\sigma}_n\tilde{\sigma}_n^T - \bar{\sigma}\bar{\sigma}^T\big(X_n\big)\right\|_F \leq \frac{1}{2}$$

we have

$$\left\|\tilde{\sigma}_n - \bar{\sigma}\big(X_n\big)\right\|_F^2 \leq \mathbb{C}\left\|\tilde{\sigma}_n\tilde{\sigma}_n^T - \bar{\sigma}\bar{\sigma}^T\big(X_n\big)\right\|_F^2,$$

where $\mathbb{C}$ depends on $\|\bar{\sigma}\|_F$, $\|\bar{\sigma}\bar{\sigma}^T\|_2$, and $\|(\bar{\sigma}\bar{\sigma}^T)^{-1}\|_2$, which by Assumptions 3.2 and 3.4 are bounded. Therefore by Assumption 3.1 we can have the following estimate:

$$
\begin{aligned}
\mathbb{E}\left\|\tilde{\sigma}_n-\bar{\sigma}\big(X_n\big)\right\|_F^2 &= \mathbb{E}\left(\chi_{\left\{\|(\bar{\sigma}\bar{\sigma}^T(X_n))^{-1}\|_2\|\tilde{\sigma}_n\tilde{\sigma}_n^T-\bar{\sigma}\bar{\sigma}^T(X_n)\|_F\le\frac{1}{2}\right\}}\left\|\tilde{\sigma}_n-\bar{\sigma}\big(X_n\big)\right\|_F^2\right)\\
&\quad+\mathbb{E}\left(\chi_{\left\{\|(\bar{\sigma}\bar{\sigma}^T(X_n))^{-1}\|_2\|\tilde{\sigma}_n\tilde{\sigma}_n^T-\bar{\sigma}\bar{\sigma}^T(X_n)\|_F>\frac{1}{2}\right\}}\left\|\tilde{\sigma}_n-\bar{\sigma}\big(X_n\big)\right\|_F^2\right)\\
&\le \mathbb{C}\mathbb{E}\left\|\tilde{\sigma}_n\tilde{\sigma}_n^T-\bar{\sigma}\bar{\sigma}^T\big(X_n\big)\right\|_F^2\\
&\quad+\mathbb{C}\mathbb{P}\left\{\left\|\tilde{\sigma}_n\tilde{\sigma}_n^T-\bar{\sigma}\bar{\sigma}^T\big(X_n\big)\right\|_F\left\|\Big(\bar{\sigma}\bar{\sigma}^T\big(X_n\big)\Big)^{-1}\right\|_2>\frac{1}{2}\right\}.
\end{aligned}
$$

Using Chebyshev's inequality and Assumption 3.4, we obtain

$$
\begin{aligned}
\mathbb{E}\left\|\tilde{\sigma}_n-\bar{\sigma}\big(X_n\big)\right\|_F^2 &\le \mathbb{C}\mathbb{E}\left\|\tilde{\sigma}_n\tilde{\sigma}_n^T-\bar{\sigma}\bar{\sigma}^T\big(X_n\big)\right\|_F^2\\
&\quad+4\mathbb{E}\left\|\tilde{\sigma}_n\tilde{\sigma}_n^T-\bar{\sigma}\bar{\sigma}^T\big(X_n\big)\right\|_F^2\left\|\Big(\bar{\sigma}\bar{\sigma}^T\big(X_n\big)\Big)^{-1}\right\|_2^2\\
&\le \mathbb{C}\mathbb{E}\left\|\tilde{\sigma}_n\tilde{\sigma}_n^T-\bar{\sigma}\bar{\sigma}^T\big(X_n\big)\right\|_F^2,
\end{aligned}
$$

which, with the first inequality in (3.18), suggests the second inequality in (3.18). ∎

Proof of Theorem 3.5. We will prove (3.4) for the case when the macro solver is the Euler–Maruyama method. The extension to general stable explicit macro solvers mentioned in section 2 is straightforward. Letting $e_n=X_n-\bar{X}_n$, we have

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

and

$$
\begin{aligned}
\mathbb{E}\,e_{n+1}^2 &= \mathbb{E}\Big|e_n+\Delta t\,\Big(\bar{a}\big(X_n\big)-\bar{a}\big(\bar{X}_n\big)\Big)\Big|^2\\
&\quad+2\Delta t\,\mathbb{E}\Big(e_n+\Delta t\,\Big(\bar{a}\big(X_n\big)-\bar{a}\big(\bar{X}_n\big)\Big)\Big)\cdot\Big(\tilde{a}_n-\bar{a}\big(X_n\big)\Big)\\
&\quad+\Delta t^2\,\mathbb{E}\Big|\tilde{a}_n-\bar{a}\big(X_n\big)\Big|^2\\
&\quad+\Delta t\mathbb{E}\left\|\tilde{\sigma}_n-\bar{\sigma}\big(\bar{X}_n\big)\right\|_F^2\\
&=: I_1+I_2+I_3+I_4.
\end{aligned}
$$

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

$$I_1\le\big(1+\mathbb{C}\Delta t\big)e_n^2.$$

For I_2, we have

$$
\begin{aligned}
|I_2| &\le 2\Delta t\,\mathbb{E}\Big|e_n+\Delta t\,\Big(\bar{a}\big(X_n\big)-\bar{a}\big(\bar{X}_n\big)\Big)\Big|\,\Big|\mathbb{E}_{X_n}\tilde{a}_n-\bar{a}\big(X_n\big)\Big|\\
&\le \Delta t\,\mathbb{E}\Big|e_n+\Delta t\,\Big(\bar{a}\big(X_n\big)-\bar{a}\big(\bar{X}_n\big)\Big)\Big|^2+\mathbb{C}\Delta t\,\mathbb{E}\Big|\mathbb{E}_{X_n}\tilde{a}_n-\bar{a}\big(X_n\big)\Big|^2\\
&\le \Delta t\big(1+\mathbb{C}\Delta t\big)\mathbb{E}e_n^2+\mathbb{C}\Delta t\,\mathbb{E}\Big|\mathbb{E}_{X_n}\tilde{a}_n-\bar{a}\big(X_n\big)\Big|^2.
\end{aligned}
$$

I_4 can be estimated as

$$
\begin{aligned}
|I_4| &\le 2\Delta t \mathbb{E}\Big\|\tilde{\sigma}_n - \bar{\sigma}\big(X_n\big)\Big\|_F^2 + 2\Delta t \mathbb{E}\Big\|\bar{\sigma}\big(\bar{X}_n\big) - \bar{\sigma}\big(X_n\big)\Big\|_F^2 \\
&\le 2\Delta t \mathbb{E}\Big\|\tilde{\sigma}_n - \bar{\sigma}\big(X_n\big)\Big\|_F^2 + \mathbb{C}\Delta t \mathbb{E}\, e_n^2.
\end{aligned}
$$

It follows from (3.14), (3.15), and (3.18) that

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

Therefore, we have

$$
\mathbb{E} e_n^2 \le \mathbb{C}\Bigg(\left(\frac{\delta t}{\varepsilon}\right)^{2\ell} + \varepsilon^2 + \frac{e^{-2\beta(n_r+1)(\delta t/\varepsilon)}}{\big(N(\delta t/\varepsilon)+1\big)^2} R + \frac{1}{M(N(\delta t/\varepsilon)+1)} \Bigg),
$$

which completes the proof of Theorem 3.5. ∎

3.3. Weak convergence rate. Now we can give the rate for the weak convergence of the HMM scheme under the assumptions in Theorem 3.5, except that the macro solver is only supposed to be of weak order k and does not have to be strongly convergent.

THEOREM 3.8. *For any $f \in \mathbb{C}_0^\infty$ and $T_0 > 0$, there exists a constant $\mathbb{C} > 0$ such that*

$$
\begin{aligned}
&\sup_{n \le T_0/\Delta t} \big| f\big(\bar{X}_{t_n}\big) - \mathbb{E} f\big(X_n\big) \big| \\
&\qquad \le \mathbb{C}\Bigg(\Delta t^k + \varepsilon + \left(\frac{\delta t}{\varepsilon}\right)^{\ell} + \frac{e^{-\beta(n_r+1)(\delta t/\varepsilon)}}{N(\delta t/\varepsilon)+1} Q \Bigg),
\end{aligned} \tag{3.19}
$$

where

$$
Q = \frac{\Delta t}{1 - e^{-\beta(n_r+N)(\delta t/\varepsilon)}}. \tag{3.20}
$$

Proof. Again for simplicity we will discuss only the case when the macro solver is the Euler–Maruyama method ($k = 1$). Generalizations to other stable explicit solvers will require little modification. Using the same argument as before, the first term on the right-hand side of (3.19) arises from $\big|\mathbb{E} f(\bar{X}_{t_n}) - \mathbb{E} f(\bar{X}_n)\big|$ and the fact that the macro solver is weakly convergent. To estimate $\big|\mathbb{E} f(\bar{X}_n) - \mathbb{E} f(X_n)\big|$, we define an auxiliary function $u(m, x)$ for $m \le n$ as follows:

$$
u(n, x) = f(x), \qquad u(m, x) = u\Big(m+1, x + \bar{a}(x)\Delta t + \bar{\sigma}(x)\Delta W_{m+1}\Big).
$$

By this definition, $u(m, x)$ depends only on $\big\{\Delta W_{m+1}, \dots, \Delta W_n\big\}$ and $\mathbb{E} u(0, x) = \mathbb{E} f(\bar{X}_n)$. By the smoothness of $\bar{a}(\cdot)$ and the fact the $f \in \mathbb{C}_0^\infty$, it is easy to show that, for each index I, $|\nabla_x^I u(m,x)|$ is uniformly bounded for different x, m, and Δt. Hence we have

$$
\begin{aligned}
&\Big|\mathbb{E}\Big(u\big(m+1,X_{m+1}\big)-u\big(k,X_m\big)\Big)\Big| \\
&\quad=\Big|\mathbb{E}\Big(u\Big(m+1,X_m+\tilde a_m\Delta t+\tilde\sigma_m\Delta W_{m+1}\Big) \\
&\qquad\qquad -u\Big(m+1,X_m+\bar a\big(X_m\big)\Delta t+\bar\sigma\big(X_m\big)\Delta W_{m+1}\Big)\Big)\Big| \\
&\quad\le \Delta t\Big|\mathbb{E}\,\nabla_x u\big(m+1,X_m\big)\cdot\Big(\mathbb{E}_{X_m}\tilde a_m-\bar a\big(X_m\big)\Big)\Big| \\
&\qquad + \Delta t\mathbb{C}\Big|\mathbb{E}\nabla_x^2 u\big(m+1,X_m\big)\cdot\Big(\mathbb{E}_{X_m}\tilde\sigma_m^T\tilde\sigma_m-\bar\sigma^T\bar\sigma\big(X_m\big)\Big)\Big| \\
&\qquad + \mathbb{C}\Delta t^2.
\end{aligned}
$$

Therefore, by Propositions 3.6 and 3.7 and the fact that $\sqrt{R}\le Q$, we have

$$
\begin{aligned}
\Big|\mathbb{E}f\big(X_n\big)-f\big(\bar X_n\big)\Big| &= \big|\mathbb{E}\,u\big(n,X_n\big)-u(0,x)\big| \\
&= \Bigg|\sum_{0\le m\le n-1}\mathbb{E}\Big(u\big(m+1,X_{m+1}\big)-u\big(m,X_m\big)\Big)\Bigg| \\
&\le \mathbb{C}\left(\varepsilon+\left(\frac{\delta t}{\varepsilon}\right)^{\ell}\right) \\
&\quad + \mathbb{C}\sum_{0\le m\le n-1}\Delta t\frac{e^{-\beta(n_r+1)(\delta t/\varepsilon)}}{N(\delta t/\varepsilon)+1}\Big(e^{-\beta m(n_r+N)(\delta t/\varepsilon)}+Q\Big) \\
&\quad + \mathbb{C}\Delta t,
\end{aligned}
$$

which gives (3.19). ∎

The optimality of (3.19) can also be shown by system (3.7). Suppose we choose $f(x)=x^2$. The weak error can be calculated such that

$$
\begin{aligned}
\mathbb{E}\big(\bar X_{t_n}\big)^2-\mathbb{E}\big(X_n\big)^2 &= \Delta t\sum_{k=0}^{n-1}\left(\mathbb{E}\frac{1}{MN}\sum_{m=n_r+1}^{n_r+N}\sum_{j=1}^{M}\left(Y_{k,m,j}-\frac{1}{2}\right)^2-\frac{3}{4}\right) \\
&= \left(\Delta\tau+\frac{e^{-(n_r+1)\Delta\tau}}{\big(N\Delta\tau+1\big)}\frac{\Delta t}{\big(1-e^{-(n_r+N)\Delta\tau}\big)}\right).
\end{aligned}
$$

4. A numerical example.

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

$$
\begin{aligned}
\dot X_t^\varepsilon &= -\big(Y_t^\varepsilon\big)^3+\sin(2\pi t)+Y_t^\varepsilon\,\dot W_t, &\qquad X_0^\varepsilon &= x, \\
\dot Y_t^\varepsilon &= -\frac{1}{\varepsilon}\big(Y_t^\varepsilon-X_t^\varepsilon\big)+\frac{1}{\sqrt{\varepsilon}}\,\dot B_t, & Y_0^\varepsilon &= y.
\end{aligned}
\tag{4.1}
$$

From (1.2), the effective dynamics for X_t^ε is

$$
\dot{\bar X}_t = -\big(\bar X_t\big)^3-\frac{3}{2}\bar X_t+\sin(2\pi t)+\sqrt{\big(\bar X_t\big)^2+\frac{1}{2}}\,\dot W_t, \qquad \bar X_0 = x. \tag{4.2}
$$

Since the error caused by the macro solver is standard, we analyze only the difference between the numerical solution X_n given by the multiscale scheme and the numerical

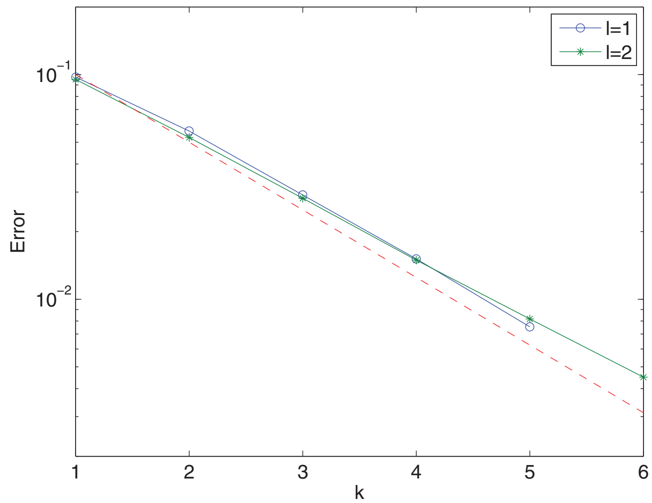

FIG. 4.1. *The error $E_p^\ell = 1/K\sum_{k=1}^K(\Delta t/T_0)\sum_{n\le\lfloor T_0/\Delta t\rfloor}|\tilde{X}_n-\bar{X}_n|$ in the function of p when $\ell=1$ (circles) and $\ell=2$ (stars). Also shown is the predicted estimate, 0.1×2^{-p} (red line).*

TABLE 4.1
The computed values for the error E_p^ℓ.

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

solution $\bar{X}_n$ given by the macro solver for (4.2). In other words, we analyze the error caused by using $\tilde{a}_n$ and $\tilde{\sigma}_n$ instead of $\bar{a}(X_n)$ and $\bar{\sigma}(X_n)$. As explained before, since the coefficients of the slow dynamics in (4.1) are independent of ε, the corresponding ε terms in the error (3.4) and (3.19) do not exist. From Theorem 3.5, we know that when the numerical parameters are chosen to satisfy (3.12) the strong error has the form

$$\mathbb{E}\big|X_n-\bar{X}_n\big|\le\left(\left(\frac{\delta t}{\varepsilon}\right)^\ell+\frac{e^{-\beta(n_r+1)(\delta t/\varepsilon)}}{N(\delta t/\varepsilon)}\sqrt{\Delta t}+\frac{1}{\sqrt{MN(\delta t/\varepsilon)}}\right).$$

Notice that Δt is a fixed parameter here.

Suppose we want to bound the error by $O(2^{-p})$ for $p=0,1,2,\dots$ and, for that purpose, we choose the parameters such that

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

In the numerical calculations, we compute the following error estimate averaged with an ensemble of $K=10^4$ realizations of the macro dynamics numerically obtained by the Euler–Maruyama method:

$$E_p^\ell=\frac{1}{K}\sum_{k=1}^K\left(\frac{\Delta t}{T_0}\sum_{n\le T_0/\Delta t}\big|\tilde{X}_n-\bar{X}_n\big|\right).$$

Figure 4.1 shows the error for different values of p. It can be seen that the error decays logarithmically as p increases. The deviation from the error estimate when p is large

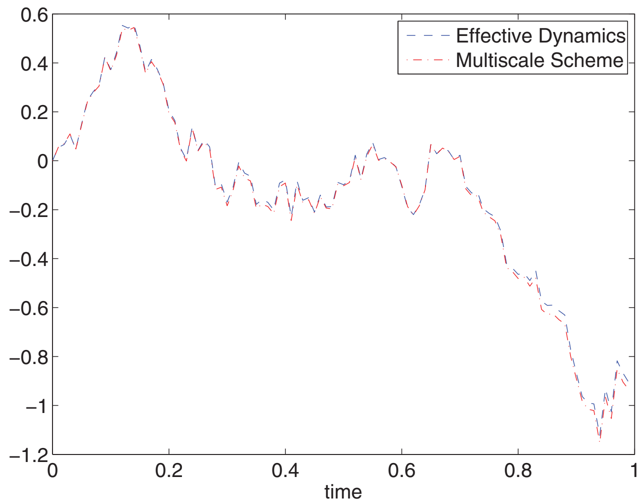

FIG. 4.2. *The comparison between $\bar{X}_n$ (blue curve) and X_n produced by the multiscale scheme with $\ell = 2$, $p = 4$ (red curve).*

is due to the sampling with finite realizations of the macro dynamics. According to the central limit theorem, this error is $O(10^{-2})$ if $K = 10^4$, which is of the same order as the error of the multiscale scheme for large p. One realization of the effective dynamics and the numerical result from the multiscale scheme is shown in Figure 4.2.

Appendix. Limiting properties of the fast processes. Here we want to provide some properties for the fast process Y_t defined by the following dynamics:

$$\dot{Y}_t = \frac{1}{\varepsilon}B\big(x, Y_t, \varepsilon\big) + \frac{1}{\sqrt{\varepsilon}}C\big(x, Y_t, \varepsilon\big)\dot{W}_t, \qquad Y_0 = y, \tag{A.1}$$

which is the same process that the micro solver simulates in the HMM scheme. The ergodicity of Y_t with a unique invariant measure $\mu_x^\varepsilon(dy)$ for each fixed (x, ε) under Assumptions 3.1–3.3 has been established in [6]. We want to provide some sharper estimates for the purpose of proving our theorems. First, we give an energy estimate for Y_t and its invariant measure.

LEMMA A.1. *There exists a constant $\mathbb{C}$ such that $\forall\ t \geq 0$ we have*

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

and the invariant measure μ_x^ε for Y_t has a finite second order moment

$$\int y^2 \mu_x^\varepsilon(dy) \leq \mathbb{C}\big(|x|^2 + \varepsilon^2 + 1\big), \tag{A.3}$$

where $\mathbb{C}$ is the same constant as in (A.2).

Proof. Fixing $y_1 = y$ and $y_2 = 0$ in Assumption 3.3 will give us

$$\big\langle y, B(x, y, \varepsilon) - B(x, 0, \varepsilon)\big\rangle + \big\|C(x, y, \varepsilon) - C(x, 0, \varepsilon)\big\|^2 \leq -\beta\big|y\big|^2. \tag{A.4}$$

By Assumption 3.1, we have, for any $\gamma > 0$,

$$\begin{aligned}\big\|C(x,y,\varepsilon)\big\|^2 &\le (1+\gamma)\big\|C(x,y,\varepsilon)-C(x,0,\varepsilon)\big\|^2 \\ &\qquad + (1+1/\gamma)\|C(x,0,\varepsilon)\|^2 \\ &\le \big\|C(x,y,\varepsilon)-C(x,0,\varepsilon)\big\|^2 \\ &\qquad + \mathbb{C}_1\gamma|y|^2 + \mathbb{C}_2(1+1/\gamma)\big(|x|^2+\varepsilon^2+1\big),\end{aligned} \tag{A.5}$$

where $\mathbb{C}_1$ and $\mathbb{C}_2$ are the Lipschitz and linear growth constants of function $C(\cdot)$, respectively. If we choose an appropriate value for γ such that $\mathbb{C}_1\gamma \le \beta/4$, by Assumption 3.3 and (A.4)–(A.5), we can get

$$\begin{aligned}&\big\langle y, B(x,y,\varepsilon)\big\rangle + \big\|C(x,y,\varepsilon)\big\|^2 \\ &\quad\le \big\langle y, B(x,y,\varepsilon)-B(x,0,\varepsilon)\big\rangle + \big\langle y, B(x,0,\varepsilon)\big\rangle \\ &\qquad + \big\|C(x,y,\varepsilon)-C(x,0,\varepsilon)\big\|^2 + \frac{\beta}{4}|y|^2 + \mathbb{C}'\big(|x^2|+\varepsilon^2+1\big) \\ &\quad\le -\frac{\beta}{2}|y|^2 + \mathbb{C}''\big(|x|^2+\varepsilon^2+1\big).\end{aligned}$$

Itô's formula then suggests that

$$\begin{aligned}d\mathbb{E}\big|Y_t\big|^2 &= \frac{2}{\varepsilon}\mathbb{E}\big\langle Y_t, B\big(x,Y_t\big)\big\rangle dt + \frac{1}{\varepsilon}\mathbb{E}\big\|C\big(x,Y_t,\varepsilon\big)\big\|^2 dt \\ &\le -\frac{\beta}{\varepsilon}\mathbb{E}\big|Y_t\big|^2 dt + \frac{\mathbb{C}}{\varepsilon}\big(|x|^2+\varepsilon^2+1\big)dt.\end{aligned}$$

Gronwall's inequality then suggests that

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

Taking $t \to \infty$, ergodicity gives (A.3). ∎

Based on ergodicity and Lemma A.1, we can prove the existence of a stationary solution for (A.1) over the whole time domain.

Lemma A.2. *For each fixed (x,ε), there exists a stationary process Z_t defined over $t \in (-\infty,\infty)$ satisfying*

$$\dot{Z}_t = \frac{1}{\varepsilon}B\big(x,Z_t,\varepsilon\big) + \frac{1}{\sqrt{\varepsilon}}C\big(x,Z_t,\varepsilon\big)\dot{W}_t \tag{A.6}$$

such that the law of Z_t is invariant, i.e.,

$$\mathcal{L}(Z_t) = \mu_x^\varepsilon.$$

Proof. Define the process $Z_{\tau,t}$ to be the solution of the following equation on time domain (τ,∞):

$$dZ_{\tau,t} = \frac{1}{\varepsilon}B\big(x,Z_{\tau,t},\varepsilon\big)dt + \frac{1}{\sqrt{\varepsilon}}C\big(x,Z_{\tau,t},\varepsilon\big)dW_t, \qquad Z_{\tau,\tau} = z.$$

For $\tau_2 < \tau_1 \le 0$, by Itô's formula and Assumption 3.4, we have

$$\begin{aligned}d\mathbb{E}\big|Z_{\tau_1,t}-Z_{\tau_2,t}\big|^2 &= \frac{2}{\varepsilon}\mathbb{E}\big\langle Z_{\tau_1,t}-Z_{\tau_2,t},\ B\big(x,Z_{\tau_1,t},\varepsilon\big)-B\big(x,Z_{\tau_2,t},\varepsilon\big)\big\rangle dt \\ &\qquad + \frac{1}{\varepsilon}\mathbb{E}\,\big\|C\big(x,Z_{\tau_1,t},\varepsilon\big)-C\big(x,Z_{\tau_2,t},\varepsilon\big)\big\|^2 dt \\ &\le -\frac{2\beta}{\varepsilon}\mathbb{E}\big|Z_{\tau_1,t}-Z_{\tau_2,t}\big|^2 dt.\end{aligned}$$

Using Lemma A.1, we have

$$\begin{aligned}\mathbb{E}\big|Z_{\tau_1,t} - Z_{\tau_2,t}\big|^2 &\le \mathbb{E}\big|Z_{\tau_2,\tau_1} - z\big|^2 e^{-2\beta(t-\tau_1)/\varepsilon}\\ &\le \mathbb{C}\big(x^2 + z^2 + \varepsilon^2 + 1\big)e^{-2\beta(t-\tau_1)/\varepsilon}.\end{aligned}$$

For any monotonically decreasing sequence $\{\tau_n\}$ such that $|\tau_{n+1} - \tau_n| \ge 1$ and $\tau_n \to -\infty$ as $n \to \infty$, we have

$$\begin{aligned}\mathbb{P}\big\{\big|Z_{\tau_{n+1},t} - Z_{\tau_n,t}\big| > 1/|\tau_n|^2\big\} &\le |\tau_n|^2\mathbb{E}\big|Z_{\tau_{n+1},t} - Z_{\tau_n,t}\big|^2\\ &\le \mathbb{C}\big(x^2 + z^2 + \varepsilon^2 + 1\big)|\tau_n|^2 e^{-\beta(t-\tau_n)/\varepsilon}.\end{aligned}$$

Since $\sum_n |\tau_n|^2 e^{\beta\tau_n/\varepsilon} < \infty$, by the Borel–Cantelli lemma, we know that, with probability one, $Z_{\tau_n,t}$ satisfies

$$\big|Z_{\tau_{n+1},t} - Z_{\tau_n,t}\big| \le 1/|\tau_n|^2$$

when $n \ge N(\omega)$. Therefore we know that $\{Z_{\tau_n,t}\}$ almost surely is a converging sequence when $n \to \infty$. By the arbitrariness of $\{\tau_n\}$, we know that, with probability one, $Z_{\tau,t}$ converges when $\tau \to -\infty$. So we can define

$$Z_t = \lim_{\tau\to-\infty} Z_{\tau,t}.$$

Notice that, for any $s < t$, $Z_{\tau,t}$ satisfies the integral equation

$$Z_{\tau,t} = Z_{\tau,s} + \frac{1}{\varepsilon}\int_s^t B\big(x, Z_{\tau,\omega}, \varepsilon\big)d\omega + \frac{1}{\varepsilon}\int_s^t C\big(x, Z_{\tau,\omega}, \varepsilon\big)dW_\omega.$$

Taking $\tau \to -\infty$ in the above equation suggests that Z_t satisfies (A.6). By ergodicity and translation invariance of (A.6), we have

$$\mathcal{L}\big(Z_t\big) = \mu_x^\varepsilon. \qquad \square$$

The following lemmas describe how the solution of (A.1) behaves when we perturb the initial condition.

LEMMA A.3. *Suppose Y_t is the solution of* (A.1) *with initial value $Y_0 = y$. We have that the following estimate holds uniformly $\forall\ (x, \varepsilon)$:*

$$\mathbb{E}\big|\nabla_y Y_t\big|^2 \le e^{-2\beta t/\varepsilon}. \tag{A.7}$$

Proof. Letting $y_1 = y + \theta z$ and $y_2 = y$ in Assumption 3.3, we have

$$\begin{aligned}&\Big\langle z,\ \frac{1}{\theta}\Big(B(x, y+\theta z, \varepsilon) - B(x,y,\varepsilon)\Big)\Big\rangle\\ &\qquad + \frac{1}{\theta^2}\big\|C(x, y+\theta z, \varepsilon) - C(x,y,\varepsilon)\big\|^2\\ &\le -\beta|z|^2.\end{aligned}$$

Taking $\theta \to 0$ in the above inequality, we get

$$\big\langle z,\ \nabla_y B(x,y,\varepsilon)z\big\rangle + \big\|\nabla_y C(x,y,\varepsilon)z\big\|^2 \le -\beta|z|^2. \tag{A.8}$$

Meanwhile, the equation satisfied by $U_t = \nabla_y Y_t$ has the following form:

$$dU_t = \frac{1}{\varepsilon}\nabla_y B(x, Y_t, \varepsilon)U_t\, dt + \frac{1}{\sqrt{\varepsilon}}\nabla_y C(x, Y_t, \varepsilon)U_t\, dW_t, \qquad U_t = I.$$

Then, by Itô's formula and (A.8), we have

$$\begin{aligned} d\big|U_t\big|^2 &= \frac{2}{\varepsilon}\langle U_t, \nabla B(x, Y_t, \varepsilon)U_t\rangle dt + \frac{1}{\varepsilon}\|\nabla_y C(x, Y_t, \varepsilon)U_t\|^2 dt \\ &\quad + \frac{2}{\sqrt{\varepsilon}}\big\langle U_t, \nabla_y C(x, Y_t, \varepsilon)U_t\, dW_t\big\rangle \\ &\le -\frac{2}{\varepsilon}\beta\big|U_t\big|^2 dt + \frac{2}{\sqrt{\varepsilon}}\big\langle U_t, \nabla_y C(x, Y_t, \varepsilon)U_t\, dW_t\big\rangle, \end{aligned}$$

which gives (A.7). ∎

Now we want to prove the stability of Z_t under perturbation of parameter x.

LEMMA A.4. *Suppose Z_t is the solution of* (A.6) *with parameter x. Then, for any multi-index I, we have*

$$\mathbb{E}\big|\nabla_x^I Z_t\big|^2 \le \mathbb{C}_I. \tag{A.9}$$

Proof. We first focus on the first order derivative. By definition, the equation satisfied by $V_t = \nabla_x Z_t$ has the following form:

$$\begin{aligned} dV_t &= \frac{1}{\varepsilon}\nabla_y B\big(x, Z_t, \varepsilon\big)V_t dt + \frac{1}{\varepsilon}\nabla_x B\big(x, Z_t, \varepsilon\big)dt \\ &\quad + \frac{1}{\sqrt{\varepsilon}}\nabla_y C\big(x, Z_t, \varepsilon\big)V_t dW_t + \frac{1}{\sqrt{\varepsilon}}\nabla_x C\big(x, Z_t, \varepsilon\big)dW_t. \end{aligned} \tag{A.10}$$

Using Assumption 3.1 and (A.8), by Itô's formula, we have

$$\begin{aligned} d\mathbb{E}\big|V_t\big|^2 &= \frac{2}{\varepsilon}\mathbb{E}\langle V_t, \nabla_y B\big(x, Z_t, \varepsilon\big)V_t\rangle dt + \frac{2}{\varepsilon}\mathbb{E}\big\langle V_t, \nabla_x B\big(x, Z_t, \varepsilon\big)\big\rangle dt \\ &\quad + \frac{1}{\varepsilon}\mathbb{E}\big\|\nabla_y C\big(x, Z_t, \varepsilon\big)V_t + \nabla_x C\big(x, Z_t, \varepsilon\big)\big\|^2 dt \\ &\le -\frac{2\beta}{\varepsilon}\mathbb{E}\big|V_t\big|^2 dt + \frac{2}{\varepsilon}\mathbb{E}\big\langle V_t, \nabla_x B\big(x, Z_t, \varepsilon\big)\big\rangle dt \\ &\quad + \frac{2}{\varepsilon}\mathbb{E}\big\|\nabla_x C\big(x, Z_t, \varepsilon\big)\big\|^2 dt \\ &\le -\frac{\beta}{\varepsilon}\mathbb{E}\big|V_t\big|^2 dt + \frac{\mathbb{C}}{\varepsilon}dt, \end{aligned} \tag{A.11}$$

which gives (A.9) for V_t. Recursively repeating the same argument for higher order derivatives gives (A.9) for arbitrary index I. ∎

Suppose Y_m is a numerical solution of Y_t using either micro solver (2.5) or (2.6) with a time step of $\Delta\tau = (\delta t/\varepsilon)$. The ergodicity of Y_m when $\Delta\tau$ is sufficiently small under Assumptions 3.1–3.3 can be easily established by the theorems in [9, 10, 11]. Suppose $\mu^\varepsilon_{x,\Delta\tau}$ is the unique invariant measure of Y_m. Using the same arguments for Lemmas A.1–A.3, we can parallel results for Y_m.

LEMMA A.5. *Suppose $(\delta t/\varepsilon)$ is sufficiently small; then we have the following:*

(i) *For each fixed (x, ε), there exists a process Z_m defined over $m \in (-\infty, \infty)$ satisfying the same equations as Y_m with the stationary distribution $\mathcal{L}(Z_m) = \mu^\varepsilon_{x,\delta t/\varepsilon}$.*

(ii) *Suppose the initial value of* $Y_0 = y$*; then the following estimate holds* $\forall\ (x,\varepsilon)$*:*

$$\mathbb{E}\big|\nabla_y Y_m\big|^2 \le e^{-2\beta t/\varepsilon}. \tag{A.12}$$

(iii) *For any multi-index* I*, we have*

$$\mathbb{E}\big|\nabla_x^I Z_m\big|^2 \le \mathbb{C}_I. \tag{A.13}$$

It is also proved in [13] that under Assumptions 3.1–3.3, for any arbitrary smooth function f with its derivatives having at most polynomial growth at infinity, for sufficiently small $(\delta t/\varepsilon)$, we have

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

where ℓ is the order of the micro solver (2.5)–(2.6) and $\mathbb{C}$ is a constant depending on the derivatives of f.

Now we prove some regularity for the coefficients of the effective dynamics (1.2). Define

$$\bar{a}(x,\varepsilon) = \int a(x,y,\varepsilon)\mu^\varepsilon(dy) \tag{A.15}$$

and

$$\bar{\Sigma}(x,\varepsilon) = \int \sigma(x,y,\varepsilon)\sigma^T(x,y,\varepsilon)\mu^\varepsilon(dy). \tag{A.16}$$

LEMMA A.6. *Functions* $\bar{a}(x,\varepsilon)$ *and* $\bar{\Sigma}(x,\varepsilon)$ *defined by* (A.15) *and* (A.16) *are smooth functions in terms of* (x,ε) *with bounded derivatives.*

Proof. The differentiability of the solution of (A.1) with respect to the initial condition and parameters under Assumption 3.1 is established in [7]. Therefore we have the smoothness of $\bar{a}(\cdot)$. Here we want to show the boundedness of the derivatives. Let us now focus on the first order derivative $\nabla_x\bar{a}(\cdot)$. Remember that Y_t denotes the solution of (A.1) with initial condition y and parameter (x,ε). Defining

$$H_t = \nabla_x Y_t,$$

we have

$$\begin{aligned} dH_t = {} & \frac{1}{\varepsilon}\nabla_y B\big(x,Y_t,\varepsilon\big)H_t dt + \frac{1}{\varepsilon}\nabla_x B\big(x,Y_t,\varepsilon\big)dt \\ & + \frac{1}{\sqrt{\varepsilon}}\nabla_y C\big(x,Y_t,\varepsilon\big)H_t dW_t + \frac{1}{\sqrt{\varepsilon}}\nabla_x C\big(x,Y_t,\varepsilon\big)dW_t. \end{aligned}$$

Using Assumption 3.1 and (A.8), by the same argument for Lemma A.4, we get

$$\mathbb{E}\big|H_t\big|^2 \le \mathbb{C}.$$

The boundedness of $\nabla_x\bar{a}(\cdot)$ follows from the above estimate and Assumption 3.2, since

$$\begin{aligned} \nabla_x\bar{a}(x,\varepsilon) &= \lim_{t\to\infty}\nabla_x\mathbb{E}a\big(x,Y_t,\varepsilon\big) \\ &= \lim_{t\to\infty}\big(\mathbb{E}\nabla_x a\big(x,Y_t,\varepsilon\big) + \mathbb{E}\nabla_y a\big(x,Y_t,\varepsilon\big)H_t\big). \end{aligned}$$

Letting

$$G_t = \nabla_\varepsilon Y_t,$$

the boundedness of $\nabla_\varepsilon \bar{a}(\cdot)$ can be obtained by repeating the above argument for G_t. Reiterating the same argument as above, we have have boundedness for higher order derivatives of $\bar{a}(x, \varepsilon)$. The smoothness of $\bar{\Sigma}(\cdot)$ and the boundedness of its derivatives can be established in the same way. $\square$

Acknowledgments. The author is grateful to Weinan E and Eric Vanden-Eijnden for stimulating discussions.

REFERENCES

[1] W. E and B. Engquist, *The heterogeneous multiscale methods*, Commun. Math. Sci., 1 (2003), pp. 87–132.

[2] W. E, B. Engquist, X. Li, W. Ren, and E. Vanden-Eijnden, *Heterogeneous multiscale methods: A review*, Commun. Comput. Phys., 2 (2007), pp. 367–450.

[3] W. E, D. Liu, and E. Vanden-Eijnden, *Analysis of multiscale methods for stochastic differential equations*, Comm. Pure Appl. Math., 58 (2005), pp. 1544–1585.

[4] D. Givon, I. G. Kevrekidis, and R. Kupferman, *Strong convergence of projective integration schemes for singularly perturbed stochastic differential systems*, Commun. Math. Sci., 4 (2006), pp. 707–729.

[5] R. Z. Khasminskii, *On an averaging principle for Itô stochastic differential equations*, Kybernetika, 4 (1968), pp. 260–279.

[6] R. Z. Khasminskii, *Stochastic Stability of Differential Equations*, Sijthoff & Noordhoff, Alphen aan den Rijn, The Netherlands, Germantown, MD, 1980.

[7] H. Kunita, *Stochastic Flows and Stochastic Differential Equations*, Cambridge University Press, Cambridge, UK, 1990.

[8] D. Liu, *Strong convergence of principle of averaging for multiscale stochastic dynamical systems*, Commun. Math. Sci., to appear.

[9] S. P. Meyn and R. L. Tweedie, *Stability of Markovian processes,* I, Adv. in Appl. Probab., 24 (1992), pp. 542–574.

[10] S. P. Meyn and R. L. Tweedie, *Stability of Markovian processes,* II, Adv. in Appl. Probab., 25 (1993), pp. 487–517.

[11] S. P. Meyn and R. L. Tweedie, *Stability of Markovian processes,* III, Adv. in Appl. Probab., 25 (1993), pp. 518–548.

[12] J. G. Sun, *Perturbation bounds for the Cholesky and QR factorization*, BIT, 31 (1991), pp. 341–352.

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

[14] E. Vanden-Eijnden, *Numerical techniques for multi-scale dynamical systems with stochastic effects*, Commun. Math. Sci., 1 (2003), pp. 377–384.

[15] E. Vanden-Eijnden, *On HMM-like integrators and projective integration methods for systems with multiple time scales*, Commun. Math. Sci., 5 (2007), pp. 495–505.