Front. Math. China 2012, 7(2): 305–320
DOI 10.1007/s11464-012-0193-6

RESEARCH ARTICLE

Strong convergence rate of principle of averaging for jump-diffusion processes

Di LIU

Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA

Abstract We study jump-diffusion processes with two well-separated time scales. It is proved that the rate of strong convergence to the averaged effective dynamics is of order $O(\varepsilon^{1/2})$, where $\varepsilon \ll 1$ is the parameter measuring the disparity of the time scales in the system. The convergence rate is shown to be optimal through examples. The result sheds light on the designing of efficient numerical methods for multiscale stochastic dynamics.

Keywords Stochastic differential equation, time scale separation, averaging of perturbations
MSC 60H10, 70K65

1 Introduction

Consider the following jump-diffusion system with a time scale separation measured by $\varepsilon \ll 1$:

$$\begin{aligned} \dot{X}_t^\varepsilon &= a(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon) + b(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon)\dot{B}_t + c(X_{t-}^\varepsilon, Y_{t-}^\varepsilon, \varepsilon)\dot{P}_t, \quad X_0^\varepsilon = x, \\ \dot{Y}_t^\varepsilon &= \frac{1}{\varepsilon} f(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon) + \frac{1}{\sqrt{\varepsilon}} g(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon)\dot{W}_t + h(X_{t-}^\varepsilon, Y_{t-}^\varepsilon, \varepsilon)\dot{N}_t^\varepsilon, \quad Y_0^\varepsilon = y, \end{aligned} \tag{1.1}$$

where $X_t^\varepsilon \in \mathbb{R}^n$ and $Y_t^\varepsilon \in \mathbb{R}^m$ are variables in vector spaces; B_t and W_t are independent standard Wiener processes of dimension d_1 and d_2, respectively; P_t is a scalar Poisson process with constant intensity λ_1; and N_t^ε is an uncorrelated scalar Poisson process with intensity λ_2/ε. Notice that as functions, $a, c \in \mathbb{R}^n$, $f, h \in \mathbb{R}^m$, $b \in \mathbb{R}^{n\times d_1}$, and $g \in \mathbb{R}^{m\times d_2}$ are of $O(1)$ magnitude. Systems in form of (1.1) arise from a wide range of applications including chemical kinetics, complex fluids, and financial engineering. We have assumed that the phase space can be decomposed into slow degrees of freedom x and fast degrees of freedom y. Under appropriate assumptions on f, g, and h, the dynamics for

Received January 16, 2011; accepted September 6, 2011
E-mail: richardl@math.msu.edu

Y_t^ε with $X_t^\varepsilon = x$ fixed is ergodic with a unique invariant measure $\mu_x^\varepsilon(\mathrm{d}y)$. In this case, the Principle of Averaging has been proved such that in the limit of $\varepsilon \to 0$, X_t^ε converges to a stochastic differential equation in the following form:

$$\dot{\overline{X}}_t = \overline{a}(\overline{X}_t) + \overline{b}(\overline{X}_t)\dot{B}_t + \int c(\overline{X}_{t-}, \nu)\dot{p}_{\overline{X}_{t-}}(\mathrm{d}\nu), \quad \overline{X}_0 = x, \tag{1.2}$$

where

$$\begin{aligned} \overline{a}(x) &= \lim_{\varepsilon\to 0} \int a(x, y, \varepsilon)\mu_x^\varepsilon(\mathrm{d}y), \\ \overline{b}(x)\overline{b}^{\mathrm{T}}(x) &= \lim_{\varepsilon\to 0} \int b(x, y, \varepsilon)b^{\mathrm{T}}(x, y, \varepsilon)\mu_x^\varepsilon(\mathrm{d}y). \end{aligned} \tag{1.3}$$

The jump component in (1.2) is such that

$$c(x, y) = \lim_{\varepsilon\to 0} c(x, y, \varepsilon)$$

and p_x is a Poisson random measure with jump rate λ_1 and jump size distribution

$$\mu_x = \lim_{\varepsilon\to 0} \mu_x^\varepsilon.$$

From the point of view of numerical analysis, an important question is in which sense, as well as how fast, the multiscale system (1.1) will converge to the effective dynamics (1.2). The recent motivation for this problem is the progress on numerical methods for dynamical systems with multiple time scales. In [17], 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], the scheme consists of a macro solver to evolve (1.2) and a micro solver for the fast dynamics in (1.1). An estimator is chosen to estimate the coefficients $\overline{a}(\cdot)$ and $\overline{b}(\cdot)$ on-the-fly at each time step of the macro solver using data obtained from the fast simulations with the micro solver. Without having to resolve all the details of the fast process on the $O(\varepsilon)$ time scale, the method is able to overcome the numerical stiffness induced by the time scale separation. To fully justify this strategy, we need an accurate quantitative estimate on the validity of the effective dynamics (1.2).

Much progress has been made for the situation when there is no jump ($c = h = 0$) in multiscale system (1.1). The convergence in probability of (1.1) to (1.2) has been proved in [3,18] with no explicit convergence rate being given. The weak convergence has been proved using the asymptotic expansion of the the backward operator [6,8], which implies that the convergence rate is $O(\varepsilon)$. The strong convergence for (1.1) has been studied under the condition that the diffusion in the slow dynamics is independent of the fast variable, i.e.,

$$b(x, y, \varepsilon) = b(x, \varepsilon). \tag{1.4}$$

Assuming (1.4), the strong convergence rates were provided as $O(\varepsilon^{1/6})$ in [9] and as $O(\varepsilon^{1/4})$ in [5]. In [2,11], it is shown that the optimal strong convergence rate should be $O(\varepsilon^{1/2})$ in this case. For systems with jumps, the weak convergence was established in [10]. A strong convergence rate of $O(\ln \varepsilon)$ was proved in [4]. In this paper, we will show that the optimal strong convergence rate for (1.1) should still be $O(\varepsilon^{1/2})$ even when there are jump components in (1.1), which is a significant improvement of the result in [4].

Throughout the paper, we denote by $\mathscr{C}$ a generic constant that does not have to have the same value. In chains of inequalities, we will adopt $\mathscr{C}$, $\mathscr{C}'$, $\mathscr{C}''$, ... or $\mathscr{C}_1$, $\mathscr{C}_2$, $\mathscr{C}_3$ to avoid confusions.

2 A simple example

In this section, we want to illustrate the main result of the paper through a simple example, for which the effective dynamics can be explicitly obtained and the strong convergence rate can be easily calculated. Let us consider the following linear equation:

$$\begin{aligned} \dot{X}_t^\varepsilon &= Y_t^\varepsilon + \dot{P}_t, \quad X_0 = x, \\ \dot{Y}_t^\varepsilon &= -\frac{1}{\varepsilon} Y_t^\varepsilon + \frac{1}{\sqrt{\varepsilon}} \dot{W}_t, \quad Y_0 = y. \end{aligned} \tag{2.1}$$

The above equation can be solved analytically such that

$$\begin{aligned} X_t^\varepsilon &= x + \int_0^t Y_s^\varepsilon \mathrm{d}s + P_t, \\ Y_t^\varepsilon &= \mathrm{e}^{-t/\varepsilon} y + \frac{1}{\sqrt{\varepsilon}} \int_0^t \mathrm{e}^{-(t-s)/\varepsilon} \mathrm{d}W_s. \end{aligned}$$

Since the fast process is an Ornstein-Uhlenbeck process which admits a unique invariant measure with mean zero, the effective dynamics prescribed by (1.2) can be simply written as

$$\overline{X}_t = x + P_t.$$

The strong rate for X_t^ε to converge to $\overline{X}_t$ can be obtained through

$$\mathbb{E}|X_t^\varepsilon - \overline{X}_t| = \mathbb{E}\Big|\int_0^t Y_s^\varepsilon \mathrm{d}s\Big|. \tag{2.2}$$

Notice that the process Y_t^ε is a Gaussian process. Therefore, its time integral, as a limit of sums of Gaussian random variables, is also Gaussian with mean

$$\mathbb{E}\int_0^t Y_s^\varepsilon \mathrm{d}s = y \int_0^t \mathrm{e}^{-s/\varepsilon} \mathrm{d}s = O(\varepsilon).$$

We can calculate its variance such that

$$\mathbb{E}\Big(\int_0^t Y_s^\varepsilon \mathrm{d}s\Big)^2 = 2\mathbb{E}\int_0^t Y_s^\varepsilon \mathrm{d}s \int_s^t Y_\tau^\varepsilon \mathrm{d}\tau$$

$$
\begin{aligned}
&= 2\int_0^t \mathrm{d}s \int_s^t \mathrm{d}\tau \Big(\frac{\mathrm{e}^{-(\tau-s)/\varepsilon} - \mathrm{e}^{-(\tau+s)/\varepsilon}}{2} + \mathrm{e}^{-(\tau+s)/\varepsilon} y^2\Big) \\
&= t\varepsilon + O(\varepsilon^2),
\end{aligned}
$$

which, together with (2.2), implies that

$$
\mathbb{E}|X_t^\varepsilon - \overline{X}_t| = O(\sqrt{\varepsilon}\,).
$$

The above example shows that the $O(\varepsilon^{1/2})$ convergence rate is optimal in the sense that any sharper rate can be counter-exampled by (2.1). It can also be seen from this example that it is the exponential decay of the correlation function of the fast dynamics that is guaranteeing the $O(\varepsilon^{1/2})$ standard deviation of the time average of the fast process, which leads to the $O(\varepsilon^{1/2})$ strong convergence of the effective dynamics.

3 Strong convergence rate

In this section, we want to prove the theorem for the strong convergence. We will first make some assumptions on system (1.1). Then we will elaborate on the proof.

3.1 Assumptions and convergence theorem

Define $\mathbb{C}_b^\infty$ to be the space of smooth functions with bounded derivatives of any order. We assume the following conditions for system (1.1).

Assumption 3.1 The coefficients a, b, c, f, g, and h, viewed as functions of (x, y, ε), are in $\mathbb{C}_b^\infty$. Moreover, a, b, and c are bounded.

Assumption 3.2 There exists a constant $\alpha > 0$ such that for any (x, y, ε),

$$
y^{\mathrm{T}} g(x, y, \varepsilon) g^{\mathrm{T}}(x, y, \varepsilon) y \geqslant \alpha |y|^2. \tag{3.1}
$$

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

$$
\begin{aligned}
&\langle y_1 - y_2,\ f(x, y_1, \varepsilon) - f(x, y_2, \varepsilon) + \lambda_2 (h(x, y_1, \varepsilon) - h(x, y_2, \varepsilon))\rangle \\
&\quad + \|g(x, y_1, \varepsilon) - g(x, y_2, \varepsilon)\|^2 + \lambda_2 |h(x, y_1, \varepsilon) - h(x, y_2, \varepsilon)|^2 \\
&\leqslant -\beta |y_1 - y_2|^2,
\end{aligned} \tag{3.2}
$$

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

Suppose that X_t^ε and $\overline{X}_t$ are solutions to (1.1) and (1.2), respectively. We will prove the following theorem for the strong convergence rate.

Theorem 3.4 *Suppose that Assumptions* 3.1–3.3 *hold and the following is true*:

$$
b = b(x, \varepsilon), \quad c = c(x, \varepsilon). \tag{3.3}
$$

Then we have for any $T_0 > 0$, there exists a constant $\mathscr{C} > 0$ independent of ε such that

$$\sup_{0\leqslant t\leqslant T_0} \mathbb{E}|X_t^\varepsilon - \overline{X}_t|^2 \leqslant \mathscr{C}\varepsilon. \tag{3.4}$$

Condition (3.3) in Theorem 3.4 implies that the effective dynamics (1.2) takes a simpler form such that

$$\dot{\overline{X}}_t = \overline{a}(\overline{X}_t) + b(\overline{X}_t)\dot{W}_t + c(\overline{X}_{t-})\dot{P}_t, \quad \overline{X}_0 = x,$$

in which the diffusion and jump terms are obtained simply by taking the limit of $\varepsilon \to 0$, without averaging with respect to the equilibrium of the fast dynamics:

$$b(x) = \lim_{\varepsilon\to 0} b(x,\varepsilon), \quad c(x) = \lim_{\varepsilon\to 0} c(x,\varepsilon). \tag{3.5}$$

As shown in [4,11], condition (3.3) is necessary for the strong convergence of Principle of Averaging in the form of (3.4). It can be seen from the example in Section 2 that Assumption 3.1 can be relaxed. The main purpose of Assumptions 3.2 and 3.3 is to guarantee the exponential convergence of the fast processes to the equilibrium. Based on the recent progress on the stability of Markov processes [14–16], we believe that Assumptions 3.2 and 3.3 can also be relaxed. We will leave finding necessary conditions for Theorem 3.4 to future investigations.

3.2 Proof of strong convergence theorem

Under Assumptions 3.1–3.3, it has been shown [7,13] that for each fixed (x,ε), the dynamics

$$\dot{Z}_t = \frac{1}{\varepsilon} f(x, Z_t, \varepsilon) + \frac{1}{\sqrt{\varepsilon}} g(x, Z_t, \varepsilon)\dot{W}_t + h(x, Z_{t-})N_t^\varepsilon, \quad Z_0 = z, \tag{3.6}$$

is ergodic with a unique invariant probability measure $\mu_x^\varepsilon(\cdot)$. To facilitate our proof, we define

$$\overline{a}(x,\varepsilon) = \int_{\mathbb{R}^m} a(x,y,\varepsilon)\mu_x^\varepsilon(\mathrm{d}y). \tag{3.7}$$

Notice that the following relation holds between the above function and that defined by (1.3):

$$\overline{a}(x) = \lim_{\varepsilon\to 0} \overline{a}(x,\varepsilon).$$

Let $Z_{x,z,t}^\varepsilon$ denote the solution of (3.6) with initial condition z and parameter (x,ε). We define the following function:

$$\hat{a}(x,z,t,\varepsilon) = \mathbb{E}\{a(x, Z_{x,z,t}^\varepsilon, \varepsilon)\}. \tag{3.8}$$

We also define the following auxiliary process which is a modification of the process defined in [3]. Partitioning $[0, T_0]$ into subintervals of the same length

Δ, we construct for $t \in [k\Delta, (k+1)\Delta),\ k \geqslant 0$, the process $(\widetilde{X}_t^\varepsilon, \widetilde{Y}_t^\varepsilon)$ such that

$$\begin{aligned} \dot{\widetilde{X}}_t^\varepsilon &= a(X_{k\Delta}^\varepsilon, \widetilde{Y}_t^\varepsilon, \varepsilon) + b(X_t^\varepsilon, \varepsilon)\dot{B}_t, + c(X_{t-}^\varepsilon, \varepsilon)\dot{P}_t, \\ \dot{\widetilde{Y}}_t^\varepsilon &= \frac{1}{\varepsilon} f(X_{k\Delta}^\varepsilon, \widetilde{Y}_t^\varepsilon, \varepsilon) + \frac{1}{\sqrt{\varepsilon}} g(X_{k\Delta}^\varepsilon, \widetilde{Y}_t^\varepsilon, \varepsilon)\dot{W}_t + h(X_{k\Delta}^\varepsilon, \widetilde{Y}_{t-}^\varepsilon, \varepsilon)\dot{N}_t^\varepsilon, \end{aligned} \tag{3.9}$$

with the initial condition

$$\widetilde{X}_0^\varepsilon = x, \quad \widetilde{Y}_0^\varepsilon = y, \tag{3.10}$$

and at the left end of each subinterval the initial value is re-set such that

$$\widetilde{X}_{(k+1)\Delta-}^\varepsilon = \lim_{t\to(k+1)\Delta^-} \widetilde{X}_t^\varepsilon, \quad \widetilde{Y}_{(k+1)\Delta-}^\varepsilon = \lim_{t\to(k+1)\Delta^-} \widetilde{Y}_t^\varepsilon. \tag{3.11}$$

Denote by $\lfloor x \rfloor$ the largest integer less than or equal to x, we can also write (3.9)-(3.10) in the following integral form:

$$\begin{aligned} \widetilde{X}_t^\varepsilon &= x + \int_0^t a(X_{\lfloor s/\Delta \rfloor \Delta}^\varepsilon, \widetilde{Y}_s^\varepsilon, \varepsilon)\mathrm{d}s + \int_0^t b(X_s^\varepsilon, \varepsilon)\mathrm{d}B_s + \int_0^t c(X_{s-}^\varepsilon, \varepsilon)\mathrm{d}P_s, \\ \widetilde{Y}_t^\varepsilon &= y + \frac{1}{\varepsilon}\int_0^t f(X_{\lfloor s/\Delta \rfloor \Delta}^\varepsilon, \widetilde{Y}_s^\varepsilon, \varepsilon)\mathrm{d}s + \frac{1}{\sqrt{\varepsilon}}\int_0^t g(X_{\lfloor s/\Delta \rfloor \Delta}^\varepsilon, \widetilde{Y}_s^\varepsilon, \varepsilon)\mathrm{d}W_s \\ &\quad + \int_0^t h(X_{\lfloor s/\Delta \rfloor \Delta}^\varepsilon, \widetilde{Y}_{s-}^\varepsilon, \varepsilon)\mathrm{d}N_s^\varepsilon. \end{aligned}$$

By Assumption 3.1 on the smoothness of the coefficients a, b, and c, $\mathbb{E}|X_t^\varepsilon|^2$ is bounded over finite time intervals. Here, we want to show the stability of the fast processes implied by Assumption 3.3 on the dissipative structure of the fast processes.

Lemma 3.5 *For any $T_0 > 0$, there exists a constant $\mathscr{C}$ independent of (ε, Δ) such that*

$$\mathbb{E}|Y_t^\varepsilon|^2,\ \mathbb{E}|\widetilde{Y}_t^\varepsilon|^2 \leqslant \mathscr{C}. \tag{3.12}$$

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

$$\begin{aligned} &\langle y, f(x,y,\varepsilon) - f(x,0,\varepsilon) + \lambda_2(h(x,y,\varepsilon) - h(x,0,\varepsilon))\rangle \\ &\qquad + \|g(x,y,\varepsilon) - g(x,0,\varepsilon)\|^2 + \lambda_2|h(x,y,\varepsilon) - h(x,0,\varepsilon)|^2 \\ &\leqslant -\beta|y|^2. \end{aligned} \tag{3.13}$$

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

$$\begin{aligned} &\|g(x,y,\varepsilon)\|^2 + \lambda_2|h(x,y,\varepsilon)|^2 \\ &\quad \leqslant (1+\gamma)(\|g(x,y,\varepsilon) - g(x,0,\varepsilon)\|^2 + \lambda_2|h(x,y,\varepsilon) - h(x,0,\varepsilon)|^2) \\ &\qquad + \Big(1 + \frac{1}{\gamma}\Big)(\|g(x,0,\varepsilon)\|^2 + \lambda_2|h(x,0,\varepsilon)|^2) \\ &\quad \leqslant \|g(x,y,\varepsilon) - g(x,0,\varepsilon)\|^2 + \lambda_2|h(x,y,\varepsilon) - h(x,0,\varepsilon)|^2 \end{aligned}$$

$$+\mathscr{C}_1\gamma|y|^2+\mathscr{C}_2\Big(1+\frac{1}{\gamma}\Big)(|x|^2+\varepsilon^2+1), \tag{3.14}$$

where $\mathscr{C}_1$ and $\mathscr{C}_2$ depend on the Lipschitz and linear growth constants of function g and h. If we choose an appropriate value for γ such that

$$\mathscr{C}_1\gamma\leqslant\frac{\beta}{4},$$

by (3.13), (3.14), and Assumption 3.1, we can get

$$\begin{aligned}
&\langle y,f(x,y,\varepsilon)+\lambda_2h(x,y,\varepsilon)\rangle+\|g(x,y,\varepsilon)\|^2+\lambda_2|h(x,y,\varepsilon)|^2\\
&\quad\leqslant\langle y,f(x,y,\varepsilon)-f(x,0,\varepsilon)+\lambda_2(h(x,y,\varepsilon)-h(x,0,\varepsilon))\rangle\\
&\qquad+\langle y,f(x,0,\varepsilon)+\lambda_2h(x,0,\varepsilon)\rangle+\|g(x,y,\varepsilon)-g(x,0,\varepsilon)\|^2\\
&\qquad+\lambda_2|h(x,y,\varepsilon)-h(x,0,\varepsilon)|^2+\frac{\beta}{4}|y|^2+\mathscr{C}'(|x^2|+\varepsilon^2+1)\\
&\quad\leqslant-\frac{\beta}{2}|y|^2+\mathscr{C}''(|x|^2+\varepsilon^2+1).
\end{aligned}$$

Itô formula then suggests that

$$\begin{aligned}
\mathrm{d}\mathbb{E}|Y_t^\varepsilon|^2&=\frac{2}{\varepsilon}\mathbb{E}\langle Y_t^\varepsilon,f(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)+\lambda_2h(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)\rangle\mathrm{d}t\\
&\quad+\frac{1}{\varepsilon}\mathbb{E}\|g(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)\|^2\mathrm{d}t+\frac{\lambda_2}{\varepsilon}\mathbb{E}|h(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)|^2\mathrm{d}t\\
&\leqslant-\frac{\beta}{\varepsilon}\mathbb{E}|Y_t^\varepsilon|^2+\frac{\mathscr{C}}{\varepsilon}\mathbb{E}(|X_t^\varepsilon|^2+\varepsilon^2+1),
\end{aligned}$$

which, by the Gronwall inequality, implies the boundedness of $\mathbb{E}|Y_t^\varepsilon|^2$. Repeating the same argument, we can also obtain the boundedness of $\mathbb{E}|\widetilde{Y}_t^\varepsilon|^2$. □

The following lemma describes how $\widetilde{Y}_t^\varepsilon$ deviates from Y_t^ε.

Lemma 3.6 *For any $T_0>0$, there exists a constant $\mathscr{C}>0$ independent of (ε,Δ) such that*

$$\sup_{0\leqslant t\leqslant T_0}\mathbb{E}|Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon|^2\leqslant\mathscr{C}\Delta. \tag{3.15}$$

Proof For each $k\geqslant 0$ and $t\in[k\Delta,(k+1)\Delta)$, direct computation with Itô formula gives that

$$\begin{aligned}
\mathrm{d}\mathbb{E}|Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon|^2&=\frac{2}{\varepsilon}\mathbb{E}\langle Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon,f(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)-f(X_{k\Delta}^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)\\
&\quad+\lambda_2(h(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)-h(X_{k\Delta}^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon))\rangle\mathrm{d}t\\
&\quad+\frac{1}{\varepsilon}\mathbb{E}\|g(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)-g(X_{k\Delta}^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)\|^2\mathrm{d}t\\
&\quad+\frac{\lambda_2}{\varepsilon}\mathbb{E}|h(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)-h(X_{k\Delta}^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)|^2\mathrm{d}t.
\end{aligned} \tag{3.16}$$

By Assumptions 3.1 and 3.3, we have

$$
\begin{aligned}
&\langle Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon, f(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)-f(X_{k\Delta}^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)+\lambda_2(h(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)-h(X_{k\Delta}^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon))\rangle\\
&\quad+\frac{1}{2}\,\|g(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)-g(X_{k\Delta}^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)\|^2+\frac{\lambda_2}{2}\,|h(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)-h(X_{k\Delta}^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)|^2\\
&\leqslant\langle Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon, f(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)-f(X_t^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)+f(X_t^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)-f(X_{k\Delta}^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)\rangle\\
&\quad+\lambda_2\langle Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon, h(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)-h(X_t^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)+h(X_t^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)-h(X_{k\Delta}^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)\rangle\\
&\quad+\|g(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)-g(X_t^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)\|^2+\|g(X_t^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)-g(X_{k\Delta}^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)\|^2\\
&\quad+\lambda_2|h(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)-h(X_t^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)|^2+\lambda_2|h(X_t^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)-h(X_{k\Delta}^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)|^2\\
&\leqslant-\beta|Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon|^2+\mathscr{C}(|Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon|\,|X_t^\varepsilon-X_{k\Delta}^\varepsilon|+|X_t^\varepsilon-X_{k\Delta}^\varepsilon|^2).
\end{aligned}
$$

Since $\beta>0$, we have

$$
\mathscr{C}|Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon|\,|X_t^\varepsilon-X_{k\Delta}^\varepsilon|\leqslant\frac{\beta}{2}\,|Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon|^2+\frac{\mathscr{C}^2}{2\beta}\,|X_t^\varepsilon-X_{k\Delta}^\varepsilon|^2,
$$

which easily leads to

$$
\begin{aligned}
&\langle Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon, f(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)-f(X_{k\Delta}^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)+\lambda_2(h(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)-h(X_{k\Delta}^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon))\rangle\\
&\quad+\frac{1}{2}\,\|g(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)-g(X_{k\Delta}^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)\|^2+\frac{\lambda_2}{2}\,|h(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon)-h(X_{k\Delta}^\varepsilon,\widetilde{Y}_t^\varepsilon,\varepsilon)|^2\\
&\leqslant-\frac{\beta}{2}|Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon|^2+\mathscr{C}|X_t^\varepsilon-X_{k\Delta}^\varepsilon|^2. \qquad (3.17)
\end{aligned}
$$

By the boundedness of a, b, and c, and the quadratic variations of the Brownian motion and the Poisson process, for $t\in[k\Delta,(k+1)\Delta)$, we have

$$
\mathbb{E}|X_t^\varepsilon-X_{k\Delta}^\varepsilon|^2\leqslant\mathscr{C}\Delta. \tag{3.18}
$$

Combining (3.16)–(3.18), it follows that

$$
\mathrm{d}\mathbb{E}|Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon|^2\leqslant-\frac{\beta}{\varepsilon}\,\mathbb{E}|Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon|^2\mathrm{d}t+\mathscr{C}\frac{\Delta}{\varepsilon}\,\mathrm{d}t.
$$

The Gronwall inequality implies that

$$
\mathbb{E}|Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon|^2\leqslant \mathrm{e}^{-\beta(t-k\Delta)/\varepsilon}\mathbb{E}|Y_{k\Delta}^\varepsilon-\widetilde{Y}_{k\Delta}^\varepsilon|^2+\mathscr{C}(1-\mathrm{e}^{-\beta(t-k\Delta)/\varepsilon})\Delta. \tag{3.19}
$$

By the continuity condition (3.11), we can take $t=(k+1)\Delta$ in the above inequality, which gives

$$
\mathbb{E}|Y_{(k+1)\Delta}^\varepsilon-\widetilde{Y}_{(k+1)\Delta}^\varepsilon|^2\leqslant \mathrm{e}^{-\beta\Delta/\varepsilon}\mathbb{E}|Y_{k\Delta}^\varepsilon-\widetilde{Y}_{k\Delta}^\varepsilon|^2+\mathscr{C}(1-\mathrm{e}^{-\beta\Delta/\varepsilon})\Delta.
$$

Applying the above inequality recursively for reducing value of k until $k=0$, we can obtain by the initial condition (3.10) that

$$
\mathbb{E}|Y_{(k+1)\Delta}^\varepsilon-\widetilde{Y}_{(k+1)\Delta}^\varepsilon|^2\leqslant\mathscr{C}(1-\mathrm{e}^{-\beta\Delta/\varepsilon})\Delta\sum_{0\leqslant\ell\leqslant k}\mathrm{e}^{-\ell\beta\Delta/\varepsilon}\leqslant\mathscr{C}\Delta,
$$

which, together with (3.19), gives (3.15). □

Lemma 3.6 easily gives the following asymptotic behavior of $\widetilde{X}_t^\varepsilon$.

Proposition 3.7 *For any $T_0 > 0$, there exists a constant $\mathscr{C} > 0$ independent of (ε, Δ) such that*

$$\sup_{0\leqslant t\leqslant T_0} \mathbb{E}|X_t^\varepsilon - \widetilde{X}_t^\varepsilon|^2 \leqslant \mathscr{C}\Delta. \tag{3.20}$$

Proof By the smoothness of the coefficients and Lemma 3.6, we can write

$$\begin{aligned}\mathbb{E}|X_t^\varepsilon - \widetilde{X}_t^\varepsilon|^2 &= \mathbb{E}\Big|\int_0^t (a(X_s^\varepsilon, Y_s^\varepsilon, \varepsilon) - a(X_{\lfloor s/\Delta\rfloor\Delta}^\varepsilon, \widetilde{Y}_s^\varepsilon, \varepsilon))\mathrm{d}s\Big|^2\\ &\leqslant \mathscr{C}\mathbb{E}\int_0^t (|X_s^\varepsilon - X_{\lfloor s/\Delta\rfloor\Delta}^\varepsilon|^2 + |Y_s^\varepsilon - \widetilde{Y}_s^\varepsilon|^2)\mathrm{d}s\\ &\leqslant \mathscr{C}'\Delta.\end{aligned}$$

□

Now, we want to give an estimate for the expectation $\mathbb{E}|\widetilde{X}_t^\varepsilon - \overline{X}_t|^2$.

Proposition 3.8 *For any $T_0 > 0$, there exists a constant $\mathscr{C} > 0$ independent of (ε, Δ) such that*

$$\sup_{0\leqslant t\leqslant T_0} \mathbb{E}|\widetilde{X}_t^\varepsilon - \overline{X}_t|^2 \leqslant \mathscr{C}(\Delta + \varepsilon). \tag{3.21}$$

Proof First of all, we notice that

$$\begin{aligned}&\mathbb{E}|\widetilde{X}_t^\varepsilon - \overline{X}_t|^2\\ &\quad\leqslant 3\,\mathbb{E}\bigg(\int_0^t (a(X_{\lfloor s/\Delta\rfloor\Delta}^\varepsilon, \widetilde{Y}_s^\varepsilon, \varepsilon) - \overline{a}(\overline{X}_s) + \lambda_1(c(X_s^\varepsilon, \varepsilon) - \overline{c}(\overline{X}_s)))\mathrm{d}s\bigg)^2\\ &\qquad + 3\,\mathbb{E}\bigg(\int_0^t (b(X_s^\varepsilon, \varepsilon) - b(\overline{X}_s))\mathrm{d}W_s\bigg)^2\\ &\qquad + 3\,\mathbb{E}\bigg(\int_0^t (c(X_s^\varepsilon, \varepsilon) - c(\overline{X}_s))\mathrm{d}\hat{P}_s\bigg)^2,\end{aligned} \tag{3.22}$$

where $\hat{P}_t = P_t - \lambda_1 t$. For the second and third terms on the right-hand side of the above inequality, by the Itô isometry, we have

$$\mathbb{E}\bigg(\int_0^t (b(X_s^\varepsilon, \varepsilon) - b(\overline{X}_s^\varepsilon))\mathrm{d}W_s\bigg)^2 = \int_0^t \mathbb{E}(b(X_s^\varepsilon, \varepsilon) - b(\overline{X}_s))^2\mathrm{d}s,$$

and

$$\mathbb{E}\bigg(\int_0^t (c(X_s^\varepsilon, \varepsilon) - c(\overline{X}_s^\varepsilon))\mathrm{d}\hat{P}_s\bigg)^2 = \lambda_1\int_0^t \mathbb{E}(c(X_s^\varepsilon, \varepsilon) - c(\overline{X}_s))^2\mathrm{d}s,$$

where functions b and c are defined as (3.5). Proposition 3.7 and the smoothness of the coefficients give the following estimate:

$$\mathbb{E}(b(X_s^\varepsilon, \varepsilon) - b(\overline{X}_s))^2,\ \mathbb{E}(c(X_s^\varepsilon, \varepsilon) - c(\overline{X}_s))^2 \leqslant \mathscr{C}(\mathbb{E}|\widetilde{X}_s^\varepsilon - \overline{X}_s|^2 + \Delta + \varepsilon^2).$$

Therefore, we have

$$\mathbb{E}\Big(\int_0^t (b(X_s^\varepsilon,\varepsilon)-b(\overline{X}_s^\varepsilon))\mathrm{d}W_s\Big)^2,\ \mathbb{E}\Big(\int_0^t (c(X_s^\varepsilon,\varepsilon)-c(\overline{X}_s^\varepsilon))\mathrm{d}\hat{P}_s\Big)^2$$
$$\leqslant \mathscr{C}\Big(\int_0^t \mathbb{E}|\widetilde{X}_s^\varepsilon-\overline{X}_s|^2\mathrm{d}s+\Delta+\varepsilon^2\Big). \tag{3.23}$$

Now, we want to give an estimate for the first term on the right-hand side of (3.22). Notice that

$$\begin{aligned}
&\mathbb{E}\Big(\int_0^t (a(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta},\widetilde{Y}_s^\varepsilon,\varepsilon)-\overline{a}(\overline{X}_s))\mathrm{d}s\Big)^2\\
&\quad\leqslant 3\,\mathbb{E}\Big(\int_0^t (a(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta},\widetilde{Y}_s^\varepsilon,\varepsilon)-\overline{a}(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta}))\mathrm{d}s\Big)^2\\
&\qquad+3\,\mathbb{E}\Big(\int_0^t (\overline{a}(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta})-\overline{a}(X_s^\varepsilon))\mathrm{d}s\Big)^2\\
&\qquad+3\,\mathbb{E}\Big(\int_0^t (\overline{a}(X_s^\varepsilon)-\overline{a}(\overline{X}_s))\mathrm{d}s\Big)^2.
\end{aligned}$$

By the smoothness of $\overline{a}(x,\varepsilon)$ proved in Appendix, we can have

$$\begin{aligned}
&\mathbb{E}\Big(\int_0^t (a(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta},\widetilde{Y}_s^\varepsilon,\varepsilon)-\overline{a}(\overline{X}_s))\mathrm{d}s\Big)^2\\
&\quad\leqslant 3\,\mathbb{E}\Big(\int_0^t (a(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta},\widetilde{Y}_s^\varepsilon,\varepsilon)-\overline{a}(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta}))\mathrm{d}s\Big)^2\\
&\qquad+\mathscr{C}\Big(\Delta+\mathbb{E}\int_0^t |X_s^\varepsilon-\overline{X}_s|^2\mathrm{d}s\Big).
\end{aligned}\tag{3.24}$$

We can evaluate the integral in (3.24) as

$$\begin{aligned}
&\mathbb{E}\Big(\int_0^t (a(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta},\widetilde{Y}_s^\varepsilon,\varepsilon)-\overline{a}(\overline{X}_{\lfloor s/\Delta\rfloor\Delta}))\mathrm{d}s\Big)^2\\
&\leqslant \mathbb{E}\sum_{0\leqslant k\leqslant\lfloor t/\Delta\rfloor}\Big(\int_{k\Delta}^{((k+1)\Delta)\wedge t}(a(X^\varepsilon_{k\Delta},\widetilde{Y}_s^\varepsilon,\varepsilon)-\overline{a}(X^\varepsilon_{k\Delta}))\mathrm{d}s\Big)^2\\
&\quad+2\,\mathbb{E}\sum_{0\leqslant i<j\leqslant\lfloor t/\Delta\rfloor}\int_{i\Delta}^{((i+1)\Delta)\wedge t}\int_{j\Delta}^{((j+1)\Delta)\wedge t}\mathrm{d}s\mathrm{d}\tau\\
&\quad\cdot\langle a(X^\varepsilon_{i\Delta},\widetilde{Y}_s^\varepsilon,\varepsilon)-\overline{a}(X^\varepsilon_{i\Delta}),a(X^\varepsilon_{j\Delta},\widetilde{Y}_\tau^\varepsilon,\varepsilon)-\overline{a}(X^\varepsilon_{j\Delta})\rangle\\
&=: A_1+2A_2.
\end{aligned}$$

For A_1, we can have the following estimate:

$$\begin{aligned}
&\mathbb{E}\bigg(\int_{k\Delta}^{((k+1)\Delta)\wedge t}(a(X_{k\Delta}^\varepsilon,\widetilde{Y}_s^\varepsilon,\varepsilon)-\overline{a}(X_{k\Delta}^\varepsilon))\mathrm{d}s\bigg)^2\\
&\quad=2\,\mathbb{E}\int_{k\Delta}^{((k+1)\Delta)\wedge t}\int_s^{(k+1)\Delta\wedge t}\mathrm{d}s\mathrm{d}\tau\\
&\qquad\cdot\langle a(X_{k\Delta}^\varepsilon,\widetilde{Y}_s^\varepsilon,\varepsilon)-\overline{a}(X_{k\Delta}^\varepsilon),a(X_{k\Delta}^\varepsilon,\widetilde{Y}_\tau^\varepsilon,\varepsilon)-\overline{a}(X_{k\Delta}^\varepsilon)\rangle,\\
&\quad\leqslant 2\int_{k\Delta}^{((k+1)\Delta)\wedge t}\int_s^{(k+1)\Delta\wedge t}\mathrm{d}s\mathrm{d}\tau\\
&\qquad\cdot|\mathbb{E}\langle a(X_{k\Delta}^\varepsilon,\widetilde{Y}_s^\varepsilon,\varepsilon)-\overline{a}(X_{k\Delta}^\varepsilon),\mathbb{E}_s(a(X_{k\Delta}^\varepsilon,\widetilde{Y}_\tau^\varepsilon,\varepsilon)-\overline{a}(X_{k\Delta}^\varepsilon))\rangle|,
\end{aligned}$$

where $\mathbb{E}_s$ denotes the conditional probability for information up to time s adapted to the filtration generated by noises in (1.1). By the smoothness of $\overline{a}(x,\varepsilon)$, we have

$$\begin{aligned}
&|\mathbb{E}_s(a(X_{k\Delta}^\varepsilon,\widetilde{Y}_s^\varepsilon,\varepsilon)-\overline{a}(X_{k\Delta}^\varepsilon))|\\
&\quad\leqslant|\mathbb{E}_s(a(X_{k\Delta}^\varepsilon,\widetilde{Y}_s^\varepsilon,\varepsilon)-\overline{a}(X_{k\Delta}^\varepsilon,\varepsilon))|+|\mathbb{E}_s(\overline{a}(X_{k\Delta}^\varepsilon,\varepsilon)-\overline{a}(X_{k\Delta}^\varepsilon))|\\
&\quad\leqslant|\mathbb{E}_s(a(X_{k\Delta}^\varepsilon,\widetilde{Y}_s^\varepsilon,\varepsilon)-\overline{a}(X_{k\Delta}^\varepsilon,\varepsilon))|+\mathscr{C}\varepsilon.
\end{aligned}$$

By the exponential mixing (A5) given in Appendix and Lemma 3.5 for the boundedness of $\mathbb{E}|Y_t^\varepsilon|$, we can write

$$\begin{aligned}
\mathbb{E}|\mathbb{E}_s(a(X_{k\Delta}^\varepsilon,\widetilde{Y}_\tau^\varepsilon,\varepsilon)-\overline{a}(X_{k\Delta}^\varepsilon,\varepsilon))|&\leqslant\mathscr{C}\mathbb{E}(|X_s^\varepsilon|+|Y_s^\varepsilon|+\varepsilon+1)\mathrm{e}^{-\beta(\tau-s)/\varepsilon}\\
&\leqslant\mathscr{C}'\mathrm{e}^{-\beta(\tau-s)/\varepsilon}.
\end{aligned}$$

Therefore, by the assumption that a is bounded, we have

$$A_1\leqslant\mathscr{C}\sum_{0\leqslant k\leqslant\lfloor t/\Delta\rfloor}\int_{k\Delta}^{(k+1)\Delta\wedge t}\mathrm{d}s\int_s^{(k+1)\Delta\wedge t}\mathrm{d}\tau(\mathrm{e}^{-\beta(\tau-s)/\varepsilon}+\varepsilon)\leqslant\mathscr{C}'\varepsilon.\qquad(3.25)$$

To estimate A_2, we define auxiliary process $Z_{i,\tau}^\varepsilon$ for $i\Delta\leqslant\tau$ such that it satisfies (3.6) with parameter $x=X_{i\Delta}^\varepsilon$ and initial condition $\widetilde{Y}_{i\Delta-}^\varepsilon$, i.e.,

$$\begin{aligned}
\dot{Z}_{i,\tau}^\varepsilon=\frac{1}{\varepsilon}f(X_{i\Delta}^\varepsilon,Z_{i,\tau}^\varepsilon,\varepsilon)+\frac{1}{\sqrt{\varepsilon}}g(X_{i\Delta}^\varepsilon,Z_{i,\tau}^\varepsilon,\varepsilon)\dot{W}_\tau+h(X_{i\Delta}^\varepsilon,Z_{i,\tau-}^\varepsilon,\varepsilon)\dot{N}_\tau^\varepsilon,\\
Z_{i,i\Delta-}=\widetilde{Y}_{i\Delta-}^\varepsilon.
\end{aligned}$$

Notice that by the above definition, we have

$$Z_{k,\tau}^\varepsilon=\widetilde{Y}_\tau^\varepsilon,\quad\tau\in[k\Delta,(k+1)\Delta).$$

And continuity implies that

$$Z_{(k+1),(k+1)\Delta}^\varepsilon=Z_{k,(k+1)\Delta}^\varepsilon=\widetilde{Y}_{(k+1)\Delta}^\varepsilon.$$

Using the boundedness of $\overline{a}(x,\varepsilon)$, we get

$$\begin{aligned}
&|\mathbb{E}\langle a(X^\varepsilon_{i\Delta},\widetilde{Y}^\varepsilon_s,\varepsilon)-\overline{a}(X^\varepsilon_{i\Delta}),\ a(X^\varepsilon_{j\Delta},\widetilde{Y}^\varepsilon_\tau,\varepsilon)-\overline{a}(X^\varepsilon_{j\Delta})\rangle| \\
&\quad = |\mathbb{E}\langle a(X^\varepsilon_{i\Delta},\widetilde{Y}^\varepsilon_s,\varepsilon)-\overline{a}(X^\varepsilon_{i\Delta}),\mathbb{E}_{i\Delta}(a(X^\varepsilon_{j\Delta},\widetilde{Y}^\varepsilon_\tau,\varepsilon)-\overline{a}(X^\varepsilon_{j\Delta}))\rangle| \\
&\quad \leqslant \mathscr{C}\mathbb{E}|\mathbb{E}_{i\Delta}(a(X^\varepsilon_{j\Delta},\widetilde{Y}^\varepsilon_\tau,\varepsilon)-\overline{a}(X^\varepsilon_{j\Delta}))| \\
&\quad \leqslant \mathscr{C}\mathbb{E}|\mathbb{E}_{i\Delta}\{(a(X^\varepsilon_{j\Delta},\widetilde{Y}^\varepsilon_\tau,\varepsilon)-\overline{a}(X^\varepsilon_{j\Delta}))-(a(X^\varepsilon_{i\Delta},Z^\varepsilon_{i,\tau},\varepsilon)-\overline{a}(X^\varepsilon_{i\Delta}))\}| \\
&\qquad +\mathscr{C}\mathbb{E}|\mathbb{E}_{i\Delta}(a(X^\varepsilon_{i\Delta},Z^\varepsilon_{i,\tau},\varepsilon)-\overline{a}(X^\varepsilon_{i\Delta}))|. \qquad (3.26)
\end{aligned}$$

By the exponential mixing (A5) and the smoothness of $\overline{a}(x,\varepsilon)$, we have

$$\mathbb{E}|\mathbb{E}_{i\Delta}(a(X^\varepsilon_{i\Delta},Z^\varepsilon_{i,\tau},\varepsilon)-\overline{a}(X^\varepsilon_{i\Delta}))|\leqslant\mathscr{C}(\mathrm{e}^{-\beta(\tau-i\Delta)/\varepsilon}+\varepsilon). \qquad (3.27)$$

The smoothness of $\overline{a}(x,\varepsilon)$ suggests that

$$\begin{aligned}
&\mathbb{E}|\mathbb{E}_{i\Delta}\{(a(X^\varepsilon_{j\Delta},\widetilde{Y}^\varepsilon_\tau,\varepsilon)-\overline{a}(X^\varepsilon_{j\Delta}))-(a(X^\varepsilon_{i\Delta},Z^\varepsilon_{i,\tau},\varepsilon)-\overline{a}(X^\varepsilon_{i\Delta}))\}| \\
&\quad \leqslant \mathbb{E}|\mathbb{E}_{i\Delta}\{(a(X^\varepsilon_{j\Delta},\widetilde{Y}^\varepsilon_\tau,\varepsilon)-\overline{a}(X^\varepsilon_{j\Delta},\varepsilon)) \\
&\qquad -(a(X^\varepsilon_{i\Delta},Z^\varepsilon_{i,\tau},\varepsilon)-\overline{a}(X^\varepsilon_{i\Delta},\varepsilon))\}|+\mathscr{C}\varepsilon \\
&\quad \leqslant \sum_{k=i}^{j-1}\mathbb{E}|\mathbb{E}_{k\Delta}\{(a(X^\varepsilon_{(k+1)\Delta},Z^\varepsilon_{k+1,\tau},\varepsilon)-\overline{a}(X^\varepsilon_{(k+1)\Delta},\varepsilon)) \\
&\qquad -(a(X^\varepsilon_{k\Delta},Z^\varepsilon_{k,\tau},\varepsilon)-\overline{a}(X^\varepsilon_{k\Delta},\varepsilon))\}|+\mathscr{C}\varepsilon. \qquad (3.28)
\end{aligned}$$

From the definition of $\hat{a}(x,z,t,\varepsilon)$ by (3.8), we can see that for $i\leqslant k$ and $(k+1)\Delta\leqslant t$,

$$\begin{aligned}
\mathbb{E}_{i\Delta}a(X^\varepsilon_{k\Delta},Z^\varepsilon_{k,t},\varepsilon)&=\mathbb{E}_{i\Delta}\hat{a}(X^\varepsilon_{k\Delta},\widetilde{Y}^\varepsilon_{k\Delta},t-k\Delta,\varepsilon) \\
&=\mathbb{E}_{i\Delta}\hat{a}(X^\varepsilon_{k\Delta},\widetilde{Y}^\varepsilon_{(k+1)\Delta},t-(k+1)\Delta,\varepsilon).
\end{aligned}$$

Let

$$\widetilde{a}(x,z,t,\varepsilon)=\hat{a}(x,z,t,\varepsilon)-\overline{a}(x,\varepsilon).$$

By the smoothness of $\hat{a}(x,z,t,\varepsilon)$ given in Appendix, we can perform the following Taylor expansion:

$$\begin{aligned}
&\mathbb{E}|\mathbb{E}_{k\Delta}\{(a(X^\varepsilon_{(k+1)\Delta},Z^\varepsilon_{k+1,\tau},\varepsilon)-\overline{a}(X^\varepsilon_{(k+1)\Delta},\varepsilon)) \\
&\qquad -(a(X^\varepsilon_{k\Delta},Z^\varepsilon_{k,\tau},\varepsilon)-\overline{a}(X^\varepsilon_{k\Delta},\varepsilon))\}| \\
&\quad =\mathbb{E}|\mathbb{E}_{k\Delta}\{\widetilde{a}(X^\varepsilon_{(k+1)\Delta},\widetilde{Y}^\varepsilon_{(k+1)\Delta},\tau-(k+1)\Delta,\varepsilon) \\
&\qquad -\widetilde{a}(X^\varepsilon_{k\Delta},\widetilde{Y}^\varepsilon_{(k+1)\Delta},\tau-(k+1)\Delta,\varepsilon)\}| \\
&\quad \leqslant \mathbb{E}\bigg|\mathbb{E}_{k\Delta}\sum_{|I|=1}^{3}\nabla^I_x\widetilde{a}(X^\varepsilon_{k\Delta},\widetilde{Y}^\varepsilon_{(k+1)\Delta},\tau-(k+1)\Delta,\varepsilon)(X^\varepsilon_{(k+1)\Delta}-X^\varepsilon_{k\Delta})^I\bigg| \\
&\qquad +\mathscr{C}\Delta^2. \qquad (3.29)
\end{aligned}$$

For any multi-index I, We can further have

$$\begin{aligned}
&\mathbb{E}|\mathbb{E}_{k\Delta}\nabla_x^I\widetilde{a}(X_{k\Delta}^\varepsilon,\widetilde{Y}_{(k+1)\Delta}^\varepsilon,\tau-(k+1)\Delta,\varepsilon)(X_{(k+1)\Delta}^\varepsilon-X_{k\Delta}^\varepsilon)^I|\\
&\quad\leqslant \mathbb{E}|\mathbb{E}_{k\Delta}\nabla_x^I\widetilde{a}(X_{k\Delta}^\varepsilon,\widetilde{Y}_{k\Delta}^\varepsilon,\tau-(k+1)\Delta,\varepsilon)(X_{(k+1)\Delta}^\varepsilon-X_{k\Delta}^\varepsilon)^I|\\
&\qquad+\mathbb{E}|\mathbb{E}_{k\Delta}\nabla_x^I\{\hat{a}(X_{k\Delta}^\varepsilon,\widetilde{Y}_{(k+1)\Delta}^\varepsilon,\tau-(k+1)\Delta,\varepsilon)\\
&\qquad-\hat{a}(X_{k\Delta}^\varepsilon,\widetilde{Y}_{k\Delta}^\varepsilon,\tau-(k+1)\Delta,\varepsilon)\}(X_{(k+1)\Delta}^\varepsilon-X_{k\Delta}^\varepsilon)^I|. \qquad (3.30)
\end{aligned}$$

Using independent increments and exponential mixing for the derivatives of $\overline{a}(x,\varepsilon)$ given by (A6) in Appendix, and the fact that for any jump-diffusion process x_t,

$$\mathbb{E}(\mathrm{d}x_t)^I=O(\mathrm{d}t),$$

we have

$$\begin{aligned}
&\mathbb{E}|\mathbb{E}_{k\Delta}\nabla_x^I\widetilde{a}(X_{k\Delta}^\varepsilon,\widetilde{Y}_{k\Delta}^\varepsilon,\tau-(k+1)\Delta,\varepsilon)(X_{(k+1)\Delta}^\varepsilon-X_{k\Delta}^\varepsilon)^I|\\
&\quad\leqslant \Delta\mathbb{E}|\mathbb{E}_{k\Delta}\nabla_x^I\{\hat{a}(X_{k\Delta}^\varepsilon,\widetilde{Y}_{k\Delta}^\varepsilon,\tau-(k+1)\Delta,\varepsilon)-\overline{a}(X_{k\Delta}^\varepsilon,\varepsilon)\}|\\
&\quad\leqslant \mathscr{C}\Delta\mathrm{e}^{-\beta(\tau-(k+1)\Delta)/\varepsilon}, \qquad (3.31)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}|\mathbb{E}_{k\Delta}\nabla_x^I\{\hat{a}(X_{k\Delta}^\varepsilon,\widetilde{Y}_{(k+1)\Delta}^\varepsilon,\tau-(k+1)\Delta,\varepsilon)\\
&\qquad-\hat{a}(X_{k\Delta}^\varepsilon,\widetilde{Y}_{k\Delta}^\varepsilon,\tau-(k+1)\Delta,\varepsilon)\}(X_{(k+1)\Delta}^\varepsilon-X_{k\Delta}^\varepsilon)^I|\\
&\quad\leqslant \Delta\mathbb{E}|\mathbb{E}_{k\Delta}\nabla_x^I\sum_{|J|=1}\nabla_z^J\hat{a}(X_{k\Delta}^\varepsilon,\widetilde{Y}_{k\Delta}^\varepsilon,\tau-(k+1)\Delta,\varepsilon)|+\mathscr{C}\Delta^2\\
&\quad\leqslant \mathscr{C}'(\Delta\mathrm{e}^{-\beta(\tau-(k+1)\Delta)/\varepsilon}+\Delta^2). \qquad (3.32)
\end{aligned}$$

Combining (3.29)–(3.32), we have

$$\begin{aligned}
&\mathbb{E}|\mathbb{E}_{k\Delta}\{(a(X_{(k+1)\Delta}^\varepsilon,Z_{k+1,\tau}^\varepsilon,\varepsilon)-\overline{a}(X_{(k+1)\Delta}^\varepsilon,\varepsilon))\\
&\qquad-(a(X_{k\Delta}^\varepsilon,Z_{k,\tau}^\varepsilon,\varepsilon)-\overline{a}(X_{k\Delta}^\varepsilon,\varepsilon))\}|\\
&\quad\leqslant \mathscr{C}(\Delta^2+\Delta\mathrm{e}^{-\beta(\tau-(k+1)\Delta)/\varepsilon}).
\end{aligned}$$

Substituting the above inequality into (3.38) gives

$$\begin{aligned}
&\mathbb{E}|\mathbb{E}_{i\Delta}\{(a(X_{j\Delta}^\varepsilon,\widetilde{Y}_\tau^\varepsilon,\varepsilon)-\overline{a}(X_{j\Delta}^\varepsilon))-(a(X_{i\Delta}^\varepsilon,Z_{i,\tau}^\varepsilon,\varepsilon)-\overline{a}(X_{i\Delta}^\varepsilon))\}|\\
&\quad\leqslant \mathscr{C}\bigg(\sum_{k=i}^{j-1}(\Delta^2+\Delta\mathrm{e}^{-\beta(\tau-(k+1)\Delta)/\varepsilon})+\varepsilon\bigg)\\
&\quad\leqslant \mathscr{C}\Big((j-i)\Delta^2+\Delta\frac{\mathrm{e}^{-\beta(\tau-j\Delta)/\varepsilon}}{1-\mathrm{e}^{-\beta\Delta/\varepsilon}}+\varepsilon\Big). \qquad (3.33)
\end{aligned}$$

Using (3.26), (3.27), and (3.33), we have

$$\begin{aligned}
&|\mathbb{E}\langle a(X_{i\Delta}^\varepsilon,\widetilde{Y}_s^\varepsilon,\varepsilon)-\overline{a}(X_{i\Delta}^\varepsilon),a(X_{j\Delta}^\varepsilon,\widetilde{Y}_\tau^\varepsilon,\varepsilon)-\overline{a}(X_{j\Delta}^\varepsilon)\rangle|\\
&\quad\leqslant \mathscr{C}\Big((j-i)\Delta^2+\Delta\frac{\mathrm{e}^{-\beta(\tau-j\Delta)/\varepsilon}}{1-\mathrm{e}^{-\beta\Delta/\varepsilon}}+\mathrm{e}^{-\beta(\tau-i\Delta)/\varepsilon}+\varepsilon\Big).
\end{aligned}$$

Therefore, by the boundedness of a, we have

$$
\begin{aligned}
A_2 &\leqslant \mathscr{C} \sum_{0\leqslant i<j\leqslant \lfloor t/\Delta\rfloor} \int_{i\Delta}^{(i+1)\Delta} \mathrm{d}s \int_{j\Delta}^{(j+1)\Delta} \mathrm{d}\tau \\
&\quad \cdot \Big((j-i)\Delta^2 + \Delta \frac{\mathrm{e}^{-\beta(\tau - j\Delta)/\varepsilon}}{1-\mathrm{e}^{-\beta\Delta/\varepsilon}} + \mathrm{e}^{-\beta(\tau - i\Delta)/\varepsilon} + \varepsilon \Big) \\
&\leqslant \mathscr{C} \sum_{0\leqslant i<j\leqslant \lfloor t/\Delta\rfloor} ((j-i)\Delta^4 + \varepsilon\Delta^2 + \mathrm{e}^{-\beta(j-i)\Delta/\varepsilon}(1-\mathrm{e}^{-\beta\Delta/\varepsilon})\varepsilon\Delta) \\
&\leqslant \mathscr{C}(\Delta + \varepsilon).
\end{aligned} \tag{3.34}
$$

Combining (3.23)–(3.25) and (3.34), we have

$$
\mathbb{E}|\widetilde{X}_t^\varepsilon - \overline{X}_t|^2 \leqslant \mathscr{C}\Big(\int_0^t \mathbb{E}|\widetilde{X}_s^\varepsilon - \overline{X}_s|^2 \mathrm{d}s + \Delta + \varepsilon \Big).
$$

The Gronwall inequality then implies (3.21). □

Now, we can finish the proof for the $O(\varepsilon^{1/2})$ strong convergence rate.

Proof of Theorem 3.4 In Propositions 3.7 and 3.8, taking $\Delta = \varepsilon$, we have

$$
\mathbb{E}|X_t^\varepsilon - \overline{X}_t|^2 \leqslant 2\,\mathbb{E}|X_t^\varepsilon - \widetilde{X}_t^\varepsilon|^2 + 2\,\mathbb{E}|\widetilde{X}_t^\varepsilon - \overline{X}_t|^2 \leqslant \mathscr{C}\varepsilon.
$$
□

4 HMM scheme

For 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 $\overline{X}_t$, as described by equation (1.2). The difficulty of dealing with $\overline{X}_t$ lies in the fact that the coefficients $\overline{a}(\cdot)$ and $\overline{b}(\cdot)$ are given via expectations with respect to measure $\mu_x^\varepsilon(\mathrm{d}y)$, which is usually difficult or impossible to obtain analytically, especially when the dimension m is large. The basic idea of HMM is to solve $\overline{X}_t$ with $\overline{a}(\cdot)$ and $\overline{\sigma}(\cdot)$ being estimated on-the-fly using a time-ensemble average of the original slow coefficients $a(\cdot)$ and $b(\cdot)$ with respect to numerical solutions of the fast processes. The method consists of a three-step procedure [11,12,17]. At each time step, we evolve the effective dynamics (1.2) with *a macro solver* using approximate coefficients, which is obtained by *a micro solver* for the fast dynamics in (1.1) with fixed slow variables being the same as those in the macro solver. Then $\widetilde{a}$ and $\widetilde{b}$, which approximate $\overline{a}$ and $\overline{b}$, can be given by *an estimator* taking the form of a time and ensemble average of the fast solutions. The new challenge proposed by the jump components in (1.1) is to efficiently generate Poisson random variables with size distribution μ_x^ε. We will leave this to future investigations.

5 Conclusion

We proved the strong convergence of the Principle of Averaging for jump-diffusion processes with two well separated time scales. Optimal rate of convergence is provided. The analytical results will help to design efficient and accurate numerical schemes for systems of this type.

Acknowledgements The author wants to thank Weinan E and Eric Vanden-Eijnden for stimulating discussion. This work was supported by the NSF-DMS 0609315.

Appendix Limiting properties of fast processes

Here, we want to provide some properties for the fast process $Z^\varepsilon_{x,t}$ defined in (3.6) on the infinite time horizon. The ergodicity and uniqueness of invariant measures of $Z^\varepsilon_{x,t}$ for each fixed (x,ε) under Assumptions 3.1–3.3 have been established in [7,13]. Some sharper estimates are needed for the purpose of proving our theorems here. We will skip the details of the proofs since they take little modifications of those in [11]. First, we give an energy estimate for $Z^\varepsilon_{x,t}$ and its invariant measure.

Lemma A1 *There exists a constant $\mathscr{C}$ such that for all $t \geqslant 0$, we have*

$$\mathbb{E}|Z^\varepsilon_{x,t}|^2 \leqslant \mathrm{e}^{-\beta t/\varepsilon}|z|^2 + \mathscr{C}(|x|^2 + \varepsilon^2 + 1), \tag{A1}$$

and the invariant measure μ^ε_x has a finite second order moment:

$$\int z^2 \mu^\varepsilon_x(\mathrm{d}z) \leqslant \mathscr{C}(|x|^2 + \varepsilon^2 + 1). \tag{A2}$$

Based on ergodicity and Lemma A1, we can prove the existence of a stationary solution for equation (3.6) satisfied by $Z^\varepsilon_{x,t}$.

Lemma A2 *For each fixed (x,ε), there exists a process $\xi^\varepsilon_{x,t}$ defined over the whole time domain $t \in (-\infty,\infty)$ such that it satisfies* (3.6) *with a stationary probability distribution that agrees with the invariant measure of* (3.6), *i.e.,*

$$\dot{\xi}^\varepsilon_{x,t} = \frac{1}{\varepsilon} f(x,\xi^\varepsilon_{x,t},\varepsilon) + \frac{1}{\sqrt{\varepsilon}} g(x,\xi^\varepsilon_{x,t},\varepsilon)\dot{W}_t + h(x,\xi^\varepsilon_{x,t-},\varepsilon)\dot{N}^\varepsilon_t, \quad \mathscr{L}(\xi^\varepsilon_{x,t}) = \mu^\varepsilon_x. \tag{A3}$$

Moreover, we have

$$\mathbb{E}|Z^\varepsilon_{x,t} - \xi^\varepsilon_{x,t}|^2 \leqslant \mathscr{C}(|z|^2 + |x|^2 + \varepsilon^2 + 1)\mathrm{e}^{-2\beta t/\varepsilon}. \tag{A4}$$

We can also give the strong rate of convergence to the equilibrium for the process $Z^\varepsilon_{x,t}$.

Lemma A3 *There exists a constant $\mathscr{C}$ such that for any function f with bounded derivatives,*

$$\left|\mathbb{E}f(Z^\varepsilon_{x,t}) - \int f(z)\mu^\varepsilon_x(\mathrm{d}z)\right| \leqslant \mathscr{C}\sup|f'|(|x| + |z| + \varepsilon + 1)\mathrm{e}^{-\beta t/\varepsilon}. \tag{A5}$$

The following lemma describes the behavior of the solution $Z^{\varepsilon}_{x,z,t}$ of (3.6) at equilibrium under the perturbation in x.

Lemma A4 *Functions* $\hat{a}(x,z,t,\varepsilon)$ *and* $\overline{a}(x,\varepsilon)$ *defined by* (3.7) *and* (3.8) *are smooth functions with bounded derivatives. In addition, for any multi-index* $I=(I_1,\ldots,I_n)$, *we have*

$$|\nabla_x^I(\hat{a}(x,z,t,\varepsilon)-\overline{a}(x,\varepsilon))| \leqslant \mathscr{C}_I \mathrm{e}^{-\beta t/\varepsilon}, \tag{A6}$$

and for any multi-indices $J=(J_1,\ldots,J_m)$, *we have*

$$|\nabla_z^J\nabla_x^I(\hat{a}(x,z,t,\varepsilon))| \leqslant \mathscr{C}_{I,J} \mathrm{e}^{-\beta t/\varepsilon}. \tag{A7}$$

References

1. E W, Engquist B. The heterogeneous multiscale methods. Commun Math Sci, 2003, 1(1): 87–133
2. E W, Liu D, Vanden-Eijnden E. Analysis of multiscale methods for stochastic differential equations. Commun Pure Appl Math, 2005, 58(11): 1544–1585
3. Freidlin M I, Wentzell A D. Random Perturbations of Dynamical Systems. 2nd ed. New York: Springer-Verlag, 1998
4. Givon D. Strong convergence rate for two-time-scale jump-diffusion stochastic differential systems. SIAM Mul Mod Simu, 2007, 6: 577–594
5. Givon D, Kevrekidis I G, Kupferman R. Strong convergence of projective integration schemes for singularly perturbed stochastic differential systems. Commun Math Sci, 2006, 4(4): 707–729
6. Khasminskii R Z. Principle of averaging for parabolic and elliptic differential equations and for Markov processes with small diffusion. Theory Probab Appl, 1963, 8: 1–21
7. Khasminskii R Z. Stochastic Stability of Differential Equations. Alphen aan den Rijn: Sijthoff & Noordhoff, 1980
8. Khasminskii R Z, Yin G. On averaging principles: An asymptotic expansion approach. SIAM J Math Anal, 2004, 35(6): 1534–1560, 2004
9. Kifer Y. Stochastic versions of Anosov and Neistadt theorems on averaging. Stoch Dyn, 2001, 1(1): 1–21
10. Kushner H J. Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems, 3, Systems & Control: Foundations & Applications. Boston: Birkhäuser, 1990
11. Liu D. Strong convergence of principle of averaging for multiscale stochastic dynamical systems. Commun Math Sci, 2010, 8: 999–1020
12. Liu D. Analysis of multiscale methods for stochastic dynamical systems with multiple time scales. SIAM Mul Mod Simu, 2010, 8: 944–964
13. Menaldi J L, Robin M. Invariant measure for diffusions with jumps. Appl Math Optim, 1999, 40: 105–140
14. Meyn S P, Tweedie R L. Stability of Markovian processes, I. Adv Appl Probab, 1992, 24(3): 542–574
15. Meyn S P, Tweedie R L. Stability of Markovian processes, II. Adv Appl Probab, 1993, 25(3): 487–517
16. Meyn S P, Tweedie R L. Stability of Markovian processes, III. Adv Appl Probab, 1993, 25(3): 518–548
17. Vanden-Eijnden E. Numerical techniques for multi-scale dynamical systems with stochastic effects. Commun Math Sci, 2003, 1: 377–384
18. Veretennikov A Yu. On an averaging principle for systems of stochastic differential equations. Math Sbornik, 1990, 181(2): 256–268