Explicit multiscale numerical method for super-linear slow-fast stochastic differential equations

Yuanping Cui [*] Xiaoyue Li [†] Xuerong Mao [‡]

Abstract

This work focuses on solving super-linear stochastic differential equations (SDEs) involving different time scales numerically. Taking advantages of being explicit and easily implementable, a multiscale truncated Euler-Maruyama scheme is proposed for slow-fast SDEs with local Lipschitz coefficients. By virtue of the averaging principle, the strong convergence of its numerical solutions to the exact ones in pth moment is obtained. Furthermore, under mild conditions on the coefficients, the corresponding strong error estimate is also provided. Finally, two examples and some numerical simulations are given to verify the theoretical results.

Keywords. Slow-fast stochastic differential equations; Super-linearity; Explicit multiscale scheme; pth moment; Strong convergence.

2020 Mathematics Subject Classification: 65C30; 60H35

1 Introduction

Stochastic modelling plays an essential role in many branches of science and industry. Especially, super-linear stochastic differential equations (SDEs) are usually used to describe real-world systems in various applications, for examples, the stochastic Lotka-Volterra model in biology for the population growth [33], the elasticity of volatility model arising in finance for the asset price [26] and the stochastic Ginzburg-Landau equation stemming from statistical physics in the study of phase transitions [24]. In many fields, various factors change at different rates: some vary rapidly whereas others evolve slowly. As a result the separation of fast and slow time scales arises in chemistry, fluid dynamics, biology, physics, finance and other fields [5, 14, 17, 25]. Stochastic

[*]School of Mathematical Sciences, Tiangong University, Tianjin, 300387, China.

[†]School of Mathematical Sciences, Tiangong University, Tianjin, 300387, China. Research of this author was supported by the National Natural Science Foundation of China (No. 12371402), the National Key R&D Program of China (2020YFA0714102) and the Natural Science Foundation of Jilin Province (No. YDZJ202101ZYTS154).

[‡]Department of Mathematics and Statistics, University of Strathclyde, Glasgow G1 1XH, UK. Research of this author was supported by the Royal Society (WM160014, Royal Society Wolfson Research Merit Award), the Royal Society of Edinburgh (RSE1832), and Shanghai Administration of Foreign Experts Affairs (21WZ2503700, the Foreign Expert Program).

systems with this characteristic are studied extensively [10, 36, 37, 48] and are often modeled by the slow-fast SDEs (SFSDEs)

$$\begin{cases} \mathrm{d}x^\varepsilon(t) = b(x^\varepsilon(t), y^\varepsilon(t))\mathrm{d}t + \sigma(x^\varepsilon(t))\mathrm{d}W^1(t), \\ \mathrm{d}y^\varepsilon(t) = \dfrac{1}{\varepsilon} f(x^\varepsilon(t), y^\varepsilon(t))\mathrm{d}t + \dfrac{1}{\sqrt{\varepsilon}} g(x^\varepsilon(t), y^\varepsilon(t))\mathrm{d}W^2(t) \end{cases} \tag{1.1}$$

with initial value $(x^\varepsilon(0), y^\varepsilon(0)) = (x_0, y_0) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$. Here, coefficients

$$\begin{aligned} &b : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}^{n_1}, \qquad \sigma : \mathbb{R}^{n_1} \to \mathbb{R}^{n_1 \times d_1}, \\ &f : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}^{n_2}, \qquad g : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}^{n_2 \times d_2} \end{aligned}$$

are continuous, while $\{W^1(t)\}_{t\geq 0}$ and $\{W^2(t)\}_{t\geq 0}$ represent mutually independent d_1-dimensional and d_2-dimensional Brownian motions, respectively. The parameter $\varepsilon > 0$ represents the ratio of nature time scales between $x^\varepsilon(t)$ and $y^\varepsilon(t)$. Especially, as $\varepsilon \ll 1$, $x^\varepsilon(t)$ and $y^\varepsilon(t)$ are called the slow component and fast component, respectively.

In various applications the time evolution of the slow component $x^\varepsilon(t)$ is under the spotlight. Due to the existence of super-linear coefficients and multiple time scales as well as the coupling of fast and slow components, it is almost impossible to anticipate the dynamics of slow-fast components directly by solving the full system. Therefore, numerical methods or approximation techniques become efficient tools. Hence, our main aim is to construct an appropriate numerical scheme to approximate the slow component.

The averaging principle is one of the key techniques in the theoretical analysis of SFSDEs. It essentially describes the asymptotic behavior of the slow component as $\varepsilon \to 0$. Precisely, let the frozen equation described by

$$\mathrm{d}y^{x,y_0}(t) = f(x, y^{x,y_0}(t))\mathrm{d}s + g(x, y^{x,y_0}(t))\mathrm{d}W^2(t) \tag{1.2}$$

with initial value $y^x(0) = y_0$, where x is regarded as a parameter. If the transition semigroup of $y^{x,y_0}(t)$ has a unique invariant probability measure μ^x , which is independent of y_0 of course, and the following integral

$$\bar{b}(x) = \int_{\mathbb{R}^{n_2}} b(x, y)\mu^x(\mathrm{d}y). \tag{1.3}$$

exists, the averaging principle reveals that under suitable assumptions on the coefficients the slow component $x^\varepsilon(t)$ converges to $\bar{x}(t)$, which is the solution of

$$\begin{cases} \mathrm{d}\bar{x}(t) = \bar{b}(\bar{x}(t))\mathrm{d}t + \sigma(\bar{x}(t))\mathrm{d}W^1(t), \\ \bar{x}(0) = x_0. \end{cases} \tag{1.4}$$

The averaging principle was originally developed by Khasminskii [23]. Subsequently, fruitful results on the averaging principle have been developed in the linear framework [8, 13, 15, 16, 30, 45, 46]. Recently, growing interests have been drawn to the study of the averaging principle for SFSDEs with super-linear growth coefficients. Liu et al. [32] proved the strong convergence of the averaging principle as the drift coefficients are locally Lipschitz continuous with respect to the slow and fast variables. Hong et al. [19] gave the 1/6-order strong convergence rate for a class of nonlinear stochastic partial differential equations (SPDEs). Shi et al. [41] obtained the optimal convergence rate for SFSDEs driven by Lévy processes, which slow drift coefficient satisfies the monotonicity condition and grows polynomially. Furthermore, the strong averaging principles have been developed for various kinds of slow-fast stochastic systems, such as jump-diffusion processes [11, 49], SPDEs [2, 4, 9], McKean-Vlasov SDEs [38], and so on.

The averaged equation derived from the averaging principle provides a substantial simplification of the original system. However, it is almost impossible to get the explicit form of the invariant measure μ^x in the averaged equation (1.4) due to the complicated dynamics of the frozen equation (1.2). Thus, the heterogeneous multiscale method (HMM) [6, 7] was proposed to approximate the averaged equation numerically. This facilitated the development of the numerical approximation theory for the SFSDEs. In 2003, Vanden-Eijnden [43] proposed a numerical scheme for the deterministic multi-scale system without rigorous analysis. E et al. [8] provided a thorough analysis of the convergence and efficiency of the HMM scheme for SFSDEs without slow diffusion term, where the slow drift and fast diffusion coefficients are bounded and the fast drift coefficient is a smooth function with bounded derivatives of any order. In 2006, Givon et al. [13] developed the projective integration schemes for SFSDEs in which the slow drift and diffusion coefficients satisfy the Lipschitz condition and the fast drift and diffusion coefficients are bounded. In 2008, Givon et al. [12] went a further step to extend the projective integration schemes for jump-diffusion systems. In 2010, Liu [29] established the HMM numerical theory for the fully coupled SFSDEs, where the slow drift and diffusion coefficients are bounded, all coefficients are smooth and have bounded derivatives with any order. Bréhier [2, 3] developed the HMM scheme for the slow-fast parabolic stochastic partial differential equations. All of the above studies was carried out under the linear growth condition, so the Euler-Maruyama (EM) sheme is used as a macro solver to simulate the evolution of the slow component owing to its simple algebraic structure and the cheap computational cost.

The super-linear growth coefficients of the slow-fast stochastic systems bring the super-linear structure to the averaged equation. For an example, consider a SFSDE with a super-linear slow drift

$$\begin{cases} \mathrm{d}x^\varepsilon(t) = \big(-(x^\varepsilon(t))^3 - y^\varepsilon(t)\big)\mathrm{d}t + x^\varepsilon(t)\mathrm{d}W^1(t), \\ \mathrm{d}y^\varepsilon(t) = \dfrac{1}{\varepsilon}\big(x^\varepsilon(t) - y^\varepsilon(t)\big)\mathrm{d}t + \dfrac{1}{\sqrt{\varepsilon}}\mathrm{d}W^2(t) \end{cases} \tag{1.5}$$

with $(x^\varepsilon(0), y^\varepsilon(0)) = (x_0, y_0)$. The corresponding frozen equation is described by

$$\mathrm{d}y^x(s) = (x - y^x(s))\mathrm{d}s + \mathrm{d}W^2(s). \tag{1.6}$$

By solving the Fokker-Planck equation, the invariant probability density of (1.6) is $\mu^x(dy) =$

$\frac{e^{-(y-x)^2}}{\sqrt{\pi}}\mathrm{d}y$. Then the averaged equation is described by

$$\mathrm{d}\bar{x}(t) = \big(-\bar{x}^3(t) - \bar{x}(t)\big)\mathrm{d}t + \bar{x}(t)\mathrm{d}W^1(t). \tag{1.7}$$

As pointed out by [21] the EM approximation error of the above equation diverges to infinity in pth moment for any $p \geq 1$. In fact, we implement the Projective Integration (PI) scheme with the EM scheme as the macro-solver, as detailed in [13, (4.1)-(4.4)], to simulate the averaged equation (1.7). However, the numerical solution generated by the PI scheme blows up quickly, see Figure 1, differing from the dynamics of the underlying exact solution. Therefore, using the EM scheme as a macro solver to simulate the averaged equation of SFSDE with super-linear coefficients leads to the divergence possibly. Although implicit numerical methods are feasible as the macro solver for the super-linear averaged equation, their application may make the algorithm and implementation more involved and expensive [8]. As a consequence, to construct an appropriate explicit numerical scheme for super-linear SFSDEs to overcome the numerical stiffness becomes an urgent target.

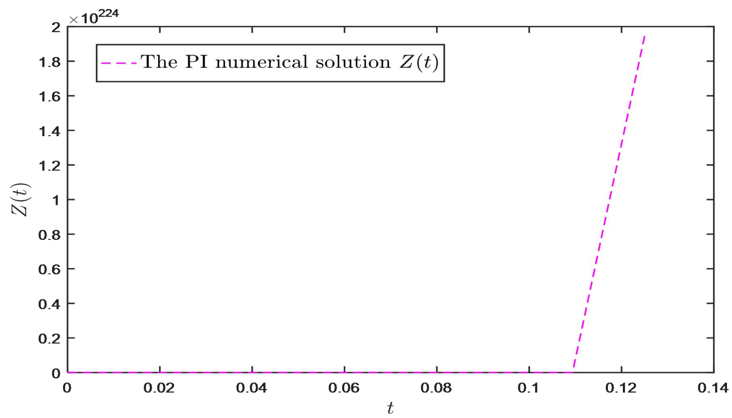

Figure 1: The sample paths of the PI numerical solution $Z(t)$ on $t \in [0,3]$ with $\Delta_1 = 2^{-6}$, $\Delta_2 = 2^{-6}$ and $M = 2^{18}$.

Fortunately, great achievements have been made in the research of explicit numerical methods for super-linear SDEs, for examples, the tamed EM scheme [20, 22, 39, 40], the tamed Milstein scheme [47], the stopped EM scheme [31], the truncated EM scheme [27, 28, 34] and therein. So far the ability of these modified EM methods to approximate the solutions of super-linear diffusion systems has been displayed comprehensively. Inspired by the above works, we are devoted to constructing an explicit multiscale numerical method suitable for super-linear SFSDEs.

In fact, HMM relies heavily on the structure of the averaged equation for the slow variable. Thus we have to overcome two major obstacles: the unknown form and super-linear structure of $\bar{b}(\cdot)$. Borrowing the idea from [34], we design a truncation device to modify the super-linear coefficient b of original slow system in advance, so as to achieve the modification of $\bar{b}$. This modification can avoid possible large excursion from the super-linearity of $\bar{b}(\cdot)$. Then fitting into the framework of HMM, we construct an explicit multiscale numerical scheme involving three subroutines as follows.

1. The truncated EM (TEM) scheme is selected as the macro solver to predict the macro dynamics $\bar{x}(t) \approx x^{\varepsilon}(t)$ in which the modified averaged coefficient is required to be estimated at each macro time step.

2. An appropriate numerical scheme is chosen as the micro solver to solve the frozen equation to produce the data used for approximating the modified averaged coefficient.

3. An estimator is established to obtain the desired approximation of the modified averaged coefficient.

Following this line, we construct an easily implementable explicit multiscale numerical scheme for a class of super-linear SFSDEs and obtain its strong convergence.

The rest of this paper is organized as follows. Section 2 gives some notations, hypotheses and preliminaries. Section 3 proposes an explicit multiscale numerical method. Section 4 provides some important pre-estimates. Section 5 yields the strong convergence of MTEM scheme. Section 6 focuses on the error analysis of the explicit MTEM scheme and presents an important example. Section 7 shows two numerical examples and carries out some numerical experiments to verify our theoretical results. Section 8 concludes this paper.

2 Preliminary

Throughout this paper, we use the following notations. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space with a natural filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions (i.e. it is right continuous and increasing while $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets), and $\mathbb{E}$ be the expectation corresponding to $\mathbb{P}$. Let $|\cdot|$ denote the Euclidean norm in $\mathbb{R}^n$ and the trace norm in $\mathbb{R}^{n\times d}$. If A is a vector or matrix, we denote its transpose by A^T. For a set $\mathbb{D}$, let $I_{\mathbb{D}}(x) = 1$ if $x \in \mathbb{D}$ and 0 otherwise. We set $\inf \emptyset = \infty$, where $\emptyset$ is empty set. Moreover, for any $a, b \in \mathbb{R}$, we define $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. We use C and C_l to denote the generic positive constants, which may take different values at different appearances, where the subscript l in C_l is used to to highlight that this constant depends on the l. In addition, C, C_l are independent of parameters Δ_1, Δ_2, n and M that occur in the next section. In particular, C_R usually denotes some positive function increasing with respect to R.

Let $\mathcal{P}(\mathbb{R}^{n_2})$ denote the family of all probability measures on $\mathbb{R}^{n_2}$. For any $p \geq 1$, let $\mathcal{P}_p(\mathbb{R}^{n_2})$ be the set in $\mathcal{P}(\mathbb{R}^{n_2})$ with finite p-th moment, i.e.,

$$\mathcal{P}_p(\mathbb{R}^{n_2}) := \Big\{\mu \in \mathcal{P}(\mathbb{R}^{n_2}) : \int_{\mathbb{R}^{n_2}} |y|^p \mu(\mathrm{d}y) < \infty\Big\},$$

which is a Polish space under the Wasserstein distance

$$\mathbb{W}_p(\mu_1, \mu_2) = \inf_{\pi \in \mathcal{C}(\mu_1,\mu_2)} \Big(\int_{\mathbb{R}^{n_2}\times\mathbb{R}^{n_2}} |y_1 - y_2|^p \pi(\mathrm{d}y_1, \mathrm{d}y_2)\Big)^{\frac{1}{p}},$$

where $\mathcal{C}(\mu_1, \mu_2)$ stands for the set of all probability measures on $\mathbb{R}^{n_2} \times \mathbb{R}^{n_2}$ with marginals μ_1 and μ_2, respectively.

To state the main results, we impose some hypotheses on the coefficients b, σ of slow equation and f and g of fast equation.

(**S1**) There exists a constant $\theta_1 \geq 1$ such that for any $R > 0$, $x_1, x_2 \in \mathbb{R}^{n_1}$ with $|x_1| \vee |x_2| \leq R$ and $y \in \mathbb{R}^{n_2}$,

$$|b(x_1, y) - b(x_2, y)| + |\sigma(x_1) - \sigma(x_2)| \leq L_R |x_1 - x_2|(1 + |y|^{\theta_1}),$$

here L_R is a positive constant dependent on R.

(**S2**) There exist constants $\theta_2 > 0$ and $K_1 > 0$ such that for any $x \in \mathbb{R}^{n_1}$ and $y_1, y_2 \in \mathbb{R}^{n_2}$.

$$|b(x, y_1) - b(x, y_2)| \leq K_1 |y_1 - y_2| \big(1 + |x|^{\theta_2} + |y_1|^{\theta_2} + |y_2|^{\theta_2}\big).$$

(**S3**) There exists a constant $K_2 > 0$ such that for any $x \in \mathbb{R}^{n_1}$,

$$|\sigma(x)| \leq K_2(1 + |x|).$$

(**S4**) There exist constants $\theta_3, \theta_4 \geq 1$ and $K_3 > 0$ such that for any $x \in \mathbb{R}^{n_1}, y \in \mathbb{R}^{n_2}$,

$$|b(x, y)| \leq K_3(1 + |x|^{\theta_3} + |y|^{\theta_4}).$$

(**S5**) There exist constants $K_4 > 0$ and $\lambda > 0$ such that for any $x \in \mathbb{R}^{n_1}, y \in \mathbb{R}^{n_2}$,

$$x^T b(x, y) \leq K_4(1 + |x|^2) + \lambda |y|^2.$$

(**F1**) The functions f and g are globally Lipschitz continuous, namely, for any $x_1, x_2 \in \mathbb{R}^{n_1}$ and $y_1, y_2 \in \mathbb{R}^{n_2}$, there exists a positive constant L such that

$$|f(x_1, y_1) - f(x_2, y_2)| \vee |g(x_1, y_1) - g(x_2, y_2)| \leq L(|x_1 - x_2| + |y_1 - y_2|).$$

(**F2**) There exists a constant $\beta > 0$ such that for any $x \in \mathbb{R}^{n_1}$ and $y_1, y_2 \in \mathbb{R}^{n_2}$,

$$2(y_1 - y_2)^T \big(f(x, y_1) - f(x, y_2)\big) + |g(x, y_1) - g(x, y_2)|^2 \leq -\beta |y_1 - y_2|^2.$$

(**F3**) For some fixed $k \geq 2$, there exist constants $\alpha_k > 0$ and $L_k > 0$ such that for any $x \in \mathbb{R}^{n_1}$, $y \in \mathbb{R}^{n_2}$,

$$y^T f(x, y) + \frac{k-1}{2} |g(x, y)|^2 \leq -\alpha_k |y|^2 + L_k(1 + |x|^2).$$

Remark 2.1. Referring to [32, Theorem 2.2], system (1.1) admits a unique global solution $(x^\varepsilon(t), y^\varepsilon(t))$ under (**S1**)-(**S5**) and (**F1**)-(**F3**). Obviously, (**F1**) guarantees that the frozen equation (1.2) has a unique global solution $y^{x,y_0}(s)$, which is a time homogeneous Markov process.

Lemma 2.1. If (**F1**)-(**F3**) hold with some $k \geq 2$, then for any fixed $x \in \mathbb{R}^{n_1}$, the transition semigroup $\{\mathbb{P}_t^x\}_{t\geq 0}$ of equation (1.2) has a unique invariant probability measure $\mu^x \in \mathcal{P}_k(\mathbb{R}^{n_2})$, which satisfies that

$$\int_{\mathbb{R}^{n_2}} |y|^k \mu^x(\mathrm{d}y) \leq C(1+|x|^k). \tag{2.1}$$

Furthermore, for any $x_1, x_2 \in \mathbb{R}^{n_1}$,

$$\mathbb{W}_2(\mu^{x_1}, \mu^{x_2}) \leq C|x_1 - x_2|. \tag{2.2}$$

Proof. For any fixed $x \in \mathbb{R}^{n_1}$ and $y_0 \in \mathbb{R}^{n_2}$, under (**F3**) it follows from [32, Lemma 3.6] that

$$\sup_{t\geq 0} \mathbb{E}|y^{x,y_0}(t)|^k \leq C(1+|x|^k) < \infty,$$

which implies that $\delta_{y_0}\mathbb{P}_t^x \in \mathcal{P}_k(\mathbb{R}^{n_2}) \subset \mathcal{P}_2(\mathbb{R}^{n_2})$. It is well known that for any $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^{n_2})$

$$\begin{aligned}
\mathbb{W}_2(\mu_1\mathbb{P}_t^x, \mu_2\mathbb{P}_t^x) &\leq \int_{\mathbb{R}^{n_2}\times\mathbb{R}^{n_2}} \mathbb{W}_2(\delta_{y_1}\mathbb{P}_t^x, \delta_{y_2}\mathbb{P}_t^x)\pi(\mathrm{d}y_1, \mathrm{d}y_2) \\
&\leq \int_{\mathbb{R}^{n_2}\times\mathbb{R}^{n_2}} \left(\mathbb{E}|y^{x,y_1}(t) - y^{x,y_2}(t)|^2\right)^{\frac{1}{2}} \pi(\mathrm{d}y_1, \mathrm{d}y_2),
\end{aligned}$$

here $\pi \in \mathcal{C}(\mu_1, \mu_2)$. Then under (**F2**), by virtue of [32, Lemma 3.7] we derive that

$$\begin{aligned}
\mathbb{W}_2(\mu_1\mathbb{P}_t^x, \mu_2\mathbb{P}_t^x) &\leq Ce^{-\frac{\beta t}{2}} \int_{\mathbb{R}^{n_2}\times\mathbb{R}^{n_2}} |y_1 - y_2|\pi(\mathrm{d}y_1, \mathrm{d}y_2) \\
&\leq Ce^{-\frac{\beta t}{2}} \Big(\int_{\mathbb{R}^{n_2}\times\mathbb{R}^{n_2}} |y_1 - y_2|^2\pi(\mathrm{d}y_1, \mathrm{d}y_2)\Big)^{\frac{1}{2}}.
\end{aligned}$$

Then due to the arbitrariness of $\pi \in C(\mu_1, \mu_2)$, we have

$$\mathbb{W}_2(\mu_1\mathbb{P}_t^x, \mu_2\mathbb{P}_t^x) \leq Ce^{-\frac{\beta t}{2}}\mathbb{W}_2(\mu_1, \mu_2),$$

which yields the uniqueness of invariant measure if it exists. Next we shall prove the existence of invariant probability measure. In fact, it is sufficient to prove that for any fixed $x \in \mathbb{R}^{n_1}$ and $y_0 \in \mathbb{R}^{n_2}$, $\{\delta_{y_0}\mathbb{P}_t^x\}_{s\geq 0}$ is a $\mathbb{W}_2$-Cauchy sequence due to the completeness of $\mathcal{P}_2(\mathbb{R}^{n_2})$ space, where δ_{y_0} is the Dirac measure with mass at point $y_0 \in \mathbb{R}^{n_2}$. Using the Kolmogorov-Chapman equation and [32, Lemma 3.7], one derives that for any $t, s > 0$,

$$\begin{aligned}
\mathbb{W}_2(\delta_{y_0}\mathbb{P}_t^x, \delta_{y_0}\mathbb{P}_{t+s}^x) &= \mathbb{W}_2(\delta_{y_0}\mathbb{P}_t^x, \delta_{y_0}\mathbb{P}_t^x\mathbb{P}_s^x) \leq Ce^{-\frac{\beta t}{2}}\mathbb{W}_2(\delta_{y_0}, \delta_{y_0}\mathbb{P}_s^x) \\
&\leq Ce^{-\frac{\beta t}{2}}\left(|y_0|^2 + \mathbb{E}|y^{x,y_0}(s)|^2\right)^{\frac{1}{2}} \leq Ce^{-\frac{\beta s}{2}}(1 + |x| + |y_0|),
\end{aligned}$$

which implies that as $t \to \infty$, $\{\delta_{y_0}\mathbb{P}_t^x\}_{s\geq 0}$ is a $\mathbb{W}_2$-Cauchy sequence whose limit is denoted by μ^x. Furthermore, in view of the continuity of $\mathbb{W}_2$-distance(see, [44, Corollary 6.1]) we derive for any $t > 0$

$$\mathbb{W}_2(\mu^x\mathbb{P}_t^x, \mu^x) = \lim_{s\to\infty} \mathbb{W}_2(\delta_{y_0}\mathbb{P}_{s+t}^x, \delta_y\mathbb{P}_s^x) = 0,$$

which implies that μ^x is indeed an invariant probability measure of $y^x(s)$. On the other hand, it follows from [32, Proposition 3.8] that

$$\int_{\mathbb{R}^{n_2}} |y|^k \mu^x(\mathrm{d}y) \leq C(1+|x|^k).$$

In addition, using the continuity of $\mathbb{W}_2$ again yields that

$$\begin{aligned}\mathbb{W}_2^2(\mu^{x_1},\mu^{x_2}) &= \lim_{t\to\infty} \mathbb{W}_2^2(\delta_{y_0}\mathbb{P}_t^{x_1},\delta_{y_0}\mathbb{P}_t^{x_2})\\ &\leq \lim_{t\to\infty} \mathbb{E}|y^{x_1,y_0}(t)-y^{x_2,y_0}(t)|^2 \leq C|x_1-x_2|^2,\end{aligned}$$

where the last step follows from the [32, Lemma 3.10]. The proof is complete. □

The averaged equation (1.4), obtained via the averaging principle, substantially reduces the complexity of the original system (1.1). Therefore, a numerical approach for system (1.1) can be developed by formulating a numerical method for the averaged equation (1.4). To facilitate this, we cite some known results on the averaging principle firstly.

Lemma 2.2 ([32, Theorem 2.3])**.** If (**S1**)-(**S5**) and (**F1**)-(**F3**) hold with $k > 4\theta_1 \vee 2(\theta_2+1) \vee 2\theta_3 \vee 2\theta_4$, then for any $0<p<k$ and $T>0$,

$$\lim_{\varepsilon\to 0} \mathbb{E}\Big(\sup_{t\in[0,T]} |x^\varepsilon(t)-\bar{x}(t)|^p\Big)=0,$$

where $x^\varepsilon(t)$ and $\bar{x}(t)$ are the solutions of (1.1) and (1.4), respectively.

Lemma 2.3 ([32, Lemma 3.11])**.** If (**S1**)-(**S3**), (**S5**) and (**F1**)-(**F3**) hold with $k \geq 2\vee\theta_1\vee 2\theta_2\vee\theta_4$, then for any $x_0\in\mathbb{R}^{n_1}$, the averaged equation (1.4) has a unique global solution $\bar{x}(t)$ satisfying

$$\mathbb{E}\Big(\sup_{0\leq t\leq T} |\bar{x}(t)|^p\Big) \leq C_{x_0,T,p}, \qquad \forall\ p>0,\ T>0.$$

Remark 2.2. For any constant $R>|x_0|$, define the stopping time

$$\tau_R = \inf\{t\geq 0 : |\bar{x}(t)|\geq R\}.$$

Then it follows from Lemma 2.3 that

$$R^p\mathbb{P}(\tau_R\leq T) \leq \mathbb{E}|\bar{x}(T\wedge\tau_R)|^p \leq \mathbb{E}\Big(\sup_{0\leq t\leq T}|\bar{x}(t)|^p\Big) \leq C_{x_0,T,p},$$

which implies that

$$\mathbb{P}(\tau_R\leq T) \leq \frac{C_{x_0,T,p}}{R^p}.$$

3 The construction of explicit multiscale scheme

With the help of the strong averaging principle, this section is devoted to constructing an easily implementable multiscale numerical scheme for the slow component of original SFSDE (1.1). One notices from (**S4**) that for any $y\in\mathbb{R}^{n_2}$,

$$|b(x,y)| \leq K_3(1+|x|)(1+|x|^{\theta_3-1}) + K_3|y|^{\theta_4}.$$

Then for any $u \geq 1$ and $x \in \mathbb{R}^{n_1}$ with $|x| \leq u$

$$|b(x, y)| \leq K_3 \sup_{|x|\leq u} \varphi(u)(1 + |x|) + K_3|y|^{\theta_4}, \tag{3.1}$$

where $\varphi(u) = 1 + u^{(\theta_3\vee\theta_4-1)}$, and θ_3, θ_4 are given in (**S4**). Then for any step size $\Delta_1 \in (0, 1]$, define

$$x^* = \Big(|x| \wedge \varphi^{-1}\big(K\Delta_1^{-\frac{1}{2}}\big)\Big)\frac{x}{|x|}, \qquad x \in \mathbb{R}^{n_1},$$

where $x/|x| = \mathbf{0} \in \mathbb{R}^{n_1}$ if $x = \mathbf{0}$, and K is a constant satisfying $K \geq\ 1 + \varphi(|x_0| \vee |y_0|)$. Thus,

$$|x_0| \vee |y_0| \leq \varphi^{-1}(K) \leq \varphi^{-1}(K\Delta^{-\frac{1}{2}}), \quad \forall \Delta \in (0, 1]. \tag{3.2}$$

Furthermore, for any $x \in \mathbb{R}^{n_1}$,

$$|b(x^*, y)| \leq C\Delta_1^{-\frac{1}{2}}(1 + |x^*|) + K_3|y|^{\theta_4}. \tag{3.3}$$

Moreover, under (**S4**), (**F1**)-(**F3**) with $k \geq 2 \vee \theta_4$ by the definition (1.3) we derive from the above inequality and (2.1) that

$$\begin{aligned}
|\bar{b}(x^*)| &= \Big|\int_{\mathbb{R}^{n_2}} b(x^*, y)\mu^{x^*}(\mathrm{d}y)\Big| \leq \int_{\mathbb{R}^{n_2}} |b(x^*, y)|\mu^{x^*}(\mathrm{d}y) \\
&\leq C\Delta_1^{-\frac{1}{2}}(1 + |x^*|) + K_3 \int_{\mathbb{R}^{n_2}} |y|^{\theta_4}\mu^{x^*}(\mathrm{d}y) \\
&\leq C\Delta_1^{-\frac{1}{2}}(1 + |x^*|) + C(1 + |x^*|^{\theta_4-1})(1 + |x^*|) \\
&\leq C\Delta_1^{-\frac{1}{2}}(1 + |x^*|), \quad x \in \mathbb{R}^{n_1},
\end{aligned} \tag{3.4}$$

where μ^{x^*} is the unique invariant probability measure of the frozen equation (1.2) with the fixed parameter x^*, and the last step used the increasing of φ.

Because the analytical form of $\bar{b}(x^*)$ is unobtainable, using the ergodicity of the frozen equation (1.2), we approximate $\bar{b}(x^*)$ by the time average of $b(x^*, \cdot)$ with respect to the numerical solution of the frozen equation (1.2) with fixed parameter x^*. For convenience, for an integer $M > 0$, we introduce an average function

$$B_M(x, h) = \frac{1}{M}\sum_{m=1}^{M} b(x, h_m), \quad \forall x \in \mathbb{R}^{n_1}, \tag{3.5}$$

where $h = \{h_m\}_{m=1}^{\infty}$ is an $\mathbb{R}^{n_2}$-valued sequence. Within the framework of HMM, we design an easily implementable multiscale numerical scheme involving a macro solver and a micro solver as well as an estimator. For clarity, we illustrate it as follows. Let Δ_1 and Δ_2 denote macro time step size and micro time step size, respectively.

(1) Macro solver: For the known X_n, since the drift coefficient $\bar{b}$ of the averaged equation may be sup-linear, the truncated EM scheme is selected as macro solver to make a macro step and get X_{n+1}. Then we have

$$X_{n+1} = X_n + B_n\Delta_1 + \sigma(X_n)\Delta W_n^1,$$

where B_n ia an approximation of $\bar{b}(X_n^*)$ that we obtain in third step, and $n_{\Delta_1}(u) := \lfloor u/\Delta_1 \rfloor$ for any $u \geq 0$ with $\lfloor t/\delta \rfloor$ the integer part of t/δ, and $\Delta W_n^1 = W^1((n+1)\Delta_1) - W^1(n\Delta_1)$.

(2) Micro solver: To obtain B_n at each macro time step, for the known $X_n \in \mathbb{R}^{n_1}$, use the EM method to solve the frozen equation (1.2) with parameter $x = X_n^*$ fixed. Therefore, the micro solver is given by

$$\begin{cases} Y_0^{X_n^*, y_0} = y_0, \\ Y_{m+1}^{X_n^*, y_0} = Y_m^{X_n^*, y_0} + f(X_n^*, Y_m^{X_n^*, y_0})\Delta_2 + g(X_n^*, Y_m^{X_n^*, y_0})\Delta W_{n,m}^2, \quad m = 0, 1, \cdots, \end{cases}$$

where $\{W_n^2(\cdot)\}_{n \geq 0}$ is a mutually independent Brownian motion sequence and also independent of $W^1(t)$, and $\Delta W_{n,m}^2 = W_n^2((m+1)\Delta_2) - W_n^2(m\Delta_2)$.

(3) Estimator: For the known X_n and $Y^{X_n^*, y_0} := \{Y_m^{X_n^*, y_0}\}_{m \geq 1}$, let

$$B_n = B_M(X_n^*, Y^{X_n^*, y_0})$$

as an approximation of $\bar{b}(X_n^*)$, where $B_M(\cdot, \cdot)$ is defined by (3.5) and M denotes the number of micro time steps used for this approximation.

Overall, for any given $\Delta_1, \Delta_2 \in (0, 1]$ and integer $M \geq 1$, define the multiscale TEM scheme (MTEM) as follows: for any $n \geq 0$,

$$\begin{cases} X_0 = x_0, \ X_n^* = \Big(|X_n| \wedge \varphi^{-1}\big(K\Delta_1^{-\frac{1}{2}}\big)\Big)\dfrac{X_n}{|X_n|}, \ Y_0^{X_n^*, y_0} = y_0, & \text{(3.6a)} \\ Y_{m+1}^{X_n^*, y_0} = Y_m^{X_n^*, y_0} + f(X_n^*, Y_m^{X_n^*, y_0})\Delta_2 + g(X_n^*, Y_m^{X_n^*, y_0})\Delta W_{n,m}^2, & \text{(3.6b)} \\ \qquad\qquad m = 0, 1, \cdots, M-1, & \\ X_{n+1} = X_n + B_M(X_n^*, Y^{X_n^*, y_0})\Delta_1 + \sigma(X_n)\Delta W_n^1. & \text{(3.6c)} \end{cases}$$

By this scheme we define the continuous approximation processes

$$X(t) = X_n, \qquad t \in [n\Delta_1, (n+1)\Delta_1), \tag{3.7}$$

$$\bar{X}(t) = x_0 + \int_0^t B_M(X^*(s), Y^{X^*(s), y_0})\mathrm{d}s + \int_0^t \sigma(X(s))\mathrm{d}W^1(s). \tag{3.8}$$

Note that $\bar{X}(n\Delta_1) = X(n\Delta_1) = X_n$, that is, $\bar{X}(t)$ and $X(t)$ coincide with the discrete solution at the grid points, respectively.

4 Some pre-estimates

In order to better study the strong convergence of the MTEM scheme, we need to study some important properties of the averaged coefficient $\bar{b}(x)$ and its estimator $B_M(x, Y_n^{x, y_0})$ (defined in later) in advance. In this section, we mainly provide some pre-estimates for $\bar{b}(x)$ and $B_M(x, Y_n^{x, y_0})$.

By virtue of Lemma 2.1, we show that the drift term $\bar{b}$ of the averaged equation (1.4) inherits the local Lipschitz continuity. □

Lemma 4.1. Under (**S1**), (**S2**), (**S4**) and (**F1**)-(**F3**) with $k \geq 2 \vee \theta_1 \vee 2\theta_2 \vee \theta_4$, for any $R > 0$ and $x_1, x_2 \in \mathbb{R}^{n_1}$ with $|x_1| \vee |x_2| \leq R$, there exists a constant $\bar{L}_R$ such that

$$|\bar{b}(x_1) - \bar{b}(x_2)| \leq \bar{L}_R |x_1 - x_2|.$$

Proof. For any $x_1, x_2 \in \mathbb{R}^{n_1}$, according to (1.3) we have

$$\begin{aligned}
|\bar{b}(x_1) - \bar{b}(x_2)| &= \Big| \int_{\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}} (b(x_1, y_1) - b(x_2, y_2)) \pi(\mathrm{d}y_1, \mathrm{d}y_2) \Big| \\
&\leq \int_{\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}} \big| b(x_1, y_1) - b(x_2, y_2) \big| \pi(\mathrm{d}y_1, \mathrm{d}y_2) \\
&\leq \int_{\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}} \big| b(x_1, y_1) - b(x_2, y_1) \big| \pi(\mathrm{d}y_1, \mathrm{d}y_2) \\
&\quad + \int_{\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}} \big| b(x_2, y_1) - b(x_2, y_2) \big| \pi(\mathrm{d}y_1, \mathrm{d}y_2),
\end{aligned}$$

where $\pi \in C(\mu^{x_1}, \mu^{x_2})$ is arbitrary. Then for any $R > 0$ and $x_1, x_2 \in \mathbb{R}^{n_1}$ with $|x_1| \vee |x_2| \leq R$, by the Hölder inequality it follows from (**S1**) and (**S2**) that

$$\begin{aligned}
|\bar{b}(x_1) - \bar{b}(x_2)| &\leq L_R |x_1 - x_2| \int_{\mathbb{R}^{n_2}} (1 + |y_1|^{\theta_1}) \mu^{x_1}(\mathrm{d}y_1) \\
&\quad + K_1 \int_{\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}} |y_1 - y_2| \big(1 + |x_2|^{\theta_2} + |y_1|^{\theta_2} + |y_2|^{\theta_2}\big) \pi(\mathrm{d}y_1, \mathrm{d}y_2) \\
&\leq L_R |x_1 - x_2| \int_{\mathbb{R}^{n_2}} (1 + |y_1|^{\theta_1}) \mu^{x_1}(\mathrm{d}y_1) + K_1 \Big(\int_{\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}} |y_1 - y_2|^2 \pi(\mathrm{d}y_1, \mathrm{d}y_2) \Big)^{\frac{1}{2}} \\
&\quad \times \Big(\int_{\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}} \big(1 + |x_2|^{2\theta_2} + |y_1|^{2\theta_2} + |y_2|^{2\theta_2}\big) \pi(\mathrm{d}y_1, \mathrm{d}y_2) \Big)^{\frac{1}{2}}.
\end{aligned}$$

Then due to the arbitrariness of $\pi \in C(\mu^{x_1}, \mu^{x_2})$, under (**F1**)-(**F3**) with $k \geq \theta_1 \vee 2\theta_2$, applying Lemma 2.1 yields that for any $x_1, x_2 \in \mathbb{R}^{n_1}$ with $|x_1| \vee |x_2| \leq R$,

$$\begin{aligned}
|\bar{b}(x_1) - \bar{b}(x_2)| &\leq C_R |x_1 - x_2| (1 + |x_1|^{\theta_1}) + C \mathbb{W}_2(\mu^{x_1}, \mu^{x_2}) (1 + |x_1|^{\theta_2} + |x_2|^{\theta_2}) \\
&\leq C_R |x_1 - x_2| + C_R \mathbb{W}_2(\mu^{x_1}, \mu^{x_2}) \leq C_R |x_1 - x_2|,
\end{aligned}$$

which implies the desired result. □

Next we reveal that the modified coefficient $\bar{b}(x^*)$ preserves the Khasminskii-like condition for all $\Delta_1 \in (0, 1]$, which is used to obtain the moment bound of the auxiliary process $\bar{Z}(t)$.

Lemma 4.2. If (**S4**), (**S5**) and (**F1**)-(**F3**) hold with $k \geq 2 \vee \theta_4$, then for any $x \in \mathbb{R}^{n_1}$, $\Delta_1 \in (0, 1]$,

$$x^T \bar{b}(x^*) \leq C(1 + |x|^2), \qquad x \in \mathbb{R}^{n_1}.$$

Proof. For $x \in \mathbb{R}^{n_1}$ with $|x| \leq \varphi^{-1}(K \Delta_1^{-1/2})$, $x = x^*$. Using (**S5**) implies that

$$x^T \bar{b}(x^*) = x^T \int_{\mathbb{R}^{n_2}} b(x, y) \mu^x(\mathrm{d}y) \ \leq K_4 (1 + |x|^2) + \lambda \int_{\mathbb{R}^{n_2}} |y|^2 \mu^x(\mathrm{d}y).$$

Using the Hölder inequality and Lemma 2.1, we yield that

$$x^T\bar{b}(x^*) \le C(1+|x|^2). \tag{4.1}$$

On the other hand, for any $x \in \mathbb{R}^{n_1}$ with $|x| > \varphi^{-1}\big(K\Delta_1^{-1/2}\big)$, it follows from the definition of x^* that $x = \Big(|x|/\varphi^{-1}(K\Delta_1^{-1/2})\Big)x^*$. This, together with (4.1), implies that

$$\begin{aligned} x^T\bar{b}(x^*) &= \frac{|x|}{\varphi^{-1}(K\Delta_1^{-\frac{1}{2}})}(x^*)^T\bar{b}(x^*) \le \frac{|x|}{\varphi^{-1}(K\Delta_1^{-\frac{1}{2}})}C(1+|x^*|^2) \\ &\le C|x|\Big(\big(\varphi^{-1}(K)\big)^{-1} + |x|\Big), \end{aligned}$$

where the last inequality used the increasing of φ^{-1}. Thus the desired assertion follows. □

For any fixed $x \in \mathbb{R}^{n_1}$, $y_0 \in \mathbb{R}^{n_2}$, and integer $n \ge 0$, define an auxiliary process $y_n^{x,y_0}(t)$ described by

$$\mathrm{d}y_n^{x,y_0}(t) = f(x, y_n^{x,y_0}(t))\mathrm{d}t + g(x, y_n^{x,y_0}(t))\mathrm{d}W_n^2(t) \tag{4.2}$$

on $t \ge 0$ with initial value $y_n^{x,y_0}(0) = y_0$. Thanks to the weak uniqueness of the solution of the frozen equation (1.2), for any $t \ge 0$, the distribution of $y_n^{x,y_0}(t)$ coincides with that of $y^{x,y_0}(t)$ for any $n \ge 0$. Consequently, according to Lemma 2.1, μ^x is also the unique invariant probability measure of transition semigroup of $y_n^{x,y_0}(t)$ for any $n \ge 0$. Then use the EM scheme for (4.2)

$$\begin{cases} Y_{n,0}^{x,y_0} = y_0, \\ Y_{n,m+1}^{x,y_0} = Y_{n,m}^{x,y_0} + f(x, Y_{n,m}^{x,y_0})\Delta_2 + g(x, Y_{n,m}^{x,y_0})\Delta W_{n,m}^2, \ m = 0, 1, \cdots. \end{cases} \tag{4.3}$$

Furthermore, define

$$\begin{aligned} &Y_n^{x,y_0}(t) = Y_{n,m}^{x,y_0}, \qquad t \in [m\Delta_2, (m+1)\Delta_2), \\ &\bar{Y}_n^{x,y_0}(t) = y_0 + \int_0^t f(x, Y_n^{x,y_0}(s))\mathrm{d}s + \int_0^t g(x, Y_n^{x,y_0}(s))\mathrm{d}W_n^2(s). \end{aligned} \tag{4.4}$$

Let Y_n^{x,y_0} denote the discrete EM solution sequence generated by (4.3). Then, one observes that $Y^{X_n^*,y_0} = Y_n^{X_n^*,y_0}$ a.s. Thus,

$$B_M(X_n^*, Y^{X_n^*,y_0}) = B_M(X_n^*, Y_n^{X_n^*,y_0}) \ \text{ a.s.} \tag{4.5}$$

Next, we first give several properties of Y_n^{x,y_0} in order for the estimation of $B_M(x, Y_n^{x,y_0})$.

Lemma 4.3 ([35, Lemmas 3.7])**.** If (**F1**) and (**F3**) hold with some $k \ge 2$, then there exists a $\hat{\Delta}_2 \in (0,1]$ such that for any $x \in \mathbb{R}^{n_1}$, $y_0 \in \mathbb{R}^{n_2}$, integer $n \ge 0$ and $\Delta_2 \in (0, \hat{\Delta}_2]$,

$$\sup_{m\ge 0} \mathbb{E}|Y_{n,m}^{x,y_0}|^k \le C(1+|y_0|^k+|x|^k),$$

and

$$\sup_{t\ge 0} \mathbb{E}|\bar{Y}_n^{x,y_0}(t) - Y_n^{x,y_0}(t)|^k \le C(1+|y_0|^k+|x|^k)\Delta_2^{\frac{k}{2}}.$$

Lemma 4.4 ([35, Lemmas 3.8])**.** Under (**F1**) and (**F2**), there exists a constant $\bar{\Delta}_2 \in (0,1]$ such that for any $\Delta_2 \in (0, \bar{\Delta}_2)$, $y_1, y_2 \in \mathbb{R}^{n_2}$, $x \in \mathbb{R}^{n_1}$, integers $n \geq 0$ and $m \geq 0$,

$$\mathbb{E}|Y_{n,m}^{x,y_1} - Y_{n,m}^{x,y_2}|^2 \leq C|y_1 - y_2|^2 e^{\frac{-\beta m \Delta_2}{4}}.$$

The results of the above two lemmas can be obtained by the same way as [35, Lemmas 3.7, 3.8]. To avoid duplication we omit the detailed proofs.

Lemma 4.5. If (**F1**)-(**F3**) hold with some $k \geq 2$, then for any fixed $x \in \mathbb{R}^{n_1}$, $y_0 \in \mathbb{R}^{n_2}$, integer $n \geq 0$ and $\Delta_2 \in (0, \bar{\Delta}_2]$, Y_n^{x,y_0} determined by (4.3) admits a unique invariant measure $\mu^{x,\Delta_2} \in \mathcal{P}_k(\mathbb{R}^{n_2})$, which is independent of y_0 and n, and satisfies

$$\int_{\mathbb{R}^{n_2}} |y|^k \mu^{x,\Delta_2}(\mathrm{d}y) \leq C(1 + |x|^k).$$

Proof. Since the EM numerical solutions Y_n^{x,y_0}, $n = 1, \cdots, \infty$ are i.i.d and have Markov property, for any $\Delta_2 \in (0,1)]$, we use $\mathbb{P}_{m\Delta_2}^{x,\Delta_2}$ to denote the same discrete Markov semigroup of Y_n^{x,y_0}. Under (**F1**)-(**F3**), with the help of Lemmas 4.3-4.4, proceeding a similar argument to [1, Theorem 3,1] we derive that for any $y_0 \in \mathbb{R}^{n_2}$ and integer $n \geq 0$, Y_n^{x,y_0} has a unique invariant measure μ^{x,Δ_2}, which is independent of y_0 and n. Furthermore, applying Lemma 4.3 yields that

$$\begin{aligned}
\int_{\mathbb{R}^{n_2}} (|y|^k \wedge N) \mu^{x,\Delta_2}(\mathrm{d}y) &= \int_{\mathbb{R}^{n_2}} \mathbb{E}(|Y_{n,m}^{x,y}|^k \wedge N) \mu^{x,\Delta_2}(\mathrm{d}y) \\
&\leq \int_{\mathbb{R}^{n_2}} (\mathbb{E}|Y_{n,m}^{x,y}|^k \wedge N) \mu^{x,\Delta_2}(\mathrm{d}y) \\
&\leq \int_{\mathbb{R}^{n_2}} \Big(|y|^k e^{-\frac{q\alpha_k m \Delta_2}{8}} \wedge N \Big) \mu^{x,\Delta_2}(\mathrm{d}y) + C(1 + |x|^k),
\end{aligned}$$

where the identity is due to the invariance of invariant measure μ^{x,Δ_2} and the first inequality holds by Jensen's inequality since $x \mapsto N \wedge x, x \in \mathbb{R}$ is a concave function. Then, taking $m \to \infty$ and using the dominated convergence theorem, we deduce that

$$\int_{\mathbb{R}^{n_2}} (|y|^k \wedge N) \mu^{x,\Delta_2}(\mathrm{d}y) \leq C(1 + |x|^k).$$

Letting $N \to \infty$ and applying the monotone convergence theorem, we get

$$\int_{\mathbb{R}^{n_2}} |y|^k \mu^{x,\Delta_2}(\mathrm{d}y) \leq C(1 + |x|^k).$$

The proof is complete. □

Lemma 4.6. If (**F1**)-(**F3**) hold with some $k \geq 2$, then for any fixed $x \in \mathbb{R}^{n_1}, y_0 \in \mathbb{R}^{n_2}$, integer $n \geq 0$ and $\Delta_2 \in (0, \hat{\Delta}_2]$,

$$\sup_{m \geq 0} \mathbb{E}\big|Y_{n,m}^{x,y_0} - y_n^{x,y_0}(m\Delta_2)\big|^2 \leq C(1 + |x|^2)\Delta_2.$$

Proof. In view of (4.2) and (4.4), define $\bar{v}_n^{x,y_0}(t) := \bar{Y}_n^{x,y_0}(t) - y_n^{x,y_0}(t)$ described by

$$\mathrm{d}\bar{v}_n^{x,y_0}(t) = \Big(f(x, Y_n^{x,y_0}(t)) - f(x, y_n^{x,y_0}(t)) \Big)\mathrm{d}t + \Big(g(x, Y_n^{x,y_0}(t)) - g(x, y_n^{x,y_0}(t)) \Big)\mathrm{d}W_n^2(t).$$

Using the Itô formula we arrive at

$$\mathbb{E}\Big(e^{\frac{\beta t}{4}}|\bar{v}_n^{x,y_0}(t)|^2\Big) \leq \mathbb{E}\int_0^t \bigg[\frac{\beta}{4}e^{\frac{\beta s}{4}}\big|\bar{v}_n^{x,y_0}(s)\big|^2 + e^{\frac{\beta s}{4}}\Big(2(\bar{v}_n^{x,y_0}(s))^T\big[f(x,Y_n^{x,y_0}(s)) - f(x,y_n^{x,y_0}(s))\big] + \big|g(x,Y_n^{x,y_0}(s)) - g(x,y_n^{x,y_0}(s))\big|^2\Big)\bigg]\mathrm{d}s. \tag{4.6}$$

Invoking (**F1**), (**F2**) and the Young inequality yields that

$$\begin{aligned}
&2(\bar{v}_n^{x,y_0}(s))^T\big[f(x,Y_n^{x,y_0}(s)) - f(x,y_n^{x,y_0}(s))\big] + \big|g(x,Y_n^{x,y_0}(s)) - g(x,y_n^{x,y_0}(s))\big|^2\\
\leq& 2(\bar{v}_n^{x,y_0}(s))^T\big[f(x,\bar{Y}_n^{x,y_0}(s)) - f(x,y_n^{x,y_0}(s))\big] + \big|g(x,\bar{Y}_n^{x,y_0}(s)) - g(x,y_n^{x,y_0}(s))\big|^2\\
&+ 2(\bar{v}_n^{x,y_0}(s))^T\big[f(x,Y_n^{x,y_0}(s)) - f(x,\bar{Y}_n^{x,y_0}(s))\big] + \big|g(x,Y_n^{x,y_0}(s)) - g(x,\bar{Y}_n^{x,y_0}(s))\big|^2\\
&+ 2\big|g(x,\bar{Y}_n^{x,y_0}(s)) - g(x,y_n^{x,y_0}(s))\big|\big|g(x,Y_n^{x,y_0}(s)) - g(x,\bar{Y}_n^{x,y_0}(s))\big|\\
\leq& -\beta|\bar{v}_n^{x,y_0}(s)|^2 + |\bar{v}_n^{x,y_0}(s)||Y_n^{x,y_0}(s) - \bar{Y}_n^{x,y_0}(s)| + C|Y_n^{x,y_0}(s) - \bar{Y}_n^{x,y_0}(s)|^2\\
\leq& -\frac{\beta}{2}|\bar{v}_n^{x,y_0}(s)|^2 + C|Y_n^{x,y_0}(s) - \bar{Y}_n^{x,y_0}(s)|^2.
\end{aligned}$$

Then inserting the above inequality into (4.6) and using (**F1**) and (**F3**), we derive from the result of lemma 4.3 that

$$e^{\frac{\beta t}{4}}\mathbb{E}|\bar{v}_n^{x,y_0}(t)|^2 \leq C\int_0^t e^{\frac{\beta s}{4}}\mathbb{E}|Y_n^{x,y_0}(s) - \bar{Y}_n^{x,y_0}(s)|^2\mathrm{d}s \leq C(1+|x|^2)\Delta_2 e^{\frac{\beta t}{4}},$$

which yields the desired result. □

Taking Lemma 4.6 into consideration, we deduce the convergence rate between numerical invariant measure μ^{x,Δ_2} and underlying invariant measure μ^x in $\mathbb{W}_2$-distance.

Lemma 4.7. Under (**F1**)-(**F3**) with some $k \geq 2$, for any fixed $x \in \mathbb{R}^{n_1}$ and $\Delta_2 \in (0, \bar{\Delta}_2]$,

$$\mathbb{W}_2(\mu^x, \mu^{x,\Delta_2}) \leq C(1+|x|)\Delta_2^{\frac{1}{2}}.$$

Proof. From the proofs of Lemmas 2.1 and 4.5, we know that

$$\lim_{m\to\infty}\mathbb{W}_2(\delta_0\mathbb{P}^x_{m\Delta_2}, \mu^x) = 0$$

and

$$\lim_{m\to\infty}\mathbb{W}_2(\delta_0\mathbb{P}^{x,\Delta_2}_{m\Delta_2}, \mu^{x,\Delta_2}) = 0.$$

The above inequalities together with Lemma 4.6 derive that

$$\begin{aligned}
\mathbb{W}_2(\mu^x, \mu^{x,\Delta_2}) &\leq \lim_{m\to\infty}\mathbb{W}_2(\mu^x, \delta_0\mathbb{P}^x_{m\Delta_2}) + \lim_{m\to\infty}\mathbb{W}_2(\delta_0\mathbb{P}^x_{m\Delta_2}, \delta_0\mathbb{P}^{x,\Delta_2}_{m\Delta_2}) + \lim_{m\to\infty}\mathbb{W}_2(\delta_0\mathbb{P}^{x,\Delta_2}_{m\Delta_2}, \mu^{x,\Delta_2})\\
&\leq \lim_{m\to\infty}\big(\mathbb{E}|y_n^{x,0}(m\Delta_2) - Y_{n,m}^{x,0}|^2\big)^{\frac{1}{2}} \leq C(1+|x|)\Delta^{\frac{1}{2}}.
\end{aligned}$$

□

Now we turn to analyze the property of the estimator $B_M(x, Y_n^{x,y_0})$.

Lemma 4.8. If (**S5**), (**F1**) and (**F3**) hold, then for any $2 \le q \le k$, $x \in \mathbb{R}^{n_1}$, $y_0 \in \mathbb{R}^{n_2}$, $\Delta_2 \in (0, \hat{\Delta}_2]$, integers $n \ge 0$ and $M \ge 1$,

$$\mathbb{E}\big|x^T B_M(x^*, Y_n^{x^*,y_0})\big|^{\frac{q}{2}} \le C(1 + |x|^q + |y_0|^q).$$

Proof. For any $x \in \mathbb{R}^{n_1}$ with $|x| \le \varphi^{-1}(K\Delta_1^{-1/2})$, $x = x^*$. Making use of (**S5**) and the elementary inequality yields that

$$\begin{aligned}\mathbb{E}\big|x^T B_M(x^*, Y_n^{x^*,y_0})\big|^{\frac{q}{2}} &= \mathbb{E}\Big|\frac{1}{M}\sum_{m=1}^{M} x^T b(x, Y_{n,m}^{x,y_0})\Big|^{\frac{q}{2}} \le \mathbb{E}\Big[K_4(1+|x|^2) + \frac{\lambda}{M}\sum_{m=1}^{M}|Y_{n,m}^{x,y_0}|^2\Big]^{\frac{q}{2}} \\ &\le C(1+|x|^q) + \frac{C}{M}\sum_{m=1}^{M}\mathbb{E}|Y_{n,m}^{x,y_0}|^q.\end{aligned}$$

Then by (**F1**) and (**F3**), applying Lemma 4.3 and the Hölder inequality implies that for any $\Delta_2 \in (0, \hat{\Delta}_2]$,

$$\mathbb{E}\big|x^T B_M(x^*, Y_n^{x^*,y_0})\big|^{\frac{q}{2}} \le C(1+|x|^q) + \frac{C}{M}\sum_{m=1}^{M}\big(\mathbb{E}|Y_{n,m}^{x,y_0}|^k\big)^{\frac{q}{k}} \le C(1+|x|^q+|y_0|^q). \tag{4.7}$$

On the other hand, for $x \in \mathbb{R}^{n_1}$ with $|x| > \varphi^{-1}(K\Delta_1^{-1/2})$, $x = |x|x^*/\varphi^{-1}(K\Delta_1^{-1/2})$. One observes that

$$x^T B_M(x^*, Y_n^{x^*,y_0}) = \frac{|x|}{\varphi^{-1}(K\Delta_1^{-\frac{1}{2}})}(x^*)^T B_M(x^*, Y_n^{x^*,y_0}).$$

Due to (4.7) and (3.2), we obtain that

$$\mathbb{E}\big|x^T B_M(x^*, Y_n^{x^*,y_0})\big|^{\frac{q}{2}} \le \frac{|x|^{\frac{q}{2}}}{\big(\varphi^{-1}\big(K\Delta_1^{-\frac{1}{2}}\big)\big)^{\frac{q}{2}}}C(1+|x^*|^q+|y_0|^q) \le C(1+|x|^q+|y_0|^q).$$

Thus the desired assertion follows. □

The error between $\bar{b}(x)$ and $B_M(x, Y_n^{x,y_0})$ is the key to obtain the convergence of the MTEM scheme numerical solution. By introducing an auxiliary function $\bar{b}^{\Delta_2}(\cdot)$ defined in the below we use $|\bar{b}(x) - \bar{b}^{\Delta_2}(x)|^2$ and $\mathbb{E}|\bar{b}^{\Delta_2}(x) - B_M(x, Y_n^{x,y_0})|^2$ to estimate $\mathbb{E}|\bar{b}(x) - B_M(x, Y_n^{x,y_0})|^2$. In fact, under (**S4**) and (**F1**)-(**F3**) with $k \ge \theta_4$, by virtue of Lemma 4.5, for any fixed $x \in \mathbb{R}^{n_1}$ and $\Delta_2 \in (0, \bar{\Delta}_2]$,

$$\begin{aligned}\int_{\mathbb{R}^{n_2}}|b(x,y)|\mu^{x,\Delta_2} &\le K_3\int_{\mathbb{R}^{n_2}}(1+|x|^{\theta_3}+|y|^{\theta_4})\mu^{x,\Delta_2}(\mathrm{d}y) \\ &\le C(1+|x|^{\theta_3\vee\theta_4}) < \infty,\end{aligned} \tag{4.8}$$

which implies that $b(x,\cdot)$ is integrable with respect to μ^{x,Δ_2}. Thus we define

$$\bar{b}^{\Delta_2}(x) = \int_{\mathbb{R}^{n_2}} b(x,y)\mu^{x,\Delta_2}(\mathrm{d}y). \tag{4.9}$$

Next we estimate $|\bar{b}(x) - \bar{b}^{\Delta_2}(x)|^2$ and $\mathbb{E}|\bar{b}^{\Delta_2}(x) - B_M(x, Y_n^{x,y_0})|^2$, respectively.

Lemma 4.9. Under (**S2**), (**S4**) and (**F1**)-(**F3**) with $k \geq 2 \vee 2\theta_2 \vee \theta_4$, for any $x \in \mathbb{R}^{n_1}$ and $\Delta_2 \in (0, \bar{\Delta}_2]$,

$$|\bar{b}(x) - \bar{b}^{\Delta_2}(x)| \leq C(1 + |x|^{\theta_2+1})\Delta_2.$$

Proof. Under (**S4**), (**F1**)-(**F3**) with $k \geq 2 \vee \theta_4$, in view of (1.3) and (4.9), using (**S2**) and the Hölder inequality yields that

$$\begin{aligned}
|\bar{b}(x) - \bar{b}^{\Delta_2}(x)| &= \Big| \int_{\mathbb{R}^{n_2} \times \mathbb{R}^{n_2}} \big(b(x, y_1) - b(x, y_2)\big)\pi(\mathrm{d}y_1, \mathrm{d}y_2)\Big| \\
&\leq \int_{\mathbb{R}^{n_2} \times \mathbb{R}^{n_2}} \big|b(x, y_1) - b(x, y_2)\big|\pi(\mathrm{d}y_1, \mathrm{d}y_2) \\
&\leq C\Big(\int_{\mathbb{R}^{n_2} \times \mathbb{R}^{n_2}} |y_1 - y_2|^2\pi(\mathrm{d}y_1, \mathrm{d}y_2)\Big)^{\frac{1}{2}} \\
&\quad \times \Big(\int_{\mathbb{R}^{n_2} \times \mathbb{R}^{n_2}} (1 + |x|^{2\theta_2} + |y_1|^{2\theta_2} + |y_2|^{2\theta_2})\pi(\mathrm{d}y_2, \mathrm{d}y_2)\Big)^{\frac{1}{2}},
\end{aligned}$$

where $\pi \in \mathcal{C}(\mu^x, \mu^{x,\Delta_2})$ is arbitrary. Thus, we derive that

$$|\bar{b}(x) - \bar{b}^{\Delta_2}(x)| \leq C\mathbb{W}_2(\mu^x, \mu^{x,\Delta_2}) \times \Big(1 + |x|^{2\theta_2} + \int_{\mathbb{R}^{n_2}} |y_1|^{2\theta_2}\mu^x(\mathrm{d}y_1) + \int_{\mathbb{R}^{n_2}} |y_2|^{2\theta_2}\mu^{x,\Delta_2}(\mathrm{d}y_2)\Big)^{\frac{1}{2}}.$$

Then due to (**F1**)-(**F3**) with $k \geq 2 \vee 2\theta_2$, applying Lemmas 2.1, 4.5 and 4.7 implies that

$$|\bar{b}(x) - \bar{b}^{\Delta_2}(x)| \leq C(1 + |x|^{\theta_2+1})\Delta_2^{\frac{1}{2}}.$$

The proof is complete. □

Before the estimation of $\mathbb{E}|\bar{b}^{\Delta_2}(x) - B_M(x, Y_n^{x,y_0})|^2$, we prepare a useful result.

Lemma 4.10. Under (**S2**), (**S4**) and (**F1**)-(**F3**) with $k \geq 2 \vee 2\theta_2 \vee \theta_4$, for any $x \in \mathbb{R}^{n_1}$, $y \in \mathbb{R}^{n_2}$, $\Delta_2 \in (0, \bar{\Delta}_2]$ and integers $n \geq 0$, $M \geq 1$,

$$|\bar{b}^{\Delta_2}(x) - \mathbb{E}b(x, Y_{n,m}^{x,y})| \leq C(1 + |x|^{\theta_2+1} + |y|^{\theta_2+1})e^{\frac{-\beta m\Delta_2}{8}}.$$

Proof. Under (**S4**) and (**F1**)-(**F3**) with $k \geq 2 \vee \theta_4$, according to (4.9) and the invariance of invariant measure μ^{x,Δ_2}, we have

$$\begin{aligned}
\bar{b}^{\Delta_2}(x) &= \lim_{k\to\infty} \int_{\mathbb{R}^{n_2}} b(x, z)I_{\{|z|\leq k\}}\mu^{x,\Delta_2}(\mathrm{d}z) \\
&\leq \lim_{k\to\infty} \int_{\mathbb{R}^{n_2}} \mathbb{E}\Big(b(x, Y_{n,m}^{x,z})I_{\{|Y_{n,m}^{x,z}|\leq k\}}\Big)\mu^{x,\Delta_2}(\mathrm{d}z).
\end{aligned} \tag{4.10}$$

Obviously, $\lim_{k\to\infty} b(x, Y_{n,m}^{x,z})I_{\{|Y_{n,m}^{x,z}|\leq k\}} = b(x, Y_{n,m}^{x,z})$, a.s. for any $z \in \mathbb{R}^{n_2}$. In addition, by (**S4**) and (**F1**)-(**F3**) with $k \geq 2 \vee \theta_4$, using Lemma 4.5 yields that

$$\begin{aligned}
\int_{\mathbb{R}^{n_2}} \mathbb{E}|b(x, Y_{n,m}^{x,z})|\mu^{x,\Delta_2}(\mathrm{d}z) &\leq C\Big(1 + |x|^{\theta_3} + \int_{\mathbb{R}^{n_2}} \mathbb{E}|Y_{n,m}^{x,z}|^{\theta_4}\mu^{x,\Delta_2}(\mathrm{d}z)\Big) \\
&\leq C\Big(1 + |x|^{\theta_3\vee\theta_4} + \int_{\mathbb{R}^{n_2}} |z|^{\theta_4}\mu^{x,\Delta_2}(\mathrm{d}z)\Big) \\
&\leq C(1 + |x|^{\theta_3\vee\theta_4}) < \infty.
\end{aligned}$$

Then applying the dominated convergence theorem for (4.10) we derive that

$$\bar{b}^{\Delta_2}(x) = \int_{\mathbb{R}^{n_2}} \mathbb{E}b(x, Y_{n,m}^{x,z})\mu^{x,\Delta_2}(\mathrm{d}z).$$

As a result, we have

$$\begin{aligned}|\bar{b}^{\Delta_2}(x) - \mathbb{E}b(x, Y_{n,m}^{x,y})| &= \Big|\mathbb{E}b(x, Y_{n,m}^{x,y}) - \int_{\mathbb{R}^{n_2}} \mathbb{E}b(x, Y_{n,m}^{x,z})\mu^{x,\Delta_2}(\mathrm{d}z)\Big| \\ &\leq \int_{\mathbb{R}^{n_2}} \mathbb{E}\big|b(x, Y_{n,m}^{x,y}) - b(x, Y_{n,m}^{x,z})\big|\mu^{x,\Delta_2}(\mathrm{d}z).\end{aligned}$$

Further using (**S2**) and the Hölder inequality gives that

$$\begin{aligned}&|\bar{b}^{\Delta_2}(x) - \mathbb{E}b(x, Y_{n,m}^{x,y})| \\ \leq& K_1 \int_{\mathbb{R}^{n_2}} \mathbb{E}\Big(|Y_{n,m}^{x,y} - Y_{n,m}^{x,z}|(1 + |x|^{\theta_2} + |Y_{n,m}^{x,y}|^{\theta_2} + |Y_{n,m}^{x,z}|^{\theta_2})\Big)\mu^{x,\Delta_2}(\mathrm{d}z) \\ \leq& C \int_{\mathbb{R}^{n_2}} \Big[\big(\mathbb{E}|Y_{n,m}^{x,y} - Y_{n,m}^{x,z}|^2\big)^{\frac{1}{2}}\big(\mathbb{E}(1 + |x|^{2\theta_2} + |Y_{n,m}^{x,y}|^{2\theta_2} + |Y_{n,m}^{x,z}|^{2\theta_2})\big)^{\frac{1}{2}}\Big]\mu^{x,\Delta_2}(\mathrm{d}z).\end{aligned}$$

Under (**F1**)-(**F3**) with $k \geq 2 \vee 2\theta_2$, utilizing Lemmas 4.3 and 4.4 we get

$$\begin{aligned}|\bar{b}^{\Delta_2}(x) - \mathbb{E}b(x, Y_{n,m}^{x,y})| &\leq Ce^{\frac{-\beta m\Delta_2}{8}} \int_{\mathbb{R}^{n_2}} |y - z|(1 + |x|^{\theta_2} + |y|^{\theta_2} + |z|^{\theta_2})\mu^{x,\Delta_2}(\mathrm{d}z) \\ &\leq Ce^{\frac{-\beta m\Delta_2}{8}} \int_{\mathbb{R}^{n_2}} (1 + |x|^{\theta_2+1} + |y|^{\theta_2+1} + |z|^{\theta_2+1})\mu^{x,\Delta_2}(\mathrm{d}z) \\ &\leq C(1 + |x|^{\theta_2+1} + |y|^{\theta_2+1})e^{\frac{-\beta m\Delta_2}{8}}.\end{aligned}$$

The proof is complete. □

Lemma 4.11. Under (**S2**), (**S4**) and (**F1**)-(**F3**) with $k \geq 2\theta_2 \vee 2\theta_4 \vee (\theta_2 + \theta_4 + 1)$, for any $x \in \mathbb{R}^{n_1}$, $y_0 \in \mathbb{R}^{n_2}$, $\Delta_2 \in (0, \bar{\Delta}_2]$ and integers $n \geq 0$, $M \geq 1$,

$$\mathbb{E}|\bar{b}^{\Delta_2}(x) - B_M(x, Y_n^{x,y_0})|^2 \leq C(1 + |x|^{2\theta_3\vee 2\theta_4\vee(\theta_2+\theta_3\vee\theta_4+1)} + |y_0|^{2\theta_3\vee 2\theta_4\vee(\theta_2+\theta_3\vee\theta_4+1)})\frac{1}{M\Delta_2}.$$

Proof. In light of (3.5), we derive that for any $x \in \mathbb{R}^{n_1}$,

$$\mathbb{E}\Big|\bar{b}^{\Delta_2}(x) - B_M(x, Y_n^{x,y_0})\Big|^2 = \frac{1}{M^2}\sum_{m,l=1}^{M} \mathbb{E}U_{m,l} = \frac{1}{M^2}\sum_{m=1}^{M} \mathbb{E}U_{m,m} + \frac{2}{M^2}\sum_{l=1}^{M}\sum_{m=l+1}^{M} \mathbb{E}U_{m,l}, \quad (4.11)$$

where

$$U_{m,l} = \Big(\bar{b}^{\Delta_2}(x) - b\big(x, Y_{n,m}^{x,y_0}\big)\Big)\Big(\bar{b}^{\Delta_2}(x) - b\big(x, Y_{n,l}^{x,y_0}\big)\Big).$$

By (**S4**), (**F1**) and (**F3**) with $k \geq 2\theta_4$, invoking Lemma 4.3 and the Hölder inequality, we obtain that

$$\begin{aligned}\mathbb{E}\big|b(x, Y_{n,m}^{x,y_0})\big|^2 &\leq C\mathbb{E}\Big(1 + |x|^{2\theta_3} + |Y_{n,m}^{x,y_0}|^{2\theta_4}\Big) \leq C(1 + |x|^{2\theta_3}) + C\Big(\mathbb{E}|Y_{n,m}^{x,y_0}|^k\Big)^{\frac{2\theta_4}{k}} \\ &\leq C(1 + |x|^{2(\theta_3\vee\theta_4)} + |y_0|^{2(\theta_3\vee\theta_4)}).\end{aligned}$$

Then using the elementary inequality along with the above inequality and (4.8), for any $m, l \geq 1$, we yield that for any $x \in \mathbb{R}^{n_1}$,

$$\begin{aligned}\mathbb{E}|U_{m,l}| &\leq \mathbb{E}\big|b(x, Y_{n,m}^{x,y_0})\big|^2 + \mathbb{E}\big|b(x, Y_{n,l}^{x,y_0})\big|^2 + 2\mathbb{E}|\bar{b}^{\Delta_2}(x)|^2 \\ &\leq C(1 + |x|^{2(\theta_3 \vee \theta_4)} + |y_0|^{2(\theta_3 \vee \theta_4)}) < \infty,\end{aligned} \tag{4.12}$$

which implies that $|U_{m,l}|$ is integrable with respect to $\mathbb{P}$. To compute precisely, let $\mathcal{G}_{n,l}^2$ denote the σ-algebra generated by

$$\Big\{W_n^2(s) - W_n^2(l\Delta_2), s \geq l\Delta_2\Big\}.$$

Obviously, $\mathcal{F}_{n,l}^2$ and $\mathcal{G}_{n,l}^2$ are mutually independent. Since $Y_{n,l}^{x,y_0}$ is $\mathcal{F}_{n,l}^2$-measurable and independent of $\mathcal{G}_{n,l}^2$, using the result of [42, p.221], we derive that for any $x \in \mathbb{R}^{n_1}$ and $1 \leq l < m \leq M$,

$$\begin{aligned}\mathbb{E}U_{m,l} =& \mathbb{E}\Big[\big(\bar{b}^{\Delta_2}(x) - b(x, Y_{n,l}^{x,y_0})\big) \times \mathbb{E}\Big(\big(\bar{b}^{\Delta_2}(x) - b(x, Y_{n,m}^{x,y_0})\big)\Big|\mathcal{F}_{n,l}^2\Big)\Big] \\ \leq& \mathbb{E}\bigg[\big|\bar{b}^{\Delta_2}(x) - b(x, Y_{n,l}^{x,y_0})\big| \times \Big|\bar{b}^{\Delta_2}(x) - \mathbb{E}b\big(x, Y_{n,m-l}^{x,z}\big)\Big|_{z=Y_{n,l}^{x,y_0}}\bigg].\end{aligned} \tag{4.13}$$

For any $x \in \mathbb{R}^{n_1}$ and $y \in \mathbb{R}^{n_2}$, it follows from (**S4**) and (4.8) that

$$|\bar{b}^{\Delta_2}(x) - b(x, y)| = |\bar{b}^{\Delta_2}(x)| + |b(x, y)| \leq C(1 + |x|^{\theta_3 \vee \theta_4} + |y|^{\theta_4}). \tag{4.14}$$

Owing to (**S2**), (**S4**) and (**F1**)-(**F3**) with $k \geq 2 \vee 2\theta_2 \vee \theta_4$, using Lemma 4.10 derives that

$$\Big|\bar{b}^{\Delta_2}(x) - \mathbb{E}b\big(x, Y_{n,m-l}^{x,z}\big)\Big| \leq Ce^{-\frac{\beta(m-l)\Delta_2}{8}}\big(1 + |x|^{\theta_2+1} + |z|^{\theta_2+1}\big).$$

Using (4.14) and substituting the above inequality into (4.13) lead to that for any $x \in \mathbb{R}^{n_1}$,

$$\begin{aligned}\mathbb{E}U_{m,l} \leq& Ce^{-\frac{\beta(m-l)\Delta_2}{8}}\mathbb{E}\Big[\big(1 + |x|^{\theta_3 \vee \theta_4} + |Y_{n,l}^{x,y_0}|^{\theta_4}\big) \\ &\qquad \times \big(1 + |x|^{\theta_2+1} + |Y_{n,l}^{x,y_0}|^{\theta_2+1}\big)\Big] \\ \leq& Ce^{-\frac{\beta(m-l)\Delta_2}{8}}\mathbb{E}\Big[\big(1 + |x|^{\theta_2+\theta_3\vee\theta_4+1} + (1 + |x|^{\theta_3 \vee \theta_4})|Y_{n,l}^{x,y_0}|^{\theta_2+1} \\ &+ (1 + |x|^{\theta_2+1})|Y_{n,l}^{x,y_0}|^{\theta_4} + |Y_{n,l}^{x,y_0}|^{\theta_2+\theta_4+1}\big)\Big].\end{aligned}$$

Due to $k \geq \theta_2 + \theta_4 + 1$, using Lemma 4.3 we deduce that for any $1 \leq l < m \leq M$,

$$\mathbb{E}U_{m,l} \leq Ce^{-\frac{\beta(m-l)\Delta_2}{8}}\big(1 + |x|^{\theta_2+\theta_3\vee\theta_4+1} + |y_0|^{\theta_2+\theta_3\vee\theta_4+1}\big). \tag{4.15}$$

Hence, inserting (4.12) and (4.15) into (4.11) yields that

$$\begin{aligned}\mathbb{E}\Big|\bar{b}^{\Delta_2}(x) - B_M(x, Y_n^{x,y_0})\Big|^2 \leq& \frac{C(1 + |x|^{2(\theta_3\vee\theta_4)} + |y_0|^{2(\theta_3\vee\theta_4)})}{M} \\ &+ \frac{C(1 + |x|^{\theta_2+\theta_3\vee\theta_4+1} + |y_0|^{\theta_2+\theta_3\vee\theta_4+1})}{M^2}\sum_{l=1}^{M}\sum_{m=l+1}^{M} e^{-\frac{\beta(m-l)\Delta_2}{8}} \\ \leq& \frac{C(1 + |x|^{2(\theta_3\vee\theta_4)} + |y_0|^{2(\theta_3\vee\theta_4)})}{M} + \frac{C(1 + |x|^{\theta_2+\theta_3\vee\theta_4+1} + |y_0|^{\theta_2+\theta_3\vee\theta_4+1})}{M(e^{\beta\Delta_2/8} - 1)} \\ \leq& C\big(1 + |x|^{2\theta_3\vee2\theta_4\vee(\theta_2+\theta_3\vee\theta_4+1)} + |y_0|^{2\theta_3\vee2\theta_4\vee(\theta_2+\theta_3\vee\theta_4+1)}\big)\Big(\frac{1}{M} + \frac{1}{M\Delta_2}\Big) \\ \leq& C\big(1 + |x|^{2\theta_3\vee2\theta_4\vee(\theta_2+\theta_3\vee\theta_4+1)} + |y_0|^{2\theta_3\vee2\theta_4\vee(\theta_2+\theta_3\vee\theta_4+1)}\big)\frac{1}{M\Delta_2},\end{aligned}$$

where the second to last inequality used the fact $e^x - 1 \geq x, \forall x \geq 0$. The proof is complete. $\square$

Combining Lemmas 4.9 and 4.11, we obtain the estimate of $\mathbb{E}|\bar{b}(x) - B_M(x, Y_n^x)|^2$ directly.

Lemma 4.12. Under (**S2**), (**S4**) and (**F1**)-(**F3**) with $k \geq 2\theta_2 \vee 2\theta_4 \vee (\theta_2 + \theta_4 + 1)$, for any $x \in \mathbb{R}^{n_1}$, $y_0 \in \mathbb{R}^{n_2}$, $\Delta_2 \in (0, \bar{\Delta}_2]$ and integers $n \geq 0$, $M \geq 1$,

$$\mathbb{E}\Big|\bar{b}(x) - B_M(x, Y_n^{x,y_0})\Big|^2 \leq C\big(1 + |x|^{2(\theta_2+1)\vee 2\theta_3 \vee 2\theta_4} + |y_0|^{2(\theta_2+1)\vee 2\theta_3\vee 2\theta_4}\big)\Big(\Delta_2 + \frac{1}{M\Delta_2}\Big).$$

5 Strong convergence in pth moment

With the help of the averaging principle, this section aims to prove the strong convergence between the slow component $x^\varepsilon(t)$ of original system (1.1) and the MTEM scheme numerical solution $X(t)$ in pth moment.

Lemma 5.1. If (**S3**)-(**S5**), (**F1**) and (**F3**) hold with $k \geq 2\theta_4$, then for any $x_0 \in \mathbb{R}^{n_1}$, $y_0 \in \mathbb{R}^{n_2}$, $0 < p \leq k/\theta_4$, $T > 0$ and $M \geq 1$,

$$\sup_{\Delta_1 \in (0,1], \Delta_2 \in (0, \hat{\Delta}_2]} \mathbb{E}\Big(\sup_{t\in[0,T]} |\bar{X}(t)|^p\Big) \leq C_{x_0,y_0,T,p},$$

and

$$\mathbb{E}\Big(\sup_{0\leq t\leq T} |\bar{X}(t) - X(t)|^p\Big) \leq C_{x_0,y_0,T,p}\Delta_1^{\frac{p}{2}}.$$

Proof. For $2 \leq p \leq k/\theta_4$, using the Itô formula, we deduce from (3.8) that for any $t \geq 0$,

$$\begin{aligned} |\bar{X}(t)|^p \leq & |x_0|^p + p\int_0^t |\bar{X}(s)|^{p-2}\Big[\bar{X}^T(s)B\Big(X^*(s), Y^{X^*(s),y_0}\Big) \\ & + \frac{p-1}{2}|\sigma(X(s))|^2\Big]\mathrm{d}s + p\int_0^t |\bar{X}(s)|^{p-2}\bar{X}^T(s)\sigma(X(s))\mathrm{d}W^1(s), \end{aligned}$$

where we write $B_M(\cdot,\cdot)$ as $B(\cdot,\cdot)$ for short. Utilizing the Burkholder-Davis-Gundy inequality [33, p.40, Theorem7.2], the Young inequality and the Hölder inequality implies that for any

$T > 0$,

$$
\begin{aligned}
\mathbb{E}\Big(\sup_{t\in[0,T]}|\bar{X}(t)|^p\Big) &\le |x_0|^p + p\mathbb{E}\int_0^T |\bar{X}(t)|^{p-2}\Big[\Big|\bar{X}^T(t)B\Big(X^*(t),Y^{X^*(t),y_0}\Big)\Big| + \frac{p-1}{2}|\sigma(X(t))|^2\Big]\mathrm{d}t \\
&\quad + 4\sqrt{2}p\mathbb{E}\Big(\int_0^T |\bar{X}(t)|^{2p-2}|\sigma(X(t))|^2\mathrm{d}t\Big)^{\frac{1}{2}} \\
&\le |x_0|^p + C\int_0^T \mathbb{E}|\bar{X}(t)|^p\mathrm{d}t + C\int_0^T \mathbb{E}\Big|\bar{X}^T(t)B\Big(X^*(t),Y^{X^*(t),y_0}\Big)\Big|^{\frac{p}{2}}\mathrm{d}t \\
&\quad + C\int_0^T \mathbb{E}|\sigma(X(t))|^p\mathrm{d}t + 4\sqrt{2}p\mathbb{E}\Big[\sup_{t\in[0,T]}|\bar{X}(t)|^{p-1}\Big(\int_0^T \big|\sigma(X(t))\big|^2\mathrm{d}t\Big)^{\frac{1}{2}}\Big] \\
&\le |x_0|^p + C\int_0^T \mathbb{E}|\bar{X}(t)|^p\mathrm{d}t + C\int_0^T \mathbb{E}\Big|X^T(t)B\Big(X^*(t),Y^{X^*(t),y_0}\Big)\Big|^{\frac{p}{2}}\mathrm{d}t \\
&\quad + C\int_0^T \mathbb{E}\Big|(\bar{X}(t)-X(t))^TB\Big(X^*(t),Y^{X^*(t),y_0}\Big)\Big|^{\frac{p}{2}}\mathrm{d}t \\
&\quad + C\int_0^T \mathbb{E}|\sigma(X(t))|^p\mathrm{d}t + \frac{1}{2}\mathbb{E}\Big(\sup_{t\in[0,T]}|\bar{X}(t)|^p\Big).
\end{aligned}
$$

Then it follows from (**S3**) that

$$
\begin{aligned}
\mathbb{E}\Big(\sup_{t\in[0,T]}|\bar{X}(t)|^p\Big) &\le |x_0|^p + C\int_0^T \mathbb{E}\Big(\sup_{0\le s\le t}|\bar{X}(s)|^p\Big)\mathrm{d}t + C\int_0^T \mathbb{E}\Big|X^T(t)B\Big(X^*(t),Y^{X^*(t),y_0}\Big)\Big|^{\frac{p}{2}}\mathrm{d}t \\
&\quad + C\int_0^T \mathbb{E}\Big|(\bar{X}(t)-X(t))^TB\Big(X^*(t),Y^{X^*(t),y_0}\Big)\Big|^{\frac{p}{2}}\mathrm{d}t. \qquad (5.1)
\end{aligned}
$$

For any $t\ge 0$, one observes that $Y^{X^*(t),y_0} = Y^{X^*_{n_{\Delta_1}(t)},y_0} = Y^{X^*_{n_{\Delta_1}(t)},y_0}_{n_1(t)}$. Due to the independence of $Y^{x,y_0}_{n_{\Delta_1}(t)}$ and $X_{n_{\Delta_1}(t)}$, for any $t\ge 0$ and $2\le p\le k/\theta_4$, under (**S5**), (**F1**) and (**F3**), we obtain from (4.5) and the result of Lemma 4.8 that

$$
\begin{aligned}
&\mathbb{E}\Big|X^T(t)B\Big(X^*(t),Y^{X^*(t),y_0}\Big)\Big|^{\frac{p}{2}} = \mathbb{E}\Big|X^T_{n_{\Delta_1}(t)}B\Big(X^*_{n_{\Delta_1}(t)},Y^{X^*_{n_{\Delta_1}(t)},y_0}\Big)\Big|^{\frac{p}{2}} \\
=&\mathbb{E}\Big[\mathbb{E}\Big(\Big|X^T_{n_{\Delta_1}(t)}B\Big(X^*_{n_{\Delta_1}(t)},Y^{X^*_{n_{\Delta_1}(t)},y_0}_{n_{\Delta_1}(t)}\Big)\Big|^{\frac{p}{2}}\Big|X_{n_{\Delta_1}(t)}\Big)\Big] \le \mathbb{E}\big(\mathbb{E}\big|x^TB\big(x^*,Y^{x^*,y_0}_{n_{\Delta_1}(t)}\big)\big|^{\frac{p}{2}}\big|_{x=X_{n_{\Delta_1}(t)}}\big) \\
\le& C(1+|y_0|^p+\mathbb{E}|X_{n_{\Delta_1}(t)}|^p) \le C(1+|y_0|^p) + C\mathbb{E}\Big(\sup_{0\le s\le t}|\bar{X}(s)|^p\Big). \qquad (5.2)
\end{aligned}
$$

Under (**S4**), we derive from (3.3) and (4.5) that

$$
\begin{aligned}
\mathbb{E}\Big|B\Big(X^*(t),Y^{X^*(t),y_0}\Big)\Big|^p &= \mathbb{E}\Big|B\Big(X^*_{n_{\Delta_1}(t)},Y^{X^*_{n_{\Delta_1}(t)},y_0}_{n_{\Delta_1}}\Big)\Big|^p \le \frac{1}{M}\sum_{m=1}^M \mathbb{E}\Big|b\big(X^*_{n_{\Delta_1}(t)},Y^{X^*_{n_{\Delta_1}(t)},y_0}_{n_{\Delta_1}(t),m}\big)\Big|^p \\
&\le \frac{1}{M}\sum_{m=1}^M \mathbb{E}\Big[\Big(C\Delta_1^{-\frac{1}{2}}(1+|X^*_{n_{\Delta_1}(t)}|) + K_3\Big|Y^{X^*_{n_{\Delta_1}(t)},y_0}_{n_{\Delta_1}(t),m}\Big|^{\theta_4}\Big)\Big]^p \\
&\le C\Delta_1^{-\frac{p}{2}}\mathbb{E}\big(1+|X^*_{n_{\Delta_1}(t)}|\big)^p + \frac{C}{M}\sum_{m=1}^M \mathbb{E}\Big|Y^{X^*_{n_{\Delta_1}(t)},y_0}_{n_{\Delta_1}(t),m}\Big|^{p\theta_4}. \qquad (5.3)
\end{aligned}
$$

Owing to (**F1**) and (**F3**) with $p\theta_4\le k$, using the Young inequality and Lemma 4.3 yields that

$$
\begin{aligned}
\mathbb{E}\Big|Y^{X^*_{n_{\Delta_1}(t)},y_0}_{n_{\Delta_1}(t),m}\Big|^{p\theta_4} &= \mathbb{E}\Big(\mathbb{E}\Big(\Big|Y^{X^*_{n_{\Delta_1}(t)},y_0}_{n_{\Delta_1}(t),m}\Big|^{p\theta_4}\big|X^*_{n_{\Delta_1}(t)}\Big)\Big) = \mathbb{E}\Big(\mathbb{E}|Y^{x,y_0}_{n_{\Delta_1}(t),m}|^{p\theta_4}|_{x=X^*_{n_{\Delta_1}(t)}}\Big) \\
&\le C(1+|y_0|^{p\theta_4}+\mathbb{E}|X^*_{n_{\Delta_1}(t)}|^{p\theta_4}).
\end{aligned}
$$

Thanks to (3.2) , we derive that

$$\mathbb{E}\Big|Y_{n_{\Delta_1}(t),m}^{X^*_{n_{\Delta_1}(t)},y_0}\Big|^{p\theta_4} \le C\Delta_1^{-\frac{p}{2}}.$$

Inserting the above inequality into (5.3) implies that

$$\mathbb{E}\Big|B\Big(X^*(t),Y^{X^*(t),y_0}\Big)\Big|^p \le C\Delta_1^{-\frac{p}{2}}\mathbb{E}\big(1+|X^*_{n_{\Delta_1}(t)}|\big)^p \le C\Delta_1^{-\frac{p}{2}}\mathbb{E}\big(1+|X(t)|\big)^p. \tag{5.4}$$

This, together with (3.8), implies that for any $t \ge 0$,

$$\begin{aligned}
\mathbb{E}|\bar{X}(t)-X(t)|^p \le& 2^{p-1}\Big(\mathbb{E}\Big|\int_{n_{\Delta_1}(t)\Delta_1}^t B\Big(X^*(s),Y^{X^*(s),y_0}\Big)\mathrm{d}s\Big|^p + \mathbb{E}\Big|\int_{n_{\Delta_1}(t)\Delta_1}^t \sigma(X(s))\mathrm{d}W^1(s)\Big|^p\Big)\\
\le& C\Big(\Delta_1^{p-1}\int_{n_{\Delta_1}(t)\Delta_1}^t \mathbb{E}\Big|B\Big(X^*(s),Y^{X^*(s),y_0}\Big)\Big|^p\mathrm{d}s + \Delta_1^{\frac{p-2}{2}}\int_{n_{\Delta_1}(t)\Delta_1}^t \mathbb{E}|\sigma(X(s))|^p\mathrm{d}s\Big)\\
\le& C\Delta_1^{\frac{p}{2}}\mathbb{E}\big(1+|X(t)|\big)^p.
\end{aligned} \tag{5.5}$$

Invoking the Hölder inequality, (5.4) and (5.5) we obtain

$$\begin{aligned}
\mathbb{E}\Big(\big|\bar{X}(t)-X(t)\big|^{\frac{p}{2}}\Big|B\Big(X^*(t),Y^{X^*(t),y_0}\Big)\Big|^{\frac{p}{2}}\Big) \le& \Big(\mathbb{E}\big|\bar{X}(t)-X(t)\big|^p\Big)^{\frac{1}{2}}\Big(\mathbb{E}\Big|B\Big(X^*(t),Y^{X^*(t),y_0}\Big)\Big|^p\Big)^{\frac{1}{2}}\\
\le& C\mathbb{E}\big(1+|X(t)|\big)^p \le C_p + C_p\mathbb{E}\Big(\sup_{0\le s\le t}|\bar{X}(s)|^p\Big).
\end{aligned} \tag{5.6}$$

Inserting (5.2) and (5.6) into (5.1) yields that

$$\mathbb{E}\Big(\sup_{t\in[0,T]}|\bar{X}(t)|^p\Big) \le |x_0|^p + C_p(1+|y_0|^p) + C_p\int_0^T \mathbb{E}\Big(\sup_{0\le s\le t}|\bar{X}(s)|^p\Big)\mathrm{d}t.$$

A direct application of Gronwall's inequality derives that

$$\mathbb{E}\Big(\sup_{t\in[0,T]}|\bar{X}(t)|^p\Big) \le C_{x_0,y_0,T,p}. \tag{5.7}$$

Then the second assertion holds directly by substituting (5.7) into (5.5). The case $0 < p < 2$ follows directly by using the Hölder inequality. The proof is complete. □

Remark 5.1. For any $R > |x_0|$, define the stopping time

$$\bar{\rho}_{\Delta_1,R} = \inf\{t \ge 0 : |\bar{X}(t)| \ge R\}. \tag{5.8}$$

It follows from Lemma 5.1 that for any $T > 0$,

$$\mathbb{P}\big(\bar{\rho}_{\Delta_1,R} \le T\big) \le \frac{C_{x_0,y_0,T,p}}{R^p}.$$

To prove the strong convergence of the MTEM scheme (3.6), we introduce an auxiliary TEM numerical scheme for the averaged equation (1.4)

$$\begin{cases} Z_0 = x_0,\\ Z_{n+1} = Z_n + \bar{b}(Z_n^*)\Delta_1 + \sigma(Z_n)\Delta W_n^1, \end{cases} \tag{5.9}$$

and the corresponding continuous-time processes

$$Z(t) = Z_n, \qquad t \in [n\Delta_1, (n+1)\Delta_1),$$

and

$$\bar{Z}(t) = x_0 + \int_0^t \bar{b}(Z^*(s))\mathrm{d}s + \int_0^t \sigma(Z(s))\mathrm{d}W^1(s). \tag{5.10}$$

One observes that $\bar{Z}(n\Delta_1) = Z(n\Delta_1) = Z_n$. In what follows, we analyze the strong error $\mathbb{E}\Big(\sup_{0\le t\le T}|\bar{x}(t) - \bar{Z}(t)|^2\Big)$ and $\mathbb{E}\Big(\sup_{0\le t\le T}|\bar{Z}(t) - X(t)|^2\Big)$, respectively. To proceed we give the bound of the pth moment of $\bar{Z}(t)$.

Lemma 5.2. If (**S3**)-(**S5**) and (**F1**)-(**F3**) hold with $k \ge 2 \vee \theta_4$, then for any $x_0 \in \mathbb{R}^{n_1}$, $p > 0$ and $T > 0$,

$$\sup_{\Delta_1\in(0,1]} \mathbb{E}\Big(\sup_{0\le t\le T}|\bar{Z}(t)|^p\Big) \le C_{x_0,T,p},$$

and

$$\sup_{0\le t\le T} \mathbb{E}|\bar{Z}(t) - Z(t)|^2 \le C_{x_0,T,p}\Delta_1.$$

Proof. The case that $0 < p < 2$ follows directly from the case $p \ge 2$ by using Lyapunov's inequality. Thus we are only going to deal with the case $p \ge 2$. Applying the Itô formula and Burkholder-Davis-Gundy inequality [33, p.40, Theorem7.2], under (**S5**) and (**F1**)-(**F3**), we derive from the result of Lemma 4.2 that for $p \ge 2$ and $T > 0$,

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le t\le T}|\bar{Z}(t)|^p\Big) \le& |x_0|^p + \frac{pC}{2}\mathbb{E}\int_0^T \big|\bar{Z}(t)\big|^{p-2}\big(1 + |Z(t)|^2\big)\mathrm{d}t\\ &+ p\mathbb{E}\int_0^T |\bar{Z}(t)|^{p-2}|\bar{Z}(t) - Z(t)||\bar{b}(Z^*(t)|\mathrm{d}t\\ &+ 4\sqrt{2}p\mathbb{E}\Big(\int_0^T \big|\bar{Z}(t)\big|^{2p-2}\big|\sigma(Z(t))|^2\mathrm{d}t\Big)^{\frac{1}{2}}.\end{aligned}$$

Then by the Young inequality we obtain that for any $T > 0$,

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le t\le T}|\bar{Z}(t)|^p\Big) \le& |x_0|^p + C\int_0^T \mathbb{E}\Big(\sup_{0\le s\le t}|\bar{Z}(s)|^p\Big)\mathrm{d}t + C\int_0^T \mathbb{E}\Big(|\bar{Z}(t) - Z(t)|^{\frac{p}{2}}|\bar{b}(Z^*(t))|^{\frac{p}{2}}\Big)\mathrm{d}t\\ &+ \frac{1}{2}\mathbb{E}\Big(\sup_{0\le t\le T}|\bar{Z}(t)|^p\Big) + C\mathbb{E}\Big(\int_0^t |\sigma(Z(t))|^2\mathrm{d}t\Big)^{\frac{p}{2}}.\end{aligned} \tag{5.11}$$

For any $t \ge 0$, due to (**S4**), (**F1**)-(**F3**) with $k \ge \theta_4$, (3.4) hold. Then using (3.4) and (**S3**) yields that

$$\begin{aligned}&\mathbb{E}\big|\bar{Z}(t) - Z(t)\big|^p = \mathbb{E}\big|\bar{Z}(t) - Z_{n_{\Delta_1}(t)}\big|^p\\ \le& 2^{p-1}\Big(\mathbb{E}\Big|\int_{n_{\Delta_1}(t)\Delta_1}^t \bar{b}(Z^*(s))\mathrm{d}s\Big|^p + \mathbb{E}\Big|\int_{n_{\Delta_1}(t)\Delta_1}^t \sigma(Z(s))\mathrm{d}W^1(s)\Big|^p\Big)\\ \le& 2^{p-1}\Big(\Delta_1^p\mathbb{E}|\bar{b}(Z^*_{n_{\Delta_1}(t)})|^p + \Delta_1^{\frac{p}{2}}\mathbb{E}|\sigma(Z_{n_{\Delta_1}(t)})|^p\Big)\\ \le& C\Delta_1^{\frac{p}{2}}\big(1 + \mathbb{E}|Z_{n_{\Delta_1}(t)}|^p\big) \le C\Delta_1^{\frac{p}{2}}(1 + \mathbb{E}|Z(t)|^p).\end{aligned} \tag{5.12}$$

Then utilizing (3.4) again and the Hölder inequality implies that

$$\begin{aligned}\mathbb{E}\Big(|\bar{Z}(t)-Z(t)|^{\frac{p}{2}}|\bar{b}(Z^*(t))|^{\frac{p}{2}}\Big) &\le \big(\mathbb{E}|\bar{Z}(t)-Z(t)|^p\big)^{\frac{1}{2}}\big(\mathbb{E}|\bar{b}(Z(t))|^p\big)^{\frac{1}{2}}\\ &\le C\Delta_1^{\frac{p}{4}}\big(1+\mathbb{E}|Z(t)|^p\big)^{\frac{1}{2}}\Delta_1^{-\frac{p}{4}}\big(1+\mathbb{E}|Z(t)|^p\big)^{\frac{1}{2}}\\ &\le C+\mathbb{E}\Big(\sup_{0\le s\le t}|\bar{Z}(s)|^p\Big). \end{aligned} \tag{5.13}$$

Applying (**S3**) and the Hölder inequality we get

$$\mathbb{E}\Big(\int_0^T\big|\sigma(Z(t))|^2\mathrm{d}t\Big)^{\frac{p}{2}} \le C_{T,p}+C_{T,p}\int_0^T\mathbb{E}\Big(\sup_{0\le s\le t}|\bar{Z}(s)|^p\Big)\mathrm{d}t. \tag{5.14}$$

Hence, substituting (5.13) and (5.14) into (5.11) yields that

$$\mathbb{E}\Big(\sup_{0\le t\le T}|\bar{Z}(t)|^p\Big)\le |x_0|^p+C_{T,p}+C_{T,p}\int_0^T\mathbb{E}\Big(\sup_{0\le s\le t}|\bar{Z}(s)|^p\Big)\mathrm{d}t.$$

An application of the Gronwall inequality gives that

$$\mathbb{E}\Big(\sup_{0\le t\le T}|\bar{Z}(t)|^p\Big)\le C_{x_0,T,p}.$$

Then inserting the above inequality into (5.12) implies that the another desired assertion holds. The proof is complete. □

Remark 5.2. From Lemma 5.2, for any constant $R>|x_0|$, define a stopping time

$$\rho_{\Delta_1,R}:=\inf\{t\ge 0: |\bar{Z}(t)|\ge R\}. \tag{5.15}$$

By a similar argument as Remark 2.2, for any $T>0$, we have

$$\mathbb{P}(\rho_{\Delta_1,R}\le T)\le C_{x_0,T,p}/R^p.$$

Lemma 5.3. If (**S1**)-(**S5**) and (**F1**)-(**F3**) hold with $k>2\vee\theta_1\vee 2\theta_2\vee\theta_4$, then for any $T>0$,

$$\lim_{\Delta_1\to 0}\mathbb{E}\Big(\sup_{0\le t\le T}|\bar{x}(t)-\bar{Z}(t)|^2\Big)=0.$$

Proof. Fix any constant $R>|x_0|$. Define the truncated functions

$$\bar{b}_R(x)=\bar{b}\Big((|x|\wedge R)\frac{x}{|x|}\Big),\quad \sigma_R(x)=\sigma\Big((|x|\wedge R)\frac{x}{|x|}\Big).$$

Consider the SDE

$$\mathrm{d}\bar{u}(t)=\bar{b}_R(\bar{u}(t))\mathrm{d}t+\sigma_R(\bar{u}(t))\mathrm{d}W^1(t) \tag{5.16}$$

with initial value $\bar{u}(0)=x_0$. Under (**S1**), (**S2**) and (**F1**)-(**F3**) with $k\ge\theta_1\vee 2\theta_2$, one observes from (**S1**) and the result of Lemma 4.1 that both $\bar{b}_R(x)$ and $\sigma_R(x)$ are global Lipschitz continuous. Thus equation (5.16) has a unique global solution $\bar{u}(t)$ on $t\ge 0$. Let $\bar{U}(t)$ denote the continuous extension of the EM numerical solution of (5.16). It is well known [18, 24] that

$$\mathbb{E}\Big(\sup_{t\in[0,T]}|\bar{u}(t)-\bar{U}(t)|^2\Big)\le C_T\Delta_1,\quad \forall\, T>0.$$

On the other hand, choose a constant $\bar{\Delta}_1 \in (0,1]$ small sufficiently such that $\varphi^{-1}(K(\bar{\Delta}_1)^{-1/2}) \geq R$. One observes that for any $\Delta_1 \in (0, \bar{\Delta}_1]$

$$\bar{b}_R(x) = \bar{b}(x) = \bar{b}(x^*), \quad \forall\, x \in \mathbb{R}^{n_1} \text{ with } |x| \leq R.$$

Then it is straightforward to see that that for any $t \geq 0$

$$\bar{x}(t \wedge \tau_R) = \bar{u}(t \wedge \tau_R), \quad \bar{Z}(t \wedge \rho_{\Delta_1,R}) = \bar{U}(t \wedge \rho_{\Delta_1,R}), \text{ a.s.},$$

where τ_R and $\rho_{\Delta_1,R}$ are defined in Remarks 2.2 and 5.2, respectively. Under (**S1**)-(**S5**) and (**F1**)-(**F3**) with $k \geq \theta_1 \vee 2\theta_2 \vee \theta_4$, by virtue of Lemmas 2.3 and 5.2, the remainder of the proof follows in a similar manner to that of [34, Theorem 3.5]. To avoid duplication we omit the details. □

Then we turn to prove the strong convergence of the auxiliary process $\bar{Z}(t)$ and the MTEM numerical solution $X(t)$. By virtue of Lemma 5.1, we only need to prove strong convergence of $\bar{Z}(t)$ and $\bar{X}(t)$.

Lemma 5.4. If (**S1**)-(**S5**) and (**F1**)-(**F3**) hold with $k > \theta_1 \vee 2\theta_2 \vee 2\theta_4 \vee (\theta_2 + \theta_4 + 1)$, for any $T > 0$ and $\Delta_1 \in (0,1]$,

$$\lim_{\Delta_2 \to 0} \lim_{M\Delta_2 \to \infty} \mathbb{E}\Big(\sup_{0 \leq t \leq T} |\bar{Z}(t) - \bar{X}(t)|^2 \Big) = 0.$$

Proof. Define $\bar{e}(t) = \bar{Z}(t) - \bar{X}(t)$ for any $t \geq 0$ and $\beta_{\Delta_1,R} = \bar{\rho}_{\Delta_1,R} \wedge \rho_{\Delta_1,R}$ for any $R > 0$, where $\bar{\rho}_{\Delta_1,R}$ and $\rho_{\Delta_1,R}$ are given by (5.8) and (5.15), respectively. Due to $k > 2\theta_4$, let $2 < p \leq k/\theta_4$. Fix $T > 0$. For any $\delta > 0$, using the Young inequality yields that

$$\begin{aligned}
\mathbb{E}\Big(\sup_{0 \leq t \leq T} |\bar{e}(t)|^2 \Big) &= \mathbb{E}\Big(\sup_{0 \leq t \leq T} |\bar{e}(t)|^2 I_{\{\beta_{\Delta_1,R} > T\}} \Big) + \mathbb{E}\Big(\sup_{0 \leq t \leq T} |\bar{e}(t)|^2 I_{\{\beta_{\Delta_1,R} \leq T\}} \Big) \\
&\leq \mathbb{E}\Big(\sup_{0 \leq t \leq T} |\bar{e}(t)|^2 I_{\{\beta_{\Delta_1,R} > T\}} \Big) + \frac{2\delta}{p} \mathbb{E}\Big(\sup_{0 \leq t \leq T} |\bar{e}(t)|^p \Big) + \frac{p-2}{p\delta^{\frac{2}{p-2}}} \mathbb{P}(\beta_{\Delta_1,R} \leq T).
\end{aligned}$$

Owing to (**S3**)-(**S5**) and (**F1**)-(**F3**), it follows from the results of Lemmas 5.1 and 5.2 that

$$\mathbb{E}\Big(\sup_{0 \leq t \leq T} |\bar{e}(t)|^p \Big) \leq 2^{p-1} \mathbb{E}\Big(\sup_{0 \leq t \leq T} |\bar{Z}(t)|^p \Big) + 2^{p-1} \mathbb{E}\Big(\sup_{0 \leq t \leq T} |\bar{X}(t)|^p \Big) \leq C_{x_0,y_0,T,p}.$$

Furthermore, both Remarks 5.2 and 5.1 imply that

$$\mathbb{P}(\beta_{\Delta_1,R} \leq T) \leq \mathbb{P}(\rho_{\Delta_1,R} \leq T) + \mathbb{P}(\bar{\rho}_{\Delta_1,R} \leq T) \leq \frac{C_{x_0,y_0,T,p}}{R^p}.$$

Consequently we have

$$\mathbb{E}\Big(\sup_{0 \leq t \leq T} |\bar{e}(t)|^2 \Big) \leq \mathbb{E}\Big(\sup_{0 \leq t \leq T} |\bar{e}(t)|^2 I_{\{\beta_{\Delta_1,R} > T\}} \Big) + \frac{C_{x_0,y_0,T,p}\delta}{p} + \frac{C_{x_0,y_0,T,p}(p-2)}{p\delta^{\frac{2}{p-2}} R^p}.$$

Now, for any $\epsilon > 0$, choose $\delta > 0$ small sufficiently such that $C_{x_0,y_0,T,p}\delta/p \le \epsilon/3$. Then for this δ, choose $R > 0$ large enough such that $C_{x_0,y_0,T,p}(p-2)/(p\delta^{\frac{2}{p-2}}R^p) \le \epsilon/3$. Hence, for the desired assertion it is sufficient to prove

$$\mathbb{E}\Big(\sup_{0\le t\le T}|\mathrm{e}(t)|^2 I_{\{\beta_{\Delta_1,R}>T\}}\Big) \le \frac{\epsilon}{3}. \tag{5.17}$$

From (3.8) and (5.10) we derive that

$$\begin{aligned}\bar{e}(t\wedge\beta_{\Delta_1,R}) = &\int_0^{t\wedge\beta_{\Delta_1,R}} \Big(\bar{b}(Z^*(s)) - B_M\Big(X^*(s), Y^{X^*(s),y_0}\Big)\Big)\mathrm{d}s\\ &+\int_0^{t\wedge\beta_{\Delta_1,R}} \big(\sigma(Z(s)) - \sigma(X(s))\big)\mathrm{d}W^1(s).\end{aligned}$$

Recalling the definition of the stopping time $\beta_{\Delta_1,R}$, it is straightforward to see that

$$Z^*(s) = Z(s), \quad X^*(s) = X(s), \quad \forall s \in [0, t\wedge\beta_{\Delta_1,R}].$$

Then we have

$$\begin{aligned}\bar{e}(t\wedge\beta_{\Delta_1,R}) = &\int_0^{t\wedge\beta_{\Delta_1,R}} \Big(\bar{b}(Z(s)) - B_M\Big(X(s), Y^{X(s),y_0}\Big)\Big)\mathrm{d}s\\ &+\int_0^{t\wedge\beta_{\Delta_1,R}} \big(\sigma(Z(s)) - \sigma(X(s))\big)\mathrm{d}W^1(s).\end{aligned}$$

Using the Hölder inequality, the Burkholder-Davis-Gundy inequality [33, p.40, Theorem 7.2] and the elementary inequality, we arrive at

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le t\le T}|\bar{e}(t\wedge\beta_{\Delta_1,R})|^2\Big) \le &2T\int_0^T \mathbb{E}\Big(\Big|\bar{b}(Z(s)) - B_M\Big(X(s), Y^{X(s),y_0}\Big)\Big|^2 I_{\{s\le\beta_{\Delta_1,R}\}}\Big)\mathrm{d}s\\ &+8\int_0^T \mathbb{E}|\sigma(Z(s\wedge\beta_{\Delta_1,R})) - \sigma(X(s\wedge\beta_{\Delta_1,R}))|^2\mathrm{d}s\\ \le &4T\int_0^T \mathbb{E}\Big(\Big|\bar{b}(X(s)) - B_M\Big(X(s), Y^{X(s),y_0}\Big)\Big|^2 I_{\{s\le\beta_{\Delta_1,R}\}}\Big)\mathrm{d}s\\ &+4T\int_0^T \mathbb{E}|\bar{b}(Z(s\wedge\beta_{\Delta_1,R})) - \bar{b}(X(s\wedge\beta_{\Delta_1,R}))|^2\mathrm{d}s\\ &+8\int_0^T \mathbb{E}|\sigma(Z(s\wedge\beta_{\Delta_1,R})) - \sigma(X(s\wedge\beta_{\Delta_1,R}))|^2\mathrm{d}s.\end{aligned} \tag{5.18}$$

For any $0 \le s \le T$, one observes that for any $\omega \in \{\omega\in\Omega : s \le \beta_{\Delta_1,R}\}$, $|X(s)| \le R$. Using this fact and (4.5) implies that

$$\begin{aligned}&\mathbb{E}\Big(\Big|\bar{b}(X(s)) - B_M\Big(X(s), Y^{X(s),y_0}\Big)\Big|^2 I_{\{s\le\beta_{\Delta_1,R}\}}\Big)\\ \le&\mathbb{E}\Big(\Big|\bar{b}(X(s)) - B_M\Big(X(s), Y^{X(s),y_0}\Big)\Big|^2 I_{\{|X(s)|\le R\}}\Big)\\ =&\mathbb{E}\Big(\Big|\bar{b}(X_{n_{\Delta_1}(s)}) - B_M\Big(X_{n_{\Delta_1}(s)}, Y_{n_{\Delta_1}(s)}^{X_{n_{\Delta_1}(s)},y_0}\Big)\Big|^2 I_{\{|X_{n_{\Delta_1}(s)}|\le R\}}\Big)\\ =&\mathbb{E}\bigg(\mathbb{E}\Big[\Big(\Big|\bar{b}(X_{n_{\Delta_1}(s)}) - B_M\Big(X_{n_{\Delta_1}(s)}, Y_{n_{\Delta_1}(s)}^{X_{n_{\Delta_1}(s)},y_0}\Big)\Big|^2 I_{\{|X_{n_{\Delta_1}(s)}|\le R\}}\Big)\Big|X_{n_{\Delta_1}(s)}\Big]\bigg)\\ =&\mathbb{E}\Big(\mathbb{E}\Big|\bar{b}(x) - B_M\Big(x, Y_{n_{\Delta_1}(s)}^{x,y_0}\Big)\Big|^2_{x=X_{n_{\Delta_1}(s)}} I_{\{|X_{n_{\Delta_1}(s)}|\le R\}}\Big).\end{aligned}$$

By (**S2**), (**S4**) and (**F1**)-(**F3**) with $k \geq 2\theta_2 \vee 2\theta_4 \vee (\theta_2+\theta_4+1)$, it follows from the result of Lemma 4.12 that

$$\begin{aligned}&\mathbb{E}\Big(\Big|\bar{b}(X(s)) - B_M\Big(X(s), Y^{X(s),y_0}\Big)\Big|^2 I_{\{s\leq \beta_{\Delta_1,R}\}}\Big)\\ =&C\mathbb{E}\Big[\big(1+|X_{n_{\Delta_1}(s)}|^{2(\theta_2+1)\vee 2\theta_3\vee 2\theta_4} + |y_0|^{2(\theta_2+1)\vee 2\theta_3\vee 2\theta_4}\big) I_{\{|X_{n_{\Delta_1}(s)}|\leq R\}}\Big]\Big(\Delta_2 + \frac{1}{M} + \frac{1}{M\Delta_2}\Big)\\ \leq& C_{y_0,R}\Big(\Delta_2 + \frac{1}{M} + \frac{1}{M\Delta_2}\Big). \end{aligned} \tag{5.19}$$

Under (**S1**), (**S2**), (**S4**) and (**F1**)-(**F3**) with $k \geq \theta_1 \vee 2\theta_2 \vee \theta_4$, applying Lemma 4.1 yields that

$$\begin{aligned}&\mathbb{E}\big|\bar{b}(Z(s\wedge \beta_{\Delta_1,})) - \bar{b}(X(s\wedge\beta_{\Delta_1,R}))\big|^2 \vee \mathbb{E}|\sigma(Z(s\wedge\beta_{\Delta_1,R})) - \sigma(X(s\wedge\beta_{\Delta_1,R}))|^2\\ \leq&(\bar{L}_R^2\vee L_R^2)\mathbb{E}|\bar{e}(s\wedge\beta_{\Delta_1,R})|^2 \leq (\bar{L}_R^2\vee L_R^2)\mathbb{E}\Big(\sup_{0\leq r\leq s}|\bar{e}(s\wedge\beta_{\Delta_1,R})|^2\Big). \end{aligned}\tag{5.20}$$

Inserting (5.19) and (5.20) into (5.18) we derive that

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\leq t\leq T}|\bar{e}(t\wedge\beta_{\Delta_1,R})|^2\Big) \leq& 4T^2 C_{y_0,R}\Big(\Delta_2+\frac{1}{M}+\frac{1}{M\Delta_2}\Big)\\ &+ (8+4T)(\bar{L}_R^2\vee L_R^2)\int_0^T \mathbb{E}\Big(\sup_{0\leq r\leq s}|\bar{e}(s\wedge\beta_{\Delta_1,R})|^2\Big)\mathrm{d}s.\end{aligned}$$

An application of the Gronwall inequality implies that

$$\mathbb{E}\Big(\sup_{0\leq t\leq T}|\bar{e}(t\wedge\beta_{\Delta_1,R})|^2\Big) \leq C_{R,y_0}\Big(\Delta_2 + \frac{1}{M\Delta_2}\Big).$$

For the given R, choose $\Delta_2 \in (0,\hat{\Delta}_2]$ small sufficiently such that $C_{y_0,R}\Delta_2 \leq \epsilon/9$. For the fixed Δ_2, choose M large sufficiently such that $C_{y_0,R}/(M\Delta_2) \leq \epsilon/9$. Therefore, we have

$$\mathbb{E}\Big(\sup_{0\leq t\leq T}|\bar{e}(t\wedge\beta_{\Delta_1,R})|^2\Big) \leq \frac{\epsilon}{9} + \frac{2C_{y_0,R}}{M\Delta_2} \leq \frac{\epsilon}{3},$$

which implies that the required assertion (5.17) holds. The proof is complete. □

Obviously, combing the second result of Lemma 5.1, Lemmas 5.3 and 5.4 derives the strong convergence between $\bar{x}(t)$ and $X(t)$.

Theorem 5.1. If (**S1**)-(**S5**) and (**F1**)-(**F3**) hold with $k > \theta_1 \vee 2\theta_2 \vee 2\theta_4 \vee (\theta_2+\theta_4+1)$, then for any $x_0 \in \mathbb{R}^{n_1}$, $y_0 \in \mathbb{R}^{n_2}$, $0 < p < k/\theta_4$ and $T > 0$

$$\lim_{\Delta_1,\Delta_2\to 0}\lim_{M\Delta_2\to\infty}\mathbb{E}\Big(\sup_{0\leq t\leq T}|\bar{x}(t) - X(t)|^p\Big) = 0. \tag{5.21}$$

Proof. For any $T > 0$, combining Lemmas 5.1, 5.3 and 5.4 implies that the desired assertion holds for $p = 2$. Obviously, (5.21) holds for $0 < p < 2$ due to the Hölder inequality. Next, we consider the case $2 < p < k/\theta_4$. Choose a constant $\bar{q}$ such that $p < \bar{q} < k/\theta_4$. Utilizing the Hölder inequality, Lemmas 2.3 and 5.1 we derive that

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\leq t\leq T}|\bar{x}(t)-X(t)|^p\Big) =& \mathbb{E}\Big(\sup_{0\leq t\leq T}\big|\bar{x}(t)-X(t)\big|^{\frac{2(\bar{q}-p)}{\bar{q}-2}}\big|\bar{x}(t)-X(t)\big|^{p-\frac{2(\bar{q}-p)}{\bar{q}-2}}\Big)\\ \leq& \Big[\mathbb{E}\Big(\sup_{0\leq t\leq T}\big|\bar{x}(t)-X(t)\big|^2\Big)\Big]^{\frac{\bar{q}-p}{\bar{q}-2}}\Big[\mathbb{E}\Big(\sup_{0\leq t\leq T}|\bar{x}(t)-X(t)|^{\bar{q}}\Big)\Big]^{\frac{p-2}{\bar{q}-2}}\\ \leq& C_T\Big[\mathbb{E}\Big(\sup_{0\leq t\leq T}|\bar{x}(t)-X(t)|^2\Big)\Big]^{\frac{\bar{q}-p}{\bar{q}-2}}.\end{aligned}$$

This, together with the case of $p=2$, implies the required assertion. The proof is complete. $\square$

Theorem 5.2. If (**S1**)-(**S5**) and (**F1**)-(**F3**) hold with $k > 4\theta_1 \vee 2(\theta_2+1) \vee 2\theta_3 \vee 2\theta_4$, then for any $x_0 \in \mathbb{R}^{n_1}$, $y_0 \in \mathbb{R}^{n_2}$, $0 < p < k/\theta_4$ and $T > 0$,

$$\lim_{\varepsilon\to 0} \lim_{\Delta_1,\Delta_2\to 0} \lim_{M\Delta_2\to\infty} \mathbb{E}\Big(\sup_{0\le t\le T} |x^\varepsilon(t) - X(t)|^p\Big) = 0.$$

Proof. For any $0 < p < k/\theta_4$, using the elementary inequality, by virtue of Lemmas 2.2 and Theorem 5.1, yields that

$$\begin{aligned}&\lim_{\varepsilon\to 0} \lim_{\Delta_1,\Delta_2\to 0} \lim_{M\Delta_2\to\infty} \mathbb{E}\Big(\sup_{0\le t\le T} |x^\varepsilon(t) - X(t)|^p\Big)\\ \le& 2^p \lim_{\varepsilon\to 0} \mathbb{E}|x^\varepsilon(t) - \bar{x}(t)|^p + 2^p \lim_{\Delta_1,\Delta_2\to 0} \lim_{M\Delta_2\to\infty} \mathbb{E}\Big(\sup_{0\le t\le T} |\bar{x}(t) - X(t)|^p\Big) = 0.\end{aligned}$$

The proof is complete. $\square$

6 Strong error

This section focuses on the strong error estimate of the MTEM scheme. To obtain the rates of convergence we need somewhat stronger conditions compared with the convergence alone, which are stated as follows.

(**S1'**) For any $x_1, x_2 \in \mathbb{R}^{n_1}$ and $y \in \mathbb{R}^{n_2}$, there exist constants $\theta_1 \ge 1$ and $K > 0$ such that

$$|b(x_1,y) - b(x_2,y)| + |\sigma(x_1) - \sigma(x_2)| \le K|x_1 - x_2|(1 + |x_1|^{\theta_1} + |x_2|^{\theta_1} + |y|^{\theta_1}).$$

(**S4'**) For any $x_1, x_2 \in \mathbb{R}^{n_1}$ and $y_1, y_2 \in \mathbb{R}^{n_2}$, there is a constant $K_5 > 0$ such that

$$2(x_1 - x_2)^T(b(x_1,y_1) - b(x_2,y_2)) + |\sigma(x_1) - \sigma(x_2)|^2 \le K_5\big(|x_1 - x_2|^2 + |y_1 - y_2|^2\big).$$

Remark 6.1. It follows from (**S1'**) and (**S2**) that for any $(x,y) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$,

$$\begin{aligned}|b(x,y)| &\le |b(x,y) - b(x,0)| + |b(x,0) - b(0,0)| + |b(0,0)|\\ &\le K_1|y|(1 + |x|^{\theta_2} + |y|^{\theta_2}) + K|x|\big(1 + |x|^{\theta_1}\big) + |b(0,0)|\\ &\le C(1 + |x|^{(\theta_1\vee\theta_2)+1} + |y|^{\theta_2+1}),\end{aligned}$$

namely, combining (**S1'**) and (**S2**) leads to (**S4**) with $\theta_3 = \theta_1 \vee \theta_2 + 1$ and $\theta_4 = \theta_2 + 1$.

Remark 6.2. According to Remark 6.1, choose $\varphi(u) = 1 + u^{\theta_1\vee\theta_2}$. Then we have

$$|b(x,y)| \le C \sup_{|x|\le u} \varphi(u)(1 + |x|) + |y|^{\theta_2+1}, \quad \forall\ u \ge 1,\ |x| \le u.$$

Using the similar techniques to that of Lemma 4.1, we derive that the averaged coefficient $\bar{b}$ keeps the property of polynomial growth. To avoid duplication we omit the proof.

Lemma 6.1. If (**S1'**), (**S2**) and (**F1**)-(**F3**) hold with $k \geq 2 \vee \theta_1 \vee 2\theta_2$, then for any $x_1, x_2 \in \mathbb{R}^{n_1}$, there is a constant $\bar{L} > 0$ such that

$$\left|\bar{b}(x_1) - \bar{b}(x_2)\right| \leq \bar{L}|x_1 - x_2|(1 + |x_1|^{\theta_1 \vee \theta_2} + |x_2|^{\theta_1 \vee \theta_2}).$$

Lemma 6.2. If (**S1'**), (**S2**), (**S4'**) and (**F1**)-(**F3**) hold with $k \geq 2 \vee (\theta_2 + 1)$, then for any $x_1, x_2 \in \mathbb{R}^{n_1}$,

$$2(x_1 - x_2)^T\big(\bar{b}(x_1) - \bar{b}(x_2)\big) + |\sigma(x_1) - \sigma(x_2)|^2 \leq C|x_1 - x_2|^2.$$

Proof. Due to (**S1'**), (**S2**) and (**F1**)-(**F3**) with $k \geq \theta_2 + 1$, it follows from the definition of $\bar{b}(x)$ and (**S4'**) that

$$\begin{aligned}
&2(x_1 - x_2)^T(\bar{b}(x_1) - \bar{b}(x_2)) + |\sigma(x_1) - \sigma(x_2)|^2 \\
=&\int_{\mathbb{R}^{n_2} \times \mathbb{R}^{n_2}} \Big[2(x_1 - x_2)^T\big(b(x_1, y_1) - b(x_2, y_2)\big) + |\sigma(x_1) - \sigma(x_2)|^2\Big]\pi(\mathrm{d}y_1 \times \mathrm{d}y_2) \\
\leq& K_5|x_1 - x_2|^2 + K_5\int_{\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}} |y_1 - y_2|^2\pi(\mathrm{d}y_1, \mathrm{d}y_2),
\end{aligned}$$

here $\pi \in \mathcal{C}(\mu^{x_1}, \mu^{x_2})$ is arbitrary. Then owing to the arbitrariness of $\pi \in \mathcal{C}(\mu^{x_1}, \mu^{x_2})$,

$$2(x_1 - x_2)^T(\bar{b}(x_1) - \bar{b}(x_2)) + |\sigma(x_1) - \sigma(x_2)|^2 \leq K_5|x_1 - x_2|^2 + K_5\mathbb{W}_2^2(\mu^{x_1}, \mu^{x_2}).$$

Under (**F1**)-(**F3**), we deduce from (2.2) that

$$2(x_1 - x_2)^T(\bar{b}(x_1) - \bar{b}(x_2)) + |\sigma(x_1) - \sigma(x_2)|^2 \leq C|x_1 - x_2|^2.$$

The proof is complete. □

According to Remark 6.1 and Lemma 4.12, we give the bound of $\mathbb{E}\big|\bar{b}(x) - B_M(x, Y_n^{x,y_0})\big|^2$.

Lemma 6.3. If (**S1'**), (**S2**) and (**F1**)-(**F3**) with $k \geq 2(\theta_2 + 1)$ hold, then for any $x \in \mathbb{R}^{n_1}$, $y_0 \in \mathbb{R}^{n_2}$, $\Delta_2 \in (0, \bar{\Delta}_2]$, and integers $n \geq 0$, $M \geq 1$,

$$\mathbb{E}\big|\bar{b}(x) - B_M(x, Y_n^{x,y_0})\big|^2 \leq C\big(1 + |x|^{2(\theta_2+1)} + |y_0|^{2(\theta_2+1)}\big)\Big(\Delta_2 + \frac{1}{M\Delta_2}\Big).$$

By the same proof techniques as the strong convergence of the MTEM scheme in Section 5, we give the error estimates of $\mathbb{E}|\bar{x}(T) - \bar{Z}(T)|^2$ and $\mathbb{E}|\bar{Z}(T) - X(T)|^2$, respectively.

Lemma 6.4. If (**S1'**), (**S2**), (**S3**), (**S4'**), (**S5**) and (**F1**)-(**F3**) hold with $k \geq 2 \vee \theta_1 \vee 2\theta_2$, then for any $x_0 \in \mathbb{R}^{n_1}$, $T > 0$ and $\Delta_1 \in (0, 1]$,

$$\mathbb{E}|\bar{x}(T) - \bar{Z}(T)|^2 \leq C_{T,x_0}\Delta_1.$$

Proof. Let $e(t) = \bar{x}(t) - \bar{Z}(t)$ for any $t \geq 0$. Define the stopping time

$$\theta_{\Delta_1} = \rho_{\Delta_1, \varphi^{-1}(K\Delta_1^{-1/2})} \wedge \tau_{\varphi^{-1}(K\Delta_1^{-1/2})}.$$

Choosing $p \geq 2(\theta_1 \vee \theta_2 + 1)$ and then using the Young inequality for $p > 2$, we derive that for any $T > 0$,

$$\begin{aligned}\mathbb{E}|e(T)|^2 &= \mathbb{E}\big(|e(T)|^2 I_{\{\theta_{\Delta_1} > T\}}\big) + \mathbb{E}\big(|e(T)|^2 I_{\{\theta_{\Delta_1} \leq T\}}\big) \\ &\leq \mathbb{E}\big(|e(T)|^2 I_{\{\theta_{\Delta_1} > T\}}\big) + \frac{2\Delta_1 \mathbb{E}|e(T)|^p}{p} + \frac{(p-2)\mathbb{P}(\theta_{\Delta_1} \leq T)}{p\Delta_1^{\frac{2}{p-2}}}. \end{aligned} \tag{6.1}$$

Under (**S1'**), (**S2**), (**S3**), (**S5**) and (**F1**)-(**F3**) with $k \geq \theta_1 \vee 2\theta_2$, it follows from the results of Lemmas 2.3 and 5.2 that

$$\mathbb{E}|e(T)|^p \leq C\big(\mathbb{E}|\bar{x}(T)|^p + \mathbb{E}|\bar{Z}(T)|^p\big) \leq C_{x_0,T,p}.$$

Furthermore, by Remarks 2.2 and 5.2 we deduce that

$$\begin{aligned}\mathbb{P}(\theta_{\Delta_1} \leq T) &\leq \mathbb{P}(\tau_{\varphi^{-1}(K\Delta_1^{-1/2})} \leq T) + \mathbb{P}(\rho_{\Delta_1, \varphi^{-1}(K\Delta_1^{-1/2})} \leq T) \\ &\leq \frac{C_{x_0,T,p}}{\big(\varphi^{-1}(K\Delta_1^{-1/2})\big)^p}.\end{aligned}$$

Then inserting the above two inequalities into (6.1) and using $p \geq 2(\theta_1 \vee \theta_2 + 1)$ yield that

$$\mathbb{E}|e(T)|^2 \leq C_{x_0,T,p}\Delta_1 + \mathbb{E}|e(T \wedge \theta_{\Delta_1})|^2.$$

Thus for the desired result it is sufficient to prove

$$\mathbb{E}|e(T \wedge \theta_{\Delta_1})|^2 \leq C_{x_0,T,p}\Delta_1.$$

Recalling the definition of the stopping time θ_{Δ_1}, one observes that $Z^*(t) = Z(t),\ 0 \leq t \leq T \wedge \theta_{\Delta_1}$. Thus using the Itô formula for (1.4) and (5.10) yields that

$$\begin{aligned}\mathbb{E}|e(T \wedge \theta_{\Delta_1})|^2 =& \mathbb{E}\int_0^{T\wedge\theta_{\Delta_1}} \Big[2e^T(t)\big(\bar{b}(\bar{x}(t)) - \bar{b}(Z(t))\big) + |\sigma(\bar{x}(t)) - \sigma(Z(t))|^2\Big]\mathrm{d}t \\ \leq& \mathbb{E}\int_0^{T\wedge\theta_{\Delta_1}} \Big[2e^T(t)\big(\bar{b}(\bar{x}(t)) - \bar{b}(\bar{Z}(t))\big) + |\sigma(\bar{x}(t)) - \sigma(\bar{Z}(t))|^2\Big]\mathrm{d}t \\ &+ 2\mathbb{E}\int_0^{T\wedge\theta_{\Delta_1}} e^T(t)\big(\bar{b}(\bar{Z}(t)) - \bar{b}(Z(t))\big)\mathrm{d}t + \mathbb{E}\int_0^{T\wedge\theta_{\Delta_1}} |\sigma(\bar{Z}(t)) - \sigma(Z(t))|^2\mathrm{d}t \\ &+ 2\mathbb{E}\int_0^{T\wedge\theta_{\Delta_1}} |\sigma(\bar{x}(t)) - \sigma(\bar{Z}(t))||\sigma(\bar{Z}(t)) - \sigma(Z(t))|\mathrm{d}t.\end{aligned}$$

Under (**S4'**) and (**F1**)-(**F3**) with $k \geq \theta_2 + 1$, utilizing the Lemma 6.2 and the Young inequality we derive that

$$\mathbb{E}|e(T \wedge \theta_{\Delta_1})|^2 \leq C\mathbb{E}\int_0^{T\wedge\theta_{\Delta_1}} |e(t)|^2\mathrm{d}t + J_1 + J_2, \tag{6.2}$$

where

$$\begin{aligned}J_1 &= C\mathbb{E}\int_0^{T\wedge\theta_{\Delta_1}} \Big(|\bar{b}(\bar{Z}(t)) - \bar{b}(Z(t))|^2 + |\sigma(\bar{Z}(t)) - \sigma(Z(t))|^2\Big)\mathrm{d}t, \\ J_2 &= C\mathbb{E}\int_0^{T\wedge\theta_{\Delta_1}} |\sigma(\bar{x}(t)) - \sigma(\bar{Z}(t))||\sigma(\bar{Z}(t)) - \sigma(Z(t))|\mathrm{d}t.\end{aligned}$$

Due to (**S1'**), (**S2**), (**S3**), (**S5**) and (**F1**)-(**F3**) with $k \geq \theta_1 \vee 2\theta_2 \vee \theta_4$, it follows from the results of Lemmas 5.2 and 6.1 that

$$\begin{aligned} J_1 &\leq C \int_0^T \mathbb{E}\big[|\bar{Z}(t) - Z(t)|^2 (1 + |\bar{Z}(t)|^{2(\theta_1 \vee \theta_2)} + |Z(t)|^{2(\theta_1 \vee \theta_2)})\big] \mathrm{d}t \\ &\leq C \int_0^T \Big(\mathbb{E}|\bar{Z}(t) - Z(t)|^4\Big)^{\frac{1}{2}} \Big[\mathbb{E}\Big(1 + |\bar{Z}(t)|^{4(\theta_1 \vee \theta_2)} + |Z(t)|^{4(\theta_1 \vee \theta_2)}\Big)\Big]^{\frac{1}{2}} \mathrm{d}t \leq C_{x_0,T,p}\Delta_1. \end{aligned} \tag{6.3}$$

In addition, using the Young inequality and the Hölder inequality yields that

$$\begin{aligned} J_2 &\leq C\mathbb{E} \int_0^{T \wedge \theta_{\Delta_1}} |e(t)||\bar{Z}(t) - Z(t)|(1 + |\bar{x}(t)|^{2\theta_1} + |\bar{Z}(t)|^{2\theta_1} + |Z(t)|^{2\theta_1}) \mathrm{d}t \\ &\leq C \int_0^T \big(\mathbb{E}|\bar{Z}(t) - Z(t)|^4\big)^{\frac{1}{2}} \Big[\mathbb{E}\big(1 + |\bar{x}(t)|^{8\theta_1} + |\bar{Z}(t)|^{8\theta_1} + |Z(t)|^{8\theta_1}\big)\Big]^{\frac{1}{2}} \mathrm{d}t \\ &\quad + C \int_0^T \mathbb{E}|e(t \wedge \theta_{\Delta_1})|^2 \mathrm{d}t. \end{aligned}$$

Similarly to (6.3), applying Lemmas 2.3 and 5.2 we show that

$$J_2 \leq C_{x_0,T,p}\Delta_1 + C \int_0^T \mathbb{E}|e(t \wedge \theta_{\Delta_1})|^2 \mathrm{d}t. \tag{6.4}$$

Inserting (6.3) and (6.4) into (6.2) and then using Gronwall's inequality derive that

$$\mathbb{E}|e(T \wedge \theta_{\Delta_1})|^2 \leq C_{x_0,T,p}\Delta_1,$$

which implies the desired result. The proof is complete. □

Lemma 6.5. *If* (**S1'**), (**S2**), (**S3**), (**S4'**), (**S5**) *and* (**F1**)-(**F3**) *with* $k \geq [2(2\theta_1 + 1) \vee 2(\theta_1 \vee \theta_2 + 1)](\theta_2 + 1)$ *hold, then for any* $x_0 \in \mathbb{R}^{n_1}$, $y_0 \in \mathbb{R}^{n_2}$, $T > 0$, $\Delta_1 \in (0, 1]$, $\Delta_2 \in (0, \bar{\Delta}_2]$ *and* $M \geq 1$,

$$\mathbb{E}|\bar{Z}(T) - \bar{X}(T)|^2 \leq C_{x_0,y_0,T}\Big(\Delta_1 + \Delta_2 + \frac{1}{M\Delta_2}\Big).$$

Proof. Define the stopping time

$$\bar{\theta}_{\Delta_1} = \bar{\rho}_{\Delta_1,\varphi^{-1}(K\Delta_1^{-1/2})} \wedge \rho_{\Delta_1,\varphi^{-1}(K\Delta_1^{-1/2})},$$

where $\bar{\rho}_{\Delta_1,\varphi^{-1}(K\Delta_1^{-1/2})}$ and $\rho_{\Delta_1,\varphi^{-1}(K\Delta_1^{-1/2})}$ are given by (5.8) and (5.15). Due to $k \geq [2(2\theta_1 + 1) \vee 2(\theta_1 \vee \theta_2 + 1)](\theta_2 + 1)$, we can choose a constant p such that

$$2 < 2(\theta_1 \vee \theta_2 + 1) \vee 2(2\theta_1 + 1) \leq p \leq k/(\theta_2 + 1).$$

By (**S1'**), (**S2**), (**S3**), (**S5**) and (**F1**)-(**F3**), using Lemmas 5.1 and 5.2 as well as the Hölder inequality yields that

$$\sup_{\Delta_1 \in (0,1]} \mathbb{E}\Big(\sup_{0 \leq t \leq T} |\bar{Z}(t)|^p\Big) \vee \sup_{\Delta_1 \in (0,1], \Delta_2 \in (0, \hat{\Delta}_2]} \mathbb{E}\Big(\sup_{t \in [0,T]} |\bar{X}(t)|^p\Big) \leq C_{x_0,y_0,T,p}. \tag{6.5}$$

Then applying the Young inequality, for any $\delta > 0$ we obtain that

$$\begin{aligned}
&\mathbb{E}|\bar{e}(T)|^2 = \mathbb{E}\big(|\bar{e}(T)|^2 I_{\{\bar{\theta}_{\Delta_1} > T\}}\big) + \mathbb{E}\big(|\bar{e}(T)|^2 I_{\{\bar{\theta}_{\Delta_1} \le T\}}\big) \\
\le &\mathbb{E}\big(|\bar{e}(T)|^2 I_{\{\bar{\theta}_{\Delta_1} > T\}}\big) + \frac{2\Delta_1}{p}\mathbb{E}|\bar{e}(T)|^p + \frac{p-2}{p\Delta_1^{\frac{2}{p-2}}}\mathbb{P}(\bar{\theta}_{\Delta_1} \le T).
\end{aligned} \tag{6.6}$$

It follows from (6.5) that

$$\mathbb{E}|\bar{e}(T)|^p \le 2^{p-1}\mathbb{E}|\bar{Z}(T)|^p + 2^{p-1}\mathbb{E}|\bar{X}(t)|^p \le C_{x_0,y_0,T,p}.$$

Furthermore, by the Markov inequality and (6.5) we derive that

$$\begin{aligned}
\mathbb{P}\Big(\bar{\rho}_{\Delta_1,\varphi^{-1}(K\Delta_1^{-1/2})} \le T\Big) &\le \mathbb{P}\Big(|\bar{X}(T \wedge \bar{\rho}_{\Delta_1,\varphi^{-1}(K\Delta_1^{-1/2})})| \ge \varphi^{-1}(K\Delta_1^{-1/2})\Big) \\
&\le \frac{\mathbb{E}\Big|\bar{X}(T \wedge \bar{\rho}_{\Delta_1,\varphi^{-1}(K\Delta_1^{-1/2})})\Big|^p}{\big(\varphi^{-1}(K\Delta_1^{-1/2})\big)^p} \le \frac{C_{x_0,y_0,T,p}}{\big(\varphi^{-1}(K\Delta_1^{-1/2})\big)^p}.
\end{aligned}$$

Then combining the above inequality and Remark 5.2 gives that

$$\mathbb{P}(\bar{\theta}_{\Delta_1} \le T) \le \mathbb{P}(\rho_{\Delta_1,\varphi^{-1}(K\Delta_1^{-1/2})} \le T) + \mathbb{P}(\bar{\rho}_{\Delta_1,\varphi^{-1}(K\Delta_1^{-1/2})} \le T) \le \frac{C_{x_0,y_0,T,p}}{(\varphi^{-1}(K\Delta_1^{-\frac{1}{2}}))^p}.$$

Due to $p \ge 2(\theta_1 \vee \theta_2 + 1)$, inserting the above inequality into (6.6) shows that

$$\begin{aligned}
\mathbb{E}|\bar{e}(T)|^2 &\le \mathbb{E}\big(|\bar{e}(T)|^2 I_{\{\bar{\theta}_{\Delta_1} > T\}}\big) + \frac{C_{x_0,y_0,T,p}\Delta_1}{p} + \frac{C_{x_0,y_0,T,p}}{p\Delta^{\frac{2}{p-2}}(\varphi^{-1}(K\Delta_1^{-\frac{1}{2}}))^p} \\
&\le \mathbb{E}\big(|\bar{e}(T)|^2 I_{\{\bar{\theta}_{\Delta_1} > T\}}\big) + C_{x_0,y_0,T,p}\Delta_1.
\end{aligned}$$

Hence for the desired result it remains to prove that

$$\mathbb{E}\big(|\bar{e}(T)|^2 I_{\{\bar{\theta}_{\Delta_1} > T\}}\big) \le C_{x_0,y_0,T,p}\Delta_1.$$

Obviously, $X^*(t) = X(t)$ and $Z^*(t) = Z(t)$ for any $0 \le t \le T \wedge \bar{\theta}_{\Delta_1}$. Using the Itô formula for (3.8) and (5.10) and the Young inequality, under (**S4'**) and (**F1**)-(**F3**), by Lemma 6.2 we arrive at that for any $T > 0$,

$$\begin{aligned}
\mathbb{E}|\bar{e}(T \wedge \bar{\theta}_{\Delta_1})|^2 &= \mathbb{E}\int_0^{T\wedge\bar{\theta}_{\Delta_1}} \Big[2\bar{e}^T(t)\Big(\bar{b}(Z(t)) - B_M\Big(X(t), Y^{X(t),y_0}\Big)\Big) + |\sigma(Z(t)) - \sigma(X(t))|^2\Big]\mathrm{d}t \\
&\le \mathbb{E}\int_0^{T\wedge\bar{\theta}_{\Delta_1}} \Big[2\bar{e}^T(t)\Big(\bar{b}(\bar{Z}(t)) - \bar{b}(\bar{X}(t))\Big) + |\sigma(\bar{Z}(t)) - \sigma(\bar{X}(t))|^2\Big]\mathrm{d}t \\
&\quad + \int_0^T C\mathbb{E}|\bar{e}(t \wedge \bar{\theta}_{\Delta_1})|^2\mathrm{d}t + I_1 + I_2 + I_3 + I_4 \\
&\le \int_0^T C\mathbb{E}|\bar{e}(t \wedge \bar{\theta}_{\Delta_1})|^2\mathrm{d}t + I_1 + I_2 + I_3 + I_4,
\end{aligned} \tag{6.7}$$

where

$$I_1=\int_0^T\mathbb{E}\Big|\bar{b}(X(t))-B_M\Big(X(t),Y^{X(t),y_0}\Big)\Big|^2\mathrm{d}t,$$
$$I_2=C\int_0^T\mathbb{E}\big(|\bar{b}(\bar{X}(t))-\bar{b}(X(t))|^2+|\sigma(\bar{X}(t))-\sigma(X(t))|^2\big)\mathrm{d}t,$$
$$I_3=C\int_0^T\mathbb{E}\big(|\bar{b}(Z(t))-\bar{b}(\bar{Z}(t))|^2+|\sigma(Z(t))-\sigma(\bar{Z}(t))|^2\big)\mathrm{d}t,$$
$$I_4=C\mathbb{E}\int_0^{T\wedge\bar{\theta}_{\Delta_1}}|\sigma(\bar{Z}(t))-\sigma(\bar{X}(t))|\big(|\sigma(\bar{X}(t))-\sigma(X(t))|+|\sigma(Z(t))-\sigma(\bar{Z}(t))|\big)\mathrm{d}t.$$

In addition, owing to (**S1'**), (**S2**) and (**F1**)-(**F3**) with $k\geq 2(\theta_2+1)$, applying (4.5) and Lemma 6.3 implies that for any $0\leq t\leq T$,

$$\begin{aligned}\mathbb{E}\Big|\bar{b}(X(t))-B_M\Big(X(t),Y^{X(t),y_0}\Big)\Big|^2&=\mathbb{E}\Big|\bar{b}(X_{n_{\Delta_1}(t)})-B_M\Big(X_{n_{\Delta_1}(t)},Y^{X_{n_{\Delta_1}(t)},y_0}\Big)\Big|^2\\&=\mathbb{E}\Big[\mathbb{E}\Big(\Big|\bar{b}(X_{n_{\Delta_1}(t)})-B_M\Big(X_{n_{\Delta_1}(t)},Y_{n_{\Delta_1}(t)}^{X_{n_{\Delta_1}(t)},y_0}\Big)\Big|^2\Big|X_{n_{\Delta_1}(t)}\Big)\Big]\\&=\mathbb{E}\Big(\mathbb{E}\Big|\bar{b}(x)-B_M\Big(x,Y_{n_{\Delta_1}(t)}^{x,y_0}\Big)\Big|^2\Big|_{x=X_{n_{\Delta_1}(t)}}\Big)\\&\leq C\Big(\Delta_2+\frac{1}{M\Delta_2}\Big)\Big(1+|y_0|^{2(\theta_2+1)}+\mathbb{E}|X_{n_{\Delta_1}(t)}|^{2(\theta_2+1)}\Big).\end{aligned}$$

Furthermore, due to $p>2(\theta_2+1)$, utilizing (6.5) and the Hölder inequality we deduce that

$$\begin{aligned}I_1&\leq C\Big(\Delta_2+\frac{1}{M\Delta_2}\Big)\int_0^T\big(1+|y_0|^{2(\theta_2+1)}+\mathbb{E}|X_{n_{\Delta_1}(t)}|^{2(\theta_2+1)}\big)\mathrm{d}t\\&\leq C_{y_0}\Big(\Delta_2+\frac{1}{M\Delta_2}\Big)\int_0^T\big(1+\big(\mathbb{E}|X_{n_{\Delta_1}(t)}|^p\big)^{\frac{2(\theta_2+1)}{p}}\big)\mathrm{d}t\\&\leq C_{y_0,T}\Big(\Delta_2+\frac{1}{M\Delta_2}\Big).\end{aligned}\tag{6.8}$$

Under (**S1'**), (**S2**) and (**F1**)-(**F3**) with $k\geq\theta_1\vee 2\theta_2$, by Lemma 6.1 and the Hölder inequality we derive that

$$\begin{aligned}I_2+I_3\leq{}&C\int_0^T\mathbb{E}\Big(|\bar{X}(t)-X(t)|^2\big(1+|X(t)|^{2(\theta_1\vee\theta_2)}+|\bar{X}(t)|^{2(\theta_1\vee\theta_2)}\big)\Big)\mathrm{d}t\\&+C\int_0^T\mathbb{E}\Big(|Z(t)-\bar{Z}(t)|^2\big(1+|Z(t)|^{2(\theta_1\vee\theta_2)}+|\bar{Z}(t)|^{2(\theta_1\vee\theta_2)}\big)\Big)\mathrm{d}t\\\leq{}&C\int_0^T\Big(\mathbb{E}|\bar{X}(t)-X(t)|^p\Big)^{\frac{2}{p}}\Big(\mathbb{E}\big(1+|X(t)|^{\frac{2p(\theta_1\vee\theta_2)}{p-2}}+|\bar{X}(t)|^{\frac{2p(\theta_1\vee\theta_2)}{p-2}}\big)\Big)^{\frac{p-2}{p}}\mathrm{d}t\\&+C\int_0^T\Big(\mathbb{E}|Z(t)-\bar{Z}(t)|^p\Big)^{\frac{2}{p}}\Big(\mathbb{E}\big(1+|Z(t)|^{\frac{2p(\theta_1\vee\theta_2)}{p-2}}+|\bar{Z}(t)|^{\frac{2p(\theta_1\vee\theta_2)}{p-2}}\big)\Big)^{\frac{p-2}{p}}\mathrm{d}t.\end{aligned}$$

Thanks to $2(\theta_1\vee\theta_2+1)\leq p\leq k/(\theta_2+1)$, we have $2p(\theta_1\vee\theta_2)/(p-2)\leq p\leq k/(\theta_2+1)$. Then applying Lemmas 5.1 and 5.2 and the Hölder inequality yields that

$$\begin{aligned}I_2+I_3\leq{}&C\int_0^T\Big(\mathbb{E}|\bar{X}(t)-X(t)|^p\Big)^{\frac{2}{p}}\Big(\mathbb{E}(1+|X(t)|^p+|\bar{X}(t)|^p)\Big)^{\frac{p-2}{p}}\mathrm{d}t\\&+C\int_0^T\Big(\mathbb{E}|Z(t)-\bar{Z}(t)|^p\Big)^{\frac{2}{p}}\Big(\mathbb{E}(1+|Z(t)|^p+|\bar{Z}(t)|^p)\Big)^{\frac{p-2}{p}}\mathrm{d}t\leq C_{x_0,y_0,T,p}\Delta_1.\end{aligned}\tag{6.9}$$

In view of (**S1'**), together with using the Young inequality and the Hölder inequality, we also obtain that

$$
\begin{aligned}
I_4 \le & C\mathbb{E}\int_0^{T\wedge\bar{\theta}_{\Delta_1}} |\bar{e}(t)|\big(|\bar{X}(t)-X(t)|+|Z(t)-\bar{Z}(t)|\big) \\
& \times \big(1+|X(t)|^{2\theta_1}+|\bar{X}(t)|^{2\theta_1}+|\bar{Z}(t)|^{2\theta_1}+|Z(t)|^{2\theta_1}\big)\mathrm{d}t \\
\le & C\int_0^T \mathbb{E}|\bar{e}(t\wedge\bar{\theta}_{\Delta_1})|^2\mathrm{d}t + C\int_0^T \Big[\mathbb{E}\big(|\bar{X}(t)-X(t)|^p+|Z(t)-\bar{Z}(t)|^p\big)\Big]^{\frac{2}{p}} \\
& \times \Big[\mathbb{E}\big(1+|X(t)|^{\frac{4p\theta_1}{p-2}}+|\bar{X}(t)|^{\frac{4p\theta_1}{p-2}}+|Z(t)|^{\frac{4p\theta_1}{p-2}}+|\bar{Z}(t)|^{\frac{4p\theta_1}{p-2}}\big)\Big]^{\frac{p-2}{p}}\mathrm{d}t.
\end{aligned}
$$

Similarly, owing to $2(2\theta_1+1)\le p\le k/(\theta_2+1)$, $4p\theta_1/(p-2)\le p\le k/(\theta_2+1)$. By means of Lemmas 5.1 and 5.2 and using the Hölder inequality we deduce that

$$
\begin{aligned}
I_4 \le & C\int_0^T \mathbb{E}|\bar{e}(t\wedge\bar{\theta}_{\Delta_1})|^2\mathrm{d}t + C\int_0^T \Big[\mathbb{E}\big(|\bar{X}(t)-X(t)|^p+|Z(t)-\bar{Z}(t)|^p\big)\Big]^{\frac{2}{p}} \\
& \times \Big[\mathbb{E}\big(1+|X(t)|^p+|\bar{X}(t)|^p+|Z(t)|^p+|\bar{Z}(t)|^p\big)\Big]^{\frac{p-2}{p}}\mathrm{d}t \\
\le & C\int_0^T \mathbb{E}|\bar{e}(t\wedge\bar{\theta}_{\Delta_1})|^2\mathrm{d}t + C_{x_0,y_0,T,p}\Delta_1.
\end{aligned}
\tag{6.10}
$$

Then inserting (6.8)-(6.10) into (6.7) implies that

$$
\mathbb{E}|\bar{e}_{\Delta_1}(T\wedge\beta_{\Delta_1})|^2 \le C\int_0^T \mathbb{E}|\bar{e}(t\wedge\beta_{\Delta_1})|^2\mathrm{d}t + C_{x_0,y_0,T,p}\Big(\Delta_1+\Delta_2+\frac{1}{M\Delta_2}\Big).
$$

Using the Gronwall inequality shows that

$$
\mathbb{E}|\bar{e}_{\Delta_1}(T\wedge\beta_{\Delta_1})|^2 \le C_{x_0,y_0,T,p}\Big(\Delta_1+\Delta_2+\frac{1}{M\Delta_2}\Big),
$$

which implies the desired result. The proof is complete. □

Combining Lemmas 5.1, 6.4 and 6.5, the error estimate of the MTEM scheme is yielded directly.

Theorem 6.1. If (**S1'**), (**S2**), (**S3**), (**S4'**), (**S5**) and (**F1**)-(**F3**) hold with $k\ge[2(2\theta_1+1)\vee 2(\theta_1\vee\theta_2+1)](\theta_2+1)$, then for any $x_0\in\mathbb{R}^{n_1}$, $y_0\in\mathbb{R}^{n_2}$, $T>0$, $\Delta_1\in(0,\bar{\Delta}_1]$, $\Delta_2\in(0,\bar{\Delta}_2]$ and $M\ge 1$,

$$
\mathbb{E}|\bar{x}(T)-X(T)|^2 \le C_{x_0,y_0,T}\Big(\Delta_1+\Delta_2+\frac{1}{M\Delta_2}\Big).
$$

Theorem 6.1 gives the strong error estimate between the exact solution of the averaged equation (1.4) and the numerical solution generated by MTEM scheme. The determination of the strong convergence rate of the averaging principle further allows us to ascertain the strong error estimate between the slow component of the original system and the MTEM numerical solution. An important case is presented here to illustrate this. Let us assume that the slow drift term $b = b_1 + b_2$ and satisfies that

(**B1**) There exist constants $C_1 > 0$, $\alpha > 0$ and $\theta \geq 2$ such that for any $x \in \mathbb{R}^{n_1}$,

$$x^T b_1(x) \leq -\alpha|x|^\theta + C_1(1 + |x|^2).$$

(**B2**) There exists a constants $L > 0$ such that for any $x, x_i \in \mathbb{R}^{n_1}$ and $y_i \in \mathbb{R}^{n_2}$, $i = 1, 2$,

$$\begin{aligned} &|b_1(x)| \leq L(1 + |x|^{\theta-1}), \\ &|b_2(x_1, y_1) - b_2(x_2, y_2)| + |\sigma(x_1) - \sigma(x_2)| \leq L(|x_1 - x_2| + |y_1 - y_2|), \end{aligned}$$

where the constant θ is given in (**B1**).

(**B3**) There exists a constant $K > 0$ such that for any $x_1, x_2 \in \mathbb{R}^{n_1}$,

$$(x_1 - x_2)^T (b_1(x_1) - b_1(x_2)) \leq K|x_1 - x_2|^2.$$

Meanwhile, Assumptions (**F1**)-(**F3**) are preserved without modification. Subsequently, the subsequent strong averaging principle can be inferred from [19, Theorem 2.2].

Lemma 6.6 ([19, Theorem 2.2])**.** Suppose that (**B1**)-(**B3**) and (**F1**)-(**F3**) hold. Then for any $(x_0, y_0) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ and $T > 0$,

$$\mathbb{E}\Big(\sup_{t\in[0,T]} |x^\varepsilon(t) - \bar{x}(t)|^2\Big) \leq C\varepsilon^{\frac{1}{3}}.$$

Theorem 6.2. Suppose that (**B1**)-(**B3**) and (**F1**)-(**F3**) hold with $k \geq 4(2\theta - 1)$. Then for any $T > 0$, $\Delta \in (0, 1]$, $\Delta_2 \in (0, \bar{\Delta}_2]$ and $M > 1$,

$$\mathbb{E}|x^\varepsilon(T) - X(T)|^2 \leq C_T \left(\varepsilon^{\frac{1}{3}} + \Delta_1 + \Delta_2 + \frac{1}{M\Delta_2}\right).$$

7 Numerical examples

This section gives two examples and carries out some numerical experiments by the MTEM scheme to verify the theoretical results.

Example 7.1. Recall the SFSDE (1.5). The exact solution of the averaged equation with initial value $\bar{x}(0) = x_0$ has the closed form (see, e.g., [21, 24])

$$\bar{x}(t) = \frac{x_0 \exp(-\frac{3}{2}t + W^1(t))}{\sqrt{1 + 2x_0^2 \int_0^t \exp(-3s + 2W^1(s))\mathrm{d}s}}.$$

It can be verified that (**S1'**), (**S2**), (**S3**), (**S4'**), (**S5**) and (**F1**)-(**F3**) hold with $\theta_1 = 2, \theta_2 = 1$ and any $k \geq 2$. According to Remark 6.2, we can choose $\varphi(u) = 1 + u^2$, $\forall\ u \geq 1$. For the fixed $\Delta_1, \Delta_2 \in (0, 1]$ and integer $M \geq 1$, define the MTEM scheme for (1.5): for any $n \geq 0$,

$$\begin{cases} X_0 = x_0, X_n^* = \left(|X_n| \wedge \left(2\Delta_1^{-\frac{1}{2}} - 1\right)^{\frac{1}{2}}\right)\frac{X_n}{|X_n|},\ Y_0^{X_n^*} = y_0, \\ Y_{m+1}^{X_n^*} = Y_m^{X_n^*} + (X_n^* - Y_m^{X_n^*})\Delta_2 + \Delta W_{n,m}^2, \quad m = 0, 1, \dots, M-1, \\ B_M(X_n^*, Y^{X_n^*}) = -\left(X_n^*\right)^3 - \dfrac{1}{M}\displaystyle\sum_{m=1}^{M} Y_m^{X_n^*}, \\ X_{n+1} = X_n + B_M(X_n^*, Y^{X_n^*})\Delta_1 + X_n \Delta W_n^1. \end{cases} \tag{7.1}$$

Figure 2 predicts the numerical solution generated by the MTEM scheme and the exact solution of the averaged equation (1.7). Comparing Figure 1 and 2 one observes that the truncation device in the MTEM scheme effectively suppresses the explosive divergence phenomenon of the PI iteration process. Correcting the grid points by using the truncation mapping, the MTEM numerical solution rapidly converges to the exact solution of the averaged equation after going through the initial transient oscillation phase.

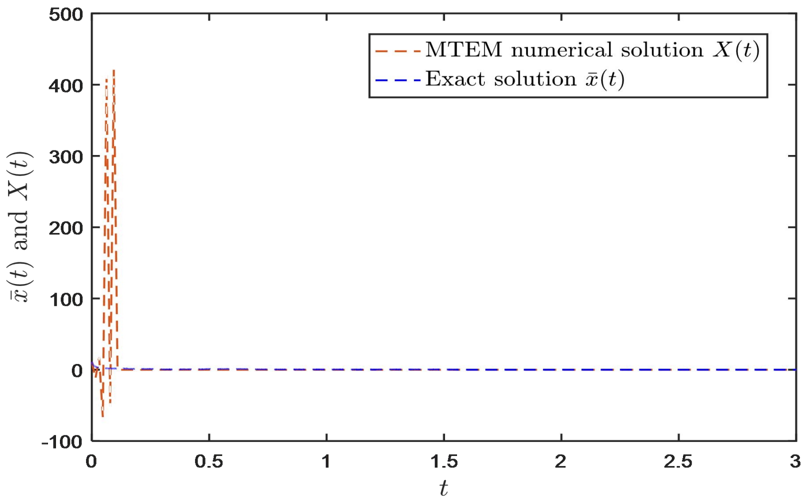

Figure 2: The sample paths of the MTEM numerical solution $X(t)$ on $t \in [0,3]$ with $\Delta_1 = 2^{-6}$, $\Delta_2 = 2^{-6}$ and $M = 2^{18}$.

Owing to Theorem 2.2, one notices that $x^{\varepsilon}(t)$ converges to $\bar{x}(t)$ as $\varepsilon \to 0$. Next we pay attention to the strong convergence between $\bar{x}(t)$ and the numerical solution $X(t)$ by the MTEM scheme (7.1) as $\Delta_1, \Delta_2 \to 0$ and $M\Delta_2 \to \infty$ revealed by Theorem 6.1. To verify this result, we carry out some numerical experiments by the MTEM scheme. Provided that we want to bound the error by $\mathcal{O}(2^{-q})(q > 0)$, the optimal parameters are derived by Theorem 6.1 as follows:

$$\Delta_1 = \mathcal{O}(2^{-q}), \quad \Delta_2 = \mathcal{O}(2^{-q}), \quad M = \mathcal{O}(2^{2q}).$$

In the numerical calculations, using 500 sample points we compute the sample mean square of the error (SMSE)

$$\mathbb{E}|\bar{x}(t) - X(t)|^2 \approx \frac{1}{500}\sum_{j=1}^{500} |\bar{x}^{(j)}(n\Delta_1) - X_n^{(j)}|^2, \tag{7.2}$$

where $\bar{x}^{(j)}(n\Delta_1)$ and $X_n^{(j)}$ are sequences of independent copies of $\bar{x}(n\Delta_1)$ and X_n, respectively. Note that for the fixed n and j, $\bar{x}^{(j)}(n\Delta_1)$ and $X_n^{(j)}$ are generated by a same Brownian motion. Then we carry out numerical experiments by implementing (7.1) using MATLAB. In Figure 3, the blue solid line depicts the SMSE for $q = 2, 3, 4, 5, 6, 7$ with 500 sample points. The red dotted line plots the reference line with the slope -1. In addition, we plot 10 groups of sample paths of $\bar{x}(t)$ and $X(t)$ for $t \in [0, 5]$ with $(\Delta_1, \Delta_2, M) = (2^{-8}, 2^{-6}, 2^{12})$. The Figure 4 only depicts four groups of them.

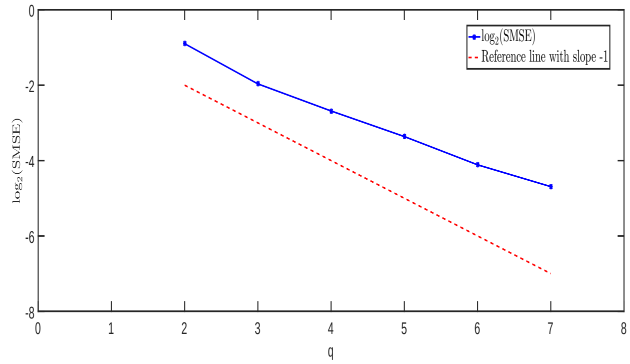

Figure 3: The SMSE for $q = 2, 3, 4, 5, 6, 7$ with 500 sample points. The red dashed line is the reference with slope -1.

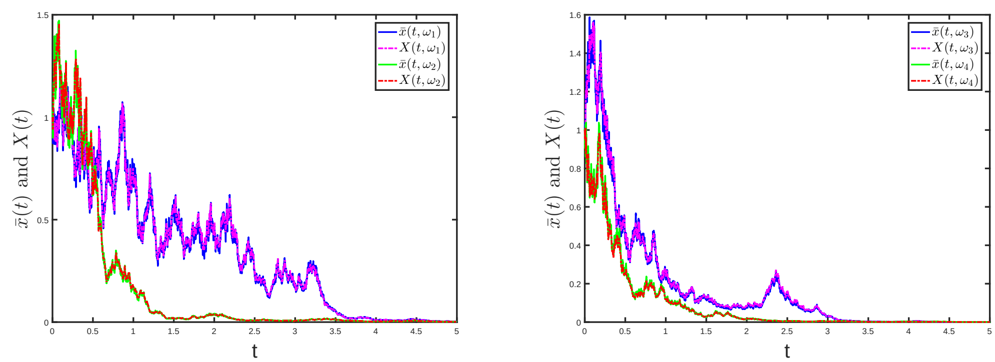

Figure 4: Four pairs of sample paths of $\bar{x}(t)$ and $X(t)$ for $t \in [0, 5]$ with $(\Delta_1, \Delta_2, M) = (2^{-8}, 2^{-6}, 2^{12})$.

Example 7.2. Consider the following SFSDE

$$\begin{cases} dx^\varepsilon(t) = \left[x^\varepsilon(t) - x^\varepsilon(t)(y^\varepsilon(t))^2 + y^\varepsilon(t)\right]dt + x^\varepsilon(t)dW^1(t), \\ dy^\varepsilon(t) = \dfrac{1}{\varepsilon}(x^\varepsilon(t) - 4y^\varepsilon(t))dt + \dfrac{1}{\sqrt{\varepsilon}}(x^\varepsilon(t) + y^\varepsilon(t))dW^2(t) \end{cases} \tag{7.3}$$

with the initial value $(x_0, y_0) = (1, 1)$. Assume that

$$b(x, y) = x - xy^2 + y, \quad \sigma(x) = x, \quad f(x, y) = x - 4y, \quad g(x, y) = x + y. \tag{7.4}$$

It can be verified that (**S1**)-(**S5**) and (**F1**)-(**F3**) hold with $\theta_1 = \theta_2 = 2, \theta_3 = \theta_4 = 2$ and $6 < k < 9$. Then using lemma 2.2 yields that the strong convergence between $x^\varepsilon(t)$ and the averaged equation $\bar{x}(t)$ in pth $(0 < p < k)$ moment. Although the averaged equation provides a substantial simplification for SFSDE (7.3), the closed form of the averaged equation is unavailable. Then classical numerical approximation techniques can't be used directly. This is where MTEM scheme defined by (3.6) comes in.

First, by (3.1) we take $\varphi(u) = 1 + u^2, u \geq 1$. Then for any $\Delta_1 \in (0, 1], \Delta_2 \in (0, 1]$ and integer

$M > 0$, define the MTEM scheme for (7.3): for any $n \geq 0$,

$$\begin{cases} X_0 = 1, X_n^* = \Big(|X_n| \wedge \big(2\Delta_1^{-\frac{1}{2}} - 1\big)^{\frac{1}{2}}\Big)\frac{X_n}{|X_n|}, Y_0^{X_n^*} = 1, \\ Y_{m+1}^{X_n^*} = Y_m^{X_n^*} + (X_n^* - 4Y_m^{X_n^*})\Delta_2 + Y_m^{X_n^*}\Delta W_{n,m}^2, \quad m = 0, 1, \dots, M-1, \\ B_M(X_n^*, Y^{X_n^*}) = X_n^* + \dfrac{1}{M}\displaystyle\sum_{m=1}^{M}\big(- X_n^*(Y_m^{X_n^*})^2 + Y_m^{X_n^*}\big), \\ X_{n+1} = X_n + B_M(X_n^*, Y^{X_n^*})\Delta_1 + X_n\Delta W_n^1. \end{cases}$$

Therefore, by Theorem 5.2, using this scheme we can approximate the slow component of SFSDE (7.3) in the pth ($0 < p < k/2$) moment. In order to test the efficiency of the scheme, we carry out numerical experiments by implementing (7.5) using MATLAB. Let $(\Delta_1, \Delta_2, M) = (2^{-10}, 2^{-8}, 2^{16})$. The Figure 5 depicts the five sample paths of $|X(t)|$ and sample mean value of 100 sample points in different time interval $[0, T]$, where $T = 5$ (left), $T = 10$ (middle) and $T = 20$(right), respectively.

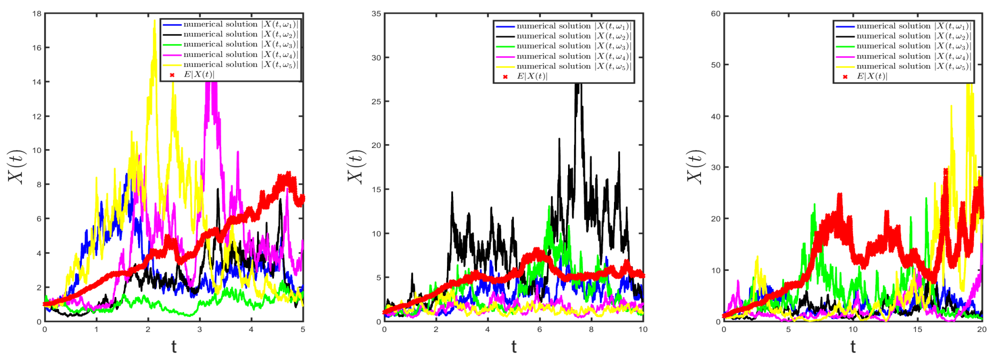

Figure 5: Five sample paths and sample mean value of $|X(t)|$ for 100 sample points in different time intervals

8 Concluding remarks

In this paper, we have developed an explicit numerical scheme tailored for a category of super-linear SFSDEs wherein the slow drift coefficient exhibits polynomial growth. An explicit multiscale numerical scheme, termed MTEM, has been proposed through the application of a truncation mechanism. The strong convergence of the numerical solutions yielded by the MTEM scheme has been rigorously established. Furthermore, the convergence rate has been determined under weakly restrictive conditions. The construction of an explicit scheme to approximate the dynamical behaviors of the exact solutions for more generic SFSDEs featuring a super-linear fast component remains an intriguing topic for future investigation. This direction will inform our subsequent research endeavors.

References

[1] J. Bao, J. Shao, C. Yuan, Approximation of invariant measures for regime-switching diffusions, Potential Anal., 44 (2016), pp. 707-727.

[2] C.-E. Bréhier, Analysis of an HMM time-discretization scheme for a system of stochastic PDEs, SIAM J. Numer. Anal., 51 (2013), pp. 1185–1210.

[3] C.-E. Bréhier, Orders of convergence in the averaging principle for SPDEs: the case of a stochastically forced slow component, Stochastic Process. Appl., 130 (2020), pp. 3325–3368.

[4] S. Cerrai, A. Lunardi, Averaging principle for nonautonomous slow-fast systems of stochastic reaction-diffusion equations: the almost periodic case, SIAM J. Math. Anal., 49 (2017), pp. 2843-2884.

[5] W. E, Principles of Multiscale Modeling, Cambridge University Press, Cambridge, 2011.

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

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

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

[9] H. Fu, J. Liu, Strong convergence in stochastic averaging principle for two time-scales stochastic partial differential equations, J. Math. Anal. Appl., 384 (2011), pp. 70-86.

[10] H. Gao, Y. Shi, Averaging principle for a stochastic coupled fast-slow atmosphere-ocean model, J. Differential Equations, 298 (2021), pp. 248-297.

[11] D. Givon, Strong convergence rate for two-time-scale jump-diffusion stochastic differential systems, Multiscale Model. Simul., 6 (2007), pp. 577–594.

[12] D. Givon, I. G. Kevrekidis, Multiscale integration schemes for jump-diffusion systems, Multiscale Model. Simul., 7 (2008), pp. 495–516.

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

[14] J. Glimm, D. H. Sharp, Multiscale science: A challenge for the twenty-first century, Advances in Mechanics, 28 (1998), pp. 545–551.

[15] J. Golec, Stochastic averaging principle for systems with pathwise uniqueness, Stochastic Anal. Appl., 13 (1995), pp. 307–322.

[16] J. Golec, G. Ladde, Averaging principle and systems of singularly perturbed stochastic differential equations, J. Math. Phys., 31 (1990), pp. 1116–1123.

[17] E. Harvey, V. Kirk, M. Wechselberger, J. Sneyd, Multiple timescales, mixed mode oscillations and canards in models of intracellular calcium dynamics, J. Nonlinear Sci., 21 (2011), pp. 639–683.

[18] D. J. Higham, X. Mao, A. M. Stuart, Strong convergence of Euler-type methods for nonlinear stochastic differential equations, SIAM J. Numer. Anal., 40 (2002), pp. 1041–1063.

[19] W. Hong, S. Li, W. Liu, Strong convergence rates in averaging principle for slow-fast McKean-Vlasov SPDEs, J. Differential Equations, 316 (2022), pp. 94-135.

[20] M. Hutzenthaler, A. Jentzen, Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients, Mem. Amer. Math. Soc., 236 (2015), 99 pp.

[21] M. Hutzenthaler, A. Jentzen, P. E. Kloeden, Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 467 (2011), pp. 1563–1576.

[22] M. Hutzenthaler, A. Jentzen, P. E. Kloeden, Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients, Ann. Appl. Probab., 22 (2012), pp. 1611–1641.

[23] R. Z. Khassminskii, On the principle of averaging the Itô's stochastic differential equations, Kybernetika (Prague), 4 (1968), pp. 260–279.

[24] P. E. Kloeden, E. Platen, Numerical Solution of Stochastic Differential Equations, Springer, Berlin, 1992.

[25] C. Kuehn, Multiple Time Scale Dynamics, Springer, Cham, 2015.

[26] A. L. Lewis, Option Valuation under Stochastic Volatility, Finance Press, Newport Beach, CA, 2000.

[27] X. Li, X. Mao, H. Yang, Strong convergence and asymptotic stability of explicit numerical schemes for nonlinear stochastic differential equations, Math. Comp., 90 (2021), pp. 2827–2872.

[28] X. Li, X. Mao, G. Yin, Explicit numerical approximations for stochastic differential equations in finite and infinite horizons: truncation methods, convergence in pth moment and stability, IMA J. Numer. Anal., 39 (2019), pp. 847–892.

[29] D. Liu, Analysis of multiscale methods for stochastic dynamical systems with multiple time scales, Multiscale Model. Simul., 8 (2010), pp. 944–964.

[30] D. Liu, Strong convergence of principle of averaging for multiscale stochastic dynamical systems, Commun. Math. Sci., 8 (2010), pp. 999–1020.

[31] W. Liu, X. Mao, Strong convergence of the stopped Euler-Maruyama method for nonlinear stochastic differential equations, Appl. Math. Comput., 223 (2013), pp. 389–400.

[32] W. Liu, M. Röckner, X. Sun, Y. Xie, Averaging principle for slow-fast stochastic differential equations with time dependent locally Lipschitz coefficients, J. Differential Equations, 268 (2020), pp. 2910–2948.

[33] X. Mao, Stochastic Differential Equations and Applications, second ed., Horwood Publishing Limited, Chichester, 2008.

[34] X. Mao, The truncated Euler-Maruyama method for stochastic differential equations, J. Comput. Appl. Math., 290 (2015), pp. 370–384.

[35] X. Mao, C. Yuan, G. Yin, Numerical method for stationary distribution of stochastic differential equations with Markovian switching, J. Comput. Appl. Math., 174 (2005), pp. 1–27.

[36] E. Papageorgiou, R. Sircar, Multiscale intensity models for single name credit derivatives, Appl. Math. Finance, 15 (2008), pp. 73–105.

[37] G. A. Pavliotis, A. M. Stuart, Multiscale Methods: Averaging and Homogenization, Springer, 2007.

[38] M. Röckner, X. Sun, Y. Xie, Strong convergence order for slow-fast McKean-Vlasov stochastic differential equations, Ann. Inst. Henri Poincaré Probab. Stat., 57 (2021), pp. 547-576.

[39] S. Sabanis, A note on tamed Euler approximations, Electron. Commun. Probab., 18 (2013), 10 pp.

[40] S. Sabanis, Euler approximations with varying coefficients: the case of superlinearly growing diffusion coefficients, Ann. Appl. Probab., 26 (2016), pp. 2083–2105.

[41] Y. Shi, X. Sun, L. Wang, Y. Xie, Asymptotic behavior for multi-scale SDEs with monotonicity coefficients driven by Levy processes, Potential Anal, 61 (2023), 111–152.

[42] A. N. Shiryaev, Probability, second ed., Springer-Verlag, New York, 1996.

[43] E. Vanden-Eijnden, Numerical techniques for multi-scale dynamical systems with stochastic effects, Commun. Math. Sci., 1 (2003), pp. 385–391.

[44] C. Villani, Optimal transport, old and new, Springer, 2009.

[45] A. Y. Veretennikov, On the averaging principle for systems of stochastic differential equations, Math. USSR Sb., 69 (1991), pp. 271–284.

[46] A. Y. Veretennikov, On large deviations in the averaging principle for SDEs with "a full dependence", Ann. Probab., 27 (1999), pp. 284-296.

[47] X. Wang, S. Gan, The tamed Milstein method for commutative stochastic differential equations with non-globally Lipschitz continuous coefficients, J. Difference Equ. Appl., 19 (2013), pp. 466–490.

[48] F. Wu, T. Tian, J. B. Rawlings, G. Yin, Approximate method for stochastic chemical kinetics with two-time scales by chemical langevin equations, J. Chem. Phys., 144 (2016): 174112.

[49] Y. Xu, J. Duan, W. Xu, An averaging principle for stochastic dynamical systems with Lévy noise, Phys. D, 240 (2011), pp. 1395–1401.