Convergence Rate of Euler-Maruyama Scheme for SDEs with Hölder-Dini Continuous Drifts *

Jianhai Bao[b)], **Xing Huang**[a)], **Chenggui Yuan**[b)]

[a)]School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China

hxsc19880409@163.com

[b)]Department of Mathematics, Swansea University, Singleton Park, SA2 8PP, UK

Jianhai.Bao@Swansea.ac.uk, c.yuan@swansea.ac.uk

Abstract

In this paper, we are concerned with convergence rate of Euler-Maruyama scheme for stochastic differential equations with Hölder-Dini continuous drift. The key contributions lie in (i), by means of regularity of non-degenerate Kolmogrov equation, we investigate convergence rate of Euler-Maruyama scheme for a class of stochastic differential equations, which allow the drifts to be Dini-continuous and unbounded; (ii) by the aid of regularization properties of degenerate Kolmogrov equation, we discuss convergence rate of Euler-Maruyama scheme for a range of degenerate stochastic differential equations, where the drift is Hölder-Dini continuous of order $\frac{2}{3}$ with respect to the first component, and is merely Dini-continuous concerning the second component.

AMS subject Classification: 60H35 · 41A25 · 60H10 · 60C30
Keywords: Euler-Maruyama scheme · convergence rate · Hölder-Dini continuity · degenerate stochastic differential equation · Kolmogorov equation

1 Introduction and Main Results

In their paper [24], Wang and Zhang studied existence and uniqueness for a class of stochastic differential equations (SDEs) with Hölder-Dini continuous drifts; Wang [23] also investigated the strong Feller property, log-Harnack inequality and gradient estimates for SDEs with Dini continuous drift. So far there are no numerical schemes available for SDEs

*Supported in part by NNSFC(11431014)

with Hölder-Dini continuous drifts. So the aim of this paper is to prove the convergence of Euler-Maruyama (EM) scheme and obtain the rate of convergence for these equations under reasonable conditions.

It is well-known that convergence rate of Euler-Maruyama for SDEs with regular coefficients is one-half, see, e.g., [12]. With regard to convergence rate of EM scheme under various settings, we refer to, e.g., [1] for stochastic differential delay equations (SDDEs) with polynomial growth with respect to (w.r.t.) the delay variables, [6] for SDDEs under local Lipschitz and also under monotonicity condition, [15] for SDEs with discontinuous coefficients, and [26] for SDEs under log-Lipschitz condition. Whereas, for SDEs with non-globally Lipschitz continuous coefficients, see, e.g., [3, 8, 9, 10], to name a few. On the other hand, Hairer et al. [7] have established the first result in the literature that Euler's method converges to the solution of an SDE with smooth coefficients in the strong and numerical weak sense without any arbitrarily small polynomial rate of convergence, and Jentzen et al. [11] have further given a counterexample that no approximation method converges to the true solution in the mean square sense with polynomial rate.

The rate of convergence of EM scheme for SDEs with irregular coefficients has also gained much attention. For instance, Adopting the Yamada-Watanabe approximation approach, [5] discussed strong convergence rate in L^p-norm sense; Using the Yamada-Watanabe approximation trick and heat kernel estimate, [17] studied strong convergence rate in L^1-norm sense for a class of non-degenerate SDEs, where the bounded drift term satisfies a weak monotonicity and is of bounded variation w.r.t. a Gaussian measure and the diffusion term is Hölder continuous; Applying the Zvonkin transformation, [19] discussed strong convergence rate in L^p-norm sense for SDEs with additive noise, where the drift coefficient is bounded and Hölder continuous.

It is worth pointing out that [17, 19] focused on convergence rate of EM for SDEs with *Hölder continuous and bounded drifts*, which rules out Hölder-Dini continuous and unbounded drifts. On the other hand, most of the existing literature on convergence rate of EM scheme is concerned with *non-degenerate SDEs*. Yet the corresponding issue for *degenerate SDEs* is scarce, to the best of our knowledge. So, in this work, we will not only investigate the convergence of the EM scheme for SDEs with Hölder-Dini continuous drift, but will also study the degenerate cases. For wellposedness of SDEs with singular coefficients, we refer to, e.g., [14, 23, 24, 28] for more details.

Throughout the paper, the following notation will be used. Let n, m be positive integers, $(\mathbb{R}^n, \langle\cdot,\cdot\rangle, |\cdot|)$ the n-dimensional Euclidean space, and $\mathbb{R}^n \otimes \mathbb{R}^m$ the family of all $n \times m$ matrices. Let $\|\cdot\|$ and $\|\cdot\|_{\mathrm{HS}}$ stand for the usual operator norm and the Hilbert-Schmidt norm, respectively. Fix $T > 0$ and set $\|f\|_{T,\infty} := \sup_{t\in[0,T],x\in\mathbb{R}^m} \|f(t,x)\|$ for an operator-valued map f on $[0,T]\times\mathbb{R}^m$. $C(\mathbb{R}^m;\mathbb{R}^n)$ means the continuous functions $f : \mathbb{R}^m \to \mathbb{R}^n$. Let $C^2(\mathbb{R}^n;\mathbb{R}^n\otimes\mathbb{R}^n)$ be the family of all continuously twice differentiable functions $f : \mathbb{R}^n \to \mathbb{R}^n \otimes \mathbb{R}^n$. Denote $\mathbb{M}^n_{\mathrm{non}}$ by the collection of all nonsingular $n\times n$-matrices. Let $\mathscr{S}_0$ be the collection of all slowly varying functions $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ at zero in Karamata's sense (i.e., $\lim_{t\to0} \frac{\phi(\lambda t)}{\phi(t)} = 1$ for any $\lambda > 0$), which are bounded from 0 and ∞ on $[\varepsilon,\infty)$ for any $\varepsilon > 0$.

Let $\mathscr{D}_0$ be the family of Dini functions, i.e.,

$$\mathscr{D}_0 := \left\{\phi\middle|\phi : \mathbb{R}_+ \to \mathbb{R}_+ \text{ is increasing and } \int_0^1 \frac{\phi(s)}{s}\mathrm{d}s < \infty\right\}.$$

A function $f : \mathbb{R}^m \to \mathbb{R}^n$ is called Dini-continuity if there exists $\phi \in \mathscr{D}_0$ such that $|f(x) - f(y)| \le \phi(|x-y|)$ for any $x, y \in \mathbb{R}^m$. We remark that every Dini-continuous function is continuous and every Lipschitz continuous function is Dini-continuous; Moreover, if f is Hölder continuous, then f is Dini-continuous. Nevertheless, there are numerous Dini-continuous functions, which are not Hölder continuous at all; see, e.g.,

$$\phi(x) = \begin{cases} \frac{1}{(\log(c+x^{-1}))^{(1+\delta)}}, & x > 0 \\ 0, & x = 0 \end{cases}$$

for some constants $\delta > 0$ and $c \ge \mathrm{e}^{3+2\delta}$. Set

$$\mathscr{D} := \{\phi \in \mathscr{D}_0 | \phi^2 \text{ is concave}\} \quad \text{and} \quad \mathscr{D}^\varepsilon := \{\phi \in \mathscr{D} | \phi^{2(1+\varepsilon)} \text{ is concave}\}$$

for some $\varepsilon \in (0,1)$ sufficiently small. Clearly, ϕ constructed above belongs to $\mathscr{D}^\varepsilon$. A function $f : \mathbb{R}^m \to \mathbb{R}^n$ is called Hölder-Dini continuity of order $\alpha \in [0,1)$ if

$$|f(x) - f(y)| \le |x-y|^\alpha \phi(|x-y|), \qquad |x-y| \le 1$$

for some $\phi \in \mathscr{D}_0$; see, for instance,

$$f(x) = \begin{cases} \frac{1}{(1+x)^\alpha(\log(c+x^{-1}))^{(1+\delta)}}, & x > 0 \\ 0, & x = 0 \end{cases}$$

for some constants $c, \delta > 0$ and $\alpha \in (0,1)$.

Before proceeding further, a few words about the notation are in order. Generic constants will be denoted by c; we use the shorthand notation $a \lesssim b$ to mean $a \le cb$. If the constant c depends on a parameter p, we shall also write c_p and $a \lesssim_p b$.

1.1 Non-degenerate SDEs with Bounded Coefficients

In this sebsection, we consider an SDE on $(\mathbb{R}^n, \langle\cdot,\cdot\rangle, |\cdot|)$

$$\text{(1.1)} \qquad \mathrm{d}X_t = b_t(X_t)\mathrm{d}t + \sigma_t(X_t)\mathrm{d}W_t, \quad t > 0, \quad X_0 = x,$$

where $b : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n \otimes \mathbb{R}^n$, and $(W_t)_{t\ge0}$ is an n-dimensional Brownian motion defined on a complete filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\ge0}, \mathbb{P})$.

With regard to (1.1), we suppose that there exists $\phi \in \mathscr{D}$ such that for any $s, t \in [0, T]$ and $x, y \in \mathbb{R}^n$,

(A1) $\sigma_t \in C^2(\mathbb{R}^n;\mathbb{R}^n\otimes\mathbb{R}^n)$, $\sigma_t(x)\in\mathbb{M}^n_{\rm non}$, and

$$\|b\|_{T,\infty}+\sum_{i=0}^{2}\|\nabla^i\sigma\|_{T,\infty}+\|\nabla\sigma^{-1}\|_{T,\infty}+\|\sigma^{-1}\|_{T,\infty}<\infty, \tag{1.2}$$

where ∇^i means the i-th order gradient operator;

(A2) (Regularity of b w.r.t. spatial variables)

$$|b_t(x)-b_t(y)|\le\phi(|x-y|);$$

(A3) (Regularity of b and σ w.r.t. time variables)

$$|b_s(x)-b_t(x)|+\|\sigma_s(x)-\sigma_t(x)\|_{\rm HS}\le\phi(|s-t|).$$

Under **(A1)** and **(A2)**, (1.1) admits a unique non-explosive strong solution $(X_t)_{t\in[0,T]}$; see, e.g., [23, Theorem 1.1].

Without loss of generality, we take an integer $N>0$ sufficiently large such that the stepsize $\delta:=T/N\in(0,1)$. The continuous-time EM scheme corresponding to (1.1) is

$$\mathrm{d}Y_t=b_{t_\delta}(Y_{t_\delta})\mathrm{d}t+\sigma_{t_\delta}(Y_{t_\delta})\mathrm{d}W_t,\quad t>0,\quad Y_0=X_0=x. \tag{1.3}$$

Herein, $t_\delta:=\lfloor t/\delta\rfloor\delta$ with $\lfloor t/\delta\rfloor$ being the integer part of t/δ.

The first contribution in this paper is stated as follows.

Theorem 1.1. Let **(A1)-(A3)** hold. Then

$$\Big(\mathbb{E}\Big(\sup_{0\le t\le T}|X_t-Y_t|^2\Big)\Big)^{1/2}\lesssim_T\phi(C_T\sqrt{\delta})$$

for some constant $C_T\ge1$.

Remark 1.2. In Theorem 1.1, by taking $\phi(x)=x^\beta$ for $x>0$ and $\beta\in(0,1]$, and inspecting closely the argument of Theorem 1.1, the concave property of ϕ^2 can be dropped. Moreover, we have

$$\mathbb{E}\Big(\sup_{0\le t\le T}|X_t-Y_t|^2\Big)\lesssim_T\delta^\beta.$$

So, our present result covers [19, Theorem 2.13], where the drift is only Hölder continuous. In particular, for $\beta=1$, it reduces to the classical result on strong convergence of EM scheme for SDEs with regular coefficients; see, e.g., [12].

1.2 Non-degenerate SDEs with Unbounded Coefficients

As we can see, in Theorem 1.1, the coefficients are uniformly bounded, and that the drift term b satisfies the global Dini-continuous condition (see **(A2)** above), which seems to be a little bit stringent. Therefore, concerning the coefficients, it is quite natural to replace uniform boundedness by local boundedness and global Dini continuity by local Dini continuity, respectively.

In lieu of **(A1)-(A3)**, as for (1.1) we assume that for any $s,t\in[0,T]$ and $k\ge 1$,

(A1') $\sigma_t\in C^2(\mathbb{R}^n;\mathbb{R}^n\otimes\mathbb{R}^n)$, for every $x\in\mathbb{R}^n$, $\sigma_t(x)\in\mathbb{M}^n_{\rm non}$, and

$$|b_t(x)|+\sum_{i=0}^2\|\nabla^i\sigma_t(x)\|_{\rm HS}+\|\nabla\sigma_t^{-1}(x)\|_{\rm HS}+\|\sigma_t^{-1}(x)\|_{\rm HS}\le K_T(1+|x|),\quad x\in\mathbb{R}^n$$

for some constant $K_T>0$;

(A2') (Regularity of b w.r.t. spatial variables) There exists $\phi_k\in\mathscr{D}$ such that

$$|b_t(x)-b_t(y)|\le\phi_k(|x-y|),\quad |x|\vee|y|\le k;$$

(A3') (Regularity of b and σ w.r.t. time variables) For $\phi_k\in\mathscr{D}$ such that **(A2')**,

$$|b_s(x)-b_t(x)|+\|\sigma_s(x)-\sigma_t(x)\|_{\rm HS}\le\phi_k(|s-t|),\qquad |x|\le k.$$

By the cut-off approach, Theorem 1.1 can be extended to include SDEs with local Dini-continuous coefficients, which is presented as below.

Theorem 1.3. Assume **(A1')-(A3')** hold. Then it holds that

$$\lim_{\delta\to0}\mathbb{E}\Big(\sup_{0\le t\le T}|X_t-Y_t|^2\Big)=0. \tag{1.4}$$

In particular, if $\phi_k(s)=\mathrm{e}^{\mathrm{e}^{c_0k^4}}s^\alpha$, $s\ge0$, for some $\alpha\in(0,1]$ and $c_0>0$, then

$$\mathbb{E}\Big(\sup_{0\le t\le T}|X_t-Y_t|^2\Big)\lesssim\inf_{\varepsilon\in(0,1)}\Big\{(\log\log(\delta^{-\alpha\varepsilon}))^{-\frac14}+\delta^{\alpha(1-\varepsilon)}\Big\}. \tag{1.5}$$

Remark 1.4. Theorem 1.3 has improved the result in [18] since the drift involved is allowed to be unbounded and local Dini continuous, while the drift in [18] is merely bounded and Hölder continuous.

1.3 Degenerate SDEs

So far, most of the existing literature on convergence of EM scheme for SDEs with irregular coefficients is concerned with non-degenerate SDEs; see, e.g., [17, 18, 19] for SDEs driven by Brownian motions, and [19] for SDEs driven by jump processes. The issue for the setup of degenerate SDEs has not yet been considered to date to the best of our knowledge. Nevertheless, in this subsection, we make an attempt to discuss the topic for degenerate SDEs with Hölder-Dini continuous drift.

Consider the following degenerate SDE on $\mathbb{R}^{2n}$

$$\begin{cases} \mathrm{d}X_t^{(1)} = b_t^{(1)}(X_t^{(1)}, X_t^{(2)})\mathrm{d}t, & X_0^{(1)} = x^{(1)} \in \mathbb{R}^n, \\ \mathrm{d}X_t^{(2)} = b_t^{(2)}(X_t^{(1)}, X_t^{(2)})\mathrm{d}t + \sigma_t(X_t^{(1)}, X_t^{(2)})\mathrm{d}W_t, & X_0^{(2)} = x^{(2)} \in \mathbb{R}^n, \end{cases} \tag{1.6}$$

where $b_t^{(1)}, b_t^{(2)} : \mathbb{R}^{2n} \to \mathbb{R}^n$, $\sigma_t : \mathbb{R}^{2n} \to \mathbb{R}^n \otimes \mathbb{R}^n$, and $(W_t)_{t\geq 0}$ is an n-dimensional Brownian motion defined on a complete filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$. (1.6) is also called a stochastic Hamiltonian system, which has been investigated extensively in [25, 27] on Bismut formulae, in [16] on ergodicity, in [22] on hypercontractivity, and in [24] on wellposedness, to name a few. For applications of the model (1.6), we refer to, e.g., Soize [21].

For notational simplicity, we shall write $\mathbb{R}^{2n}$ instead of $\mathbb{R}^n \times \mathbb{R}^n$. Write the gradient operator on $\mathbb{R}^{2n}$ as $\nabla = (\nabla^{(1)}, \nabla^{(2)})$, where $\nabla^{(1)}$ and $\nabla^{(2)}$ stand for the gradient operators w.r.t. the first and the second components, respectively.

We assume that there exists $\phi \in \mathscr{D}^\varepsilon \cap \mathscr{S}_0$ such that for any $x = (x^{(1)}, x^{(2)}), y = (y^{(1)}, y^{(2)}) \in \mathbb{R}^{2n}$ and $s, t \in [0, T]$,

(C1) (Hypoellipticity) $[\nabla^{(2)} b_t^{(1)}(x)], \sigma_t(x) \in \mathbb{M}^n_{\text{non}}$, and

$$\begin{aligned} &\|b^{(1)}\|_{T,\infty} + \|b^{(2)}\|_{T,\infty} + \|\nabla^{(2)} b^{(1)}\|_{T,\infty} + \big\|[\nabla^{(2)} b^{(1)}]^{-1}\big\|_{T,\infty} \\ &+ \|\sigma\|_{T,\infty} + \|\nabla\sigma\|_{T,\infty} + \|\sigma^{-1}\|_{T,\infty} < \infty; \end{aligned}$$

(C2) (Regularity of $b^{(1)}$ w.r.t. spatial variables)

$$\begin{aligned} &|b_t^{(1)}(x) - b_t^{(1)}(y)| \leq |x^{(1)} - y^{(1)}|^{\frac{2}{3}} \phi(|x^{(1)} - y^{(1)}|) && \text{if } x^{(2)} = y^{(2)}, \\ &\|(\nabla^{(2)} b_t^{(1)})(x) - (\nabla^{(2)} b_t^{(1)})(y)\|_{\text{HS}} \leq \phi(|x^{(2)} - y^{(2)}|) && \text{if } x^{(1)} = y^{(1)}; \end{aligned}$$

(C3) (Regularity of $b^{(2)}$ w.r.t. spatial variables)

$$|b_t^{(2)}(x) - b_t^{(2)}(y)| \leq |x^{(1)} - y^{(1)}|^{\frac{2}{3}} \phi(|x^{(1)} - y^{(1)}|) + \phi^{\frac{7}{2}}(|x^{(2)} - y^{(2)}|);$$

(C4) (Regularity of $b^{(1)}, b^{(2)}$ and σ w.r.t. time variables)

$$|b_t^{(1)}(x) - b_s^{(1)}(x)| + |b_t^{(2)}(x) - b_s^{(2)}(x)| + \|\sigma_t(x) - \sigma_s(x)\|_{\text{HS}} \leq \phi(|t - s|).$$

Observe from **(C2)** and **(C3)** that $b^{(1)}(\cdot, x^{(2)})$ and $b^{(2)}(\cdot, x^{(2)})$ with fixed $x^{(2)}$ are locally Hölder-Dini continuous of order $\frac{2}{3}$, and $(\nabla^{(2)}b^{(1)})(x^{(1)}, \cdot)$ and $b^{(2)}(x^{(1)}, \cdot)$ with fixed $x^{(1)}$ are merely Dini continuous. According to [24, Theorem 1.2], (1.6) admits a unique strong solution under the assumptions **(C1)-(C3)**. In fact, (1.6) is wellposed under **(C1)-(C3)** with $\phi \in \mathscr{D}_0 \cap \mathscr{S}_0$ in lieu of $\phi \in \mathscr{D}^\varepsilon \cap \mathscr{S}_0$. Nevertheless, the requirement $\phi \in \mathscr{D}^\varepsilon \cap \mathscr{S}_0$ is imposed in order to reveal the order of convergence for the EM scheme below.

The continuous-time EM scheme associated with (1.6) is as follows:

$$
\begin{cases} \mathrm{d}Y_t^{(1)} = b_{t_\delta}^{(1)}(Y_{t_\delta}^{(1)}, Y_{t_\delta}^{(2)})\mathrm{d}t, & X_0^{(1)} = x^{(1)} \in \mathbb{R}^n, \\ \mathrm{d}Y_t^{(2)} = b_{t_\delta}^{(2)}(Y_{t_\delta}^{(1)}, Y_{t_\delta}^{(2)})\mathrm{d}t + \sigma_{t_\delta}(Y_{t_\delta}^{(1)}, Y_{t_\delta}^{(2)})\mathrm{d}W_t, & X_0^{(2)} = x^{(2)} \in \mathbb{R}^n, \end{cases} \tag{1.7}
$$

where t_δ is defined as in (1.3).

Another contribution in this paper reads as below.

Theorem 1.5. Let **(C1)-(C4)** hold. Then

$$
\Big(\mathbb{E}\Big(\sup_{0\le t\le T}|X_t - Y_t|^2\Big)\Big)^{1/2} \lesssim_T \phi(C_T\sqrt{\delta})
$$

for some constant $C_T \ge 1$, in which

$$
X_t := \begin{pmatrix} X_t^{(1)} \\ X_t^{(2)} \end{pmatrix} \text{ and } Y_t := \begin{pmatrix} Y_t^{(1)} \\ Y_t^{(2)} \end{pmatrix}.
$$

Remark 1.6. By applying the cut-off approach and refining the argument of [24, Theorem 2.3] (see also Lemma 5.1 below), the boundedness of coefficients can be removed. We herein do not go into details since the corresponding trick is quite similar to the proof of Theorem 1.3.

The outline of this paper is organized as follows: In Section 2, we elaborate regularity of nondegenerate Kolmogorov equation, which plays an important role in dealing with convergence rate of EM scheme for nondegenerate SDEs with Hölder-Dini continuous and unbounded drifts; In Sections 3, 4 and 5, we complete the proofs of Theorems 1.1, 1.3 and 1.5, respectively.

2 Regularity of Non-degenerate Kolmogorov Equation

Let $(e_i)_{i\ge1}$ be an orthogonal basis of $\mathbb{R}^n$. For any $\lambda > 0$, consider the following $\mathbb{R}^n$-valued parabolic equation:

$$
\partial_t u_t^\lambda + L_t u_t^\lambda + b_t + \nabla_{b_t} u_t^\lambda = \lambda u_t^\lambda, \quad u_T^\lambda = \mathbf{0_n}, \tag{2.1}
$$

where $\nabla_{b_t} u_t^\lambda$ means the directional derivative along the direction b_t, $\mathbf{0_n}$ is the zero vector in $\mathbb{R}^n$ and

$$
L_t := \frac{1}{2}\sum_{i,j}\langle(\sigma_t\sigma_t^*)(\cdot)e_i, e_j\rangle\nabla_{e_i}\nabla_{e_j}
$$

with σ_t^* standing for the transpose of σ_t. Consider the coupled forward-backward SDE

$$\begin{cases}\mathrm{d}Z_t^{s,x}=\sigma_t(Z_t^{s,x})\mathrm{d}W_t, \quad Z_s^{s,x}=x,\\ \mathrm{d}Y_t^{\lambda,s,x}=\{\lambda Y_t^{\lambda,s,x}-b(Z_t^{s,x})-\nabla_{b(Z_t^{s,x})}Y_t^{\lambda,s,x}\}\mathrm{d}t+Z_t^{s,x}\mathrm{d}W_t, \quad Y_T^{\lambda,s,x}=\mathbf{0_n}\end{cases} \tag{2.2}$$

for any $t\in[s,T]\subset[0,T]$. By the chain rule, it follows from (2.2) that

$$\begin{aligned}\mathrm{d}(\mathrm{e}^{-\lambda t}Y_t^{\lambda,s,x})&=-\lambda\mathrm{e}^{-\lambda t}Y_t^{\lambda,s,x}\mathrm{d}t+\mathrm{e}^{-\lambda t}\mathrm{d}Y_t^{\lambda,s,x}\\ &=-\mathrm{e}^{\lambda t}\{b(Z_t^{s,x})+\nabla_{b(Z_t^{s,x})}Y_t^{\lambda,s,x}\}\mathrm{d}t+\mathrm{e}^{\lambda t}Z_t^{s,x}\mathrm{d}W_t.\end{aligned}$$

Integrating from t to T and, in particular, taking $s=t$ yields that

$$\mathbb{E}Y_t^{\lambda,t,x}=\int_t^T\mathrm{e}^{-\lambda(r-t)}\mathbb{E}\{b(Z_r^{t,x})+\nabla_{b(Z_r^{t,x})}Y_r^{\lambda,t,x}\}\mathrm{d}r.$$

Noting from [2, Theorem 5.5] that $Y_t^{\lambda,t,x}=u_t^\lambda(Z_t^{t,x})$, $\mathbb{P}$-a.s., and $Z_t^{t,x}=x$, so we arrive at

$$u_s^\lambda=\int_s^T\mathrm{e}^{-\lambda(t-s)}P_{s,t}^0\{b_t+\nabla_{b_t}u_t^\lambda\}\mathrm{d}t, \tag{2.3}$$

where the semigroup $(P_{s,t}^0)_{0\le s\le t}$ is generated by $(Z_t^{s,x})_{0\le s\le t}$ which solves an SDE below

$$\mathrm{d}Z_t^{s,x}=\sigma_t(Z_t^{s,x})\mathrm{d}W_t, \quad t>s, \quad Z_s^{s,x}=x. \tag{2.4}$$

For notational simplicity, let

$$\Lambda_{T,\sigma}=\mathrm{e}^{\frac{T}{2}\|\nabla\sigma\|_{T,\infty}^2}\|\sigma^{-1}\|_{T,\infty} \tag{2.5}$$

and

$$\begin{aligned}\tilde\Lambda_{T,\sigma}=48\mathrm{e}^{288\,T^2\|\nabla\sigma\|_{T,\infty}^4}\Big\{&6\sqrt{2}\mathrm{e}^{T\|\nabla\sigma\|_{T,\infty}^2}\|\sigma^{-1}\|_{T,\infty}^4+T\|\nabla\sigma^{-1}\|_{T,\infty}^2\\ &+2T^2\|\nabla^2\sigma\|_{T,\infty}^2\|\sigma^{-1}\|_{T,\infty}^2\mathrm{e}^{2T\|\nabla\sigma\|_{T,\infty}^2}\Big\}.\end{aligned} \tag{2.6}$$

Moreover, set

$$\Upsilon_{T,\sigma}:=\sqrt{\tilde\Lambda_{T,\sigma}}\Big\{3+2\|b\|_{T,\infty}+28\Big(\Lambda_{T,\sigma}+\sqrt{\tilde\Lambda_{T,\sigma}}\Big)\|b\|_{T,\infty}^2\Big\}. \tag{2.7}$$

The lemma below plays a crucial role in investigating error analysis.

Lemma 2.1. Under **(A1)** and **(A2)**, for any $\lambda\ge 9\pi\Lambda_{T,\sigma}^2\|b\|_{T,\infty}^2+4(\|b\|_{T,\infty}+\Lambda_{T,\sigma})^2$,

- **(i)** (2.1) (i.e., (2.3)) enjoys a unique strong solution $u^\lambda\in C([0,T];C_b^1(\mathbb{R}^n;\mathbb{R}^n))$;
- **(ii)** $\|\nabla u^\lambda\|_{T,\infty}\le\frac{1}{2}$;
- **(iii)** $\|\nabla^2u^\lambda\|_{T,\infty}\le\Upsilon_{T,\sigma}\int_0^T\frac{\mathrm{e}^{-\lambda t}}{t}\tilde\phi(\|\sigma\|_{T,\infty}\sqrt{t})\mathrm{d}t$, where $\tilde\phi(s):=\sqrt{\phi^2(s)+s}$, $s\ge0$.

Proof. To show **(i)**-**(iii)**, it boils down to refine the argument of [23, Lemma 2.1]. **(i)** holds for any $\lambda \geq 4(\|b\|_{T,\infty} + \Lambda_{T,\sigma})^2$ via the Banach fixed-point theorem.

In what follows, we aim to show **(ii)** and **(iii)** one-by-one. Observe from [13, Theorem 3.1, p.218] that

$$\mathrm{d}\nabla_\eta Z_t^{s,x} = (\nabla_{\nabla_\eta Z_t^{s,x}}\sigma_t)(Z_t^{s,x})\mathrm{d}W_t, \quad t \geq s, \quad \nabla_\eta Z_s^{s,x} = \eta \in \mathbb{R}^n. \tag{2.8}$$

Using Itô's isometry and Gronwall's inequality, one has

$$\mathbb{E}|\nabla_\eta Z_t^{s,x}|^2 \leq |\eta|^2 \mathrm{e}^{T\|\nabla\sigma\|_{T,\infty}^2}. \tag{2.9}$$

Utilizing the BDG inequality, we deduce that

$$\mathbb{E}|\nabla_\eta Z_t^{s,x}|^4 \leq 8\Big\{|\eta|^4 + 36(t-s)\|\nabla\sigma\|_{T,\infty}^4 \int_s^t \mathbb{E}|\nabla_\eta Z_u^{s,x}|^4 \mathrm{d}u\Big\},$$

which, combining with Gronwall's inequality, yields that

$$\mathbb{E}|\nabla_\eta Z_t^{s,x}|^4 \leq 8|\eta|^4 \mathrm{e}^{288\,T^2\|\nabla\sigma\|_{T,\infty}^4}. \tag{2.10}$$

Recall from [23, (2.8)] that the Bismut formula below

$$\nabla_\eta P_{s,t}^0 f(x) = \mathbb{E}\Big(\frac{f(Z_t^{s,x})}{t-s}\int_s^t \langle \sigma_r^{-1}(Z_r^{s,x})\nabla_\eta Z_r^{s,x}, \mathrm{d}W_r\rangle\Big), \quad f \in \mathscr{B}_b(\mathbb{R}^n) \tag{2.11}$$

holds. By the Cauchy-Schwartz inequality, the Itô isometry and (2.9), we obtain that

$$|\nabla_\eta P_{s,t}^0 f|^2(x) \leq \frac{\Lambda_{T,\sigma}^2|\eta|^2 P_{s,t}^0 f^2(x)}{t-s}, \quad f \in \mathscr{B}_b(\mathbb{R}^n), \tag{2.12}$$

where $\Lambda_{T,\sigma} > 0$ is defined in (2.5). So, one infers from (2.3) and (2.12) that

$$\begin{aligned}\|\nabla u_s^\lambda\| &\leq \int_s^T \mathrm{e}^{-\lambda(t-s)}\|\nabla P_{s,t}^0\{b_t + \nabla_{b_t}u_t^\lambda\}\|\mathrm{d}t \\ &\leq \Lambda_{T,\sigma}(1 + \|\nabla u^\lambda\|_{T,\infty})\|b\|_{T,\infty}\int_0^T \frac{\mathrm{e}^{-\lambda t}}{\sqrt{t}}\mathrm{d}t \\ &\leq \lambda^{-\frac{1}{2}}\sqrt{\pi}\Lambda_{T,\sigma}\|b\|_{T,\infty}(1 + \|\nabla u^\lambda\|_{T,\infty}).\end{aligned}$$

Thus, **(ii)** follows by taking $\lambda \geq 9\pi\Lambda_{T,\sigma}^2\|b\|_{T,\infty}^2$.

In the sequel, we intend to verify **(iii)**. Set $\gamma_{s,t} := \nabla_\eta\nabla_{\eta'}Z_t^{s,x}$ for any $\eta, \eta' \in \mathbb{R}^n$. Notice from (2.8) that

$$\mathrm{d}\gamma_{s,t} = \big\{\nabla_{\gamma_{s,t}}\sigma_t(Z_t^{s,x}) + \nabla_{\nabla_\eta Z_t^{s,x}}\nabla_{\nabla_{\eta'}Z_t^{s,x}}\sigma_t(Z_t^{s,x})\big\}\mathrm{d}W_t, \quad t \geq s, \quad \gamma_{s,s} = \mathbf{0_n}.$$

By the Doob submartingale inequality and the Itô isometry, besides the Gronwall inequality and (2.9), we get that

$$\sup_{s\leq t\leq T}\mathbb{E}|\gamma_{s,t}|^2 \leq 16T\|\nabla^2\sigma\|_{T,\infty}^2 \mathrm{e}^{288T^2\|\nabla\sigma\|_{T,\infty}^4 + 2T\|\nabla\sigma\|_{T,\infty}^2}|\eta|^2|\eta'|^2. \tag{2.13}$$

From (2.11) and the Markov property, we have

$$\nabla_\eta P^0_{s,t} f(x) = \mathbb{E}\bigg(\frac{(P^0_{\frac{t+s}{2},t} f)(Z^{s,x}_{\frac{t+s}{2}})}{(t-s)/2} \int_s^{\frac{t+s}{2}} \langle \sigma_r^{-1}(Z^{s,x}_r) \nabla_\eta Z^{s,x}_r, \mathrm{d}W_r \rangle \bigg).$$

This further gives that

$$\begin{aligned}
&\frac{1}{2}(\nabla_{\eta'}\nabla_\eta P^0_{s,t} f)(x)\\
&= \mathbb{E}\bigg(\frac{(\nabla_{\nabla_{\eta'} Z^{s,x}_{\frac{t+s}{2}}} P^0_{\frac{t+s}{2},t} f)(Z^{s,x}_{\frac{t+s}{2}})}{t-s} \int_s^{\frac{t+s}{2}} \langle \sigma_r^{-1}(Z^{s,x}_r) \nabla_\eta Z^{s,x}_r, \mathrm{d}W_r \rangle \bigg)\\
&\quad + \mathbb{E}\bigg(\frac{(P^0_{\frac{t+s}{2},t} f)(Z^{s,x}_{\frac{t+s}{2}})}{t-s} \int_s^{\frac{t+s}{2}} \langle (\nabla_{\nabla_{\eta'} Z^{s,x}_r} \sigma_r^{-1})(Z^{s,x}_r) \nabla_\eta Z^{s,x}_r, \mathrm{d}W_r \rangle \bigg)\\
&\quad + \mathbb{E}\bigg(\frac{(P^0_{\frac{t+s}{2},t} f)(Z^{s,x}_{\frac{t+s}{2}})}{t-s} \int_s^{\frac{t+s}{2}} \langle \sigma_r^{-1}(Z^{s,x}_r) \nabla_{\eta'}\nabla_\eta Z^{s,x}_r, \mathrm{d}W_r \rangle \bigg).
\end{aligned}$$

Thus, applying Cauchy-Schwartz's inequality and Itô's isometry and taking (2.10), (2.12) and (2.13) into consideration, we derive that

$$\begin{aligned}
&|\nabla_{\eta'}\nabla_\eta P^0_{s,t} f|^2(x)\\
&\le 12 \bigg\{ 6\|\sigma^{-1}\|^2_{T,\infty} \frac{\mathbb{E}|\nabla P^0_{\frac{t+s}{2},t} f|^2(Z^{s,x}_{\frac{t+s}{2}})}{(t-s)^{5/2}}\\
&\quad \times (\mathbb{E}|\nabla_{\eta'} Z^{s,x}_{\frac{t+s}{2}}|^4)^{1/2} \bigg(\int_s^{\frac{t+s}{2}} \mathbb{E}|\nabla_\eta Z^{s,x}_r|^4 \mathrm{d}r \bigg)^{1/2}\\
&\quad + \frac{P^0_{s,t} f^2(x)}{(t-s)^2} \|\nabla\sigma^{-1}\|^2_{T,\infty} \int_s^{\frac{t+s}{2}} (\mathbb{E}|\nabla_{\eta'} Z^{s,x}_r|^4)^{1/2} (\mathbb{E}|\nabla_\eta Z^{s,x}_r|^4)^{1/2} \mathrm{d}r\\
&\quad + \frac{P^0_{s,t} f^2(x)}{(t-s)^2} \|\sigma^{-1}\|^2_{T,\infty} \int_s^{\frac{t+s}{2}} \mathbb{E}|\nabla_{\eta'}\nabla_\eta Z^{s,x}_r|^2 \mathrm{d}r \bigg\}\\
&\le \tilde{\Lambda}_{T,\sigma} |\eta|^2 |\eta'|^2 \frac{P^0_{s,t} f^2(x)}{(t-s)^2},
\end{aligned} \tag{2.14}$$

where $\tilde{\Lambda}_{T,\sigma} > 0$ is defined as in (2.6).

Set $\tilde{f}(\cdot) := f(\cdot) - f(x)$ for fixed $x \in \mathbb{R}^n$ and $f \in \mathscr{B}_b(\mathbb{R}^n)$ which verifies

$$|f(x) - f(y)| \le \phi(|x-y|), \quad x, y \in \mathbb{R}^n \tag{2.15}$$

for some $\phi \in \mathscr{D}$. For $f \in \mathscr{B}_b(\mathbb{R}^n)$ such that (2.15), (2.14) implies that

$$\begin{aligned}
|\nabla_{\eta'}\nabla_\eta P^0_{s,t} f|^2(x) = |\nabla_{\eta'}\nabla_\eta P^0_{s,t} \tilde{f}|^2(x) &\le \frac{\tilde{\Lambda}_{T,\sigma} |\eta|^2 |\eta'|^2}{(t-s)^2} \mathbb{E}|f(Z^{s,x}_t) - f(x)|^2\\
&\le \frac{\tilde{\Lambda}_{T,\sigma} |\eta|^2 |\eta'|^2}{(t-s)^2} \phi^2(\|\sigma\|_{T,\infty}(t-s)^{1/2}),
\end{aligned} \tag{2.16}$$

where in the second display we have used that

$$Z_t^{s,x} - x = \int_s^t \sigma_r(Z_r^{s,x})\mathrm{d}W_r,$$

and utilized Jensen's inequality as well as Itô's isometry.

Let $f_t = b_t + \nabla_{b_t} u_t^\lambda$. For any $\lambda \ge 9\pi\Lambda_{T,\sigma}^2\|b\|_{T,\infty}^2 + 4(\|b\|_{T,\infty} + \Lambda_{T,\sigma})^2$, note from **(ii)**, (2.12) and (2.14) that

$$\begin{aligned}
|f_t(x) - f_t(y)| &\le (1 + \|\nabla u^\lambda\|_{T,\infty})\phi(|x-y|) + \|b\|_{T,\infty}\|\nabla u_t^\lambda(x) - \nabla u_t(y)\|\mathbf{1}_{\{|x-y|\ge 1\}} \\
&\quad + \|b\|_{T,\infty}\|\nabla u_t^\lambda(x) - \nabla u_t(y)\|\mathbf{1}_{\{|x-y|\le 1\}} \\
&\le \frac{3}{2}\phi(|x-y|) + \|b\|_{T,\infty}\sqrt{|x-y|}\mathbf{1}_{\{|x-y|\ge 1\}} \\
&\quad + 10\Big(\Lambda_{T,\sigma} + \sqrt{\tilde{\Lambda}_{T,\sigma}}\Big)\|b\|_{T,\infty}^2\sqrt{|x-y|}\sqrt{|x-y|}\log\Big(\mathrm{e} + \frac{1}{|x-y|}\Big)\mathbf{1}_{\{|x-y|\le 1\}} \\
&\le \Big\{3 + 2\|b\|_{T,\infty} + 28\Big(\Lambda_{T,\sigma} + \sqrt{\tilde{\Lambda}_{T,\sigma}}\Big)\|b\|_{T,\infty}^2\Big\}\tilde{\phi}(|x-y|)
\end{aligned}$$

with $\tilde{\phi}(s) := \sqrt{\phi^2(s) + s}$, $s \ge 0$, where in the second inequality we have used [23, Lemma 2.2 (1)], and the fact that the function $[0,1] \ni x \mapsto \sqrt{x}\log(\mathrm{e} + \frac{1}{x})$ is non-decreasing. As a result, **(iii)** follows from (2.16). □

Remark 2.2. By checking carefully the argument of Lemma 2.1, the concave property of ϕ^2 can be removed whenever $\phi(x) = x^\beta$ for $x \ge 0$ and $\beta \in (0,1]$, i.e., the drift b is Hölder continuous.

3 Proof of Theorem 1.1

With Lemma 2.1 in hand, we now in a position to complete the

Proof of Theorem 1.1. Throughout the whole proof, we assume $\lambda \ge 9\pi\Lambda_{T,\sigma}^2\|b\|_{T,\infty}^2 + 4(\|b\|_{T,\infty} + \Lambda_{T,\sigma})^2$ so that **(i)-(iii)** in Lemma 2.1 hold. For any $t \in [0,T]$, applying Itô's formula to $x + u_t^\lambda(x), x \in \mathbb{R}^n$, we deduce from (2.1) that

$$X_t + u_t^\lambda(X_t) = x + u_0^\lambda(x) + \lambda\int_0^t u_s^\lambda(X_s)\mathrm{d}s + \int_0^t \{\mathbf{I_{n\times n}} + (\nabla u_s^\lambda)(\cdot)\}(X_s)\sigma_s(X_s)\mathrm{d}W_s, \tag{3.1}$$

where $\mathbf{I_{n\times n}}$ is an $n \times n$ identity matrix, and that

$$\begin{aligned}
Y_t + u_t^\lambda(Y_t) &= x + u_0^\lambda(x) + \lambda\int_0^t u_s^\lambda(Y_s)\mathrm{d}s + \int_0^t \{\mathbf{I_{n\times n}} + (\nabla u_s^\lambda)(\cdot)\}(Y_s)\sigma_{s_\delta}(Y_{s_\delta})\mathrm{d}W_s \\
&\quad + \int_0^t \{\mathbf{I_{n\times n}} + (\nabla u_s^\lambda)(\cdot)\}(Y_s)\{b_{s_\delta}(Y_{s_\delta}) - b_s(Y_s)\}\mathrm{d}s \\
&\quad + \frac{1}{2}\int_0^t \sum_{k,j}\langle\{(\sigma_{s_\delta}\sigma_{s_\delta}^*)(Y_{s_\delta}) - (\sigma_s\sigma_s^*)(Y_s)\}e_k, e_j\rangle(\nabla_{e_k}\nabla_{e_j}u_s^\lambda)(Y_s)\mathrm{d}s.
\end{aligned} \tag{3.2}$$

For notational simplicity, set

$$M_t^\lambda := X_t - Y_t + u_t^\lambda(X_t) - u_t^\lambda(Y_t). \tag{3.3}$$

Using the elementary inequality: $(a+b)^2 \le (1+\varepsilon)(a^2+\varepsilon^{-1}b^2)$ for arbitrary $\varepsilon, a, b > 0$, we derive from **(ii)** that

$$\begin{aligned}|X_t - Y_t|^2 &\le (1+\varepsilon)(|M_t^\lambda|^2 + \varepsilon^{-1}|u_t^\lambda(X_t) - u_t^\lambda(Y_t)|^2)\\ &\le (1+\varepsilon)\Big(|M_t^\lambda|^2 + \frac{\varepsilon^{-1}}{4}|X_t - Y_t|^2\Big).\end{aligned}$$

In particular, taking $\varepsilon = 1$ leads to

$$|X_t - Y_t|^2 \le \frac{1}{2}|X_t - Y_t|^2 + 2|M_t^\lambda|^2.$$

As a consequence,

$$\mathbb{E}\Big(\sup_{0\le s\le t}|X_s - Y_s|^2\Big) \le 4\mathbb{E}\Big(\sup_{0\le s\le t}|M_s^\lambda|^2\Big). \tag{3.4}$$

In what follows, our goal is to estimate the term on the right hand side of (3.4). Observe from the definition of the Hilbert-Schmidt norm that

$$\begin{aligned}&\int_0^t \mathbb{E}\Big|\sum_{k,j}\langle[(\sigma_{s_\delta}\sigma_{s_\delta}^*)(Y_{s_\delta}) - (\sigma_s\sigma_s^*)(Y_s)]e_k, e_j\rangle(\nabla_{e_k}\nabla_{e_j}u_s^\lambda)(Y_s)\Big|^2\mathrm{d}s\\ &\lesssim_T \|\nabla^2 u^\lambda\|_{T,\infty}^2\int_0^t \mathbb{E}\|(\sigma_{s_\delta}\sigma_{s_\delta}^*)(Y_{s_\delta}) - (\sigma_s\sigma_s^*)(Y_s)\|_{\mathrm{HS}}^2\mathrm{d}s.\end{aligned} \tag{3.5}$$

Thus, by Hölder's inequality, Doob's submartingale inequality and Itô's isometry, it follows from (3.1), (3.2) and (3.5) that

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le s\le t}|M_s^\lambda|^2\Big) \le C_T\Big\{&\lambda^2\int_0^t \mathbb{E}|u_s^\lambda(X_s) - u_s^\lambda(Y_s)|^2\mathrm{d}s\\ &+ (1+\|\nabla u\|_{T,\infty}^2)\int_0^t \mathbb{E}|b_{s_\delta}(Y_s) - b_{s_\delta}(Y_{s_\delta})|^2\mathrm{d}s\\ &+ (1+\|\nabla u\|_{T,\infty}^2)\int_0^t \mathbb{E}|b_s(Y_s) - b_{s_\delta}(Y_s)|^2\mathrm{d}s\\ &+ \int_0^t \mathbb{E}\|\{(\nabla u_s^\lambda)(X_s) - (\nabla u_s^\lambda)(Y_s)\}\sigma_s(X_s)\|_{\mathrm{HS}}^2\mathrm{d}s\\ &+ (1+\|\nabla u\|_{T,\infty}^2)\int_0^t \mathbb{E}\|\sigma_{s_\delta}(X_s) - \sigma_{s_\delta}(Y_{s_\delta})\|_{\mathrm{HS}}^2\mathrm{d}s\\ &+ \|\nabla^2 u^\lambda\|_{T,\infty}^2\int_0^t \mathbb{E}\|\{\sigma_{s_\delta}(Y_s) - \sigma_{s_\delta}(Y_{s_\delta})\}\sigma_{s_\delta}^*(Y_{s_\delta})\|_{\mathrm{HS}}^2\mathrm{d}s\end{aligned}$$

$$
\begin{aligned}
&+\|\nabla^2 u^\lambda\|_{T,\infty}^2\int_0^t\mathbb{E}\|\sigma_s(Y_s)\{\sigma^*_{s_\delta}(Y_s)-\sigma^*_{s_\delta}(Y_{s_\delta})\}\|_{\mathrm{HS}}^2\mathrm{d}s\\
&+(1+\|\nabla u\|_{T,\infty}^2)\int_0^t\mathbb{E}\|\sigma_s(X_s)-\sigma_{s_\delta}(X_s)\|_{\mathrm{HS}}^2\mathrm{d}s\\
&+\|\nabla^2 u^\lambda\|_{T,\infty}^2\int_0^t\mathbb{E}\|\sigma_s(Y_s)\{\sigma^*_s(Y_s)-\sigma^*_{s_\delta}(Y_s)\}\|_{\mathrm{HS}}^2\mathrm{d}s\\
&+\|\nabla^2 u^\lambda\|_{T,\infty}^2\int_0^t\mathbb{E}\|\{\sigma_s(Y_s)-\sigma_{s_\delta}(Y_s)\}\sigma^*_{s_\delta}(Y_{s_\delta})\|_{\mathrm{HS}}^2\mathrm{d}s\Big\}\\
=:&\, C_T\Big(\sum_{i=1}^{10}I_i(t)\Big)
\end{aligned}
$$

for some constant $C_T>0$. Also, applying Hölder's inequality and Itô's isometry, we deduce from (**A1**) that

$$
\mathbb{E}|Y_t-Y_{t_\delta}|^2\le\beta_T\delta \tag{3.6}
$$

for some constant $\beta_T\ge1$. By Taylor's expansion, it is readily to see that

$$
I_1(t)+I_4(t)\lesssim\{\lambda^2\|\nabla u^\lambda\|_{T,\infty}^2+\|\nabla^2 u^\lambda\|_{T,\infty}^2\|\sigma\|_{T,\infty}^2\}\int_0^t\mathbb{E}|X_s-Y_s|^2\mathrm{d}s. \tag{3.7}
$$

From (**A3**) and due to the fact that $\phi(\cdot)$ is increasing and $\delta\in(0,1)$, one has

$$
I_3(t)+\sum_{i=8}^{10}I_i(t)\lesssim_T\{1+\|\nabla u^\lambda\|_{T,\infty}^2+\|\nabla^2 u^\lambda\|_{T,\infty}^2\|\sigma\|_{T,\infty}^2\}\phi^2(\sqrt{\delta}). \tag{3.8}
$$

In view of (**A2**), we derive that

$$
\begin{aligned}
&I_2(t)+\sum_{i=5}^{7}I_i(t)\\
&\lesssim\{1+\|\nabla u^\lambda\|_{T,\infty}^2\}\int_0^t\mathbb{E}\phi(|Y_s-Y_{s_\delta}|)^2\mathrm{d}s\\
&\quad+\{1+\|\nabla u^\lambda\|_{T,\infty}^2\}\|\nabla\sigma\|_{T,\infty}^2\int_0^t\mathbb{E}|X_s-Y_s|^2\mathrm{d}s\\
&\quad+\{1+\|\nabla u^\lambda\|_{T,\infty}^2+\|\nabla^2 u^\lambda\|_{T,\infty}^2\|\sigma\|_{T,\infty}^2\}\|\nabla\sigma\|_{T,\infty}^2\int_0^t\mathbb{E}|Y_s-Y_{s_\delta}|^2\mathrm{d}s.
\end{aligned} \tag{3.9}
$$

Thus, taking (3.6)-(3.9) into account and applying Jensen's inequality gives that

$$
\mathbb{E}\Big(\sup_{0\le s\le t}|M_s^\lambda|^2\Big)\lesssim_T C_{T,\sigma,\lambda}\{\delta+\phi^2(\beta_T\sqrt{\delta})\}+C_{T,\sigma,\lambda}\int_0^t\mathbb{E}|X_s-Y_s|^2\mathrm{d}s,
$$

where

$$C_{T,\sigma,\lambda} := \{1+\|\nabla\sigma\|_{T,\infty}^2\}\Big\{\frac{5}{4}+(1+\lambda^2)\|\nabla^2 u^\lambda\|_{T,\infty}^2\|\sigma\|_{T,\infty}^2\Big\}. \tag{3.10}$$

Owing to $\phi \in \mathscr{D}$, we conclude that $\phi(0)=0$, $\phi'>0$ and $\phi''<0$ so that, for any $c>0$ and $\delta\in(0,1)$,

$$\phi(c\delta)=\phi(0)+\phi'(\xi)c\delta \geq \phi'(c)c\delta,$$

where $\xi\in(0,c\delta)$. This further implies that

$$\mathbb{E}\Big(\sup_{0\le s\le t}|M_s^\lambda|^2\Big) \lesssim_T C_{T,\sigma,\lambda}\phi^2(\beta_T\sqrt{\delta}) + C_{T,\sigma,\lambda}\int_0^t \mathbb{E}|X_s-Y_s|^2\mathrm{d}s.$$

Substituting this into (3.4) gives that

$$\mathbb{E}\Big(\sup_{0\le s\le t}|X_s-Y_s|^2\Big) \lesssim_T C_{T,\sigma,\lambda}\phi^2(\beta_T\sqrt{\delta}) + C_{T,\sigma,\lambda}\int_0^t \mathbb{E}|X_s-Y_s|^2\mathrm{d}s.$$

Thus, Gronwall's inequality implies that there exists $\tilde{C}_T>0$ such that

$$\mathbb{E}\Big(\sup_{0\le s\le t}|X_s-Y_s|^2\Big) \le \tilde{C}_T C_{T,\sigma,\lambda}\mathrm{e}^{\tilde{C}_T C_{T,\sigma,\lambda}}\phi^2(\beta_T\sqrt{\delta}). \tag{3.11}$$

So the desired assertion holds immediatelly.

4 Proof of Theorem 1.3

We shall adopt the cut-off approach to finish the

Proof of Theorem 1.3. Take $\psi\in C_b^\infty(\mathbb{R}_+)$ such that $0\le\psi\le 1$, $\psi(r)=1$ for $r\in[0,1]$ and $\psi(r)=0$ for $r\ge 2$. For any $t\in[0,T]$ and $k\ge 1$, define the cut-off functions

$$b_t^{(k)}(x)=b_t(x)\psi(|x|/k) \qquad \text{and} \qquad \sigma_t^{(k)}(x)=\sigma_t(x)\psi(|x|/k), \qquad x\in\mathbb{R}^n.$$

It is easy to see that $b^{(k)}$ and $\sigma^{(k)}$ satisfy **(A1)**. For fixed $k\ge 1$, consider the following SDE

$$\mathrm{d}X_t^{(k)} = b_t^{(k)}(X_t^{(k)})\mathrm{d}t + \sigma_t^{(k)}(X_t^{(k)})\mathrm{d}W_t, \quad t>0, \quad X_0^{(k)}=X_0=x. \tag{4.1}$$

The corresponding continuous-time EM of (4.1) is defined by

$$\mathrm{d}Y_t^{(k)} = b_{t_\delta}^{(k)}(Y_{t_\delta}^{(k)})\mathrm{d}t + \sigma_{t_\delta}^{(k)}(Y_{t_\delta}^{(k)})\mathrm{d}W_t, \quad t>0, \quad Y_0^{(k)}=X_0=x. \tag{4.2}$$

Applying BDG's inequality, Hölder's inequality and Gronwall's inequality, we deduce from **(A1')** that

$$\mathbb{E}\Big(\sup_{0\le t\le T}|X_t|^4\Big)+\mathbb{E}\Big(\sup_{0\le t\le T}|Y_t|^4\Big)+\mathbb{E}\Big(\sup_{0\le t\le T}|X_t^{(k)}|^4\Big)+\mathbb{E}\Big(\sup_{0\le t\le T}|Y_t^{(k)}|^4\Big)\le C_T \tag{4.3}$$

for some constant $C_T>0$. Note that

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le t\le T}|X_t-Y_t|^2\Big)&\le 2\mathbb{E}\Big(\sup_{0\le t\le T}|X_t-X_t^{(k)}|^2\Big)+2\mathbb{E}\Big(\sup_{0\le t\le T}|X_t^{(k)}-Y_t^{(k)}|^2\Big)\\&\quad+2\mathbb{E}\Big(\sup_{0\le t\le T}|Y_t-Y_t^{(k)}|^2\Big)\\&=:I_1+I_2+I_3.\end{aligned}$$

For the terms I_1 and I_3, in terms of the Chebyshev inequality we find from (4.3) that

$$\begin{aligned}I_1+I_3&\lesssim\mathbb{E}\Big(\sup_{0\le t\le T}|X_t-X_t^{(k)}|^2\mathbf{1}_{\{\sup_{0\le t\le T}|X_t|\ge k\}}\Big)\\&\quad+\mathbb{E}\Big(\sup_{0\le t\le T}|Y_t-Y_t^{(k)}|^2\mathbf{1}_{\{\sup_{0\le t\le T}|Y_t|\ge k\}}\Big)\\&\lesssim\sqrt{\mathbb{E}\Big(\sup_{0\le t\le T}|X_t|^4\Big)+\mathbb{E}\Big(\sup_{0\le t\le T}|X_t^{(k)}|^4\Big)}\frac{\sqrt{\mathbb{E}\Big(\sup_{0\le t\le T}|X_t|^2\Big)}}{k}\\&\quad+\sqrt{\mathbb{E}\Big(\sup_{0\le t\le T}|Y_t|^4\Big)+\mathbb{E}\Big(\sup_{0\le t\le T}|Y_t^{(k)}|^4\Big)}\frac{\sqrt{\mathbb{E}\Big(\sup_{0\le t\le T}|Y_t|^2\Big)}}{k}\\&\lesssim_T\frac{1}{k},\end{aligned}$$

where in the first display we have used the facts that $\{X_t\neq X_t^{(k)}\}\subset\{\sup_{0\le s\le t}|X_s|\ge k\}$ and $\{Y_t\neq Y_t^{(k)}\}\subset\{\sup_{0\le s\le t}|Y_s|\ge k\}$. Observe from **(A1')** that $9\pi\Lambda^2_{T,\sigma^{(k)}}\|b^{(k)}\|^2_{T,\infty}+4(\|b^{(k)}\|_{T,\infty}+\Lambda_{T,\sigma^{(k)}})^2\le \mathrm{e}^{ck^2}$ for some $c>0$. Next, according to (3.11), by taking $\lambda=\mathrm{e}^{ck^2}$ there exits $C_T>0$ such that

$$I_2\le \mathrm{e}^{C_TC_{T,\sigma^{(k)},\lambda}}\phi_k^2(\beta_T\sqrt{\delta}).$$

Herein, $C_{T,\sigma^{(k)},\lambda}>0$ is defined as in (3.10) with σ and u^λ replaced by $\sigma^{(k)}$ and $u^{\lambda,k}$, respectively, where $u^{\lambda,k}$ solves (2.3) by writing $b^{(k)}$ instead of b. Consequently, we conclude that

$$\mathbb{E}\Big(\sup_{0\le t\le T}|X_t-Y_t|^2\Big)\le\frac{\bar{c}_0}{k}+\bar{c}_0\mathrm{e}^{C_TC_{T,\sigma^{(k)},\lambda}}\phi_k^2(\beta_T\sqrt{\delta})\tag{4.4}$$

for some $\bar{c}_0>0$. For any $\varepsilon>0$, taking $k=\frac{2\bar{c}_0}{\varepsilon}$ and letting δ go to zero implies that

$$\lim_{\delta\to0}\mathbb{E}\Big(\sup_{0\le t\le T}|X_t-Y_t|^2\Big)\le\varepsilon.$$

Thus, (1.4) follows due to the arbitrariness of ε.

For $\phi_k(s)=\mathrm{e}^{\mathrm{e}^{c_0k^4}}s^\alpha$, $s\ge 0$, with $\alpha\in(0,1]$, we deduce from Lemma 2.1 **(iii)** and Remark 2.2 that

$$\|\nabla^2 u^{\lambda,k}\|_{T,\infty}\le \frac{1}{2} \tag{4.5}$$

whenever

$$\begin{aligned}\lambda\ge \Big\{2\Upsilon_{T,\sigma^{(k)}}\Big(\mathrm{e}^{\mathrm{e}^{c_0k^4}}\|\sigma^{(k)}\|_{T,\infty}^{\alpha}\Gamma(\alpha/2)+\|\sigma^{(k)}\|_{T,\infty}^{1/2}\Gamma(1/4)\Big)\Big\}^{2/\alpha}\\ +9\pi(\Lambda_{T,\sigma^{(k)}})^2\|b^{(k)}\|_{T,\infty}^2+4(\|b^{(k)}\|_{T,\infty}+\Lambda_{T,\sigma^{(k)}})^2.\end{aligned} \tag{4.6}$$

Since the right hand side of (4.6) can be bounded by $\mathrm{e}^{\mathrm{e}^{\bar C_T k^4}}$ for some constant $\bar C_T>0$ due to **(A1')**, we can take $\lambda=\mathrm{e}^{\mathrm{e}^{\bar C_T k^4}}$ so that (4.5) holds. Thus, (4.4), together with (4.5) and **(A1')**, yields that

$$\mathbb{E}\Big(\sup_{0\le t\le T}|X_t-Y_t|^2\Big)\le \frac{\hat C_T}{k}+\hat C_T\mathrm{e}^{\mathrm{e}^{\tilde C_T k^4}}\delta^\alpha$$

for some constants $\hat C_T,\tilde C_T>0$. Thus, (1.5) follows immediately by taking

$$k=(\tilde C_T\log\log\delta^{-\alpha\varepsilon})^{\frac{1}{4}}.$$

□

5 Proof of Theorem 1.5

For simplicity, for any $f:\mathbb{R}^{m_1}\to\mathbb{R}^{m_2}$, let

$$[f]_L=\sup_{x\neq y}\frac{|f(x)-f(y)|}{|x-y|},\quad \|f\|_\infty=\sup_{x\in\mathbb{R}^{m_1}}|f(x)|,$$

$[\cdot]_L$ is the Lipschitz constant of f.

The proof of Theorem 1.5 relies on regularization properties of the following $\mathbb{R}^{2n}$-valued *degenerate* parabolic equation

$$\partial_t u_t^\lambda+\mathscr{L}_t^{b,\sigma}u_t^\lambda+b_t=\lambda u_t^\lambda,\qquad u_T^\lambda=\mathbf{0_{2n}},\quad t\in[0,T],\quad \lambda>0, \tag{5.1}$$

where $\mathbf{0_{2n}}$ is the zero vector in $\mathbb{R}^{2n}$,

$$b_t:=\begin{pmatrix} b_t^{(1)}\\ b_t^{(2)}\end{pmatrix}\quad\text{and}\quad \mathscr{L}_t^{b,\sigma}u^\lambda:=\frac{1}{2}\sum_{i,j=1}^n\langle(\sigma_t\sigma_t^*)(\cdot)e_i,e_j\rangle\nabla_{e_i}^{(2)}\nabla_{e_j}^{(2)}u^\lambda+\nabla_{b_t^{(1)}}^{(1)}u^\lambda+\nabla_{b_t^{(2)}}^{(2)}u^\lambda.$$

The following lemma on regularity estimate of solution to (5.1) is taken from [24, Theorem 3.10, (4.4)] and is an essential ingredient in analyzing numerical approximation.

Lemma 5.1. Under **(C1)-(C3)**, (5.1) has a unique solution $u^\lambda \in C([0,T];C_b^1(\mathbb{R}^{2n};\mathbb{R}^{2n}))$ such that for all $t\in[0,T]$,

$$\|\nabla u_t^\lambda\|_\infty + \|\nabla^{(2)}\nabla^{(2)}u_t^\lambda\|_\infty + [\nabla^{(2)}u_t]_L \le C\int_0^T \mathrm{e}^{-\lambda t}\frac{\phi(t^{\frac{1}{2}})}{t}\mathrm{d}t, \tag{5.2}$$

where $C>0$ is a constant.

From now on, we move forward to complete the

Proof of Theorem 1.5. For notational simplicity, set

$$X_t := \begin{pmatrix} X_t^{(1)} \\ X_t^{(2)} \end{pmatrix}, \quad Y_t := \begin{pmatrix} Y_t^{(1)} \\ Y_t^{(2)} \end{pmatrix} \text{ and } \quad b_t(x) := \begin{pmatrix} b_t^{(1)}(x) \\ b_t^{(2)}(x) \end{pmatrix}, \quad x\in\mathbb{R}^{2n}.$$

Then (1.6) and (1.7) can be reformulated respectively as

$$\mathrm{d}X_t = b_t(X_t)\mathrm{d}t + \begin{pmatrix} \mathbf{0_{n\times n}} \\ \sigma_t \end{pmatrix}(X_t)\mathrm{d}W_t, \quad t>0, \quad X_0 = x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\in\mathbb{R}^{2n},$$

where $\mathbf{0_{n\times n}}$ is an $n\times n$ zero matrix, and

$$\mathrm{d}Y_t = b_{t_\delta}(Y_{t_\delta})\mathrm{d}t + \begin{pmatrix} \mathbf{0_{n\times n}} \\ \sigma_{t_\delta} \end{pmatrix}(Y_{t_\delta})\mathrm{d}W_t, \quad t>0, \quad Y_0 = x\in\mathbb{R}^{2n}.$$

Note from (5.2) that there exists $\lambda_0>0$ sufficiently large such that for any $t\in[0,T]$,

$$\|\nabla u_t^\lambda\|_\infty + \|\nabla^{(2)}\nabla^{(2)}u_t^\lambda\|_\infty + [\nabla^{(2)}u_t^\lambda]_L \le \frac{1}{2}, \quad \lambda\ge\lambda_0. \tag{5.3}$$

Applying Itô's formula to $x+u_t^\lambda(x)$ for any $x\in\mathbb{R}^{2n}$, we deduce that

$$X_t + u_t^\lambda(X_t) = x + u_0^\lambda(x) + \lambda\int_0^t u_s^\lambda(X_s)\mathrm{d}s + \int_0^t \begin{pmatrix} \mathbf{0_{n\times n}} \\ \sigma_s \end{pmatrix}(X_s)\mathrm{d}W_s + \int_0^t (\nabla^{(2)}_{\sigma_s\mathrm{d}W_s}u_s^\lambda)(X_s), \tag{5.4}$$

and that

$$\begin{aligned} Y_t + u_t^\lambda(Y_t) &= x + u_0^\lambda(x) + \lambda\int_0^t u_s^\lambda(Y_s)\mathrm{d}s \\ &\quad + \int_0^t \{\mathbf{I_{2n\times 2n}} + (\nabla u_s)(\cdot)\}(Y_s)\{b_{s_\delta}(Y_{s_\delta}) - b_s(Y_s)\}\mathrm{d}s \\ &\quad + \int_0^t \begin{pmatrix} \mathbf{0_{n\times n}} \\ \sigma_{s_\delta} \end{pmatrix}(Y_{s_\delta})\mathrm{d}W_s + \int_0^t (\nabla^{(2)}_{\sigma_{s_\delta}(Y_{s_\delta})\mathrm{d}W_s}u_s^\lambda)(Y_s) \\ &\quad + \frac{1}{2}\int_0^t \sum_{k,j=1}^n \langle\{(\sigma_{s_\delta}\sigma_{s_\delta}^*)(Y_{s_\delta}) - (\sigma_s\sigma_s^*)(Y_s)\}e_k, e_j\rangle(\nabla^{(2)}_{e_k}\nabla^{(2)}_{e_j}u_s^\lambda)(Y_s)\mathrm{d}s, \end{aligned} \tag{5.5}$$

where $\mathbf{I_{2n\times 2n}}$ is an $2n\times 2n$ identity matrix. Thus, using Hölder's inequality, Doob's submartingale inequality and Itô's isometry and taking (3.5) into consideration gives that

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le s\le t}|M_s^\lambda|^2\Big)\le C_{0,T}\Big\{&\int_0^t\mathbb{E}|u_s^\lambda(X_s)-u_s^\lambda(Y_s)|^2\mathrm{d}s\\&+(1+\|\nabla u^\lambda\|_{T,\infty}^2)\int_0^t\mathbb{E}|b_{s_\delta}(Y_s)-b_{s_\delta}(Y_{s_\delta})|^2\mathrm{d}s\\&+(1+\|\nabla u^\lambda\|_{T,\infty}^2)\int_0^t\mathbb{E}|b_s(Y_s)-b_{s_\delta}(Y_s)|^2\mathrm{d}s\\&+\int_0^t\mathbb{E}\|\{(\nabla^{(2)}u_s^\lambda)(X_s)-\nabla^{(2)}u_s^\lambda(Y_s)\}\sigma_s(X_s)\|_{\mathrm{HS}}^2\mathrm{d}s\\&+(1+\|\nabla^{(2)}u^\lambda\|_{T,\infty}^2)\int_0^t\mathbb{E}\|\sigma_{s_\delta}(X_s)-\sigma_{s_\delta}(Y_{s_\delta})\|_{\mathrm{HS}}^2\mathrm{d}s\\&+(1+\|\nabla^{(2)}u^\lambda\|_{T,\infty}^2)\int_0^t\mathbb{E}\|\sigma_s(X_s)-\sigma_{s_\delta}(X_s)\|_{\mathrm{HS}}^2\mathrm{d}s\\&+\|\nabla^{(2)}\nabla^{(2)}u^\lambda\|_{T,\infty}^2\int_0^t\mathbb{E}\|\{\sigma_{s_\delta}(Y_s)-\sigma_{s_\delta}(Y_{s_\delta})\}\sigma_{s_\delta}^*(Y_{s_\delta})\|_{\mathrm{HS}}^2\mathrm{d}s\\&+\|\nabla^{(2)}\nabla^{(2)}u^\lambda\|_{T,\infty}^2\int_0^t\mathbb{E}\|\sigma_s(Y_s)\{\sigma_{s_\delta}^*(Y_s)-\sigma_{s_\delta}^*(Y_{s_\delta})\}\|_{\mathrm{HS}}^2\mathrm{d}s\\&+\|\nabla^{(2)}\nabla^{(2)}u^\lambda\|_{T,\infty}^2\int_0^t\mathbb{E}\|\sigma_s(Y_s)\{\sigma_s^*(Y_s)-\sigma_{s_\delta}^*(Y_s)\}\|_{\mathrm{HS}}^2\mathrm{d}s\\&+\|\nabla^{(2)}\nabla^{(2)}u^\lambda\|_{T,\infty}^2\int_0^t\mathbb{E}\|\{\sigma_s(Y_s)-\sigma_{s_\delta}(Y_s)\}\sigma_{s_\delta}^*(Y_{s_\delta})\|_{\mathrm{HS}}^2\mathrm{d}s\Big\}\\=:C_{0,T}\Big(\sum_{i=1}^{10}J_i(t)\Big)\end{aligned}$$

for some constant $C_{0,T}>0$, where M_t^λ is defined as in (3.3). By using Hölder's inequality and the BDG inequality, **(C1)** implies that

$$\mathbb{E}|Y_t-Y_{t_\delta}|^p\lesssim\delta^{\frac{p}{2}},\qquad p\ge 1. \tag{5.6}$$

Utilizing Taylor's expansion, one gets from (3.6), (5.3) and (5.6) that

$$\begin{aligned}J_1(t)+J_4(t)+J_5(t)&\lesssim\{1+\|\nabla u^\lambda\|_{T,\infty}^2+\|\nabla\nabla^{(2)}u^\lambda\|_{T,\infty}^2\|\sigma\|_{T,\infty}^2\}\int_0^t\mathbb{E}|X_s-Y_s|^2\mathrm{d}s\\&\quad+\{1+\|\nabla^{(2)}u^\lambda\|_{T,\infty}^2\}\int_0^t\mathbb{E}|Y_s-Y_{s_\delta}|^2\mathrm{d}s\\&\lesssim\delta+\int_0^t\mathbb{E}|X_s-Y_s|^2\mathrm{d}s.\end{aligned}$$

Next, **(C1)**, **(C5)** and (5.3) yield that

$$J_3(t)+J_6(t)+J_9(t)+J_{10}(t)\lesssim\phi^2(\sqrt{\delta}),$$

where we have also used that $\phi(\cdot)$ is increasing and $\delta \in (0,1)$. Additionally, by virtue of **(C1)**, **(C2)**, and (5.3), we infer from **(C3)** that

$$\begin{aligned}J_2(t)+J_7(t)+J_8(t) &\lesssim \delta + \int_0^t \mathbb{E}|b_{s_\delta}(Y_s^{(1)},Y_s^{(2)}) - b_{s_\delta}(Y_{s_\delta}^{(1)},Y_s^{(2)})|^2\mathrm{d}s \\ &\quad + \int_0^t \mathbb{E}|b_{s_\delta}(Y_{s_\delta}^{(1)},Y_s^{(2)}) - b_{s_\delta}(Y_{s_\delta}^{(1)},Y_{s_\delta}^{(2)})|^2\mathrm{d}s \\ &\le C_{1,T}\bigg\{\delta + \int_0^t \mathbb{E}|b^{(1)}_{s_\delta}(Y_s^{(1)},Y_s^{(2)}) - b^{(1)}_{s_\delta}(Y_{s_\delta}^{(1)},Y_s^{(2)})|^2\mathrm{d}s \\ &\quad + \int_0^t \mathbb{E}|b^{(2)}_{s_\delta}(Y_s^{(1)},Y_s^{(2)}) - b^{(2)}_{s_\delta}(Y_{s_\delta}^{(1)},Y_s^{(2)})|^2\mathrm{d}s \\ &\quad + \int_0^t \mathbb{E}|b^{(1)}_{s_\delta}(Y_{s_\delta}^{(1)},Y_s^{(2)}) - b^{(1)}_{s_\delta}(Y_{s_\delta}^{(1)},Y_{s_\delta}^{(2)})|^2\mathrm{d}s \\ &\quad + \int_0^t \mathbb{E}|b^{(2)}_{s_\delta}(Y_{s_\delta}^{(1)},Y_s^{(2)}) - b^{(2)}_{s_\delta}(Y_{s_\delta}^{(1)},Y_{s_\delta}^{(2)})|^2\mathrm{d}s\bigg\} \\ &=: C_{1,T}\Big(\delta + \sum_{i=1}^4 \Lambda_i(t)\Big)\end{aligned}$$

for some constant $C_{1,T}>0$. From **(C2)**, **(C3)**, (5.6) and $\phi \in \mathscr{D}^\varepsilon$, we derive from Hölder's inequality and Jensen's inequality that

$$\begin{aligned}\Lambda_1(t)+\Lambda_2(t) &\lesssim \sum_{i=1}^2 \int_0^t \mathbb{E}\bigg(\frac{|b^{(i)}_{s_\delta}(Y_s^{(1)},Y_s^{(2)}) - b^{(i)}_{s_\delta}(Y_{s_\delta}^{(1)},Y_s^{(2)})|}{|Y_s^{(1)}-Y_{s_\delta}^{(1)}|^{\frac23}\phi(|Y_s^{(1)}-Y_{s_\delta}^{(1)}|)}\mathbf{1}_{\{\mathbf{Y}_\mathbf{s}^{(1)}\neq \mathbf{Y}_{\mathbf{s}_\delta}^{(1)}\}} \\ &\qquad \times |Y_s^{(1)}-Y_{s_\delta}^{(1)}|^{\frac23}\phi(|Y_s^{(1)}-Y_{s_\delta}^{(1)}|)\bigg)^2\mathrm{d}s \\ &\lesssim \int_0^t \mathbb{E}(|Y_s^{(1)}-Y_{s_\delta}^{(1)}|^{\frac23}\phi(|Y_s^{(1)}-Y_{s_\delta}^{(1)}|))^2\mathrm{d}s \\ &\lesssim \int_0^t \Big(\mathbb{E}\phi(|Y_s^{(1)}-Y_{s_\delta}^{(1)}|)^{2(1+\varepsilon)}\Big)^{\frac{1}{1+\varepsilon}}\Big(\mathbb{E}|Y_s^{(1)}-Y_{s_\delta}^{(1)}|^{\frac{4(1+\varepsilon)}{3\varepsilon}}\Big)^{\frac{\varepsilon}{1+\varepsilon}}\mathrm{d}s \\ &\lesssim \delta^{\frac23}\phi^2(C_{2,T}\sqrt{\delta})\end{aligned} \tag{5.7}$$

for some constant $C_{2,T}>0$. With regard to the term $\Lambda_3(t)$, **(C1)** and (5.6) leads to

$$\Lambda_3(t) \lesssim \|\nabla^{(2)}b^{(1)}\|_{T,\infty}^2 \int_0^t \mathbb{E}|Y_s^{(1)}-Y_{s_\delta}^{(1)}|^2\mathrm{d}s \lesssim \delta. \tag{5.8}$$

Due to **(C3)**, observe from Jensen's inequality and (5.6) that

$$\begin{aligned}\Lambda_4(t) &\lesssim \int_0^t \mathbb{E}\bigg(\frac{|b^{(2)}_{s_\delta}(Y_{s_\delta}^{(1)},Y_s^{(2)}) - b^{(2)}_{s_\delta}(Y_{s_\delta}^{(1)},Y_{s_\delta}^{(2)})|}{\phi(|Y_s^{(2)}-Y_{s_\delta}^{(2)}|)}\mathbf{1}_{\{Y_s^{(2)}\neq Y_{s_\delta}^{(2)}\}} \times \phi(|Y_s^{(2)}-Y_{s_\delta}^{(2)}|)\bigg)^2\mathrm{d}s \\ &\lesssim \int_0^t \mathbb{E}\phi(|Y_s^{(2)}-Y_{s_\delta}^{(2)}|)^2\mathrm{d}s \\ &\lesssim \phi^2(C_{3,T}\sqrt{\delta})\end{aligned}$$

for some constant $C_{3,T} > 0$. Consequently, we arrive at

$$\mathbb{E}\Big(\sup_{0\le s\le t}|X_s - Y_s|^2\Big) \lesssim_T \phi^2(C_{4,T}\sqrt{\delta}) + \int_0^t \mathbb{E}\sup_{0\le r\le s}|X_r - Y_r|^2 \mathrm{d}s$$

for some constant $C_{4,T} \ge 1$. Thus, the desired assertion follows from the Gronwall inequality. □

Acknowledgement. The authors would like to thank Professor Feng-Yu Wang for helpful comments.

References

[1] Bao, J., Yuan, C., Convergence rate of EM scheme for SDDEs, *Proc. Amer. Math. Soc.*, **141** (2013), 3231–3243.

[2] Cordonia, F., Di Persiob, L., Oliva, I., A nonlinear Kolmogorov equation for stochastic functional delay differential equations with jumps, arXiv:1602.03851v1.

[3] Dareiotis, K., Kumar, C., Sabanis, S., On tamed Euler approximations of SDEs driven by Lévy noise with applications to delay equations, *SIAM J. Numer. Anal.*, **54** (2016), 1840–1872.

[4] Da Prato, G., Flandoli, F., Röckner, M., Veretennikov, A. Yu., Strong uniqueness for SDEs in Hilbert spaces with nonregular drift, *Ann. Probab.*, **44** (2016), 1985–2023.

[5] Gyÿgy, I., Rásonyi, M., A note on Euler approximations for SDEs with Hölder continuous diffusion coefficients, *Stochastic Process. Appl.*, **121** (2011), 2189–2200.

[6] Gyöngy, I., Sabanis, S., A note on Euler approximations for stochastic differential equations with delay, *Appl. Math. Optim.*, **68** (2013), 391–412.

[7] Hairer, M., Hutzenthaler, M., Jentzen, A., Loss of regularity for Kolmogorov equations, *Ann. Probab.,* **43** (2015), 468-527.

[8] Higham, Desmond J., Mao, X., Yuan, C., Almost sure and moment exponential stability in the numerical simulation of stochastic differential equations, *SIAM J. Numer. Anal.*, **45** (2007), 592–609.

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

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

[11] Jentzen, A., Müller-Gronbach, T., Yaroslavtseva, L., On stochastic differential equations with arbitrary slow convergence rates for strong approximation, *Commun. Math. Sci.,* **14** (2016), 1477-1500.

[12] Kloeden, P. E., Platen, E., *Numerical solution of stochastic differential equations*, Springer, 1992, Berlin.

[13] Kunita, H., Stochastic differential equations and stochastic flows of diffeomorphisms, École d'été de probabilités de Saint-Flour, XII-1982, 143–303, Lecture Notes in Math., 1097, Springer, Berlin, 1984.

[14] Krylov, N. V., Röckner, M., Strong solutions of stochastic equations with singular time dependent drift, *Probab. Theory Related Fields*, **131** (2005), 154–196.

[15] Leobacher, G., Szögyenyi, M., Convergence of the Euler-Maruyama method for multidimensional SDEs with discontinuous drift and degenerate diffusion coefficient, arXiv:1610.07047v2.

[16] Mattingly, J. C., Stuart, A. M., Higham, D. J., Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise, *Stochastic Process. Appl.*, **101** (2002), 185–232.

[17] Ngo, H.-L., Taguchi, D., Strong rate of convergence for the Euler-Maruyama approximation of stochastic differential equations with irregular coefficients, *Math. Comp.*, **85** (2016), 1793–1819.

[18] Ngo, H.-L., Taguchi, D., On the Euler-Maruyama approximation for one-dimensional stochastic differential equations with irregular coefficients, arXiv:1509.06532v1.

[19] Pamen, O. M., Taguchi, D., Strong rate of convergence for the Euler-Maruyama approximation of SDEs with Hölder continuous drift coefficient, arXiv1508.07513v1.

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

[21] Soize, C., *The Fokker-Planck equation for stochastic dynamical systems and its explicit steady state solutions*, World Scientific Publishing Co., Inc., River Edge, NJ, 1994.

[22] Wang, F.-Y., Hypercontractivity for Stochastic Hamiltonian Systems, arXiv:1409.1995.

[23] Wang, F.-Y., Gradient estimates and applications for SDEs in Hilbert space with multiplicative noise and Dini continuous drift, *J. Differential Equations*, **260** (2016), 2792–2829.

[24] Wang, F.-Y., Zhang, X., Degenerate SDE with Hölder-Dini Drift and Non-Lipschitz Noise Coefficient, *SIAM J. Math. Anal.*, **48** (2016), 2189–2226.

[25] Wang, F.-Y., Zhang, X., Derivative formula and applications for degenerate diffusion semigroups, *J. Math. Pures Appl.*, **99** (2013), 726–740.

[26] Yuan, C., Mao, X., A note on the rate of convergence of the Euler-Maruyama method for stochastic differential equations, *Stoch. Anal. Appl.*, **26** (2008), 325–333.

[27] Zhang, X., Stochastic flows and Bismut formulas for stochastic Hamiltonian systems, *Stoch. Proc. Appl.*, **120** (2010), 1929–1949.

[28] Zvonkin, A. K., A transformation of the phase space of a diffusion process that removes the drift, *Math. Sb.* , 93 (1974), 129-149.