Available online at www.sciencedirect.com

ScienceDirect

Journal of Differential Equations 459 (2026) 114104

Journal of Differential Equations

www.elsevier.com/locate/jde

Densities of stochastic functional differential equation and its discretizations ☆

Chuchu Chen [a,b], Tonghe Dang [a,d,*], Jialin Hong [a,b], Guoting Song [c]

[a] State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

[b] School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

[c] School of Mathematics and Statistics, Changchun University, Changchun 130022, China

[d] Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China

Received 18 October 2024; revised 6 December 2025; accepted 6 January 2026

Abstract

This paper studies densities for solutions of the stochastic functional differential equation (SFDE) and of its Euler-type discretizations. First, by means of the Malliavin calculus, we prove the existence of densities for the exact solution and its discretizations. Then we establish the $L^1(\mathbb{R}^d)$-convergence for the density of discretizations by implementing a dimensionality reduction argument and a localization argument. Further, we prove that the pointwise convergence rate of the density is 1 when the noise is of additive type. The convergence results indicate that the total variation distance between laws of solutions for the SFDE and its discretizations vanishes to zero as the discretization parameter diminishes, while that between laws of functional solutions fails to vanish due to the high degeneracy of the equation. This finding highlights one of the main distinctions in asymptotic behaviors of the corresponding discretized systems when compared to stochastic ordinary (partial) differential equations.

☆ This work is funded by the National Key R&D Program of China under Grant (No. 2020YFA0713701), the National Natural Science Foundation of China (No. 12031020, No. 12461160278, and No. 12471386), and by Youth Innovation Promotion Association CAS, China.

* Corresponding author at: State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China, and Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China.

E-mail addresses: chenchuchu@lsec.cc.ac.cn (C. Chen), dangth@lsec.cc.ac.cn (T. Dang), hjl@lsec.cc.ac.cn (J. Hong), songgt606@lsec.cc.ac.cn (G. Song).

https://doi.org/10.1016/j.jde.2026.114104

MSC: 34K50; 60H07; 65C30

Keywords: Density; Stochastic functional differential equation; Euler-type discretizations; Convergence

1. Introduction

Stochastic functional differential equations (SFDEs) serve as essential mathematical models for capturing the intricate dynamics of systems influenced by both randomness and time delays. They have wide applications in various fields including finance, biology, and engineering. Understanding the intrinsic behaviors of solutions of SFDEs is of great importance in both theoretical and practical contexts. It is known that the density of a solution process at a given time, describing the probability law, is one of the most essential characteristics that reveals behaviors of the solution; see [1,2] and references therein. However, the research on densities for solutions of SFDEs remains underexplored, and the existing works mainly focus on the globally Lipschitz drift case. For instance, authors in [3,4] study the existence and smoothness for the density of exact solution of SFDEs; Authors in [5] investigate the asymptotic behavior for the perturbed densities for SFDEs with small noise.

In this paper, we consider the following SFDE

$$\begin{cases} \mathrm{d}x^{\xi}(t) = b(x_t^{\xi})\mathrm{d}t + \sigma(x_t^{\xi})\mathrm{d}W(t), & t \in (0, T], \\ x^{\xi}(t) = \xi(t), & t \in [-\tau, 0], \end{cases} \tag{1}$$

where $T > 0$, the delay $\tau > 0$, the initial datum $\xi \in \mathcal{C}([-\tau, 0]; \mathbb{R}^d)$, $\{W(t)\}_{t\geq 0}$ is an m-dimensional standard Brownian motion defined on a filtered complete probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$, and for $t \geq 0$, $x_t^{\xi} : r \mapsto x^{\xi}(t + r)$ is the $\mathcal{C}([-\tau, 0]; \mathbb{R}^d)$-valued functional solution. Here, the drift coefficient $b : \mathcal{C}([-\tau, 0]; \mathbb{R}^d) \to \mathbb{R}^d$ and the diffusion coefficient $\sigma : \mathcal{C}([-\tau, 0]; \mathbb{R}^d) \to \mathbb{R}^{d\times m}$ are continuous and measurable functions; see Section 2 for concrete assumptions. The aim of this paper is twofold:

(i) to investigate the existence of densities for solutions of (1) and its discretized systems when the drift grows superlinearly;
(ii) to study the convergence of the density for solutions of discretized systems, with the hope of revealing the distinction in corresponding asymptotic behaviors between SFDEs and stochastic ordinary (partial) differential equations.

For the first aim, we establish that the exact solution of (1) with superlinearly growing drift admits a density, by analyzing the invertibility of the Malliavin covariance matrix of the exact solution. Next, we introduce a class of Euler-type discretizations, namely the θ-Euler–Maruyama (θ-EM) discretization with $\theta \in (\frac{1}{2}, 1]$ for (1), to obtain the discretized systems. Then we investigate the properties of the corresponding density. Due to the time-lag effect, the SFDE and its discretizations exhibit high degeneracy, characterized as infinite-dimensional Markov processes influenced by finite-dimensional noises, presenting challenges in analyzing their densities. To deal with this challenge, we implement a dimensionality reduction argument by constructing basis functions of interpolation in $\mathcal{C}([-\tau, 0]; \mathbb{R}^d)$. This argument enables us to demonstrate the invertibility of the Malliavin covariance matrix of the discretization and further establish the existence of its density when the drift grows superlinearly.

For the second aim, we study the convergence of the density for the θ-EM discretization, by employing a localization argument to overcome the obstacle caused by the superlinear growth of coefficients. We show that the density of the discretization converges to the exact density in $L^1(\mathbb{R}^d)$ over a finite time horizon. Especially, in the case of linearly growing drift and additive noise, we analyze the error between densities and obtain the pointwise convergence rate 1, based on the test-functional-independent weak convergence analysis of the discretization. A key ingredient in this analysis lies in the full use of the Malliavin integration by parts formula. This requires negative moment estimates of the determinant for the corresponding Malliavin covariance matrix of the discretization, which is derived by presenting a discrete comparison principle for the SFDE with additive noise. The convergence result for densities yields that the total variation distance between laws of the solutions for the SFDE and its discretizations vanishes to zero as the discretization parameter diminishes. In contrast, we find that the total variation distance between laws of functional solutions does not vanish, highlighting the distinctive behaviors for solutions of SFDEs that come from the time-lag effect. This phenomenon introduces more complexity to the system, leading to a rich interaction between historical influences and stochastic perturbations.

At last, in order to illustrate the distinctions in asymptotic behaviors between the discretized systems of SFDEs and those of stochastic ordinary or partial differential equations, we mention some related results regarding the convergence of discretizations in total variation distance. Building on existing studies concerning the convergence of densities for discretizations, one can derive convergence results in total variation distance for solutions of these discretizations. For stochastic ordinary differential equations, the convergence in total variation distance for laws of Itô–Taylor-type discretizations can be obtained from the corresponding density convergence results studied in [6–11]. By further leveraging properties of densities, one can also derive the convergence rate in total variation distance; see e.g. [12–14]. For stochastic partial differential equations, the solution can be understood as the real-valued random field or the Hilbert-valued stochastic process. The convergence in total variation distance differs between these two types of solutions. Utilizing the density convergence of discretizations, one can establish the convergence in total variation distance for random field solutions of discretizations at fixed spatial and temporal variables; see e.g. [15,16]. The convergence behavior in total variation distance for the Hilbert-valued solutions of temporal semi-discretizations depends on the choice of discretizations, with some choices failing to converge; see e.g. [17,18].

The outline of this paper is as follows. In Section 2, we focus on the SFDE (1) with superlinearly growing drift coefficient and the corresponding θ-EM discretization, and obtain the existence of densities for both the exact solution and its discretization. In Section 3, we investigate the convergence of densities of discretizations. Section 4 is devoted to proofs of moment estimates for the exact solution and its discretizations.

2. Densities of exact solution and its discretization

In this section, we focus on the SFDE (1) with superlinearly growing drift coefficient and the corresponding θ-EM discretization, and investigate the existence of densities for both the exact solution and its discretizations, based on the technique of the Malliavin calculus. We first present some preliminaries, including some notation used in this paper, a brief introduction to the Malliavin calculus, and precise assumptions on coefficients.

2.1. Preliminaries

Throughout this paper, the following notation is used. Let $\mathbb{N}$ and $\mathbb{N}_+$ denote the sets of the non-negative integers and the positive integers, respectively. We use $|\cdot|$ to denote both the Euclidean norm in $\mathbb{R}^d$ and the trace norm in $\mathbb{R}^{d\times m}$, and use $\langle x, y\rangle$ to denote the inner product of x, y in $\mathbb{R}^d$. Let $\mathbf{1}_A(\cdot)$ be the indicator function of the set A, i.e., $\mathbf{1}_A(x) = 1$ for $x \in A$ and $\mathbf{1}_A(x) = 0$ for $x \in A^c$. Denote by $\mathcal{D}$ (resp. $\mathscr{D}$) the Gâteaux (resp. Fréchet) derivative operator and by $\mathcal{D}^\alpha$ the Gâteaux derivative operator of order $\alpha \in \mathbb{N}_+$. Denote by D the Malliavin derivative operator and by D^α the Malliavin derivative operator of order $\alpha \in \mathbb{N}_+$. To simplify the notation, we denote by $\mathcal{C}^d := (\mathcal{C}([-\tau, 0]; \mathbb{R}^d), \|\cdot\|)$ the space of all continuous functions $\phi : [-\tau, 0] \to \mathbb{R}^d$ equipped with the norm $\|\phi\| = \sup_{s\in[-\tau,0]} |\phi(s)|$. For an integer $k \geq 2$, the space $\mathcal{L}\big((\mathcal{C}^d)^{\otimes k}; \mathbb{R}^d\big)$ denotes the collection of all bounded k-linear operators from $(\mathcal{C}^d)^{\otimes k}$ to $\mathbb{R}^d$. Namely, for every operator $B : (\mathcal{C}^d)^{\otimes k} \to \mathbb{R}^d$ with $B \in \mathcal{L}\big((\mathcal{C}^d)^{\otimes k}; \mathbb{R}^d\big)$,

$$B(\lambda_1 x_1, \dots, \lambda_k x_k) = \lambda_1 \cdots \lambda_k\, B(x_1, \dots, x_k) \qquad \forall\, \lambda_i \in \mathbb{R},\ x_i \in \mathcal{C}^d,$$

and there exists a constant $K > 0$ such that $|B(x_1, \dots, x_k)| \leq K \prod_{i=1}^k \|x_i\|_{\mathcal{C}^d}$ for $x_i \in \mathcal{C}^d$. Throughout this paper, K denotes a generic positive constant independent of the step size, whose value may vary at different occurrences.

Now we give a brief introduction to the Malliavin calculus. Let $\mathbb{T} := [0, T]$ with $T > 0$. Let H be the Hilbert space $L^2(\mathbb{T}; \mathbb{R}^m)$ endowed with the inner product $\langle g, h\rangle_H := \int_{\mathbb{T}} g(t)^\top h(t)\mathrm{d}t$ for $g, h \in H$, and $\mathcal{C}_0(\mathbb{T}; \mathbb{R}^m)$ be the space of all continuous functions $u : \mathbb{T} \to \mathbb{R}^m$ with $u(0) = 0$. By identifying $W(t, \omega)$ with the value $\omega(t)$ at time t of an element $\omega \in \mathcal{C}_0(\mathbb{T}; \mathbb{R}^m)$, we take $\Omega = \mathcal{C}_0(\mathbb{T}; \mathbb{R}^m)$ as the Wiener space and $\tilde{\mathbb{P}}$ as the Wiener measure. For $g = (g^1, \dots, g^m)^\top \in H$, we set $W(g) := \sum_{k=1}^m \int_{\mathbb{T}} g^k(t)\mathrm{d}W^k(t)$, where $W(t) = (W^1(t), \dots, W^m(t))^\top$. Denote by $\mathcal{S}$ the class of smooth random variables such that $G \in \mathcal{S}$ has the form $G = f(W(g_1), \dots, W(g_n))$, where $f \in \mathcal{C}^\infty_{pol}(\mathbb{R}^n; \mathbb{R})$, $g_i \in H$, $i = 1, \dots, n$. Here, $\mathcal{C}^\infty_{pol}(\mathbb{R}^n; \mathbb{R})$ is the space of all real-valued smooth functions on $\mathbb{R}^n$ whose partial derivatives have at most polynomial growth. The Malliavin derivative of a smooth random variable G is an H-valued random variable given by $DG = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(W(g_1), \dots, W(g_n))g_i$, which is also an m-dimensional stochastic process $DG = \{D_r G, r \in \mathbb{T}\}$ with $D_r G = \sum_{i=1}^n \partial_i f(W(g_1), \dots, W(g_n))g_i(r)$. For any $p \geq 1$, we denote the domain of D in $L^p(\Omega)$ by $\mathbb{D}^{1,p}(\mathbb{R})$, meaning that $\mathbb{D}^{1,p}(\mathbb{R})$ is the closure of $\mathcal{S}$ with respect to the norm $\|G\|_{1,p} := (\mathbb{E}[|G|^p + \|DG\|_H^p])^{\frac{1}{p}}$.

For $\alpha \in \mathbb{N}_+$, the iterated derivative $D^\alpha G$ is a random variable with value in $H^{\otimes\alpha}$. For any $p \geq 1$ and $\alpha \in \mathbb{N}_+$, denote by $\mathbb{D}^{\alpha,p}(\mathbb{R})$ the completion of $\mathcal{S}$ with respect to the norm $\|G\|_{\alpha,p} := \big(\mathbb{E}\big[|G|^p + \sum_{j=1}^\alpha \|D^j G\|^p_{H^{\otimes j}}\big]\big)^{\frac{1}{p}}$. Define $\mathbb{D}^{\alpha,\infty}(\mathbb{R}) := \cap_{p\geq 1}\mathbb{D}^{\alpha,p}(\mathbb{R})$, $\mathbb{D}^\infty(\mathbb{R}) := \cap_{p\geq 1}\cap_{\alpha\geq 1}\mathbb{D}^{\alpha,p}(\mathbb{R})$, $L^{\infty-}(\Omega; \mathbb{R}) := \cap_{p\geq 1} L^p(\Omega; \mathbb{R})$. Similarly, let V be a real separable Banach space and define the space $\mathbb{D}^{k,p}(V)$ as the completion of V-valued smooth random variables with respect to the norm $\|G\|_{k,p,V} = (\mathbb{E}[\|G\|_V^p + \sum_{j=1}^k \|D^j G\|^p_{H^{\otimes j}\otimes V}])^{\frac{1}{p}}$. In this case, the corresponding spaces are denoted by $\mathbb{D}^{k,\infty}(V)$, $\mathbb{D}^\infty(V)$, and $L^{\infty-}(\Omega; V)$ respectively. For simplicity of notation, when $V = \mathbb{R}^d$, we abbreviate them as $\mathbb{D}^{\alpha,p}$, $\mathbb{D}^{k,\infty}$, $\mathbb{D}^\infty$, $L^{\infty-}(\Omega)$.

We introduce the one-sided Lipschitz condition on the drift coefficient b; see e.g. [19–21] for such conditions on SFDEs.

Assumption 2.1. *There exist a constant $L_1 > 0$ and a probability measure ν_1 on $[-\tau, 0]$ such that for any $\phi_1, \phi_2 \in \mathcal{C}^d$,*

$$\langle b(\phi_1) - b(\phi_2), \phi_1(0) - \phi_2(0)\rangle \leq L_1\Big(|\phi_1(0) - \phi_2(0)|^2 + \int_{-\tau}^{0} |\phi_1(r) - \phi_2(r)|^2 \mathrm{d}\nu_1(r)\Big).$$

In addition, assume that b has continuous Gâteaux derivative, and that there exists a constant $\beta \geq 0$ such that $|\mathcal{D}b(\phi_1)\phi_2| \leq K(1 + \|\phi_1\|^{\beta})\|\phi_2\|$, where $\phi_1, \phi_2 \in \mathcal{C}^d$ and $K > 0$.

We impose the globally Lipschitz condition and the uniform non-degeneracy condition on the diffusion coefficient σ.

Assumption 2.2. *There exist a constant $L_2 > 0$ and a probability measure ν_2 on $[-\tau, 0]$ such that for any $\phi_1, \phi_2 \in \mathcal{C}^d$,*

$$|\sigma(\phi_1) - \sigma(\phi_2)|^2 \leq L_2\Big(|\phi_1(0) - \phi_2(0)|^2 + \int_{-\tau}^{0} |\phi_1(r) - \phi_2(r)|^2 \mathrm{d}\nu_2(r)\Big).$$

In addition, assume that σ has continuous Gâteaux derivative, and that there exists some $\sigma_0 > 0$ such that

$$\inf_{\phi \in \mathcal{C}^d} \min_{u \in \mathbb{R}^d, |u|=1} u^{\top}\sigma(\phi)\sigma(\phi)^{\top}u \geq \sigma_0.$$

Remark 2.1. Once coefficients b and σ have continuous Gâteaux derivatives, it follows from Assumptions 2.1 and 2.2 that for any $\phi, \phi_1 \in \mathcal{C}^d$,

$$\langle \mathcal{D}b(\phi_1)\phi, \phi(0)\rangle \leq L_1\big(|\phi(0)|^2 + \int_{-\tau}^{0} |\phi(r)|^2 \mathrm{d}\nu_1(r)\big), \tag{2}$$

$$|\mathcal{D}\sigma(\phi_1)\phi|^2 \leq L_2\big(|\phi(0)|^2 + \int_{-\tau}^{0} |\phi(r)|^2 \mathrm{d}\nu_2(r)\big). \tag{3}$$

We also remark that the classical example

$$b(\phi) = -|\phi(0)|^2\phi(0) + \int_{-\tau}^{0} \phi(r)\mathrm{d}\nu_1(r), \quad \phi \in \mathcal{C}^d \tag{4}$$

for the superlinearly growing drift coefficient is included, and in this case,

$$\mathcal{D}b(\phi_1)\phi = \int_{-\tau}^{0} \phi(s)\mathrm{d}\nu_1(s) - 2\phi_1(0)^{\top}\phi(0)\phi_1(0) - |\phi_1(0)|^2\phi(0).$$

2.2. Density of exact solution

In this subsection, we present that the exact solution of (1) admits a density. Under Assumptions 2.1 and 2.2, the existence and uniqueness of the solution of (1) can be obtained by using [22, Theorem 2]. In addition, the functional solution of (1) has the following moment boundedness, whose proof is given in Section 4.

Lemma 2.1. *Let Assumptions 2.1 and 2.2 hold. Then for any* $p \geq 2$,

$$\mathbb{E}\Big[\sup_{t\in[0,T]} \|x_t^\xi\|^p\Big] \leq K_T, \qquad \sup_{r\in[0,T]} \mathbb{E}\Big[\sup_{t\in[r,T]} \|D_r x_t^\xi\|^p\Big] \leq K_T.$$

Based on Lemma 2.1 and $D_r x^\xi(t) = 0$ for $r > t$, we derive that $x^\xi(t) \in \mathbb{D}^{1,p}$ for all $p \geq 1$. Hence, according to [1, Theorem 2.1.2], in order to obtain the existence of the density of the solution of (1), it suffices to show the a.s. invertibility of the Malliavin covariance matrix $\gamma^E(t)$ of $x^\xi(t)$, where

$$\gamma^E(t) := \int_0^t D_r x^\xi(t)(D_r x^\xi(t))^\top \mathrm{d}r, \quad t \in [0, T].$$

To this end, by virtue of [23, Section 3.3], we first prove that for some $q > 0$, there exists a small number $\varepsilon_0(q)$ such that for any $\varepsilon < \varepsilon_0(q)$, $\sup_{u\in\mathbb{R}^d,|u|=1} \mathbb{P}(u^\top \gamma^E(t) u \leq \varepsilon) \leq K_T \varepsilon^q$, $t \in [0, T]$. This is stated as follows with $q = 1$.

Proposition 2.1. *Under Assumptions 2.1 and 2.2, for any* $\varepsilon \in (0, 1)$, *it holds that*

$$\sup_{u\in\mathbb{R}^d,|u|=1} \mathbb{P}(u^\top \gamma^E(t) u \leq \varepsilon) \leq K_T \varepsilon,\ t \in [0, T].$$

Proof. For any $r \leq t$, by the chain rule of the Malliavin derivative (see e.g. [5]),

$$D_r x^\xi(t) = \int_r^t \mathcal{D}b(x_s^\xi) D_r x_s^\xi \mathrm{d}s + \int_r^t \mathcal{D}\sigma(x_s^\xi) D_r x_s^\xi \mathrm{d}W(s) + \sigma(x_r^\xi)\mathbf{1}_{[0,t]}(r).$$

Fixing $\varepsilon \in (0, 1)$ and letting $\varepsilon_1 := \frac{2\varepsilon}{\sigma_0}$, we obtain

$$\begin{aligned} u^\top \gamma^E(t) u &\geq \int_{t-\varepsilon_1}^t u^\top D_r x^\xi(t)(D_r x^\xi(t))^\top u \mathrm{d}r \geq \int_{t-\varepsilon_1}^t u^\top \sigma(x_r^\xi)(\sigma(x_r^\xi))^\top u \mathrm{d}r \\ &+ 2\int_{t-\varepsilon_1}^t u^\top \Big(\int_r^t \mathcal{D}b(x_s^\xi) D_r x_s^\xi \mathrm{d}s + \int_r^t \mathcal{D}\sigma(x_s^\xi) D_r x_s^\xi \mathrm{d}W(s)\Big)(\sigma(x_r^\xi))^\top u \mathrm{d}r. \end{aligned}$$

It follows from Assumption 2.2 that

$$\int_{t-\varepsilon_1}^{t} u^\top \sigma(x_r^\xi)\sigma(x_r^\xi)^\top u \mathrm{d}r \geq \varepsilon_1 \sigma_0 = 2\varepsilon.$$

This, along with the Chebyshev inequality, the Hölder inequality, and the Burkholder–Davis–Gundy inequality implies that for $u \in \mathbb{R}^d$ with $|u| = 1$,

$$\begin{aligned}
&\mathbb{P}(u^\top \gamma^E(t) u \leq \varepsilon) \\
&\leq \mathbb{P}\Big(2\Big|\int_{t-\varepsilon_1}^{t} u^\top \Big(\int_r^t \mathcal{D}b(x_s^\xi) D_r x_s^\xi \mathrm{d}s + \int_r^t \mathcal{D}\sigma(x_s^\xi) D_r x_s^\xi \mathrm{d}W(s)\Big)(\sigma(x_r^\xi))^\top u \mathrm{d}r\Big| \geq \varepsilon\Big) \\
&\leq 8\varepsilon^{-2}\mathbb{E}\Big[\Big|\int_{t-\varepsilon_1}^{t} u^\top \int_r^t \mathcal{D}b(x_s^\xi) D_r x_s^\xi \mathrm{d}s (\sigma(x_r^\xi))^\top u \mathrm{d}r\Big|^2\Big] \\
&\quad + 8\varepsilon^{-2}\mathbb{E}\Big[\Big|\int_{t-\varepsilon_1}^{t} u^\top \int_r^t \mathcal{D}\sigma(x_s^\xi) D_r x_s^\xi \mathrm{d}W(s)(\sigma(x_r^\xi))^\top u \mathrm{d}r\Big|^2\Big] \\
&\leq K\varepsilon^2\Big(\sup_{0\leq r\leq T} \mathbb{E}\Big[\sup_{r\leq s\leq T} |\mathcal{D}b(x_s^\xi) D_r x_s^\xi|^4\Big]\Big)^{\frac{1}{2}}\Big(\sup_{0\leq r\leq T} \mathbb{E}[|\sigma(x_r^\xi)|^4]\Big)^{\frac{1}{2}} \\
&\quad + K\varepsilon\Big(\sup_{0\leq r\leq T} \mathbb{E}\Big[\sup_{r\leq s\leq T} |\mathcal{D}\sigma(x_s^\xi) D_r x_s^\xi|^4\Big]\Big)^{\frac{1}{2}}\Big(\sup_{0\leq r\leq T} \mathbb{E}[|\sigma(x_r^\xi)|^4]\Big)^{\frac{1}{2}} \leq \varepsilon K_T,
\end{aligned}$$

where in the last inequality we used (2), (3) and Lemma 2.1. The proof is finished. □

The existence of a density for the solution of (1) is stated as follows.

Theorem 2.1. *Under conditions in Proposition 2.1, for any $t \in (0, T]$, the law of $x^\xi(t)$ admits a density, denoted by $\mathfrak{p}(t, \cdot)$.*

Proof. Based on Proposition 2.1, we have

$$\sup_{u\in\mathbb{R}^d, |u|=1} \mathbb{P}(u^\top \gamma^E(t) u = 0) \leq \sup_{u\in\mathbb{R}^d, |u|=1} \mathbb{P}(u^\top \gamma^E(t) u \leq \varepsilon) \leq K_T \varepsilon \quad \forall \varepsilon \in (0, 1),$$

which shows the a.s. invertibility of $\gamma^E(t)$ due to the arbitrariness of ε. This, together with $x^\xi(t) \in \mathbb{D}^{1,p}$ and [1, Theorem 2.1.2] completes the proof. □

2.3. Density of θ-EM discretization

In this subsection, we introduce the θ-EM discretization with $\theta \in (\frac{1}{2}, 1]$ for the SFDE (1), and present the existence of its density. Without loss of generality, it is assumed that τ is a multiple of the step size Δ, and that T is a multiple of τ. Then there exist two numbers $N, N^\Delta \in \mathbb{N}_+$ such that $\tau = N\Delta \in (0, 1]$ and $T = N^\Delta \Delta$. Let $t_k = k\Delta$ for $k \in \{-N, \ldots, N^\Delta\}$. Introduce the θ-EM discretization as follows: for any $\xi \in \mathcal{C}^d$,

$$\begin{cases} y^{\xi,\Delta}(t_k) = \xi(t_k), \quad -N \le k \le 0, \\ y^{\xi,\Delta}(t_{k+1}) = y^{\xi,\Delta}(t_k) + (1-\theta)b(y^{\xi,\Delta}_{t_k})\Delta + \theta b(y^{\xi,\Delta}_{t_{k+1}})\Delta + \sigma(y^{\xi,\Delta}_{t_k})\delta W_k, \ 0 \le k \le N^\Delta - 1, \end{cases} \tag{5}$$

where $\theta \in (\frac{1}{2}, 1]$, $\delta W_k = W(t_{k+1}) - W(t_k)$, and $y^{\xi,\Delta}_{t_k}$ is a $\mathcal{C}^d$-valued random variable defined by the linear interpolation

$$y^{\xi,\Delta}_{t_k}(s) = \frac{t_{j+1}-s}{\Delta} y^{\xi,\Delta}(t_{k+j}) + \frac{s-t_j}{\Delta} y^{\xi,\Delta}(t_{k+j+1}) \tag{6}$$

for $s \in [t_j, t_{j+1}],\ j \in \{-N, \dots, -1\}$. We call $\{y^{\xi,\Delta}(t_k)\}_{k=-N}^{\infty}$ the θ-*EM solution* and $\{y^{\xi,\Delta}_{t_k}\}_{k=0}^{\infty}$ the θ-*EM functional solution*. Under Assumptions 2.1 and 2.2, the solution of (5) exists uniquely for any $\Delta \in (0, \frac{1}{2\theta L_1})$, whose proof is similar to that of [21, Lemma 3.2] and thus is omitted.

Note that the θ-EM functional solution depends on the past state, causing the main difficulty in the analysis of the density for discretizations. To deal with this difficulty, we implement a dimensionality reduction argument of the interpolation. To this end, we introduce an $(N+1)$-dimensional linear interpolation space $\mathcal{C}^{Int}$, which consists of piecewise linear functions from $[-\tau, 0]$ to $\mathbb{R}$, with basis functions given as follows:

$$\begin{aligned} I^{[-N]}(s) &:= \frac{N}{\tau}(t_{-N+1} - s)\mathbf{1}_{\Delta_{-N}}(s), \\ I^{[j]}(s) &:= \frac{N}{\tau}(s - t_{j-1})\mathbf{1}_{\Delta_{j-1}}(s) + \frac{N}{\tau}(t_{j+1} - s)\mathbf{1}_{\Delta_j}(s), \quad j = -N+1, \dots, -1, \\ I^{[0]}(s) &:= \frac{N}{\tau}(s - t_{-1})\mathbf{1}_{\Delta_{-1}}(s), \end{aligned} \tag{7}$$

where $\Delta_j := [t_j, t_{j+1})$ for $j = -N, \dots, -2$, and $\Delta_{-1} := [t_{-1}, 0]$. Then by (6), the θ-EM functional solution $y^{\xi,\Delta}_{t_k}$ can be represented as

$$\begin{aligned} y^{\xi,\Delta}_{t_k}(s) &= \sum_{j=-N}^{-1} \mathbf{1}_{\Delta_j}(s)\Big[y^{\xi,\Delta}_{t_k}(t_{j+1})\frac{N}{\tau}(s-t_j) + y^{\xi,\Delta}_{t_k}(t_j)\frac{N}{\tau}(t_{j+1}-s)\Big] \\ &= \sum_{j=-N}^{0} I^{[j]}(s) y^{\xi,\Delta}(t_{k+j}), \quad s \in [-\tau, 0]. \end{aligned} \tag{8}$$

Thus, the $\mathcal{C}^d$-valued random variable $y^{\xi,\Delta}_{t_k}$ is transformed into a $\mathcal{C}^{Int} \otimes \mathbb{R}^d$-valued random variable. We call it the dimensionality reduction. Here, $\mathcal{C}^{Int} \otimes \mathbb{R}^d$ denotes the tensor product space of $\mathcal{C}^{Int}$ and $\mathbb{R}^d$. An element $\mathfrak{u} : [-\tau, 0] \to \mathbb{R}^d$ in this space is of the form $\sum_{j=-N}^{0} I^{[j]}(\cdot)\mathfrak{u}_j$, where $I^{[j]} \in \mathcal{C}^{Int}$ and $\mathfrak{u}_j \in \mathbb{R}^d$.

The following lemma shows the moment boundedness of the θ-EM functional solution, whose proof is postponed to Section 4.

Lemma 2.2. *Let Assumptions 2.1 and 2.2 hold. Then for $\Delta \in (0, \Delta_0)$ with some $\Delta_0 \in (0, 1)$, $p \ge 2$, and $T > 0$,*

$$\mathbb{E}\Big[\sup_{t_k\in[0,T]}\|y^{\xi,\Delta}_{t_k}\|^p\Big]\le K_T,\qquad \sup_{r\in[0,T]}\mathbb{E}\Big[\sup_{t_k\in[r,T]}\|D_r y^{\xi,\Delta}_{t_k}\|^p\Big]\le K_T.$$

As a consequence, we obtain $y^{\xi,\Delta}(t_k)\in\mathbb{D}^{1,p}$. Below we show the existence of the density of discretizations.

Theorem 2.2. *Under Assumptions 2.1 and 2.2, for $\Delta\in(0,\Delta_0)$ with some $\Delta_0\in(0,1)$ and $k\in\{1,\dots,N^{\Delta}\}$, the law of $y^{\xi,\Delta}(t_k)$ admits a density, denoted by $\mathfrak{p}^{\Delta}(t_k,\cdot)$.*

Proof. Similar to the proof of Theorem 2.1, it suffices to show that the Malliavin covariance matrix of $y^{\xi,\Delta}(t_k)$, defined by

$$\gamma_k:=\int_0^{t_k} D_r y^{\xi,\Delta}(t_k)(D_r y^{\xi,\Delta}(t_k))^{\top}\mathrm{d}r,$$

is a.s. invertible.

Taking the Malliavin derivatives on both sides of (5), we derive from $D_r\delta W_k=\mathbf{1}_{[t_k,t_{k+1}]}(r)$ and the chain rule of the Malliavin derivative that for any $r\in[0,T]$,

$$\begin{aligned}D_r y^{\xi,\Delta}(t_{k+1})&=D_r y^{\xi,\Delta}(t_k)+(1-\theta)\mathcal{D}b(y^{\xi,\Delta}_{t_k})D_r y^{\xi,\Delta}_{t_k}\Delta+\theta\mathcal{D}b(y^{\xi,\Delta}_{t_{k+1}})D_r y^{\xi,\Delta}_{t_{k+1}}\Delta\\&\quad+\mathcal{D}\sigma(y^{\xi,\Delta}_{t_k})D_r y^{\xi,\Delta}_{t_k}\delta W_k+\sigma(y^{\xi,\Delta}_{t_k})\mathbf{1}_{[t_k,t_{k+1}]}(r).\end{aligned}$$

It follows from (8) that $D_r y^{\xi,\Delta}_{t_{k+1}}=D_r\Big(\sum_{j=-N}^{-1}I^{[j]}y^{\xi,\Delta}_{t_{k+1}}(t_j)\Big)+D_r y^{\xi,\Delta}(t_{k+1})I^{[0]}$. Hence,

$$\begin{aligned}&D_r y^{\xi,\Delta}(t_{k+1})-\theta\mathcal{D}b(y^{\xi,\Delta}_{t_{k+1}})D_r y^{\xi,\Delta}(t_{k+1})I^{[0]}\Delta\\&=D_r y^{\xi,\Delta}(t_k)+(1-\theta)\mathcal{D}b(y^{\xi,\Delta}_{t_k})D_r\Big(\sum_{j=-N}^{0}I^{[j]}y^{\xi,\Delta}_{t_k}(t_j)\Big)\Delta+\theta\mathcal{D}b(y^{\xi,\Delta}_{t_{k+1}})I^{[-1]}D_r y^{\xi,\Delta}(t_k)\Delta\\&\quad+\theta\mathcal{D}b(y^{\xi,\Delta}_{t_{k+1}})\sum_{j=-N}^{-2}I^{[j]}D_r y^{\xi,\Delta}_{t_{k+1}}(t_j)\Delta+\mathcal{D}\sigma(y^{\xi,\Delta}_{t_k})D_r\Big(\sum_{j=-N}^{0}I^{[j]}y^{\xi,\Delta}_{t_k}(t_j)\Big)\delta W_k\\&\quad+\sigma(y^{\xi,\Delta}_{t_k})\mathbf{1}_{[t_k,t_{k+1}]}(r).\end{aligned}$$

Denote by $\mathrm{Id}_{d\times d}$ the $(d\times d)$-dimensional identity operator. It is straightforward to see that

$$\begin{aligned}&\Big(\mathrm{Id}_{d\times d}-\theta\mathcal{D}b(y^{\xi,\Delta}_{t_{k+1}})(I^{[0]}\mathrm{Id}_{d\times d})\Delta\Big)D_r y^{\xi,\Delta}(t_{k+1})\\&=\sum_{j=-N}^{0}A_{j,k}+\sigma(y^{\xi,\Delta}_{t_k})\mathbf{1}_{[t_k,t_{k+1}]}(r),\end{aligned}\tag{9}$$

where

$$A_{0,k} := \Big(\mathrm{Id}_{d\times d} + (1-\theta)\mathcal{D}b(y^{\xi,\Delta}_{t_k})(I^{[0]}\mathrm{Id}_{d\times d})\Delta$$
$$+ \theta\mathcal{D}b(y^{\xi,\Delta}_{t_{k+1}})(I^{[-1]}\mathrm{Id}_{d\times d})\Delta\Big)D_r y^{\xi,\Delta}(t_k) + \mathcal{D}\sigma(y^{\xi,\Delta}_{t_k})(I^{[0]}D_r y^{\xi,\Delta}(t_k))\delta W_k,$$
$$A_{-N,k} := \Big((1-\theta)\mathcal{D}b(y^{\xi,\Delta}_{t_k})(I^{[-N]}\mathrm{Id}_{d\times d})\Delta\Big)D_r y^{\xi,\Delta}(t_{k-N})$$
$$+ \mathcal{D}\sigma(y^{\xi,\Delta}_{t_k})(I^{[-N]}D_r y^{\xi,\Delta}(t_{k-N}))\delta W_k,$$

and

$$A_{j,k} := \Big((1-\theta)\mathcal{D}b(y^{\xi,\Delta}_{t_k})(I^{[j]}\mathrm{Id}_{d\times d})\Delta + \theta\mathcal{D}b(y^{\xi,\Delta}_{t_{k+1}})(I^{[j-1]}\mathrm{Id}_{d\times d})\Delta\Big)D_r y^{\xi,\Delta}(t_{k+j})$$
$$+ \mathcal{D}\sigma(y^{\xi,\Delta}_{t_k})(I^{[j]}D_r y^{\xi,\Delta}(t_{k+j}))\delta W_k$$

for $j = -1, \ldots, -N+1$. Here, $I^{[j]}\mathrm{Id}_{d\times d}$, $j = -N, \ldots, 0$ are elements of $\mathcal{C}([-\tau, 0]; \mathbb{R}^{d\times d}) \cong \mathcal{C}^d \otimes \mathbb{R}^d$ and thus can be acted upon by the operators $\mathcal{D}b(y^{\xi,\Delta}_{t_k})$, $\mathcal{D}b(y^{\xi,\Delta}_{t_{k+1}})$. It follows from (2) that for any $u \in \mathbb{R}^d$ with $u \neq 0$, one has

$$u^\top \mathcal{D}b(y^{\xi,\Delta}_{t_{k+1}})(I^{[0]}\mathrm{Id}_{d\times d})u \le L_1\Big(|u|^2 + \int_{-\tau}^{0} |I^{[0]}\mathrm{Id}_{d\times d}u|^2 \mathrm{d}\nu_1(r)\Big)$$
$$\le L_1\Big(|u|^2 + \int_{-\tau}^{0} \Big|\frac{s-t_{-1}}{\Delta}\mathbf{1}_{\Delta_{-1}}(s)\Big|^2 |u|^2 \mathrm{d}\nu_1(r)\Big) \le 2L_1|u|^2,$$

where we used $I^{[0]}(0) = 1$. Then for $\Delta \in (0, \frac{1}{4\theta L_1})$,

$$u^\top\Big(\mathrm{Id}_{d\times d} - \theta\mathcal{D}b(y^{\xi,\Delta}_{t_{k+1}})(I^{[0]}\mathrm{Id}_{d\times d})\Delta\Big)u \ge (1-2\theta L_1\Delta)|u|^2 > 0,$$

which implies that $\mathrm{Id}_{d\times d} - \theta\mathcal{D}b(y^{\xi,\Delta}_{t_{k+1}})(I^{[0]}\mathrm{Id}_{d\times d})\Delta$ is invertible. Denoting $A_{1,k} := \big(\mathrm{Id}_{d\times d} - \theta\mathcal{D}b(y^{\xi,\Delta}_{t_{k+1}})(I^{[0]}\mathrm{Id}_{d\times d})\Delta\big)^{-1}$, we derive $\|A_{1,k}\|_{\mathcal{L}(\mathbb{R}^d;\mathbb{R}^d)} \in (0, \frac{1}{1-2\theta L_1\Delta})$. According to (9), we arrive at

$$\gamma_{k+1} = \int_0^{t_{k+1}} D_r y^{\xi,\Delta}(t_{k+1})(D_r y^{\xi,\Delta}(t_{k+1}))^\top \mathrm{d}r$$
$$= \int_0^{t_k} A_{1,k}\sum_{j=-N}^{0} A_{j,k}\Big(\sum_{j=-N}^{0} A_{j,k}\Big)^\top (A_{1,k})^\top \mathrm{d}r$$
$$+ \int_{t_k}^{t_{k+1}} A_{1,k}\sigma(y^{\xi,\Delta}_{t_k})\sigma(y^{\xi,\Delta}_{t_k})^\top (A_{1,k})^\top \mathrm{d}r, \tag{10}$$

where we used $D_r y^{\xi,\Delta}(t_k) = 0$ for $r > t_k$. Since $\sigma(x)\sigma(x)^\top$ is positive definite (see Assumption 2.2), we deduce

$$u^\top \gamma_{k+1} u \geq u^\top A_{1,k}\sigma(y_{t_k}^{\xi,\Delta})\sigma(y_{t_k}^{\xi,\Delta})^\top (A_{1,k})^\top u\Delta > 0 \quad a.s.$$

Moreover, utilizing again the invertibility of $\sigma\sigma^\top$, we have that γ_1 is also a.s. invertible. This finishes the proof. □

Remark 2.2. Under conditions in Theorem 2.2, if in addition assume that coefficients b and σ are smooth with bounded derivatives of arbitrary orders, then for each $k \in \mathbb{N}_+$, the law of $y^{\xi,\Delta}(t_k)$ admits a smooth density.

3. Convergence of density for discretizations

In this section, we investigate the convergence of the density of the θ-EM discretization. We first give the convergence of the density in $L^1(\mathbb{R}^d)$ when SFDE (1) has superlinearly growing drift coefficient and multiplicative noise. Then we show that for the case of linearly growing drift coefficient and the additive noise, the convergence rate of the density is 1.

3.1. Convergence of density

In this subsection, we focus on the SFDE (1) with superlinearly growing drift coefficient and the multiplicative noise, and investigate the convergence of the corresponding density for the θ-EM discretization. We present the following assumptions on the second-order derivatives of coefficients and on the Hölder continuity of the initial value.

Assumption 3.1. *The coefficients b and σ have continuous Gâteaux derivatives up to order 2 satisfying*

$$|\mathcal{D}^2 b(\phi_1)(\phi_2,\phi_3)| \leq K(1 + \|\phi_1\|^{(\beta-1)\vee 0})\|\phi_2\|\|\phi_3\|,$$
$$|\mathcal{D}^2 \sigma(\phi_1)(\phi_2,\phi_3)| \leq K\|\phi_2\|\|\phi_3\|,$$

where $\phi_1, \phi_2, \phi_3 \in \mathcal{C}^d$, $K > 0$, and β is given in Assumption 2.1.

We note that the example given in (4) satisfies Assumption 3.1 with $\beta = 2$.

Assumption 3.2. *There exist constants $K > 0$ and $\rho \geq 1/2$ such that*

$$|\xi(s_1) - \xi(s_2)| \leq K|s_1 - s_2|^\rho, \quad s_1, s_2 \in [-\tau, 0].$$

The main result on the convergence of the density concerned in this subsection is stated as follows.

Theorem 3.1. *Let Assumptions 2.1, 2.2, 3.1 and 3.2 hold. Then we have*

$$\lim_{\Delta\to 0} \sup_{0<t_k\leq T} \int_{\mathbb{R}^d} |\mathfrak{p}^\Delta(t_k,x) - \mathfrak{p}(t_k,x)|\mathrm{d}x = 0.$$

Proof. The proof is based on the localization argument and is divided into three steps.

Step 1. Introduce the smooth cut-off functional $\Theta_R : \mathcal{C}^d \to [0, 1]$ with continuous derivatives and compact support. For $R > \|\xi\|$, let $\Theta_R(x) = 1$ for $\|x\| \le R$ and $\Theta_R(x) = 0$ for $\|x\| > R+1$. Then we consider the truncated version of (1) on $[0, T]$,

$$\begin{cases} \mathrm{d}x^{\xi,R}(t) = b^R(x_t^{\xi,R})\mathrm{d}t + \sigma^R(x_t^{\xi,R})\mathrm{d}W(t), \quad t \in (0, T], \\ x^{\xi,R}(t) = \xi, \quad t \in [-\tau, 0], \end{cases} \tag{11}$$

where $b^R(\cdot) := \Theta_R(\cdot)b(\cdot)$ and $\sigma^R(\cdot) := \Theta_R(\cdot)\sigma(\cdot)$ are globally Lipschitz continuous for each R. Moreover, the coefficients b^R and σ^R satisfy

$$|\mathcal{D}b^R(\phi_1)\phi_2|^2 \le K_R\Big(|\phi_2(0)|^2 + \int_{-\tau}^{0} |\phi_2(r)|^2 \mathrm{d}\nu_1(r)\Big), \tag{12}$$

$$|\mathcal{D}\sigma^R(\phi_1)\phi_2|^2 \le K_R\Big(|\phi_2(0)|^2 + \int_{-\tau}^{0} |\phi_2(r)|^2 \mathrm{d}\nu_2(r)\Big), \tag{13}$$

where $\phi_1, \phi_2 \in \mathcal{C}^d$. In addition, the θ-EM solution and θ-EM functional solution for (11) are denoted by $\{y^{\xi,\Delta,R}(t_k)\}_{k=-N}^{\infty}$ and $\{y_{t_k}^{\xi,\Delta,R}\}_{k=0}^{\infty}$, respectively. We need to estimate the error between $x^{\xi,R}(t)$ and $y^{\xi,\Delta,R}(t)$ in $\|\cdot\|_{1,2}$. Here we recall the norm $\|G\|_{1,2} = (\mathbb{E}[|G|^2 + \|DG\|_H^2])^{\frac{1}{2}}$ for an $\mathbb{R}^d$-valued random variable G (see Section 2). For the strong convergence errors of the θ-EM solution and its truncated version, similar to the proof of [21, Theorem 5.3], using Lemmas 2.1 and 2.2, we deduce from Assumptions 2.1, 2.2 and 3.2 that

$$\mathbb{E}\Big[\sup_{0 \le t_k \le T} |x^{\xi}(t_k) - y^{\xi,\Delta}(t_k)|^4\Big] \le K_T \Delta^2 \tag{14}$$

and

$$\mathbb{E}\Big[\sup_{0 \le t_k \le T} |x^{\xi,R}(t_k) - y^{\xi,\Delta,R}(t_k)|^4\Big] \le K_{T,R} \Delta^2. \tag{15}$$

Now we estimate the term

$$\mathbb{E}\big[\|Dx^{\xi,R}(t_k) - Dy^{\xi,\Delta,R}(t_k)\|_H^2\big] = \int_0^T \mathbb{E}\big[|D_r x^{\xi,R}(t_k) - D_r y^{\xi,\Delta,R}(t_k)|^2\big]\mathrm{d}r.$$

Define the auxiliary process of the θ-EM discretization as follows:

$$\begin{cases} z^{\xi,\Delta,R}(t_k) = \xi(t_k), \quad -N \le k \le -1, \\ z^{\xi,\Delta,R}(t_k) = y^{\xi,\Delta,R}(t_k) - \theta b^R(y_{t_k}^{\xi,\Delta,R})\Delta, \quad 0 \le k \le N^{\Delta}. \end{cases}$$

Then for $1 \le k \le N^{\Delta}$,

$$z^{\xi,\Delta,R}(t_k)= z^{\xi,\Delta,R}(t_{k-1})+b^R(y^{\xi,\Delta,R}_{t_{k-1}})\Delta+\sigma^R(y^{\xi,\Delta,R}_{t_{k-1}})\delta W_{k-1},\ 1\le k\le N^\Delta,$$

and the continuous version $\{z^{\xi,\Delta,R}(t)\}_{t\ge -\tau}$ satisfies

$$z^{\xi,\Delta,R}(t)=z^{\xi,\Delta,R}(t_k)+b(y^{\xi,\Delta,R}_{t_k})(t-t_k)+\sigma(y^{\xi,\Delta,R}_{t_k})(W(t)-W(t_k)),$$

for $t\in[t_k,t_{k+1}),\ 0\le k\le N^\Delta$ with the initial datum

$$z^{\xi,\Delta,R}(t)=\frac{t_{k+1}-t}{\Delta}z^{\xi,\Delta,R}(t_k)+\frac{t-t_k}{\Delta}z^{\xi,\Delta,R}(t_{k+1})$$

for $t\in[t_k,t_{k-1}), -N\le k\le -1$. Then by the Hölder inequality and Burkholder–Davis–Gundy inequality, we have

$$\begin{aligned}
&\mathbb{E}[|D_r x^{\xi,R}(t)-D_r z^{\xi,\Delta,R}(t)|^2]\\
&\le K(t-r)\mathbb{E}\Big[\int_r^t \big|\mathcal{D}b^R(x^{\xi,R}_s)D_r x^{\xi,R}_s-\mathcal{D}b^R(y^{\xi,\Delta,R}_{\lfloor s\rfloor})D_r y^{\xi,\Delta,R}_{\lfloor s\rfloor}|^2\mathrm{d}s\Big]\\
&\quad+K\mathbb{E}\Big[\int_r^t \big|\mathcal{D}\sigma^R(x^{\xi,R}_s)D_r x^{\xi,R}_s-\mathcal{D}\sigma^R(y^{\xi,\Delta,R}_{\lfloor s\rfloor})D_r y^{\xi,\Delta,R}_{\lfloor s\rfloor}|^2\mathrm{d}s\Big]\\
&\quad+K\mathbb{E}\Big[|\sigma^R(x^{\xi,R}_r)\mathbf{1}_{[0,t]}(r)-\sigma^R(y^{\xi,\Delta,R}_{\lfloor r\rfloor})\mathbf{1}_{[0,t]}(r)|^2\Big],
\end{aligned}\tag{16}$$

where we used the notation $\lfloor s\rfloor:=t_k$ for $s\in[t_k,t_{k+1})$. It follows from the Taylor formula that

$$\begin{aligned}
&\Big|\mathcal{D}b^R(x^{\xi,R}_s)D_r x^{\xi,R}_s-\mathcal{D}b^R(y^{\xi,\Delta,R}_{\lfloor s\rfloor})D_r y^{\xi,\Delta,R}_{\lfloor s\rfloor}\Big|\\
&\le\Big|\int_0^1 \mathcal{D}^2b^R\big(\varsigma x^{\xi,R}_{\lfloor s\rfloor}+(1-\varsigma)y^{\xi,\Delta,R}_s\big)(x^{\xi,R}_s-y^{\xi,\Delta,R}_{\lfloor s\rfloor},D_r x^{\xi,R}_s)\mathrm{d}\varsigma\Big|\\
&\quad+\Big|\mathcal{D}b^R(y^{\xi,\Delta,R}_{\lfloor s\rfloor})(D_r x^{\xi,R}_s-D_r y^{\xi,\Delta,R}_{\lfloor s\rfloor})\Big|\\
&\le K_R\|x^{\xi,R}_s-y^{\xi,\Delta,R}_{\lfloor s\rfloor}\|\|D_r x^{\xi,R}_s\|+K_R\Big(|D_r x^{\xi,R}(s)-D_r y^{\xi,\Delta,R}(\lfloor s\rfloor)|^2\\
&\quad+\int_{-\tau}^0|D_r x^{\xi,R}_s(v)-D_r y^{\xi,\Delta,R}_{\lfloor s\rfloor}(v)|^2\mathrm{d}\nu_1(v)\Big)^{\frac12},
\end{aligned}$$

where in the last inequality we used the boundedness of operators $\mathcal{D}b^R(\cdot)$ (see (12)) and $\mathcal{D}^2b^R(\cdot)$ on their compact support $\{x\in\mathcal{C}^d:\|x\|\le R+1\}$. Similar to the proof of [24, Lemma 3.3], we obtain $\sup_{r\le t\le T}\mathbb{E}[\|D_r y^{\xi,\Delta,R}_{\lfloor t\rfloor}-D_r z^{\xi,\Delta,R}_t\|^2]\le K_{T,R}\Delta$. Then

$$\begin{aligned}&\mathbb{E}\Big[\Big|\mathcal{D}b^R(x_s^{\xi,R})D_r x_s^{\xi,R}-\mathcal{D}b^R(y_{\lfloor s\rfloor}^{\xi,\Delta,R})D_r y_{\lfloor s\rfloor}^{\xi,\Delta,R}\Big|^2\Big]\\&\leq K_R\mathbb{E}\big[\|x_s^{\xi,R}-y_{\lfloor s\rfloor}^{\xi,\Delta,R}\|^2\|D_r x_s^{\xi,R}\|^2\big]+K_R\mathbb{E}\Big[|D_r x^{\xi,R}(s)-D_r z^{\xi,\Delta,R}(s)|^2\\&\quad+\int_{-\tau}^{0}|D_r x_s^{\xi,R}(r)-D_r z_s^{\xi,\Delta,R}(r)|^2\mathrm{d}\nu_1(r)\Big]+K_{T,R}\Delta.\end{aligned}\tag{17}$$

Similarly, by using (13), we deduce

$$\begin{aligned}&\mathbb{E}\Big[\Big|\mathcal{D}\sigma^R(x_s^{\xi,R})D_r x_s^{\xi,R}-\mathcal{D}\sigma^R(y_{\lfloor s\rfloor}^{\xi,\Delta,R})D_r y_{\lfloor s\rfloor}^{\xi,\Delta,R}\Big|^2\Big]\\&\leq K_R\mathbb{E}\big[\|x_s^{\xi,R}-y_{\lfloor s\rfloor}^{\xi,\Delta,R}\|^2\|D_r x_s^{\xi,R}\|^2\big]+K_R\mathbb{E}\Big[|D_r x^{\xi,R}(s)-D_r z^{\xi,\Delta,R}(s)|^2\\&\quad+\int_{-\tau}^{0}|D_r x_s^{\xi,R}(r)-D_r z_s^{\xi,\Delta,R}(r)|^2\mathrm{d}\nu_2(r)\Big]+K_{T,R}\Delta.\end{aligned}\tag{18}$$

Inserting (17) and (18) into (16), and then using Assumption 2.2, Lemmas 2.1 and 2.2, and (15), we obtain that for $\Delta\in(0,\Delta_1]$ with some $\Delta_1\in(0,1)$,

$$\begin{aligned}&\mathbb{E}\Big[|D_r x^{\xi,R}(t)-D_r z^{\xi,\Delta,R}(t)|^2\Big]\\&\leq K_{T,R}\Delta+K_{T,R}\int_r^T\big(\mathbb{E}[\|x_{\lfloor s\rfloor}^{\xi,R}-y_{\lfloor s\rfloor}^{\xi,\Delta,R}\|^4]\big)^{\frac12}\mathrm{d}s+K_{T,R}\int_r^T\big(\mathbb{E}[\|x_{\lfloor s\rfloor}^{\xi,R}-x_s^{\xi,R}\|^4]\big)^{\frac12}\mathrm{d}s\\&\quad+K_{T,R}\int_r^T\mathbb{E}\big[|D_r x^{\xi,R}(s)-D_r z^{\xi,\Delta,R}(s)|^2\big]\mathrm{d}s+K_{T,R}\sup_{0\leq s\leq T}\mathbb{E}[\|x_{\lfloor s\rfloor}^{\xi,R}-x_s^{\xi,R}\|^2]\\&\leq K_{T,R}\Delta+K_{T,R}\int_r^T\mathbb{E}\big[|D_r x^{\xi,R}(s)-D_r z^{\xi,\Delta,R}(s)|^2\big]\mathrm{d}s.\end{aligned}$$

Applying the Grönwall inequality, we arrive at $\mathbb{E}\Big[|D_r x^{\xi,R}(t)-D_r z^{\xi,\Delta,R}(t)|^2\Big]\leq K_{T,R}\Delta$, which implies for $t_k\in[0,T]$,

$$\int_0^T\mathbb{E}\Big[|D_r x^{\xi,R}(t_k)-D_r y^{\xi,\Delta,R}(t_k)|^2\Big]\mathrm{d}r\leq K_{T,R}\Delta.\tag{19}$$

It follows from (15) and (19) that

$$\sup_{0\leq t_k\leq T}\|y^{\xi,\Delta,R}(t_k)-x^{\xi,R}(t_k)\|_{1,2}\leq K_{T,R}\Delta^{\frac12}.$$

Step 2. In this step, we aim to apply [16, Lemma A.1] to estimate the error in $L^1(\mathbb{R}^d)$ between densities of $x^{\xi,R}(\cdot)$ and $y^{\xi,\Delta,R}(\cdot)$. To this end, we first claim that for $\varrho \in (0,1)$,

$$\mathbb{E}\Big[\Big(\int_0^T |D_r x^{\xi,R}(t)|^2 \mathrm{d}r\Big)^{-\varrho}\Big] \le K_{T,R,\varrho}. \tag{20}$$

In fact, by means of Proposition 2.1 with $u = \frac{1}{d}(1,\dots,1) \in \mathbb{R}^d$, one has that $\mathbb{P}\big(\int_0^T |D_r x^{\xi,R}(t)|^2 \mathrm{d}r \le \varepsilon\big) \le K_{T,R}\varepsilon$ for any $\varepsilon \in (0,1)$. Hence for any $\varrho \in (0,1)$, we derive that

$$\begin{aligned}&\sum_{n=1}^{\infty} n^{\varrho-1}\mathbb{P}\Big(\big(\int_0^T |D_r x^{\xi,R}(t)|^2 \mathrm{d}r\big)^{-1} \ge n\Big)\\ &\le 1 + K_{T,R}\sum_{n=2}^{\infty} n^{\varrho-1}n^{-1} \le K_{T,R,\varrho}.\end{aligned} \tag{21}$$

Then for any $\varrho \in (0,1)$,

$$\begin{aligned}&\mathbb{E}\Big[\big(\int_0^T |D_r x^{\xi,R}(t)|^2 \mathrm{d}r\big)^{-\varrho}\Big]\\ &\le 1 + \sum_{n=1}^{\infty}(n+1)^{\varrho}\mathbb{P}\Big(n \le \big(\int_0^T |D_r x^{\xi,R}(t)|^2 \mathrm{d}r\big)^{-1} \le n+1\Big)\\ &\le 2 + \sum_{n=1}^{\infty}((n+1)^{\varrho} - n^{\varrho})\mathbb{P}\Big(\big(\int_0^T |D_r x^{\xi,R}(t)|^2 \mathrm{d}r\big)^{-1} \ge n\Big)\\ &\le 2 + \varrho\sum_{n=1}^{\infty} n^{\varrho-1}\mathbb{P}\Big(\big(\int_0^T |D_r x^{\xi,R}(t)|^2 \mathrm{d}r\big)^{-1} \ge n\Big) \le K_{T,R,\varrho},\end{aligned}$$

where in the last step we used (21). This proves the claim (20).

Furthermore, similar to the proof of Lemma 2.1, it follows from Assumptions 2.1, 2.2 and 3.1 that $x^{\xi}(t) \in \mathbb{D}^{2,4}$ for all $t \in [0,T]$. Then, by [16, Eq. (6.1) and Lemma A.1], we arrive at

$$\begin{aligned}&\sup_{0<t_k\le T}\int_{\mathbb{R}^d} |\mathfrak{p}^{\Delta,R}(t_k,x) - \mathfrak{p}^R(t_k,x)|\mathrm{d}x = \sup_{0<t_k\le T} \mathrm{d}_{TV}(y^{\xi,\Delta,R}(t_k), x^{\xi,R}(t_k))\\ &\le K_{T,R}\sup_{0\le t_k\le T}\|y^{\xi,\Delta,R}(t_k) - x^{\xi,R}(t_k)\|_{1,2}^{\frac{2\varrho}{2\varrho+2}} \le K_{T,R}\Delta^{\frac{\varrho}{2\varrho+2}},\end{aligned} \tag{22}$$

where $\mathrm{d}_{TV}(X,Y)$ denotes the total variation distance between two $\mathbb{R}^d$-valued random variables X and Y, and $\mathfrak{p}^R(t_k,\cdot)$, $\mathfrak{p}^{\Delta,R}(t_k,\cdot)$ are densities of $x^{\xi,R}(t_k)$ and $y^{\xi,\Delta,R}(t_k)$, respectively.

Step 3. Denoting two sequences of events by

$$\Omega_R := \{\omega\in\Omega : \sup_{t\in[0,T]} |x^\xi(t)| \le R\},$$

$$\Omega_{R,y} := \{\omega\in\Omega : \sup_{t_k\in[0,T]} |y^{\xi,\Delta}(t_k)| \le R\},$$

we have $\lim_{R\to\infty}\mathbb{P}(\Omega_R)=\mathbb{P}(\Omega)=\lim_{R\to\infty}\mathbb{P}(\Omega_{R,y})=1$. According to [16, Eq. (6.1)], we derive

$$\begin{aligned}
&\sup_{0<t_k\le T}\int_{\mathbb{R}^d} |\mathfrak{p}^\Delta(t_k,x)-\mathfrak{p}(t_k,x)|\mathrm{d}x = \sup_{0<t_k\le T} \mathrm{d}_{TV}(y^{\Delta,\xi}(t_k),x^\xi(t_k))\\
&\le 4\mathbb{P}(\Omega^c_{R-1}) + 2\mathbb{P}(\sup_{0\le t_k\le T}|x^\xi(t_k)-y^{\xi,\Delta}(t_k)|\ge 1) + \sup_{0\le t_k\le T}\mathrm{d}_{TV}(y^{\xi,\Delta,R}(t_k),x^{\xi,R}(t_k))\\
&\le 4\frac{\mathbb{E}[\sup_{0\le t\le T}|x^\xi(t)|^2]}{(R-1)^2} + 2\mathbb{E}[\sup_{0\le t_k\le T}|x^\xi(t_k)-y^{\xi,\Delta}(t_k)|^2] + K_{T,R}\Delta^{\frac{\varrho}{2\varrho+2}}\\
&\le \frac{K_T}{(R-1)^2} + K_T\Delta + K_{T,R}\Delta^{\frac{\varrho}{2\varrho+2}},
\end{aligned}$$

where we used Lemma 2.1, (14), and (22). Letting $\Delta\to 0$, $R\to\infty$, we obtain the desired argument. □

Remark 3.1. It follows from the proof of Theorem 3.1 that when the coefficients b and σ are globally Lipschitz continuous, the convergence rate of the density of discretizations is almost $1/4$ in $L^1(\mathbb{R}^d)$. We will show that for the additive noise case, the convergence rate could attain 1 in the pointwise sense (see Theorem 3.2).

It is known that the estimate between densities of random variables is closely related to the total variation distance of random variables. Let

$$\mathrm{d}_{TV}(X,Y) := \sup_{\Phi\in\mathcal{B}_b(\mathcal{H}),\|\Phi\|_\infty\le 1} |\mathbb{E}[\Phi(X)]-\mathbb{E}[\Phi(Y)]|$$

denote the total variation distance of two $\mathcal{H}$-valued random variables X, Y, where $\mathcal{H}=\mathbb{R}^d$ or $\mathcal{H}=\mathcal{C}^d$, $\mathcal{B}_b(\mathcal{H})$ is the set of bounded and measurable mappings from $\mathcal{H}$ to $\mathbb{R}$, and $\|\Phi\|_\infty := \sup_{x\in\mathcal{H}}|\Phi(x)|$. As a result of Theorem 3.1, we obtain the convergence in the total variation distance

$$\lim_{\Delta\to 0}\mathrm{d}_{TV}(x^\xi(t_n), y^{\xi,\Delta}(t_n)) = 0. \tag{23}$$

While for the θ-EM functional solution, as a $\mathcal{C}^d$-valued random variable, the law does not converge in total variation distance, namely,

$$\limsup_{\Delta\to 0} \mathrm{d}_{TV}(x^{\xi}_{t_n}, y^{\xi,\Delta}_{t_n}) > 0. \tag{24}$$

In order to illustrate (24), we consider the test function $\mathbf{1}_{\{\|x\|_{\mathcal{C}_{1/2}([-\frac{\Delta}{2},0];\mathbb{R}^d)}<\infty\}}$, where $\mathcal{C}_{1/2}([-\frac{\Delta}{2},0];\mathbb{R}^d)$ is the space of all continuous functions $f:[-\frac{\Delta}{2},0]\to\mathbb{R}^d$ that is $\frac{1}{2}$-Hölder continuous, equipped with the norm $\|x\|_{\mathcal{C}_{1/2}([-\frac{\Delta}{2},0];\mathbb{R}^d)} := \sup_{t,s\in[-\frac{\Delta}{2},0],t\neq s}\frac{|x(t)-x(s)|}{|t-s|^{\frac{1}{2}}}$. On the one hand, the θ-EM functional solution satisfies

$$\begin{aligned}\|y^{\xi,\Delta}_{t_n}\|_{\mathcal{C}_{1/2}([-\frac{\Delta}{2},0];\mathbb{R}^d)} &\le \sup_{t,s\in[-\frac{\Delta}{2},0],t\neq s}\frac{|t-s|^{\frac{1}{2}}}{\Delta}(|y^{\xi,\Delta}(t_{n-1})|+|y^{\xi,\Delta}(t_n)|)\\ &\le C(\omega)\Delta^{-\frac{1}{2}}<\infty,\ \ \omega\in\Omega,\end{aligned}$$

due to the definition (6). On the other hand, by the Kolmogorov continuous theorem, there is a modification of the exact solution such that the path is almost surely $(\frac{1}{2}-\epsilon)$-Hölder continuous with any small constant $\epsilon\in(0,\frac{1}{2})$. This leads to

$$\|x^{\xi}_{t_n}\|_{\mathcal{C}_{1/2}([-\frac{\Delta}{2},0];\mathbb{R}^d)} = \sup_{t,s\in[-\frac{\Delta}{2},0],t\neq s}\frac{|x(t_n+t)-x(t_n+s)|}{|t-s|^{\frac{1}{2}}} = \infty,\ a.s.$$

Hence, recalling the definition of the total variation distance, we arrive at

$$\mathrm{d}_{TV}(x^{\xi}_{t_n}, y^{\xi,\Delta}_{t_n}) \ge |\mathbb{P}(\|x^{\xi}_{t_n}\|_{\mathcal{C}_{1/2}([-\frac{\Delta}{2},0];\mathbb{R}^d)}<\infty) - \mathbb{P}(\|y^{\xi,\Delta}_{t_n}\|_{\mathcal{C}_{1/2}([-\frac{\Delta}{2},0];\mathbb{R}^d)}<\infty)| = 1,$$

which shows (24) by taking the upper limit.

3.2. Convergence rate for linearly growing drift and additive noise case

In this subsection, we consider the additive noise case and study the convergence rate for the density of the θ-EM discretization, based on the test-functional-independent weak convergence analysis of the discretization. Assumptions on the coefficients considered in this subsection are given below.

Assumption 3.3. *The noise of* (1) *is of additive type, i.e.,* $d=m$*, and there exists some constant* $\tilde{\sigma}>0$ *such that for any* $\phi\in\mathcal{C}^d$, $\sigma(\phi)\equiv\tilde{\sigma}\mathrm{Id}_{d\times d}$.

Since $\{-W(t)\}_{t\ge 0}$ is also a Brownian motion whenever $\{W(t)\}_{t\ge 0}$ is a Brownian motion, we only consider the case $\tilde{\sigma}>0$. The next assumption imposes higher-order regularity on the coefficients, which is needed to obtain estimates for the higher-order Malliavin derivatives of the solutions. Recall that the definition of $\mathcal{L}((\mathcal{C}^d)^{\otimes k};\mathbb{R}^d)$ is given in Section 2.

Assumption 3.4. *The coefficient* b *has continuous and bounded derivatives up to order* 4 *satisfying* $\sup_{\phi\in\mathcal{C}^d}\|\mathcal{D}^k b(\phi)\|_{\mathcal{L}((\mathcal{C}^d)^{\otimes k};\mathbb{R}^d)}\le K$ *for* $k\in\mathbb{N}_+$ *with* $k\le 4$. *In addition, there exist a constant* $n_b\in\mathbb{N}_+$ *and probability measures* ν_3^i *on* $[-\tau,0]$ *for* $i\in\{1,\dots,n_b\}$*, such that the coefficient* b *satisfies that for any* $\phi_1,\phi_2\in\mathcal{C}^d$,

$$\mathcal{D}b(\phi_1)\phi_2 = \sum_{i=1}^{n_b} \int_{-\tau}^{0} k_b^i(\phi_1)\phi_2(s)\mathrm{d}\nu_3^i(s), \tag{25}$$

where $k_b^i : \mathcal{C}^d \to \mathbb{R}^{d\times d}$ *satisfies*

$$\sup_{i\in\{1,\dots,n_b\}} \sup_{l\in\{1,2,3\}} \sup_{\phi\in\mathcal{C}^d} \|\mathcal{D}^l k_b^i(\phi)\|_{\mathcal{L}((\mathcal{C}^d)^{\otimes l};\mathbb{R}^{d\times d})} + \sup_{i\in\{1,\dots,n_b\}} \sup_{\phi\in\mathcal{C}^d} |k_b^i(\phi)| \le K.$$

We give an example of coefficient b that satisfies (25). Let b have the form: $b(\phi) = \tilde{b}(\int_{-\tau}^{0} \phi(r)\mathrm{d}\nu_3^1(r))$ for $\phi \in \mathcal{C}^d$ with some function $\tilde{b} : \mathbb{R}^d \to \mathbb{R}^d$. Then for $\phi_1, \phi_2 \in \mathcal{C}^d$, we have

$$\mathcal{D}b(\phi_1)\phi_2 = \int_{-\tau}^{0} \mathcal{D}\tilde{b}\Big(\int_{-\tau}^{0} \phi_1(r)\mathrm{d}\nu_3^1(r)\Big)\phi_2(s)\mathrm{d}\nu_3^1(s).$$

The convergence rate of the density for the θ-EM discretization is stated as follows.

Theorem 3.2. *Let Assumption 2.1, Assumption 3.2 with* $\rho = 1$*, Assumptions 3.3 and 3.4 hold. Then there exists* $\tilde{\Delta} > 0$ *such that for* $\Delta \in (0, \tilde{\Delta}]$ *and* $T \ge T_0$*,*

$$\sup_{z\in\mathbb{R}^d} |\mathfrak{p}(T,z) - \mathfrak{p}^{\Delta}(T,z)| \le K_T \Delta,$$

where $T_0 := \frac{\ln(\frac{3}{2})}{2L_b n_b(\theta+2)}$ *with* $L_b := \sup_{i\in\{1,\dots,n_b\}} \sup_{\phi\in\mathcal{C}^d} |k_b^i(\phi)| < \infty$.

Remark 3.2. Under conditions in Theorem 3.2, we can also obtain (23). In fact, it follows from the Scheffé lemma and [15, Section 3.1] that for $T \ge T_0$,

$$\int_{\mathbb{R}^d} |\mathfrak{p}(T,z) - \mathfrak{p}^{\Delta}(T,z)|\mathrm{d}z \to 0 \text{ as } \Delta \to 0.$$

The proof of Theorem 3.2 is based on a weak convergence analysis of the θ-EM discretization. We would like to mention that there have been some works devoted to the weak convergence of the EM discretization for SFDEs. For instance, weak error estimates have been obtained with the upper bound depending on a given test functional; see e.g. [25,26]. When analyzing the convergence rate of the density of discretizations, an effective approach is to apply the Malliavin integration by parts formula to derive a test-functional-independent weak convergence analysis; see [15] for the relevant study of the stochastic heat equation. For SFDEs, the high degeneracy of coefficients makes the derivation of the Malliavin integration by parts formula challenging. In this subsection, we will fully utilize the dimensionality reduction argument presented in Section 2 to establish the negative moment estimates of the determinant for the Malliavin covariance matrix, which allows us to derive the integration by parts formula; see Lemmas 3.3 and 3.4 for details. Then combining *a priori* estimates of both the exact solution and the discretization, we obtain the test-functional-independent weak convergence analysis for the θ-EM discretizations.

We begin with presenting the relation between the error of densities and the weak error of discretizations. According to [10,15], we have the following approximation of densities for both the exact solution and discretizations: for fixed $T > 0$ and $z \in \mathbb{R}^d$,

$$\mathfrak{p}(T, z) = \lim_{n\to\infty} \int_{\mathbb{R}^d} g_{n^{-1}}(y - z)\mathfrak{p}(T, y)\mathrm{d}y = \lim_{n\to\infty} \mathbb{E}[g_{n^{-1}}(x^{\xi}(T) - z)],$$

$$\mathfrak{p}^{\Delta}(T, z) = \lim_{n\to\infty} \int_{\mathbb{R}^d} g_{n^{-1}}(y - z)\mathfrak{p}^{\Delta}(T, y)\mathrm{d}y = \lim_{n\to\infty} \mathbb{E}[g_{n^{-1}}(y^{\xi,\Delta}(T) - z)],$$

where g_ζ denotes the Gaussian density with mean 0 and covariance matrix $\zeta \mathrm{Id}_{d\times d}$. This gives that

$$|\mathfrak{p}(T, z) - \mathfrak{p}^{\Delta}(T, z)| = \lim_{n\to\infty} |\mathbb{E}[g_{n^{-1}}(x^{\xi}(T) - z)] - \mathbb{E}[g_{n^{-1}}(y^{\xi,\Delta}(T) - z)]|. \tag{26}$$

Noting that for any $n \geq 1$, the function $g_{n^{-1}}(\cdot - z)$, $z \in \mathbb{R}^d$ belongs to

$$\mathscr{C} := \Big\{ f : \mathbb{R}^d \to \mathbb{R} \,\Big|\, f \in C^{\infty}_{pol}(\mathbb{R}^d; \mathbb{R}), \exists F : \mathbb{R}^d \to \mathbb{R} \text{ with } 0 \leq F \leq 1 \text{ such that } F(x_1, \ldots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(y_1, \ldots, y_d)\mathrm{d}y_d \cdots \mathrm{d}y_1 \Big\}, \tag{27}$$

we have

$$|\mathfrak{p}(T, z) - \mathfrak{p}^{\Delta}(T, z)| \leq \sup_{f\in\mathscr{C}} |\mathbb{E}[f(x^{\xi}(T))] - \mathbb{E}[f(y^{\xi,\Delta}(T))]|. \tag{28}$$

Hence, to obtain the convergence rate of the density, it suffices to estimate the error $|\mathbb{E}[f(x^{\xi}(T))] - \mathbb{E}[f(y^{\xi,\Delta}(T))]|$.

To this end, we give some frequently used notation hereafter. Let $\varphi(t; t_i, \eta)$ and $\varphi_t(t_i, \eta)$ denote the solution and functional solution of (1) at time t with initial value $\eta \in \mathcal{C}^d$ at t_i, $\Phi(t_k; t_i, \eta)$ and $\Phi_{t_k}(t_i, \eta)$ denote the θ-EM solution and θ-EM functional solution at time t_k with initial value η at t_i. For $Y \in \mathcal{C}([-\tau, T]; \mathbb{R}^d)$, we can define $Y^{Int}(\cdot)$ as the linear interpolation with respect to $\{(t_k, Y(t_k))\}_{k=-N}^{N^{\Delta}}$, whose segment process $Y^{Int}_r(\cdot) \in \mathcal{C}^d$ for $r \in [0, T]$ is defined by $Y^{Int}_r(s) := Y^{Int}(r + s)$, $s \in [-\tau, 0]$. In addition, for $Y \in \mathcal{C}([-\tau, T]; \mathbb{R}^d)$, we know that Y_r is $\mathcal{C}^d$-valued for $r \in [0, T]$. In particular, when Y is a $\mathcal{C}([-\tau, T]; \mathbb{R}^d)$-valued random variable, Y_r is a $\mathcal{C}^d$-valued random variable for $r \in [0, T]$. We always use the relation $\varphi^{Int}(t_l; t_j, \varphi_{t_j}(0, \xi^{Int})) = \varphi^{Int}(t_l; 0, \xi^{Int}) = \varphi(t_l; 0, \xi^{Int})$ for $l \geq j \geq 0$ in the decomposition of the weak error.

Let $f \in \mathscr{C}$. By relation $\varphi(T; 0, \xi) = \varphi^{Int}_{t_{N^{\Delta}}}(0, \xi)(0) = \Phi(T; t_{N^{\Delta}}, \varphi^{Int}_{t_{N^{\Delta}}}(0, \xi))$, we have

$$\begin{aligned}
&\mathbb{E}[f(x^{\xi}(T))] - \mathbb{E}[f(y^{\xi,\Delta}(T))] = \mathbb{E}[f(\varphi(T; 0, \xi))] - \mathbb{E}[f(\Phi(T; 0, \xi^{Int}))] \\
&= \mathbb{E}[f(\Phi(T; t_{N^{\Delta}}, \varphi^{Int}_{t_{N^{\Delta}}}(0, \xi)))] - \mathbb{E}[f(\Phi(T; t_{N^{\Delta}}, \varphi^{Int}_{t_{N^{\Delta}}}(0, \xi^{Int})))] \\
&\quad + \mathbb{E}[f(\Phi(T; t_{N^{\Delta}}, \varphi^{Int}_{t_{N^{\Delta}}}(0, \xi^{Int})))] - \mathbb{E}[f(\Phi(T; 0, \xi^{Int}))].
\end{aligned} \tag{29}$$

We split $\mathbb{E}[f(\Phi(T;t_{N^\Delta},\varphi^{Int}_{t_{N^\Delta}}(0,\xi^{Int})))]-\mathbb{E}[f(\Phi(T;0,\xi^{Int}))]$ further as

$$\begin{aligned}
&\mathbb{E}[f(\Phi(T;t_{N^\Delta},\varphi^{Int}_{t_{N^\Delta}}(0,\xi^{Int})))]-\mathbb{E}[f(\Phi(T;0,\xi^{Int}))]\\
&=\sum_{i=1}^{N^\Delta}\Big\{\mathbb{E}[f(\Phi(T;t_i,\varphi^{Int}_{t_i}(0,\xi^{Int})))]-\mathbb{E}[f(\Phi(T;t_{i-1},\varphi^{Int}_{t_{i-1}}(0,\xi^{Int})))]\Big\}\\
&=\sum_{i=1}^{N^\Delta}\mathbb{E}\Big[\mathbb{E}\Big[f(\Phi(T;t_i,\varphi^{Int}_{t_i}(t_{i-1},\varphi_{t_{i-1}}(0,\xi^{Int}))))\\
&\qquad -f(\Phi(T;t_i,\Phi_{t_i}(t_{i-1},\varphi^{Int}_{t_{i-1}}(0,\xi^{Int}))))\Big|\mathcal{F}_{t_i}\Big]\Big]\\
&=\sum_{i=1}^{N^\Delta}\int_0^1\mathbb{E}\Big[f'(\Phi(T;t_i,Y_i^\varsigma))\mathcal{D}\Phi(T;t_i,Y_i^\varsigma)\\
&\qquad\Big(\varphi^{Int}_{t_i}(t_{i-1},\varphi_{t_{i-1}}(0,\xi^{Int}))-\Phi_{t_i}(t_{i-1},\varphi^{Int}_{t_{i-1}}(0,\xi^{Int}))\Big)\Big]\mathrm{d}\varsigma,
\end{aligned}\tag{30}$$

where we used $\varphi^{Int}_{t_i}(t_{i-1},\varphi_{t_{i-1}}(0,\xi^{Int}))=\varphi^{Int}_{t_i}(0,\xi^{Int})$, and

$$Y_i^\varsigma:=\varsigma\varphi^{Int}_{t_i}(t_{i-1},\varphi_{t_{i-1}}(0,\xi^{Int}))+(1-\varsigma)\Phi_{t_i}(t_{i-1},\varphi^{Int}_{t_{i-1}}(0,\xi^{Int})).\tag{31}$$

It follows from $\varphi^{Int}(t_j;0,\xi^{Int})=\varphi(t_j;0,\xi^{Int})$ that

$$\begin{aligned}
&\varphi^{Int}_{t_i}(t_{i-1},\varphi_{t_{i-1}}(0,\xi^{Int}))-\Phi_{t_i}(t_{i-1},\varphi^{Int}_{t_{i-1}}(0,\xi^{Int}))\\
&=\sum_{j=-N}^{0}I^{[j]}\Big(\varphi^{Int}(t_i+t_j;t_{i-1},\varphi_{t_{i-1}}(0,\xi^{Int}))-\Phi(t_i+t_j;t_{i-1},\varphi^{Int}_{t_{i-1}}(0,\xi^{Int}))\Big)\\
&=I^{[0]}\Big(\varphi(t_i;t_{i-1},\varphi_{t_{i-1}}(0,\xi^{Int}))-\Phi(t_i;t_{i-1},\varphi^{Int}_{t_{i-1}}(0,\xi^{Int}))\Big),
\end{aligned}\tag{32}$$

where $I^{[j]}(\cdot)$, $j\in\{-N,\dots,0\}$ are basis functions given in (7). Note that for any $i=1,\dots,N^\Delta$,

$$\begin{aligned}
&\varphi(t_i;t_{i-1},\varphi_{t_{i-1}}(0,\xi^{Int}))-\Phi(t_i;t_{i-1},\varphi^{Int}_{t_{i-1}}(0,\xi^{Int}))\\
&=\int_{t_{i-1}}^{t_i}b(\varphi_r(t_{i-1},\varphi_{t_{i-1}}(0,\xi^{Int})))\mathrm{d}r-\int_{t_{i-1}}^{t_i}\Big[(1-\theta)b(\varphi^{Int}_{t_{i-1}}(0,\xi^{Int}))\\
&\qquad+\theta b(\Phi_{t_i}(t_{i-1},\varphi^{Int}_{t_{i-1}}(0,\xi^{Int})))\Big]\mathrm{d}r=\mathcal{I}_b^i+\mathcal{I}_{b,\theta}^i,
\end{aligned}\tag{33}$$

where

$$\mathcal{I}_b^i := \int_{t_{i-1}}^{t_i} \big[b(\varphi_r(t_{i-1}, \varphi_{t_{i-1}}(0, \xi^{Int}))) - b(\varphi_{t_{i-1}}^{Int}(0, \xi^{Int}))\big]\mathrm{d}r,$$

$$\mathcal{I}_{b,\theta}^i := \theta \int_{t_{i-1}}^{t_i} \big[b(\varphi_{t_{i-1}}^{Int}(0, \xi^{Int})) - b(\Phi_{t_i}(t_{i-1}, \varphi_{t_{i-1}}^{Int}(0, \xi^{Int})))\big]\mathrm{d}r.$$

According to (29)–(33), we derive the decomposition of the weak error

$$\begin{aligned}
&\mathbb{E}[f(x^{\xi}(T))] - \mathbb{E}[f(y^{\xi,\Delta}(T))] \\
&= \mathbb{E}[f(\Phi(T; t_{N^\Delta}, \varphi_{t_{N^\Delta}}^{Int}(0, \xi)))] - \mathbb{E}[f(\Phi(T; t_{N^\Delta}, \varphi_{t_{N^\Delta}}^{Int}(0, \xi^{Int})))] \\
&\quad + \sum_{i=1}^{N^\Delta} \int_0^1 \mathbb{E}\Big[\langle f'(\Phi(T; t_i, Y_i^{\varsigma}))\mathcal{D}\Phi(T; t_i, Y_i^{\varsigma}), I^{[0]}(\mathcal{I}_b^i + \mathcal{I}_{b,\theta}^i)\rangle\Big]\mathrm{d}\varsigma \\
&= \mathcal{I}_0 + \mathcal{I}_b + \mathcal{I}_{b,\theta}, \qquad (34)
\end{aligned}$$

where

$$\mathcal{I}_0 := \mathbb{E}[f(\Phi(T; t_{N^\Delta}, \varphi_{t_{N^\Delta}}^{Int}(0, \xi)))] - \mathbb{E}[f(\Phi(T; t_{N^\Delta}, \varphi_{t_{N^\Delta}}^{Int}(0, \xi^{Int})))],$$

$$\mathcal{I}_b := \sum_{i=1}^{N^\Delta} \int_0^1 \mathbb{E}\Big[\Big\langle f'(\Phi(T; t_i, Y_i^{\varsigma}))\mathcal{D}\Phi(T; t_i, Y_i^{\varsigma}), I^{[0]}\mathcal{I}_b^i\Big\rangle\Big]\mathrm{d}\varsigma,$$

$$\mathcal{I}_{b,\theta} := \sum_{i=1}^{N^\Delta} \int_0^1 \mathbb{E}\Big[\Big\langle f'(\Phi(T; t_i, Y_i^{\varsigma}))\mathcal{D}\Phi(T; t_i, Y_i^{\varsigma}), I^{[0]}\mathcal{I}_{b,\theta}^i\Big\rangle\Big]\mathrm{d}\varsigma.$$

Based on this decomposition of the weak error, we will prove that the test-functional-independent weak convergence rate is 1 (see Theorem 3.3). Then the proof of Theorem 3.2 follows from (28). To proceed, we make some preparations in the following subsection.

3.2.1. A priori estimates

This subsection gives some moment estimates for derivatives of the exact solution and discretizations. In addition, we also present the negative moment estimates of the determinant of the Malliavin covariance matrix for discretizations. We begin by showing that the moments of high-order derivatives of the exact solution and discretizations are bounded, which are stated in Lemmas 3.1 and 3.2. The proofs are omitted since they are similar to those of Lemmas 2.1 and 2.2; see also [15] for related results.

Lemma 3.1. *Let $\alpha \in \{1, 2\}$. Under Assumptions 2.1, 3.3 and 3.4, we have that for any fixed $p \geq 2$, there exist $\tilde{p} \geq p$ such that for any $\eta \in \mathcal{C}^d$,*

$$\sup_{t\in(0,T]} \mathbb{E}[\|\mathcal{D}x_t^{\tilde{\xi}} \cdot \eta\|^{2p}] \leq K_T \|\eta\|^{2p}(1 + \mathbb{E}[\|\tilde{\xi}\|^{2\tilde{p}}]),$$

$$\sup_{t\in(0,T]} \mathbb{E}[\|D_{(r_1,\dots,r_\alpha)}\mathcal{D}x_t^{\tilde{\xi}}\cdot\eta\|^{2p}]$$

$$\le K_T\|\eta\|^{2p}\Big(1+\sum_{k=0}^{\alpha}\sum_{1\le i_1<\cdots<i_k\le\alpha}\mathbb{E}[\|D_{(r_{i_1},\dots,r_{i_k})}\tilde{\xi}\|^{2\tilde{p}}]\Big),\quad r_1,\dots,r_\alpha\in[0,T],$$

$$\sup_{t\in(r_1\vee r_2,T]} \mathbb{E}[\|D_{(r_1,r_2)}x_t^{\tilde{\xi}}\|^{2p}]$$

$$\le K_T\Big(1+\sum_{k=0}^{2}\sum_{1\le i_1<\cdots<i_k\le 2}\mathbb{E}[\|D_{(r_{i_1},\dots,r_{i_k})}\tilde{\xi}\|^{2\tilde{p}}]\Big),\quad r_1,r_2\in[0,T],$$

where $\tilde{\xi}(r)\in\mathbb{D}^{2,2\tilde{p}}$ *for* $r\in[-\tau,0]$ *and we adopt the convention that* $D_{(r_{i_1},\dots,r_{i_k})}\tilde{\xi}:=\tilde{\xi}$ *for* $k=0$.

Lemma 3.2. *Let* $\alpha\in\{1,2,3\}$ *and* $\tilde{\alpha}\in\{2,3,4\}$. *Under Assumptions 2.1, 3.3 and 3.4, for* $p\ge 2$, *there exist* $\tilde{p}\ge p$ *and* $\tilde{\Delta}:=\tilde{\Delta}(p)>0$ *such that for any* $\Delta\in(0,\tilde{\Delta}]$ *and* $\eta\in\mathcal{C}^d$,

$$\sup_{t_k\in(0,T]} \mathbb{E}[\|\mathcal{D}y_{t_k}^{\tilde{\xi},\Delta}\cdot\eta\|^{2p}]\le K_T\|\eta\|^{2p}(1+\mathbb{E}[\|\tilde{\xi}\|^{2\tilde{p}}]),$$

$$\sup_{t_k\in(0,T]} \mathbb{E}[\|D_{(r_1,\dots,r_\alpha)}\mathcal{D}y_{t_k}^{\tilde{\xi},\Delta}\cdot\eta\|^{2p}]$$

$$\le K_T\|\eta\|^{2p}\Big(1+\sum_{k=0}^{\alpha}\sum_{1\le i_1<\cdots<i_k\le\alpha}\mathbb{E}[\|D_{(r_{i_1},\dots,r_{i_k})}\tilde{\xi}\|^{2\tilde{p}}]\Big),\quad r_1,\dots,r_\alpha\in[0,T],$$

$$\sup_{t_k\in(r_1\vee\cdots\vee r_{\tilde{\alpha}},T]} \mathbb{E}[\|D_{(r_1,\dots,r_{\tilde{\alpha}})}y_{t_k}^{\tilde{\xi},\Delta}\|^{2p}]$$

$$\le K_T\Big(1+\sum_{k=0}^{\tilde{\alpha}}\sum_{1\le i_1<\cdots<i_k\le\tilde{\alpha}}\mathbb{E}[\|D_{(r_{i_1},\dots,r_{i_k})}\tilde{\xi}\|^{2\tilde{p}}]\Big),\quad r_1,\dots,r_{\tilde{\alpha}}\in[0,T],$$

where $\tilde{\xi}(r)\in\mathbb{D}^{4,2\tilde{p}}$ *for* $r\in[-\tau,0]$ *and we adopt the convention that* $D_{(r_{i_1},\dots,r_{i_k})}\tilde{\xi}:=\tilde{\xi}$ *for* $k=0$.

The following lemma shows the negative moment estimates of the determinant of the Malliavin covariance matrix for the discretizations.

Lemma 3.3. *Let Assumptions 3.3 and 3.4 hold. Then for any* $u\in\mathbb{R}^d$ *with* $|u|=1$ *and* $\Delta\in(0,\frac{1}{2L_bn_b\theta}]$,

$$|u^\top\gamma_{\Phi(T;t_j,Y_j^\varsigma)}u|\ge\frac{1}{4}(T\wedge T_0)\tilde{\sigma}^2,\quad j=1,\dots,N^\Delta, \tag{35}$$

where L_b *and* T_0 *are given in Theorem 3.2, and* Y_j^ς *is given in* (31). *In particular, we have* $\det(\gamma_{\Phi(T;t_j,Y_j^\varsigma)})^{-1}\in L^{\infty-}(\Omega)$.

Proof. We first prove (35) for the case of $d = 1$. For $r \in (t_k, t_{k+1}]$ with $j \le k \le i-1$, we have

$$D_r\Phi(t_i; t_j, Y_j^\varsigma) = \sum_{n=k}^{i-1}\Big[(1-\theta)\mathcal{D}b(\Phi_{t_n}(t_j, Y_j^\varsigma))D_r\Phi_{t_n}(t_j, Y_j^\varsigma)\Delta \\ + \theta\mathcal{D}b(\Phi_{t_{n+1}}(t_j, Y_j^\varsigma))D_r\Phi_{t_{n+1}}(t_j, Y_j^\varsigma)\Delta\Big] + \tilde{\sigma}.$$

It follows from (8) that $D_r\Phi_{t_n}(t_j, Y_j^\varsigma) = \sum_{l=-N}^{0} I^{[l]} D_r\Phi(t_{n+l}; t_j, Y_j^\varsigma)$, which together with Assumption 3.4 implies that

$$D_r\Phi(t_i; t_j, Y_j^\varsigma) \\ = \sum_{n=k}^{i-1}\Big[(1-\theta)\sum_{\ell=1}^{n_b}\sum_{l=-N}^{0}\int_{-\tau}^{0} k_b^\ell(\Phi_{t_n}(t_j, Y_j^\varsigma))I^{[l]}(s)\mathrm{d}\nu_3^\ell(s)D_r\Phi(t_{n+l}; t_j, Y_j^\varsigma)\Delta \\ + \theta\sum_{\ell=1}^{n_b}\sum_{l=-N}^{0}\int_{-\tau}^{0} k_b^\ell(\Phi_{t_{n+1}}(t_j, Y_j^\varsigma))I^{[l]}(s)\mathrm{d}\nu_3^\ell(s)D_r\Phi(t_{n+l+1}; t_j, Y_j^\varsigma)\Delta\Big] + \tilde{\sigma}. \quad (36)$$

Step 1. To derive a lower bound of $D_r\Phi(t_i; t_j, Y_j^\varsigma)$, we need to present a discrete comparison principle. Define a two-parameter nonnegative sequence $\{A_i^k\}_{0\le k,i\le N^\Delta}$ as follows: when $i \le k$, define $A_i^k = 0$; when $0 \le k \le i-1$, define

$$A_i^k = L_b\Delta\sum_{n=k}^{i-1}\sum_{\ell=1}^{n_b}\sum_{l=-N}^{0}\nu_3^\ell(\Delta_{l-1}+\Delta_l)\big((1-\theta)A_{n+l}^k + \theta A_{n+l+1}^k\big) + \tilde{\sigma},$$

where we let $\nu_3^\ell(\Delta_{-N-1}) = \nu_3^\ell(\Delta_0) = 0$. It follows from the definition that when $i_1 - k_1 = i_2 - k_2 > 0$, $A_{i_1}^{k_1} = A_{i_2}^{k_2} =: \mathcal{A}_{i_1-k_1}$. Then

$$\mathcal{A}_{i-k} = A_i^k = L_b\Delta\sum_{\ell=1}^{n_b}\sum_{l=-N}^{0}\nu_3^\ell(\Delta_{l-1}+\Delta_l)\big((1-\theta)\mathcal{A}_{i-1+l-k} + \theta\mathcal{A}_{i+l-k}\big) \\ + L_b\Delta\sum_{n=k}^{i-2}\sum_{\ell=1}^{n_b}\sum_{l=-N}^{0}\nu_3^\ell(\Delta_{l-1}+\Delta_l)\big((1-\theta)\mathcal{A}_{n+l-k} + \theta\mathcal{A}_{n+1+l-k}\big) + \tilde{\sigma} \\ = L_b\Delta\sum_{\ell=1}^{n_b}\sum_{l=-N}^{0}\nu_3^\ell(\Delta_{l-1}+\Delta_l)\big((1-\theta)\mathcal{A}_{i-1+l-k} + \theta\mathcal{A}_{i+l-k}\big) + \mathcal{A}_{i-k-1}.$$

This gives that for any $\Delta \in (0, \frac{1}{L_b\theta n_b})$,

$$\mathcal{A}_{i-k}=\big(1-L_b\Delta\theta\sum_{\ell=1}^{n_b}\nu_3^{\ell}(\Delta_{-1})\big)^{-1}\Big\{\mathcal{A}_{i-k-1}+L_b\Delta\Big[(1-\theta)\sum_{\ell=1}^{n_b}\sum_{l=-N}^{0}\nu_3^{\ell}(\Delta_{l-1}+\Delta_l)\times$$
$$\mathcal{A}_{i-1-k+l}+\theta\sum_{\ell=1}^{n_b}\sum_{l=-N+1}^{0}\nu_3^{\ell}(\Delta_{l-2}+\Delta_{l-1})\mathcal{A}_{i-k+l-1}\Big]\Big\}.$$

By the relation $(1-L_b\Delta\theta n_b)^{-1}=1+\frac{L_b\theta\Delta n_b}{1-L_b\Delta\theta n_b}$ and the iteration, we obtain

$$\begin{aligned}\mathcal{A}_{i-k}&\le(1+\frac{L_b\theta\Delta n_b}{1-L_b\Delta\theta n_b})\mathcal{A}_{i-k-1}+(1-L_b\Delta\theta n_b)^{-1}L_b\Delta\Big[(1-\theta)\sum_{\ell=1}^{n_b}\sum_{l=-N}^{0}\\&\quad\nu_3^{\ell}(\Delta_{l-1}+\Delta_l)\mathcal{A}_{i-1-k+l}+\theta\sum_{\ell=1}^{n_b}\sum_{l=-N+1}^{0}\nu_3^{\ell}(\Delta_{l-2}+\Delta_{l-1})\mathcal{A}_{i-k+l-1}\Big]\\&\le\mathcal{A}_1+\sum_{n=1}^{i-k-1}\frac{L_b\theta\Delta n_b}{1-L_b\Delta\theta n_b}\mathcal{A}_n+(1-L_b\Delta\theta n_b)^{-1}L_b\Delta\sum_{\ell=1}^{n_b}\sum_{n=1}^{i-k-1}\Big[(1-\theta)\times\\&\quad\sum_{l=-N}^{0}\nu_3^{\ell}(\Delta_{l-1}+\Delta_l)\mathcal{A}_{n+l}+\theta\sum_{l=-N+1}^{0}\nu_3^{\ell}(\Delta_{l-2}+\Delta_{l-1})\mathcal{A}_{n+l}\Big].\end{aligned}$$

Noting that

$$\begin{aligned}\sum_{n=1}^{i-k-1}\sum_{l=-N}^{0}\nu_3^{\ell}(\Delta_{l-1}+\Delta_l)\mathcal{A}_{n+l}&=\sum_{l=-N}^{0}\sum_{n=1}^{i-k-1}\nu_3^{\ell}(\Delta_{l-1}+\Delta_l)\mathcal{A}_{n+l}\\&\le\sum_{l=-N}^{0}\sum_{m=-N}^{i-k-1}\nu_3^{\ell}(\Delta_{l-1}+\Delta_l)\mathcal{A}_m\le 2\sum_{n=1}^{i-k-1}\mathcal{A}_n\end{aligned}$$

and $\mathcal{A}_1=(1-L_b\sum_{\ell=1}^{n_b}\nu_3^{\ell}(\Delta_{-1})\Delta\theta)^{-1}\tilde{\sigma}\le(1-L_b\Delta\theta n_b)^{-1}\tilde{\sigma}$, we arrive at

$$\mathcal{A}_{i-k}\le(1-L_b\Delta\theta n_b)^{-1}\tilde{\sigma}+\frac{L_bn_b(\theta+2)}{1-L_b\Delta\theta n_b}\sum_{n=1}^{i-k-1}\mathcal{A}_n\Delta.$$

Applying the discrete Grönwall inequality, we deduce

$$\begin{aligned}\mathcal{A}_{i-k}&\le(1-L_b\Delta\theta n_b)^{-1}\tilde{\sigma}\exp\Big\{\frac{L_bn_b(\theta+2)\Delta(i-k-1)}{1-L_b\Delta\theta n_b}\Big\}\\&=:K_0\exp\Big\{\frac{L_bn_b(\theta+2)\Delta(i-k-1)}{1-L_b\Delta\theta n_b}\Big\}.\end{aligned}\tag{37}$$

When $j\le i_1\le k_1$ and $r\in(t_{k_1},t_{k_1+1}]$, we have $D_r\Phi(t_{i_1};t_j,Y_j^{\varsigma})=0=A_{i_1}^{k_1}$. This gives $|D_r\Phi(t_{i_1};t_j,Y_j^{\varsigma})|=A_{i_1}^{k_1}$ for $i_1-k_1\le 0$ and $r\in(t_k,t_{k+1}]$. Now we claim that

$$|D_r\Phi(t_i;t_j,Y_j^\varsigma)| \le A_i^k, \text{ for any } r\in(t_k,t_{k+1}],\ j\le k\le i-1. \tag{38}$$

We prove the claim by the induction argument on $i-k$. Suppose that (38) holds for integers k, i satisfying $0\le i-k\le i-k'-1$. Then we show that (38) holds for $i-k=i-k'$. By (36) and the induction assumption, we have $|D_r\Phi(t_n;t_j,Y_j^\varsigma)|\le A_n^{k'}$ holds for all $n\le i-1$, which yields

$$\begin{aligned}
&|D_r\Phi(t_i;t_j,Y_j^\varsigma)|\\
&\le \sum_{n=k'}^{i-2}\sum_{\ell=1}^{n_b}\sum_{l=-N}^{0} L_b\nu_3^\ell(\Delta_{l-1}+\Delta_l)\big((1-\theta)A_{n+l}^{k'}\Delta+\theta A_{n+l+1}^{k'}\Delta\big)+\tilde\sigma\\
&\quad+(1-\theta)\sum_{\ell=1}^{n_b}\sum_{l=-N}^{0} L_b\nu_3^\ell(\Delta_{l-1}+\Delta_l)A_{i-1+l}^{k'}\Delta\\
&\quad+\theta\sum_{\ell=1}^{n_b}\sum_{l=-N}^{-1} L_b\nu_3^\ell(\Delta_{l-1}+\Delta_l)A_{i+l}^{k'}\Delta+L_b\theta\Delta\sum_{\ell=1}^{n_b}\nu_3^\ell(\Delta_{-1})|D_r\Phi(t_i;t_j,Y_j^\varsigma)|.
\end{aligned}$$

Thus we derive

$$\begin{aligned}
|D_r\Phi(t_i;t_j,Y_j^\varsigma)| &\le (1-L_b\Delta\theta\sum_{\ell=1}^{n_b}\nu_3^\ell(\Delta_{-1}))^{-1}\Big[A_{i-1}^{k'}+L_b\Delta(1-\theta)\sum_{\ell=1}^{n_b}\sum_{l=-N}^{0}\nu_3^\ell(\Delta_{l-1}\\
&\quad+\Delta_l)A_{i-1+l}^{k'}+L_b\Delta\theta\sum_{\ell=1}^{n_b}\sum_{l=-N}^{-1}\nu_3^\ell(\Delta_{l-1}+\Delta_l)A_{i+l}^{k'}\Big]=A_i^{k'},\qquad r\in(t_{k'},t_{k'+1}].
\end{aligned}$$

This finishes the proof of the claim (38).

Step 2. From (36) and (37), we obtain

$$\begin{aligned}
&|D_r\Phi(t_i;t_j,Y_j^\varsigma)|\\
&\ge\tilde\sigma-L_b\Delta\Big[\sum_{n=k}^{i-1}\sum_{\ell=1}^{n_b}\sum_{l=-N}^{0}\Big((1-\theta)\nu_3^\ell(\Delta_{l-1}+\Delta_l)A_{n+l}^k+\theta\nu_3^\ell(\Delta_{l-1}+\Delta_l)A_{n+l+1}^k\Big)\Big]\\
&\ge\tilde\sigma-2L_bn_b\Delta\Big((1-\theta)K_0e^{-\frac{L_bn_b(\theta+2)\Delta}{1-L_b\Delta\theta n_b}}+\theta K_0\Big)\frac{\exp\{\frac{L_bn_b(\theta+2)\Delta(i-k)}{1-L_b\Delta\theta n_b}\}-1}{\exp\{\frac{L_bn_b(\theta+2)\Delta}{1-L_b\Delta\theta n_b}\}-1}\\
&\ge\tilde\sigma-\frac{2}{\theta+2}\Big(\exp\Big\{\frac{L_bn_b(\theta+2)\Delta(i-k)}{1-L_b\Delta\theta n_b}\Big\}-1\Big)\tilde\sigma\\
&\ge\tilde\sigma-\Big(\exp\Big\{\frac{L_bn_b(\theta+2)\Delta(i-k)}{1-L_b\Delta\theta n_b}\Big\}-1\Big)\tilde\sigma,
\end{aligned}$$

where we used $e^x\ge x+1, x\ge 0$. Then $|D_r\Phi(t_i;t_j,Y_j^\varsigma)|\ge\frac{1}{2}\tilde\sigma$, when $\Delta\le\frac{1}{2L_b\theta n_b}$ and $i-k\le\frac{\ln(\frac{3}{2})}{2L_bn_b(\theta+2)\Delta}$. Thus,

$$\gamma_{\Phi(T;t_j,Y_j^\varsigma)} := \int_0^T D_r\Phi(T;t_j,Y_j^\varsigma)(D_r\Phi(T;t_j,Y_j^\varsigma))^\top \mathrm{d}r$$
$$\geq \int_{(T-T_0)\vee 0}^T |D_r\Phi(T;t_j,Y_j^\varsigma)|^2 \mathrm{d}r \geq \frac{1}{4}(T\wedge T_0)\tilde{\sigma}^2,$$

where $T_0 := \frac{\ln(\frac{3}{2})}{2L_b n_b(\theta+2)}$. This finishes the proof of (35) for $d=1$.

For the case of $d\geq 2$, by replacing $D_r\Phi$ with $u^\top D_r\Phi$ in the above argument, one can also obtain (35).

Moreover, it follows from (35) that

$$\lambda_{\min}(\gamma_{\Phi(T;t_j,Y_j^\varsigma)}) = \min_{u\in\mathbb{R}^d,|u|=1} u^\top \gamma_{\Phi(T;t_j,Y_j^\varsigma)} u \geq \frac{1}{4}(T\wedge T_0)\tilde{\sigma}^2,$$

which implies

$$|\det(\gamma_{\Phi(T;t_j,Y_j^\varsigma)})^{-1}| \leq \Big[\frac{1}{4}(T\wedge T_0)\tilde{\sigma}^2\Big]^{-d}.$$

Thus the proof is completed. □

With Lemmas 3.2 and 3.3 in hand, we present the following Malliavin integration by parts formula, which plays an important role in the test-functional-independent weak convergence analysis.

Lemma 3.4. *Let $\alpha=(\alpha_1,\ldots,\alpha_d)$ be the multi-index with $\alpha_j\in\mathbb{N}$, $j=1,\ldots,d$ and $|\alpha| := \sum_{j=1}^d \alpha_j \leq 2$, $f\in\mathscr{C}$, and $G_1\in\mathbb{D}^{|\alpha|+1,\infty}$. Then under conditions in Lemmas 3.2 and 3.3, there exist a constant $\Delta_2\in(0,1]$ and an element $H_{|\alpha|+1}$ such that for $\Delta\in(0,\Delta_2]$, $i\in\mathbb{N}$ with $i\leq N^\Delta$,*

$$\mathbb{E}[\partial_\alpha f(\Phi(T;t_i,Y_i^\varsigma))G_1] = \mathbb{E}[F(\Phi(T;t_i,Y_i^\varsigma))H_{|\alpha|+1}(\Phi(T;t_i,Y_i^\varsigma),G_1)], \tag{39}$$

where F is an antiderivative of f given in the definition of $\mathscr{C}$ (see (27)). Moreover, for $T\geq T_0$,

$$|\mathbb{E}[\partial_\alpha f(\Phi(T;t_i,Y_i^\varsigma))G_1]| \leq K\|G_1\|_{|\alpha|+1,2}, \tag{40}$$

where T_0 is given in Theorem 3.2, and Y_i^ς is given in (31).

Proof. According to the definition of $\mathscr{C}$, we apply [1, Proposition 2.1.4, (2.29)–(2.32)] to obtain (39). Moreover, for $q>q_1\geq 1$, there exist constants $\eta_1,\eta_2>0$ and integers $n_1,n_2>0$ such that

$$\|H_{|\alpha|+1}(\Phi(T;t_i,Y_i^\varsigma),G_1)\|_{q_1}$$
$$\leq K(q_1,q)\|\det(\gamma_{\Phi(T;t_i,Y_i^\varsigma)})^{-1}\|_{0,\eta_1}^{n_1}\|D\Phi(T;t_i,Y_i^\varsigma)\|_{|\alpha|+1,\eta_2}^{n_2}\|G_1\|_{|\alpha|+1,q}.$$

It follows from Lemmas 3.2 and 3.3 that for any $\Delta\in(0,\Delta_2]$ with $\Delta_2 := \tilde{\Delta}\wedge\frac{1}{2L_b n_b\theta}$,

$$\|H_{|\alpha|+1}(\Phi(T;t_i,Y_i^{\varsigma}),G_1)\|_{q_1}\le K\Big[\frac{1}{4}(T\wedge T_0)\tilde{\sigma}^2\Big]^{-dn_1}\|G_1\|_{|\alpha|+1,q},$$

where $\tilde{\Delta}:=\tilde{\Delta}(\eta_2)$ is given in Lemma 3.2. Taking $q_1=1$ and $q=2$, we finish the proof. □

3.2.2. Weak convergence analysis

In this subsection, we present the test-functional-independent weak convergence rate of the θ-EM discretization.

Theorem 3.3. *Let conditions in Theorem 3.2 hold. Then there exists $\tilde{\Delta}>0$ such that for $\Delta\in(0,\tilde{\Delta}]$ and $T\ge T_0$,*

$$\sup_{f\in\mathscr{C}}\big|\mathbb{E}[f(x^{\xi}(T))]-\mathbb{E}[f(y^{\xi,\Delta}(T))]\big|\le K_T\Delta,$$

where T_0 is given in Theorem 3.2.

Proof. Without loss of generality, we take the parameter n_b in Assumption 3.4 to be $n_b=1$. The case of $n_b>1$ can be proved similarly. In order to obtain the test-functional-independent weak convergence rate, we need to estimate terms $\mathcal{I}_0,\mathcal{I}_b$, and $\mathcal{I}_{b,\theta}$ in (34) by means of the Malliavin integration by parts formula (39) and the inequality (40).

Estimate of term $\mathcal{I}_0$. Recalling

$$\mathcal{I}_0=\mathbb{E}[f(\Phi(T;t_{N^\Delta},\varphi^{Int}_{t_{N^\Delta}}(0,\xi)))]-\mathbb{E}[f(\Phi(T;t_{N^\Delta},\varphi^{Int}_{t_{N^\Delta}}(0,\xi^{Int})))],$$

it follows from (40), Lemma 3.1, and $\|\xi-\xi^{Int}\|\le K\Delta$ that

$$\begin{aligned}|\mathcal{I}_0|&=\big|\mathbb{E}[f(\varphi^{Int}(t_{N^\Delta};0,\xi)]-\mathbb{E}[f(\varphi^{Int}(t_{N^\Delta};0,\xi^{Int}))]\big|\\ &\le\int_0^1\Big|\mathbb{E}\Big[f'(\varphi^{Int}(t_{N^\Delta};0,\varsigma\xi+(1-\varsigma)\xi^{Int}))\mathcal{D}\varphi^{Int}(t_{N^\Delta};0,\varsigma\xi+(1-\varsigma)\xi^{Int})(\xi-\xi^{Int})\Big]\Big|\mathrm{d}\varsigma\\ &\le K\int_0^1\|\mathcal{D}\varphi^{Int}(t_{N^\Delta};0,\varsigma\xi+(1-\varsigma)\xi^{Int})(\xi-\xi^{Int})\|_{2,2}\mathrm{d}\varsigma\le K_T\Delta.\end{aligned}$$

Estimate of term $\mathcal{I}_b$. Recalling the definition of $\mathcal{I}_b$, we have that

$$\begin{aligned}\mathcal{I}_b=\sum_{i=1}^{N^\Delta}\int_0^1\mathbb{E}\Big[\Big\langle f'(\Phi(T;t_i,Y_i^{\varsigma}))\mathcal{D}\Phi(T;t_i,Y_i^{\varsigma}),I^{[0]}\int_{t_{i-1}}^{t_i}\int_0^1\mathcal{D}b(Z_{i,r}^{\beta_1})\\ \big(\varphi_r(t_{i-1},\varphi_{t_{i-1}}(0,\xi^{Int}))-\varphi^{Int}_{t_{i-1}}(0,\xi^{Int})\big)\mathrm{d}\beta_1\mathrm{d}r\Big\rangle\Big]\mathrm{d}\varsigma,\end{aligned}$$

where $Z_{i,r}^{\beta_1}:=\beta_1\varphi_r(t_{i-1},\varphi_{t_{i-1}}(0,\xi^{Int}))+(1-\beta_1)\varphi^{Int}_{t_{i-1}}(0,\xi^{Int})$. To estimate $\mathcal{I}_b$, we need to split $\varphi_r(t_{i-1},\varphi_{t_{i-1}}(0,\xi^{Int}))-\varphi^{Int}_{t_{i-1}}(0,\xi^{Int})$ for $r\in[t_{i-1},t_i)$, based on the definitions of the exact functional solution and its linear interpolation. For $s\in[t_j,t_{j+1}]\subset[-\tau,0]$, we have

$$\varphi_{t_{i-1}}^{Int}(0,\xi^{Int})(s) = \frac{t_{j+1}-s}{\Delta}\varphi(t_{i+j-1};0,\xi^{Int}) + \frac{s-t_j}{\Delta}\varphi(t_{i+j};0,\xi^{Int}). \tag{41}$$

We also have $r+s \in [t_{i+j-1}, t_{i+j+1})$, which leads to the following two cases. For notational simplicity and to illustrate the main idea of the proof, we suppose $t_{i+j-1} \geq 0$. The case $t_{i+j-1} < 0$, which involves contributions from the initial values on $[-\tau, 0]$, can be treated similarly.

Case 1: $r+s \in [t_{i+j-1}, t_{i+j})$. In this case, by the integral form of the exact solution of (1), we have

$$\begin{aligned}
&\varphi_r(t_{i-1}, \varphi_{t_{i-1}}(0,\xi^{Int}))(s) = \varphi(r+s;0,\xi^{Int}) \\
&= \varphi(t_{i+j-1};0,\xi^{Int}) + \int_{t_{i+j-1}}^{r+s} b(\varphi_v(0,\xi^{Int}))\mathrm{d}v + \int_{t_{i+j-1}}^{r+s} \tilde{\sigma}\mathrm{d}W(v).
\end{aligned}$$

Hence, combining this with (41) yields

$$\begin{aligned}
&\varphi_r(t_{i-1}, \varphi_{t_{i-1}}(0,\xi^{Int}))(s) - \varphi_{t_{i-1}}^{Int}(0,\xi^{Int})(s) \\
&= \frac{t_j - s}{\Delta}(\varphi(t_{i+j};0,\xi^{Int}) - \varphi(t_{i+j-1};0,\xi^{Int})) \\
&\quad + \int_{t_{i+j-1}}^{r+s} b(\varphi_v(0,\xi^{Int}))\mathrm{d}v + \int_{t_{i+j-1}}^{r+s} \tilde{\sigma}\mathrm{d}W(v).
\end{aligned}$$

Case 2: $r+s \in [t_{i+j}, t_{i+j+1})$. In this case, by the integral form of the exact solution of (1) again, we have

$$\varphi_r(t_{i-1}, \varphi_{t_{i-1}}(0,\xi^{Int}))(s) = \varphi(t_{i+j};0,\xi^{Int}) + \int_{t_{i+j}}^{r+s} b(\varphi_v(0,\xi^{Int}))\mathrm{d}v + \int_{t_{i+j}}^{r+s} \tilde{\sigma}\mathrm{d}W(v).$$

This, together with (41) leads to

$$\begin{aligned}
&\varphi_r(t_{i-1}, \varphi_{t_{i-1}}(0,\xi^{Int}))(s) - \varphi_{t_{i-1}}^{Int}(0,\xi^{Int})(s) \\
&= \frac{t_{j+1} - s}{\Delta}(\varphi(t_{i+j};0,\xi^{Int}) - \varphi(t_{i+j-1};0,\xi^{Int})) \\
&\quad + \int_{t_{i+j}}^{r+s} b(\varphi_v(0,\xi^{Int}))\mathrm{d}v + \int_{t_{i+j}}^{r+s} \tilde{\sigma}\mathrm{d}W(v).
\end{aligned}$$

Combining Case 1 and Case 2, we deduce

$$\varphi_r(t_{i-1}, \varphi_{t_{i-1}}(0,\xi^{Int})) - \varphi_{t_{i-1}}^{Int}(0,\xi^{Int})$$

$$= \sum_{j=-N}^{-1} \Big\{ \mathbf{1}_{[t_{i+j-1}-r, t_{i+j}-r)}(\cdot) \Big[\frac{t_j - \cdot}{\Delta} \Big(\int_{t_{i+j-1}}^{t_{i+j}} b(\varphi_v(0, \xi^{Int})) \mathrm{d}v$$

$$+ \int_{t_{i+j-1}}^{t_{i+j}} \tilde{\sigma} \mathrm{d}W(v) \Big) + \int_{t_{i+j-1}}^{r+\cdot} b(\varphi_v(0, \xi^{Int})) \mathrm{d}v + \int_{t_{i+j-1}}^{r+\cdot} \tilde{\sigma} \mathrm{d}W(v) \Big]$$

$$+ \mathbf{1}_{[t_{i+j}-r, t_{i+j+1}-r)}(\cdot) \Big[\frac{t_{j+1} - \cdot}{\Delta} \Big(\int_{t_{i+j-1}}^{t_{i+j}} b(\varphi_v(0, \xi^{Int})) \mathrm{d}v + \int_{t_{i+j-1}}^{t_{i+j}} \tilde{\sigma} \mathrm{d}W(v) \Big)$$

$$+ \int_{t_{i+j}}^{r+\cdot} b(\varphi_v(0, \xi^{Int})) \mathrm{d}v + \int_{t_{i+j}}^{r+\cdot} \tilde{\sigma} \mathrm{d}W(v) \Big] \Big\}.$$

Inserting the above equality into $\mathcal{I}_b$, we are in the position to estimate $\mathcal{I}_b$. We only estimate the sub-term

$$\mathcal{I}_b^0 := \sum_{i=1}^{N^\Delta} \int_0^1 \mathbb{E}\Big[\Big\langle f'(\Phi(T; t_i, Y_i^\varsigma)) \mathcal{D}\Phi(T; t_i, Y_i^\varsigma), I^{[0]} \int_{t_{i-1}}^{t_i} \int_0^1 \mathcal{D}b(Z_{i,r}^{\beta_1})$$

$$\sum_{j=-N}^{-1} \mathbf{1}_{[t_{i+j-1}-r, t_{i+j}-r)}(\cdot) \frac{t_j - \cdot}{\Delta} \int_{t_{i+j-1}}^{t_{i+j}} b(\varphi_v(0, \xi^{Int})) \mathrm{d}v \mathrm{d}\beta_1 \mathrm{d}r \Big\rangle \Big] \mathrm{d}\varsigma,$$

the sub-term

$$\mathcal{I}_b^1 := \sum_{i=1}^{N^\Delta} \int_0^1 \mathbb{E}\Big[\Big\langle f'(\Phi(T; t_i, Y_i^\varsigma)) \mathcal{D}\Phi(T; t_i, Y_i^\varsigma), I^{[0]} \int_{t_{i-1}}^{t_i} \int_0^1 \mathcal{D}b(Z_{i,r}^{\beta_1})$$

$$\sum_{j=-N}^{-1} \mathbf{1}_{[t_{i+j-1}-r, t_{i+j}-r)}(\cdot) \frac{t_j - \cdot}{\Delta} \int_{t_{i+j-1}}^{t_{i+j}} \tilde{\sigma} \mathrm{d}W(v) \mathrm{d}\beta_1 \mathrm{d}r \Big\rangle \Big] \mathrm{d}\varsigma,$$

and the sub-term

$$\mathcal{I}_b^2 := \sum_{i=1}^{N^\Delta} \int_0^1 \mathbb{E}\Big[\Big\langle f'(\Phi(T; t_i, Y_i^\varsigma)) \mathcal{D}\Phi(T; t_i, Y_i^\varsigma), I^{[0]} \int_{t_{i-1}}^{t_i} \int_0^1 \mathcal{D}b(Z_{i,r}^{\beta_1})$$

$$\sum_{j=-N}^{-1} \mathbf{1}_{[t_{i+j-1}-r, t_{i+j}-r)}(\cdot) \int_{t_{i+j-1}}^{r+\cdot} \tilde{\sigma} \mathrm{d}W(v) \mathrm{d}\beta_1 \mathrm{d}r \Big\rangle \Big] \mathrm{d}\varsigma,$$

since other sub-terms in $\mathcal{I}_b$ can be estimated similarly.

For the term $\mathcal{I}_b^0$, it follows from (39) and (40) that

$$|\mathcal{I}_b^0| \le K \sum_{i=1}^{N^\Delta} \int_0^1 \Big\| \mathcal{D}\Phi(T; t_i, Y_i^\varsigma) I^{[0]} \int_{t_{i-1}}^{t_i} \int_0^1 \mathcal{D}b(Z_{i,r}^{\beta_1}) \sum_{j=-N}^{-1} \mathbf{1}_{[t_{i+j-1-r}, t_{i+j}-r)}(\cdot) \times \frac{t_j - \cdot}{\Delta} \int_{t_{i+j-1}}^{t_{i+j}} b(\varphi_v(0, \xi^{Int})) \mathrm{d}v \mathrm{d}\beta_1 \mathrm{d}r \Big\|_{2,2} \mathrm{d}\varsigma. \tag{42}$$

According to Lemma 3.2 and $\sup_{\phi_1 \in \mathcal{C}^d} |\mathcal{D}b(\phi_1)\phi_2| \le K\|\phi_2\|$, we have

$$\Big\| \mathcal{D}\Phi(T; t_i, Y_i^\varsigma) I^{[0]} \int_{t_{i-1}}^{t_i} \int_0^1 \mathcal{D}b(Z_{i,r}^{\beta_1}) \sum_{j=-N}^{-1} \mathbf{1}_{[t_{i+j-1}-r, t_{i+j}-r)}(\cdot) \times \frac{t_j - \cdot}{\Delta} \int_{t_{i+j-1}}^{t_{i+j}} b(\varphi_v(0, \xi^{Int})) \mathrm{d}v \mathrm{d}\beta_1 \mathrm{d}r \Big\|_{0,2} \le K_T \Delta^2. \tag{43}$$

In addition, using the chain rule of the Malliavin derivative, Lemma 3.2, and Assumption 3.4, we derive that for some $\tilde{p} \ge 2$,

$$\begin{aligned}
&\Big(\int_{t_i}^T \mathbb{E}\Big[\Big\| D_{r_1} \Big[\mathcal{D}\Phi(T; t_i, Y_i^\varsigma) I^{[0]} \int_{t_{i-1}}^{t_i} \int_0^1 \mathcal{D}b(Z_{i,r}^{\beta_1}) \sum_{j=-N}^{-1} \mathbf{1}_{[t_{i+j-1}-r, t_{i+j}-r)}(\cdot) \times \\
&\frac{t_j - \cdot}{\Delta} \int_{t_{i+j-1}}^{t_{i+j}} b(\varphi_v(0, \xi^{Int})) \mathrm{d}v \mathrm{d}\beta_1 \mathrm{d}r \Big] \Big\|^2 \Big] \mathrm{d}r_1 \Big)^{\frac{1}{2}} \\
&\le K_T \Big\{ \int_{t_i}^T \Big(\mathbb{E}\Big[\Big\| I^{[0]} \int_{t_{i-1}}^{t_i} \int_0^1 \mathcal{D}b(Z_{i,r}^{\beta_1}) \sum_{j=-N}^{-1} \mathbf{1}_{[t_{i+j-1}-r, t_{i+j}-r)}(\cdot) \times \\
&\frac{t_j - \cdot}{\Delta} \int_{t_{i+j-1}}^{t_{i+j}} b(\varphi_v(0, \xi^{Int})) \mathrm{d}v \mathrm{d}\beta_1 \mathrm{d}r \Big\|^2 \big(1 + \|D_{r_1} Y_i^\varsigma\|^{\tilde{p}} \big) \Big] \\
&+ \mathbb{E}\Big[\Big\| \mathcal{D}\Phi(T; t_i, Y_i^\varsigma) I^{[0]} \int_{t_{i-1}}^{t_i} \int_0^1 \mathcal{D}^2 b(Z_{i,r}^{\beta_1}) \Big(D_{r_1} Z_{i,r}^{\beta_1}, \sum_{j=-N}^{-1} \mathbf{1}_{[t_{i+j-1}-r, t_{i+j}-r)}(\cdot) \times \\
&\frac{t_j - \cdot}{\Delta} \int_{t_{i+j-1}}^{t_{i+j}} b(\varphi_v(0, \xi^{Int})) \mathrm{d}v \Big) \mathrm{d}\beta_1 \mathrm{d}r \Big\|^2 \Big]
\end{aligned}$$

$$+\mathbb{E}\Big[\Big\|\mathcal{D}\Phi(T;t_i,Y_i^{\varsigma})I^{[0]}\int_{t_{i-1}}^{t_i}\int_0^1 \mathcal{D}b(Z_{i,r}^{\beta_1})\sum_{j=-N}^{-1}\mathbf{1}_{[t_{i+j-1}-r,t_{i+j}-r)}(\cdot)\times$$

$$\frac{t_j-\cdot}{\Delta}\int_{t_{i+j-1}}^{t_{i+j}}\mathcal{D}b(\varphi_v(0,\xi^{Int}))D_{r_1}\varphi_v(0,\xi^{Int})\mathrm{d}v\mathrm{d}\beta_1\mathrm{d}r\Big\|^2\Big]\Big)\mathrm{d}r_1\Big\}^{\frac{1}{2}}\le K_T\Delta^2. \tag{44}$$

Similarly,

$$\Big\{\int_{t_i}^{T}\int_{t_i}^{T}\mathbb{E}\Big[\Big\|D_{r_1,r_2}\Big[\mathcal{D}\Phi(T;t_i,Y_i^{\varsigma})I^{[0]}\int_{t_{i-1}}^{t_i}\int_0^1\mathcal{D}b(Z_{i,r}^{\beta_1})\sum_{j=-N}^{-1}\mathbf{1}_{[t_{i+j-1}-r,t_{i+j}-r)}(\cdot)\times$$

$$\frac{t_j-\cdot}{\Delta}\int_{t_{i+j-1}}^{t_{i+j}}b(\varphi_v(0,\xi^{Int}))\mathrm{d}v\mathrm{d}\beta_1\mathrm{d}r\Big]\Big\|^2\Big]\mathrm{d}r_1\mathrm{d}r_2\Big\}^{\frac{1}{2}}\le K_T\Delta^2. \tag{45}$$

Inserting (43)–(45) into (42), one has $|\mathcal{I}_b^0|\le K_T\Delta$.

For the sub-term $\mathcal{I}_b^1$, we have

$$\mathcal{I}_b^1=\sum_{i=1}^{N^\Delta}\int_0^1\int_{t_{i-1}}^{t_i}\int_0^1\sum_{j=-N}^{-1}\mathbf{1}_{[t_{i+j-1}-r,t_{i+j}-r)}(\cdot)\frac{t_j-\cdot}{\Delta}\times$$

$$\mathbb{E}\Big[\Big\langle(I^{[0]}\mathcal{D}b(Z_{i,r}^{\beta_1}))^*f'(\Phi(T;t_i,Y_i^{\varsigma}))\mathcal{D}\Phi(T;t_i,Y_i^{\varsigma}),\int_{t_{i+j-1}}^{t_{i+j}}\tilde{\sigma}\mathrm{d}W(v)\Big\rangle\Big]\mathrm{d}\beta_1\mathrm{d}r\mathrm{d}\varsigma$$

$$=\sum_{i=1}^{N^\Delta}\sum_{j=-N}^{-1}\int_0^1\int_{t_{i-1}}^{t_i}\int_0^1\int_{t_{i+j-1}}^{t_{i+j}}\mathbf{1}_{[t_{i+j-1}-r,t_{i+j}-r)}(\cdot)\frac{t_j-\cdot}{\Delta}\times$$

$$\mathbb{E}\Big[\Big\langle D_v\big[(I^{[0]}\mathcal{D}b(Z_{i,r}^{\beta_1}))^*f'(\Phi(T;t_i,Y_i^{\varsigma}))\mathcal{D}\Phi(T;t_i,Y_i^{\varsigma})\big],\tilde{\sigma}\mathrm{I}_d\Big\rangle\Big]\mathrm{d}v\mathrm{d}\beta_1\mathrm{d}r\mathrm{d}\varsigma$$

$$=\sum_{i=1}^{N^\Delta}\sum_{j=-N}^{-1}\int_0^1\int_{t_{i-1}}^{t_i}\int_0^1\int_{t_{i+j-1}}^{t_{i+j}}\Big\{\mathbb{E}\Big[\Big\langle f'(\Phi(T;t_i,Y_i^{\varsigma}))\mathcal{D}\Phi(T;t_i,Y_i^{\varsigma}),$$

$$I^{[0]}\mathcal{D}^2b(Z_{i,r}^{\beta_1})\Big(D_vZ_{i,r}^{\beta_1},\mathbf{1}_{[t_{i+j-1}-r,t_{i+j}-r)}(\cdot)\frac{t_j-\cdot}{\Delta}\tilde{\sigma}\mathrm{I}_d\Big)\Big\rangle\Big]$$

$$+\mathbb{E}\Big[\Big\langle f'(\Phi(T;t_i,Y_i^{\varsigma}))D_v\mathcal{D}\Phi(T;t_i,Y_i^{\varsigma}),I^{[0]}\mathcal{D}b(Z_{i,r}^{\beta_1})$$

$$\mathbf{1}_{[t_{i+j-1}-r,t_{i+j}-r)}(\cdot)\frac{t_j-\cdot}{\Delta}\tilde{\sigma}\mathrm{I}_d\Big\rangle\Big]+\mathbb{E}\Big[\Big\langle f''(\Phi(T;t_i,Y_i^{\varsigma}))D_v\Phi(T;t_i,Y_i^{\varsigma})$$

$$\mathcal{D}\Phi(T;t_i,Y_i^{\varsigma}),I^{[0]}\mathcal{D}b(Z_{i,r}^{\beta_1})\mathbf{1}_{[t_{i+j-1}-r,t_{i+j}-r)}(\cdot)\frac{t_j-\cdot}{\Delta}\tilde{\sigma}\mathrm{I}_d\Big\rangle\Big]\Big\}\mathrm{d}v\mathrm{d}\beta_1\mathrm{d}r\mathrm{d}\varsigma,$$

where I_d is the d-dimensional all-ones vector. Similarly, together with (39) and (40), we can derive that $|\mathcal{I}_b^1| \leq K_T \Delta$.

For the sub-term $\mathcal{I}_b^2$, based on Assumption 3.4, we have

$$
\begin{aligned}
\mathcal{I}_b^2 = & \sum_{i=1}^{N^\Delta} \int_0^1 \mathbb{E}\Big[\Big\langle f'(\Phi(T;t_i,Y_i^\varsigma))\mathcal{D}\Phi(T;t_i,Y_i^\varsigma), I^{[0]} \int_{t_{i-1}}^{t_i} \int_0^1 \int_{-\tau}^0 k_b^1(Z_{i,r}^{\beta_1}) \\
& \sum_{j=-N}^{-1} \mathbf{1}_{[t_{i+j-1}-r, t_{i+j}-r)}(s) \int_{t_{i+j-1}}^{r+s} \tilde{\sigma} \mathrm{d}W(v) \mathrm{d}\nu_3^1(s) \mathrm{d}\beta_1 \mathrm{d}r \Big\rangle\Big] \mathrm{d}\varsigma \\
= & \sum_{i=1}^{N^\Delta} \sum_{j=-N}^{-1} \int_0^1 \int_{t_{i-1}}^{t_i} \int_0^1 \int_{-\tau\vee(t_{i+j-1}-r)}^{(t_{i+j}-r)\wedge 0} \int_{t_{i+j-1}}^{r+s} \mathbb{E}\Big[\Big\langle D_v[(I^{[0]} k_b^1(Z_{i,r}^{\beta_1}))^* \\
& f'(\Phi(T;t_i,Y_i^\varsigma))\mathcal{D}\Phi(T;t_i,Y_i^\varsigma)], \tilde{\sigma}\mathrm{I}_d \Big\rangle\Big] \mathrm{d}v \mathrm{d}\nu_3^1(s) \mathrm{d}\beta_1 \mathrm{d}r \mathrm{d}\varsigma \\
= & \sum_{i=1}^{N^\Delta} \sum_{j=-N}^{-1} \int_0^1 \int_{t_{i-1}}^{t_i} \int_0^1 \int_{-\tau\vee(t_{i+j-1}-r)}^{(t_{i+j}-r)\wedge 0} \int_{t_{i+j-1}}^{r+s} \Big\{\mathbb{E}\Big[\Big\langle f'(\Phi(T;t_i,Y_i^\varsigma)) \\
& \mathcal{D}\Phi(T;t_i,Y_i^\varsigma), I^{[0]} D_v k_b^1(Z_{i,r}^{\beta_1}) \tilde{\sigma}\mathrm{I}_d \Big\rangle\Big] + \mathbb{E}\Big[\Big\langle f'(\Phi(T;t_i,Y_i^\varsigma)) D_v \mathcal{D}\Phi(T;t_i,Y_i^\varsigma), \\
& I^{[0]} k_b^1(Z_{i,r}^{\beta_1}) \tilde{\sigma}\mathrm{I}_d \Big\rangle\Big] + \mathbb{E}\Big[\Big\langle f''(\Phi(T;t_i,Y_i^\varsigma)) D_v \Phi(T;t_i,Y_i^\varsigma) \mathcal{D}\Phi(T;t_i,Y_i^\varsigma), \\
& I^{[0]} k_b^1(Z_{i,r}^{\beta_1}) \tilde{\sigma}\mathrm{I}_d \Big\rangle\Big]\Big\} \mathrm{d}v \mathrm{d}\nu_3^1(s) \mathrm{d}\beta_1 \mathrm{d}r \mathrm{d}\varsigma.
\end{aligned}
$$

Combining (39), (40), and Lemma 3.2, we deduce

$$
\begin{aligned}
|\mathcal{I}_b^2| \leq K & \sum_{i=1}^{N^\Delta} \sum_{j=-N}^{-1} \int_0^1 \int_{t_{i-1}}^{t_i} \int_0^1 \int_{-\tau\vee(t_{i+j-1}-r)}^{(t_{i+j}-r)\wedge 0} \int_{t_{i+j-1}}^{r+s} \\
& \Big\{ \Big\| \mathcal{D}\Phi(T;t_i,Y_i^\varsigma) I^{[0]} D_v k_b^1(Z_{i,r}^{\beta_1}) \tilde{\sigma}\mathrm{I}_d \Big\|_{2,2} + \Big\| D_v \mathcal{D}\Phi(T;t_i,Y_i^\varsigma) I^{[0]} k_b^1(Z_{i,r}^{\beta_1}) \tilde{\sigma}\mathrm{I}_d \Big\|_{2,2} \\
& + \big\| D_v \Phi(T;t_i,Y_i^\varsigma) \mathcal{D}\Phi(T;t_i,Y_i^\varsigma) I^{[0]} k_b^1(Z_{i,r}^{\beta_1}) \tilde{\sigma}\mathrm{I}_d \big\|_{3,2} \Big\} \mathrm{d}v \mathrm{d}\nu_3^1(s) \mathrm{d}\beta_1 \mathrm{d}r \mathrm{d}\varsigma \\
& \leq K_T \Delta,
\end{aligned}
$$

where we used the assumption

$$
\sup_{l\in\{1,2,3\}} \sup_{\phi\in\mathcal{C}^d} \Big(\|\mathcal{D}^l k_b^1(\phi)\|_{\mathcal{L}((\mathcal{C}^d)^{\otimes l};\mathbb{R}^{d\times d})} + |k_b^1(\phi)| \Big) \leq K.
$$

Hence, we have $|\mathcal{I}_b| \leq K_T \Delta$.

Estimate of $\mathcal{I}_{b,\theta}$. Similar to proof of $\mathcal{I}_b$, we derive that $|\mathcal{I}_{b,\theta}| \leq K_T \Delta$.

Combining estimates of terms $\mathcal{I}_0, \mathcal{I}_b$ and $\mathcal{I}_{b,\theta}$, we complete the proof. □

4. Proofs of Lemmas 2.1 and 2.2

Proof of Lemma 2.1. Similar to the proof of [22, Theorem 2], we can obtain the first inequality in Lemma 2.1. For the second inequality in Lemma 2.1, when $v \le t$, $D_v x^\xi(t)$ satisfies

$$D_v x^\xi(t) = \int_v^t \mathcal{D}b(x_s^\xi) D_v x_s^\xi \mathrm{d}s + \int_v^t \mathcal{D}\sigma(x_s^\xi) D_v x_s^\xi \mathrm{d}W(s) + \sigma(x_v^\xi)\mathbf{1}_{[0,t]}(v).$$

By the Itô formula, (2), (3), the Burkholder–Davis–Gundy inequality, and the Young inequality, we deduce

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{v\le t\le T} |D_v x^\xi(t)|^{2p}\Big] \\
&\le \mathbb{E}[|\sigma(x_v^\xi)|^{2p}] + K\int_v^T \mathbb{E}\Big[|D_v x^\xi(s)|^{2p} + \int_{-\tau}^0 |D_v x_s^\xi(r)|^{2p}\mathrm{d}\nu_1(r) \\
&\quad + \int_{-\tau}^0 |D_v x_s^\xi(r)|^{2p}\mathrm{d}\nu_2(r)\Big]\mathrm{d}s + K\mathbb{E}\Big[\Big(\int_v^T |D_v x^\xi(s)|^{4p-2}|\mathcal{D}\sigma(x_s^\xi)D_v x_s^\xi|^2\mathrm{d}s\Big)^{\frac{1}{2}}\Big] \\
&\le K_T + K\int_v^T \mathbb{E}\Big[\sup_{v\le r\le s}|D_v x^\xi(r)|^{2p}\Big]\mathrm{d}s + \frac{1}{2}\mathbb{E}\Big[\sup_{v\le t\le T}|D_v x^\xi(t)|^{2p}\Big].
\end{aligned}$$

Using the Grönwall inequality finishes the proof. □

Proof of Lemma 2.2. We first show that $\mathbb{E}\Big[\sup_{0\le t_k\le T}\|y_{t_k}^{\xi,\Delta}\|^p\Big] \le K_T$. Introduce the auxiliary process

$$\begin{cases} z^{\xi,\Delta}(t_k) = \xi(t_k), & -N \le k \le -1, \\ z^{\xi,\Delta}(t_k) = y^{\xi,\Delta}(t_k) - \theta b(y_{t_k}^{\xi,\Delta})\Delta, & 0 \le k \le N^\Delta. \end{cases}$$

Then for $1 \le k \le N^\Delta$,

$$z^{\xi,\Delta}(t_k) = z^{\xi,\Delta}(t_{k-1}) + b(y_{t_{k-1}}^{\xi,\Delta})\Delta + \sigma(y_{t_{k-1}}^{\xi,\Delta})\delta W_{k-1}.$$

For $k \in \mathrm{N}$ and $\Delta \in (0, \frac{1}{2\theta L_1})$, it follows from Assumption 2.1 that

$$\begin{aligned}
&|z^{\xi,\Delta}(t_{k+1})|^2 \\
&= |z^{\xi,\Delta}(t_k)|^2 + |b(y_{t_k}^{\xi,\Delta})|^2\Delta^2 + |\sigma(y_{t_k}^{\xi,\Delta})\delta W_k|^2 + 2\langle y^{\xi,\Delta}(t_k) - \theta b(y_{t_k}^{\xi,\Delta})\Delta, b(y_{t_k}^{\xi,\Delta})\rangle\Delta \\
&\quad + 2\langle z^{\xi,\Delta}(t_k), \sigma(y_{t_k}^{\xi,\Delta})\delta W_k\rangle + 2\langle b(y_{t_k}^{\xi,\Delta}), \sigma(y_{t_k}^{\xi,\Delta})\delta W_k\rangle\Delta
\end{aligned}$$

$$
\begin{aligned}
&\le |z^{\xi,\Delta}(t_k)|^2 + (1-2\theta)|b(y^{\xi,\Delta}_{t_k})|^2\Delta^2 + |\sigma(y^{\xi,\Delta}_{t_k})\delta W_k|^2 + K\Delta\Big(1+|y^{\xi,\Delta}(t_k)|^2 \\
&\quad + \int_{-\tau}^{0} |y^{\xi,\Delta}_{t_k}(r)|^2 \mathrm{d}\nu_1(r)\Big) + \mathcal{M}_k \\
&\le |z^{\xi,\Delta}(t_k)|^2 + |\sigma(y^{\xi,\Delta}_{t_k})\delta W_k|^2 + K\Delta\Big(1+|y^{\xi,\Delta}(t_k)|^2 + \int_{-\tau}^{0} |y^{\xi,\Delta}_{t_k}(r)|^2 \mathrm{d}\nu_1(r)\Big) + \mathcal{M}_k,
\end{aligned}
$$

where $\{\mathcal{M}_k\}_{k\in\mathbb{N}}$ is a martingale defined by

$$
\mathcal{M}_k := (2-\frac{2}{\theta})\langle z^{\xi,\Delta}(t_k), \sigma(y^{\xi,\Delta}_{t_k})\delta W_k\rangle + \frac{2}{\theta}\langle y^{\xi,\Delta}(t_k), \sigma(y^{\xi,\Delta}_{t_k})\delta W_k\rangle.
$$

Then for any $p\in\mathbb{N}_+$,

$$
\begin{aligned}
|z^{\xi,\Delta}(t_{k+1})|^{2p} \le |z^{\xi,\Delta}(t_k)|^{2p} &+ \sum_{l=1}^{p} C_p^l |z^{\xi,\Delta}(t_k)|^{2(p-l)}\Big(|\sigma(y^{\xi,\Delta}_{t_k})\delta W_k|^2 \\
&+ K\Delta\big(1+|y^{\xi,\Delta}(t_k)|^2 + \int_{-\tau}^{0} |y^{\xi,\Delta}_{t_k}(r)|^2 \mathrm{d}\nu_1(r)\big) + \mathcal{M}_k\Big)^l,
\end{aligned}
$$

where C_p^l is the binomial coefficient. This implies

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{0\le(k+1)\Delta\le T} |z^{\xi,\Delta}(t_{k+1})|^{2p}\Big] \\
&= \mathbb{E}\Big[\sup_{\Delta\le(k+1)\Delta\le T} \sum_{i=0}^{k}\big(|z^{\xi,\Delta}(t_{i+1})|^{2p} - |z^{\xi,\Delta}(t_i)|^{2p}\big)\Big] + |z^{\xi,\Delta}(0)|^{2p} \\
&\le |z^{\xi,\Delta}(0)|^{2p} + \mathbb{E}\Big[\sup_{\Delta\le(k+1)\Delta\le T}\sum_{i=0}^{k}\sum_{l=1}^{p} C_p^l |z^{\xi,\Delta}(t_i)|^{2(p-l)}\Big(|\sigma(y^{\xi,\Delta}_{t_i})\delta W_i|^2 \\
&\quad + K\Delta\big(1+|y^{\xi,\Delta}(t_i)|^2 + \int_{-\tau}^{0} |y^{\xi,\Delta}_{t_i}(r)|^2 \mathrm{d}\nu_1(r)\big) + \mathcal{M}_i\Big)^l\Big] \\
&= |z^{\xi,\Delta}(0)|^{2p} + \sum_{l=1}^{p} C_p^l I_l, \qquad (46)
\end{aligned}
$$

where

$$
I_l := \mathbb{E}\Big[\sup_{\Delta\le(k+1)\Delta\le T}\sum_{i=0}^{k} |z^{\xi,\Delta}(t_i)|^{2(p-l)}\Big(|\sigma(y^{\xi,\Delta}_{t_i})\delta W_i|^2
$$

$$+ K\Delta\big(1+|y^{\xi,\Delta}(t_i)|^2+\int_{-\tau}^{0}|y^{\xi,\Delta}_{t_i}(r)|^2\mathrm{d}\nu_1(r)\big)+\mathcal{M}_i\Big)^l\Big].$$

For the term I_1, using the property of the conditional expectation and applying the Burkholder–Davis–Gundy inequality, we arrive at

$$\begin{aligned}I_1 \le{}& K\Delta\sum_{i=0}^{N^\Delta-1}\mathbb{E}\Big[|z^{\xi,\Delta}(t_i)|^{2(p-1)}\Big(|\sigma(y^{\xi,\Delta}_{t_i})|^2+1+|y^{\xi,\Delta}(t_i)|^2\\&+\int_{-\tau}^{0}|y^{\xi,\Delta}_{t_i}(r)|^2\mathrm{d}\nu_1(r)\Big)\Big]+K\mathbb{E}\Big[\Big(\sum_{i=0}^{N^\Delta-1}|z^{\xi,\Delta}(t_i)|^{2(2p-1)}|\sigma(y^{\xi,\Delta}_{t_i})|^2\Delta\Big)^{\frac{1}{2}}\Big]\\&+K\mathbb{E}\Big[\Big(\sum_{i=0}^{N^\Delta-1}|z^{\xi,\Delta}(t_i)|^{4(p-1)}|y^{\xi,\Delta}(t_i)|^2|\sigma(y^{\xi,\Delta}_{t_i})|^2\Delta\Big)^{\frac{1}{2}}\Big].\end{aligned}$$

By Assumptions 2.1 and 2.2, and the Young inequality, we derive that for $\varepsilon\in(0,1)$,

$$\begin{aligned}I_1 \le{}& K_T+K_T(\varepsilon)\Delta\sum_{i=0}^{N^\Delta-1}\mathbb{E}\Big[|z^{\xi,\Delta}(t_i)|^{2p}+|y^{\xi,\Delta}(t_i)|^{2p}+\int_{-\tau}^{0}|y^{\xi,\Delta}_{t_i}(r)|^{2p}\mathrm{d}\nu_1(r)\\&+\int_{-\tau}^{0}|y^{\xi,\Delta}_{t_i}(r)|^{2p}\mathrm{d}\nu_2(r)\Big]+\varepsilon\mathbb{E}\Big[\sup_{0\le k\Delta\le T}|z^{\xi,\Delta}(t_k)|^{2p}\Big].\end{aligned}\tag{47}$$

It follows from (6) and the convex property of $|\cdot|^{2p}$ that

$$\begin{aligned}\sum_{i=0}^{N^\Delta-1}\int_{-\tau}^{0}|y^{\xi,\Delta}_{t_i}(r)|^{2p}\mathrm{d}\nu_\ell(r)\le{}&\sum_{i=0}^{N^\Delta-1}\sum_{j=-N}^{-1}\Big(\int_{t_j}^{t_{j+1}}\frac{t_{j+1}-r}{\Delta}\mathrm{d}\nu_\ell(r)|y^{\xi,\Delta}(t_{i+j})|^{2p}\\&+\int_{t_j}^{t_{j+1}}\frac{r-t_j}{\Delta}\mathrm{d}\nu_\ell(r)|y^{\xi,\Delta}(t_{i+j+1})|^{2p}\Big)\le N\|\xi\|^{2p}+\sum_{i=0}^{N^\Delta-1}|y^{\xi,\Delta}(t_i)|^{2p}\end{aligned}\tag{48}$$

for $\ell=1,2$, which combining (47) leads to

$$\begin{aligned}I_1 \le{}& K_T(1+\|\xi\|^{2p})+K_T(\varepsilon)\Delta\sum_{i=0}^{N^\Delta-1}\mathbb{E}\Big[|z^{\xi,\Delta}(t_i)|^{2p}+|y^{\xi,\Delta}(t_i)|^{2p}\Big]\\&+\varepsilon\mathbb{E}\Big[\sup_{0\le k\Delta\le T}|z^{\xi,\Delta}(t_k)|^{2p}\Big].\end{aligned}\tag{49}$$

Similarly, we derive that for $l=2,\dots,p$,

$$I_l \le K_T(1+\|\xi\|^{2p}) + K_T(\varepsilon)\Delta \sum_{i=0}^{N^\Delta-1} \mathbb{E}\Big[|z^{\xi,\Delta}(t_i)|^{2p} + |y^{\xi,\Delta}(t_i)|^{2p}\Big] + \varepsilon\mathbb{E}\Big[\sup_{0\le k\Delta\le T} |z^{\xi,\Delta}(t_k)|^{2p}\Big]. \tag{50}$$

Inserting (49) and (50) into (46) yields

$$\begin{aligned}\mathbb{E}\Big[\sup_{0\le k\Delta\le T} |z^{\xi,\Delta}(t_k)|^{2p}\Big] &\le K_T(1+\|\xi\|^{2p}+|b(\xi)|^{2p}) \\ &+ K_T(\varepsilon)\Delta \sum_{i=0}^{N^\Delta-1} \mathbb{E}\Big[|z^{\xi,\Delta}(t_i)|^{2p} + |y^{\xi,\Delta}(t_i)|^{2p}\Big] \\ &+ \varepsilon(2^p-1)\Big[\sup_{0\le k\Delta\le T} |z^{\xi,\Delta}(t_k)|^{2p}\Big],\end{aligned} \tag{51}$$

where we used $\sum_{l=1}^p C_p^l = 2^p - 1$. It follows from $z^{\xi,\Delta}(t_k) = y^{\xi,\Delta}(t_k) - \theta b(y_{t_k}^{\xi,\Delta})\Delta$, Assumption 2.1, and $\langle a, b\rangle \le \frac{1}{2}(|a|^2+|b|^2)$, $a, b \in \mathbb{R}^d$ that

$$\begin{aligned}|y^{\xi,\Delta}(t_k)|^2 &\le |z^{\xi,\Delta}(t_k)|^2 + 2\big\langle y^{\xi,\Delta}(t_k), \theta b(y_{t_k}^{\xi,\Delta})\Delta\big\rangle \\ &\le |z^{\xi,\Delta}(t_k)|^2 + 2\theta\Delta L_1\Big(|y^{\xi,\Delta}(t_k)|^2 + \int_{-\tau}^0 |y_{t_k}^{\xi,\Delta}(r)|^2 \mathrm{d}\nu_1(r) + \langle y(t_k), b(0)\rangle\Big) \\ &\le |z^{\xi,\Delta}(t_k)|^2 + 2\theta\Delta L_1\Big(\frac{5}{2}\sup_{0\le k\Delta\le T} |y^{\xi,\Delta}(t_k)|^2 + \|\xi\|^2 + \frac{1}{2}|b(0)|^2\Big) \\ &\le \sup_{0\le k\Delta\le T} |z^{\xi,\Delta}(t_k)|^2 + 5\theta\Delta L_1 \sup_{0\le k\Delta\le T} |y^{\xi,\Delta}(t_k)|^2 + K\Delta.\end{aligned} \tag{52}$$

By $1 - 5\theta\Delta L_1 > \frac{1}{6}$ for $\Delta < \frac{1}{6\theta L_1}$, we obtain

$$\sup_{0\le k\Delta\le T} |y^{\xi,\Delta}(t_k)|^{2p} \le K\big(1 + \sup_{0\le k\Delta\le T} |z^{\xi,\Delta}(t_k)|^{2p}\big).$$

Combining (51) and letting $\varepsilon > 0$ be sufficiently small, we conclude

$$\begin{aligned}&\mathbb{E}\Big[\sup_{0\le k\Delta\le T}\Big(|z^{\xi,\Delta}(t_k)|^{2p} + |y^{\xi,\Delta}(t_k)|^{2p}\Big)\Big] \\ &\le K_T(1+\|\xi\|^{2p}+|b(\xi)|^{2p}) + K_T\Delta \sum_{i=0}^{N^\Delta-1} \mathbb{E}\Big[|z^{\xi,\Delta}(t_i)|^{2p} + |y^{\xi,\Delta}(t_i)|^{2p}\Big].\end{aligned}$$

Applying the discrete Grönwall inequality, and noticing the relation

$$\mathbb{E}\Big[\sup_{0\le t_k\le T} \|y_{t_k}^{\xi,\Delta}\|^p\Big] \le \mathbb{E}\Big[\sup_{0\le t_k\le T} |y^{\xi,\Delta}(t_k)|^p\Big] + \|\xi\|^p,$$

we finish the proof of $\mathbb{E}\big[\sup_{0\le t_k\le T}\|y_{t_k}^{\xi,\Delta}\|^p\big]\le K_T$.

Next, we prove that $\sup_{0\le v\le T}\mathbb{E}\big[\sup_{v\le t_k\le T}\|D_v y_{t_k}^{\xi,\Delta}\|^p\big]\le K_T$. Similarly, for any $k\in\mathbb{N}$ with $0\le k\Delta\le T$, $v\in[0,T]$, and $p\ge 2$, we have

$$|D_v z^{\xi,\Delta}(t_{k+1})|^{2p} \le |D_v z^{\xi,\Delta}(t_k)|^{2p}+\sum_{l=1}^{p}C_p^l I_{k,l}\le\sum_{i=0}^{k}\sum_{l=1}^{p}C_p^l I_{i,l},$$

where

$$\begin{aligned}I_{i,l}:=\ &|D_v z^{\xi,\Delta}(t_i)|^{2(p-l)}\Big(|\mathcal{D}\sigma(y_{t_i}^{\xi,\Delta})D_v y_{t_i}^{\xi,\Delta}\delta W_i|^2+|\sigma(y_{t_i}^{\xi,\Delta})\mathbf{1}_{[t_i,t_{i+1}]}(v)|^2\\&+K\Delta\big(|D_v y^{\xi,\Delta}(t_i)|^2+\int_{-\tau}^{0}|D_v y_{t_i}^{\xi,\Delta}(r)|^2\mathrm{d}\nu_1(r)\big)+\big\langle\frac{2(\theta-1)}{\theta}D_v z^{\xi,\Delta}(t_i)\\&+\frac{2}{\theta}D_v y^{\xi,\Delta}(t_i),\sigma(y_{t_i}^{\xi,\Delta})\mathbf{1}_{[t_i,t_{i+1}]}(v)\big\rangle+\tilde{\mathcal{M}}_i\Big)^l\end{aligned}$$

with

$$\tilde{\mathcal{M}}_k:=2\big\langle D_v z^{\xi,\Delta}(t_k)+\mathcal{D}b(y_{t_k}^{\xi,\Delta})D_v y_{t_k}^{\xi,\Delta}\Delta+\sigma(y_{t_k}^{\xi,\Delta})\mathbf{1}_{[t_k,t_{k+1}]}(v),\mathcal{D}\sigma(y_{t_k}^{\xi,\Delta})D_v y_{t_k}^{\xi,\Delta}\delta W_k\big\rangle.$$

The remaining proof is similar as before, and thus is omitted. □

Data availability

No data was used for the research described in the article.

References

[1] D. Nualart, The Malliavin Calculus and Related Topics, 2nd edition, Probability and Its Applications, Springer-Verlag, Berlin, 2006.

[2] A. Kohatsu-Higa, N. Privault, S.-J. Sheu (Eds.), Stochastic Analysis with Financial Applications, Progress in Probability, vol. 65, Birkhäuser/Springer Basel AG, Basel, 2011.

[3] D. Bell, S.-E. Mohammed, Smooth densities for degenerate stochastic delay equations with hereditary drift, Ann. Probab. 23 (4) (1995) 1875–1894, http://links.jstor.org/sici=0091-1798(199510)23:4<1875:SDFDSD>2.0.CO;2-J&origin=MSN.

[4] R. Chhaibi, I. Ekren, A Hörmander condition for delayed stochastic differential equations, Ann. H. Lebesgue 3 (2020) 1023–1048, https://doi.org/10.5802/ahl.53.

[5] M. Ferrante, C. Rovira, M. Sanz-Solé, Stochastic delay equations with hereditary drift: estimates of the density, J. Funct. Anal. 177 (1) (2000) 138–177, https://doi.org/10.1006/jfan.2000.3631.

[6] V. Bally, D. Talay, The law of the Euler scheme for stochastic differential equations. II. Convergence rate of the density, Monte Carlo Methods Appl. 2 (2) (1996) 93–128, https://doi.org/10.1515/mcma.1996.2.2.93.

[7] J. Guyon, Euler scheme and tempered distributions, Stochastic Process. Appl. 116 (6) (2006) 877–904, https://doi.org/10.1016/j.spa.2005.11.011.

[8] Y. Hu, S. Watanabe, Donsker's delta functions and approximation of heat kernels by the time discretization methods, J. Math. Kyoto Univ. 36 (3) (1996) 499–518, https://doi.org/10.1215/kjm/1250518506.

[9] A. Kohatsu-Higa, High order Itô–Taylor approximations to heat kernels, J. Math. Kyoto Univ. 37 (1) (1997) 129–150, https://doi.org/10.1215/kjm/1250518401.

[10] A. Kebaier, A. Kohatsu-Higa, An optimal control variance reduction method for density estimation, Stochastic Process. Appl. 118 (12) (2008) 2143–2180, https://doi.org/10.1016/j.spa.2008.01.006.

[11] J. Cui, J. Hong, D. Sheng, Density function of numerical solution of splitting AVF scheme for stochastic Langevin equation, Math. Comp. 91 (337) (2022) 2283–2333, https://doi.org/10.1090/mcom/3752.
[12] O. Bencheikh, B. Jourdain, Convergence in total variation of the Euler–Maruyama scheme applied to diffusion processes with measurable drift coefficient and additive noise, SIAM J. Numer. Anal. 60 (4) (2022) 1701–1740, https://doi.org/10.1137/20M1371774.
[13] P. Bras, G. Pagès, F. Panloup, Total variation distance between two diffusions in small time with unbounded drift: application to the Euler–Maruyama scheme, Electron. J. Probab. 27 (2022) 153, https://doi.org/10.1214/22-ejp881.
[14] L. Li, M. Wang, Y. Wang, Error estimates of the Euler's method for stochastic differential equations with multiplicative noise via relative entropy, arXiv:2409.04991, 2024, https://arxiv.org/abs/2409.04991.
[15] C. Chen, J. Cui, J. Hong, D. Sheng, Accelerated exponential Euler scheme for stochastic heat equation: convergence rate of the density, IMA J. Numer. Anal. 43 (2) (2023) 1181–1220, https://doi.org/10.1093/imanum/drac011.
[16] J. Hong, D. Jin, D. Sheng, Density convergence of a fully discrete finite difference method for stochastic Cahn–Hilliard equation, Math. Comp. 93 (349) (2024) 2215–2264, https://doi.org/10.1090/mcom/3928.
[17] C.-E. Bréhier, Total variation error bounds for the accelerated exponential Euler scheme approximation of parabolic semilinear SPDEs, SIAM J. Numer. Anal. 62 (3) (2024) 1171–1190, https://doi.org/10.1137/22M152596X.
[18] C.-E. Bréhier, Influence of the regularity of the test functions for weak convergence in numerical discretization of SPDEs, J. Complexity 56 (2020) 101424, https://doi.org/10.1016/j.jco.2019.101424.
[19] X. Li, X. Mao, G. Song, An explicit approximation for super-linear stochastic functional differential equations, Stochastic Process. Appl. 169 (2024) 104275, https://doi.org/10.1016/j.spa.2023.104275, https://doi-org.ezproxy.lb.polyu.edu.hk/10.1016/j.spa.2023.104275.
[20] J. Bao, G. Yin, C. Yuan, Ergodicity for functional stochastic differential equations and applications, Nonlinear Anal. 98 (2014) 66–82, https://doi.org/10.1016/j.na.2013.12.001, https://doi-org.ezproxy.lb.polyu.edu.hk/10.1016/j.na.2013.12.001.
[21] B. Shi, Y. Wang, X. Mao, F. Wu, Approximation of invariant measures of a class of backward Euler–Maruyama scheme for stochastic functional differential equations, J. Differential Equations 389 (2024) 415–456, https://doi.org/10.1016/j.jde.2024.01.025.
[22] Q. Luo, X. Mao, Y. Shen, Generalised theory on asymptotic stability and boundedness of stochastic functional differential equations, Automatica J. IFAC 47 (9) (2011) 2075–2081, https://doi.org/10.1016/j.automatica.2011.06.014.
[23] J. Cui, J. Hong, Absolute continuity and numerical approximation of stochastic Cahn–Hilliard equation with unbounded noise diffusion, J. Differential Equations 269 (11) (2020) 10143–10180, https://doi.org/10.1016/j.jde.2020.07.007.
[24] X. Mao, Numerical solutions of stochastic functional differential equations, LMS J. Comput. Math. 6 (2003) 141–161, https://doi.org/10.1112/S1461157000000425.
[25] E. Buckwar, T. Shardlow, Weak approximation of stochastic differential delay equations, IMA J. Numer. Anal. 25 (1) (2005) 57–86, https://doi.org/10.1093/imanum/drh012.
[26] E. Clément, A. Kohatsu-Higa, D. Lamberton, A duality approach for the weak approximation of stochastic differential equations, Ann. Appl. Probab. 16 (3) (2006) 1124–1154, https://doi.org/10.1214/105051606000000060.