



# Mean-square convergence of numerical methods for random ordinary differential equations

Peng Wang<sup>1</sup> · Yanzhao Cao<sup>2</sup> · Xiaoying Han<sup>2</sup>  · Peter Kloeden<sup>2</sup>

Received: 30 January 2020 / Accepted: 15 June 2020 / Published online: 29 July 2020  
© Springer Science+Business Media, LLC, part of Springer Nature 2020

## Abstract

The aim of this work is to analyze the mean-square convergence rates of numerical schemes for random ordinary differential equations (RODEs). First, a relation between the global and local mean-square convergence order of one-step explicit approximations is established. Then, the global mean-square convergence rates are investigated for RODE-Taylor schemes for general RODEs, Affine-RODE-Taylor schemes for RODEs with affine noise, and Itô-Taylor schemes for RODEs with Itô noise, respectively. The theoretical convergence results are demonstrated through numerical experiments.

**Keywords** Random ordinary differential equations · Mean-square convergence · Integral Taylor expansion · Affine noise · Brownian motions · Fractional Brownian motion

**Mathematics Subject Classification (2010)** 65C30 · 65L05 · 65L20 · 34F05

---

✉ Xiaoying Han  
xzh0003@auburn.edu

Peng Wang  
pwang@jlu.edu.cn

Yanzhao Cao  
yzc0009@auburn.edu

Peter Kloeden  
kloeden@na.uni-tuebingen.de

<sup>1</sup> Institute of Mathematics, Jilin University, Changchun, 130012, People's Republic of China

<sup>2</sup> Department of Mathematics and Statistics, Auburn University, Auburn, AL, 36849, USA

## 1 Introduction

Random ordinary differential equations (RODEs) are ordinary differential equations (ODEs) that include a stochastic process in their vector field. More precisely, let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space and let  $\eta : [0, T] \times \Omega \rightarrow \mathbb{R}^m$  be an  $\mathbb{R}^m$ -valued stochastic process with Hölder continuous sample paths. A RODE in  $\mathbb{R}^d$

$$\frac{dy}{dt} = f(y, \eta_t(\omega)), \quad y \in \mathbb{R}^d \quad (1)$$

is essentially a nonautonomous ODE:

$$\frac{dy}{dt} = f_\omega(y, t) := f(y, \eta_t(\omega)) \quad (2)$$

for almost all realizations  $\omega \in \Omega$ . Therefore, RODEs can be analyzed pathwise with deterministic calculus, but require further treatment beyond that of classical ordinary differential equation (ODE) theory. While the rules of deterministic calculus apply pathwise to RODEs for a fixed sample path, the vector field function in (2) is not smooth in its temporal variable. In fact, it is at most Hölder continuous in time when the driving stochastic process  $\eta_t$  is Hölder continuous and thus lacks sufficient smoothness needed to justify the usual Taylor expansions and the error analysis of traditional numerical methods for ODEs. Such methods can still be used (see, e.g., [5, 11]), but will attain at best a fractional low convergence order.

Similar to stochastic ordinary differential equations (SODEs), higher order numerical schemes for RODEs can be developed using Taylor-like expansions derived by iterated applications of the proper chain rule in the integral form (see e.g., [1–4, 8–12]). The mean-square convergence of numerical solutions of ODEs with a random initial value is discussed in [7]. The mean-square order for a class of backward stochastic differential equations was investigated in [6]. The pathwise convergence of numerical schemes for RODEs has received much attention (see, e.g., [3, 9, 10, 12]), as the numerical calculations of the approximating random variables are carried out path by path. However, there is no existing result on the mean-square convergence of numerical schemes for RODEs. In fact, the pathwise convergence rate does not imply the same mean-square convergence rate in general.

The goal of this work is to investigate the mean-square convergence rates of numerical schemes for various types of RODEs with different structures and noise. The mean-square order of convergence, also referred to as the strong order, is an important convergence index for numerical schemes of stochastic systems. A key criterion to estimate the mean-square convergence order of numerical approximations for SODEs was presented in (p. 12 [14]), where a fundamental convergence theorem was established for the mean-square convergence order of a method resting on properties of its one-step approximation. Here, we follow a similar idea to that used in (p. 12 [14]), and establish a general mean-square convergence theorem for numerical solutions of the RODE (1). In particular, we develop the relation between the local mean-square convergence order and the global mean-square convergence order of a general one-step approximation for the RODE (1) (see Theorem 1 below).

The paper is organized as follows. First, a generic theorem on the mean-square convergence order of one-step approximations for RODEs is presented Section 2.

The theorem is then applied to establish the global mean-square convergence order of various numerical schemes for RODEs. In particular, the global mean-square convergence orders of RODE-Taylor schemes for general RODEs are discussed in Section 3; and the global mean-square convergence orders of affine-RODE-Taylor schemes for RODEs with affine noise are discussed in Section 4. Numerical experiments are carried out in Section 6 to demonstrate the theoretical convergence results and some closing remarks are provided in Section 7.

## 2 The mean-square convergence of a generic numerical approximation for RODEs

When the vector field function  $f$  of the RODE (1) is continuous in both of its variables and the sample paths of the noise process  $\eta_t$  are also continuous, the vector field function  $f_\omega$  in the corresponding nonautonomous ODE (2) is continuous in both of its variables for each fixed  $\omega \in \Omega$ . Therefore, classical existence and uniqueness theorems for ODEs can be applied pathwise.

Throughout this paper, it is assumed that:

- (A1) The vector field function  $f : \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}^d$  is at least continuously differentiable in both of its variables and satisfies a one-sided Lipschitz condition, i.e., there exists a constant  $L \in \mathbb{R}$  such that:

$$\langle f(x, w) - f(y, w), x - y \rangle \leq L|x - y|^2, \quad \forall x, y \in \mathbb{R}^d, \quad w \in \mathbb{R}^m;$$

- (A2) There exist constants  $a, b \in \mathbb{R}$  and  $p \in \mathbb{N}$  such that:

$$|f(y, w)|^2 \leq a|w|^p + b|y|^2, \quad \forall y \in \mathbb{R}^d, \quad w \in \mathbb{R}^m;$$

- (A3) The noise process  $\eta_t$  is Hölder continuous with finite  $p$ th moments; i.e., there exists  $M_\eta = M_\eta(T) > 0$  such that  $\mathbb{E}[|\eta_t|^p] < M_\eta$  for all  $t \in [0, T]$ .

Here and in the rest of the paper,  $|\cdot|$  denotes the Euclidean norm of a vector or a matrix and  $\langle \cdot, \cdot \rangle$  denotes the dot product on  $\mathbb{R}^d$ .

Given any initial condition  $y(t_0) = y_0 \in \mathcal{L}^2(\Omega)$ , Assumption (A1) ensures the existence of a unique solution  $y(t; t_0, y_0)$  of (1) for all future time  $t \geq t_0$  (see, e.g., [4, 16]). More generally, for any  $s \geq t_0$  and  $\xi \in \mathbb{R}^d$ , let  $y(t; s, \xi)$  be the solution for (1) satisfying  $y(s) = \xi$ .

We consider a general numerical method to approximate the exact solution  $y(t; t_0, y_0)$  of (1). To start with, we assume a uniform partition on  $[t_0, T]$  with partition size  $h = \frac{T-t_0}{N}$ , and denote:

$$t_{n+1} = t_n + h \quad \text{or} \quad t_n = t_0 + nh \quad \text{for} \quad n = 0, 1, \dots, N.$$

Let  $\Phi^{(h)}$  be a one-step numerical scheme with step size  $h$  and for  $n = 0, 1, \dots, N - 1$  let  $y_{n+1}$  be the one-step approximation of  $y(t_n; t_0, y_0)$ :

$$y_{n+1} = \Phi^{(h)}(t_n, y_n). \tag{3}$$

Denote by  $\mathbb{E}$  the expectation with respect to the probability measure  $\mathbb{P}$ . Then, the local mean-square error (MSE) for the approximation (3) reads:

$$\mathcal{E}_{n+1}^L := \left( \mathbb{E}|y_{n+1} - y(t_{n+1}; t_n, y_n)|^2 \right)^{1/2}, \quad n = 0, 1, \dots, N - 1,$$

and the global MSE for the approximation (3) reads:

$$\mathcal{E}_{n+1}^G := \left( \mathbb{E}|y_{n+1} - y(t_{n+1}; t_0, y_0)|^2 \right)^{1/2}, \quad n = 0, 1, \dots, N - 1.$$

The main goal of this section is to establish the relation between  $\mathcal{E}_{n+1}^G$  and  $\mathcal{E}_{n+1}^L$ . In particular, we show that  $\mathcal{E}_{n+1}^G$  is one order (of the step size  $h$ ) lower than  $\mathcal{E}_{n+1}^L$ , presented in the following theorem.

**Theorem 1** *Assume that the assumptions (A1)–(A3) hold, and let  $h \in (0, 1]$ . If for some  $C > 0$  and  $\gamma > 0$ , the local MSE satisfies the estimation:*

$$\mathcal{E}_{n+1}^L \leq C \left( 1 + \mathbb{E}|y_n|^2 \right)^{1/2} h^{\gamma+1}, \quad n = 0, 1, \dots, N - 1, \tag{4}$$

*then for sufficiently small  $h$ , there exists positive constant  $K$  (independent of  $h$ ) such that the global MSE satisfies the estimation:*

$$\mathcal{E}_{n+1}^G \leq K \left( 1 + \mathbb{E}|y_0|^2 \right)^{1/2} h^\gamma, \quad n = 0, 1, \dots, N - 1.$$

*Remark 1* Theorem 1 states that the global mean-square convergence rate of any one-step approximation for RODEs is one order less than the local mean-square convergence rate. This result is consistent with the pathwise convergence rate for RODEs [9]. The local MSE condition (4) is parallel to those used in Milstein’s convergence theorem for SODEs [14, p. 12] and our result generalizes the deterministic case (ODEs) to the pathwise random case (RODEs).

The proof of Theorem 1 is based on a sequence of lemmas stated below. In particular, Lemma 1 provides a sequence of mean-square estimates of the true solution  $y(t_{n+1}; t_n, \xi)$  given  $y(t_n; t_0, y_0) = \xi$ , Lemma 2 presents a Gronwall type of iterative difference inequalities, and Lemma 3 provides a mean-square upper bound for the numerical solution  $y_n$ . In the sequel,  $K$  represents a generic constant that may depend on  $a, b, L$ , and  $M_\eta$ , but not on  $h$ , that may change from line to line.

**Lemma 1** *For any  $\xi \in L^2(\Omega)$ ,  $\zeta \in L^2(\Omega)$  and  $h < 1$ , there exists  $K = K(a, b, M_\eta) > 0$  such that the following inequalities hold for every  $n = 0, 1, \dots, N$ :*

- (i)  $\mathbb{E}|y(t_{n+1}; t_n, \xi)|^2 \leq K(1 + \mathbb{E}|\xi|^2)$ ;
- (ii)  $\mathbb{E}|y(t_{n+1}; t_n, \xi) - \xi|^2 \leq K(1 + \mathbb{E}|\xi|^2)h^2$ ;
- (iii)  $|\mathbb{E}\langle y(t_{n+1}; t_n, \xi), y(t_{n+1}; t_n, \xi) - \xi \rangle| \leq K(1 + \mathbb{E}|\xi|^2)h$ ;
- (iv)  $\mathbb{E}|y(t_{n+1}; t_n, \xi) - y(t_{n+1}; t_n, \zeta)|^2 \leq e^{2Lh} \mathbb{E}|\xi - \zeta|^2$ .

*Proof* (i) First, note that solutions to (1) satisfy the integral equation:

$$y(t) = y(t_n) + \int_{t_n}^t f(y(s), \eta_s) ds$$

on every subinterval  $[t_n, t_{n+1}]$ . Thus,

$$y(t_{n+1}; t_n, \xi) = \xi + \int_{t_n}^{t_{n+1}} f(y(s; t_n, \xi), \eta_s) ds. \tag{5}$$

Writing  $f(y(s; t_n, \xi), \eta_s)$  as  $f(y(s), \eta_s)$  in short, taking the mean-square of both sides of (5), and using the relationship  $(a + b)^2 \leq 2a^2 + 2b^2$ , we have:

$$\mathbb{E}|y(t_{n+1}; t_n, \xi)|^2 \leq 2\mathbb{E}|\xi|^2 + 2\mathbb{E} \left| \int_{t_n}^{t_{n+1}} f(y(s), \eta_s) ds \right|^2.$$

Then using the Cauchy-Schwarz inequality and assumptions (A2)–(A3), we obtain:

$$\begin{aligned} \mathbb{E}|y(t_{n+1}; t_n, \xi)|^2 &\leq 2\mathbb{E}|\xi|^2 + 2h \int_{t_n}^{t_{n+1}} \mathbb{E}|f(y(s), \eta_s)|^2 ds \\ &\leq 2\mathbb{E}|\xi|^2 + 2aM_\eta h^2 + 2bh \int_{t_n}^{t_{n+1}} \mathbb{E}|y(s)|^2 ds. \end{aligned}$$

Applying Gronwall’s inequality to the above inequality gives:

$$\mathbb{E}|y(t_{n+1}; t_n, \xi)|^2 \leq 2e^{2bh^2} (aM_\eta h^2 + \mathbb{E}|\xi|^2) \leq 2e^{2b} \max\{aM_\eta, 1\} \cdot (1 + \mathbb{E}|\xi|^2). \tag{6}$$

(ii) Using the (5), the Cauchy-Schwarz inequality and Assumptions (A2)–(A3), we have:

$$\begin{aligned} \mathbb{E}|y(t_{n+1}; t_n, \xi) - \xi|^2 &\leq h \int_{t_n}^{t_{n+1}} \mathbb{E}|f(y(s), \eta_s)|^2 ds \\ &\leq aM_\eta h^2 + bh \int_{t_n}^{t_{n+1}} (\mathbb{E}|y(s) - \xi|^2 + \mathbb{E}|\xi|^2) ds \\ &\leq (aM_\eta + \mathbb{E}|\xi|^2)h^2 + bh \int_{t_n}^{t_{n+1}} \mathbb{E}|y(s) - \xi|^2 ds. \end{aligned}$$

Applying Gronwall’s Lemma to the above inequality gives:

$$\mathbb{E}|y(t_{n+1}; t_n, \xi) - \xi|^2 \leq (aM_\eta + \mathbb{E}|\xi|^2)h^2 e^{bh^2} \leq \max\{aM_\eta, 1\}e^b \cdot (1 + \mathbb{E}|\xi|^2)h^2. \tag{7}$$

(iii) Using Jensen’s inequality, Hölder’s inequality, and inequalities (6) and (7), we obtain:

$$\begin{aligned} |\mathbb{E}\langle y(t_{n+1}; t_n, \xi), y(t_{n+1}; t_n, \xi) - \xi \rangle| &\leq \mathbb{E}|\langle y(t_{n+1}; t_n, \xi), y(t_{n+1}; t_n, \xi) - \xi \rangle| \\ &\leq (\mathbb{E}|y(t_{n+1}; t_n, \xi)|^2)^{1/2} \cdot (\mathbb{E}|y(t_{n+1}; t_n, \xi) - \xi|^2)^{1/2} \\ &\leq \sqrt{2}e^{3b/2} \max\{aM_\eta, 1\} \cdot (1 + \mathbb{E}|\xi|^2)h. \end{aligned}$$

(iv) For any  $\xi, \zeta \in L^2(\Omega)$ , let  $y(t; t_n, \xi)$  and  $y(t; t_n, \zeta)$  be the two solutions to (1) with initial conditions  $y(t_n) = \xi$  and  $y(t_n) = \zeta$ , respectively. Then, it follows directly from (1) that:

$$\frac{d}{dt} (y(t; t_n, \xi) - y(t; t_n, \zeta)) = f(y(t; t_n, \xi), \eta_t) - f(y(t; t_n, \zeta), \eta_t).$$

Taking the dot product of the above equation with  $y(t; t_n, \xi) - y(t; t_n, \zeta)$  gives:

$$\frac{d}{dt} |y(t; t_n, \xi) - y(t; t_n, \zeta)|^2 = 2(f(y(t; t_n, \xi), \eta_t) - f(y(t; t_n, \zeta), \eta_t), y(t; t_n, \xi) - y(t; t_n, \zeta))$$

Then, by Assumption (A1),

$$\frac{d}{dt} |y(t; t_n, \xi) - y(t; t_n, \zeta)|^2 \leq 2L |y(t; t_n, \xi) - y(t; t_n, \zeta)|^2. \tag{8}$$

Integrating (8) from  $t_n$  to  $t_{n+1}$ , we obtain:

$$|y(t_{n+1}; t_n, \xi) - y(t_{n+1}; t_n, \zeta)|^2 \leq e^{2Lh} |\xi - \zeta|^2,$$

which implies immediately

$$\mathbb{E} |y(t_{n+1}; t_n, \xi) - y(t_{n+1}; t_n, \zeta)|^2 \leq e^{2Lh} \mathbb{E} |\xi - \zeta|^2.$$

The proof is complete. □

**Lemma 2** [15, p. 7] *Suppose that for an arbitrary  $N \in \mathbb{N}$ , we have:*

$$u_{n+1} \leq (1 + \nu h)u_n + \delta h^r, \quad n = 0, 1, \dots, N$$

where  $\nu > 0, \delta \geq 0, r \geq 1, u_n \geq 0$  for all  $n = 0, 1, \dots, N$ . Then

$$u_{n+1} \leq e^{\nu N h} u_0 + \frac{\delta}{\nu} (e^{\nu N h} - 1) h^{r-1}.$$

**Lemma 3** *Let assumptions (A1)–(A3) and the local mean-square error estimate (4) hold. Then, there exists  $K = K(a, b, L, M_\eta, T) > 0$  such that:*

$$\mathbb{E}|y_n|^2 \leq K(1 + \mathbb{E}|y_0|^2), \quad n = 0, 1, \dots, N. \tag{9}$$

*Proof* First notice that  $\mathbb{E}|y_n|^2 < \infty$  for every  $n = 1, 2, \dots, N$ , provided  $\mathbb{E}|y_0|^2 < \infty$ . In fact, assuming that  $\mathbb{E}|y_k|^2 < \infty$  for every  $k = 0, \dots, n$ , then by the local MSE estimate (4) and Lemma 1-(i):

$$\begin{aligned} \mathbb{E}|y_{n+1}|^2 &\leq 2\mathbb{E}|y_{n+1} - y(t_{n+1}; t_n, y_n)|^2 + 2\mathbb{E}|y(t_{n+1}; t_n, y_n)|^2 \\ &\leq 2C^2 \left(1 + \mathbb{E}|y_n|^2\right) h^{2\gamma+2} + 2K(1 + \mathbb{E}|y_n|^2) < \infty. \end{aligned}$$

It then follows from induction that  $\mathbb{E}|y_n|^2 < \infty$  for all  $n = 0, \dots, N$ .

It is straightforward to check that the assertion (9) holds for  $n = 0$ . To estimate  $\mathbb{E}|y_n|^2$  for  $n \geq 1$ , we split  $y_n$  as:

$$y_n = (y_n - y(t_n; t_{n-1}, y_{n-1})) + (y(t_n; t_{n-1}, y_{n-1}) - y_{n-1}) + y_{n-1}. \tag{10}$$

For simplicity, here we write  $y(t_n; t_{n-1}, y_{n-1}) = y(t_n)$  when there is no confusion. Then, taking the dot product of (10) with  $y_n$  gives:

$$|y_n|^2 = |y_{n-1}|^2 + |y_n - y(t_n)|^2 + |y(t_n) - y_{n-1}|^2 + 2\langle y_n - y(t_n), y_{n-1} \rangle + 2\langle y_{n-1}, y(t_n) - y_{n-1} \rangle + 2\langle y_n - y(t_n), y(t_n) - y_{n-1} \rangle. \tag{11}$$

We next estimate the expectation of the terms on the right-hand side of the above equation.

By the local MSE estimation (4), one has:

$$\mathbb{E}|y_n - y(t_n)|^2 \leq C^2 \left(1 + \mathbb{E}|y_{n-1}|^2\right) h^{2\gamma+2}. \tag{12}$$

From Lemma 1-(ii), we have that:

$$\mathbb{E}|y(t_n) - y_{n-1}|^2 \leq K(1 + \mathbb{E}|y_{n-1}|^2)h^2. \tag{13}$$

The Hölder inequality and the local MSE estimation (4) lead to:

$$\mathbb{E}\langle y_n - y(t_n), y_{n-1} \rangle \leq \left(\mathbb{E}|y_n - y(t_n)|^2\right)^{1/2} \left(\mathbb{E}|y_{n-1}|^2\right)^{1/2} \leq K \left(1 + \mathbb{E}|y_{n-1}|^2\right) h^{\gamma+1}.$$

By Lemma 1-(iii), we have that:

$$\begin{aligned} \mathbb{E}\langle y_{n-1}, y(t_n) - y_{n-1} \rangle &= -\mathbb{E}|y_{n-1} - y(t_n)|^2 + \mathbb{E}\langle y(t_n), y(t_n) - y_{n-1} \rangle \\ &\leq K(1 + \mathbb{E}|y_{n-1}|^2)h. \end{aligned}$$

Finally, using the Hölder inequality and the estimations (12) and (13) above, we obtain:

$$\begin{aligned} \mathbb{E}\langle y_n - y(t_n), y(t_n) - y_{n-1} \rangle &\leq \left(\mathbb{E}|y_n - y(t_n)|^2\right)^{1/2} \left(\mathbb{E}|y(t_n) - y_{n-1}|^2\right)^{1/2} \\ &\leq K(1 + \mathbb{E}|y_{n-1}|^2)h^{\gamma+2}. \end{aligned} \tag{14}$$

Collecting estimations (11)–(14) and inserting into the expectation of (10), we have that:

$$\begin{aligned} \mathbb{E}|y_n|^2 &\leq \mathbb{E}|y_{n-1}|^2 + K(1 + \mathbb{E}|y_{n-1}|^2)(h + h^2 + h^{\gamma+1} + h^{2\gamma+2}) \\ &\leq \mathbb{E}|y_{n-1}|^2 + K(1 + \mathbb{E}|y_{n-1}|^2)h = (1 + Kh)\mathbb{E}|y_{n-1}|^2 + Kh, \end{aligned}$$

in which  $K$  is independent of  $h$ . Taking into account Lemma 2, we obtain:

$$\mathbb{E}|y_n|^2 \leq e^{KT} \mathbb{E}|y_0|^2 + e^{KT} - 1 \leq K(1 + \mathbb{E}|y_0|^2), \quad n = 0, \dots, N,$$

where  $K$  depends on  $T, a, b, M_\eta$ , but is independent of  $h$ . The proof is complete.  $\square$

We are now ready to prove Theorem 1 as follows.

*Proof of Theorem 1* First, we split the difference between  $y_{n+1} = y_{n+1}(t_n, y_n)$  and  $y(t_{n+1}; t_0, y_0)$  as:

$$y_{n+1} - y(t_{n+1}; t_0, y_0) = y_{n+1} - y(t_{n+1}; t_n, y_n) + y(t_{n+1}; t_n, y_n) - y(t_{n+1}; t_0, y_0). \tag{15}$$

Noticing that  $y(t_{n+1}; t_0, y_0) = y(t_{n+1}; t_n, y(t_n; t_0, y_0))$ , (15) then becomes:

$$y_{n+1} - y(t_{n+1}; t_0, y_0) = y_{n+1} - y(t_{n+1}; t_n, y_n) + y(t_{n+1}; t_n, y_n) - y(t_{n+1}; t_n, y(t_n; t_0, y_0)).$$

Taking the dot product of the above equation with  $y_{n+1} - y(t_{n+1}; t_0, y_0)$  and then taking the expectation of the resulting relation gives:

$$\begin{aligned} (\mathcal{E}_{n+1}^G)^2 &= \mathbb{E} \left[ |y_{n+1} - y(t_{n+1}; t_n, y_n)|^2 + |y(t_{n+1}; t_n, y_n) - y(t_{n+1}; t_n, y(t_n; t_0, y_0))|^2 \right. \\ &\quad \left. + 2(y_{n+1} - y(t_{n+1}; t_n, y_n), y(t_{n+1}; t_n, y_n) - y(t_{n+1}; t_n, y(t_n; t_0, y_0))) \right] \\ &\leq \mathbb{E}|y_{n+1} - y(t_{n+1}; t_n, y_n)|^2 + \mathbb{E}|y(t_{n+1}; t_n, y_n) - y(t_{n+1}; t_n, y(t_n; t_0, y_0))|^2 \\ &\quad + h^{-1} \mathbb{E}|y_{n+1} - y(t_{n+1}; t_n, y_n)|^2 + h \mathbb{E}|y(t_{n+1}; t_n, y_n) - y(t_{n+1}; t_n, y(t_n; t_0, y_0))|^2. \end{aligned} \tag{16}$$

By the local MSE assumption (4) and Lemma 3:

$$\mathbb{E}|y_{n+1} - y(t_{n+1}; t_n, y_n)|^2 \leq K \left( 1 + \mathbb{E}|y_n|^2 \right) h^{2\gamma+2} \leq K \left( 1 + \mathbb{E}|y_0|^2 \right) h^{2\gamma+2}. \tag{17}$$

By Lemma 1-(iv),

$$\mathbb{E}|y(t_{n+1}; t_n, y_n) - y(t_{n+1}; t_n, y(t_n; t_0, y_0))|^2 \leq e^{2Lh} \mathbb{E}|y_n - y(t_n; t_0, y_0)|^2 = e^{2Lh} (\mathcal{E}_n^G)^2. \tag{18}$$

Inserting inequalities (17)–(18) into (16), we obtain:

$$(\mathcal{E}_{n+1}^G)^2 \leq K \left( 1 + \mathbb{E}|y_0|^2 \right) \left( h^{2\gamma+2} + h^{2\gamma+1} \right) + e^{2Lh} (1 + h) (\mathcal{E}_n^G)^2.$$

For sufficiently small  $h$ , there exists a positive constant  $\nu > 0$  such that

$$e^{2Lh} (1 + h) < 1 + \nu h,$$

we then obtain

$$(\mathcal{E}_{n+1}^G)^2 \leq [1 + \nu h] (\mathcal{E}_n^G)^2 + K (1 + \mathbb{E}|y_0|^2) h^{2\gamma+1}.$$

It then follows directly from Lemma 2 that

$$(\mathcal{E}_{n+1}^G)^2 \leq K (1 + \mathbb{E}|y_0|^2) \frac{e^{(T-t_0)} - 1}{\nu} h^{2\gamma},$$

which implies that

$$\mathcal{E}_{n+1}^G \leq K (1 + \mathbb{E}|y_0|^2)^{1/2} h^\gamma,$$

where  $K$  is constant depending on  $t_0, T, L$  and the distribution of  $\eta_t$ , but not  $h$ . The proof is complete. □

### 2.1 Applications

As an introductory example, we first consider the explicit Euler scheme for the RODE (1):



$$y_{n+1} = y_n + hf(y_n, \eta_{t_n}), \quad n = 0, 1, \dots, N - 1, \tag{19}$$

where  $\eta_t$  is assumed to be a fractional Brownian motion (fBm).

**Theorem 2** *Let Assumptions (A2)–(A3) hold and in addition replace Assumption (A1) by*

(A1') *the function  $f$  is globally Lipschitz in both of its variables.*

*Then, the mean-square convergence order of the Euler scheme (19) for (1) is  $H$ , where  $H$  is the Hurst parameter of the fBm  $\eta_t$ .*

*Proof* We first estimate the local mean-square error of the scheme (19). Using the integral representation of the solution (5), we have:

$$\begin{aligned} (\mathcal{E}_{n+1}^L)^2 &= \mathbb{E} \left| y_n + \int_{t_n}^{t_{n+1}} f(y(s; t_n, y_n), \eta_s) ds - y_n - f(y_n, \eta_{t_n})h \right|^2 \\ &= \mathbb{E} \left| \int_{t_n}^{t_{n+1}} (f(y(s; t_n, y_n), \eta_s) - f(y_n, \eta_{t_n})) ds \right|^2. \end{aligned}$$

Then by Cauchy-Schwarz inequality and Fubini’s Theorem, we have:

$$(\mathcal{E}_{n+1}^L)^2 \leq h \int_{t_n}^{t_{n+1}} \mathbb{E} |f(y(s; t_n, y_n), \eta_s) - f(y_n, \eta_{t_n})|^2 ds. \tag{20}$$

Write  $y(s; t_n, y_n)$  as  $y(s)$  in short when there is no confusion. To estimate the expectation inside the integral in the above inequality, write:

$$f(y(s), \eta_s) - f(y_n, \eta_{t_n}) = f(y(s), \eta_s) - f(y_n, \eta_s) + f(y_n, \eta_s) - f(y_n, \eta_{t_n}).$$

Then, by Hölder’s inequality and (20):

$$(\mathcal{E}_{n+1}^L)^2 \leq 2h \int_{t_n}^{t_{n+1}} \left( \mathbb{E} |f(y(s), \eta_s) - f(y_n, \eta_s)|^2 + \mathbb{E} |f(y_n, \eta_s) - f(y_n, \eta_{t_n})|^2 \right) ds. \tag{21}$$

Due to Assumption (A1'), there exist positive constants  $M_f$  such that:

$$|f(y(s), \eta_s) - f(y_n, \eta_s)| \leq M_f |y(s) - y_n|, \tag{22}$$

$$|f(y_n, \eta_s) - f(y_n, \eta_{t_n})| \leq M_f |\eta_s - \eta_{t_n}|. \tag{23}$$

Following (22), we have:

$$\mathbb{E} |f(y(s), \eta_s) - f(y_n, \eta_s)|^2 \leq M_f^2 \mathbb{E} |y(s) - y_n|^2. \tag{24}$$

Furthermore, by the interpolation approximation  $y_s = y_n + (s - t_n)f(y_n, \eta_{t_n})$  and Assumptions (A2) and (A3) we obtain:

$$\begin{aligned} \mathbb{E} |y(s) - y_n|^2 &\leq 2 \left( \mathbb{E} |y(s) - y_s|^2 + \mathbb{E} |y_s - y_n|^2 \right) \leq 2 \left( (\mathcal{E}_s^L)^2 + h^2 \mathbb{E} |f(y_n, \eta_{t_n})|^2 \right) \\ &\leq 2 \left( (\mathcal{E}_s^L)^2 + h^2 \left( a \mathbb{E} |\eta_{t_n}|^p + b \mathbb{E} |y_n|^2 \right) \right) \leq 2(\mathcal{E}_s^L)^2 + Kh^2(1 + \mathbb{E} |y_n|^2), \end{aligned} \tag{25}$$

where  $\mathcal{E}_s^L = (\mathbb{E}|y(s; t_n, y_n) - y_n|^2)^{1/2}$  for  $s \in (t_n, t_{n+1})$ , and  $K$  is independent of  $h$ . Then, it follows from (23) that:

$$\mathbb{E} |f(y_n, \eta_s) - f(y_n, \eta_{t_n})|^2 \leq M_f^2 \mathbb{E} |\eta_s - \eta_{t_n}|^2. \tag{26}$$

Collecting estimates (24)–(26) and inserting into (21) give:

$$\begin{aligned} (\mathcal{E}_{n+1}^L)^2 &\leq 2hM_f^2 \int_{t_n}^{t_{n+1}} \left( 2(\mathcal{E}_s^L)^2 + Kh^2(1 + \mathbb{E}|y_n|^2) + \mathbb{E}|\eta_s - \eta_{t_n}|^2 \right) ds \\ &= KM_f^2 h^4 (1 + \mathbb{E}|y_n|^2) + 4hM_f^2 \int_{t_n}^{t_{n+1}} (\mathcal{E}_s^L)^2 ds + 2hM_f^2 \int_{t_n}^{t_{n+1}} \mathbb{E}|\eta_s - \eta_{t_n}|^2 ds. \end{aligned} \tag{27}$$

Since  $\eta_t$  is a fractional Brownian motion with Hurst parameter  $H \in (0, 1]$ , then  $\mathbb{E}|\eta_s - \eta_{t_n}|^2 = K|s - t_n|^{2H}$  and thus:

$$\int_{t_n}^{t_{n+1}} \mathbb{E}|\eta_s - \eta_{t_n}|^2 ds = K \int_{t_n}^{t_{n+1}} (s - t_n)^{2H} ds = \frac{1}{2H + 1} h^{2H+1}.$$

Also, because  $H \leq 1$ , the inequality (27) implies:

$$\begin{aligned} (\mathcal{E}_{n+1}^L)^2 &\leq 4hM_f^2 \int_{t_n}^{t_{n+1}} (\mathcal{E}_s^L)^2 ds + KM_f^2 h^4 (1 + \mathbb{E}|y_n|^2) + \frac{2M_f^2}{2H + 1} h^{2H+2} \\ &\leq 4hM_f^2 \int_{t_n}^{t_{n+1}} (\mathcal{E}_s^L)^2 ds + Kh^{2H+2} (1 + \mathbb{E}|y_n|^2), \end{aligned}$$

and by Gronwall’s inequality again, we obtain:

$$(\mathcal{E}_{n+1}^L)^2 \leq K (1 + \mathbb{E}|y_n|^2) h^{2H+2}.$$

Theorem 1 then implies that  $\mathcal{E}_{n+1}^G \sim \mathcal{O}(h^\gamma)$  where  $\gamma = H$ . □

*Remark 2* The assumption (A1’) is imposed for convenience for exposition and may be stronger than needed. In fact, our numerical experiments (see, e.g., Example 15) indicate that the results of the theorem are still valid under conditions weaker than (A1’).

*Remark 3* The convergence order established in Theorem 2 is the *slowest* possible rate for the whole class of equations. One can always obtain better convergence with some special equations such as linear RODEs or RODEs with simple noise processes. For example, when  $\eta_t$  is a Brownian motion, the Euler scheme (19) can achieve a convergence order of 1, or even higher for special function  $f$ . This does not contradict with the result proved in Theorem 2.

In the following sections, we apply Theorem 1 to estimate the global mean-square convergence rate for various schemes for a class of general RODEs, and for the special classes of Affine-RODEs with affine noises and Itô-RODEs with Itô diffusion noise.

### 3 RODE-Taylor schemes for general RODEs

In this section, we first introduce the general framework of RODE-Taylor schemes for the RODE (1), based on RODE-Taylor approximations, then investigate the order of mean-square convergence for RODE-Taylor schemes. Here, we adopt the same set of notations as those in [9] and [10] with slight modifications. For the reader’s convenience, we summarize the necessary preliminaries below.

#### 3.1 Preliminaries

For any nonempty set  $A$  and any  $l, k \in \mathbb{N}$ , denote by  $A^{l \times k}$  the set of all  $l \times k$  matrices with entries in  $A$ . In particular, for  $i \geq 1$  denote by  $\mathbb{N}_0^{m \times i}$  the set of all  $m \times i$  matrices with nonnegative integers. In addition, write  $\mathbb{N}_0^{m \times 0} = \mathbb{N}_0^0 := \{\emptyset\}$ . Consider an element  $\mathbf{a} = (\mathbf{a}_1, \dots, \mathbf{a}_i) \in \mathbb{N}_0^{m \times i}$  with  $\mathbf{a}_j = (\mathbf{a}_{j,1}, \dots, \mathbf{a}_{j,m})^\top \in \mathbb{N}_0^m$  for  $j = 1, \dots, i$  represented as:

$$\mathbf{a} = (\mathbf{a}_1, \dots, \mathbf{a}_i) = \begin{pmatrix} \mathbf{a}_{1,1} & \dots & \mathbf{a}_{i,1} \\ \vdots & & \vdots \\ \mathbf{a}_{1,m} & \dots & \mathbf{a}_{i,m} \end{pmatrix}.$$

For such an  $\mathbf{a} \in \mathbb{N}_0^{m \times i}$ , define  $\iota(\mathbf{a}) := i$ .

Denote the set of all the above matrix-valued multi-indices by  $\mathfrak{A}_m := \cup_{i=0}^\infty \mathbb{N}_0^{m \times i}$ . Then, for any  $\mathbf{a} \in \mathfrak{A}_m \setminus \{\emptyset\}$  with  $\iota(\mathbf{a}) = i \geq 1$ , we define:

$$\begin{aligned} \mathbf{a}! &:= \mathbf{a}_{1,1}! \cdots \mathbf{a}_{1,m}! \cdots \mathbf{a}_{i,1}! \cdots \mathbf{a}_{i,m}! \\ |\mathbf{a}| &:= \mathbf{a}_{1,1} + \cdots + \mathbf{a}_{1,m} + \cdots + \mathbf{a}_{i,1} + \cdots + \mathbf{a}_{i,m}. \end{aligned}$$

In particular, for  $\mathbf{a} = \emptyset$ , define:

$$\iota(\emptyset) := 0, \quad \emptyset! := 1, \quad |\emptyset| := 0.$$

Given such a matrix valued index  $\mathbf{a}$  with  $\iota(\mathbf{a}) = i \geq 1$  and an  $m$ -dimensional stochastic process  $\eta_t = (\eta_t^1, \dots, \eta_t^m)$  with mutually independent component processes, define:

$$\Delta \eta_{t,s}^{\mathbf{a}k} = \prod_{j=1}^m (\Delta \eta_{t,s})^{\mathbf{a}_{k,j}} = \prod_{j=1}^m (\eta_s^j - \eta_t^j)^{\mathbf{a}_{k,j}}, \quad k = 1, \dots, i$$

and the iterated integrals:

$$I_{t,s}^{\mathbf{a}} := \int_t^s \int_t^{s_1} \cdots \int_t^{s_{i-1}} (\Delta \eta_{t,s_1}^{\mathbf{a}_1} \cdots \Delta \eta_{t,s_i}^{\mathbf{a}_i}) ds_i \cdots ds_1, \quad I_{t,s}^\emptyset := 1. \quad (28)$$

Then,  $I_{t,s}^{\mathbf{a}}$  is a random variable for every  $t, s \in \mathbb{R}$ .

To derive an integral equation expansion of the solutions to the RODE (1), we introduce the vector spaces over  $\mathbb{R}$  of all smooth functions from  $\mathbb{R}^d \times \mathbb{R}^{m \times i}$  to  $\mathbb{R}^d$ :

$$\mathcal{H}_i := C^\infty(\mathbb{R}^d \times \mathbb{R}^{m \times i}, \mathbb{R}^d), \quad \mathcal{H}_0 := C^\infty(\mathbb{R}^d, \mathbb{R}^d).$$

Given a matrix multi-index  $\alpha \in \mathfrak{A}_m$  with  $\iota(\alpha) = i \geq 1$  and a function  $\phi \in \mathcal{H}_i$  of the element  $(y, w) \in \mathbb{R}^d \times \mathbb{R}^{m \times i}$  with the components:

$$y = \begin{pmatrix} y^1 \\ \vdots \\ y^d \end{pmatrix} \quad \text{and} \quad w = (w_1, \dots, w_i) = \begin{pmatrix} w_{1,1} & \cdots & w_{i,1} \\ \vdots & & \vdots \\ w_{1,m} & \cdots & w_{i,m} \end{pmatrix},$$

define the linear differential operators  $\mathfrak{D}_i : \mathcal{H}_i \rightarrow \mathcal{H}_{i+1}$  by

$$\mathfrak{D}_i \phi(y, w_1, \dots, w_{i+1}) := \partial_y \phi(y, w_1, \dots, w_i) \cdot f(y, w_{i+1}),$$

and the  $\alpha$ -derivative of  $\phi$  with respect to  $w$  as

$$\partial^\alpha \phi := \partial_{w_{1,1}}^{\alpha_{1,1}} \cdots \partial_{w_{1,m}}^{\alpha_{1,m}} \cdots \partial_{w_{i,1}}^{\alpha_{i,1}} \cdots \partial_{w_{i,m}}^{\alpha_{i,m}} \phi. \tag{29}$$

Let  $\mathcal{H} := \bigcup_{i=0}^\infty \mathcal{H}_i$ , then the differential operator  $\mathfrak{D}_i$  can be generalized to  $\mathfrak{D} : \mathcal{H} \rightarrow \mathcal{H}$  by  $\mathfrak{D}\phi := \mathfrak{D}_i \phi$  if  $\phi \in \mathcal{H}_i$  for some  $i \in \mathbb{N}_0$ . In addition, define the iterated differential operator defined by:

$$\mathfrak{D}^j := \underbrace{\mathfrak{D} \circ \mathfrak{D} \circ \cdots \circ \mathfrak{D}}_{j \text{ times}}. \tag{30}$$

For each  $p \in \mathbb{N}$  and  $t_0 \leq t \leq s \leq T$ , the integral equation expansion of the solution to the RODE (1) reads [9, 10]:

$$\begin{aligned} y(s) = y(t) &+ \sum_{j=1}^p \int_t^s \int_t^{s_1} \cdots \int_t^{s_{j-1}} \mathfrak{D}^j \chi(y(t), \eta_{s_1}, \dots, \eta_{s_j}) ds_j \cdots ds_1 \\ &+ \int_t^s \int_t^{s_1} \cdots \int_t^{s_p} \mathfrak{D}^{p+1} \chi(y(s_{p+1}), \eta_{s_1}, \dots, \eta_{s_{p+1}}) ds_{p+1} \cdots ds_1, \end{aligned} \tag{31}$$

where  $\chi : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is the identity function defined by

$$\chi(y) = y, \quad \forall y \in \mathbb{R}^d.$$

### 3.2 $\gamma$ -RODE-Taylor schemes

Assume that the driving stochastic process  $\eta_t = (\eta_t^1, \dots, \eta_t^m)$  in the RODE (1) has Hölder continuous sample paths. More specifically, assume that:

(A4) For each  $j = 1, \dots, m$ , there exists  $b_j \in (0, 1]$  and  $M_q > 0$  such that the driving stochastic process  $\eta_t^j$  satisfies:

$$\mathbb{E} \left( \eta_t^j - \eta_s^j \right)^{2q} \leq M_q |t - s|^{2qb_j}, \quad q = 1, 2, \dots.$$

*Remark 4* Due to the Kolmogorov Continuity Theorem, Assumption (A4) implies that  $\eta_t^j$  has  $\beta$ -Hölder continuous paths for  $\beta \in (0, b_j)$ . In particular,  $b_j = \frac{1}{2}$  when  $\eta_t^j$  is a Brownian motion and  $b_j = H$  if  $\eta_t^j$  is a fractional Brownian motion with Hurst parameter  $H$ .

Define  $\mathbf{b} := (\mathbf{b}_1, \dots, \mathbf{b}_m)^\top \in (0, 1]^m$ . Then, for any  $\mathbf{a} \in \mathfrak{A}_m$  with  $\iota(\mathbf{a}) \geq 1$ , we have:

$$\mathbf{a}^\top \mathbf{b} = \begin{pmatrix} \mathbf{a}_{1,1} \mathbf{b}_1 + \dots + \mathbf{a}_{1,m} \mathbf{b}_m \\ \vdots \\ \mathbf{a}_{i,1} \mathbf{b}_1 + \dots + \mathbf{a}_{i,m} \mathbf{b}_m \end{pmatrix} \in \mathbb{R}_{\geq 0}^m, \quad \mathbb{R}_{\geq 0}^m := \{(x_1, \dots, x_m) \in \mathbb{R}^m : x_j \geq 0, j = 1, \dots, m\},$$

and

$$\|\mathbf{a}^\top \mathbf{b}\|_1 := \mathbf{b}_1(\mathbf{a}_{1,1} + \dots + \mathbf{a}_{i,1}) + \dots + \mathbf{b}_m(\mathbf{a}_{1,m} + \dots + \mathbf{a}_{i,m}).$$

In particular, for  $\mathbf{a} = \emptyset$ , define  $\|\emptyset^\top \mathbf{b}\|_1 := 0$ .

Consider specific subsets of matrix multi-indices of the form:

$$\mathfrak{A}_m^\gamma := \left\{ \mathbf{a} \in \mathfrak{A}_m : \iota(\mathbf{a}) + \|\mathbf{a}^\top \mathbf{b}\|_1 < \gamma + 1 \right\}, \quad \gamma > 0.$$

Use the abbreviation:

$$f_{\mathbf{a}} = \frac{1}{\mathbf{a}!} \left( \partial^{\mathbf{a}} \mathfrak{D}^{\iota(\mathbf{a})} \chi \right) \in \mathcal{H}_{\iota(\mathbf{a})}, \tag{32}$$

where  $\partial^{\mathbf{a}}$  and  $\mathfrak{D}^{\iota(\mathbf{a})}$  are defined according to (29) and (30), respectively.

Approximating  $\mathfrak{D}^j \chi : \mathbb{R}^d \times \mathbb{R}^{m \times i} \rightarrow \mathbb{R}^d$  in the integral equation expansion (31) by a Taylor expansion in its first  $m \times i$  variables, the *RODE-Taylor schemes* for the RODE (1) can be constructed to be:

$$y_{n+1} = \Phi_\gamma^{(h)}(t_n, y_n) = \sum_{\mathbf{a} \in \mathfrak{A}_m^\gamma} f_{\mathbf{a}}(y_n, \eta_{t_n}^{\times \iota(\mathbf{a})}) I_{t_n, t_{n+1}}^{\mathbf{a}}, \tag{33}$$

where  $I_{t_n, t_{n+1}}^{\mathbf{a}}$  is defined by (28),  $f_{\mathbf{a}}$  is defined by (32), and  $\eta_{t_n}^{\times \iota(\mathbf{a})}$  is the  $m \times \iota(\mathbf{a})$  matrix:

$$\eta_{t_n}^{\times \iota(\mathbf{a})} = \begin{pmatrix} \eta_{t_n}^1 & \dots & \eta_{t_n}^1 \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ \eta_{t_n}^m & \dots & \eta_{t_n}^m \end{pmatrix}.$$

### 3.3 Convergence analysis

In this subsection, we analyze the mean-square convergence rate of the general RODE-Taylor scheme (33). Here, it is assumed that all components of the driving stochastic process  $\eta_t = (\eta_t^1, \dots, \eta_t^m)$  are mutually independent.

**Theorem 3** *Assume that assumptions (A1)–(A4) hold. In addition, assume that:*

- (A5) *f is infinitely often continuously differentiable in its variables and all partial derivatives of f are bounded on  $[t_0, T]$ .*

*Then, given any  $\gamma > 0$ , the global mean-square error of the RODE-Taylor scheme (33) satisfies:*

$$\mathcal{E}_{n+1}^G = \left( \mathbb{E} |y(t_{n+1}; t_0, y_0) - y_{n+1}|^2 \right)^{1/2} \leq Kh^\gamma,$$

where  $K$  is a constant depending on  $t_0, T, \gamma, \mathbf{b}, M_\eta$ , but independent of  $h$ .

*Proof* The proof is based on the proof of Theorem 5.1 [10], but in the mean-square sense with different details. For completeness, we provide all necessary details below.

Let  $r = \lceil \gamma \rceil$  be the smallest positive number that is larger than or equal to  $\gamma$ , and write  $y(t_{n+1}; t_n, y_n)$  as  $y(t_{n+1})$  when there is no confusion. Then, the integral expansion (31) gives:

$$\begin{aligned}
 y(t_{n+1}) = & y_n + \sum_{i=1}^r \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} \cdots \int_{t_n}^{s_{i-1}} \mathfrak{D}^i \chi(y_n, \eta_{s_1}, \dots, \eta_{s_i}) ds_i \cdots ds_1 \\
 & + \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} \cdots \int_{t_n}^{s_r} \mathfrak{D}^{r+1} \chi(y(t_{n+1}), \eta_{s_1}, \dots, \eta_{s_{r+1}}) ds_{r+1} \cdots ds_1.
 \end{aligned}
 \tag{34}$$

Given  $\omega \in \Omega$  define the mapping  $G : [0, 1] \rightarrow \mathbb{R}^d$  by:

$$G(\lambda) = \mathfrak{D}^i \chi(y_n(\omega), \eta_{t_n}(\omega) + \lambda \Delta \eta_{t_n, s_1}(\omega), \dots, \eta_{t_n}(\omega) + \lambda \Delta \eta_{t_n, s_i}(\omega)).$$

Let  $\underline{b} := \min\{b_1, \dots, b_m\}$ , and set  $\kappa_i := \left\lceil \frac{\gamma-i+1}{\underline{b}} \right\rceil - 1$ . Then, applying a Taylor expansion of order  $\kappa_i$  to the function  $G$  gives:

$$G(1) = G(0) + G'(0) + \cdots + \frac{G^{(\kappa_i)}(0)}{\kappa_i!} + \int_0^1 G^{(\kappa_i+1)}(\lambda) \frac{(1-\lambda)^{\kappa_i}}{\kappa_i!} d\lambda.$$

It follows directly from the above equality that:

$$\begin{aligned}
 \mathfrak{D}^i \chi(y_n, \eta_{s_1}, \dots, \eta_{s_i}) = & \sum_{\substack{|\mathbf{a}| \leq \kappa_i \\ \mathbf{a} \in \mathbb{N}_0^{m \times i}}} f_{\mathbf{a}}(y_n, \eta_{t_n}^{\times i}) (\Delta \eta_{t_n, s_1})^{\mathbf{a}_1} \cdots (\Delta \eta_{t_n, s_i})^{\mathbf{a}_i} \\
 & + \sum_{\substack{|\mathbf{a}| = \kappa_i + 1 \\ \mathbf{a} \in \mathbb{N}_0^{m \times i}}} ((\kappa_i + 1) (\Delta \eta_{t_n, s_1})^{\mathbf{a}_1} \cdots (\Delta \eta_{t_n, s_i})^{\mathbf{a}_i} \\
 & \cdot \int_0^1 f_{\mathbf{a}}(y_n, \eta_{t_n} + \lambda \Delta \eta_{t_n, s_1}, \dots, \eta_{t_n} + \lambda \Delta \eta_{t_n, s_i}) (1-\lambda)^{\kappa_i} d\lambda).
 \end{aligned}
 \tag{35}$$

Inserting (35) into (34) results in

$$y(t_{n+1}) = y_n + \sum_{i=1}^r \sum_{\substack{|\mathbf{a}| \leq \kappa_i \\ \mathbf{a} \in \mathbb{N}_0^{m \times i}}} f_{\mathbf{a}}(y_n, \eta_{t_n}^{\times i}) I_{t_n, t_{n+1}}^{\mathbf{a}} + R_1 + R_2,
 \tag{36}$$

where

$$\begin{aligned}
 R_1 &= \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} \cdots \int_{t_n}^{s_r} \mathfrak{D}^{r+1} \chi(y(s_{r+1}), \eta_{s_1}, \dots, \eta_{s_{r+1}}) ds_{r+1} \cdots ds_1, \\
 R_2 &= \sum_{i=1}^r \sum_{\substack{|\mathbf{a}|=\kappa_i+1 \\ \mathbf{a} \in \mathbb{N}_0^{m \times i}}} \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} \cdots \int_{t_n}^{s_{i-1}} \Delta \eta_{t_n, s_1}^{\mathbf{a}_1} \cdots \Delta \eta_{t_n, s_i}^{\mathbf{a}_i} \cdot (\kappa_i + 1) \\
 &\quad \cdot \left( \int_0^1 f_{\mathbf{a}}(y_n, \eta_{t_n} + \lambda \Delta \eta_{t_n, s_1}, \dots, \eta_{t_n} + \lambda \Delta \eta_{t_n, s_i}) (1 - \lambda)^{\kappa_i} d\lambda \right) ds_i \cdots ds_1.
 \end{aligned}$$

Now, notice that since  $|\mathbf{a}| \mathbf{b} \leq \|\mathbf{a}^\top \beta\|_1$ , we have:

$$\mathfrak{A}_m^\gamma = \left\{ \mathbf{a} \in \mathfrak{A}_m : \iota(\mathbf{a}) + \|\mathbf{a}^\top \beta\|_1 \leq \gamma + 1 \right\} \subset \left\{ \mathbf{a} \in \mathfrak{A}_m : |\mathbf{a}| \leq \kappa_i(\mathbf{a}) \right\},$$

and thus (36) can be rewritten as:

$$y(t_{n+1}) = \Phi_\gamma^{(h)}(t_n, y_n) + R_1 + R_2 + R_3, \tag{37}$$

with

$$R_3 = \sum_{i=1}^r \sum_{\substack{|\mathbf{a}| \leq \kappa_i \\ \mathbf{a} \in \mathbb{N}_0^{m \times i} \setminus \mathfrak{A}_m^\gamma}} f_{\mathbf{a}}(y_n, \eta_n^{\times i}) \cdot I_{t_n^\alpha, t_{n+1}}.$$

Therefore, taking the dot product of (37) with itself then taking the expectation of the resulting relation gives:

$$\begin{aligned}
 (\mathcal{E}_{n+1}^L)^2 &= \mathbb{E} \left| y(t_{n+1}; t_n, y_n) - \Phi_\gamma^{(h)}(t_n, y_n) \right|^2 \\
 &= \mathbb{E}|R_1|^2 + \mathbb{E}|R_2|^2 + \mathbb{E}|R_3|^2 + 2\mathbb{E}(\langle R_1, R_2 \rangle + \langle R_1, R_3 \rangle + \langle R_2, R_3 \rangle) \\
 &\leq 3\mathbb{E}|R_1|^2 + 3\mathbb{E}|R_2|^2 + 3\mathbb{E}|R_3|^2. \tag{38}
 \end{aligned}$$

We next estimate  $\mathbb{E}|R_1|^2$ ,  $\mathbb{E}|R_2|^2$  and  $\mathbb{E}|R_3|^2$ . Again, we denote by  $K$  a generic constant that may change from line to line.

First, since all partial derivatives of  $f$  are bounded, there exists  $K > 0$  such that:

$$\mathbb{E}|R_1|^2 \leq K \left( \frac{h^{r+1}}{(r+1)!} \right)^2. \tag{39}$$

Second, due to Assumption (A4) and the independence among  $\eta_t^j$ , for every  $k = 1, \dots, \iota(\mathbf{a})$  and  $s_k \in [t_n, t_{n+1}]$ , we have:

$$\mathbb{E} \left[ \Delta \eta_{t_n, s_k}^{2\mathbf{a}_k} \right] = \prod_{j=1}^m \mathbb{E} \left[ \left( \Delta \eta_{t_n, s_k}^j \right)^{2\mathbf{a}_{k,j}} \right] \leq K \prod_{j=1}^m (s_k - t_n)^{2\mathbf{a}_{k,j} b_j} \leq K h^{2\sum_{j=1}^m \mathbf{a}_{k,j} b_j},$$

and therefore:

$$\prod_{k=1}^{\iota(\mathbf{a})} \mathbb{E} \left[ \Delta \eta_{t_n, s_k}^{2\mathbf{a}_k} \right] \leq K \prod_{k=1}^{\iota(\mathbf{a})} h^{2\mathbf{a}_{k,j} b_j} = h^{2\|\mathbf{a}^\top \mathbf{b}\|_1}. \tag{40}$$

For  $\mathbf{a} \in \mathfrak{A}_m$  and  $i = \iota(\mathbf{a}) \geq 1$ , define the random variable:

$$R_{\mathbf{a}} := \sup_{\substack{|w_1|, \dots, |w_i| \leq |\eta|_{\infty} \\ |u| \leq |y|_{\infty}}} |f_{\mathbf{a}}(u, w_1, \dots, w_i)|,$$

where

$$|\eta|_{\infty} := \sup_{t \in [t_0, T]} \left( (\eta_t^1)^2 + \dots + (\eta_t^m)^2 \right)^{\frac{1}{2}}, \quad |y|_{\infty} := \sup_{t \in [t_0, T]} \left( (y^1(t))^2 + \dots + (y^d(t))^2 \right)^{\frac{1}{2}}.$$

Then, there exists  $K > 0$  such that:

$$\begin{aligned} \mathbb{E}|R_2|^2 &\leq K \sum_{i=1}^r \sum_{\substack{|\mathbf{a}|=\kappa_i+1 \\ \mathbf{a} \in \mathbb{N}_0^{m \times i}}} \mathbb{E} \left[ \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} \dots \int_{t_n}^{s_{i-1}} \prod_{k=1}^{\iota(\mathbf{a})} \Delta \eta_{t_n, s_k}^{\mathbf{a}_k} R_{\mathbf{a}} ds_i \dots ds_1 \right]^2 \\ &\leq K \sum_{i=1}^r \sum_{\substack{|\mathbf{a}|=\kappa_i+1 \\ \mathbf{a} \in \mathbb{N}_0^{m \times i}}} \frac{\mathbb{E}[R_{\mathbf{a}}^2]}{i!} h^i \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} \dots \int_{t_n}^{s_{i-1}} \prod_{k=1}^{\iota(\mathbf{a})} \mathbb{E} \left[ \Delta \eta_{t_n, s_k}^{2\mathbf{a}_k} \right] ds_i \dots ds_1. \end{aligned}$$

It then follows immediately from (40) that:

$$\mathbb{E}|R_2|^2 \leq K \sum_{i=1}^r \sum_{\substack{|\mathbf{a}|=\kappa_i+1 \\ \mathbf{a} \in \mathbb{N}_0^{m \times i}}} \frac{\mathbb{E}[R_{\mathbf{a}}^2]}{(i!)^2} h^{2\|\mathbf{a}^{\top} \mathbf{b}\|_1 + 2i}. \tag{41}$$

Following similar calculations, we can obtain:

$$\mathbb{E}|R_3|^2 \leq K \sum_{i=1}^r \sum_{\substack{|\mathbf{a}|=\kappa_i \\ \mathbf{a} \in \mathbb{N}_0^{m \times i} \setminus \mathfrak{A}_m^{\gamma}}} \frac{\mathbb{E}[R_{\mathbf{a}}^2]}{(i!)^2} h^{2\|\mathbf{a}^{\top} \mathbf{b}\|_1 + 2i}. \tag{42}$$

Collecting (39), (41), and (42) into (38) and putting  $i = \iota(\mathbf{a})$  results in:

$$\left( \mathcal{E}_{n+1}^L \right)^2 \leq K \sum_{\substack{\mathbf{a} \in \mathfrak{A}_m \setminus \mathfrak{A}_m^{\gamma} \\ \iota(\mathbf{a}) \leq \lceil \gamma \rceil + 1 \\ |\mathbf{a}| \leq \lceil \gamma / \mathbf{b} \rceil}} \frac{\mathbb{E}[R_{\mathbf{a}}^2]}{[(\iota(\mathbf{a}))!]^2} h^{2(\iota(\mathbf{a}) + \|\mathbf{a}^{\top} \mathbf{b}\|_1)}.$$

Noticing that  $\iota(\mathbf{a}) + \|\mathbf{a}^{\top} \mathbf{b}\|_1 \geq \gamma + 1$  for  $\mathbf{a} \in \mathfrak{A}_m \setminus \mathfrak{A}_m^{\gamma}$  and assuming that  $h \leq 1$  we then obtain:

$$\mathcal{E}_{n+1}^L \leq K(\gamma, \mathbf{b}, f, M_{\eta})h^{\gamma+1},$$

where  $K$  depends on  $\gamma$ , the Hölder property of  $\eta$ , and smoothness of  $f$ , but independent of  $h$ .

The desired global mean-square error estimate then follows from the Lipschitz estimate for  $\Phi_{\gamma}^{(h)}$  (see Lemma 10.4 [10]) and Theorem 1. □



### 3.4 Applications

Here, we present some explicit RODE-Taylor schemes for scalar RODEs ( $d = 1$ ) with Brownian motion or fractional Brownian motion.

*Example 5* Scalar RODEs with scalar Brownian motions.

We have  $m = 1$ ,  $\mathbf{b} = (\frac{1}{2})$  and

$$\mathbb{N}_0 + \frac{1}{2}\mathbb{N}_0 = \left\{ 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, \dots \right\}.$$

The multi-index set for a  $\gamma$ -RODE-Taylor scheme is:

$$\mathfrak{A}_1^\gamma = \left\{ \mathbf{a} \in \mathfrak{A}_1 : \iota(\mathbf{a}) + \frac{1}{2}\|\mathbf{a}\|_1 < \gamma + 1 \right\}.$$

In particular consider  $\gamma = \frac{1}{2}, 1, \frac{3}{2}, 2$  with:

$$\begin{aligned} \mathfrak{A}_1^{\frac{1}{2}} &= \{\emptyset, (0)\}, & \mathfrak{A}_1^1 &= \{\emptyset, (0), (1)\}, & \mathfrak{A}_1^{1.5} &= \{\emptyset, (0), (1), (2), (0, 0)\}, \\ \mathfrak{A}_1^2 &= \{\emptyset, (0), (1), (2), (3), (0, 0), (0, 1), (1, 0)\}. \end{aligned}$$

The corresponding RODE-Taylor schemes are:

$$y_{n+1} = \Phi_{\frac{1}{2}}^{(h)} = y_n + hf \tag{43}$$

$$y_{n+1} = \Phi_1^{(h)} = y_n + hf + f_\eta \int_{t_n}^{t_{n+1}} \Delta \eta_{t_n, s} ds, \tag{44}$$

$$y_{n+1} = \Phi_{1.5}^{(h)} = y_n + hf + \sum_{i=1}^2 \frac{1}{i!} \partial_\eta^i f I_{t_n, t_{n+1}}^{(i)} + f_y f \frac{h^2}{2}, \tag{45}$$

$$y_{n+1} = \Phi_2^{(h)} = y_n + hf + \sum_{i=1}^3 \frac{1}{i!} \partial_\eta^i f I_{t_n, t_{n+1}}^{(i)} + f_y f \frac{h^2}{2} + f_{\eta, y} f I_{t_n, t_{n+1}}^{(1,0)} + f_y f_\eta I_{t_n, t_{n+1}}^{(0,1)}. \tag{46}$$

where  $I_{t_n, t_{n+1}}^\alpha$  are defined by (28), and  $f$  and all partial derivatives of  $f$  are evaluated at  $(y_n, \eta_{t_n})$ .

The above schemes (43)–(46) have global mean-square convergence rates of  $1, 1, \frac{3}{2}$ , and  $2$ , respectively (see Examples 14-(i) and 15-(i) in Section 6).

*Remark 6* The Euler scheme (43) achieves a convergence order 1 instead of  $\frac{1}{2}$  due to the special noise process of Brownian motion. This will be further explained in Example 12 and Remark 13 in Section 5. Note this does not contradict the result in Theorem 3, as stated in Remark 3.

*Example 7* Scalar RODEs with scalar fractional Brownian motion with Hurst parameter  $H$ .

We have  $m = 1$ ,  $\mathbf{b} = (H)$  and

$$\mathbb{N}_0 + H\mathbb{N}_0 = \begin{cases} \{0, H, 1, 2H, \dots\}, & H \in \left(\frac{1}{2}, 1\right) \\ \{0, H, 2H, \dots, 1, \dots\}, & H \in \left(0, \frac{1}{2}\right) \end{cases}$$

The multi-index set for a  $\gamma$ -RODE-Taylor scheme is:

$$\mathfrak{A}_1^\gamma = \{\mathbf{a} \in \mathfrak{A}_1 : \iota(\mathbf{a}) + H\|\mathbf{a}\|_1 < \gamma + 1\}.$$

- For  $H \in \left(\frac{1}{2}, 1\right)$ , we consider the particular case  $\gamma = H, 1, 2H$  with

$$\mathfrak{A}_1^H = \{\emptyset, (0)\}, \quad \mathfrak{A}_1^1 = \{\emptyset, (0), (1)\}, \quad \mathfrak{A}_1^{2H} = \{\emptyset, (0), (1), (0, 0)\}.$$

The corresponding  $\gamma$ -RODE-Taylor schemes read:

$$y_{n+1} = \Phi_H^{(h)} = y_n + hf \tag{47}$$

$$y_{n+1} = \Phi_1^{(h)} = y_n + hf + f_\eta \int_{t_n}^{t_{n+1}} \Delta\eta_{t_n,s} ds, \tag{48}$$

$$y_{n+1} = \Phi_{2H}^{(h)} = y_n + hf + f_\eta \int_{t_n}^{t_{n+1}} \Delta\eta_{t_n,s} ds + f_y f \frac{h^2}{2}, \tag{49}$$

with global mean-square convergence rates of  $H, 1$ , and  $2H$ , respectively (see Examples 14-(ii)/(iii) and 15-(ii)/(iii) in Section 6).

To achieve a convergence rate of 2 as the scheme (46) does, we need more information of  $H$ . For example, with  $H = \frac{3}{4}$ :

$$\mathfrak{A}_1^2 = \left\{ \mathbf{a} \in \mathfrak{A}_1 : \iota(\mathbf{a}) + \frac{3}{4}\|\mathbf{a}\|_1 < 3 \right\} = \{\emptyset, (0), (1), (2), (0, 0), (1, 0), (0, 1)\}$$

and the corresponding 2-RODE-Taylor scheme reads:

$$y_{n+1} = \Phi_2^{(h)} = y_n + hf + \sum_{i=1}^2 \frac{1}{i!} \partial_\eta^i f I_{t_n, t_{n+1}}^{(i)} + f_y f \frac{h^2}{2} + f_{\eta,y} f I_{t_n, t_{n+1}}^{(1,0)} + f_y f_\eta I_{t_n, t_{n+1}}^{(0,1)}. \tag{50}$$

- For  $H \in \left(0, \frac{1}{2}\right)$ , we consider the particular case  $\gamma = H, 2H$  with:

$$\mathfrak{A}_1^H = \{\emptyset, (0)\}, \quad \mathfrak{A}_1^{2H} = \{\emptyset, (0), (1), (0, 0)\}.$$

The corresponding  $\gamma$ -RODE-Taylor schemes are the same as (47) and (49) with global mean-square convergence rate of  $H$  and  $2H$ , respectively (see Examples 14-(ii) and 15-(ii) in Section 6).

To achieve a convergence rate of 1 or 2 as above, we need more information of  $H$ . For example, for  $H = \frac{1}{3}$ :

$$\mathfrak{A}_1^1 = \left\{ \mathbf{a} \in \mathfrak{A}_1 : \iota(\mathbf{a}) + \frac{1}{3}\|\mathbf{a}\|_1 < 2 \right\} = \{\emptyset, (0), (1), (2)\},$$

$$\mathfrak{A}_1^2 = \left\{ \mathbf{a} \in \mathfrak{A}_1 : \iota(\mathbf{a}) + \frac{1}{3}\|\mathbf{a}\|_1 < 3 \right\} = \left\{ \emptyset, (0), (1), (2), (3), (4), (5), (0, 0), \right. \\ \left. (0, 1), (0, 2), (1, 0), (1, 1), (2, 0) \right\},$$

and the corresponding 1- and 2-RODE-Taylor schemes read:

$$y_{n+1} = \Phi_1^{(h)} = y_n + hf + \sum_{i=1}^2 \frac{1}{i!} \partial_\eta^i f I_{t_n, t_{n+1}}^{(i)} \tag{51}$$

$$y_{n+1} = \Phi_2^{(h)} = y_n + hf + \sum_{i=1}^5 \frac{1}{i!} \partial_\eta^i f I_{t_n, t_{n+1}}^{(i)} + f_y f \frac{h^2}{2} + f_y f_\eta I_{t_n, t_{n+1}}^{(0,1)} + f_y f_\eta I_{t_n, t_{n+1}}^{(1,0)} + f_{\eta,y} f_\eta I_{t_n, t_{n+1}}^{(1,1)} + \frac{1}{2} f_y f_{\eta,\eta} I_{t_n, t_{n+1}}^{(0,2)} + \frac{1}{2} f_{\eta,\eta,y} f_\eta I_{t_n, t_{n+1}}^{(2,0)}. \tag{52}$$

The schemes (51) and (52) have mean-square convergence orders of 1 and 2, respectively (see Examples 14-(iii) and 15-(iii) in Section 6).

*Example 8* Notice that when  $m > 1$ , Assumption (A4) allows different components of  $\eta_t$  to have different Hölder properties. In this example, we consider RODEs with a 2-dimensional stochastic process  $\eta_t = (\eta_t^1, \eta_t^2)^\top$  where  $\eta_t^1$  is a Brownian motion and  $\eta_t^2$  is a fractional Brownian motion with Hurst parameter  $H = \frac{3}{4}$ , i.e.,  $\mathfrak{b} = (\frac{1}{2}, \frac{3}{4})$ . Then, we have:

$$\mathbb{N}_0 + \frac{1}{2} \mathbb{N}_0 + \frac{3}{4} \mathbb{N}_0 = \left\{ 0, \frac{1}{2}, \frac{3}{4}, 1, \frac{5}{4}, \frac{3}{2}, \frac{7}{4}, 2, \frac{9}{4}, \frac{5}{2}, \dots \right\},$$

and the multi-index set for a  $\gamma$ -RODE-Taylor scheme is:

$$\mathfrak{A}_2^\gamma = \left\{ \mathfrak{a} \in \mathfrak{A}_2 : \iota(\mathfrak{a}) + \frac{1}{2} \sum_{k=1}^{\iota(\mathfrak{a})} \mathfrak{a}_{k,1} + \frac{3}{4} \sum_{k=1}^{\iota(\mathfrak{a})} \mathfrak{a}_{k,2} < \gamma + 1 \right\}.$$

In particular, we consider  $\gamma = \frac{1}{2}, 1, \frac{3}{2}$  with

$$\begin{aligned} \mathfrak{A}_2^{\frac{1}{2}} &= \left\{ \emptyset, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}, & \mathfrak{A}_2^1 &= \left\{ \emptyset, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}, \\ \mathfrak{A}_2^{1.5} &= \left\{ \emptyset, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right\}. \end{aligned}$$

Hence, the corresponding  $\frac{1}{2}$ -, 1-, and 1.5-RODE-Taylor schemes are:

$$y_{n+1} = \Phi_{\frac{1}{2}}^{(h)} = y_n + hf, \tag{53}$$

$$y_{n+1} = \Phi_1^{(h)} = y_n + hf + f_{\eta^1} \int_{t_n}^{t_{n+1}} \Delta \eta_{t_n, s}^1 ds + f_{\eta^2} \int_{t_n}^{t_{n+1}} \Delta \eta_{t_n, s}^2 ds \tag{54}$$

$$y_{n+1} = \Phi_{\frac{3}{2}}^{(h)} = \Phi_1^{(h)} + f_{\eta^1, \eta^2} \int_{t_n}^{t_{n+1}} \Delta \eta_{t_n, s}^1 \Delta \eta_{t_n, s}^2 ds + \frac{f_{\eta^1, \eta^1}}{2} \int_{t_n}^{t_{n+1}} (\Delta \eta_{t_n, s}^1)^2 ds + f_y f \frac{h^2}{2}. \tag{55}$$

The global mean-square convergence orders for (53), (54), and (55) are  $\frac{1}{2}, 1$ , and  $\frac{3}{2}$ , respectively (see Example 16 in Section 6).

## 4 Affine-RODE-Taylor schemes for RODEs with affine noise

In this section, we consider a special type of RODEs with affine noise [3, 9], formulated as:

$$\frac{dy}{dt} = g(t, y) + \sum_{j=1}^m \sigma_j(t, y) \eta_t^j, \quad y(t_0) = y_0, \quad (56)$$

where  $y \in \mathbb{R}^d$  and the noise process  $\eta_t = (\eta_t^1, \dots, \eta_t^m)$  takes values in  $\mathbb{R}^m$ . Here, the sample paths of  $\eta_t$  are assumed to be continuous, and the coefficients  $\sigma_j$ ,  $j = 1, \dots, m$  are assumed to be sufficiently smooth real-valued functions satisfying appropriate conditions to ensure the existence and uniqueness of solutions to (56) on the whole time interval  $[t_0, T]$ .

When  $g(t, y) \equiv g(y)$  and  $\sigma_j(t, y) \equiv \sigma_j(y)$ , the RODE-Taylor approximations introduced in Section 3 can be applied to the vector field  $f(y, \eta_t) := g(y) + \sum_{j=1}^m \sigma_j(y) \eta_t^j$  to obtain RODE-Taylor schemes of various mean-square convergence orders. Here, we introduce the affine-RODE-Taylor framework due to Asai and Kloeden [3] that deals with nonautonomous affine RODEs (56) and may achieve a different rate of convergence compared with RODE-Taylor schemes when  $g$  and  $\sigma_j$  do not depend on  $t$ . For readers' convenience, the hierarchical sets and vector-valued multi-index notations used in [3, 13] are summarized below.

### 4.1 Preliminaries

Let  $\mathcal{A}_m$  be the set of all  $l$ -dimensional vectors with components valued in the set  $\{0, 1, \dots, m\}$ :

$$\mathcal{A}_m = \left\{ \alpha = (\alpha_1, \dots, \alpha_l) \in \{0, 1, \dots, m\}^l : l \in \mathbb{N}_0 \right\}, \quad (57)$$

where  $l = \ell(\alpha)$  is the length of vector  $\alpha$ . In particular,  $\ell(\emptyset) = 0$ . For any  $\alpha = (\alpha_1, \dots, \alpha_l) \in \mathcal{A}_m$  with  $\ell(\alpha) = l \geq 1$ , denote by  $-\alpha$  and  $\alpha^-$  the multi-indices in  $\mathcal{A}_m$  obtained by deleting the first and last components of  $\alpha$ , respectively, i.e.:

$$-\alpha = (\alpha_2, \dots, \alpha_l), \quad \alpha^- = (\alpha_1, \dots, \alpha_{l-1}).$$

Given  $\gamma \in \mathbb{N}$ , let  $\mathcal{A}_m^\gamma$  be the *hierarchical set* of vector-valued multi-indices given by:

$$\mathcal{A}_m^\gamma := \{ \alpha \in \mathcal{A}_m : \ell(\alpha) \leq \gamma \}.$$

In addition, define the *remainder set*  $\mathcal{R}(\mathcal{A}_m^\gamma)$  by:

$$\mathcal{R}(\mathcal{A}_m^\gamma) := \{ \alpha \in \mathcal{A}_m \setminus \mathcal{A}_m^\gamma : -\alpha \in \mathcal{A}_m^\gamma \}.$$

Then,  $\mathcal{R}(\mathcal{A}_m^\gamma)$  consists of all the next following multi-indices with respect to the hierarchical set  $\mathcal{A}_m^\gamma$  that do not belong to  $\mathcal{A}_m^\gamma$ . Obviously,  $\mathcal{R}(\mathcal{A}_m^\gamma) = \mathcal{A}_m^{\gamma+1} \setminus \mathcal{A}_m^\gamma$ .

For any  $\alpha \in \mathcal{A}_m$  with  $\ell(\alpha) = l$  and an integrable function  $\phi : [t_0, T] \rightarrow \mathbb{R}$ , define the multiple integral  $J[\phi(\cdot)]_{t,s}^\alpha$  as:

$$J[\phi(\cdot)]_{t,s}^\alpha := \begin{cases} \phi(s), & l = 0 \\ \int_t^s J[\phi(\cdot)]_{t,s}^{\alpha'} \eta_\tau^{\alpha_l} d\tau, & l \geq 1 \end{cases} \tag{58}$$

In particular, when  $\phi \equiv 1$ , write  $J[1]_{t,s}^\alpha$  as  $J_{t,s}^\alpha$  in short.

Denote by  $\phi^k$  the  $k$ -th component of an  $\mathbb{R}^d$ -valued function  $\phi$ . For  $y = (y^1, \dots, y^d) \in \mathbb{R}^d$  and the  $\mathbb{R}^d$ -valued functions  $g$  and  $\sigma_j$  in (56), define the partial differential operators  $\mathcal{D}^0$  and  $\mathcal{D}^j$  by:

$$\mathcal{D}^0 = \frac{\partial}{\partial t} + \sum_{k=1}^d g^k \frac{\partial}{\partial y^k}, \quad \mathcal{D}^j = \sum_{k=1}^d \sigma_j^k \frac{\partial}{\partial y^k}, \quad j = 1, \dots, m. \tag{59}$$

Given a vector-valued multi-index  $\alpha \in \mathcal{A}_m$  and a sufficiently smooth function  $\varphi : [t_0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ , the coefficient function  $\varphi_\alpha$  is defined recursively by:

$$\varphi_\alpha := \begin{cases} \varphi, & \ell(\alpha) = 0, \\ \mathcal{D}^{\alpha_1} \varphi_{-\alpha}, & \ell(\alpha) \geq 1 \end{cases}.$$

Then, the  $\mathcal{A}_m^\gamma$ -RODE-Taylor expansion of  $\varphi(t, y(t))$  for a solution  $y(t)$  of the RODE (56) reads:

$$\varphi(t, y(t)) = \sum_{\alpha \in \mathcal{A}_m^\gamma} \varphi_\alpha(t_0, y_0) J_{t_0,t}^\alpha + \sum_{\alpha \in \mathcal{R}(\mathcal{A}_m^\gamma)} J[\varphi_\alpha(\cdot, y(\cdot))]_{t_0,t}^\alpha.$$

In particular, let  $\varphi = \chi$ , the identity function on  $\mathbb{R}^d$  with  $\chi^k(y) = y^k$  for  $k = 1, \dots, d$ . Then, given any  $y(t_n) = y_n$ , the solution of the RODE (56) can be written componentwise as:

$$\phi^k(t; t_n, y_n) = \sum_{\alpha \in \mathcal{A}_m^\gamma} \mathcal{D}^\alpha \chi^k(t_n, y_n) J_{t_n,t}^\alpha + \sum_{\alpha \in \mathcal{R}(\mathcal{A}_m^\gamma)} J[\mathcal{D}^\alpha \chi^k(\cdot, y(\cdot))]_{t_0,t}^\alpha, \tag{60}$$

where  $J$  is defined as in (58) and  $\mathcal{D}^\alpha = \mathcal{D}^{\alpha_1} \mathcal{D}^{\alpha_2} \dots \mathcal{D}^{\alpha_l}$  with  $\mathcal{D}^j$  defined as in (59) for  $j = 0, 1, \dots, m$ .

Affine-RODE-Taylor schemes are constructed systematically using the affine-RODE-Taylor expansions (60) on each finite subinterval  $[t_n, t_{n+1}]$ , deleting the remainder term on the right-hand side of (60). More specifically, the  $\gamma$ -affine-RODE-Taylor scheme is given componentwise by:

$$y_{n+1}^k = \Psi_\gamma^k(h) := y_n^k + \sum_{\alpha \in \mathcal{A}_m^\gamma} \mathcal{D}^\alpha \chi^k(t_n, y_n) J_{t_n,t_{n+1}}^\alpha, \quad k = 1, \dots, d. \tag{61}$$

### 4.2 Convergence analysis

In this subsection, we analyze the mean-square convergence rate for the  $\gamma$ -affine-RODE-Taylor scheme. The analysis is based on Theorem 2 in [3], but in the mean-square sense with different estimations. Here, all the stochastic processes  $\eta_t^1, \dots, \eta_t^m$  are assumed to be mutually independent and satisfy:

(A6) there exists  $C > 0$  such that

$$\sup_{t \in [t_0, T]} \mathbb{E} \left| \eta_t^j \right|^2 \leq C, \quad j = 1, \dots, m.$$

**Theorem 4** *Let Assumptions (A1)–(A3) and (A6) hold, and assume that all derivatives of  $g, \sigma_1, \dots, \sigma_m$  and all the multiple integrals of stochastic processes appearing in (60) exist. Then, given any  $\gamma > 0$ , the affine-RODE-Taylor scheme (61) has a global mean-square convergence order of  $\gamma$ .*

*Proof* The local mean-square error for the  $k$ -th component of  $\gamma$ -affine-RODE-Taylor scheme (61) is given by:

$$\mathcal{E}_{n+1}^{L,k} = \left( \mathbb{E} \left| \sum_{\alpha \in \mathcal{R}(\mathcal{A}_m^\gamma)} J[\mathcal{D}^\alpha \chi^k(\cdot, y(\cdot))]_{t_n, t_{n+1}}^\alpha \right|^2 \right)^{\frac{1}{2}},$$

and satisfies

$$\left( \mathcal{E}_{n+1}^{L,k} \right)^2 \leq K \sum_{\alpha \in \mathcal{R}(\mathcal{A}_m^\gamma)} \mathbb{E} \left| J[\mathcal{D}^\alpha \chi^k(\cdot, y(\cdot))]_{t_n, t_{n+1}}^\alpha \right|^2.$$

First, note that under standard assumptions, the RODE (61) has a unique solution on the finite interval  $[t_0, T]$  with continuous path. Thus, for any  $\omega \in \Omega$ , there exists  $R = R(T, \omega)$  such that  $|y(t, \omega)| \leq R(T, \omega)$  for all  $t \in [t_0, T]$ . Moreover, there exists  $C = C(\mathcal{D}^\alpha \chi, t_0, T) > 0$  such that  $|\mathcal{D}^\alpha \chi^k(t, y(t))| \leq C$  for all  $t \in [t_0, T]$  (see [3]). Therefore:

$$\begin{aligned} \left| J[\mathcal{D}^\alpha \chi^k(\cdot, y(\cdot))]_{t_n, t_{n+1}}^\alpha \right| &= \left| \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} \dots \int_{t_n}^{s_l} \mathcal{D}^\alpha \chi^k(s_1, y(s_1)) \eta_{s_1}^{\alpha_1} \dots \eta_{s_{l-1}}^{\alpha_{l-1}} \eta_{s_l}^{\alpha_l} ds_1 \dots ds_l \right| \\ &\leq C \left| \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} \dots \int_{t_n}^{s_l} \eta_{s_1}^{\alpha_1} \dots \eta_{s_{l-1}}^{\alpha_{l-1}} \eta_{s_l}^{\alpha_l} ds_1 \dots ds_l \right|, \end{aligned}$$

where  $l = \ell(\alpha)$ . Then, by Hölder’s inequality:

$$\left| J[\mathcal{D}^\alpha \chi^k(\cdot, y(\cdot))]_{t_n, t_{n+1}}^\alpha \right|^2 \leq Ch^l \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} \dots \int_{t_n}^{s_l} |\eta_{s_1}^{\alpha_1}|^2 \dots |\eta_{s_{l-1}}^{\alpha_{l-1}}|^2 |\eta_{s_l}^{\alpha_l}|^2 ds_1 \dots ds_l.$$

Taking expectation of the above inequality and using Assumption (A6) and the independence among  $\eta_t^j$ , we obtain:

$$\begin{aligned} \mathbb{E} \left| J[\mathcal{D}^\alpha \chi^k(\cdot, y(\cdot))]_{t_n, t_{n+1}}^\alpha \right|^2 &\leq Ch^l \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} \dots \int_{t_n}^{s_l} \prod_{i=1}^l \mathbb{E} \left| \eta_{s_i}^{\alpha_i} \right|^2 ds_1 \dots ds_l \\ &\leq Kh^l \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} \dots \int_{t_n}^{s_l} ds_1 \dots ds_l \leq Kh^{2l}. \end{aligned}$$

Since  $\gamma \in \mathbb{N}$ ,  $\ell(\alpha) = l \geq \gamma + 1$  for all  $\alpha \in \mathcal{R}(\mathcal{A}_m^\gamma)$ . Therefore, assuming  $h \leq 1$  we obtain immediately that:

$$\mathbb{E} \left| J[\mathcal{D}^\alpha \chi^k(\cdot, y(\cdot))]_{t_n, t_{n+1}}^\alpha \right|^2 \leq Kh^{2(\gamma+1)} \quad \forall \alpha \in \mathcal{R}(\mathcal{A}_m^\gamma).$$

which implies that  $(\mathcal{E}_{n+1}^{L,k})^2 \leq Kh^{2(\gamma+1)}$ . It then follows immediately that:

$$\mathcal{E}_{n+1}^L = \left( \sum_{k=1}^d (\mathcal{E}_{n+1}^{L,k})^2 \right)^{\frac{1}{2}} \leq Kh^{\gamma+1},$$

and by Theorem 1 the  $\gamma$ -affine-RODE-Taylor scheme (61) has a global mean-square convergence order of  $\gamma$ . □

Next, we present some explicit  $\gamma$ -affine-RODE-Taylor schemes for scalar RODEs with affine noise. In particular, for comparison purpose, we consider autonomous examples with  $g(t, y) = g(y)$  and  $\sigma_j(t, y) = \sigma_j(y)$  for  $j = 1, \dots, m$ . Then:

$$\mathcal{D}^0 \phi = g\phi', \quad \mathcal{D}^j \phi = \sigma_j \phi', \quad \forall \phi \text{ smooth.}$$

*Example 9* Consider the RODE (56) with  $d = m = 1$ . The hierarchical sets  $\mathcal{A}_1^\gamma$  for  $\gamma = 1, 2, 3$  are, respectively:

$$\begin{aligned} \mathcal{A}_1^1 &= \{\emptyset, (0), (1)\}, & \mathcal{A}_1^2 &= \{\emptyset, (0), (1), (0, 0), (0, 1), (1, 0), (1, 1)\}, \\ \mathcal{A}_1^3 &= \left\{ \emptyset, (0), (1), (0, 0), (0, 1), (1, 0), (1, 1), (0, 0, 0), (0, 0, 1), \right. \\ & \quad \left. (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1) \right\}. \end{aligned}$$

Then, the corresponding 1-, 2-, 3-affine-RODE-Taylor schemes read:

$$y_{n+1} = \Psi_1^{(h)} = y_n + gh + \sigma \int_{t_n}^{t_{n+1}} \eta_s ds, \tag{62}$$

$$y_{n+1} = \Psi_2^{(h)} = \Psi_1^{(h)} + \frac{1}{2}gg'h^2 + g\sigma'J_n^{(0,1)} + \sigma g'J_n^{(1,0)} + \sigma\sigma'J_n^{(1,1)}, \tag{63}$$

$$\begin{aligned} y_{n+1} = \Psi_3^{(h)} &= \Psi_2^{(h)} + g \left[ \frac{1}{6}(gg')'h^3 + (g\sigma')'J_n^{(0,0,1)} + (\sigma g')'J_n^{(0,1,0)} + (\sigma\sigma')'J_n^{(0,1,1)} \right] \\ & \quad + \sigma \left[ (gg')'J_n^{(1,0,0)} + (g\sigma')'J_n^{(1,0,1)} + (\sigma g')'J_n^{(1,1,0)} + (\sigma\sigma')'J_n^{(1,1,1)} \right], \end{aligned} \tag{64}$$

where  $g, \sigma$ , and their derivatives are evaluated at  $(t_n, y_n)$ , and the integrals  $J_n^\alpha = J[1]_{t_n, t_{n+1}}^\alpha$ . The global mean-square convergence orders of (62), (63), and (64) are 1, 2, and 3, respectively (see Example 15 in Section 6).

*Example 10* Consider the RODE (4.1) with  $d = 1$  and  $m = 2$ . The hierarchical sets  $\mathcal{A}_2^\gamma$  for  $\gamma = 1, 2$  are, respectively:

$$\begin{aligned} \mathcal{A}_2^1 &= \{\emptyset, (0), (1), (2)\}, \\ \mathcal{A}_2^2 &= \{\emptyset, (0), (1), (2), (0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (1, 2), (2, 0), (2, 1), (2, 2)\}. \end{aligned}$$

Then, the corresponding 1-, 2-affine-RODE-Taylor schemes read:

$$y_{n+1} = \Psi_1^{(h)} = y_n + gh + \sigma_1 \int_{t_n}^{t_{n+1}} \eta_s^1 ds + \sigma_2 \int_{t_n}^{t_{n+1}} \eta_s^2 ds, \tag{65}$$

$$y_{n+1} = \Psi_2^{(h)} = \Psi_1^{(h)} + g \left[ \frac{1}{2} g' h^2 + \sigma_1' J_n^{(0,1)} + \sigma_2' J_n^{(0,2)} \right] + \sigma_1 \left[ g' J_n^{(1,0)} + \sigma_1' J_n^{(1,1)} \right] + \sigma_1 \sigma_2' J_n^{(1,2)} + \sigma_2 \left[ g' J_n^{(2,0)} + \sigma_1' J_n^{(2,1)} + \sigma_2' J_n^{(2,2)} \right], \tag{66}$$

where  $g, \sigma_1, \sigma_2$ , and their derivatives are evaluated at  $(t_n, y_n)$ , and the integrals  $J_n^\alpha = J[1]_{t_n, t_{n+1}}^\alpha$ . The global mean-square convergence orders of (65) and (66) are 1 and 2, respectively (see Example 16 in Section 6).

### 5 Itô-RODE-Taylor schemes for RODEs with Itô diffusion processes

In this section, we investigate mean-square convergence of RODEs driven by an Itô diffusion process, i.e., the solution of an Itô SODE. In particular, consider a RODE on  $\mathbb{R}^{d_1}$ :

$$\frac{dy}{dt} = f(y, \eta_t), \quad y(t_0) = y_0, \tag{67}$$

where  $\eta_t$  is the solution of an Itô SODE in  $\mathbb{R}^{d_2}$ :

$$d\eta_t = \mu(\eta_t)dt + \sum_{j=1}^m \rho_j(\eta_t) dW_t^j \tag{68}$$

with  $m$ -independent scalar Wiener processes  $W_t^1, \dots, W_t^m$ .

In this section, we again assume the standard assumption holds, i.e.:

- (A7) All the coefficients  $\mu, \rho_1, \dots, \rho_m$  and the vector field  $f$  are infinitely often continuously differentiable in its variables and all partial derivatives are bounded on  $[t_0, T]$

It is straightforward to verify that under Assumption (A7), the solution  $\eta_t = (\eta_t^1, \dots, \eta_t^{d_2})$  of the SDE (68) satisfies:

$$\mathbb{E}|\eta_t^j - \eta_s^j|^{2q} \leq K|t - s|^q, \quad j = 1, \dots, d_2, \quad \forall q = 1, 2, \dots$$

Therefore, Assumption (A4) is satisfied with  $b_j = \frac{1}{2}$  for all  $j = 1, \dots, d_2$  and all the RODE-Taylor schemes introduced in Section 3 can be applied here with  $b = (\frac{1}{2}, \dots, \frac{1}{2})$  and attain their orders as shown in Section 3.

On the other hand, the system (67)–(68) can also be simulated via stochastic ordinary differential equations (see [1]). For comparison purpose, we also present the convergence analysis for numerical schemes for the RODE (67) via SODEs. More precisely, for  $y(t) \in \mathbb{R}^{d_1}$  in (67) and  $\eta_t \in \mathbb{R}^{d_2}$  in (68), let  $Y(t) = (y(t), \eta_t)^\top$ . Then, (67) and (68) can be formulated as an SODE in  $d = d_1 + d_2$ :

$$dY(t) = F(Y)dt + \sum_{j=1}^m G_j(Y) dW_t^j, \tag{69}$$



with

$$F(Y) = \begin{pmatrix} f(y, \eta_t) \\ \mu(\eta_t) \end{pmatrix}, \quad G_j(Y) = \begin{pmatrix} 0 \\ \rho_j(\eta_t) \end{pmatrix}.$$

### 5.1 Preliminaries

Let  $\mathcal{A}_m$  be defined as in (57) and define the hierarchical set of multi-indices:

$$\Lambda_\gamma = \left\{ \alpha \in \mathcal{A}_m : \ell(\alpha) + n(\alpha) \leq 2\gamma \text{ or } \ell(\alpha) = n(\alpha) = \gamma + \frac{1}{2} \right\},$$

where  $\ell(\alpha)$  is the length of  $\alpha$  and  $n(\alpha)$  is the number of zero components of  $\alpha$ .

Given a multi-index  $\alpha = (\alpha_1, \dots, \alpha_l)$  with  $\ell(\alpha) = l \geq 1$ , define the multiple integrals:

$$\mathcal{I}_n^\alpha := \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} \dots \int_{t_n}^{s_l} dW_{s_1}^{\alpha_1} \dots dW_{s_{l-1}}^{\alpha_{l-1}} dW_{s_l}^{\alpha_l}, \quad \text{with } \mathcal{I}_n^\emptyset = 1. \quad (70)$$

In addition, for any smooth enough scalar-valued function  $\phi$  of  $t$  and  $Y$ , define the differential operators:

$$\begin{aligned} \mathcal{L}^0 \phi &= \frac{\partial \phi}{\partial t} + \sum_{k=1}^d F^k \frac{\partial \phi}{\partial Y^k} + \frac{1}{2} \sum_{k,l=1}^d \sum_{j=1}^m G_j^k G_j^l \frac{\partial^2 \phi}{\partial Y^k \partial Y^l}, \\ \mathcal{L}^j \phi &= \sum_{k=1}^{d_2} G_j^k \frac{\partial \phi}{\partial Y^k}, \quad j = 1, \dots, m, \end{aligned}$$

where the superscript  $k$  denotes the  $k$ -th component of a vector-valued function. Moreover,  $\mathcal{L}^\alpha := \mathcal{L}^{\alpha_1} \dots \mathcal{L}^{\alpha_l}$  for  $\alpha = (\alpha_1, \dots, \alpha_l)$ .

The  $\gamma$ -Itô-Taylor scheme for the SDE (69) reads:

$$Y_{n+1} = \sum_{\alpha \in \Lambda_\gamma^0} \mathcal{L}^\alpha \chi(t_n, Y_n) \mathcal{I}_n^\alpha,$$

and has a mean-square convergence order of  $\gamma$  [13]. In particular, the  $y$ -component of the  $\gamma$ -Itô-Taylor scheme gives the  $\gamma$ -Itô-RODE-Taylor scheme:

$$y_{n+1}^k = \Theta_{\gamma}^{(h)}(t_n, y_n) := \sum_{\alpha \in \Lambda_\gamma^0} \mathcal{L}^\alpha \chi(t_n, y_n, \eta_{t_n}) \mathcal{I}_n^\alpha, \quad k = 1, 2, \dots, d_1,$$

where  $\mathcal{I}_n^\alpha$  is defined as in (70) and

$$\Lambda_\gamma^0 = \{ \alpha \in \Lambda_\gamma \subset \mathcal{A}_1 : \alpha = \emptyset \text{ or } \ell(\alpha) \geq 1 \text{ with last component } \alpha_l = 0 \}.$$

*Remark 11* The smaller set  $\Lambda_\gamma^0$  is used because the coefficients corresponding to  $\Lambda_\gamma \setminus \Lambda_\gamma^0$  are annihilated as the diffusion coefficient component is 0 in the RODE component of the system (69) (see [1]).

### 5.2 Applications

Some explicit Itô-RODE-Taylor schemes for scalar RODEs ( $d = 1$ ) with scalar Itô diffusion processes are provided below.

*Example 12* For  $\gamma = \frac{1}{2}$ ,  $\Lambda_{\frac{1}{2}} = \{\emptyset, (0), (1)\}$  and  $\Lambda_{\frac{1}{2}}^0 = \{\emptyset, (0)\}$ . Thus the corresponding  $\frac{1}{2}$ -Itô-RODE-Taylor scheme reads:

$$y_{n+1} = \Theta_{\frac{1}{2}}^{(h)}(t_n, y_n) = y_n + hf(y_n, \eta_{t_n}), \tag{71}$$

which is the Euler scheme.

For  $\gamma = 1$ ,  $\Lambda_1 = \{\emptyset, (0), (1), (1, 1)\}$  and  $\Lambda_1^0 = \{\emptyset, (0)\}$ . The corresponding 1-Itô-RODE-Taylor scheme  $y_{n+1} = \Theta_1^{(h)}(t_n, y_n)$  coincides with the Euler scheme (71). Thus, the Euler scheme (71) has a mean-square convergence order of 1 (see Examples 14-(i) and 17 in Section 6).

*Remark 13* The Euler scheme (71) can be derived from the  $\frac{1}{2}$ -RODE-Taylor scheme  $\Phi_{\frac{1}{2}}^{(h)}$  in (43) and from  $\frac{1}{2}$ -Itô-RODE-Taylor scheme  $\Theta_{\frac{1}{2}}^{(h)}$  in (71). It coincides with the 1-Itô-RODE-Taylor scheme and thus achieves a convergence order of 1 instead of  $\frac{1}{2}$ .

For  $\gamma = \frac{3}{2}$ ,  $\Lambda_{\frac{3}{2}} = \{\emptyset, (0), (1), (0, 0), (0, 1), (1, 0), (1, 1), (1, 1, 1)\}$  and

$$\Lambda_{\frac{3}{2}}^0 = \{\emptyset, (0), (0, 0), (1, 0)\}.$$

The corresponding  $\frac{3}{2}$ -Itô-RODE-Taylor scheme reads:

$$y_{n+1} = \Theta_{\frac{3}{2}}^{(h)}(t_n, y_n) = y_n + fh + \rho f_{\eta} \mathcal{I}_n^{(1,0)} + \frac{1}{2}h^2 \left( ff_y + \mu f_{\eta} + \frac{1}{2}\rho^2 f_{\eta,\eta} \right), \tag{72}$$

where  $f$  and all its partial derivatives are evaluated at  $(y_n, \eta_{t_n})$ ,  $\rho$  and  $\mu$  are evaluated at  $\eta_{t_n}$ . The Itô-RODE-Taylor scheme (72) has a global mean-square convergence order of  $\frac{3}{2}$  (see Example 17 in Section 6).

For  $\gamma = 2$ ,  $\Lambda_2 = \Lambda_{\frac{3}{2}} \cup \{(0, 1, 1), (1, 0, 1), (1, 1, 0), (1, 1, 1, 1)\}$  and  $\Lambda_2^0 = \{\emptyset, (0), (0, 0), (1, 0), (1, 1, 0)\}$ . The corresponding 2-Itô-RODE-Taylor scheme reads:

$$y_{n+1} = \Theta_2^{(h)}(t_n, y_n) = y_n + fh + \rho f_{\eta} \mathcal{I}_n^{(1,0)} + \frac{1}{2}h^2 \left( ff_y + \mu f_{\eta} + \frac{1}{2}\rho^2 f_{\eta,\eta} \right) + \left( \rho\rho' f_{\eta} + \rho^2 f_{\eta,\eta} \right) \mathcal{I}_n^{(1,1,0)}, \tag{73}$$

where  $f$  and all its partial derivatives are evaluated at  $(y_n, \eta_{t_n})$ ,  $\rho$ ,  $\rho'$ , and  $\mu$  are evaluated at  $\eta_{t_n}$ . The Itô-RODE-Taylor scheme (73) has a global mean-square convergence order of 2 (see Example 17 in Section 6).

## 6 Numerical experiments

In this section, we conduct numerical experiments to demonstrate the mean-square orders of convergence for different schemes presented in Sections 3, 4, and 5. In particular, we simulate:

- (I) A scalar RODE with scalar noise process being Brownian motion or fBm;
- (II) A scalar RODE with scalar affine noise being either Brownian motion or fBm;
- (III) A scalar RODE with 2-dimensional affine noise consisting of one Brownian motion and one fBm;
- (IV) A scalar RODE with scalar Itô diffusion noise.

Let  $M$  be the sample size of the Monte Carlo simulation. For  $i = 1, \dots, M$ , denote by  $y_n^{(i)}$  the numerical approximation to the true solution  $y^{(i)}(t_n; t_0, y_0)$  at the time instant  $t_n$ . The mean-square error  $\mathcal{E}_{n+1}^G$  is simulated by:

$$\mathcal{E}_{n+1}^G = \left( \frac{1}{M} \sum_{i=1}^M |y_n^{(i)} - y^{(i)}(t_n; t_0, y_0)|^2 \right)^{1/2}$$

to verify the mean-square convergence order of various numerical schemes. In all the numerical experiments below, the sample size  $M$  is chosen to be 1000 and the “exact” solutions are calculated numerically with high precision (the integral is evaluated as a Riemann sum with 20,000 subintervals).

*Example 14* Consider the scalar RODE:

$$\frac{dy}{dt} = -y + \cos \eta_t, \quad y(0) = y_0. \tag{74}$$

For each realization  $\eta^{(i)}$  (74) has the explicit solution:

$$y^{(i)}(t; 0, y_0) = y_0 e^{-t} + e^{-t} \int_0^t e^s \cos \eta_s^{(i)} ds.$$

The initial valued is picked to be  $y_0 = 1$  and the final time is picked to be  $T = 1$ .

- (i) Assume that  $\eta_t$  is a Wiener process. The Euler scheme (43) of order 1, the RODE-Taylor schemes (44) of order 1, (45) of order  $\frac{3}{2}$  and (46) of order 2 are applied to simulate the RODE (74). The global mean-square errors  $\mathcal{E}_{n+1}^G$  versus the time steps  $h = 2^{-i}$ ,  $i = 3, 4, 5, 6, 7, 8$  are plotted in Fig. 1. The two dashed lines are reference lines with slopes 1 and 2, respectively.
- (ii) Assume that  $\eta_t$  is an fBm with Hurst parameter  $\frac{3}{4}$ . The RODE-Taylor schemes (47) of order  $\frac{3}{4}$ , (48) of order 1, (49) of order  $\frac{3}{2}$ , and (50) of order 2 are applied to simulate the RODE (74). The global mean-square errors  $\mathcal{E}_{n+1}^G$  versus the time steps  $h = 2^{-i}$ ,  $i = 3, 4, 5, 6, 7, 8$  are plotted in Fig. 2. The two dashed lines are reference lines with slopes 0.75 and 2, respectively.
- (iii) Assume that  $\eta_t$  is an fBm with Hurst parameter  $\frac{1}{3}$ . The RODE-Taylor schemes (51) of order 1 and (52) of order 2 are applied to simulate the RODE (74). The global mean-square errors  $\mathcal{E}_{n+1}^G$  versus the time steps  $h = 2^{-i}$ ,  $i =$

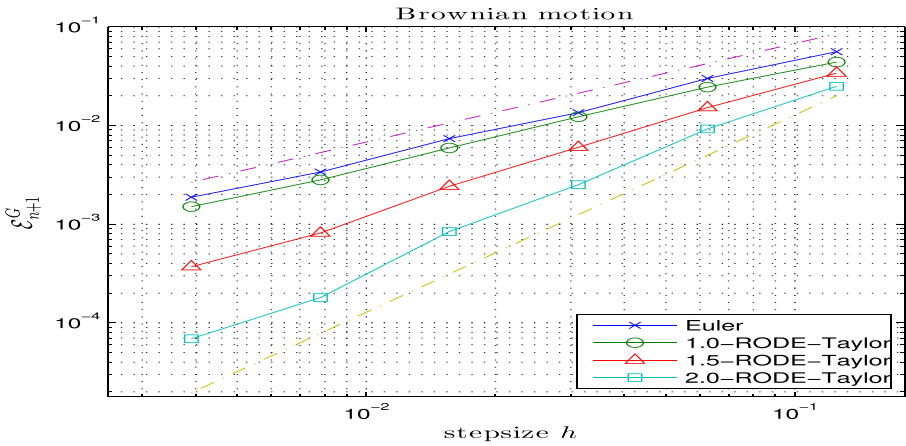


Fig. 1 RODE-Taylor schemes for (74) with Wiener process

3, 4, 5, 6, 7, 8 are plotted in Fig. 3. The two dashed lines are reference lines with slopes 1 and 2, respectively.

*Example 15* Consider the scalar RODE with affine noise:

$$\frac{dy}{dt} = -e^{-2y} + \frac{1}{2}e^{-2y} \cdot \eta_t, \quad y(0) = y_0. \tag{75}$$

For each realization  $\eta^{(i)}$  (75) has the explicit solution:

$$y^{(i)}(t; 0, y_0) = \frac{1}{2} \ln \left( e^{2y_0} + \int_0^t (\eta_s^{(i)} - 2) ds \right).$$

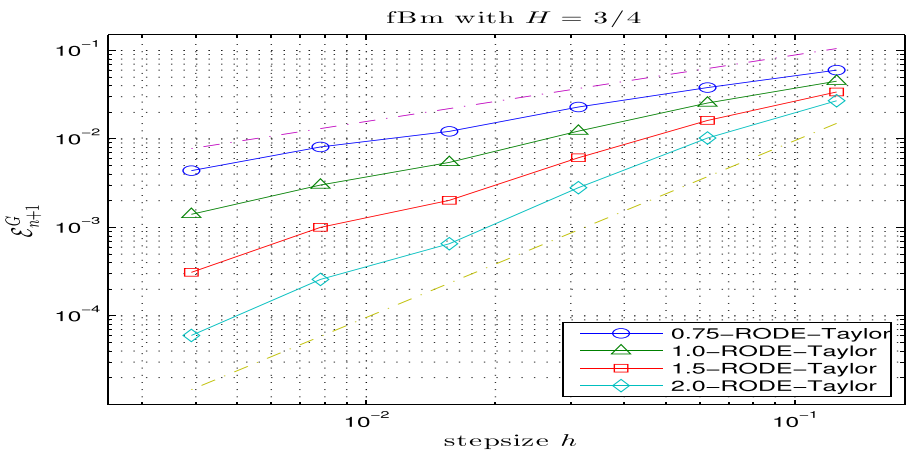


Fig. 2 RODE-Taylor schemes for (74) with fBm with Hurst parameter  $\frac{3}{4}$

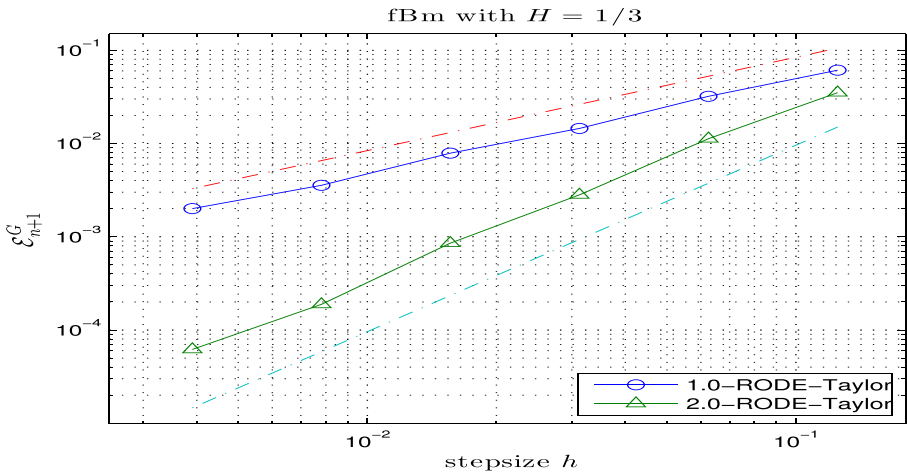


Fig. 3 RODE-Taylor schemes for (74) with fBm with Hurst parameter  $\frac{1}{3}$

The initial value is picked to be  $y_0 = 1$  and the final time is picked to be  $T = 1$ .

- (i) Assume that  $\eta_t = \sin W_t$ , where  $W_t$  is a Wiener process. The Euler scheme (43) of order 1, the RODE-Taylor schemes (44) of order 1, (45) of order  $\frac{3}{2}$  and (46) of order 2, and the affine-RODE-Taylor schemes (62) of order 1, (63) of order 2, and (64) of order 3 are applied to simulate (75).

The global mean-square errors  $E_{n+1}^G$  versus the time steps  $h = 2^{-i}$ ,  $i = 3, 4, 5, 6, 7, 8$  using the RODE-Taylor schemes and the affine-RODE-Taylor schemes are plotted in Figs. 4 and 5, respectively. The dashed reference lines in Fig. 4 have slopes 1 and 2, respectively, and the dashed reference lines in Fig. 5 have slopes 1 and 3, respectively.

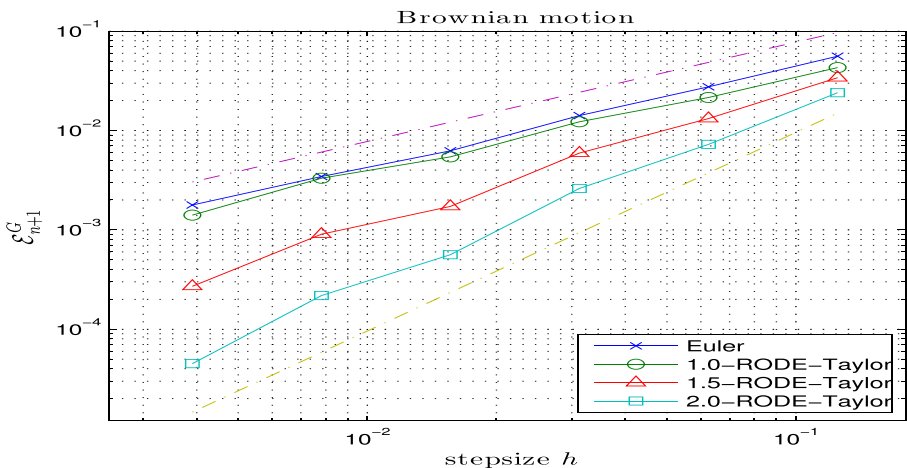


Fig. 4 RODE-Taylor schemes for (75) with Wiener process

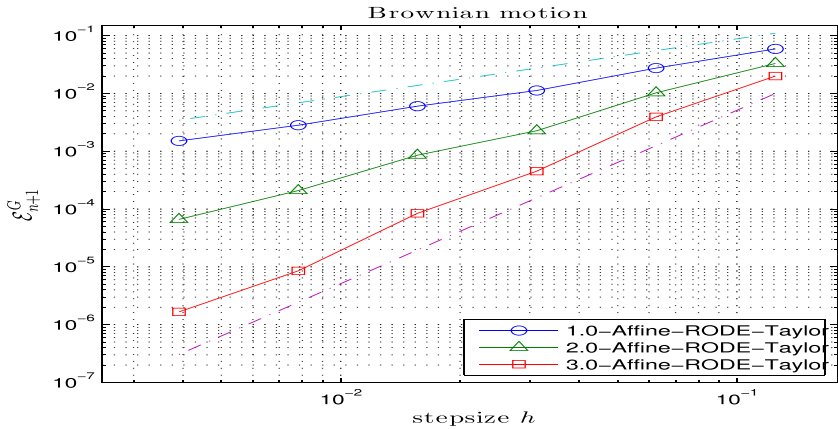


Fig. 5 Affine-RODE-Taylor schemes for (75) with Wiener process

- (ii) Assume that  $\eta_t = \sin B^{\frac{3}{4}}(t)$ , where  $B^{\frac{3}{4}}(t)$  is an fBm with Hurst parameter  $\frac{3}{4}$ . The RODE-Taylor schemes (47) of order  $\frac{3}{4}$ , (48) of order 1, (49) of order  $\frac{3}{2}$ , and (50) of order 2, and the affine-RODE-Taylor schemes (62) of order 1, (63) of order 2, and (64) of order 3 are applied to simulate (75).

The global mean-square errors  $E_{n+1}^G$  versus the time steps  $h = 2^{-i}$ ,  $i = 3, 4, 5, 6, 7, 8$  using the RODE-Taylor schemes and the affine-RODE-Taylor schemes are plotted in Figs. 4 and 5, respectively. The dashed reference lines in Fig. 6 have slopes 0.75 and 2, respectively, and the dashed reference lines in Fig. 7 have slopes 1 and 3, respectively.

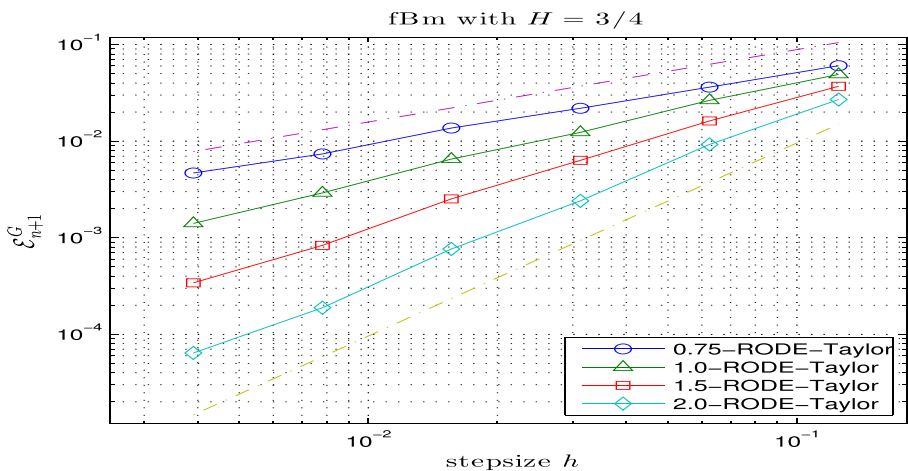


Fig. 6 RODE-Taylor schemes for (75) with fBm with Hurst parameter  $\frac{3}{4}$

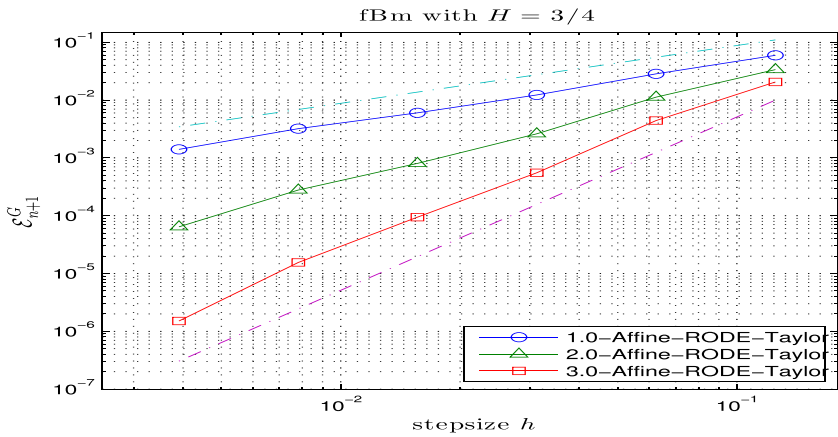


Fig. 7 Affine-RODE-Taylor schemes for (75) with fBm with Hurst parameter  $\frac{3}{4}$

- (iii) Assume that  $\eta_t = \sin B^{\frac{1}{3}}(t)$ , where  $B^{\frac{1}{3}}(t)$  is an fBm with Hurst parameter  $\frac{1}{3}$ . The RODE-Taylor schemes (51) of order 1 and (52) of order 2 and affine-RODE-Taylor schemes (62) of order 1, (63) of order 2, and (64) of order 3 are applied to simulate (75).

The global mean-square errors  $\mathcal{E}_{n+1}^G$  versus the time steps  $h = 2^{-i}$ ,  $i = 3, 4, 5, 6, 7, 8$  using the RODE-Taylor schemes and the affine-RODE-Taylor schemes are plotted in Figs. 8 and 9, respectively. The dashed reference lines in Fig. 8 have slopes 1 and 2, respectively, and the dashed reference lines in Fig. 9 have slopes 1 and 3, respectively.

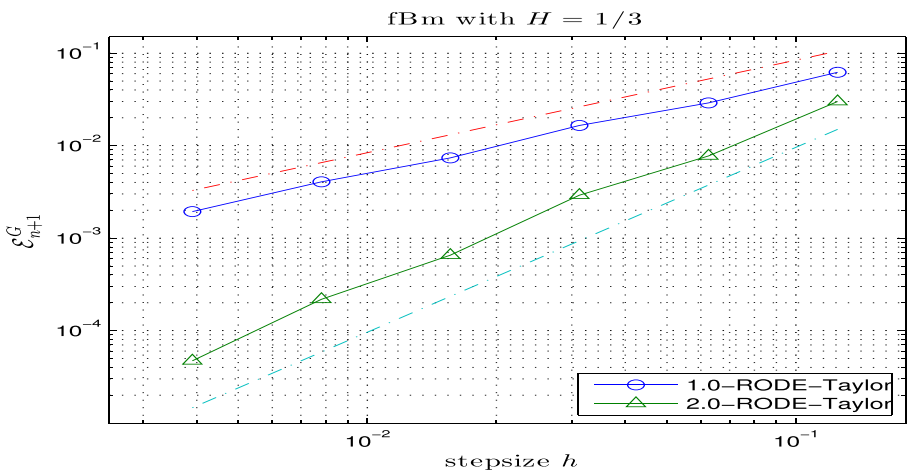


Fig. 8 RODE-Taylor schemes for (75) with fBm with Hurst parameter  $\frac{1}{3}$

*Example 16* Consider the scalar RODE with 2-dimensional affine noise process consisting of one Brownian motion and one fBm:

$$\frac{dy}{dt} = -\cos^2 y + (\cos^2 y)\eta_t^1 + 2(\cos^2 y)\eta_t^2. \tag{76}$$

For each realization  $\eta^{(i)} = (\eta_t^{1(i)}, \eta_t^{2(i)})$  (76) has the explicit solution:

$$y^{(i)}(t; 0, 0) = \tan^{-1} \left( \int_0^t (\eta_s^{1(i)} + 2\eta_s^{2(i)} - 1) ds \right).$$

The initial valued is picked to be  $y_0 = 1$  and the final time is picked to be  $T = 1$ .

Assume that  $\eta_t^1 = \sin W_t$ , where  $W_t$  is a Brownian motion and  $\eta_t^2 = \sin B_t^{\frac{3}{4}}$ , where  $B_t^{\frac{3}{4}}$  is an fBm with Hurst parameter  $\frac{3}{4}$ . The RODE-Taylor schemes (53) of order  $\frac{1}{2}$ , (54) of order 1 and (55) of order 2, and the affine-RODE-Taylor schemes (65) of order 1 and (66) or order 2, are applied to simulate (76).

The global mean-square errors  $\mathcal{E}_{n+1}^G$  versus the time steps  $h = 2^{-i}$ ,  $i = 3, 4, 5, 6, 7, 8$  using the RODE-Taylor schemes and the affine-RODE-Taylor schemes are plotted in Figs. 10 and 11, respectively. The dashed reference lines in Figs. 10 and 11 have slopes 1 and 2, respectively.

*Example 17* Consider the scalar RODE with scalar Itô diffusion noise:

$$\frac{dy}{dt} = -y + \cos \eta_t, \quad y(0) = 1, \tag{77}$$

$$d\eta_t = -\frac{1}{2}a^2\eta_t dt + a\sqrt{1 - \eta_t^2}dW_t, \quad \eta_0 = 1. \tag{78}$$

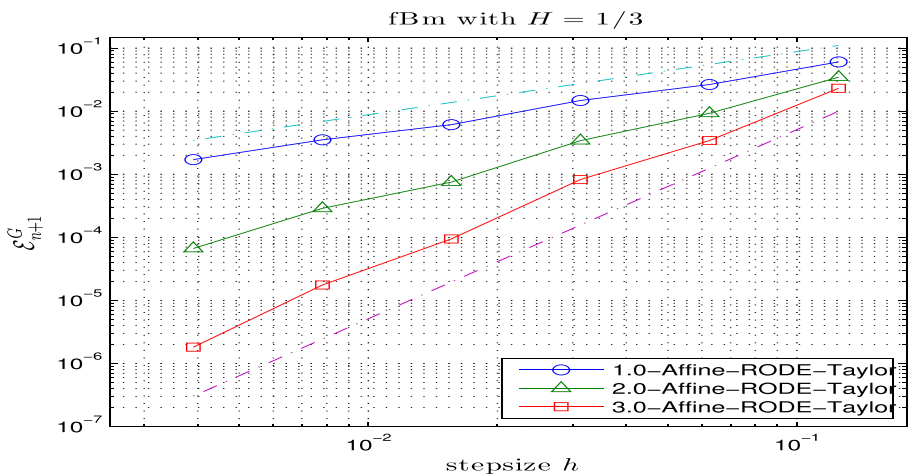


Fig. 9 Affine-RODE-Taylor schemes for (75) with fBm with Hurst parameter  $\frac{1}{3}$



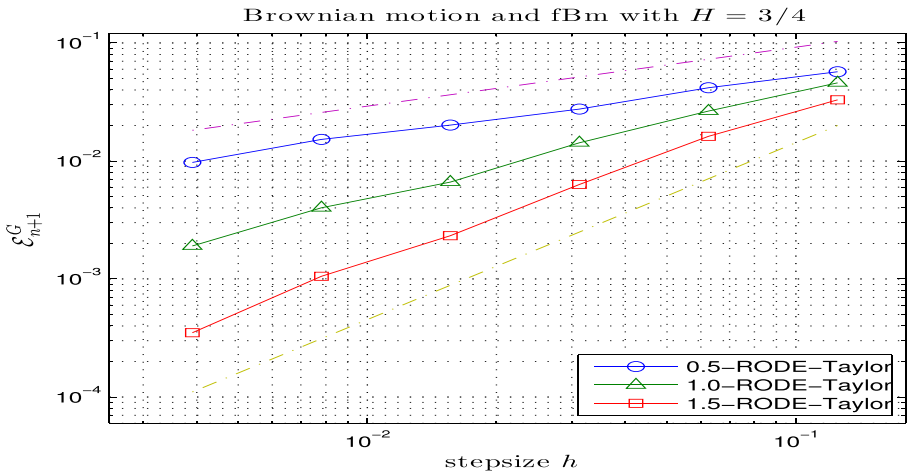


Fig. 10 RODE-Taylor schemes for (76) with Wiener process and fBm

For each realization  $W_t^{(i)}$ , the solution to (78) gives:

$$\eta_t^{(i)} = \sin \left( aW_t^{(i)} + \frac{\pi}{2} \right)$$

and the explicit solution to (77) is then:

$$y^{(i)}(t; 0, 1) = e^{-t} + e^{-t} \int_0^t e^s \cos \left( \sin \left( aW_s^{(i)} + \frac{\pi}{2} \right) \right) ds.$$

The initial value is picked to be  $y_0 = 1$  and  $\eta_0 = 1$ , the final time is picked to be  $T = 1$ , and the parameter  $a = 1$ .

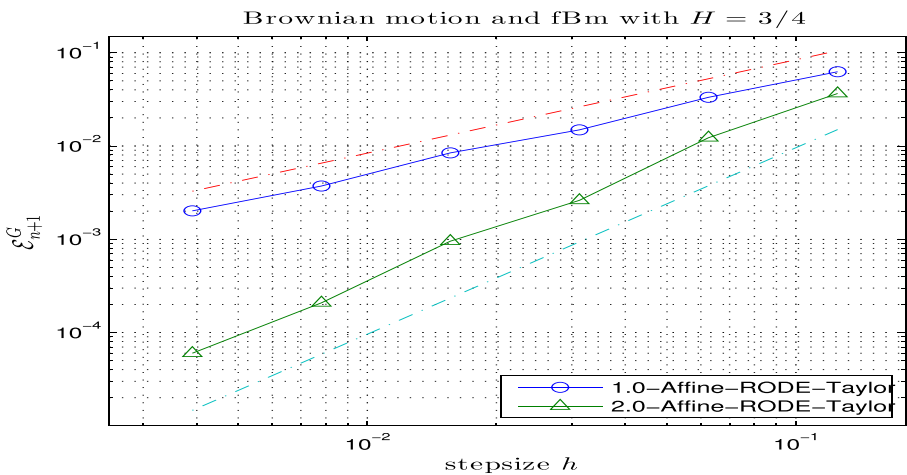
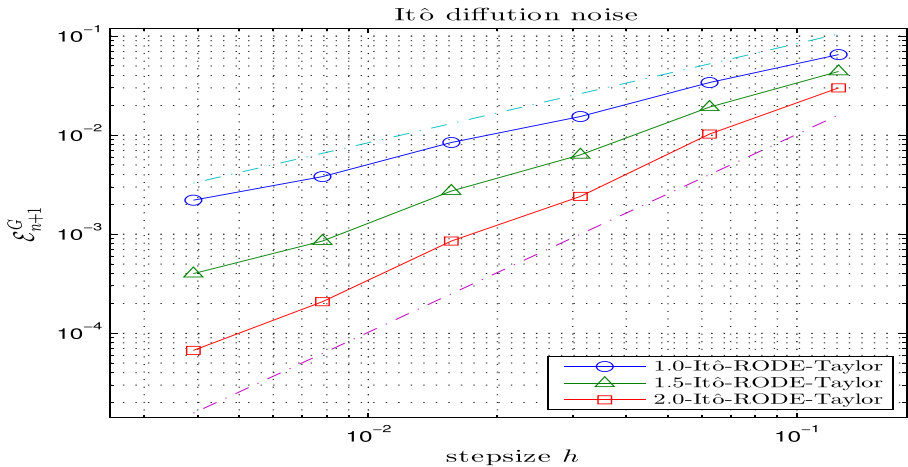


Fig. 11 Affine-RODE-Taylor schemes for (76) with Wiener process and fBm



**Fig. 12** Itô-RODE-Taylor schemes for (77)–(78) with Itô diffusion noise

The Itô-RODE-Taylor schemes (72) of order  $\frac{3}{2}$ , and (73) of order 2, are applied to simulate solutions to (77)–(78). The global mean-square errors  $\mathcal{E}_{n+1}^G$  versus the time steps  $h = 2^{-i}$ ,  $i = 3, 4, 5, 6, 7, 8$  using the Itô-RODE-Taylor schemes are plotted in Fig. 12, where the dashed reference lines have slopes 1 and 2, respectively.

## 7 Closing remarks

The pathwise convergence rates for RODE-Taylor schemes were studied by Jentzen and Kloeden in [10], and the pathwise convergence rates for affine-RODE-Taylor schemes were studied by Asai and Kloeden in [3]. In this work, we first establish a generic theorem on the relation between local and global mean-square convergence rates for RODEs. Then, we investigate the global mean-square convergence rate for each of the RODE-Taylor, affine-RODE-Taylor, and Itô-RODE-Taylor schemes. It appears that the mean-square convergence rate is  $\varepsilon$  order higher than the pathwise convergence rate (see, e.g., [9, 12]) for the same scheme.

**Acknowledgments** The authors would like to express their sincere gratitude to Prof. Dr. Arnulf Jentzen for his insightful comments. The authors would also like to thank the anonymous referees for their insightful comments, which have led to the much improved quality of this work.

**Funding information** Wang is partially supported by the National Natural Science Foundation of China (11101184, 11271151), Scientific and Technological Development Plan of Jilin Province (20200201251JC) and China Scholarship Council (201506175081). Cao is partially supported by the National Science Foundation (DMS1620150), Han is partially supported by Simons Foundation (Collaboration Grants for Mathematicians 429717), and Kloeden is partially supported by the National Natural Science Foundation of China (11571125).

## References

1. Asai, Y., Kloeden, P.E.: Numerical schemes for random ODES via stochastic differential equations. *Commun. Appl. Anal.* **17**, 511–528 (2013)
2. Asai, Y., Kloeden, P.E.: Multi-step methods for random ODES driven by Itô diffusions. *J. Comput. Appl. Math.* **294**, 210–224 (2016)
3. Asai, Y., Kloeden, P.E.: Numerical schemes for random ODES with affine noise. *Numer. Algorithms* **72**, 155–171 (2016)
4. Asai, Y., Herrmann, E., Kloeden, P.E.: Stable integration of stiff random ordinary differential equations. *Stoch. Anal. Appl.* **31**, 293–313 (2013)
5. Bunke, H.: *Gewöhnliche Differentialgleichungen mit zufälligen Parametern*. Akademie-Verlag, Berlin (1972)
6. Chen, C., Hong, J.: Mean-square convergence of numerical approximations for a class of backward stochastic differential equations. *Discrete Contin. Dyn. Syst. Ser. B* **8**, 2051–2067 (2013)
7. Cortes, J., Jordan, L., Villafuerte, L.: Numerical solution of random differential equations, a mean square approach. *Math. Comput. Model.* **45**, 757–765 (2007)
8. Grüne, L., Kloeden, P.E.: Pathwise approximation of random ordinary differential equations. *BIT* **41**, 711–721 (2001)
9. Han, X., Kloeden, P.E.: *Random Ordinary Differential Equations and Their Numerical Solution*. Springer Nature, Singapore (2017)
10. Jentzen, A., Kloeden, P.E.: Pathwise Taylor schemes for random ordinary differential equations. *BIT Numer. Math.* **49**, 113–140 (2009)
11. Jentzen, A., Kloeden, P.E.: *Taylor Approximations of Stochastic Partial Differential Equations*. SIAM, Philadelphia (2011)
12. Kloeden, P.E., Jentzen, A.: Pathwise convergent higher order numerical schemes for random ordinary. *Proc. R. Soc. A* **463**, 2929–2944 (2007)
13. Kloeden, P.E., Platen, E.: *Numerical Solution of Stochastic Differential Equations*. Springer, Heidelberg (1992)
14. Milstein, G.N.: *Numerical Integration of Stochastic Differential Equations*. Kluwer Academic Publishers Group, Dordrecht (1995)
15. Milstein, G.N.: *Tretyakov: Stochastic Numerics for Mathematical Physics*. Springer, New York (2004)
16. Stuart, A.M., Humphries, A.R.: *Dynamical Systems and Numerical Analysis*. Cambridge University Press, Cambridge (1996)

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.