A fundamental mean-square convergence theorem for SDEs with locally Lipschitz coefficients and its applications

M.V. Tretyakov[†] and Z. Zhang[‡]
[†]School of Mathematical Sciences, University of Nottingham,
Nottingham, NG7 2RD, UK
Email: Michael.Tretyakov@nottingham.ac.uk
[‡]Division of Applied Mathematics, Brown University, Providence RI, 02912
Email: Zhongqiang_Zhang@brown.edu

July 25, 2018

Abstract

A version of the fundamental mean-square convergence theorem is proved for stochastic differential equations (SDE) which coefficients are allowed to grow polynomially at infinity and which satisfy a one-sided Lipschitz condition. The theorem is illustrated on a number of particular numerical methods, including a special balanced scheme and fully implicit methods. Some numerical tests are presented.

AMS 2000 subject classification. Primary 60H35; secondary 65C30, 60H10.

Keywords. SDEs with nonglobally Lipschitz coefficients, numerical integration of SDEs in the mean-square sense; balanced methods; fully implicit methods; strong convergence; almost sure convergence.

1 Introduction

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(w(t), \mathcal{F}_t^w) = ((w_1(t), \ldots, w_m(t))^\top, \mathcal{F}_t^w)$ be an m-dimensional standard Wiener process, where $\mathcal{F}_t^w$, $0 \le t \le T$, is an increasing family of σ-subalgebras of $\mathcal{F}$ induced by $w(t)$. We consider the system of Ito stochastic differential equations (SDE):

$$dX = a(t,X)dt + \sum_{r=1}^{m} \sigma_r(t,X)dw_r(t), \quad t \in (t_0, T], \ X(t_0) = X_0, \tag{1.1}$$

where X, a, σ_r are d-dimensional column-vectors and X_0 is independent of w. We suppose that any solution $X_{t_0,X_0}(t)$ of (1.1) is regular on $[t_0, T]$. We recall [3] that a process is called regular if it is defined for all $t_0 \le t \le T$.

In traditional numerical analysis for SDE [12, 9, 15] it is assumed that the SDE coefficients are globally Lipschitz which is a significant limitation taking into account that most of the models of applicable interest have coefficients which grow faster at infinity than a linear function. If the global Lipschitz condition is violated, the convergence of many usual numerical methods can disappear (see, e.g., [22, 5, 6, 16]). This has been the motivation for the recent interest in both theoretical support of existing numerical methods and developing new methods or approaches for solving SDE under nonglobal Lipschitz assumptions on the coefficients.

In most of SDE applications (e.g., in molecular dynamics, financial engineering and other problems of mathematical physics), one is interested in simulating averages $\mathbb{E}\varphi(X(T))$ of the solution to SDE – the task for which the weak-sense SDE approximation is sufficient and effective [12, 15]. The problem with divergence of weak-sense schemes was addressed in [16] (see also [17]) for simulation of averages at finite time and also of ergodic limits when ensemble averaging is used. The concept of rejecting exploding trajectories proposed and justified in [16] allows us to use any numerical method for solving SDE with nonglobally Lipschitz coefficients for estimating averages. Following this concept, we do not take into account the approximate trajectories $X(t)$ which leave a sufficiently large ball $S_R := \{x : |x| < R\}$ during the time T. See other approaches for resolving this problem in the context of computing averages, including the case of simulating ergodic limits via time averaging, e.g. in [22, 10, 1].

In this paper we deal with mean-square (strong) approximation of SDE with nonglobal Lipschitz coefficients. Mean-square schemes have their own area of applicability (e.g. for simulating scenarios, visualization of stochastic dynamics, filtering, etc., see further discussion on this in [9, 15, 7] and references therein). Furthermore, mean-square approximation is of theoretical interest and it also provides fundamental insight for weak-sense schemes.

We note that in the case of weak approximation we often have to simulate large dimensional complicated stochastic systems using the Monte Carlo technique (or time averaging), which is typical for molecular dynamics applications, or we have to perform calculations on a daily basis, which is usual, e.g., in financial applications. Hence the cost per step of a weak numerical integrator should be low, which, in particular, essentially prohibits the use of implicit methods. In contrast, areas of applicability of mean-square schemes, as a rule, do not involve simulation of a large number of trajectories or over very long time periods and, consequently, there are more relaxed requirements on the cost per step of mean-square schemes and efficient and reliable implicit schemes have practical interest. There have been a number of recent works, including [6, 5, 8, 21, 7] (see also the references therein), where strong schemes for SDE with nonglobal Lipschitz coefficients were considered. An extended literature review on this topic is available in [7].

In this paper we give a variant of the fundamental mean-square convergence theorem in the case of SDE with nonglobal Lipschitz coefficients, which is analogous to Milstein's fundamental theorem for the global Lipschitz case [11] (see also [12, 15]). More precisely, we assume that the SDE coefficients can grow polynomially at infinity and satisfy a one-sided Lipschitz condition. The theorem is stated in Section 2 and proved in Appendix A. Its corollary on almost sure convergence is also given. In Section 2 we start discussion on applicability of the fundamental theorem, including its application to the drift-implicit Euler scheme and thus establish its order of convergence. Strong convergence (but without order) of this scheme was proved for SDE with nonglobal Lipschitz drift and diffusion in [21, 7]. A particular balanced method (see the class of balanced methods in [13, 15]) is proposed and its convergence with order 1/2 in the nonglobal Lipschitz setting is proved in Section 3. In Section 4 we revisit fully implicit (i.e., implicit both in drift and diffusion) mean-square schemes proposed in [14] (see also [15]). In [14, 15] their convergence was proved for SDE with globally Lipschitz coefficients. Here we relax these conditions as the drift is required to satisfy only a one-sided Lipschitz condition and be of not faster than polynomial growth at infinity. Some numerical experiments supporting our results are presented in Section 5.

2 Fundamental theorem

Let $X_{t_0,X_0}(t) = X(t)$, $t_0 \le t \le T$, be a solution of the system (1.1). In what follows we will assume the following.

Assumption 2.1. *(i) The initial condition is such that*

$$\mathbb{E}|X_0|^{2p} \le K < \infty, \quad \text{for all } \ p \ge 1. \tag{2.1}$$

(ii) For a sufficiently large $p_0 \ge 1$ there exists a constant $c_1 \ge 0$ such that

$$(x-y, a(t,x)-a(t,y)) + \frac{2p_0-1}{2}\sum_{r=1}^{m}|\sigma_r(t,x)-\sigma_r(t,y)|^2 \le c_1|x-y|^2,\ t \in [t_0,T],\ x,y \in \mathbb{R}^d. \tag{2.2}$$

(iii) There exist $c_2 \ge 0$ and $\varkappa \ge 1$ such that

$$|a(t,x)-a(t,y)|^2 \le c_2(1+|x|^{2\varkappa-2}+|y|^{2\varkappa-2})|x-y|^2, \quad t \in [t_0,T],\ x,y \in \mathbb{R}^d. \tag{2.3}$$

We note that (2.2) implies that

$$(x, a(t,x)) + \frac{2p_0-3}{2}\sum_{r=1}^{m}|\sigma_r(t,x)|^2 \le c_0 + c_1'|x|^2, \quad t \in [t_0,T],\ x \in \mathbb{R}^d, \tag{2.4}$$

where $c_0 = |a(t,0)|^2/2 + \frac{(2p_0-3)(2p_0-1)}{4}\sum_{r=1}^{m}|\sigma_r(t,0)|^2$ and $c_1' = c_1 + 1/2$. The inequality (2.4) together with (2.1) is sufficient to ensure finiteness of moments [3]: there is $K > 0$

$$\mathbb{E}|X_{t_0,X_0}(t)|^{2p} < K(1+\mathbb{E}|X_0|^{2p}),\ 1 \le p \le p_0-1, \quad t \in [t_0,T]. \tag{2.5}$$

Also, (2.3) implies that

$$|a(t,x)|^2 \le c_3 + c_2'|x|^{2\varkappa}, \quad t \in [t_0,T],\ x \in \mathbb{R}^d, \tag{2.6}$$

where $c_3 = 2|a(t,0))|^2 + 2c_2(\varkappa-1)/\varkappa$ and $c_2' = 2c_2(1+\varkappa)/\varkappa$.

Introduce the one-step approximation $\bar{X}_{t,x}(t+h)$, $t_0 \le t < t+h \le T$, for the solution $X_{t,x}(t+h)$ of (1.1), which depends on the initial point (t,x), a time step h, and $\{w_1(\theta) - w_1(t), \ldots, w_m(\theta) - w_m(t),\ t \le \theta \le t+h\}$ and which is defined as follows:

$$\bar{X}_{t,x}(t+h) = x + A(t,x,h; w_i(\theta) - w_i(t),\ i = 1,\ldots,m,\ t \le \theta \le t+h). \tag{2.7}$$

Using the one-step approximation (2.7), we recurrently construct the approximation $(X_k, \mathcal{F}_{t_k}),\ k = 0,\ldots,N,\ t_{k+1} - t_k = h_{k+1},\ T_N = T:$

$$X_0 = X(t_0),\ X_{k+1} = \bar{X}_{t_k,\bar{X}_k}(t_{k+1}) \tag{2.8}$$
$$= X_k + A(t_k, X_k, h_{k+1}; w_i(\theta) - w_i(t_k),\ i = 1,\ldots,m,\ t_k \le \theta \le t_{k+1}).$$

The following theorem is a generalization of Milstein's fundamental theorem [11] (see also [12, 15, Chapter 1]) from the global to nonglobal Lipschitz case. It also has similarities with a strong convergence theorem in [5] proved for the case of nonglobal Lipschitz drift, global Lipschitz diffusion and Euler-type schemes.

For simplicity, we will consider a uniform time discretization, i.e. $h_k = h$ for all k.

Theorem 2.1. *Suppose*

(i) Assumption 2.1 holds;

(ii) The one-step approximation $\bar{X}_{t,x}(t+h)$ from (2.7) has the following orders of accuracy: for some $p \ge 1$ there are $\alpha \ge 1$, $h_0 > 0$, and $K > 0$ such that for arbitrary $t_0 \le t \le T - h$, $x \in \mathbf{R}^d$, and all $0 < h \le h_0$:

$$|\mathbb{E}[X_{t,x}(t+h) - \bar{X}_{t,x}(t+h)]| \le K(1 + |x|^{2\alpha})^{1/2} h^{q_1}\,, \tag{2.9}$$

$$\left[\mathbb{E}|X_{t,x}(t+h) - \bar{X}_{t,x}(t+h)|^{2p}\right]^{1/(2p)} \le K(1 + |x|^{2\alpha p})^{1/(2p)} h^{q_2} \tag{2.10}$$

with

$$q_2 \ge \frac{1}{2}\,,\ q_1 \ge q_2 + \frac{1}{2}\,; \tag{2.11}$$

(iii) The approximation X_k from (2.8) has finite moments, i.e., for some $p \ge 1$ there are $\beta \ge 1$, $h_0 > 0$, and $K > 0$ such that for all $0 < h \le h_0$ and all $k = 0,\ldots,N$:

$$\mathbb{E}|X_k|^{2p} < K(1 + \mathbb{E}|X_0|^{2p\beta}). \tag{2.12}$$

Then for any N and $k = 0,1,\ldots,N$ the following inequality holds:

$$\left[\mathbb{E}|X_{t_0,X_0}(t_k) - \bar{X}_{t_0,X_0}(t_k)|^{2p}\right]^{1/(2p)} \le K(1 + \mathbb{E}|X_0|^{2\gamma p})^{1/(2p)} h^{q_2 - 1/2}\,, \tag{2.13}$$

where $K > 0$ and $\gamma \ge 1$ do not depend on h and k, i.e., the order of accuracy of the method (2.8) is $q = q_2 - 1/2$.

The theorem is proved in Appendix A and it uses the following lemma.

Lemma 2.1. *Suppose Assumption 2.1 holds. For the representation*

$$X_{t,x}(t+\theta) - X_{t,y}(t+\theta) = x - y + Z_{t,x,y}(t+\theta), \tag{2.14}$$

we have for $1 \leq p \leq (p_0 - 1)/\varkappa$ *:*

$$\mathbb{E}|X_{t,x}(t+h) - X_{t,y}(t+h)|^{2p} \leq |x-y|^{2p}(1+Kh)\,, \tag{2.15}$$

$$\mathbb{E}\,|Z_{t,x,y}(t+h)|^{2p} \leq K(1+|x|^{2\varkappa-2}+|y|^{2\varkappa-2})^{p/2}|x-y|^{2p}h^{p}\,. \tag{2.16}$$

This lemma is proved in Appendix B. Theorem 2.1 has the following corollary.

Corollary 2.1. *In the setting of Theorem 2.1 for* $p \geq 1/(2q)$ *in* (2.13)*, there is* $0 < \varepsilon < q$ *and an a.s. finite random variable* $C(\omega) > 0$ *such that*

$$|X_{t_0,X_0}(t_k) - X_k| \leq C(\omega)h^{q-\varepsilon},$$

i.e., the method (2.8) *for* (1.1) *converges with order* $q - \varepsilon$ *a.s.*

The corollary is proved using the Borel-Cantelli-type of arguments (see, e.g. [2, 18]).

2.1 Discussion

In this section we make a number of observations concerning Theorem 2.1.

1. As a rule, it is not difficult to check the conditions (2.9)-(2.10) following the usual routine calculations as in the global Lipschitz case [12, 9, 15]. We note that in order to achieve the optimal q_1 and q_2 in (2.9)-(2.10) additional assumptions on smoothness of $a(t,x)$ and $\sigma_r(t,x)$ are usually needed.

In contrast to the conditions (2.9)-(2.10), checking the condition (2.12) on moments of a method X_k is often rather difficult. In the case of global Lipschitz coefficients, boundedness of moments of X_k is just direct implication of the boundedness of moments of the SDE solution and the one-step properties of the method (see [15, Lemma 1.1.5]). There is no result of this type in the case of nonglobal Lipschitz SDE and each scheme requires a special consideration. For a number of strong schemes boundedness of moments in nonglobal Lipschitz cases were proved (see, e.g. [6, 5, 8, 7, 22]). In Section 3 we show boundedness of moments for a balanced method and in Section 4 for fully implicit methods.

Roughly speaking, Theorem 2.1 says that if moments of X_k are bounded and the scheme was proved to be convergent with order q in the global Lipschitz case then the scheme has the same convergence order q in the considered nonglobal Lipschitz case.

2. Assumptions and the statement of Theorem 2.1 include the famous fundamental theorem of Milstein [11] proved under the global conditions on the SDE coefficients (of course, as discussed in the previous point, this case does not need the assumption (2.12)). Though the main focus here is on cases when drift and diffusion can grow faster than a linear function at infinity, we note that the assumptions also include the case when the diffusion coefficient grows slower than linear function at infinity, e.g. they cover so-called CIR process which is used in modelling short interest rates and stochastic volatility in financial engineering.

3. Consider the drift-implicit scheme [15, p. 30]:

$$X_{k+1} = X_k + a(t_{k+1}, X_{k+1})h + \sum_{r=1}^{m} \sigma_r(t_k, X_k)\xi_{rk}\sqrt{h}, \tag{2.17}$$

where $\xi_{rk} = (w_r(t_{k+1}) - w_r(t_k))/\sqrt{h}$ are Gaussian $\mathcal{N}(0,1)$ i.i.d. random variables. Assume that the coefficients $a(t,x)$ and $\sigma_r(t,x)$ have continuous first-order partial derivatives in t and the coefficient $a(t,x)$ also has continuous first-order partial derivatives in x^i and that all these derivatives and the coefficients themselves satisfy inequalities of the form (2.3). It is not difficult to show that the one-step approximation corresponding to (2.17) satisfies (2.9) and (2.10) with $q_1 = 2$ and $q_2 = 1$, respectively. Its boundedness of moments, in particular, under the condition (2.4) for time steps $h \leq 1/(2c_1)$, is proved in [7]. Then, due to Theorem 2.1, (2.17) converges with mean-square order $q = 1/2$ (note that for $q = 1/2$, it is sufficient to have $q_1 = 3/2$ which can be obtained under lesser smoothness of a). Further, in the case of additive noise (i.e., $\sigma_r(t,x) = \sigma_r(t)$, $r = 1, \ldots, m$), $q_1 = 2$ and $q_2 = 3/2$ and (2.17) converges with mean-square order 1 due to Theorem 2.1. We note that convergence of (2.17) with order 1/2 in the global Lipschitz case is well known [12, 9, 15]; in the case of nonglobal Lipschitz drift and global Lipschitz diffusion was proved in [6, 5] (see also related results in [2, 22]); and its strong convergence without order under Assumption 2.1 was proved in [21, 7].

5. Due to the bound (2.5) on the moments of the solution $X(t)$, it would be natural to require that β in (2.12) should be equal to 1. Indeed, (2.12) with $\beta = 1$ holds for the drift-implicit method (2.17) [7] and for fully implicit methods (see Section 4). However, this is not the case for tamed-type methods (see [8]) or the balanced method from Section 3.

6. The constant K in (2.13) depends on p, t_0, T as well as on the SDE coefficients. The constant γ in (2.13) depends on α, β and $\varkappa$.

3 A balanced method

In this section we introduce a particular balanced scheme from the class of balanced methods introduced in [13] (see also [15]) and prove its mean-square convergence with order 1/2 using Theorem 2.1. As far as we know, this variant of balanced schemes has not been considered before. In Section 5 we test the balanced scheme on a model problem and demonstrate that it is more efficient than the tamed scheme (5.2) (see Section 5) from [7]. We also note that it was mentioned in [7] that a balanced scheme suitable for the nonglobal Lipschitz case could potentially be derived.

Consider the following balanced-type scheme for (1.1):

$$X_{k+1} = X_k + \frac{a(t_k, X_k)h + \sum_{r=1}^{m} \sigma_r(t_k, X_k)\xi_{rk}\sqrt{h}}{1 + h|a(t_k, X_k)| + \sqrt{h}\sum_{r=1}^{m} |\sigma_r(t_k, X_k)\xi_{rk}|}, \tag{3.1}$$

where ξ_{rk} are Gaussian $\mathcal{N}(0,1)$ i.i.d. random variables.

We prove two lemmas which show that the scheme (3.1) satisfies the conditions of Theorem 2.1. The first lemma is on boundedness of moments, which uses a stopping time technique (see also, e.g. [16, 7]).

Lemma 3.1. *Suppose Assumption 2.1 holds with sufficiently large p_0. For all natural N and all $k = 0, \dots, N$ the following inequality holds for moments of the scheme* (3.1)*:*

$$\mathbb{E}|X_k|^{2p} \le K(1 + \mathbb{E}|X_0|^{2p\beta}), \quad 1 \le p \le \frac{p_0 - 1}{4(3\varkappa - 2)} - \frac{1}{2}, \tag{3.2}$$

with some constants $\beta \ge 1$ and $K > 0$ independent of h and k.

Proof. In the proof we shall use the letter K to denote various constants which are independent of h and k. We note in passing that the case $\varkappa = 1$ (i.e., when $a(t,x)$ is globally Lipschitz) is trivial.

The following elementary consequence of the inequalities (2.4) and (2.6) will be used in the proof: for any $C_1 > 0$ and $C_2 > 0$:

$$\begin{aligned} C_1 \sum_{r=1}^{m} |\sigma_r(t,x)|^2 &\le (2C_1 c + \frac{C_1^2}{C_2})(1 + |x|^2) + C_2 |a(t,x)|^2 \\ &\le (2C_1 c + \frac{C_1^2}{C_2})(1 + |x|^2) + C_2 c \left(1 + |x|^{2\varkappa}\right), \end{aligned} \tag{3.3}$$

where $c = \max(c_0, c_1', c_2', c_3)$.

We observe that

$$|X_{k+1}| \le |X_k| + 1 \le |X_0| + (k+1). \tag{3.4}$$

Let $R > 0$ be a sufficiently large number. Introduce the events

$$\tilde{\Omega}_{R,k} := \{\omega : |X_l| \le R, \ l = 0, \dots, k\}, \tag{3.5}$$

and their compliments $\tilde{\Lambda}_{R,k}$. We first prove the lemma for integer $p \ge 1$. We have

$$\begin{aligned} &\mathbb{E}\chi_{\tilde{\Omega}_{R,k+1}}(\omega)|X_{k+1}|^{2p} \le \mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)|X_{k+1}|^{2p} = \mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)|(X_{k+1} - X_k) + X_k|^{2p} \\ \le\ &\mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)|X_k|^{2p} + \mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)\,|X_k|^{2p-2}\left[2p(X_k, X_{k+1} - X_k) + p(2p-1)|X_{k+1} - X_k|^2\right] \\ &+K \sum_{l=3}^{2p} \mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)\,|X_k|^{2p-l}\,|X_{k+1} - X_k|^l. \end{aligned} \tag{3.6}$$

Consider the second term in the right-hand side of (3.6):

$$\begin{aligned} &\mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)\,|X_k|^{2p-2}\left[2p(X_k, X_{k+1} - X_k) + p(2p-1)|X_{k+1} - X_k|^2\right] \\ = \ &2p\mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)\,|X_k|^{2p-2}\,\mathbb{E}\left[\left(X_k, \frac{a(t_k, X_k)h + \sum_{r=1}^m \sigma_r(t_k, X_k)\xi_{rk}\sqrt{h}}{1 + h|a(t_k, X_k)| + \sqrt{h}\sum_{r=1}^m |\sigma_r(t_k, X_k)\xi_{rk}|}\right)\right. \\ &\left. + \frac{2p-1}{2}\left|\frac{a(t_k, X_k)h + \sum_{r=1}^m \sigma_r(t_k, X_k)\xi_{rk}\sqrt{h}}{1 + h|a(t_k, X_k)| + \sqrt{h}\sum_{r=1}^m |\sigma_r(t_k, X_k)\xi_{rk}|}\right|^2 \middle|\, \mathcal{F}_{t_k}\right]. \end{aligned} \tag{3.7}$$

Since

$$\mathbb{E}\left[\frac{\sum_{r=1}^m \sigma_r(t_k,X_k)\xi_{rk}\sqrt{h}}{1+h|a(t_k,X_k)|+\sqrt{h}\sum_{r=1}^m|\sigma_r(t_k,X_k)\xi_{rk}|}\middle|\mathcal{F}_{t_k}\right]=0 \tag{3.8}$$

and for $l \neq r$

$$\mathbb{E}\left[\frac{\sigma_r(t_k,X_k)\xi_{rk}\sqrt{h}\sigma_l(t_k,X_k)\xi_{lk}\sqrt{h}}{(1+h|a(t_k,X_k)|+\sqrt{h}\sum_{r=1}^m|\sigma_r(t_k,X_k)\xi_{rk}|)^2}\middle|\mathcal{F}_{t_k}\right]=0, \tag{3.9}$$

the conditional expectation in (3.7) becomes

$$\begin{aligned}
A \quad : \quad &= \mathbb{E}\Bigg[\left(X_k,\frac{a(t_k,X_k)h+\sum_{r=1}^m \sigma_r(t_k,X_k)\xi_{rk}\sqrt{h}}{1+h|a(t_k,X_k)|+\sqrt{h}\sum_{r=1}^m|\sigma_r(t_k,X_k)\xi_{rk}|}\right) \qquad (3.10)\\
&\quad +\frac{2p-1}{2}\left|\frac{a(t_k,X_k)h+\sum_{r=1}^m \sigma_r(t_k,X_k)\xi_{rk}\sqrt{h}}{1+h|a(t_k,X_k)|+\sqrt{h}\sum_{r=1}^m|\sigma_r(t_k,X_k)\xi_{rk}|}\right|^2\Bigg|\mathcal{F}_{t_k}\Bigg]\\
&= \mathbb{E}\Bigg[\frac{(X_k,a(t_k,X_k)h)}{1+h|a(t_k,X_k)|+\sqrt{h}|\sum_{r=1}^m\sigma_r(t_k,X_k)\xi_{rk}|}\\
&\quad +\frac{2p-1}{2}\frac{a^2(t_k,X_k)h^2+h\sum_{r=1}^m(\sigma_r(t_k,X_k)\xi_{rk})^2}{\left(1+h|a(t_k,X_k)|+\sqrt{h}|\sum_{r=1}^m\sigma_r(t_k,X_k)\xi_{rk}|\right)^2}\Bigg|\mathcal{F}_{t_k}\Bigg]\\
&\le \mathbb{E}\Bigg[\frac{(X_k,a(t_k,X_k)h)}{1+h|a(t_k,X_k)|+\sqrt{h}\sum_{r=1}^m|\sigma_r(t_k,X_k)\xi_{rk}|}\\
&\quad +\frac{2p-1}{2}\frac{h\sum_{r=1}^m|\sigma_r(t_k,X_k)|^2\xi_{rk}^2}{1+h|a(t_k,X_k)|+\sqrt{h}\sum_{r=1}^m|\sigma_r(t_k,X_k)\xi_{rk}|}\Bigg|\mathcal{F}_{t_k}\Bigg]+\frac{2p-1}{2}a^2(t_k,X_k)h^2\\
&= \mathbb{E}\Bigg[\frac{(X_k,a(t_k,X_k)h)+\frac{2p-1}{2}h\sum_{r=1}^m|\sigma_r(t_k,X_k)|^2}{1+h|a(t_k,X_k)|+\sqrt{h}\sum_{r=1}^m|\sigma_r(t_k,X_k)\xi_{rk}|}\\
&\quad +\frac{2p-1}{2}\frac{h\sum_{r=1}^m|\sigma_r(t_k,X_k)|^2(\xi_{rk}^2-1)}{1+h|a(t_k,X_k)|+\sqrt{h}\sum_{r=1}^m|\sigma_r(t_k,X_k)\xi_{rk}|}\Bigg|\mathcal{F}_{t_k}\Bigg]+\frac{2p-1}{2}a^2(t_k,X_k)h^2.
\end{aligned}$$

Using (2.4) and (2.6), we obtain

$$\begin{aligned}
A \quad \le \quad & c_0h+c_1'|X_k|^2h \qquad (3.11)\\
&+\frac{2p-1}{2}h\sum_{r=1}^m|\sigma_r(t_k,X_k)|^2\mathbb{E}\left[\frac{(\xi_{rk}^2-1)}{1+h|a(t_k,X_k)|+\sqrt{h}\sum_{r=1}^m|\sigma_r(t_k,X_k)\xi_{rk}|}\middle|\mathcal{F}_{t_k}\right]\\
&+Kh^2+K|X_k|^{2\varkappa}h^2.
\end{aligned}$$

For the expectation in the second term in (3.11), we obtain

$$\mathbb{E}\left[\left.\frac{(\xi_{rk}^2-1)}{1+h|a(t_k,X_k)|+\sqrt{h}\sum_{r=1}^m|\sigma_r(t_k,X_k)\xi_{rk}|}\right|\mathcal{F}_{t_k}\right] \tag{3.12}$$

$$= \mathbb{E}\left[\left.(\xi_{rk}^2-1)\left[1-\frac{h|a(t_k,X_k)|+\sqrt{h}|\sum_{l=1}^m\sigma_l(t_k,X_k)\xi_{lk}|}{1+h|a(t_k,X_k)|+\sqrt{h}\sum_{l=1}^m|\sigma_l(t_k,X_k)\xi_{lk}|}\right]\right|\mathcal{F}_{t_k}\right]$$

$$= -\mathbb{E}\left[\left.(\xi_{rk}^2-1)\frac{h|a(t_k,X_k)|+\sqrt{h}|\sum_{r=1}^m\sigma_l(t_k,X_k)\xi_{lk}|}{1+h|a(t_k,X_k)|+\sqrt{h}\sum_{l=1}^m|\sigma_r(t_k,X_k)\xi_{lk}|}\right|\mathcal{F}_{t_k}\right]$$

$$\leq Kh|a(t_k,X_k)|+K\sqrt{h}\sum_{r=1}^m|\sigma_r(t_k,X_k)|.$$

Using (2.6) and (3.3), we obtain from (3.11)-(3.12):

$$\begin{aligned} A \leq\ & c_0h+c_1'|X_k|^2h+Kh\sum_{r=1}^m|\sigma_r(t_k,X_k)|^2\left[h|a(t_k,X_k)|+\sqrt{h}\sum_{r=1}^m|\sigma_r(t_k,X_k)|\right] \\ & +Kh^2+K|X_k|^{2\varkappa}h^2 \\ \leq\ & Kh(1+|X_k|^2+|X_k|^{2\varkappa}h+|X_k|^{3\varkappa}h^{1/2}) \leq Kh(1+|X_k|^2+|X_k|^{3\varkappa}h^{1/2}). \end{aligned} \tag{3.13}$$

Now consider the last term in (3.6):

$$\begin{aligned} & \mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)\,|X_k|^{2p-l}\,|X_{k+1}-X_k|^l \\ \leq\ & K\mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)\,|X_k|^{2p-l}\left[h^l|a(t_k,X_k)|^l+h^{l/2}\sum_{r=1}^m|\sigma_r(t_k,X_k)|^l|\xi_{rk}|^l\right] \\ \leq\ & K\mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)\,|X_k|^{2p-l}\,h^{l/2}\left[1+|X_k|^{l\varkappa}\right], \end{aligned} \tag{3.14}$$

where we used (2.6) and (3.3) again as well as the fact that $\chi_{\tilde{\Omega}_{R,k}}(\omega)$ and X_k are $\mathcal{F}_{t_k}$-measurable while ξ_{rk} are independent of $\mathcal{F}_{t_k}$.

Combining (3.6), (3.7), (3.10), (3.13) and (3.14), we obtain

$$\begin{aligned} & \mathbb{E}\chi_{\tilde{\Omega}_{R,k+1}}(\omega)|X_{k+1}|^{2p} \\ \leq\ & \mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)|X_k|^{2p}+Kh\mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)\,|X_k|^{2p-2}\left[1+|X_k|^2+|X_k|^{3\varkappa}h^{1/2}\right] \\ & +K\sum_{l=3}^{2p}\mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)\,|X_k|^{2p-l}\,h^{l/2}\left[1+|X_k|^{l\varkappa}\right] \\ \leq\ & \mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)|X_k|^{2p}+Kh\mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)\,|X_k|^{2p}+K\sum_{l=2}^{2p}\mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)\,|X_k|^{2p-l}\,h^{l/2} \\ & +Kh^{3/2}\mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)\,|X_k|^{2p-2+3\varkappa}+Kh\sum_{l=3}^{2p}\mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)\,|X_k|^{2p+l(\varkappa-1)}\,h^{l/2-1}. \end{aligned} \tag{3.15}$$

Choosing

$$R=R(h)=h^{-1/(6\varkappa-4)}, \tag{3.16}$$

we get $\mathbb{E}\chi_{\tilde{\Omega}_{R,k}}(\omega)\left|X_k\right|^{2p-2+3\varkappa}h^{l/2-1} \leq \chi_{\tilde{\Omega}_{R(h),k}}(\omega)\left|X_k\right|^{2p}$ and $\chi_{\tilde{\Omega}_{R(h),k}}(\omega)\left|X_k\right|^{2p+l(\varkappa-1)}h^{l/2-1} \leq \chi_{\tilde{\Omega}_{R(h),k}}(\omega)\left|X_k\right|^{2p}$, $l=3,\ldots,2p$, and hence we re-write (3.15) as

$$\begin{aligned}
& \mathbb{E}\chi_{\tilde{\Omega}_{R(h),k+1}}(\omega)|X_{k+1}|^{2p} && (3.17)\\
\leq\; & \mathbb{E}\chi_{\tilde{\Omega}_{R(h),k}}(\omega)|X_k|^{2p}+Kh\mathbb{E}\chi_{\tilde{\Omega}_{R(h),k}}(\omega)\left|X_k\right|^{2p}+K\sum_{l=1}^{p}\mathbb{E}\chi_{\tilde{\Omega}_{R(h),k}}(\omega)\left|X_k\right|^{2(p-l)}h^l \\
\leq\; & \mathbb{E}\chi_{\tilde{\Omega}_{R(h),k}}(\omega)|X_k|^{2p}+Kh\mathbb{E}\chi_{\tilde{\Omega}_{R(h),k}}(\omega)\left|X_k\right|^{2p}+Kh,
\end{aligned}$$

where in the last line we have used Young's inequality. From here, we get by Gronwall's inequality that

$$\mathbb{E}\chi_{\tilde{\Omega}_{R(h),k}}(\omega)|X_k|^{2p} \leq K(1+\mathbb{E}|X_0|^{2p}), \tag{3.18}$$

where $R(h)$ is from (3.16) and K does not depend on k and h but it depends on p.

It remains to estimate $\mathbb{E}\chi_{\tilde{\Lambda}_{R(h),k}}(\omega)|X_k|^{2p}$. We have

$$\begin{aligned}
\chi_{\tilde{\Lambda}_{R,k}} &= 1-\chi_{\tilde{\Omega}_{R,k}} = 1-\chi_{\tilde{\Omega}_{R,k-1}}\chi_{|X_k|\leq R} = \chi_{\tilde{\Lambda}_{R,k-1}}+\chi_{\tilde{\Omega}_{R,k-1}}\chi_{|X_k|>R} \\
&= \cdots = \sum_{l=0}^{k}\chi_{\tilde{\Omega}_{R,l-1}}\chi_{|X_l|>R},
\end{aligned}$$

where we put $\chi_{\tilde{\Omega}_{R,-1}} = 1$. Then, using (3.4), (3.18), (2.1), and Cauchy-Bunyakovsky's and Markov's inequalities, we obtain

$$\begin{aligned}
& \mathbb{E}\chi_{\tilde{\Lambda}_{R(h),k}}(\omega)|X_k|^{2p} = \mathbb{E}\sum_{l=0}^{k}|X_k|^{2p}\chi_{\tilde{\Omega}_{R(h),l-1}}\chi_{|X_l|>R(h)} && (3.19)\\
& \leq \left(\mathbb{E}|X_0+k|^{4p}\right)^{1/2}\sum_{l=0}^{k}\left(\mathbb{E}\left[\chi_{\tilde{\Omega}_{R(h),l-1}|X_l|>R(h)}\right]\right)^{1/2} \\
& = \left(\mathbb{E}|X_0+k|^{4p}\right)^{1/2}\sum_{l=0}^{k}\left(P(\chi_{\tilde{\Omega}_{R(h),l-1}}|X_l|>R)\right)^{1/2} \\
& \leq \left(\mathbb{E}|X_0+k|^{4p}\right)^{1/2}\sum_{l=0}^{k}\frac{\left(\mathbb{E}(\chi_{\tilde{\Omega}_{R(h),l-1}}|X_l|^{2(2p+1)(6\varkappa-4)})\right)^{1/2}}{R(h)^{(2p+1)(6\varkappa-4)}} \\
& \leq K\left(\mathbb{E}|X_0+k|^{4p}\right)^{1/2}\left(\mathbb{E}(1+|X_0|^{2(2p+1)(6\varkappa-4)})\right)^{1/2}kh^{2p+1} \leq K(1+\mathbb{E}|X_0|^{4p+2(2p+1)(6\varkappa-4)})^{1/2},
\end{aligned}$$

which together with (3.18) implies (3.2) for integer $p \geq 1$. Then, by Jensen's inequality, (3.2) holds for non-integer p as well. □

The next lemma gives estimates for the one-step error of the balanced scheme (3.1).

Lemma 3.2. *Assume that* (2.5) *holds. Assume that the coefficients* $a(t,x)$ *and* $\sigma_r(t,x)$ *have continuous first-order partial derivatives in* t *and that these derivatives and the coefficients satisfy inequalities of the form* (2.3). *Then the scheme* (3.1) *satisfies the inequalities* (2.9) *and* (2.10) *with* $q_1 = 3/2$ *and* $q_2 = 1$, *respectively.*

The proof of this lemma is a routine analysis of the one-step approximation corresponding to (3.1) using the equalities (3.8)-(3.9). Since such analysis is similar to those done in the global Lipschitz case [13, 15], we omit these routine calculations here. Lemmas 3.1 and 3.2 and Theorem 2.1 imply the following result.

Proposition 3.1. *Under the assumptions of Lemmas 3.1 and 3.2 the balanced scheme* (3.1) *has mean-square order* 1/2, *i.e., for it the inequality* (2.13) *holds with* $q = q_2 - 1/2 = 1/2$.

Remark 3.1. *In the additive noise case the mean-square order of the balanced scheme* (3.1) *does not improve (*q_1 *and* q_2 *remain* 3/2 *and* 1, *respectively).*

4 Fully implicit schemes

Fully implicit (i.e., implicit both in drift and diffusion coefficients) mean-square schemes were proposed in [14] (see also [15, Chapter 1]), where their convergence was proved under global Lipschitz conditions. Here we analyze these schemes under the following assumptions, which are stronger with respect to the diffusion coefficient than Assumption 2.1 used in the previous Sections 2 and 3.

Assumption 4.1. *(i) The initial condition is such that*

$$\mathbb{E}|X_0|^{2p} \leq K < \infty, \quad \textit{for all} \;\; p \geq 1. \tag{4.1}$$

(ii) There exists a constant $c_1 \geq 0$ *such that*

$$(x - y, a(t,x) - a(t,y)) \leq c_1 |x-y|^2, \quad t \in [t_0, T], \;\; x, y \in \mathbb{R}^d. \tag{4.2}$$

(iii) There exist $c_2 \geq 0$ *and* $\varkappa \geq 1$ *such that*

$$|a(t,x) - a(t,y)|^2 \leq c_2(1 + |x|^{2\varkappa-2} + |y|^{2\varkappa-2})|x-y|^2, \quad t \in [t_0, T], \;\; x, y \in \mathbb{R}^d. \tag{4.3}$$

(iv) The coefficients $\sigma_r(t,x)$ *have continuous bounded first-order spatial derivatives so that there are constants* $L_1 \geq 0$ *and* $L_2 \geq 0$ *:*

$$|\nabla \sigma_r(t,x)| \leq L_1, \;\; r = 1, \ldots, m, \quad t \in [t_0, T], \;\; x \in \mathbb{R}^d, \tag{4.4}$$

and

$$|\nabla \sigma_r(t,x)\sigma_r(t,x) - \nabla \sigma_r(t,y)\sigma_r(t,y)| \leq L_2 |x-y|, \;, \; r = 1, \ldots, m, \quad t \in [t_0, T], \;\; x, y \in \mathbb{R}^d. \tag{4.5}$$

In proofs which follow we will need some implications of Assumption 4.1. The condition (4.2) implies that there is $c \geq 0$

$$(x, a(t,x)) \leq c(1 + |x|^2), \quad t \in [t_0, T], \;\; x \in \mathbb{R}^d. \tag{4.6}$$

It follows from (4.4) that

$$|\sigma_r(t,x) - \sigma_r(t,y)| \leq L_1|x-y|, \quad t \in [t_0,T], \ x,y \in \mathbb{R}^d, \tag{4.7}$$

and hence

$$|\sigma_r(t,x)| \leq L_1|x| + L_0, \tag{4.8}$$

where $L_0 = \max_{t\in[t_0,T]} |\sigma_r(t,0)|$. Further, there is $L \geq 0$:

$$|\nabla\sigma_r(t,x)\sigma_r(t,x)| \leq L(1+|x|), \quad t \in [t_0,T], \ x \in \mathbb{R}^d, \tag{4.9}$$

and

$$|\sigma_r(t,x)|^2 \leq L(1+|x|^2), \quad t \in [t_0,T], \ x \in \mathbb{R}^d. \tag{4.10}$$

For definiteness, we consider the following one-parametric family of methods for (1.1) from the broader class of fully implicit schemes of [14, 15]:

$$\begin{aligned} X_{k+1} &= X_k + a(t_{k+\lambda},(1-\lambda)X_k + \lambda X_{k+1})h \\ &\quad -\lambda\sum_{r=1}^{m}\sum_{j=1}^{d}\frac{\partial\sigma_r}{\partial x^j}(t_{k+\lambda},(1-\lambda)X_k + \lambda X_{k+1})\sigma_r^j(t_{k+\lambda},(1-\lambda)X_k + \lambda X_{k+1})h \\ &\quad +\sum_{r=1}^{m}\sigma_r(t_{k+\lambda},(1-\lambda)X_k + \lambda X_{k+1})\left(\zeta_{rh}\right)_k\sqrt{h}, \end{aligned} \tag{4.11}$$

where $0 \leq \lambda \leq 1$, $t_{k+\lambda} = t_k + \lambda h$ and $\left(\zeta_{rh}\right)_k$ are i.i.d. random variables so that

$$\zeta_h = \begin{cases} \xi, \ |\xi| \leq A_h, \\ A_h, \ \xi > A_h, \\ -A_h, \ \xi < -A_h, \end{cases} \tag{4.12}$$

with $\xi \sim \mathcal{N}(0,1)$ and $A_h = \sqrt{2l|\ln h|}$ with $l \geq 1$. We recall [15, Lemma 1.3.4] that

$$E(\xi^2 - \zeta_h^2) = (1 + 2\sqrt{2l|\ln h|})h^l. \tag{4.13}$$

Remark 4.1. *Three choices of λ are most notable: $\lambda = 0$ gives the explicit Euler scheme which is divergent [5, 7] in the considered setting; $\lambda = 1$ gives the fully implicit Euler scheme; and $\lambda = 1/2$ corresponds to the mid-point rule, which in application to a system of Stratonovich SDE is derivative free [15, p. 45].*

Now we will study properties of the method (4.11).

Consider the one-step approximations corresponding to (4.11)

$$\begin{aligned} \bar{X} &= \bar{X}^{\lambda} = x + a(t+\lambda h, U^{\lambda})h - \lambda\sum_{r=1}^{m}\sum_{j=1}^{d}\frac{\partial\sigma_r}{\partial x^j}(t+\lambda h, U^{\lambda})\sigma_r^j(t+\lambda h, U^{\lambda})h \\ &\quad +\sum_{r=1}^{m}\sigma_r(t+\lambda h, U^{\lambda})\zeta_{rh}\sqrt{h}, \end{aligned} \tag{4.14}$$

where

$$U=U^{\lambda}:=(1-\lambda)x+\lambda\bar{X}^{\lambda}. \tag{4.15}$$

Note that

$$\begin{aligned} U^{\lambda} &= x+\lambda a(t+\lambda h,U^{\lambda})-\lambda^2\sum_{r=1}^{m}\sum_{j=1}^{d}\frac{\partial\sigma_r}{\partial x^j}(t+\lambda h,U^{\lambda})\sigma_r^j(t+\lambda h,U^{\lambda})h \\ &\quad +\lambda\sum_{r=1}^{m}\sigma_r(t+\lambda h,U^{\lambda})\zeta_{rh}\sqrt{h}. \end{aligned} \tag{4.16}$$

Lemma 4.1. *Let $0<\lambda\leq 1$. Assume that Assumption 4.1 holds. For an arbitrary $0<\varepsilon<1$, find $h_0>0$ such that*

$$\lambda\left[h_0c_1+m\lambda L_2h_0+mL_1\sqrt{2lh_0|\ln h_0|}\right]=1-\varepsilon. \tag{4.17}$$

Then the equation (4.14) for any $0<h\leq h_0$ has the unique solution $\bar{X}$ which satisfies the inequalities for some $K>0$:

$$|\bar{X}-x|\leq K(1+|x|^{\varkappa})h+K(1+|x|)\sqrt{h|\ln h|} \tag{4.18}$$

and

$$|\bar{X}|^2\leq\frac{16}{3\varepsilon^2\lambda}(L_0+1)\sqrt{2lh|\ln h|}+\frac{4}{\lambda^2}[(1-\lambda)^2+\frac{4}{3\varepsilon^2}]|x|^2,\ t\in[t_0,T],\ x\in\mathbb{R}^d. \tag{4.19}$$

Proof. Let

$$\tilde{a}(t,x)=a(t,x)-\lambda\sum_{r=1}^{m}\sum_{j=1}^{d}\frac{\partial\sigma_r}{\partial x^j}(t,x)\sigma_r^j(t,x). \tag{4.20}$$

For any fixed λ, t, ζ_{rh}, and h, we introduce the function

$$\psi(z)=z-\lambda\tilde{a}(t+\lambda h,z)h-\lambda\sum_{r=1}^{m}\sigma_r(t+\lambda h,z)\zeta_{rh}\sqrt{h}$$

which is continuous in z due to our assumptions. The equation (4.14) can be written as

$$\psi(U^{\lambda})=x. \tag{4.21}$$

Using (4.2), (4.5) and (4.7), we obtain

$$\begin{aligned} (z-y,\psi(z)-\psi(y)) &\geq |z-y|^2-h\lambda c_1|z-y|^2-hm\lambda^2L_2|z-y|^2 \\ &\quad -m\lambda L_1\sqrt{2lh|\ln h|}|z-y|^2 \\ &= (1-\lambda\left[hc_1+m\lambda L_2h+mL_1\sqrt{2lh|\ln h|}\right])|z-y|^2\geq\varepsilon|z-y|^2>0, \end{aligned} \tag{4.22}$$

i.e., $\psi(z)$ is uniformly monotone function for $h\leq h_0$. This implies (see, e.g. [20, Theorem 6.4.4, p. 167]) that (4.14) has a unique solution.

We obtain from (4.21) and (4.22):

$$\begin{aligned}\varepsilon|U|^2 &\leq (U, \psi(U) - \psi(0)) = (U, x - \psi(0)) \\ &\leq \frac{\varepsilon}{4}|U|^2 + \frac{2}{\varepsilon}|x|^2 + \frac{2}{\varepsilon}|\psi(0)|^2 \leq \frac{\varepsilon}{4}|U|^2 + \frac{2}{\varepsilon}|x|^2 + \frac{2\lambda(L_0+1)\sqrt{2lh|\ln h|}}{\varepsilon},\end{aligned}$$

from which (4.19) follows.

$$|U|^2 \leq \frac{8}{3\varepsilon^2}(\lambda(L_0+1)\sqrt{2lh|\ln h|} + |x|^2) \tag{4.23}$$

which implies (4.19).

Further, it follows from (4.15), (4.21) and (4.22) that

$$\begin{aligned}\lambda\varepsilon|\bar{X} - x|^2 &= \varepsilon|U - x|^2 \leq (U - x, -\lambda\tilde{a}(t+\lambda h, x)h - \lambda\sum_{r=1}^{m}\sigma_r(t+\lambda h, x)\zeta_{rh}\sqrt{h}) \\ &\leq \lambda^2|\bar{X} - x|\left(h|\tilde{a}(t+\lambda h, x)| + \sqrt{2lh|\ln h|}\sum_{r=1}^{m}|\sigma_r(t+\lambda h, x)|\right).\end{aligned}$$

Then, using (4.3) and (4.8), we obtain (4.18), which completes the proof of Lemma 4.1 for the implicit method (4.11). □

Now we consider boundedness of moments of (4.11).

Lemma 4.2. *Let $1/2 < \lambda \leq 1$. Assume that Assumption 4.1 holds. Then for all $0 < h \leq h_0$ with h_0 from (4.17) and for all $k = 0, \ldots, N$ the following inequality holds for the fully implicit scheme (4.11) for $p \geq 1$:*

$$\mathbb{E}|X_k|^{2p} \leq K(1 + \mathbb{E}|X_0|^{2p}), \tag{4.24}$$

where $K > 0$ is a constant.

Proof. We note that (4.6) and (4.9) imply

$$(x, \tilde{a}(t, x)) \leq (c + 3m\lambda L/2)(1 + |x|^2), \quad t \in [t_0, T],\ x \in \mathbb{R}, \tag{4.25}$$

which together with (4.4) ensures that the solution of (1.1) has all moments (2.5), $p \geq 1$ [3].

Let $U_{k+1} = (1-\lambda)X_k + \lambda X_{k+1}$ (cf. (4.15)). We have

$$V_{k+1} := |X_{k+1}|^2 - |X_k|^2 = 2(U_{k+1}, X_{k+1} - X_k) - (2\lambda - 1)|X_{k+1} - X_k|^2 \tag{4.26}$$
$$= 2\lambda h(U_{k+1}, \tilde{a}(t_{k+\lambda}, U_{k+1})) + 2\lambda\sqrt{h}(U_{k+1}, \sum_{r=1}^{m} \sigma_r(t_{k+\lambda}, U_{k+1})\,(\zeta_{rh})_k)$$
$$-(2\lambda-1)h^2|\tilde{a}(t_{k+\lambda}, U_{k+1})|^2 - (2\lambda - 1)h|\sum_{r=1}^{m} \sigma_r(t_{k+\lambda}, U_{k+1})\,(\zeta_{rh})_k\,|^2$$
$$-2(2\lambda-1)h^{3/2}(\tilde{a}(t_{k+\lambda}, U_{k+1}), \sum_{r=1}^{m} \sigma_r(t_{k+\lambda}, U_{k+1})\,(\zeta_{rh})_k)$$
$$= 2\lambda(U_{k+1}, h\tilde{a}(t_{k+\lambda}, U_{k+1})) + 2\lambda\sqrt{h}(X_k, \sum_{r=1}^{m} \sigma_r(t_{k+\lambda}, U_{k+1})\,(\zeta_{rh})_k)$$
$$+(2\lambda^2 - 2\lambda + 1)h|\sum_{r=1}^{m} \sigma_r(t_{k+\lambda}, U_{k+1})\,(\zeta_{rh})_k\,|^2 - (2\lambda - 1)h^2|\tilde{a}(t_{k+\lambda}, U_{k+1})|^2$$
$$+2(1-\lambda)^2 h^{3/2}(\tilde{a}(t_{k+\lambda}, U_{k+1}), \sum_{r=1}^{m} \sigma_r(t_{k+\lambda}, U_{k+1})\,(\zeta_{rh})_k).$$

Expanding $\sigma_r(t_{k+\lambda}, U_{k+1})$ at $(t_{k+\lambda}, X_k)$, we obtain

$$V_{k+1} = 2\lambda(U_{k+1}, h\tilde{a}(t_{k+\lambda}, U_{k+1})) + 2\lambda\sqrt{h}(X_k, \sum_{r=1}^{m} \sigma_r(t_{k+\lambda}, X_k)\,(\zeta_{rh})_k) \tag{4.27}$$
$$+2\lambda\sqrt{h}(X_k, \sum_{r=1}^{m} \nabla\sigma_r(t_{k+\lambda}, \theta)(U_{k+1} - X_k)\,(\zeta_{rh})_k)$$
$$+(2\lambda^2 - 2\lambda + 1)h|\sum_{r=1}^{m} \sigma_r(t_{k+\lambda}, U_{k+1})\,(\zeta_{rh})_k\,|^2 - (2\lambda - 1)h^2|\tilde{a}(t_{k+\lambda}, U_{k+1})|^2$$
$$+2(\lambda-1)^2 h^{3/2}(\tilde{a}(t_{k+\lambda}, U_{k+1}), \sum_{r=1}^{m} \sigma_r(t_{k+\lambda}, U_{k+1})\,(\zeta_{rh})_k)$$

$$= 2\lambda(U_{k+1}, h\tilde{a}(t_{k+\lambda}, U_{k+1})) + 2\lambda\sqrt{h}(X_k, \sum_{r=1}^{m} \sigma_r(t_{k+\lambda}, X_k)\,(\zeta_{rh})_k)$$
$$+2\lambda^2 h^{3/2}(X_k, \sum_{r=1}^{m} \nabla\sigma_r(t_{k+\lambda}, \theta)\tilde{a}(t_{k+\lambda}, U_{k+1})\,(\zeta_{rh})_k)$$
$$+2\lambda^2 h(X_k, \sum_{r=1}^{m} \nabla\sigma_r(t_{k+\lambda}, \theta) \sum_{l=1}^{m} \sigma_l(t_{k+\lambda}, X_k)\,(\zeta_{lh})_k\,(\zeta_{rh})_k)$$
$$+(2\lambda^2 - 2\lambda + 1)h|\sum_{r=1}^{m} \sigma_r(t_{k+\lambda}, U_{k+1})\,(\zeta_{rh})_k\,|^2 - (2\lambda - 1)h^2|\tilde{a}(t_{k+\lambda}, U_{k+1})|^2$$
$$+2(1-\lambda)^2 h^{3/2}(\tilde{a}(t_{k+\lambda}, U_{k+1}), \sum_{r=1}^{m} \sigma_r(t_{k+\lambda}, U_{k+1})\,(\zeta_{rh})_k),$$

where $\theta = \nu U_{k+1} - (1-\nu)X_k$, $\nu \in [0,1]$, is an intermediate point. Using (4.25), (4.10), Young's inequality, (4.4) and (4.19), we obtain

$$
\begin{aligned}
V_{k+1} &\leq \lambda h(2c+3\lambda mL)(1+|U_{k+1}|^2)+2\lambda\sqrt{h}(X_k,\sum_{r=1}^{m}\sigma_r(t_{k+\lambda},X_k)\left(\zeta_{rh}\right)_k) \qquad (4.28)\\
&+\frac{2\lambda-1}{2}h^2|\tilde{a}(t_{k+\lambda},U_{k+1})|^2+\frac{2\lambda^4}{2\lambda-1}h|X_k|^2m\sum_{r=1}^{m}|\nabla\sigma_r(t_{k+\lambda},\theta)|^2|\left(\zeta_{rh}\right)_k|^2\\
&+\lambda^2hm|X_k|^2\sum_{r=1}^{m}|\nabla\sigma_r(t_{k+\lambda},\theta)|^2|\left(\zeta_{rh}\right)_k|^2+\lambda^2hm\sum_{l=1}^{m}|\sigma_l(t_{k+\lambda},X_k)|^2|\left(\zeta_{lh}\right)_k|^2\\
&+(2\lambda^2-2\lambda+1)hm\sum_{r=1}^{m}|\sigma_r(t_{k+\lambda},U_{k+1})|^2|\left(\zeta_{rh}\right)_k|^2-(2\lambda-1)h^2|\tilde{a}(t_{k+\lambda},U_{k+1})|^2\\
&+\frac{2\lambda-1}{2}h^2|\tilde{a}(t_{k+\lambda},U_{k+1})|^2+\frac{2(1-\lambda)^4}{2\lambda-1}hm\sum_{r=1}^{m}|\sigma_r(t_{k+\lambda},U_{k+1})|^2|\left(\zeta_{rh}\right)_k|^2\\
&\leq \lambda h(2c+3\lambda mL)(1+|U_{k+1}|^2)+2\lambda\sqrt{h}(X_k,\sum_{r=1}^{m}\sigma_r(t_{k+\lambda},X_k)\left(\zeta_{rh}\right)_k)\\
&+\lambda^2[\frac{2\lambda^2}{2\lambda-1}+1]hL_1^2|X_k|^2m\sum_{r=1}^{m}|\left(\zeta_{rh}\right)_k|^2+\lambda^2hmL(1+|X_k|^2)\sum_{l=1}^{m}|\left(\zeta_{lh}\right)_k|^2\\
&+[2\lambda^2-2\lambda+1+\frac{2(1-\lambda)^4}{2\lambda-1}]hmL(1+|U_{k+1}|^2)\sum_{l=1}^{m}|\left(\zeta_{lh}\right)_k|^2.
\end{aligned}
$$

Then using (4.23), we arrive at

$$
V_{k+1}\leq Kh(1+|X_k|^2)\left(1+\sum_{r=1}^{m}|\left(\zeta_{rh}\right)_k|^2\right)+2\lambda\sqrt{h}(X_k,\sum_{r=1}^{m}\sigma_r(t_{k+\lambda},X_k)\left(\zeta_{rh}\right)_k),
$$

where $K>0$ is independent of h and k while it depends on λ and on constants appearing in (4.2)-(4.10).

Thus

$$
1+|X_{k+1}|^2\leq 1+|X_k|^2+Kh(1+|X_k|^2)\left(1+\sum_{r=1}^{m}|\left(\zeta_{rh}\right)_k|^2\right)+2\lambda\sqrt{h}(X_k,\sum_{r=1}^{m}\sigma_r(t_{k+\lambda},X_k)\left(\zeta_{rh}\right)_k).
$$

Then for integer $p\geq 1$ we get

$$
\begin{aligned}
\left(1+|X_{k+1}|^2\right)^p &\leq \left(1+|X_k|^2\right)^p+K\left(1+|X_k|^2\right)^p\sum_{l=1}^{p}h^l\left[1+\sum_{r=1}^{m}|\left(\zeta_{rh}\right)_k|^2\right]^l\\
&+K\sum_{l=1}^{p}\left(1+|X_k|^2\right)^{p-l}h^{l/2}\left[(X_k,\sum_{r=1}^{m}\sigma_r(t_{k+\lambda},X_k)\left(\zeta_{rh}\right)_k)\right]^l,
\end{aligned}
$$

whence, observing that X_k are $\mathcal{F}_{t_k}$-measurable while ξ_{rk} are independent of $\mathcal{F}_{t_k}$, it is not difficult to obtain

$$
\begin{aligned}
\mathbb{E}\left(1+|X_{k+1}|^2\right)^p &\leq \mathbb{E}\left(1+|X_k|^2\right)^p+Kh\mathbb{E}\left(1+|X_k|^2\right)^p\\
&+K\sum_{l=2}^{p}\mathbb{E}\left(1+|X_k|^2\right)^{p-l}h^{l/2}\left[(X_k,\sum_{r=1}^{m}\sigma_r(t_{k+\lambda},X_k)\left(\zeta_{rh}\right)_k)\right]^l\\
&\leq \mathbb{E}\left(1+|X_k|^2\right)^p+Kh\mathbb{E}\left(1+|X_k|^2\right)^p,
\end{aligned}
$$

which together with Gronwall's inequality completes the proof of the lemma for integer $p \geq 1$. Then by Jensen's inequality for non-integer $p > 1$ as well. $\square$

We have not succeeded in proving boundedness of moments for the mid-point scheme, i.e., (4.11) with $\lambda = 1/2$ under Assumption 4.1. One can observe that the proof of Lemma 4.2 is not applicable to this choice of λ as the estimate in (4.28) blows up when $\lambda \to 1/2$ and it is clear that the mid-point scheme is the boundary case. We also know [4] that for $\sigma_r = 0$ (4.11) is B-stable for $\lambda \geq 1/2$ and not B-stable (in fact, not A-stable) for $\lambda < 1/2$. It is natural to expect that for $\lambda < 1/2$ the moments of (4.11) are not bounded and hence the method with $\lambda < 1/2$ is divergent under Assumption 4.1 (see also such a conclusion for the drift-implicit θ-method in [21]). In our experiments (Section 5) the mid-point method produced accurate results.

At the same time, we proved boundedness of moments for the mid-point scheme if in addition to Assumption 4.1 we require that the diffusion coefficients $\sigma_r(t,x)$ are bounded. The proof is similar to the proof of Lemma 3.1.

Lemma 4.3. *Let the assumptions of Lemma 4.2 hold and in addition assume that the diffusion coefficients $\sigma_r(t,x)$ are uniformly bounded. Then the moments of the mid-point method* (4.11) *with $\lambda = 1/2$ has bounded moments: for $p \geq 1$:*

$$\mathbb{E}|X_k|^{2p} \leq K(1 + \mathbb{E}|X_0|^{4(p+1)\varkappa - 4})^{1/2}, \tag{4.29}$$

where $K > 0$ is a constant.

Proof. For $\varkappa = 1$ (cf. (4.3)), i.e., the global Lipschitz case, boundedness of moments of X_k is established in [14, 15]. Let $\varkappa > 1$.

From (4.26), we have

$$|X_{k+1}|^2 - |X_k|^2 = h(U_{k+1}, \tilde{a}(t_{k+1/2}, U_{k+1})) + \sqrt{h}(U_{k+1}, \sum_{r=1}^{m} \sigma_r(t_{k+1/2}, U_{k+1})\,(\zeta_{rh})_k).$$

Then using (4.25) and Young's inequality, boundedness of σ_r and (4.23), we get

$$\begin{aligned} |X_{k+1}|^2 - |X_k|^2 &\leq h(c + 3\lambda m L/2)(1 + |U_{k+1}|^2) + h|U_{k+1}|^2 + K\sum_{r=1}^{m} |\,(\zeta_{rh})_k\,|^2 \\ &\leq Kh(1 + |X_k|^2) + K\sum_{r=1}^{m} |\,(\zeta_{rh})_k\,|^2, \end{aligned}$$

from which one can obtain for integer $p \geq 1$:

$$\mathbb{E}\left(1 + |X_k|^2\right)^p \leq Kh^{-p}\mathbb{E}\left(1 + |X_0|^2\right)^p. \tag{4.30}$$

Further, using (4.25), Young's inequality, boundedness of σ_r, (4.4), (4.3) and (4.23), we get from

(4.27):

$$|X_{k+1}|^2 - |X_k|^2 = (U_{k+1}, h\tilde{a}(t_{k+1/2}, U_{k+1})) + \sqrt{h}(X_k, \sum_{r=1}^m \sigma_r(t_{k+1/2}, X_k)\,(\zeta_{rh})_k)$$

$$+\frac{1}{2}h^{3/2}(X_k, \sum_{r=1}^m \nabla\sigma_r(t_{k+1}, \theta)\tilde{a}(t_{k+1/2}, U_{k+1})\,(\zeta_{rh})_k)$$

$$+\frac{1}{2}h(X_k, \sum_{r=1}^m \nabla\sigma_r(t_{k+1/2}, \theta)\sum_{l=1}^m \sigma_l(t_{k+1/2}, X_k)\,(\zeta_{lh})_k\,(\zeta_{rh})_k) + \frac{1}{2}h|\sum_{r=1}^m \sigma_r(t_{k+1/2}, U_{k+1})\,(\zeta_{rh})_k|^2$$

$$+\frac{1}{2}h^{3/2}(\tilde{a}(t_{k+\lambda}, U_{k+1}), \sum_{r=1}^m \sigma_r(t_{k+\lambda}, U_{k+1})\,(\zeta_{rh})_k)$$

$$\leq \quad Kh(1+|X_k|^2)(1+\sum_{r=1}^m |\,(\zeta_{rh})_k\,|^2) + \sqrt{h}(X_k, \sum_{r=1}^m \sigma_r(t_{k+1/2}, X_k)\,(\zeta_{rh})_k)$$

$$+Kh^{3/2}(1+|X_k|^{\varkappa+1})\sum_{r=1}^m |\,(\zeta_{rh})_k\,|.$$

Choosing $R(h) = h^{-1/2(\varkappa-1)}$, we get after some additional calculation (cf. (3.18)):

$$\mathbb{E}\chi_{\tilde{\Omega}_{R(h),k}}(\omega)|X_k|^{2p} \leq K(1+\mathbb{E}|X_0|^{2p}), \tag{4.31}$$

where $\tilde{\Omega}_{R(h),k}$ is the event as in (3.5).

Now using (4.30), (4.31), and Cauchy-Bunyakovsky's and Markov's inequalities, we arrive at (cf. (3.19)):

$$\mathbb{E}\chi_{\tilde{\Lambda}_{R(h),k}}(\omega)|X_k|^{2p} \leq K(1+\mathbb{E}|X_0|^{4(p+1)\varkappa-4})^{1/2},$$

from which together with (4.31) the inequality (4.29) follows. □

The next lemma gives estimates for the one-step error of the method (4.11).

Lemma 4.4. *Let* $0 \leq \lambda \leq 1$. *Assume that* (2.5) *holds. Assume that the coefficient* $a(t,x)$ *has continuous first order partial derivative in* t *and in* x^i *and that the derivatives and the coefficient satisfy inequalities of the form* (4.3)*; the functions* $\sigma_r(t,x)$ *have continuous first-order partial derivatives in* t *and that the derivatives and the coefficients satisfy inequalities of the form* (4.4)−(4.5)*; and the functions* $\nabla\sigma_r(t,x)\sigma_r(t,x)$ *have continuous first partial derivatives in* t *and in* x^i *which satisfy inequalities of the form* (4.5). *Then the method* (4.11) *satisfies the inequalities* (2.9) *and* (2.10) *with* $q_1 = 2$ *and* $q_2 = 1$, *respectively.*

Proofs of this lemma is rather routine and similar to the global Lipschitz case [14, 15] and it is omitted here. Using Lemmas 4.1-4.4, the next proposition follows from Theorem 2.1.

Proposition 4.1. *Let for* $1/2 < \lambda \leq 1$ *the assumptions of Lemmas 4.2 and 4.4 hold and for* $\lambda = 1/2$ *in addition assume that the diffusion coefficients* $\sigma_r(t,x)$ *are uniformly bounded. Then the fully implicit method* (4.11) *has mean-square order* $1/2$, *i.e., for it the inequality* (2.13) *holds with* $q = 1/2$.

Remark 4.2. *Consider the commutative case, i.e., when* $\Lambda_i\sigma_r = \Lambda_r\sigma_i$ *(here the operator* $\Lambda_r := (\sigma_r, \partial/\partial x)$*) or in the case of a system with one noise (i.e.,* $m = 1$*). Then in the setting of Lemma 4.4, the mid-point method, i.e.,* (4.11) *with* $\lambda = 1/2$*, satisfies the inequalities* (2.9) *and* (2.10) *with* $q_1 = 2$ *and* $q_2 = 3/2$*, respectively (see such a result in the global Lipschitz case in [14, 15]). Therefore, it converges in this case with mean-square order* 1 *when its moments are bounded.*

5 Numerical examples

In this section we will test the following schemes: the balanced method (3.1) from Section 3; the drift-implicit scheme (2.17); the fully implicit Euler scheme (4.11) with $\lambda = 1$; the mid-point method (4.11) with $\lambda = 1/2$; the drift-tamed Euler scheme (a modified balanced method) [8]:

$$X_{k+1} = X_k + h\frac{a(X_k)}{1 + h\,|a(X_k)|} + \sum_{r=1}^{m} \sigma_r(t_k, X_k)\xi_{rk}\sqrt{h}; \tag{5.1}$$

the fully-tamed scheme [7]:

$$X_{k+1} = X_k + \frac{a(X_k)h + \sum_{r=1}^{m} \sigma_r(t_k, X_k)\xi_{rk}\sqrt{h}}{\max\left(1, h\left|ha(X_k) + \sum_{r=1}^{m} \sigma_r(t_k, X_k)\xi_{rk}\sqrt{h}\right|\right)}; \tag{5.2}$$

and the trapezoidal scheme:

$$X_{k+1} = X_k + \frac{h}{2}\left[a(X_{k+1}) + a(X_k)\right] + \sum_{r=1}^{m} \sigma_r(t_k, X_k)\xi_{rk}\sqrt{h}. \tag{5.3}$$

As before, $\xi_{rk} = (w_r(t_{k+1}) - w_r(t_k))/\sqrt{h}$ are Gaussian $\mathcal{N}(0,1)$ i.i.d. random variables. We note that under Assumption 2.1 boundedness of second moments and strong convergence (without giving order) of θ-schemes, and in particular of (5.3), can be found in [21]. Strong convergence with order 1/2 of (5.1) under Assumption 4.1 is proved in [8]. Strong convergence of (5.2) without order under Assumption 2.1 is proved in [7].

In all the experiments with fully implicit schemes, where the truncated random variables ζ are used, we took $l = 2$ (see (4.13)). The experiments were performed using Matlab R2012a on a Macintosh desktop computer with Intel Xeon CPU E5462 (quad-core, 2.80 GHz). In simulations we used the Mersenne twister random generator with seed 100. Newton's method was used to solve the nonlinear algebraic equations at each step of the implicit schemes.

We test the methods on the two model problems. The first one satisfies Assumption 4.1 (nonglobal Lipschitz drift, global Lipschitz diffusion) and has two non-commutative noises. The second example satisfies Assumption 2.1 (nonglobal Lipschitz both drift and diffusion). The aim of the tests is to compare performance of the methods: their accuracy (i.e., roughly speaking, size of prefactors at a power of h) and computational costs. We note that experiments cannot prove or disprove boundedness of moments of the schemes since experiments rely on a finite sample of trajectories run over a finite time interval while blow-up of moments in divergent methods (e.g., explicit Euler scheme) is, in general, a result of large deviations [10, 16].

Example 5.1. *Our first test model is the Stratonovich SDE of the form:*

$$dX = (1 - X^5)\, dt + X \circ dw_1 + dw_2, \quad X(0) = 0. \tag{5.4}$$

In Ito's sense, the drift of the equation becomes $a(t,x) = 1 - x^5 + x/2$. Here we tested the balanced method (3.1); the drift-tamed scheme (5.1); the fully implicit Euler scheme (4.11) with $\lambda = 1$; the mid-point method (4.11) with $\lambda = 1/2$. We note that for all the methods tested on this example except the mid-point rule mean-square convergence with order 1/2 is proved either in earlier papers or here as it was described before.

To compute the mean-square error, we run M independent trajectories $X^{(i)}(t)$, $X_k^{(i)}$:

$$\left(E\left[X(T) - X_N\right]^2\right)^{1/2} \doteq \left(\frac{1}{M}\sum_{i=1}^{M}[X^{(i)}(T) - X_N^{(i)}]^2\right)^{1/2}. \tag{5.5}$$

We took time $T = 50$ and $M = 10^4$. The reference solution was computed by the mid-point method with small time step $h = 10^{-4}$. It was verified that using a different implicit scheme for simulating a reference solution does not affect the outcome of the tests. We chose the mid-point scheme as a reference since in all the experiments it produced the most accurate results.

Table 5.1 gives the mean-square errors and experimentally observed convergence rates for the corresponding methods. We checked that the number of trajectories $M = 10^4$ was sufficiently large for the statistical errors not to significantly hinder the mean-square errors (the Monte Carlo error computed with 95% confidence was at least 10 time smaller than the reported mean-square errors except values for (5.1) at $h = 0.1$ and 0.05 where it was at least 5 time smaller than the mean-square errors). In addition to the data in the table, we evaluated errors for (3.1) for smaller time steps: $h = 0.002$ – the error is 9.27e-02 (rate 0.41), 0.001 – 6.86e-02 (0.44). The observed rates of convergence of all the tested methods are close to the predicted 1/2. For a fixed time step h, the most accurate scheme is the mid-point one, the less accurate scheme is the new balanced method (3.1). To produce the result with accuracy $\sim 0.06 - 0.07$, in our experiment of running $M = 10^4$ trajectories the scheme (5.1) required 170 sec., the mid-point (4.11) with $\lambda = 1/2$ – 329 sec., (4.11) with $\lambda = 1$ – 723 sec., and (3.1) – 1870 sec. That is, our experiments confirmed the conclusion of [8] that the drift-tamed (modified balance method) (5.1) from [8] is highly competitive. We note that (5.1) is not applicable when diffusion grows faster than a linear function and that in this case the balanced method (3.1) can outcompete implicit schemes as it is shown in the next example.

Example 5.2. *Consider the SDE in the Stratonovich sense:*

$$dX = (1 - X^5)\, dt + X^2 \circ dw, \quad X(0) = 0. \tag{5.6}$$

In Ito's sense, the drift of the equation becomes $a(t,x) = 1 - x^5 + x^3$.

Here we tested the balanced method (3.1); the fully-tamed Euler scheme (5.2); the drift-implicit scheme (2.17); the fully implicit Euler scheme (4.11) with $\lambda = 1$; the mid-point method

Table 5.1: *Example 5.1.* Mean-square errors of the selected schemes. See further details in the text.

h	(4.11), $\lambda = 1$	rate	(4.11), $\lambda = 1/2$	rate	(5.1)	rate	(3.1)	rate
0.1	1.712e-01	–	1.443e-01	–	3.748e-01	–	3.594e-01	–
0.05	1.234e-01	0.47	9.224e-02	0.65	2.103e-01	0.83	3.017e-01	0.25
0.02	7.692e-02	0.52	5.261e-02	0.61	9.472e-02	0.87	2.297e-01	0.30
0.01	5.478e-02	0.49	3.549e-02	0.57	6.104e-02	0.63	1.778e-01	0.37
0.005	3.935e-02	0.48	2.487e-02	0.51	3.959e-02	0.62	1.354e-01	0.39

(4.11) with $\lambda = 1/2$; and the trapezoidal scheme (5.3). We recall that in the case of nonglobal Lipschitz drift and diffusion, for the drift-implicit scheme (2.17) and the balanced method (3.1) mean-square convergence with order 1/2 is shown earlier in this paper; strong convergence of the trapezoidal scheme (5.3) without order is proved in [21], it is natural to expect that its mean-square order is 1/2 which is indeed supported by the experiments. Strong convergence of (5.2) without order is proved in [7]. We note that it can be proved directly that implicit algebraic equations arising from application of the mid-point and fully implicit Euler schemes to (5.6) have unique solutions under a sufficiently small time step.

The reference solution was computed by the mid-point method with small time step $h = 10^{-4}$. The time $T = 50$ and $M = 10^4$ in (5.5).

The fully-tamed scheme (5.2) did not produce accurate results until the time step size is at least $h = 0.005$ and we do not then report its errors here but see the remark below.

Remark 5.1. *The fully-tamed scheme* (5.2) *appears to be of a low practical value. If at a step* k_*, *the event* $O := \left| ha(X_k) + \sum_{r=1}^{m} \sigma_r(t_k, X_k)\xi_{rk}\sqrt{h} \right| > 1/h$ *happens, then in the case of* (5.6) *the trajectory* X_k, $k > k_*$, *oscillates approximately between* X_{k_*} *and* $X_{k_*} - signum(X_{k_*})/h$. *Since the probability of the event* O *is positive for any step size* $h > 0$ *and grows with integration time, it is unavoidable that in some scenarios (i.e., on some trajectories) such oscillatory behavior will appear. For instance, in this experiment for* $h = 0.1$ *we observed* 989 *out of* 1000 *paths for which* O *happened over the time interval* $[0, 50]$; *for* $h = 0.05$ – 866 *out of* 1000 *paths. From the practical point of view,* (5.2) *works as long as the explicit Euler scheme works (cf. [10] and also [15, p. 17]). The strong convergence (without order) of* (5.2) *[7] in comparison with the explicit Euler scheme is due to the following fact. When event* O *happens for the Euler scheme its sequence* X_k *starts oscillating with growing amplitude which leads to unboundedness of its moments and, consequently, its divergence in the mean-square sense. For* (5.2), *the oscillations are bounded by* $\sim 1/h$ *and since the probability of* O *over a finite time interval decreasing with decrease of* h, *then the moments are bounded uniformly in* h. *At the same time, the one-step approximation of* (5.2) *does not satisfy the conditions* (2.9) *and* (2.10) *of Theorem 2.1. We note that the explicit balanced-type scheme* (3.1) *does not have such drawbacks as* (5.2).

Table 5.2 gives the mean-square errors and experimentally observed convergence rates for the corresponding methods. We checked that the number of trajectories $M = 10^4$ was sufficiently large for the statistical errors not to significantly hinder the mean-square errors (the Monte Carlo error computed with 95% confidence was at least ten time smaller than the reported mean-square errors). In addition to the data in the table, we evaluated errors for (3.1) for smaller time steps: $h = 0.002$ – the error is 3.70e-02 (rate 0.41), 0.001 – 2.73e-02 (0.44), 0.0005 – 2.00e-02 (0.45), i.e., for smaller h the observed convergence rate of (3.1) becomes closer to the theoretically predicted order 1/2. Since (5.6) is with single noise, Remark 4.2 is valid here which explains why the mid-point scheme demonstrates the first order of convergence. The other implicit schemes show the order 1/2 as expected. Table 5.3 presents the time costs in seconds. Let us fix tolerance level at $0.05 - 0.06$. We highlighted in bold the corresponding values in both tables. We see that in this example the mid-point scheme is the most efficient which is due to its first order convergence in the commutative case. Among methods of order 1/2, the balanced method (3.1) is the fastest and one can expect that for multi-dimensional SDE the explicit scheme (3.1) can considerably outperform implicit methods (see a similar outcome for the drift-tamed method (5.1) supported by experiments in [8]; note that (5.1), in comparison with (3.1), is, as a rule, divergent when diffusion is growing faster than a linear function on infinity).

Table 5.2: *Example 5.2.* Mean-square errors of the selected schemes. See further details in the text.

h	(2.17)	rate	(4.11), $\lambda = 1$	rate	(4.11), $\lambda = 1/2$	rate	(5.3)	rate	(3.1)	rate
0.2	3.449e-01	–	1.816e-01	–	1.378e-01	–	4.920e-01	–	2.102e-01	–
0.1	2.441e-01	0.50	1.331e-01	0.45	8.723e-02	0.66	3.526e-01	0.48	1.637e-01	0.36
0.05	1.592e-01	0.62	9.619e-02	0.47	**5.344e-02**	0.71	2.230e-01	0.66	1.270e-01	0.37
0.02	8.360e-02	0.70	6.599e-02	0.41	2.242e-02	0.95	1.048e-01	0.82	9.170e-02	0.36
0.01	**5.460e-02**	0.61	**4.919e-02**	0.42	1.145e-02	0.97	**5.990e-02**	0.81	7.065e-02	0.38
0.005	3.682e-02	0.57	3.522e-02	0.48	5.945e-03	0.95	3.784e-02	0.66	**5.393e-02**	0.39

Table 5.3: *Example 5.2.* Comparison of computational times for the selected schemes. See further details in the text.

h	(2.17)	(4.11), $\lambda = 1$	(4.11), $\lambda = 1/2$	(5.3)	(3.1)
0.2	9.25e+00	1.10e+01	9.33e+00	1.20e+01	3.98e+00
0.1	1.77e+01	2.17e+01	1.80e+01	2.30e+01	7.49e+00
0.05	3.42e+01	4.26e+01	**3.51e+01**	4.48e+01	1.41e+01
0.02	8.33e+01	1.04e+02	8.69e+01	1.10e+02	3.37e+01
0.01	**1.64e+02**	**2.05e+02**	1.73e+02	**2.19e+02**	6.62e+01
0.005	3.25e+02	4.07e+02	3.47e+02	4.37e+02	**1.32e+02**

Acknowledgments

We are very grateful to Boris Rozovskii and George Karniadakis for fruitful discussions. MVT was partially supported by the Leverhulme Trust Fellowship SAF-2012-006 and is also grateful to ICERM (Brown University, Providence) for its hospitality. ZZ was supported by the OSD/MURI grant FA9550-09-1-0613.

References

[1] N. Bou-Rabee, E. Vanden-Eijnden. A patch that imparts unconditional stability to explicit integrators for Langevin-like equations. *J. Comp. Phys.* **231** (2012), 2565–2580.

[2] I. Gyöngy. A note on Euler's approximations. *Poten. Anal.* **8** (1998), 205–216.

[3] R.Z. Hasminskii. *Stochastic Stability of Differential Equations*. Sijthoff & Noordhoff, 1980.

[4] E. Hairer, G. Wanner. *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems.* Springer, 2004.

[5] D.J. Higham, X. Mao, A.M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM J. Num. Anal.* **40** (2003), 1041–1063.

[6] Y. Hu. Semi-implicit Euler-Maruyama scheme for stiff stochastic equations. *In: Stochastic Analysis and Related Topics V: The Silvri Workshop* (Ed.: H. Koerezlioglu), Progr. Probab. 38, Birkhauser, Boston, 1996, pp. 183–202.

[7] M. Hutzenthaler, A. Jentzen. *Numerical approximation of stochastic differential equations with non-globally Lipschitz continuous coefficients*. Preprint, 2012.

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

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

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

[11] G.N. Milstein. A theorem on the order of convergence of mean-square approximations of solutions of systems of stochastic differential equations. *Theor. Prob. Appl.* **32** (1987), 738–741.

[12] G.N. Milstein. *Numerical Integration of Stochastic Differential Equations*. Kluwer Academic Publishers, 1995.

[13] G.N. Milstein, E. Platen, H. Schurz. Balanced implicit methods for stiff stochastic systems. *SIAM J. Num. Anal.* **35** (1998), 1010–1019.

[14] G.N. Milstein, Yu.M. Repin, M.V. Tretyakov. Numerical methods for stochastic systems preserving symplectic structure. *SIAM J. Num. Anal.* **40 (**2002), 1583–1604.

[15] G.N. Milstein, M.V. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Springer, 2004.

[16] G.N. Milstein, M.V. Tretyakov. Numerical integration of stochastic differential equations with nonglobally Lipschitz coefficients. *SIAM J. Num. Anal.* **43 (**2005), 1139–1154.

[17] G.N. Milstein, M.V. Tretyakov. Computing ergodic limits for Langevin equations. *Physica D* **229** (2007), 81–95.

[18] G.N. Milstein, M.V. Tretyakov. Monte Carlo algorithms for backward equations in nonlinear filtering. *Adv. Appl. Prob.* **41** (2009), 63–100.

[19] D.S. Mitrinovic, J. Pecaric, A.M. Fink. *Inequalities Involving Functions and their Integrals and Derivatives*. Kluwer, 1994.

[20] J. M. Ortega, W.C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. SIAM, 2000.

[21] L. Szpruch, X. Mao. Strong convergence and stability of numerical methods for non-linear stochastic differential equations under monotone condition. *J. Comp. App. Math.* **238** (2013), 14–28.

[22] D. Talay. Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Proc. Relat. Fields* **8** (2002), 163–198.

A Proof of the fundamental theorem

Note that in this and the next section we shall use the letter K to denote various constants which are independent of h and k. The proof exploits the idea of the prove of this theorem in the global Lipschitz case [11].

Consider the error of the method $\bar{X}_{t_0,X_0}(t_{k+1})$ at the $(k+1)$-step:

$$\begin{aligned}\rho_{k+1} &:= X_{t_0,X_0}(t_{k+1}) - \bar{X}_{t_0,X_0}(t_{k+1}) = X_{t_k,X(t_k)}(t_{k+1}) - \bar{X}_{t_k,X_k}(t_{k+1}) \\ &= (X_{t_k,X(t_k)}(t_{k+1}) - X_{t_k,X_k}(t_{k+1})) + (X_{t_k,X_k}(t_{k+1}) - \bar{X}_{t_k,X_k}(t_{k+1}))\,.\end{aligned} \tag{A.1}$$

The first difference in the right-hand side of (A.1) is the error of the solution arising due to the error in the initial data at time t_k, accumulated at the k-th step, which we can re-write as

$$\begin{aligned} S_{t_k,X(t_k),X_k}(t_{k+1}) &= S_{k+1} := X_{t_k,X(t_k)}(t_{k+1}) - X_{t_k,X_k}(t_{k+1}) = \rho_k + Z_{t_k,X(t_k),X_k}(t_{k+1}) \\ &= \rho_k + Z_{k+1}, \end{aligned}$$

where Z is as in (2.14). The second difference in (A.1) is the one-step error at the $(k+1)$-step and we denote it as r_{k+1} :

$$r_{k+1} = X_{t_k,X_k}(t_{k+1}) - \bar{X}_{t_k,X_k}(t_{k+1}).$$

Let $p \geq 1$ be an integer. We have

$$\begin{aligned} \mathbb{E}|\rho_{k+1}|^{2p} &= \mathbb{E}\left|S_{k+1} + r_{k+1}\right|^{2p} = \mathbb{E}[(S_{k+1}, S_{k+1}) + 2(S_{k+1}, r_{k+1}) + (r_{k+1}, r_{k+1})]^p \\ &\leq \mathbb{E}\left|S_{k+1}\right|^{2p} + 2p\mathbb{E}\left|S_{k+1}\right|^{2p-2}(\rho_k + Z_{k+1}, r_{k+1}) + K\sum_{l=2}^{2p}\mathbb{E}\left|S_{k+1}\right|^{2p-l}\left|r_{k+1}\right|^l. \end{aligned} \tag{A.2}$$

Due to (2.15) of Lemma 2.1, the first term on the right-hand side of (A.2) is estimated as

$$\mathbb{E}\left|S_{k+1}\right|^{2p} \leq \mathbb{E}|\rho_k|^{2p}(1 + Kh). \tag{A.3}$$

Consider the second term on the right-hand side of (A.2):

$$\begin{aligned} \mathbb{E}\left|S_{k+1}\right|^{2p-2}(\rho_k + Z_{k+1}, r_{k+1}) = \mathbb{E}\left|\rho_k\right|^{2p-2}(\rho_k, r_{k+1}) \\ +\mathbb{E}\left(\left|S_{k+1}\right|^{2p-2} - \left|\rho_k\right|^{2p-2}\right)(\rho_k, r_{k+1}) + \mathbb{E}\left|S_{k+1}\right|^{2p-2}(Z_{k+1}, r_{k+1}). \end{aligned} \tag{A.4}$$

Due to $\mathcal{F}_{t_k}$-measurability of ρ_k and due to the conditional variant of (2.9), we get for the first term on the right-hand side of (A.4):

$$\mathbb{E}\left|\rho_k\right|^{2p-2}(\rho_k, r_{k+1}) \leq K\mathbb{E}\left|\rho_k\right|^{2p-1}(1 + |X_k|^{2\alpha})^{1/2}h^{q_1}. \tag{A.5}$$

Consider the second term on the right-hand side of (A.4) and first of all note that it is equal to zero for $p = 1$. We have for integer $p \geq 2$:

$$\mathbb{E}\left(\left|S_{k+1}\right|^{2p-2} - \left|\rho_k\right|^{2p-2}\right)(\rho_k, r_{k+1}) \leq K\mathbb{E}\left|Z_{k+1}\right|\left|\rho_k\right||r_{k+1}|\sum_{l=0}^{2p-3}\left|S_{k+1}\right|^{2p-3-l}|\rho_k|^l.$$

Further, using $\mathcal{F}_{t_k}$-measurability of ρ_k and the conditional variants of (2.10), (2.15) and (2.16) and the Cauchy-Bunyakovsky inequality (twice), we get for $p \geq 2$:

$$\begin{aligned} &\mathbb{E}\left(\left|S_{k+1}\right|^{2p-2} - \left|\rho_k\right|^{2p-2}\right)(\rho_k, r_{k+1}) \\ \leq\ &K\mathbb{E}\left|\rho_k\right|^{2p-1}(1 + |X(t_k)|^{2\varkappa-2} + |X_k|^{2\varkappa-2})^{1/4}h^{q_2+1/2}(1 + |X_k|^{2\alpha})^{1/2}. \end{aligned} \tag{A.6}$$

Due to $\mathcal{F}_{t_k}$-measurability of ρ_k, the conditional variants of (2.10) and (2.16) and the Cauchy-Bunyakovsky inequality (twice), we obtain for the third term on the right-hand side of (A.4):

$$\begin{aligned} \mathbb{E}\left|S_{k+1}\right|^{2p-2}(Z_{k+1}, r_{k+1}) \leq \mathbb{E}[\mathbb{E}\left(S_{k+1}|^{4p-4}|\mathcal{F}_{t_k}\right)^{1/2}\mathbb{E}\left(|Z_{k+1}|^4|\mathcal{F}_{t_k}\right)^{1/4}\mathbb{E}\left(|r_{k+1}|^4|\mathcal{F}_{t_k}\right)^{1/4}] \\ \leq K\mathbb{E}\left|\rho_k\right|^{2p-1}(1 + |X(t_k)|^{2\varkappa-2} + |X_k|^{2\varkappa-2})^{1/4}h^{q_2+1/2}(1 + |X_k|^{4\alpha})^{1/4}. \end{aligned} \tag{A.7}$$

Due to $\mathcal{F}_{t_k}$-measurability of ρ_k and due to the conditional variants of (2.10) and (2.15) and the Cauchy-Bunyakovsky inequality, we estimate the third term on the right-hand side of (A.2):

$$K\sum_{l=2}^{2p}\mathbb{E}\,|S_{k+1}|^{2p-l}\,|r_{k+1}|^{l} \leq K\sum_{l=2}^{2p}\mathbb{E}[\mathbb{E}(|S_{k+1}|^{4p-2l}\,|\mathcal{F}_{t_k})^{1/2}\mathbb{E}(|r_{k+1}|^{2l}|\mathcal{F}_{t_k})^{1/2}] \leq K\sum_{l=2}^{2p}\mathbb{E}[|\rho_k|^{2p-l}h^{lq_2}(1+|X_k|^{2l\alpha})^{1/2}]. \tag{A.8}$$

Substituting (A.3)-(A.8) in (A.2) and recalling that $q_1 \geq q_2 + 1/2$, we obtain

$$\begin{aligned}\mathbb{E}|\rho_{k+1}|^{2p} &\leq \mathbb{E}|\rho_k|^{2p}(1+Kh) + K\mathbb{E}\,|\rho_k|^{2p-1}\,(1+|X_k|^{2\alpha})^{1/2}h^{q_2+1/2}\\ &+K\mathbb{E}\,|\rho_k|^{2p-1}\,(1+|X(t_k)|^{2\varkappa-2}+|X_k|^{2\varkappa-2})^{1/4}h^{q_2+1/2}(1+|X_k|^{2\alpha})^{1/2}\\ &+K\mathbb{E}\,|\rho_k|^{2p-1}\,(1+|X(t_k)|^{2\varkappa-2}+|X_k|^{2\varkappa-2})^{1/4}h^{q_2+1/2}(1+|X_k|^{4\alpha})^{1/4}\\ &+K\sum_{l=2}^{2p}\mathbb{E}[|\rho_k|^{2p-l}h^{lq_2}(1+|X_k|^{2\alpha l})^{1/2}]\\ &\leq \mathbb{E}|\rho_k|^{2p}(1+Kh)+K\mathbb{E}\,|\rho_k|^{2p-1}\,(1+|X(t_k)|^{2\varkappa-2}+|X_k|^{2\varkappa-2})^{1/4}h^{q_2+1/2}(1+|X_k|^{2\alpha})^{1/2}\\ &+K\sum_{l=2}^{2p}\mathbb{E}[|\rho_k|^{2p-l}h^{lq_2}(1+|X_k|^{2l\alpha})^{1/2}].\end{aligned}$$

Then using Young's inequality and the conditions (2.5) and (2.12), we obtain

$$\mathbb{E}|\rho_{k+1}|^{2p} \leq \mathbb{E}|\rho_k|^{2p} + Kh\mathbb{E}|\rho_k|^{2p} + K(1+\mathbb{E}|X_0|^{\beta p(\varkappa-1)+2p\alpha\beta})h^{2p(q_2-1/2)+1}$$

whence (2.13) with integer $p \geq 1$ follows by application of Gronwall's inequality. Then by Jensen's inequality (2.13) holds for non-integer p as well. □

B Proof of Lemma 2.1

Lemma 2.1 is an analogue of Lemma 1.1.3 in [15].

Proof. Introduce the process $S_{t,x,y}(s) = S(s) := X_{t,x}(s) - X_{t,y}(s)$ and note that $Z(s) = S(s) - (x-y)$. We first prove (2.15). Using the Ito formula and the condition (2.2) (recall that

(2.2) implies (2.5)), we obtain for $\theta \geq 0$:

$$
\begin{aligned}
\mathbb{E}|S(t+\theta)|^{2p} = |x-y|^{2p} + 2p \int_t^{t+\theta} \mathbb{E}|S|^{2p-2} \Bigg[S^{\intercal}(a(t, X_{t,x}(s)) - a(t, X_{t,y}(s))) \\
+ \frac{1}{2} \sum_{r=1}^{m} |\sigma_r(t, X_{t,x}(s)) - \sigma_r(t, X_{t,y}(s))|^2 \Bigg] \, ds \\
+2p(p-1) \int_t^{t+\theta} \mathbb{E}|S|^{2p-4} \left| S^{\intercal}(s) \sum_{r=1}^{m} [\sigma_r(t, X_{t,x}(s)) - \sigma_r(t, X_{t,y}(s))] \right|^2 ds \\
\leq |x-y|^{2p} + 2p \int_t^{t+\theta} \mathbb{E}|S|^{2p-2} \Bigg[S^{\intercal}(a(t, X_{t,x}(s)) - a(t, X_{t,y}(s))) \\
+ \frac{2p-1}{2} \int_t^{t+\theta} \sum_{r=1}^{m} |\sigma_r(t, X_{t,x}(s)) - \sigma_r(t, X_{t,y}(s))|^2 \Bigg] \, ds \\
\leq |x-y|^{2p} + 2pc_1 \int_t^{t+\theta} \mathbb{E}|S(s)|^{2p} \, ds
\end{aligned}
$$

from which (2.15) follows after applying Gronwall's inequality.

Now we prove (2.16). Using the Ito formula and the condition (2.2), we obtain for $\theta \geq 0$:

$$
\begin{aligned}
\mathbb{E}\,|Z(t+\theta)|^{2p} = 2p \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2} \Bigg[Z^{\intercal}(a(t, X_{t,x}(s)) - a(t, X_{t,y}(s))) \\
+ \frac{1}{2} \sum_{r=1}^{m} |\sigma_r(t, X_{t,x}(s)) - \sigma_r(t, X_{t,y}(s))|^2 \Bigg] \, ds \\
+2p(p-1) \int_t^{t+\theta} \mathbb{E}|Z|^{2p-4} \left| Z^{\intercal} \sum_{r=1}^{m} [\sigma_r(t, X_{t,x}(s)) - \sigma_r(t, X_{t,y}(s))] \right|^2 ds \\
\\
\leq 2p \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2}(s) \Bigg[S^{\intercal}(a(t, X_{t,x}(s)) - a(t, X_{t,y}(s))) \\
+ \frac{2p-1}{2} \int_t^{t+\theta} \sum_{r=1}^{m} |\sigma_r(t, X_{t,x}(s)) - \sigma_r(t, X_{t,y}(s))|^2 \Bigg] \, ds \\
-2p \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2}(x-y, a(t, X_{t,x}(s)) - a(t, X_{t,y}(s))) ds \\
\leq 2pc_1 \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2}|S|^2 \, ds - 2p \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2}(x-y, a(t, X_{t,x}(s)) - a(t, X_{t,y}(s))) ds.
\end{aligned}
\tag{B.1}
$$

Using Young's inequality, we get for the first term in the right-hand side of (B.1):

$$
\begin{aligned}
2pc_1 \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2}|S|^2 \, ds \leq 4pc_1 \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2}(|Z|^2 + |x-y|^2) \, ds \\
\leq K \int_t^{t+\theta} \mathbb{E}|Z|^{2p} ds + K|x-y|^2 \int_t^{t+\theta} \mathbb{E}|Z|^{2p-2} ds.
\end{aligned}
\tag{B.2}
$$

Consider the second term in the right-hand side of (B.1). Using Hoelder's inequality (twice), (2.3), (2.15) and (2.5), we obtain

$$
\begin{aligned}
& -2p\int_t^{t+\theta}\mathbb{E}|Z|^{2p-2}(x-y,a(t,X_{t,x}(s))-a(t,X_{t,y}(s)))ds \\
\leq\ & 2p\int_t^{t+\theta}\mathbb{E}|Z|^{2p-2}|a(t,X_{t,x}(s))-a(t,X_{t,y}(s))||x-y|ds \\
\leq\ & K|x-y|\int_t^{t+\theta}\left[\mathbb{E}|Z|^{2p}\right]^{1-1/p}\left[\mathbb{E}|a(t,X_{t,x}(s))-a(t,X_{t,y}(s))|^p\right]^{1/p}ds
\end{aligned} \tag{B.3}
$$

$$
\begin{aligned}
\leq\ & K|x-y|\int_t^{t+\theta}\left[\mathbb{E}|Z|^{2p}\right]^{1-1/p} \\
& \times(\mathbb{E}[(1+|X_{t,x}(s)|^{2\varkappa-2}+|X_{t,y}(s)|^{2\varkappa-2})^{p/2}|X_{t,x}(s)-X_{t,y}(s)|^p])^{1/p}ds \\
\leq\ & K|x-y|\int_t^{t+\theta}\left[\mathbb{E}|Z|^{2p}\right]^{1-1/p}\left(\mathbb{E}[(1+|X_{t,x}(s)|^{2\varkappa-2}+|X_{t,y}(s)|^{2\varkappa-2})^p]\right)^{1/2p} \\
& \times\left(\mathbb{E}[|X_{t,x}(s)-X_{t,y}(s)|^{2p}]\right)^{1/2p}ds \\
\leq\ & K\,|x-y|^2\,(1+|x|^{2\varkappa-2}+|y|^{2\varkappa-2})^{1/2}\int_t^{t+\theta}\left[\mathbb{E}|Z|^{2p}\right]^{1-1/p}ds.
\end{aligned}
$$

Substituting (B.2) and (B.3) in (B.1) and applying Hoelder's inequality to $\mathbb{E}|Z|^{2p-2}\cdot 1$, we get

$$
\mathbb{E}\,|Z(t+\theta)|^{2p}\leq K\int_t^{t+\theta}\mathbb{E}|Z|^{2p}ds+K\,|x-y|^2\,(1+|x|^{2\varkappa-2}+|y|^{2\varkappa-2})^{1/2}\int_t^{t+\theta}\left[\mathbb{E}|Z|^{2p}\right]^{1-1/p}ds \tag{B.4}
$$

whence we obtain (2.16) for integer $p\geq 1$ using Gronwall's inequality as, e.g. in [19, p. 360], and then by Jensen's inequality for non-integer $p>1$ as well. □