

# **Lower error bounds for strong approximation of scalar SDEs with non-Lipschitzian coefficients**

**M. Hefter, A. Herzwurm, T.  
Müller-Gronbach**

**RICAM-Report 2017-38**

# LOWER ERROR BOUNDS FOR STRONG APPROXIMATION OF SCALAR SDES WITH NON-LIPSCHITZIAN COEFFICIENTS

MARIO HEFTER, ANDRÉ HERZWURM, AND THOMAS MÜLLER-GRONBACH

ABSTRACT. We study pathwise approximation of scalar stochastic differential equations at a single time point or globally in time by means of methods that are based on finitely many observations of the driving Brownian motion. We prove lower error bounds in terms of the average number of evaluations of the driving Brownian motion that hold for every such method under rather mild assumptions on the coefficients of the equation. The underlying simple idea of our analysis is as follows: the lower error bounds known for equations with coefficients that have sufficient regularity globally in space should still apply in the case of coefficients that have this regularity in space only locally, in a small neighborhood of the initial value. Our results apply to a huge variety of equations with coefficients that are not globally Lipschitz continuous in space including Cox-Ingersoll-Ross processes, equations with superlinearly growing coefficients, and equations with discontinuous coefficients. In many of these cases the resulting lower error bounds even turn out to be sharp.

## CONTENTS

1. Introduction	2
2. Notation	5
3. Setting	5
4. The Class of Methods	6
5. Lower Error Bounds for Strong Approximation at a Single Point	7
5.1. Global Assumptions	7
5.2. Local Assumptions	11
6. Lower Error Bounds for Strong Approximation Globally in Time	13
6.1. $L_\infty$ -Approximation	14
6.2. $L_p$ -Approximation	16
6.3. Maximum Pointwise Error	19
7. Examples	20
7.1. Cox-Ingersoll-Ross Processes	20
7.1.1. Strong Approximation at a Single Point	20
7.1.2. Strong Approximation Globally in Time	21
7.2. SDEs with Superlinearly Growing Coefficients	23
7.3. SDEs with Discontinuous Coefficients	23
Appendix A. Properties of the Normal Distribution	24
Appendix B. Localization Technique: Comparison Results	26
Acknowledgment	30
References	30

---

2010 *Mathematics Subject Classification.* 65C30, 60H10.

*Key words and phrases.* stochastic differential equations; non-globally Lipschitz continuous coefficients; strong (pathwise) approximation; lower error bounds.

## 1. INTRODUCTION

Let  $T > 0$  and consider a scalar stochastic differential equation (SDE)

$$(1) \quad dX(t) = a(t, X(t)) dt + b(t, X(t)) dW(t), \quad t \in [0, T],$$

with drift coefficient  $a: [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ , diffusion coefficient  $b: [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ , one-dimensional driving Brownian motion  $W$ , and initial value  $X(0)$  such that (1) has a solution  $X = (X(t))_{t \in [0, T]}$ . The computational problem we study is strong approximation of the solution  $X$ , either globally on the whole time interval  $[0, T]$  or at the final time  $T$ , by means of methods that may use the initial value  $X(0)$  and a finite number of sequentially taken evaluations  $W(\tau_1), \dots, W(\tau_\nu)$  of the driving Brownian motion  $W$  at times  $\tau_1, \dots, \tau_\nu \in [0, T]$ . Except for measurability conditions we do not impose any further restrictions. The  $k$ -th site  $\tau_k$  may depend on the previous evaluations  $X(0), W(\tau_1), \dots, W(\tau_{k-1})$ , e.g., by using a path-dependent step size control, and the total number  $\nu$  of observations of  $W$  may be determined by a stopping rule. Finally, the resulting discrete data may be used in any way to generate an approximation to  $X$  or to  $X(T)$ . See Section 4 for a formal description of such approximations. Our goal is to establish lower error bounds that hold for any such method, in terms of the average number  $\mathbb{E}[\nu]$  of evaluations of  $W$  that are used.

Lower error bounds for strong approximation of (systems of) SDEs based on evaluations of the driving Brownian motion at finitely many times were first established in 1980 by Clark and Cameron [10] in the particular case of strong approximation of Lévy areas. Meanwhile, lower error bounds have extensively been studied in the case of coefficients that are globally Lipschitz continuous in the state variable and sufficiently smooth, see Cambanis and Hu [9], Hofmann, Müller-Gronbach, and Ritter [22, 23, 24], Müller-Gronbach [36, 37, 38], Rümelin [46]. Moreover, under the assumption of global Lipschitz continuity in the state variable, lower error bound results are also available for equations with coefficients that are discontinuous in time, see Przybyłowicz [43, 44, 45], for stochastic delay differential equations, see Hofmann and Müller-Gronbach [21], and for equations driven by a fractional Brownian motion, see Neuenkirch [40, 41], Neuenkirch and Shalaiko [42].

For SDEs with coefficients that are not globally Lipschitz continuous in the state variable investigations on lower error bounds have started only recently. There seem to be two directions of research, up to now. One of them consists in establishing sub-polynomial lower error bounds for particular equations with smooth coefficients in order to come closer to a characterization of polynomial convergence in that case, see Gerencsér, Jentzen, and Salimova [13], Hairer, Hutzenthaler, and Jentzen [16], Jentzen, Müller-Gronbach, and Yaroslavl'tseva [28], Müller-Gronbach and Yaroslavl'tseva [39], Yaroslavl'tseva [50]. The other one aims at a thorough analysis of strong approximation of Cox-Ingersoll-Ross processes as a prototype of SDEs with a diffusion coefficient that is Hölder continuous in the state variable with a Hölder exponent strictly between zero and one, see Hefter and Herzwurm [17, 18], Hefter and Jentzen [19].

In the present paper we aim at scalar equations (1) with coefficients  $a$  and  $b$  that are not globally Lipschitz continuous in the state variable and we establish lower error bounds under rather mild assumptions on  $a$  and  $b$  by exploiting, essentially, the following simple idea: it is likely that the lower error bounds known for equations with coefficients that have sufficient regularity globally in space still apply in the case of coefficients that have this regularity in space only locally, in a small neighborhood of the initial value.

To give a flavour of our results we consider for simplicity the particular case of an autonomous equation (1), i.e., we assume that

- (A)  $(\Omega, \mathfrak{F}, \mathbb{P})$  is a complete probability space with a normal filtration  $(\mathcal{F}_t)_{t \in [0, T]}$ ,  $W: [0, T] \times \Omega \rightarrow \mathbb{R}$  is a standard Brownian motion on  $(\Omega, \mathfrak{F}, \mathbb{P})$  with respect to  $(\mathcal{F}_t)_{t \in [0, T]}$ ,  $a: \mathbb{R} \rightarrow \mathbb{R}$  and  $b: \mathbb{R} \rightarrow \mathbb{R}$  are Borel-measurable functions, and  $X: [0, T] \times \Omega \rightarrow \mathbb{R}$

is an  $(\mathcal{F}_t)_{t \in [0, T]}$ -adapted stochastic process with continuous sample paths such that  $\mathbb{P}$ -a.s.  $\int_0^T (|a(X(t))| + |b(X(t))|^2) dt < \infty$  and for all  $t \in [0, T]$   $\mathbb{P}$ -a.s.

$$X(t) = X(0) + \int_0^t a(X(s)) ds + \int_0^t b(X(s)) dW(s),$$

and we restrict ourselves to approximations of the solution  $X$  that are based on finitely many evaluations of the driving Brownian motion at fixed times in  $[0, T]$ . Note, however, that all of the following lower error bounds also hold for approximations that use  $n$  sequentially taken evaluations of the driving Brownian motion  $W$  on average, see Sections 5 and 6.

We first consider strong approximation of the solution globally on  $[0, T]$  with respect to the supremum-norm. The following result is an immediate consequence of Theorem 9 in Section 6.1.

**Theorem 1** ( $L_\infty$ -approximation). *Assume (A). Let  $t_0 \in [0, T]$  and let  $\emptyset \neq I \subseteq \mathbb{R}$  be an open interval such that*

- (i)  $a, b$  are once continuously differentiable on  $I$ ,
- (ii)  $\forall x \in I: b(x) \neq 0$ ,
- (iii)  $\mathbb{P}(X(t_0) \in I) > 0$ .

*Then there exist constants  $c, \gamma \in (0, \infty)$  such that for all  $n \in \mathbb{N}$ , for all  $s_1, \dots, s_n \in [0, T]$  and for all measurable mappings  $u: \mathbb{R}^{n+1} \rightarrow C([0, T])$  we have*

$$\mathbb{P}(\|X - u(X(0), W(s_1), \dots, W(s_n))\|_\infty \geq c \cdot \sqrt{\ln(n+1)/n}) \geq \gamma.$$

*In particular, we have for all  $n \in \mathbb{N}$  that*

$$\inf_{\substack{s_1, \dots, s_n \in [0, T] \\ u: \mathbb{R}^{n+1} \rightarrow C([0, T]) \text{ measurable}}} \mathbb{E}[\|X - u(X(0), W(s_1), \dots, W(s_n))\|_\infty] \geq c\gamma \cdot \sqrt{\ln(n+1)/n}.$$

Under slightly stronger smoothness assumptions on the coefficients  $a$  and  $b$  we obtain lower bounds for the error with respect to the  $L_1([0, T])$ -norm (denoted by  $\|\cdot\|_1$ ), which in turn implies a lower bound for the maximum pointwise approximation error. The following result is an immediate consequence of Theorem 12 in Section 6.2 and Theorem 13 in Section 6.3.

**Theorem 2** ( $L_1$ -approximation & maximum pointwise error). *Assume (A). Let  $t_0 \in [0, T]$  and let  $\emptyset \neq I \subseteq \mathbb{R}$  be an open interval such that*

- (i)  $a, b$  are twice continuously differentiable on  $I$ ,
- (ii)  $\forall x \in I: b(x) \neq 0$ ,
- (iii)  $\mathbb{P}(X(t_0) \in I) > 0$ .

*Then there exist constants  $c, \gamma \in (0, \infty)$  such that for all  $n \in \mathbb{N}$ , for all  $s_1, \dots, s_n \in [0, T]$  and for all measurable mappings  $u: \mathbb{R}^{n+1} \rightarrow L_1([0, T])$  we have*

$$\mathbb{P}(\|X - u(X(0), W(s_1), \dots, W(s_n))\|_1 \geq c/\sqrt{n}) \geq \gamma.$$

*In particular, we have for all  $n \in \mathbb{N}$  that*

$$\inf_{\substack{s_1, \dots, s_n \in [0, T] \\ u: \mathbb{R}^{n+1} \rightarrow L_1([0, T]) \text{ measurable}}} \mathbb{E}[\|X - u(X(0), W(s_1), \dots, W(s_n))\|_1] \geq c\gamma \cdot n^{-1/2}$$

and

$$\inf_{\substack{s_1, \dots, s_n \in [0, T] \\ u: \mathbb{R}^{n+1} \rightarrow C([0, T]) \text{ measurable}}} \left\{ \sup_{t \in [0, T]} \mathbb{E}[|X(t) - (u(X(0), W(s_1), \dots, W(s_n)))(t)|] \right\} \geq c\gamma/T \cdot n^{-1/2}.$$

Finally, we consider strong approximation of the solution at the final time. The following result is an immediate consequence of Theorem 6 in Section 5.2.

**Theorem 3** (Pointwise approximation). *Assume (A). Let  $t_0 \in [0, T]$  and let  $\emptyset \neq I \subseteq \mathbb{R}$  be an open interval such that*

- (i)  $a, b$  are three times continuously differentiable on  $I$ ,
- (ii)  $\forall x \in I: b(x) \neq 0$  and  $(a'b - ab' - \frac{1}{2}b^2b'')(x) \neq 0$ ,
- (iii)  $\mathbb{P}(X(t_0) \in I) > 0$ .

*Then there exist constants  $c, \gamma \in (0, \infty)$  such that for all  $n \in \mathbb{N}$ , for all  $s_1, \dots, s_n \in [0, T]$  and for all measurable mappings  $u: \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  we have*

$$\mathbb{P}(|X(T) - u(X(0), W(s_1), \dots, W(s_n))| \geq c/n) \geq \gamma.$$

*In particular, we have for all  $n \in \mathbb{N}$  that*

$$\inf_{\substack{s_1, \dots, s_n \in [0, T] \\ u: \mathbb{R}^{n+1} \rightarrow \mathbb{R} \text{ measurable}}} \mathbb{E}[|X(T) - u(X(0), W(s_1), \dots, W(s_n))|] \geq c\gamma \cdot n^{-1}.$$

We stress that up to now the lower bounds on the mean errors in Theorems 1-3 were known to hold only under assumptions on the coefficients  $a$  and  $b$  and the initial value  $X(0)$  that are much stronger than the conditions (i)-(iii) used in the above theorems. For instance, all derivatives of  $a$  and  $b$ , which appear in Theorems 1-3, are typically assumed to exist on the whole real line and to be bounded and the initial value  $X(0)$  is required to satisfy a moment condition, see, e.g., Müller-Gronbach [36, 37, 38] for further details and references. Moreover, Theorems 1-3 provide error bounds that hold with positive probability, uniformly in  $n$ . To the best of our knowledge such error estimates have not been established in this generality in the literature so far.

The second condition in Assumption (ii) in Theorem 3 requires some motivation, which we take from a related discussion in Müller-Gronbach [38]. First note that the function  $a'b - ab' - b^2b''/2$  is the Lie bracket  $[\tilde{a}, b]$ , where  $\tilde{a} = a - bb'/2$  is the drift coefficient of the Stratonovich equation corresponding to the Itô coefficients  $a$  and  $b$ . By a general result of Yamato [49, Theorem 2.1], which links the representability of the solution  $X$  in terms of multiple Itô integrals to the nilpotent property of the Lie algebras associated with  $\tilde{a}$  and  $b$ , it follows that  $X(T) = u(X(0), W(T))$  for some measurable function  $u: \mathbb{R}^2 \rightarrow \mathbb{R}$  if  $\tilde{a}, b \in C^\infty(\mathbb{R})$  and  $[\tilde{a}, b] = 0$ . A special example is provided by a geometric Brownian motion, where  $a(x) = \alpha x$ ,  $b(x) = \beta x$  with  $\alpha, \beta \in \mathbb{R}$ , and  $X(T) = X(0) \exp((\alpha - \beta^2/2)T + \beta W(T))$ . In particular,  $X(T)$  can then be approximated with error zero based only on  $X(0)$  and  $W(T)$  and the lower error bound in Theorem 3 cannot hold. Thus, roughly speaking, the second condition in (ii) in Theorem 3 (and its generalized version for non-autonomous equations in Proposition 4) excludes trivial approximation problems for SDEs such as a geometric Brownian motion or a one-dimensional squared Bessel process, see Section 7.1.

Theorems 1-3 yield lower error bounds for a huge variety of SDEs including Cox-Ross-Ingersoll processes, see Section 7.1, equations with superlinearly growing coefficients, see Section 7.2, and equations with discontinuous coefficients, see Section 7.3. In many of these cases these lower error bounds turn out to be sharp. Here, we illustrate our results by considering an SDE with a superlinearly growing drift coefficient, namely

$$(2) \quad \begin{aligned} dX(t) &= -(X(t))^5 dt + X(t) dW(t), \quad t \in [0, T], \\ X(0) &= x_0 \in \mathbb{R} \setminus \{0\}. \end{aligned}$$

Note that (2) has a unique strong solution since both coefficients are locally Lipschitz continuous and jointly satisfy a suitable monotone condition, see, e.g., Mao [33]. Clearly,  $(a'b - ab' - b^2b''/2)(x) = -4x^5$  for all  $x \in \mathbb{R}$ , so that all of the assumptions in Theorems 1-3 are satisfied and therefore all of the respective lower error bounds hold true. On the other hand, a tamed version of the equidistant Milstein scheme achieves the upper bound  $1/n$ , up to a constant, for the error at the final time in the  $p$ -th mean, for any  $p \in [1, \infty)$ , see

Kumar and Sabanis [30], while the piecewise linear interpolation of a tamed version of the Euler scheme achieves the upper bounds  $n^{-1/2}$  and  $\sqrt{\ln(n+1)/n}$ , up to a constant each, for the maximum pointwise error in the  $p$ -th mean (and thus the pathwise  $L_p$ -error in the  $p$ -th mean) and for the uniform error in the  $p$ -th mean, respectively, for any  $p \in [1, \infty)$ , see Hutzenthaler, Jentzen, and Kloeden [26] and Hutzenthaler, Jentzen, and Kloeden [27], respectively. Hence all of the lower error bounds from Theorems 1-3 are sharp and both methods just mentioned perform asymptotically optimal for equation (2).

The present paper only addresses scalar SDEs and strong approximations based on evaluations of the driving Brownian motion at single times in  $[0, T]$ . However, our proof techniques can also be applied in the case of systems of SDEs and approximations based on bounded linear functionals of the driving Brownian motion or iterated Itô integrals, and we therefore expect an analogue transfer of the lower error bounds known for such methods in the case of systems of SDEs with coefficients that are globally Lipschitz with respect to the state variable, see, e.g., Hofmann and Müller-Gronbach [20], Hofmann, Müller-Gronbach, and Ritter [25], Müller-Gronbach [37], to the case of systems of SDEs with coefficients that behave sufficiently well with respect to the state variable only in a neighborhood of the initial value.

Our paper is organized as follows. In Section 2 and Section 3 we introduce some notation and fix the setting with respect to equation (1), respectively. In Section 4 we thoroughly explain what kind of approximation methods are covered by the lower error bounds. In Section 5 we establish lower error bounds for pointwise approximation whereas Section 6 provides lower error bounds with respect to global approximation. In Section 7 we then use the results from the latter two sections to obtain lower error bounds for strong approximation of Cox-Ingersoll-Ross processes, equations with superlinearly growing coefficients, and equations with discontinuous coefficients. The proofs of our lower error bounds rely on a localization technique, which provides a link between equations with coefficients that are globally Lipschitz in space and equations with coefficients that behave well only in a small neighbourhood of the initial value. This tool as well as a number of needed properties of Gaussian distributions are provided in an appendix.

## 2. NOTATION

For  $T \in (0, \infty)$  and  $p \in [1, \infty)$  we use  $L_p([0, T])$  to denote the Banach space of all Borel-measurable functions  $f: [0, T] \rightarrow \mathbb{R}$  with  $\|f\|_p = (\int_0^T |f(t)|^p dt)^{1/p} < \infty$ , where two functions are identified if they coincide Lebesgue-almost everywhere on  $[0, T]$ . By  $C([0, T])$  we denote the Banach space of all continuous functions  $f: [0, T] \rightarrow \mathbb{R}$  equipped with the norm  $\|f\|_\infty = \sup_{t \in [0, T]} |f(t)|$ . Furthermore, for  $d \in \mathbb{N}$ ,  $x \in \mathbb{R}^d$ , and  $p \in [1, \infty)$  we use  $|x|_p = (\sum_{i=1}^d |x_i|^p)^{1/p}$  to denote the  $p$ -norm of  $x$  and we use  $|x|_\infty = \max_{i=1, \dots, d} |x_i|$  to denote the maximum norm of  $x$ .

## 3. SETTING

Throughout this article we fix the following scenario of a scalar SDE. Let  $(\Omega, \mathfrak{F}, \mathbb{P})$  be a complete probability space with a normal filtration  $(\mathcal{F}_t)_{t \in [0, \infty)}$  and let  $W: [0, \infty) \times \Omega \rightarrow \mathbb{R}$  be a standard Brownian motion on  $(\Omega, \mathfrak{F}, \mathbb{P})$  with respect to  $(\mathcal{F}_t)_{t \in [0, \infty)}$ . Let  $T \in (0, \infty)$ , let  $a: [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  and  $b: [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  be Borel-measurable functions, and let  $X: [0, T] \times \Omega \rightarrow \mathbb{R}$  be an  $(\mathcal{F}_t)_{t \in [0, T]}$ -adapted stochastic process with continuous sample paths such that  $\mathbb{P}$ -a.s.  $\int_0^T (|a(t, X(t))| + |b(t, X(t))|^2) dt < \infty$  and for all  $t \in [0, T]$   $\mathbb{P}$ -a.s.

$$(3) \quad X(t) = X(0) + \int_0^t a(s, X(s)) ds + \int_0^t b(s, X(s)) dW(s).$$

## 4. THE CLASS OF METHODS

Consider the setting in Section 3 and fix  $n \in \mathbb{N}$ . We formally explain what we mean by an approximation of  $X(T)$  or of  $X$  that is based on the initial value  $X(0)$  and on at most  $n$  sequential evaluations of the Brownian motion  $W$  at points in  $[0, \infty)$  on average.

To this end we consider, more generally, a measurable space  $(S, \mathfrak{F})$  and we introduce the class  $\mathcal{A}_n(S, X(0), W)$  of all  $\mathfrak{F}$ - $S$ -measurable mappings  $V: \Omega \rightarrow S$  that can be constructed using  $X(0)$  and at most  $n$  sequential evaluations of the Brownian motion  $W$  in  $[0, \infty)$  on average. In Section 5 we take  $S = \mathbb{R}$  to study approximation of  $X(T)$ . In Section 6 we take  $S = L_p([0, T])$  or  $S = C([0, T])$  to study approximation of the whole process  $X$ . In either of these cases  $\mathfrak{F}$  is taken to be the Borel  $\sigma$ -field generated by the respective canonical norm.

Every random variable  $V \in \mathcal{A}_n(S, X(0), W)$  is determined by three sequences

$$\psi = (\psi_k)_{k \in \mathbb{N}}, \quad \chi = (\chi_k)_{k \in \mathbb{N}}, \quad \varphi = (\varphi_k)_{k \in \mathbb{N}}$$

of measurable mappings

$$\begin{aligned} \psi_k &: \mathbb{R}^k \rightarrow [0, \infty), \\ \chi_k &: \mathbb{R}^{k+1} \rightarrow \{\text{STOP}, \text{GO}\}, \\ \varphi_k &: \mathbb{R}^{k+1} \rightarrow S. \end{aligned}$$

The sequence  $\psi$  is used to determine the sequential evaluation sites for the Brownian motion  $W$  in  $[0, \infty)$ . The sequence  $\chi$  determines when to stop the evaluation of  $W$ . The sequence  $\varphi$  is used to obtain the outcome of  $V$  in  $S$  once the evaluation of  $W$  has stopped.

To be more precise, let  $x \in \mathbb{R}$  be a possible realization of  $X(0)$  and let  $w \in C([0, \infty))$  be a possible realization of  $W$ . We put  $D_0(x, w) = x$ , and for  $k \in \mathbb{N}$  we recursively define

$$\tau_k(x, w) = \psi_k(D_{k-1}(x, w)), \quad Y_k(x, w) = w(\tau_k(x, w)), \quad D_k(x, w) = (D_{k-1}(x, w), Y_k(x, w)).$$

Thus  $\tau_k(x, w) \in [0, \infty)$  is the  $k$ -th evaluation node for the actual path  $w$  of  $W$  and  $Y_k(x, w) \in \mathbb{R}$  is the corresponding evaluation of  $w$ . The vector  $D_k(x, w) \in \mathbb{R}^{k+1}$  contains the available data about  $x$  and  $w$  after  $k$  steps and we decide whether to stop or to go on with the sequential evaluation of  $w$  according to the value of  $\chi_k(D_k(x, w))$ . The total number of evaluations of  $w$  is thus given by

$$\nu(x, w) = \min\{k \in \mathbb{N}: \chi_k(D_k(x, w)) = \text{STOP}\} \in \mathbb{N} \cup \{\infty\}.$$

We require that the triple of sequences  $(\psi, \chi, \varphi)$  satisfies  $\mathbb{P}(\nu(X(0), W) < \infty) = 1$  and we define  $V: \Omega \rightarrow S$  to be a random variable that satisfies  $\mathbb{P}$ -a.s.

$$(4) \quad V = \varphi_{\nu(X(0), W)}(D_{\nu(X(0), W)}(X(0), W)).$$

Now we put

$$\mathcal{A}_n(S, X(0), W) = \{V: \Omega \rightarrow S: V \text{ is of the form (4) with } \mathbb{E}[\nu(X(0), W)] \leq n\}.$$

The class  $\mathcal{A}_n(S, X(0), W)$  is called the class of adaptive methods (with varying cardinality). For  $V \in \mathcal{A}_n(S, X(0), W)$ , the corresponding average number  $\mathbb{E}[\nu(X(0), W)]$  of evaluation nodes of the Brownian motion  $W$  can be seen as a rough measure for the computational cost of the method  $V$ . Prominent examples of such adaptive methods are Euler- or Milstein-type methods with a step size control where the average number of evaluation nodes is bounded by  $n$ , see, e.g., Burrage, Herdiana, and Burrage [7], Gaines and Lyons [12], Hofmann, Müller-Gronbach, and Ritter [22, 23, 24], Mauthner [34], Moon, Szepessy, Tempone, and Zouraris [35], Müller-Gronbach [36, 37, 38].

An important subclass of  $\mathcal{A}_n(S, X(0), W)$  is the class  $\mathcal{A}_n^{\text{eq}}(S, X(0), W)$  of all methods that are based on  $n$  evaluations of  $W$  at the equidistant times  $kT/n$ ,  $k = 0, 1, \dots, n$ . Formally, we have  $\chi_k = \text{GO}$  for  $k < n$ ,  $\chi_n = \text{STOP}$  and  $\psi_k = Tk/\nu$  for  $k \leq n$ . Thus

$\mathcal{A}_n^{\text{eq}}(S, X(0), W) = \{u(X(0), W(T/n), W(2T/n), \dots, W(T)): u: \mathbb{R}^{n+1} \rightarrow S \text{ is measurable}\}$  and we have  $\mathcal{A}_n^{\text{eq}}(S, X(0), W) \subseteq \mathcal{A}_n(S, X(0), W)$ . The lower error bounds established in Sections 5 and 6 hold for any method from the class  $\mathcal{A}_n(S, X(0), W)$  for the respective choice of  $S$ . In Section 7 we will see that matching upper error bounds are often (but not always) achieved by methods from the class  $\mathcal{A}_n^{\text{eq}}(S, X(0), W)$ .

For technical reasons we introduce a further class of algorithms. By  $\mathbb{A}_n(S, W)$  we denote the set of all random variables  $V: \Omega \rightarrow S$  that are measurable with respect to the  $\sigma$ -algebra generated by  $\mathcal{F}_0, W(\tau_1), \dots, W(\tau_n)$ , where  $\tau_1, \dots, \tau_n: \Omega \rightarrow [0, \infty)$  are any random variables such that for all  $k = 1, \dots, n$  the random variable  $\tau_k$  is measurable with respect to the  $\sigma$ -algebra generated by  $\mathcal{F}_0, W(\tau_1), \dots, W(\tau_{k-1})$ .

## 5. LOWER ERROR BOUNDS FOR STRONG APPROXIMATION AT A SINGLE POINT

In this section we consider strong approximation of  $X(T)$ . In view of Section 4, we study lower error bounds for approximation methods  $\widehat{X}_n(T)$  belonging to the class  $\mathcal{A}_n(\mathbb{R}, X(0), W)$ . In Section 5.1, we extend results from Müller-Gronbach [38] and prove probability bounds on the error of  $\widehat{X}_n(T) \in \mathbb{A}_n(\mathbb{R}, W)$  under rather restrictive global smoothness assumptions on the coefficients  $a$  and  $b$  of the SDE (3). In Section 5.2, we switch to rather mild local smoothness assumptions on  $a$  and  $b$  and prove our main result on lower error bounds for pointwise approximation, see Theorem 6.

**5.1. Global Assumptions.** The following proposition shows that for sufficiently smooth coefficients  $a$  and  $b$ , the probability for an adaptive method  $\widehat{X}_n(T) \in \mathbb{A}_n(\mathbb{R}, W)$  of having an error at least of order one is arbitrarily close to one.

**Proposition 4** (Probability bounds on the pointwise error of adaptive methods I). *Assume the setting in Section 3 with  $\mathbb{E}[|X(0)|^{16}] < \infty$ . Assume further that there exists a Borel set  $I \subseteq \mathbb{R}$  such that*

- (A1)  $\forall i \in \{0, 1, 2\}, j \in \{0, 1, 2, 3\}$ : the partial derivatives  $a^{(i,j)}, b^{(i,j)}$  exist on  $[0, T] \times \mathbb{R}$  and are continuous and bounded,
- (A2)  $\forall x \in I$ :  $(a^{(0,1)} b - b^{(1,0)} - a b^{(0,1)} - \frac{1}{2} b^2 b^{(0,2)})(0, x) \neq 0$ ,
- (A3)  $\mathbb{P}(X(0) \in I) = 1$ .

Then for every  $\varepsilon \in (0, 1)$  there exists  $c \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  and for all  $\widehat{X}_n(T) \in \mathbb{A}_n(\mathbb{R}, W)$  we have

$$\mathbb{P}(|X(T) - \widehat{X}_n(T)| \geq c/n) \geq 1 - \varepsilon.$$

*Proof.* This proof is based on the techniques and the notation developed in Müller-Gronbach [38] for the  $p$ -th mean error analysis of pointwise approximation of scalar SDEs. We provide the corresponding tools and results from the latter work that are needed in the present context. The assumption  $\mathbb{E}[|X(0)|^{16}] < \infty$  corresponds to the case  $p = 1$  in Müller-Gronbach [38, Condition (B), p. 1609].

Without loss of generality we may assume  $T = 1$ . For every  $t \in [0, 1]$  we put

$$\mathcal{Y}(t) = \mathcal{M}(t) \cdot \mathcal{G}(t, X(t))$$

where

$$\mathcal{G} = a^{(0,1)} b - b^{(1,0)} - a b^{(0,1)} - \frac{1}{2} b^2 b^{(0,2)}$$



and

$$\mathcal{M}(t) = \exp\left(\int_t^1 \left(a^{(0,1)} - 1/2 \cdot (b^{(0,1)})^2\right)(u, X(u)) du + \int_t^1 b^{(0,1)}(u, X(u)) dW(u)\right).$$

The analysis in Müller-Gronbach [38] shows that the problem of pathwise approximation of  $X(1)$  based on finitely many evaluations of the driving Brownian motion  $W$  is strongly connected to an integration problem for  $W$  with a random weight given by the process  $\mathcal{Y}$ .

For  $k \in \mathbb{N}$  we put  $t_\ell = \ell/k$  for  $\ell = 0, \dots, k$  and we define a truncated Wagner-Platen scheme  $\widehat{X}_k^{\text{WPt}} = (\widehat{X}_k^{\text{WPt}}(t_\ell))_{\ell=0, \dots, k}$  by  $\widehat{X}_k^{\text{WPt}}(0) = X(0)$  and for  $\ell = 0, \dots, k-1$ ,

$$\begin{aligned} & \widehat{X}_k^{\text{WPt}}(t_{\ell+1}) \\ &= \widehat{X}_k^{\text{WPt}}(t_\ell) + a(t_\ell, \widehat{X}_k^{\text{WPt}}(t_\ell)) \cdot (t_{\ell+1} - t_\ell) + b(t_\ell, \widehat{X}_k^{\text{WPt}}(t_\ell)) \cdot (W(t_{\ell+1}) - W(t_\ell)) \\ & \quad + 1/2 \cdot (bb^{(0,1)})(t_\ell, \widehat{X}_k^{\text{WPt}}(t_\ell)) \cdot ((W(t_{\ell+1}) - W(t_\ell))^2 - (t_{\ell+1} - t_\ell)) \\ & \quad + (b^{(1,0)} + ab^{(0,1)} - 1/2 \cdot b(b^{(0,1)})^2)(t_\ell, \widehat{X}_k^{\text{WPt}}(t_\ell)) \cdot (W(t_{\ell+1}) - W(t_\ell)) \cdot (t_{\ell+1} - t_\ell) \\ & \quad + 1/6 \cdot (b(b^{(0,1)})^2 + b^2b^{(0,2)})(t_\ell, \widehat{X}_k^{\text{WPt}}(t_\ell)) \cdot (W(t_{\ell+1}) - W(t_\ell))^3 \\ & \quad + 1/2 \cdot (a^{(1,0)} + aa^{(0,1)} + 1/2 \cdot b^2a^{(0,2)})(t_\ell, \widehat{X}_k^{\text{WPt}}(t_\ell)) \cdot (t_{\ell+1} - t_\ell)^2. \end{aligned}$$

The scheme  $\widehat{X}_k^{\text{WPt}}$  is used for two purposes. First, we obtain a discrete-time approximation  $\widehat{\mathcal{Y}}_k = (\widehat{\mathcal{Y}}_k(t_\ell))_{\ell=0, \dots, k-1}$  to the random weight  $\mathcal{Y}$  in the following way. Put

$$\widehat{m}_\ell = 1 + a^{(0,1)}(t_\ell, \widehat{X}_k^{\text{WPt}}(t_\ell)) \cdot (t_{\ell+1} - t_\ell) + b^{(0,1)}(t_\ell, \widehat{X}_k^{\text{WPt}}(t_\ell)) \cdot (W(t_{\ell+1}) - W(t_\ell))$$

for  $\ell = 0, \dots, k-1$ , define a discrete-time Euler-type approximation  $\widehat{\mathcal{M}}_k = (\widehat{\mathcal{M}}_k(t_\ell))_{\ell=0, \dots, k}$  of the process  $\mathcal{M}$  by

$$\widehat{\mathcal{M}}_k(t_\ell) = \begin{cases} \widehat{m}_\ell \cdots \widehat{m}_{k-1}, & \text{if } \ell \leq k-1, \\ 1, & \text{if } \ell = k, \end{cases}$$

and define for  $\ell = 0, \dots, k-1$ ,

$$\widehat{\mathcal{Y}}_k(t_\ell) = \widehat{\mathcal{M}}_k(t_{\ell+1}) \cdot \mathcal{G}(t_\ell, \widehat{X}_k^{\text{WPt}}(t_\ell)).$$

Second, we define an auxiliary scheme  $\overline{X}_k^{\text{aux}} = (\overline{X}_k^{\text{aux}}(t_\ell))_{\ell=0, \dots, k}$  by

$$\overline{X}_k^{\text{aux}}(t_\ell) = \widehat{X}_k^{\text{WPt}}(t_\ell) + \overline{Q}_k(t_\ell),$$

where  $\overline{Q}_k = (\overline{Q}_k(t_\ell))_{\ell=0, \dots, k}$  is defined by  $\overline{Q}_k(0) = 0$  and

$$\overline{Q}_k(t_{\ell+1}) = \widehat{m}_\ell \cdot \overline{Q}_k(t_\ell) + \mathcal{G}(t_\ell, \widehat{X}_k^{\text{WPt}}(t_\ell)) \cdot \int_{t_\ell}^{t_{\ell+1}} (W(t) - W(t_\ell)) dt$$

for  $\ell = 0, \dots, k-1$ . Observe that for  $\ell = 0, \dots, k$  it holds that

$$\overline{Q}_k(t_\ell) = \sum_{r=0}^{\ell-1} \left[ \left( \mathcal{G}(t_r, \widehat{X}_k^{\text{WPt}}(t_r)) \cdot \int_{t_r}^{t_{r+1}} (W(t) - W(t_r)) dt \right) \cdot \prod_{j=r+1}^{\ell-1} \widehat{m}_j \right].$$

In particular, we have

$$\begin{aligned}\bar{Q}_k(1) &= \sum_{r=0}^{k-1} \left[ \left( \mathcal{G}(t_r, \hat{X}_k^{\text{WPt}}(t_r)) \cdot \int_{t_r}^{t_{r+1}} (W(t) - W(t_r)) dt \right) \cdot \prod_{j=r+1}^{k-1} \hat{m}_j \right] \\ &= \sum_{r=0}^{k-1} \left[ \left( \mathcal{G}(t_r, \hat{X}_k^{\text{WPt}}(t_r)) \cdot \int_{t_r}^{t_{r+1}} (W(t) - W(t_r)) dt \right) \cdot \hat{\mathcal{M}}_k(t_{r+1}) \right] \\ &= \sum_{r=0}^{k-1} \left[ \hat{\mathcal{Y}}_k(t_r) \cdot \int_{t_r}^{t_{r+1}} (W(t) - W(t_r)) dt \right].\end{aligned}$$

Hence we get

$$(5) \quad \bar{X}_k^{\text{aux}}(1) = \hat{X}_k^{\text{WPt}}(1) + \sum_{r=0}^{k-1} \left[ \hat{\mathcal{Y}}_k(t_r) \cdot \int_{t_r}^{t_{r+1}} (W(t) - W(t_r)) dt \right].$$

We have

$$(6) \quad \exists c_1 \in (0, \infty) \quad \forall k \in \mathbb{N}: \quad \max_{\ell=0, \dots, k-1} \mathbb{E}[|\mathcal{Y}(t_\ell) - \hat{\mathcal{Y}}_k(t_\ell)|] \leq c_1 \cdot k^{-1/2},$$

see Müller-Gronbach [38, Lemma 2, p. 1626], and

$$(7) \quad \exists c_2 \in (0, \infty) \quad \forall k \in \mathbb{N}: \quad \mathbb{E}[|X(1) - \bar{X}_k^{\text{aux}}(1)|] \leq c_2 \cdot k^{-3/2},$$

see Müller-Gronbach [38, Lemma 12, p. 1639].

For  $k \in \mathbb{N}$  we put

$$R_k = \frac{1}{k} \cdot \sum_{\ell=0}^{k-1} |\mathcal{Y}(t_\ell)|^{2/3}, \quad \hat{R}_k = \frac{1}{k} \cdot \sum_{\ell=0}^{k-1} |\hat{\mathcal{Y}}_k(t_\ell)|^{2/3}.$$

Note that the process  $\mathcal{Y}$  has continuous paths. Hence,  $\mathbb{P}$ -a.s.,

$$(8) \quad \lim_{k \rightarrow \infty} R_k = \int_0^1 |\mathcal{Y}(t)|^{2/3} dt.$$

Moreover, since

$$\begin{aligned}\mathbb{E}[|R_k - \hat{R}_k|] &= \frac{1}{k} \cdot \mathbb{E} \left[ \left| \sum_{\ell=0}^{k-1} |\mathcal{Y}(t_\ell)|^{2/3} - |\hat{\mathcal{Y}}_k(t_\ell)|^{2/3} \right| \right] \\ &\leq \frac{1}{k} \cdot \sum_{\ell=0}^{k-1} \mathbb{E} \left[ \left| |\mathcal{Y}(t_\ell)|^{2/3} - |\hat{\mathcal{Y}}_k(t_\ell)|^{2/3} \right| \right] \\ &\leq \frac{1}{k} \cdot \sum_{\ell=0}^{k-1} \mathbb{E} \left[ |\mathcal{Y}(t_\ell) - \hat{\mathcal{Y}}_k(t_\ell)|^{2/3} \right] \leq \max_{\ell=0, \dots, k-1} \left( \mathbb{E}[|\mathcal{Y}(t_\ell) - \hat{\mathcal{Y}}_k(t_\ell)|] \right)^{2/3},\end{aligned}$$

we have

$$(9) \quad \lim_{k \rightarrow \infty} \mathbb{E}[|R_k - \hat{R}_k|] = 0$$

due to (6). Combining (8) with (9) we conclude that

$$(10) \quad \lim_{k \rightarrow \infty} \hat{R}_k = \int_0^1 |\mathcal{Y}(t)|^{2/3} dt \quad \text{in probability.}$$

By definition of  $\mathcal{Y}$ , (A2), and (A3) we have

$$\mathbb{P}(\mathcal{Y}(0) \neq 0) = \mathbb{P}(\mathcal{G}(0, X(0)) \neq 0) \geq \mathbb{P}(X(0) \in I) = 1.$$

Observing the continuity of the process  $\mathcal{Y}$  we thus obtain

$$(11) \quad \mathbb{P}\left(\int_0^1 |\mathcal{Y}(t)|^{2/3} dt > 0\right) = 1.$$

Combining (10) with (11) yields

$$(12) \quad \forall \varepsilon \in (0, 1) \exists k_0 \in \mathbb{N}, \beta \in (0, 1) \forall k \geq k_0: \mathbb{P}(\widehat{R}_k \geq \beta) \geq 1 - \varepsilon.$$

Let  $k \in \mathbb{N}$  and consider an approximation  $\widehat{X}_k(1) \in \mathbb{A}_k(\mathbb{R}, W)$ . Observe that there exist random variables  $\tau_0, \dots, \tau_{2k}: \Omega \rightarrow [0, \infty)$  with

(i)  $\tau_\ell$  is measurable with respect to  $\sigma(\mathcal{F}_0, W(\tau_0), \dots, W(\tau_{\ell-1}))$  for all  $\ell = 0, \dots, 2k$ ,

(ii)  $\tau_\ell = t_\ell$  for all  $\ell = 0, \dots, k$ ,

(iii)  $\widehat{X}_k(1)$  is measurable with respect to the  $\sigma$ -algebra  $\mathfrak{A} = \sigma(\mathcal{F}_0, W(\tau_0), \dots, W(\tau_{2k}))$ .

Let  $B = (B(t))_{t \in [0,1]}$  denote the piecewise linear interpolation of  $(W(t))_{t \in [0,1]}$  at the nodes  $\tau_0, \dots, \tau_{2k}$ . We define the process  $Z = (Z(t))_{t \in [0,1]}$  by  $Z(t) = W(t) - B(t)$  and put

$$U = \widehat{X}_k(1) - \widehat{X}_k^{\text{WPt}}(1) - \sum_{\ell=0}^{k-1} \widehat{\mathcal{Y}}_k(t_\ell) \cdot \int_{t_\ell}^{t_{\ell+1}} (B(t) - W(t_\ell)) dt.$$

From (5) we get

$$\overline{X}_k^{\text{aux}}(1) - \widehat{X}_k(1) = \sum_{\ell=0}^{k-1} \widehat{\mathcal{Y}}_k(t_\ell) \cdot \int_{t_\ell}^{t_{\ell+1}} Z(t) dt - U.$$

Conditioned on  $\mathfrak{A}$ , the values of  $\widehat{\mathcal{Y}}_k$ ,  $U$ , and the evaluation sites  $\tau_0, \dots, \tau_{2k}$  are fixed and the process  $Z$  consists of independent Brownian bridges (from 0 to 0) on the subintervals corresponding to  $\tau_0, \dots, \tau_{2k}$ , cf. Yaroslavtseva [50, Lemma 1-2]. Note that

$$\mathbb{E}\left[\int_{t_\ell}^{t_{\ell+1}} Z(t) dt \mid \mathfrak{A}\right] = 0$$

for  $\ell = 0, \dots, k-1$ . We conclude that conditioned on  $\mathfrak{A}$  the random variable  $\overline{X}_k^{\text{aux}}(1) - \widehat{X}_k(1)$  is normally distributed with variance

$$(13) \quad \text{Var}\left[\overline{X}_k^{\text{aux}}(1) - \widehat{X}_k(1) \mid \mathfrak{A}\right] = \mathbb{E}\left[\left|\sum_{\ell=0}^{k-1} \widehat{\mathcal{Y}}_k(t_\ell) \cdot \int_{t_\ell}^{t_{\ell+1}} Z(t) dt\right|^2 \mid \mathfrak{A}\right].$$

For  $\ell \in \{0, \dots, k-1\}$  we define  $d_\ell = \#\{i \in \{0, \dots, 2k\}: \tau_i \in (t_\ell, t_{\ell+1})\}$ . By using Müller-Gronbach [38, Eq. (17), p. 1624] we obtain

$$(14) \quad \begin{aligned} \mathbb{E}\left[\left|\sum_{\ell=0}^{k-1} \widehat{\mathcal{Y}}_k(t_\ell) \cdot \int_{t_\ell}^{t_{\ell+1}} Z(t) dt\right|^2 \mid \mathfrak{A}\right] &= \sum_{\ell=0}^{k-1} \left(\widehat{\mathcal{Y}}_k(t_\ell)\right)^2 \cdot \mathbb{E}\left[\left|\int_{t_\ell}^{t_{\ell+1}} Z(t) dt\right|^2 \mid \mathfrak{A}\right] \\ &\geq \sum_{\ell=0}^{k-1} \left(\widehat{\mathcal{Y}}_k(t_\ell)\right)^2 \cdot \frac{1}{12k^3} \cdot \frac{1}{(d_\ell + 1)^2}. \end{aligned}$$

Clearly,  $\sum_{\ell=0}^{k-1} d_\ell \leq k$ , and therefore, by the Hölder inequality,

$$(k \cdot \widehat{R}_k)^3 = \left(\sum_{\ell=0}^{k-1} (d_\ell + 1)^{2/3} \cdot \frac{|\widehat{\mathcal{Y}}_k(t_\ell)|^{2/3}}{(d_\ell + 1)^{2/3}}\right)^3 \leq 4k^2 \cdot \sum_{\ell=0}^{k-1} (\widehat{\mathcal{Y}}_k(t_\ell)/(d_\ell + 1))^2,$$

which jointly with (13) and (14) implies

$$\text{Var}\left[k \cdot |\overline{X}_k^{\text{aux}}(1) - \widehat{X}_k(1)| \mid \mathfrak{A}\right] \geq \frac{1}{48} \widehat{R}_k^3.$$

Employing Lemma 17 we conclude that

$$\forall \varepsilon \in (0, 1): \mathbb{P}(k \cdot |\bar{X}_k^{\text{aux}}(1) - \hat{X}_k(1)| \geq \varepsilon (\hat{R}_k^3/48)^{1/2} \mid \mathfrak{A}) \geq 1 - \varepsilon$$

and hence

$$(15) \quad \forall \varepsilon \in (0, 1): \mathbb{P}(k \cdot |\bar{X}_k^{\text{aux}}(1) - \hat{X}_k(1)| \geq \varepsilon \hat{R}_k^{3/2}/7) \geq 1 - \varepsilon.$$

Fix  $\varepsilon \in (0, 1)$  and choose  $k_0 \in \mathbb{N}$  and  $\beta \in (0, 1)$  according to (12). By (12) and (15) we obtain for all  $k \geq k_0$  and  $\hat{X}_k(1) \in \mathbb{A}_k(\mathbb{R}, W)$  that

$$(16) \quad \begin{aligned} & \mathbb{P}\left(|\bar{X}_k^{\text{aux}}(1) - \hat{X}_k(1)| \geq \frac{\varepsilon \beta^{3/2}}{7k}\right) \\ & \geq \mathbb{P}(\{k \cdot |\bar{X}_k^{\text{aux}}(1) - \hat{X}_k(1)| \geq \varepsilon \hat{R}_k^{3/2}/7\} \cap \{\hat{R}_k \geq \beta\}) \\ & \geq \mathbb{P}(k \cdot |\bar{X}_k^{\text{aux}}(1) - \hat{X}_k(1)| \geq \varepsilon \hat{R}_k^{3/2}/7) - \mathbb{P}(\hat{R}_k < \beta) \\ & \geq 1 - 2\varepsilon. \end{aligned}$$

Put  $c_3 = \varepsilon \beta^{3/2}/14 \in (0, \infty)$ . Using the Markov inequality we derive from (7) that for all  $k \in \mathbb{N}$  we have

$$(17) \quad \mathbb{P}(|X(1) - \bar{X}_k^{\text{aux}}(1)| \geq c_3/k) \leq k/c_3 \cdot c_2 \cdot k^{-3/2} = c_2/c_3 \cdot k^{-1/2}.$$

We conclude from (16) and (17) that for all  $k \geq \max(k_0, (c_2/(c_3\varepsilon))^2)$  and  $\hat{X}_k(1) \in \mathbb{A}_k(\mathbb{R}, W)$  we have

$$\begin{aligned} & \mathbb{P}(|X(1) - \hat{X}_k(1)| \geq c_3/k) \\ & \geq \mathbb{P}(\{|\bar{X}_k^{\text{aux}}(1) - \hat{X}_k(1)| \geq 2c_3/k\} \cap \{|X(1) - \bar{X}_k^{\text{aux}}(1)| < c_3/k\}) \\ & \geq \mathbb{P}(|\bar{X}_k^{\text{aux}}(1) - \hat{X}_k(1)| \geq 2c_3/k) - \mathbb{P}(|X(1) - \bar{X}_k^{\text{aux}}(1)| \geq c_3/k) \\ & \geq 1 - 2\varepsilon - c_2/c_3 \cdot k^{-1/2} \geq 1 - 3\varepsilon, \end{aligned}$$

which completes the proof.  $\square$

**5.2. Local Assumptions.** The next proposition shows that even under very mild local regularity conditions on the coefficients  $a$  and  $b$ , the probability for an adaptive method  $\hat{X}_n(T) \in \mathbb{A}_n(\mathbb{R}, W)$  of having an error at least of order one is still uniformly bounded away from zero. Its proof exploits a comparison result for SDEs, see Proposition 22 in the appendix, to reduce the general case to the case treated in Proposition 4.

**Proposition 5** (Probability bounds on the pointwise error of adaptive methods II). *Assume the setting in Section 3. Let  $t_0 \in [0, T]$  and let  $\emptyset \neq I \subseteq \mathbb{R}$  be an open interval such that*

(A1\*)  $\forall i \in \{0, 1, 2\}, j \in \{0, 1, 2, 3\}$ : *the partial derivatives  $a^{(i,j)}, b^{(i,j)}$  exist on  $[t_0, T] \times I$  and are continuous,*

(A2\*)  $\forall x \in I$ :  $(a^{(0,1)}b - b^{(1,0)} - ab^{(0,1)} - \frac{1}{2}b^2b^{(0,2)})(t_0, x) \neq 0$ , *and*

$\forall (t, x) \in [t_0, T] \times I$ :  $b(t, x) \neq 0$ ,

(A3\*)  $\mathbb{P}(X(t_0) \in I) > 0$ .

*Then there exist constants  $c, \gamma \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  and for all  $\hat{X}_n(T) \in \mathbb{A}_n(\mathbb{R}, W)$  we have*

$$\mathbb{P}(|X(T) - \hat{X}_n(T)| \geq c/n) \geq \gamma.$$

*Proof.* By considering the SDE (3) starting from time  $t_0$  we may assume  $t_0 = 0$ . According to the openness of  $I$  and (A3\*), there exist bounded open intervals  $I_1, I_2, I_3 \subseteq \mathbb{R}$  such that

$$\emptyset \neq I_3 \subseteq \bar{I}_3 \subseteq I_2 \subseteq \bar{I}_2 \subseteq I_1 \subseteq \bar{I}_1 \subseteq I$$

and  $\mathbb{P}(X(0) \in I_3) > 0$ . Here,  $\bar{I}_1, \bar{I}_2, \bar{I}_3$  denote the closures of the intervals  $I_1, I_2, I_3$ , respectively. Due to the continuity of  $b$  on  $[0, T] \times \bar{I}_1$  and the second condition in (A2\*) we may without loss of generality assume that

$$(18) \quad \inf_{(t,x) \in [0,T] \times \bar{I}_1} b(t,x) > 0.$$

Let  $\eta_1, \eta_2: \mathbb{R} \rightarrow \mathbb{R}$  be infinitely differentiable functions such that  $0 \leq \eta_1, \eta_2 \leq 1$  and

$$\eta_1(x) = \begin{cases} 1, & \text{if } x \in I_2, \\ 0, & \text{if } x \in I_1^c, \end{cases} \quad \eta_2(x) = \begin{cases} 0, & \text{if } x \in I_3, \\ 1, & \text{if } x \in I_2^c. \end{cases}$$

Furthermore, define  $\tilde{a}: [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  and  $\tilde{b}: [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  by

$$\tilde{a}(t, x) = \eta_1(x) \cdot a(t, x), \quad \tilde{b}(t, x) = \eta_1(x) \cdot b(t, x) + \eta_2(x).$$

Due to (A1\*) and (18) it holds that

- (i)  $\forall i \in \{0, 1, 2\}, j \in \{0, 1, 2, 3\}$ : the partial derivatives  $\tilde{a}^{(i,j)}, \tilde{b}^{(i,j)}$  exist on  $[0, T] \times \mathbb{R}$  are continuous and bounded,
- (ii)  $\inf_{(t,x) \in [0,T] \times \mathbb{R}} |\tilde{b}(t, x)| > 0$ ,
- (iii)  $\forall t \in [0, T], x \in I_3$ :  $a(t, x) = \tilde{a}(t, x)$  and  $b(t, x) = \tilde{b}(t, x)$ .

Let  $x_0 \in I_3$  and define the bounded random variable  $\bar{X}(0): \Omega \rightarrow \mathbb{R}$  by

$$\bar{X}(0)(\omega) = \begin{cases} X(0)(\omega), & \text{if } X(0)(\omega) \in I_3, \\ x_0, & \text{otherwise.} \end{cases}$$

Furthermore, let  $\bar{X}: [0, T] \times \Omega \rightarrow \mathbb{R}$  be an  $(\mathcal{F}_t)_{t \in [0, T]}$ -adapted stochastic process with continuous sample paths such that for all  $t \in [0, T]$  it holds  $\mathbb{P}$ -a.s. that

$$\bar{X}(t) = \bar{X}(0) + \int_0^t \tilde{a}(s, \bar{X}(s)) ds + \int_0^t \tilde{b}(s, \bar{X}(s)) dW(s).$$

Applying Proposition 22 with  $I = I_3$  shows

$$\mathbb{P}(\forall t \in [0, T]: \bar{X}(t) = X(t)) > 0$$

and hence

$$(19) \quad \mathbb{P}(\bar{X}(T) = X(T)) > 0.$$

Moreover, Proposition 4 with  $X = \bar{X}$ ,  $a = \tilde{a}$ ,  $b = \tilde{b}$ ,  $I = I_3$ ,  $\varepsilon = \mathbb{P}(\bar{X}(T) = X(T))/2$  yields the existence of a constant  $c \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  and for all  $\hat{X}_n(T) \in \mathbb{A}_n(\mathbb{R}, W)$  we have

$$(20) \quad \mathbb{P}(|\bar{X}(T) - \hat{X}_n(T)| \geq c/n) \geq 1 - \mathbb{P}(\bar{X}(T) = X(T))/2.$$

Combining (19) with (20) shows that for all  $n \in \mathbb{N}$  and for all  $\hat{X}_n(T) \in \mathbb{A}_n(\mathbb{R}, W)$  it holds that

$$\begin{aligned} \mathbb{P}(|X(T) - \hat{X}_n(T)| \geq c/n) &\geq \mathbb{P}(\{|X(T) - \hat{X}_n(T)| \geq c/n\} \cap \{X(T) = \bar{X}(T)\}) \\ &= \mathbb{P}(\{|\bar{X}(T) - \hat{X}_n(T)| \geq c/n\} \cap \{X(T) = \bar{X}(T)\}) \\ &\geq \mathbb{P}(|\bar{X}(T) - \hat{X}_n(T)| \geq c/n) + \mathbb{P}(X(T) = \bar{X}(T)) - 1 \\ &\geq \mathbb{P}(\bar{X}(T) = X(T))/2 > 0, \end{aligned}$$

which completes the proof.  $\square$

Proposition 5 provides a lower error bound for methods from the class  $\mathbb{A}_n(\mathbb{R}, W)$ , i.e., for adaptive methods that are based on  $n$  evaluations of  $W$ . We now extend this result to the class  $\mathcal{A}_n(\mathbb{R}, X(0), W)$  of adaptive methods that may use  $n$  evaluations of  $W$ , on average, which is our main result for pointwise approximation.

**Theorem 6** (Lower error bound for pointwise approximation). *Assume the setting in Section 3. Let  $t_0 \in [0, T)$  and let  $\emptyset \neq I \subseteq \mathbb{R}$  be an open interval such that the conditions (A1\*), (A2\*), and (A3\*) from Proposition 5 are satisfied. Then there exist constants  $\bar{c}, \bar{\gamma} \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  and for all  $\widehat{X}_n(T) \in \mathcal{A}_n(\mathbb{R}, X(0), W)$  we have*

$$\mathbb{P}(|X(T) - \widehat{X}_n(T)| \geq \bar{c}/n) \geq \bar{\gamma}.$$

In particular, for  $\hat{c} = \bar{c} \cdot \bar{\gamma} \in (0, \infty)$  we have for all  $n \in \mathbb{N}$  that

$$\inf_{\widehat{X}_n(T) \in \mathcal{A}_n(\mathbb{R}, X(0), W)} \left\{ \mathbb{E} \left[ |X(T) - \widehat{X}_n(T)| \right] \right\} \geq \hat{c} \cdot n^{-1}.$$

*Proof.* Choose  $c, \gamma \in (0, \infty)$  according to Proposition 5 and choose  $k \in \mathbb{N}$  such that  $1/k \leq \gamma/2$ . Let  $n \in \mathbb{N}$  and  $\widehat{X}_n(T) \in \mathcal{A}_n(\mathbb{R}, X(0), W)$ , and let  $\nu(X(0), W)$  denote the number of evaluations nodes used by  $\widehat{X}_n(T)$ , see Section 4. By assumption we have  $\mathbb{E}[\nu(X(0), W)] \leq n$ . Hence the Markov inequality shows

$$(21) \quad \mathbb{P}(\nu(X(0), W) \geq kn) \leq \frac{\mathbb{E}[\nu(X(0), W)]}{kn} \leq 1/k \leq \gamma/2.$$

Define  $\widehat{X}_{kn}^*(T) \in \mathbb{A}_{kn}(\mathbb{R}, W)$  by

$$\widehat{X}_{kn}^*(T)(\omega) = \begin{cases} \widehat{X}_n(T)(\omega), & \text{if } \nu(X(0), W)(\omega) < kn, \\ 0, & \text{otherwise,} \end{cases}$$

Then

$$\mathbb{P}(\widehat{X}_n(T) = \widehat{X}_{kn}^*(T)) \geq \mathbb{P}(\nu(X(0), W) < kn) \geq 1 - \gamma/2$$

due to (21). Combining the latter fact with Proposition 5 yields

$$\begin{aligned} & \mathbb{P}(|X(T) - \widehat{X}_n(T)| \geq (c/k)/n) \\ & \geq \mathbb{P}(\{|X(T) - \widehat{X}_n(T)| \geq c/(kn)\} \cap \{\widehat{X}_n(T) = \widehat{X}_{kn}^*(T)\}) \\ & = \mathbb{P}(\{|X(T) - \widehat{X}_{kn}^*(T)| \geq c/(kn)\} \cap \{\widehat{X}_n(T) = \widehat{X}_{kn}^*(T)\}) \\ & \geq \mathbb{P}(|X(T) - \widehat{X}_{kn}^*(T)| \geq c/(kn)) + \mathbb{P}(\widehat{X}_n(T) = \widehat{X}_{kn}^*(T)) - 1 \\ & \geq \gamma/2 > 0, \end{aligned}$$

which completes the proof.  $\square$

## 6. LOWER ERROR BOUNDS FOR STRONG APPROXIMATION GLOBALLY IN TIME

In this section we consider strong approximation of  $X$ . In view of Section 4, we study lower error bounds of approximation methods  $\widehat{X}_n$  belonging to the class  $\mathcal{A}_n(C([0, T]), X(0), W)$  or  $\mathcal{A}_n(L_p([0, T]), X(0), W)$ . In both cases we proceed similar to our analysis of one-point approximation in Section 5. We first prove probability bounds on the error of  $\widehat{X}_n$  under restrictive global smoothness assumptions on the coefficients  $a$  and  $b$  of the SDE (3) and then switch to the setting of mild local smoothness conditions by employing the localization technique from Appendix B.

**6.1.  $L_\infty$ -Approximation.** In this section we consider approximation with respect to the maximum distance on the time interval  $[0, T]$ . The following proposition extends mean error bounds from Müller-Gronbach [36] to probability bounds for the error of  $\widehat{X}_n$  from  $\mathbb{A}_n(C([0, T]), W)$ .

**Proposition 7** (Probability bounds on the  $L_\infty$ -error of adaptive methods I). *Assume the setting in Section 3 with  $\mathbb{E}[|X(0)|^2] < \infty$ . Assume further that*

(B1)  $\forall i \in \{0, 1\}, j \in \{0, 1\}$ : *the partial derivatives  $a^{(i,j)}, b^{(i,j)}$  exist on  $[0, T] \times \mathbb{R}$  and are continuous and bounded,*

(B2)  $\inf_{(t,x) \in [0,T] \times \mathbb{R}} |b(t,x)| > 0$ .

*Then for every  $\varepsilon \in (0, 1)$  there exists  $c \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  and for all  $\widehat{X}_n \in \mathbb{A}_n(C([0, T]), W)$  we have*

$$\mathbb{P}(\|X - \widehat{X}_n\|_\infty \geq c \sqrt{\ln(n+1)/n}) \geq 1 - \varepsilon.$$

*Proof.* Due to a scaling in time we may assume  $T = 1$ . For  $k \in \mathbb{N}$  we put  $t_\ell = \ell/k$  for  $\ell = 0, \dots, k$  and we define a continuous-time Euler scheme  $\widehat{X}_k^E = (\widehat{X}_k^E(t))_{t \in [0,1]}$  by  $\widehat{X}_k^E(0) = X(0)$  and for  $\ell = 0, \dots, k-1$  and  $t \in (t_\ell, t_{\ell+1}]$  by

$$\widehat{X}_k^E(t) = \widehat{X}_k^E(t_\ell) + a(t_\ell, \widehat{X}_k^E(t_\ell)) \cdot (t - t_\ell) + b(t_\ell, \widehat{X}_k^E(t_\ell)) \cdot (W(t) - W(t_\ell)).$$

Moreover, for  $\ell = 0, \dots, k$  we put

$$\widehat{b}_{k,\ell} = b(t_\ell, \widehat{X}_k^E(t_\ell)).$$

By (B1) and  $\mathbb{E}[|X(0)|^2] < \infty$  we have

$$(22) \quad \exists c_1 \in (0, \infty) \quad \forall k \in \mathbb{N}: \mathbb{E}[\|X - \widehat{X}_k^E\|_\infty] \leq c_1/\sqrt{k},$$

see, e.g., Hofmann, Müller-Gronbach, and Ritter [23, Theorem 3, page 631].

Let  $\varepsilon \in (0, 1)$  and

$$\delta = \inf_{(t,x) \in [0,1] \times \mathbb{R}} |b(t,x)| \in (0, \infty),$$

see (B2), and choose  $c_2 \in (0, \infty)$  according to Lemma 18.

Let  $k \in 2\mathbb{N}$  and consider an approximation  $\widehat{X}_k \in \mathbb{A}_{k/2}(C([0, 1]), W)$ . Observe that there exist random variables  $\tau_0, \dots, \tau_{3k/2}: \Omega \rightarrow [0, \infty)$  with

- (i)  $\tau_\ell$  is measurable with respect to  $\sigma(\mathcal{F}_0, W(\tau_0), \dots, W(\tau_{\ell-1}))$  for all  $\ell = 0, \dots, 3k/2$ ,
- (ii)  $\tau_\ell = t_\ell$  for all  $\ell = 0, \dots, k$ ,
- (iii)  $\widehat{X}_k$  is measurable with respect to the  $\sigma$ -algebra  $\mathfrak{A} = \sigma(\mathcal{F}_0, W(\tau_0), \dots, W(\tau_{3k/2}))$ .

Let  $B = (B(t))_{t \in [0,1]}$  denote the piecewise linear interpolation of  $(W(t))_{t \in [0,1]}$  at the nodes  $\tau_0, \dots, \tau_{3k/2}$ . We define the process  $Z = (Z(t))_{t \in [0,1]}$  by  $Z(t) = W(t) - B(t)$ . Moreover, we define the process  $U = (U(t))_{t \in [0,1]}$  by

$$U(t) = \widehat{X}_k(t) - \left[ \mathbf{1}_{\{0\}}(t) \cdot X(0) + \sum_{\ell=0}^{k-1} \mathbf{1}_{(t_\ell, t_{\ell+1})}(t) \cdot \left( \widehat{X}_k^E(t_\ell) + a(t_\ell, \widehat{X}_k^E(t_\ell)) \cdot (t - t_\ell) + \widehat{b}_{k,\ell} \cdot (B(t) - W(t_\ell)) \right) \right].$$

By definition of  $\widehat{X}_k^E$  it holds for all  $t \in [0, 1]$  that

$$\widehat{X}_k^E(t) - \widehat{X}_k(t) = \sum_{\ell=0}^{k-1} \left[ \mathbf{1}_{(t_\ell, t_{\ell+1})}(t) \cdot \widehat{b}_{k,\ell} \cdot Z(t) \right] - U(t).$$

Let

$$\mathcal{L}_k = \{\ell \in \{0, \dots, k-1\} : (t_\ell, t_{\ell+1}) \cap \{\tau_0, \dots, \tau_{3k/2}\} = \emptyset\}$$

and observe that

$$(23) \quad \#\mathcal{L}_k \geq k/2.$$

Conditioned on  $\mathfrak{A}$ , the values of  $(\widehat{X}_k^E(t_\ell))_{\ell \in \{0, \dots, k\}}$ ,  $U$ ,  $(\widehat{b}_{k,\ell})_{\ell \in \{0, \dots, k\}}$ , and the evaluation sites  $\tau_0, \dots, \tau_{3k/2}$  used by  $\widehat{X}_k$  are fixed and the process  $Z$  consists of independent Brownian bridges (from 0 to 0) on the subintervals corresponding to  $\tau_0, \dots, \tau_{3k/2}$ , cf. Yaroslavtseva [50, Lemma 1-2]. We conclude that conditioned on  $\mathfrak{A}$  the set  $\mathcal{L}_k$  is fixed and the random variables

$$V_\ell = (\widehat{X}_k^E - \widehat{X}_k)((t_\ell + t_{\ell+1})/2),$$

where  $\ell \in \{0, \dots, k-1\}$ , are independent and normally distributed with variances satisfying

$$(24) \quad \text{Var}(V_\ell | \mathfrak{A}) \geq \mathbf{1}_{\mathcal{L}_k}(\ell) \cdot \frac{\delta^2}{4k}.$$

Since  $\|\widehat{X}_k^E - \widehat{X}_k\|_\infty \geq \max_{\ell \in \mathcal{L}_k} |V_\ell|$ , we get

$$(25) \quad \mathbb{P}(2\sqrt{k} \cdot \|\widehat{X}_k^E - \widehat{X}_k\|_\infty \geq c_2 \sqrt{\ln(k/2)} | \mathfrak{A}) \geq \mathbb{P}(2\sqrt{k} \cdot \max_{\ell \in \mathcal{L}_k} |V_\ell| \geq c_2 \sqrt{\ln(k/2)} | \mathfrak{A}).$$

Lemma 18 and (23) imply

$$(26) \quad \mathbb{P}(2\sqrt{k} \cdot \max_{\ell \in \mathcal{L}_k} |V_\ell| \geq c_2 \sqrt{\ln(k/2)} | \mathfrak{A}) \geq 1 - \varepsilon.$$

Combing (25) and (26) yields

$$\mathbb{P}(2\sqrt{k} \cdot \|\widehat{X}_k^E - \widehat{X}_k\|_\infty \geq c_2 \sqrt{\ln(k/2)} | \mathfrak{A}) \geq 1 - \varepsilon$$

and hence

$$(27) \quad \mathbb{P}(2\sqrt{k} \cdot \|\widehat{X}_k^E - \widehat{X}_k\|_\infty \geq c_2 \sqrt{\ln(k/2)}) \geq 1 - \varepsilon.$$

By (27), there exists a constant  $c_3 \in (0, \infty)$  such that for all  $k \in \mathbb{N}$  and  $\widehat{X}_k \in \mathbb{A}_k(C([0, 1]), W)$  it holds that

$$\mathbb{P}(\|\widehat{X}_k^E - \widehat{X}_k\|_\infty \geq c_3 \sqrt{\ln(k+1)/k}) \geq 1 - \varepsilon.$$

The latter fact and the Markov inequality combined with (22) imply that for all  $k \in \mathbb{N}$  and  $\widehat{X}_k \in \mathbb{A}_k(C([0, 1]), W)$  it holds that

$$\begin{aligned} & \mathbb{P}(\|X - \widehat{X}_k\|_\infty \geq c_3/2 \cdot \sqrt{\ln(k+1)/k}) \\ & \geq \mathbb{P}(\{\|\widehat{X}_k^E - \widehat{X}_k\|_\infty \geq c_3 \sqrt{\ln(k+1)/k}\} \cap \{\|X - \widehat{X}_k^E\|_\infty < c_3/2 \cdot \sqrt{\ln(k+1)/k}\}) \\ & \geq \mathbb{P}(\|\widehat{X}_k^E - \widehat{X}_k\|_\infty \geq c_3 \sqrt{\ln(k+1)/k}) - \mathbb{P}(\|X - \widehat{X}_k^E\|_\infty \geq c_3/2 \cdot \sqrt{\ln(k+1)/k}) \\ & \geq 1 - \varepsilon - 2/c_3 \cdot \sqrt{k/\ln(k+1)} \cdot c_1/\sqrt{k} \\ & = 1 - \varepsilon - 2c_1/(c_3 \sqrt{\ln(k+1)}), \end{aligned}$$

which completes the proof.  $\square$

Combining the localization technique from Appendix B with Proposition 7 leads to the following result.

**Proposition 8** (Probability bounds on the  $L_\infty$ -error of adaptive methods II). *Assume the setting in Section 3. Let  $t_0 \in [0, T)$ ,  $T_0 \in (t_0, T]$  and let  $\emptyset \neq I \subseteq \mathbb{R}$  be an open interval such that*

(B1\*)  $\forall i \in \{0, 1\}, j \in \{0, 1\}$ : the partial derivatives  $a^{(i,j)}, b^{(i,j)}$  exist on  $[t_0, T_0] \times I$  and are continuous,



$$(B2^*) \quad \forall (t, x) \in [t_0, T_0] \times I: b(t, x) \neq 0,$$

$$(B3^*) \quad \mathbb{P}(X(t_0) \in I) > 0.$$

Then there exist constants  $c, \gamma \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  and for all  $\widehat{X}_n \in \mathbb{A}_n(C([0, T]), W)$  we have

$$\mathbb{P}(\|X - \widehat{X}_n\|_\infty \geq c \sqrt{\ln(n+1)/n}) \geq \gamma.$$

*Proof.* Since the error is measured globally in time, we may assume  $T_0 = T$ . Now, the reasoning is almost identical to the reasoning in the proof of Proposition 5. Instead of Proposition 4 we rely on Proposition 7.  $\square$

Proposition 8 provides a lower error bound for methods from the class  $\mathbb{A}_n(C([0, T]), W)$ , i.e., for adaptive methods that are based on  $n$  evaluations of  $W$ . We now extend this result to the class  $\mathcal{A}_n(C([0, T]), X(0), W)$  of adaptive methods that may use  $n$  evaluations of  $W$ , on average. The following theorem is our main result for  $L_\infty$ -Approximation.

**Theorem 9** (Lower error bound for  $L_\infty$ -approximation). *Assume the setting in Section 3. Let  $t_0 \in [0, T]$ ,  $T_0 \in (t_0, T]$  and let  $\emptyset \neq I \subseteq \mathbb{R}$  be an open interval such that the conditions  $(B1^*)$ ,  $(B2^*)$ , and  $(B3^*)$  from Proposition 8 are satisfied. Then there exist constants  $\widehat{c}, \widehat{\gamma} \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  and for all  $\widehat{X}_n \in \mathcal{A}_n(C([0, T]), X(0), W)$  we have*

$$\mathbb{P}(\|X - \widehat{X}_n\|_\infty \geq \widehat{c} \sqrt{\ln(n+1)/n}) \geq \widehat{\gamma}.$$

In particular, for  $\widehat{c} = \widehat{c} \cdot \widehat{\gamma} \in (0, \infty)$  we have for all  $n \in \mathbb{N}$  that

$$\inf_{\widehat{X}_n \in \mathcal{A}_n(C([0, T]), X(0), W)} \left\{ \mathbb{E}[\|X - \widehat{X}_n\|_\infty] \right\} \geq \widehat{c} \cdot \sqrt{\ln(n+1)/n}.$$

*Proof.* The proof of Theorem 9 is almost identical to the proof of Theorem 6. Instead of Proposition 5 we rely on Proposition 8.  $\square$

**6.2.  $L_p$ -Approximation.** In this section we consider approximation with respect to the  $L_p$ -distance on the time interval  $[0, T]$  for  $p \in [1, \infty)$ . The following proposition extends mean error bounds from Hofmann, Müller-Gronbach, and Ritter [22], Müller-Gronbach [37] to probability bounds for the error of  $\widehat{X}_n$  from  $\mathbb{A}_n(L_p([0, T]), W)$ .

**Proposition 10** (Probability bounds on the  $L_p$ -error of adaptive methods I). *Assume the setting in Section 3 with  $\mathbb{E}[|X(0)|^4] < \infty$ . Assume further that*

$$(C1) \quad \forall i \in \{0, 1\}, j \in \{0, 1, 2\}: \text{the partial derivatives } a^{(i,j)}, b^{(i,j)} \text{ exist on } [0, T] \times \mathbb{R} \text{ and are continuous and bounded,}$$

$$(C2) \quad \inf_{(t,x) \in [0, T] \times \mathbb{R}} |b(t, x)| > 0.$$

Then for every  $\varepsilon \in (0, 1)$  there exists  $c \in (0, \infty)$  such that for all  $p \in [1, \infty)$ , for all  $n \in \mathbb{N}$ , and for all  $\widehat{X}_n \in \mathbb{A}_n(L_p([0, T]), W)$  we have

$$\mathbb{P}(\|X - \widehat{X}_n\|_p \geq c/\sqrt{n}) \geq 1 - \varepsilon.$$

*Proof.* Due to monotonicity in  $p$  we may assume  $p = 1$ . Moreover, due to a scaling in time we may assume  $T = 1$ .

For  $k \in \mathbb{N}$  we put  $t_\ell = \ell/k$  for  $\ell = 0, \dots, k$  and we define a continuous-time Milstein scheme  $\widehat{X}_k^M = (\widehat{X}_k^M(t))_{t \in [0, 1]}$  by  $\widehat{X}_k^M(0) = X(0)$  and for  $\ell = 0, \dots, k-1$  and  $t \in (t_\ell, t_{\ell+1}]$  by

$$\begin{aligned} \widehat{X}_k^M(t) = & \widehat{X}_k^M(t_\ell) + a(t_\ell, \widehat{X}_k^M(t_\ell)) \cdot (t - t_\ell) + b(t_\ell, \widehat{X}_k^M(t_\ell)) \cdot (W(t) - W(t_\ell)) \\ & + 1/2 \cdot (bb^{(0,1)})(t_\ell, \widehat{X}_k^M(t_\ell)) \cdot ((W(t) - W(t_\ell))^2 - (t - t_\ell)). \end{aligned}$$

By (C1) and  $\mathbb{E}[|X(0)|^4] < \infty$  we have

$$(28) \quad \exists c_1 \in (0, \infty) \quad \forall k \in \mathbb{N}: \quad \sup_{t \in [0,1]} \mathbb{E}[|X(t) - \widehat{X}_k^M(t)|] \leq c_1 \cdot k^{-1},$$

see, e.g., Hofmann, Müller-Gronbach, and Ritter [22, Theorem 4]. For  $k \in \mathbb{N}$  and  $\ell = 0, \dots, k-1$  we define the process  $U_k^{(\ell)} = (U_k^{(\ell)}(t))_{t \in [t_\ell, t_{\ell+1}]}$  by

$$U_k^{(\ell)}(t) = \widehat{X}_k^M(t) - 1/2 \cdot (bb^{(0,1)})(t_\ell, \widehat{X}_k^M(t_\ell)) \cdot ((W(t) - W(t_\ell))^2 - (t - t_\ell)).$$

Furthermore, for  $k \in \mathbb{N}$  we define an auxiliary scheme  $\overline{X}_k^M = (\overline{X}_k^M(t))_{t \in [0,1]}$  by  $\overline{X}_k^M(0) = X(0)$  and for  $\ell = 0, \dots, k-1$  and  $t \in (t_\ell, t_{\ell+1}]$  by

$$\overline{X}_k^M(t) = U_k^{(\ell)}(t).$$

By (C1) we have

$$c_2 = \sup_{(t,x) \in [0,1] \times \mathbb{R}} |bb^{(0,1)}(t,x)| \in [0, \infty).$$

Hence we get from (28) for  $c_3 = c_1 + c_2 \in (0, \infty)$  and all  $k \in \mathbb{N}$  that

$$(29) \quad \begin{aligned} & \mathbb{E}[\|X - \overline{X}_k^M\|_1] \\ & \leq \sup_{t \in [0,1]} \mathbb{E}[|X(t) - \overline{X}_k^M(t)|] = \sup_{\substack{\ell \in \{0, \dots, k-1\} \\ t \in (t_\ell, t_{\ell+1}]}} \mathbb{E}[|X(t) - U_k^{(\ell)}(t)|] \\ & \leq \sup_{\substack{\ell \in \{0, \dots, k-1\} \\ t \in (t_\ell, t_{\ell+1}]}} \mathbb{E}[|X(t) - \widehat{X}_k^M(t)|] \\ & \quad + \sup_{\substack{\ell \in \{0, \dots, k-1\} \\ t \in (t_\ell, t_{\ell+1}]}} \mathbb{E}\left[|1/2 \cdot (bb^{(0,1)})(t_\ell, \widehat{X}_k^M(t_\ell)) \cdot ((W(t) - W(t_\ell))^2 - (t - t_\ell))|\right] \\ & \leq c_1 \cdot k^{-1} + \frac{c_2}{2} \cdot 2 \cdot k^{-1} = c_3 \cdot k^{-1}. \end{aligned}$$

Let  $\varepsilon \in (0, 1)$  and

$$\delta = \inf_{(t,x) \in [0,1] \times \mathbb{R}} |b(t,x)| \in (0, \infty),$$

see (C2), and choose  $c_4 \in (0, \infty)$  according to Lemma 19.

Let  $k \in 2\mathbb{N}$  and consider an approximation  $\widehat{X}_k \in \mathbb{A}_{k/2}(L_1([0,1]), W)$ . Observe that there exist random variables  $\tau_0, \dots, \tau_{3k/2}: \Omega \rightarrow [0, \infty)$  with

- (i)  $\tau_\ell$  is measurable with respect to  $\sigma(\mathcal{F}_0, W(\tau_0), \dots, W(\tau_{\ell-1}))$  for all  $\ell = 0, \dots, 3k/2$ ,
- (ii)  $\tau_\ell = t_\ell$  for all  $\ell = 0, \dots, k$ ,
- (iii)  $\widehat{X}_k$  is measurable with respect to the  $\sigma$ -algebra  $\mathfrak{A} = \sigma(\mathcal{F}_0, W(\tau_0), \dots, W(\tau_{3k/2}))$ .

For  $\ell \in \{0, \dots, k-1\}$  define  $d_\ell = \#\{i \in \{0, \dots, 3k/2\}: \tau_i \in (t_\ell, t_{\ell+1})\}$ . Let  $\ell_1, \dots, \ell_{k/2} \in \{0, \dots, k-1\}$  such that  $\ell_1 < \dots < \ell_{k/2}$ ,  $d_{\ell_i} = 0$  for all  $i = 1, \dots, k/2$ , and  $d_\ell > 0$  for all

$\ell \in \{0, \dots, \ell_{k/2}\} \setminus \{\ell_1, \dots, \ell_{k/2}\}$ . We then have

$$\begin{aligned} \|\widehat{X}_k - \overline{X}_k^M\|_1 &\geq \sum_{i=1}^{k/2} \int_{t_{\ell_i}}^{t_{\ell_i+1}} |\widehat{X}_k(t) - \overline{X}_k^M(t)| dt \\ &= \sum_{i=1}^{k/2} \int_{t_{\ell_i}}^{t_{\ell_i+1}} |\widehat{X}_k(t) - U_k^{(\ell_i)}(t)| dt \\ &= \sum_{i=1}^{k/2} \int_0^{1/k} |\widehat{X}_k(t + t_{\ell_i}) - U_k^{(\ell_i)}(t + t_{\ell_i})| dt \end{aligned}$$

and thus

$$\begin{aligned} (30) \quad &\mathbb{P}(\|\widehat{X}_k - \overline{X}_k^M\|_1 \geq c_4/\sqrt{k/2} \mid \mathfrak{A}) \\ &\geq \mathbb{P}\left(\sum_{i=1}^{k/2} \int_0^{1/k} |\widehat{X}_k(t + t_{\ell_i}) - U_k^{(\ell_i)}(t + t_{\ell_i})| dt \geq c_4/\sqrt{k/2} \mid \mathfrak{A}\right). \end{aligned}$$

Conditioned on  $\mathfrak{A}$ , the values of  $(\widehat{X}_k^M(t_{\ell}))_{\ell \in \{0, \dots, k\}}$ ,  $W(t_0), \dots, W(t_k)$ ,  $\ell_1, \dots, \ell_{k/2}$ , and  $\widehat{X}_k$  are fixed and the processes  $(B_1(t))_{t \in [0, 1/k]}, \dots, (B_{k/2}(t))_{t \in [0, 1/k]}$ , given by

$$B_i(t) = W(t + t_{\ell_i}) - ((1 - k \cdot t) \cdot W(t_{\ell_i}) + k \cdot t \cdot W(t_{\ell_i+1})),$$

are independent Brownian bridges (from 0 to 0), cf. Yaroslavtseva [50, Lemmas 1,2]. Furthermore, for  $i \in \{1, \dots, k/2\}$  and  $t \in [0, 1/k]$  it holds that

$$\begin{aligned} U_k^{(\ell_i)}(t + t_{\ell_i}) - \widehat{X}_k(t + t_{\ell_i}) &= b(t_{\ell_i}, \widehat{X}_k^M(t_{\ell_i})) \cdot B_i(t) + \left[ \widehat{X}_k^M(t_{\ell_i}) + a(t_{\ell_i}, \widehat{X}_k^M(t_{\ell_i})) \cdot t \right. \\ &\quad \left. + b(t_{\ell_i}, \widehat{X}_k^M(t_{\ell_i})) \cdot ((-k \cdot t) \cdot W(t_{\ell_i}) + k \cdot t \cdot W(t_{\ell_i+1})) - \widehat{X}_k(t + t_{\ell_i}) \right]. \end{aligned}$$

Hence we get from Lemma 19 that

$$\mathbb{P}\left(\sum_{i=1}^{k/2} \int_0^{1/k} |\widehat{X}_k(t + t_{\ell_i}) - U_k^{(\ell_i)}(t + t_{\ell_i})| dt \geq c_4/\sqrt{k/2} \mid \mathfrak{A}\right) \geq 1 - \varepsilon.$$

Combining this with (30) yields

$$(31) \quad \mathbb{P}(\|\widehat{X}_k - \overline{X}_k^M\|_1 \geq c_4/\sqrt{k/2}) \geq 1 - \varepsilon.$$

Put  $c_5 = c_4/\sqrt{2} \in (0, \infty)$ . Using the Markov inequality we derive from (29) that for all  $k \in \mathbb{N}$  we have

$$(32) \quad \mathbb{P}(\|X - \overline{X}_k^M\|_1 \geq c_5/\sqrt{k}) \leq \sqrt{k}/c_5 \cdot c_3 \cdot k^{-1} = c_3/c_5 \cdot k^{-1/2}.$$

From (31) and (32) we get that for all  $k \in 2\mathbb{N} \cap [(c_3/(c_5\varepsilon))^2, \infty)$  and  $\widehat{X}_k \in \mathbb{A}_{k/2}(L_1([0, 1]), W)$  it holds that

$$\begin{aligned} \mathbb{P}(\|X - \widehat{X}_k\|_1 \geq c_5/\sqrt{k}) &\geq \mathbb{P}(\{\|\overline{X}_k^M - \widehat{X}_k\|_1 \geq 2c_5/\sqrt{k}\} \cap \{\|X - \overline{X}_k^M\|_1 < c_5/\sqrt{k}\}) \\ &\geq \mathbb{P}(\|\overline{X}_k^M - \widehat{X}_k\|_1 \geq 2c_5/\sqrt{k}) - \mathbb{P}(\|X - \overline{X}_k^M\|_1 \geq c_5/\sqrt{k}) \\ &\geq 1 - \varepsilon - c_3/c_5 \cdot k^{-1/2} \geq 1 - 2\varepsilon, \end{aligned}$$

which completes the proof.  $\square$

Combining the localization technique from Appendix B with Proposition 10 leads to the following result.

**Proposition 11** (Probability bounds on the  $L_p$ -error of adaptive methods II). *Assume the setting in Section 3. Let  $t_0 \in [0, T)$ ,  $T_0 \in (t_0, T]$  and let  $\emptyset \neq I \subseteq \mathbb{R}$  be an open interval such that*

- (C1\*)  $\forall i \in \{0, 1\}, j \in \{0, 1, 2\}$ : the partial derivatives  $a^{(i,j)}, b^{(i,j)}$  exist on  $[t_0, T_0] \times I$  and are continuous,
- (C2\*)  $\forall (t, x) \in [t_0, T_0] \times I$ :  $b(t, x) \neq 0$ ,
- (C3\*)  $\mathbb{P}(X(t_0) \in I) > 0$ .

Then there exist constants  $c, \gamma \in (0, \infty)$  such that for all  $p \in [1, \infty)$ , for all  $n \in \mathbb{N}$ , and for all  $\widehat{X}_n \in \mathbb{A}_n(L_p([0, T]), W)$  we have

$$\mathbb{P}(\|X - \widehat{X}_n\|_p \geq c/\sqrt{n}) \geq \gamma.$$

*Proof.* Since the error is measured globally in time, we may assume  $T_0 = T$ . Now, the reasoning is almost identical to the reasoning in the proof of Proposition 5. Instead of Proposition 4 we rely on Proposition 10.  $\square$

Proposition 11 provides a lower error bound for methods from the class  $\mathbb{A}_n(L_p([0, T]), W)$ , i.e., for adaptive methods that are based on  $n$  evaluations of  $W$ . We now extend this result to the class  $\mathcal{A}_n(L_p([0, T]), X(0), W)$  of adaptive methods that may use  $n$  evaluations of  $W$ , on average. The following theorem is our main result for  $L_p$ -Approximation with  $p \in [1, \infty)$ .

**Theorem 12** (Lower error bound for  $L_p$ -approximation). *Assume the setting in Section 3. Let  $t_0 \in [0, T)$ ,  $T_0 \in (t_0, T]$  and let  $\emptyset \neq I \subseteq \mathbb{R}$  be an open interval such that the conditions (C1\*), (C2\*), and (C3\*) from Proposition 11 are satisfied. Then there exist constants  $\bar{c}, \bar{\gamma} \in (0, \infty)$  such that for all  $p \in [1, \infty)$ , for all  $n \in \mathbb{N}$ , and for all  $\widehat{X}_n \in \mathcal{A}_n(L_p([0, T]), X(0), W)$  we have*

$$\mathbb{P}(\|X - \widehat{X}_n\|_p \geq \bar{c}/\sqrt{n}) \geq \bar{\gamma}.$$

In particular, for  $\hat{c} = \bar{c} \cdot \bar{\gamma} \in (0, \infty)$  we have for all  $p \in [1, \infty)$  and for all  $n \in \mathbb{N}$  that

$$\inf_{\widehat{X}_n \in \mathcal{A}_n(L_p([0, T]), X(0), W)} \left\{ \mathbb{E}[\|X - \widehat{X}_n\|_p] \right\} \geq \hat{c} \cdot n^{-1/2}.$$

*Proof.* The proof of Theorem 12 is almost identical to the proof of Theorem 6. Instead of Proposition 5 we rely on Proposition 11.  $\square$

**6.3. Maximum Pointwise Error.** The following theorem is a consequence of Theorem 12 and Fubini's theorem.

**Theorem 13** (Lower bound for the maximum pointwise error). *Assume the setting in Section 3. Let  $t_0 \in [0, T)$ ,  $T_0 \in (t_0, T]$  and let  $\emptyset \neq I \subseteq \mathbb{R}$  be an open interval such that the conditions (C1\*), (C2\*), and (C3\*) from Proposition 11 are satisfied. Then there exists a constant  $\hat{c} \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  we have*

$$\inf_{\widehat{X}_n \in \mathcal{A}_n(C([0, T]), X(0), W)} \left\{ \sup_{t \in [0, T]} \mathbb{E}[|X(t) - \widehat{X}_n(t)|] \right\} \geq \hat{c} \cdot n^{-1/2}.$$

*Proof.* Let  $n \in \mathbb{N}$  and consider an approximation  $\widehat{X}_n \in \mathcal{A}_n(C([0, T]), X(0), W)$ . Fubini's theorem shows

$$\mathbb{E}[\|X - \widehat{X}_n\|_1] = \int_0^T \mathbb{E}[|X(t) - \widehat{X}_n(t)|] dt \leq T \cdot \sup_{t \in [0, T]} \mathbb{E}[|X(t) - \widehat{X}_n(t)|].$$

Applying Theorem 12 for  $p = 1$  completes the proof.  $\square$

## 7. EXAMPLES

We study Cox-Ingersoll-Ross processes in Section 7.1, equations with superlinearly growing coefficients in Section 7.2, and equations with discontinuous coefficients in Section 7.3. Throughout this section we assume the setting in Section 3.

**7.1. Cox-Ingersoll-Ross Processes.** Cox-Ingersoll-Ross processes are widely used in mathematical finance, e.g., in the Cox-Ingersoll-Ross model for short term interest rates or as the instantaneous variance in the Heston model. Such processes are described by the SDE

$$(33) \quad dX(t) = (\delta - \beta X(t)) dt + \sigma \sqrt{|X(t)|} dW(t), \quad X(0) = x_0,$$

with initial value  $x_0 \in (0, \infty)$  and parameters  $\delta, \sigma \in (0, \infty)$ ,  $\beta \in [0, \infty)$ . In this case the coefficients  $a, b$  are given by

$$a(t, x) = a(x) = \delta - \beta \cdot x \quad \text{and} \quad b(t, x) = b(x) = \sigma \cdot \sqrt{|x|}$$

for every  $(t, x) \in [0, T] \times \mathbb{R}$ . It is well-known that strong existence and pathwise uniqueness hold for the SDE (33). Furthermore, one has

$$\mathbb{P}(\forall t \in [0, \infty): X(t) \geq 0) = 1.$$

Due to a simple scaling we may restrict ourselves to the case

$$\sigma = 2,$$

which we assume in the following.

In the context of strong approximation, it turns out that  $\delta$  is the crucial parameter with respect to the rate of convergence, see, e.g., Corollary 15(iv) below.

**7.1.1. Strong Approximation at a Single Point.** Note that the coefficients  $a, b$  of the SDE (33) are infinitely differentiable on  $(0, \infty)$ . For  $x \in (0, \infty)$  we obtain

$$(a'b - ab' - \frac{1}{2}b^2b'')(x) = -\beta\sqrt{x} + \frac{1-\delta}{\sqrt{x}}.$$

In particular, it holds that

$$(34) \quad \begin{aligned} \exists \emptyset \neq I \subseteq (0, \infty) \text{ open interval } \forall x \in I: (a'b - ab' - \frac{1}{2}b^2b'')(x) \neq 0 \\ \Leftrightarrow \\ \delta \neq 1 \text{ or } \beta \neq 0. \end{aligned}$$

The marginal distributions of the SDE (33) are known explicitly in terms of Lebesgue-densities, see, e.g., Göing-Jaeschke and Yor [14, Equation (4) and Section A.2]. The latter results immediately yield

$$(35) \quad \forall \emptyset \neq I \subseteq (0, \infty) \text{ open interval } \forall t \in (0, \infty): \mathbb{P}(X(t) \in I) > 0.$$

In case of  $\delta \neq 1$  or  $\beta \neq 0$  we get from (34) and (35) that the assumptions of Theorem 6 are fulfilled for every  $t_0 \in (0, T)$ , which yields the lower bound stated in part (i) of the following corollary. Combining this lower bound with the upper bound of Alfonsi [1, Theorem 2] yields a sharp result for  $\delta > 4$ , see part (ii).

**Corollary 14** (CIR processes, pointwise approximation).

(i) If  $\delta \neq 1$  or  $\beta \neq 0$ , then there exists a constant  $c \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  we have

$$\inf_{\widehat{X}_n(T) \in \mathcal{A}_n(\mathbb{R}, X(0), W)} \left\{ \mathbb{E} \left[ |X(T) - \widehat{X}_n(T)| \right] \right\} \geq c \cdot n^{-1}.$$

(ii) If  $\delta > 4$ , then there exist constants  $c, C \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  we have

$$\begin{aligned} c \cdot n^{-1} &\leq \inf_{\widehat{X}_n(T) \in \mathcal{A}_n(\mathbb{R}, X(0), W)} \left\{ \mathbb{E} \left[ |X(T) - \widehat{X}_n(T)| \right] \right\} \\ &\leq \inf_{\widehat{X}_n(T) \in \mathcal{A}_n^{\text{eq}}(\mathbb{R}, X(0), W)} \left\{ \mathbb{E} [|X(T) - \widehat{X}_n(T)|] \right\} \leq C \cdot n^{-1}. \end{aligned}$$

Let us stress that Corollary 14(ii) shows that, in case of  $\delta > 4$ , adaptive algorithms are not superior (up to multiplicative constants) to algorithms that are based on fixed equidistant grids, similar to the case of SDEs with coefficients that satisfy global Lipschitz assumptions. Moreover, we recover the usual optimal convergence rate of 1.

If  $\delta < 1$  then combining the upper bound of Hefter and Herzwurm [18, Theorem 2] with the lower bound of Hefter and Jentzen [19, Theorem 1] shows that there exist constants  $c, C \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  it holds that

$$(36) \quad c \cdot n^{-\delta/2} \leq \inf_{\widehat{X}_n(T) \in \mathcal{A}_n^{\text{eq}}(\mathbb{R}, X(0), W)} \left\{ \mathbb{E} [|X(T) - \widehat{X}_n(T)|] \right\} \leq C \cdot n^{-\delta/2}.$$

Hence, for this range of values of  $\delta$  the lower bound in Corollary 14(i) cannot be attained by algorithms that are based on equidistant grids and it is unclear, up to now, whether the bound in Corollary 14(i) is sharp, i.e., whether adaptive algorithms can achieve the convergence rate  $c/n$  in terms of the average number  $n$  of evaluations of the driving Brownian motion that are used. Note, however, that the result (37) on the power of adaptive methods in the case  $\delta = 1$  and  $\beta = 0$  provides a positive indication in that sense.

Combining Corollary 14(ii) with (36) shows that the optimal convergence rate for algorithms that are based on equidistant grids is exactly  $\min(\delta/2, 1)$  if  $\delta \in (0, 1) \cup (4, \infty)$ . For  $\delta \in (1, 4)$ , no matching upper and lower bounds are known, see Hefter and Herzwurm [17, Figure 1]. Nevertheless, we expect  $\min(\delta/2, 1)$  to be the best possible convergence rate for algorithms that are based on equidistant grids for all  $\delta \in (0, \infty)$ .

Finally, we consider the case of

$$\delta = 1 \quad \text{and} \quad \beta = 0,$$

i.e., the solution of the SDE (33) is a one-dimensional squared Bessel process. In this case none of the results of Section 5 is applicable, see (34). Indeed, in Hefter and Herzwurm [17, Theorem 4] it is shown that adaptive methods are able to achieve any polynomial convergence order and hence are not restricted to rate 1. See also Calvin, Hefter, and Herzwurm [8]. More precisely, Hefter and Herzwurm [17, Theorem 4] shows that for all  $r \in [1, \infty)$  there exists a constant  $C_r \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  it holds that

$$(37) \quad \inf_{\widehat{X}_n(T) \in \mathcal{A}_n(\mathbb{R}, X(0), W)} \left\{ \mathbb{E} [|X(T) - \widehat{X}_n(T)|] \right\} \leq C_r \cdot n^{-r}.$$

On the other hand, Hefter and Herzwurm [17, Corollary 1] shows that the best algorithm based on equidistant nodes converges at rate  $1/2$ , i.e., there exist constants  $c, C \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  we have

$$c \cdot n^{-1/2} \leq \inf_{\widehat{X}_n(T) \in \mathcal{A}_n^{\text{eq}}(\mathbb{R}, X(0), W)} \left\{ \mathbb{E} [|X(T) - \widehat{X}_n(T)|] \right\} \leq C \cdot n^{-1/2}.$$

*7.1.2. Strong Approximation Globally in Time.* We now turn to lower error bounds that are based on Theorem 9 and Theorem 12. Moreover, these lower bounds are complemented by upper error bounds from the literature.

**Corollary 15** (CIR processes, global approximation).

(i) For all  $\delta \in (0, \infty)$  and  $\beta \in [0, \infty)$  there exists a constant  $c \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  we have

$$\inf_{\widehat{X}_n \in \mathcal{A}_n(C([0, T]), X(0), W)} \left\{ \mathbb{E}[\|X - \widehat{X}_n\|_\infty] \right\} \geq c \cdot \sqrt{\ln(n+1)/n}$$

and

$$\inf_{\widehat{X}_n \in \mathcal{A}_n(L_1([0, T]), X(0), W)} \left\{ \mathbb{E}[\|X - \widehat{X}_n\|_1] \right\} \geq c \cdot n^{-1/2}.$$

(ii) If  $\delta > 2$  and  $\beta > 0$  or if  $\delta = 1$ , then there exist constants  $c, C \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  we have

$$\begin{aligned} c \cdot \sqrt{\ln(n+1)/n} &\leq \inf_{\widehat{X}_n \in \mathcal{A}_n(C([0, T]), X(0), W)} \left\{ \mathbb{E}[\|X - \widehat{X}_n\|_\infty] \right\} \\ &\leq \inf_{\widehat{X}_n \in \mathcal{A}_n^{\text{eq}}(C([0, T]), X(0), W)} \left\{ \mathbb{E}[\|X - \widehat{X}_n\|_\infty] \right\} \leq C \cdot \sqrt{\ln(n+1)/n}. \end{aligned}$$

(iii) If  $\delta > 1$  then there exist constants  $c, C \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  we have

$$\begin{aligned} c \cdot n^{-1/2} &\leq \inf_{\widehat{X}_n \in \mathcal{A}_n(L_1([0, T]), X(0), W)} \left\{ \mathbb{E}[\|X - \widehat{X}_n\|_1] \right\} \\ &\leq \inf_{\widehat{X}_n \in \mathcal{A}_n^{\text{eq}}(L_1([0, T]), X(0), W)} \left\{ \mathbb{E}[\|X - \widehat{X}_n\|_1] \right\} \leq C \cdot n^{-1/2}. \end{aligned}$$

(iv) For all  $\delta \in (0, \infty)$  and  $\beta \in [0, \infty)$  there exist constants  $c, C \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  we have

$$\begin{aligned} c \cdot n^{-\min(1/2, \delta/2)} &\leq \inf_{\widehat{X}_n \in \mathcal{A}_n^{\text{eq}}(C([0, T]), X(0), W)} \left\{ \sup_{t \in [0, T]} \mathbb{E}[|X(t) - \widehat{X}_n(t)|] \right\} \\ &\leq C \cdot (1 + \mathbf{1}_{\{1\}}(\delta)) \cdot \mathbf{1}_{\mathbb{R} \setminus \{0\}}(\beta) \cdot \sqrt{\ln(n)} \cdot n^{-\min(1/2, \delta/2)}. \end{aligned}$$

*Proof.* Similar to the reasoning in the previous section we conclude that the assumptions of Theorem 9 and Theorem 12 are fulfilled without any restriction on the parameters  $\delta$  and  $\beta$ . This shows part (i).

The lower bound in (ii) clearly follows from the first statement in (i). The upper bound in (ii) is due to Dereich, Neuenkirch, and Szpruch [11, Theorem 1.1] if  $\delta > 2$  and  $\beta > 0$  and due to Hefter and Herzwurm [17, Remark 7] (with a linearly interpolated version of the corresponding numerical scheme) if  $\delta = 1$ .

The lower bound in (iii) clearly follows from the second statement in (i). The upper bound in (iii) is a consequence of the upper bound in (iv).

The lower bound in (iv) follows from the second statement in (i) and (36). The upper bound in (iv) follows from Hefter and Herzwurm [18, Theorem 2] (with a linearly interpolated version of the corresponding numerical scheme) if  $\delta \neq 1$ , from the upper bound in (ii) if  $\delta = 1$  and  $\beta \neq 0$ , and from Hefter and Herzwurm [17, Corollary 1] (with a linearly interpolated version of the corresponding numerical scheme) if  $\delta = 1$  and  $\beta = 0$ .  $\square$

Concerning Corollary 15(ii) we add that no matching upper and lower bounds are known for  $\delta \in (0, 2) \setminus \{1\}$ , cf. Hefter and Herzwurm [18, Figure 1]. Concerning Corollary 15(iii) we add that no matching upper and lower bounds are known for  $\delta \in (0, 1)$ .

We stress that Corollary 15(iv) is the first result in the literature that provides matching upper and lower bounds (up to logarithmic terms) for a particular error criterion without any restriction on the parameters.

**7.2. SDEs with Superlinearly Growing Coefficients.** For simplicity, we assume throughout this section that

$$a(t, x) = a(x) \quad \text{and} \quad b(t, x) = b(x) \quad \text{are polynomials.}$$

Furthermore, for a polynomial  $h: \mathbb{R} \rightarrow \mathbb{R}$  we put

$$\text{zeros}(h) = \{x \in \mathbb{R}: h(x) = 0\}.$$

The following result is an immediate consequence of Theorem 6 as well as Theorem 9 and Theorem 12.

**Corollary 16** (Superlinearly growing coefficients).

(i) Assume that

$$\mathbb{P}(X(0) \notin (\text{zeros}(b) \cup \text{zeros}(a'b - ab' - \frac{1}{2}b^2b''))) > 0.$$

Then there exists a constant  $c \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  we have

$$\inf_{\widehat{X}_n(T) \in \mathcal{A}_n(\mathbb{R}, X(0), W)} \left\{ \mathbb{E} \left[ |X(T) - \widehat{X}_n(T)| \right] \right\} \geq c \cdot n^{-1}.$$

(ii) Assume that

$$\mathbb{P}(X(0) \notin \text{zeros}(b)) > 0.$$

Then there exists a constant  $c \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  we have

$$\inf_{\widehat{X}_n \in \mathcal{A}_n(C([0, T]), X(0), W)} \left\{ \mathbb{E} [\|X - \widehat{X}_n\|_\infty] \right\} \geq c \cdot \sqrt{\ln(n+1)/n}$$

and

$$\inf_{\widehat{X}_n \in \mathcal{A}_n(L_1([0, T]), X(0), W)} \left\{ \mathbb{E} [\|X - \widehat{X}_n\|_1] \right\} \geq c \cdot n^{-1/2}.$$

There are matching upper error bounds under certain monotone conditions on the coefficients. These bounds are achieved by tamed or projected or implicit versions of the Euler scheme in case of the mean  $L_p$ -error and the mean  $L_\infty$ -error, see Beyn, Isaak, and Kruse [3], Hutzenthaler, Jentzen, and Kloeden [26, 27], Sabanis [47], and of the Milstein scheme in case of the error at a single time, see Beyn, Isaak, and Kruse [4], Kumar and Sabanis [30], Wang and Gan [48]. In all of these cases, the corresponding methods are non-adaptive and based on a fixed equidistant grid, and therefore, adaptive algorithms are not superior to non-adaptive ones. For an example see equation (2) and the subsequent discussion in the introduction.

**7.3. SDEs with Discontinuous Coefficients.** As an illustrating example we consider the SDE

$$(38) \quad dX(t) = \text{sgn}(X(t)) \cdot (1 + X(t)) dt + dW(t), \quad X(0) = x_0,$$

with initial value  $x_0 \in \mathbb{R}$ , where  $\text{sgn}(x) = 1$  for  $x \in [0, \infty)$  and  $\text{sgn}(x) = -1$  for  $x \in (-\infty, 0)$ . Here, the coefficients  $a, b$  are given by

$$a(t, x) = a(x) = \text{sgn}(x) \cdot (1 + x) \quad \text{and} \quad b(t, x) = b(x) = 1$$

for every  $(t, x) \in [0, T] \times \mathbb{R}$ . Strong existence and pathwise uniqueness for the SDE (38) follows from, e.g., Zvonkin [51, Theorem 4]. Clearly, the coefficients  $a, b$  of the SDE (38) are infinitely often differentiable on  $\mathbb{R} \setminus \{0\}$ . Moreover, for every  $x \in \mathbb{R} \setminus \{0\}$  it holds that

$$(a'b - ab' - \frac{1}{2}b^2b'')(x) = \text{sgn}(x)$$

and it is easy to see that

$$\mathbb{P}(\forall t \in [0, T/2]: X(t) = 0) \neq 1.$$



Hence the assumptions of Theorems 6, 9, 12 are satisfied. Combining this with an upper bound of Leobacher and Szölgyenyi [31, Theorem 3.1] (with a linearly interpolated version of the corresponding numerical scheme) shows that there exist constants  $c, C \in (0, \infty)$  such that for all  $n \in \mathbb{N}$  we have

$$\begin{aligned} \inf_{\widehat{X}_n \in \mathcal{A}_n(\mathbb{R}, X(0), W)} \mathbb{E}[|X(T) - \widehat{X}_n(T)|] &\geq c \cdot n^{-1}, \\ \inf_{\widehat{X}_n \in \mathcal{A}_n(C([0, T]), X(0), W)} \mathbb{E}[\|X - \widehat{X}_n\|_\infty] &\geq c \cdot \sqrt{\ln(n+1)/n} \end{aligned}$$

as well as

$$\begin{aligned} c \cdot n^{-1/2} &\leq \inf_{\widehat{X}_n \in \mathcal{A}_n(L_1([0, T]), X(0), W)} \mathbb{E}[\|X - \widehat{X}_n\|_1] \\ &\leq \inf_{\widehat{X}_n \in \mathcal{A}_n^{\text{eq}}(L_1([0, T]), X(0), W)} \mathbb{E}[\|X - \widehat{X}_n\|_1] \leq C \cdot n^{-1/2}. \end{aligned}$$

In particular, adaptive algorithms are not superior (up to multiplicative constants) to algorithms that are based on fixed equidistant grids for global approximation with respect to the  $L_1$ -norm.

A further example of an SDE with discontinuous coefficients that is often considered in the literature is given by

$$dX(t) = \text{sgn}(X(t)) dt + dW(t), \quad X(0) = x_0,$$

with initial value  $x_0 \in \mathbb{R}$ , see, e.g., Göttlich, Lux, and Neuenkirch [15, Table 1]. Observe that in this case the assumptions of Theorem 6 for pointwise approximation are not fulfilled since  $a'b - ab' - \frac{1}{2}b^2b'' = 0$  on  $\mathbb{R} \setminus \{0\}$ . As a consequence, it could be possible that adaptive algorithms are able to achieve a convergence rate greater than 1 for this SDE.

#### APPENDIX A. PROPERTIES OF THE NORMAL DISTRIBUTION

We collect a number of properties of the Gaussian distribution on the real line, which are employed in the proofs of Proposition 4, Proposition 7, and Proposition 10. We suppose that these properties are well-known but for the convenience of the reader we provide proofs of these facts.

**Lemma 17.** *Let  $Z$  be a real-valued random variable that is normally distributed. Then for all  $\sigma \in [0, \infty)$  with  $\sigma^2 \leq \text{Var}[Z]$  and  $\varepsilon \in [0, \infty)$  we have*

$$\mathbb{P}(|Z| \geq \varepsilon \sigma) \geq 1 - \varepsilon.$$

*Proof.* We may assume  $\text{Var}[Z] > 0$  and  $\sigma^2 = \text{Var}[Z]$ . Let  $\varepsilon \in [0, \infty)$  and  $\mu = \mathbb{E}[Z]$ . By the Anderson inequality we have  $\mathbb{P}(|Z| < \varepsilon \sigma) \leq \mathbb{P}(|Z - \mu| < \varepsilon \sigma)$ , see Anderson [2] or Lifshits [32, Corollary 7.1, p. 47]. Hence

$$\mathbb{P}(|Z| \geq \varepsilon \sigma) \geq 1 - 2 \int_0^\varepsilon \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx \geq 1 - \frac{2}{\sqrt{2\pi}} \varepsilon \geq 1 - \varepsilon.$$

□

**Lemma 18.** *Let  $\varepsilon, \delta \in (0, \infty)$ . Then there exists a constant  $c \in (0, \infty)$  with the following property. If  $N \in \mathbb{N}$  and  $Z_1, \dots, Z_N$  are independent real-valued random variables each being normally distributed with variance at least  $\delta^2$ , then*

- (i)  $\mathbb{P}\left(\sum_{i=1}^N |Z_i| \geq cN\right) \geq 1 - \varepsilon,$
- (ii)  $\mathbb{P}\left(\max_{i=1, \dots, N} |Z_i| \geq c\sqrt{\ln(N)}\right) \geq 1 - \varepsilon.$

*Proof.* Let  $\varepsilon, \delta \in (0, \infty)$  and let  $(Y_i)_{i \in \mathbb{N}}$  be a sequence of independent, real-valued, standard normal random variables.

Put  $M_N = N^{-1} \cdot \sum_{i=1}^N |Y_i|$  for  $N \in \mathbb{N}$ . By the strong law of large numbers we have  $\lim_{N \rightarrow \infty} M_N = \mathbb{E}[|Y_1|] = \sqrt{2/\pi}$  with probability one. Since  $\sqrt{2/\pi} > 1/2$  we conclude that  $\lim_{N \rightarrow \infty} \mathbb{P}(M_N \leq 1/2) = 0$ . Hence

$$(39) \quad \exists N_1 \in \mathbb{N} \quad \forall N \geq N_1: \mathbb{P}\left(\sum_{i=1}^N |Y_i| \leq N/2\right) \leq \varepsilon.$$

Furthermore, since  $\mathbb{P}(\min_{N=1, \dots, N_1-1} M_N > 0) = 1$ , there exists  $c_1 \in (0, \infty)$  such that

$$(40) \quad \forall N < N_1: \mathbb{P}\left(\sum_{i=1}^N |Y_i| \leq c_1 N\right) \leq \varepsilon.$$

Next, define  $\varphi: (1, \infty) \rightarrow (0, \infty)$  by

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \cdot \left(\frac{1}{t} - \frac{1}{t^3}\right) \cdot \exp(-t^2/2).$$

Using a well-known bound for the tails of the standard normal distribution, see, e.g., Bogachev [6, Lemma 1.1.3], we have

$$\forall N \in \mathbb{N}, t \in (1, \infty): \mathbb{P}\left(\max_{i=1, \dots, N} |Y_i| \leq t\right) = \mathbb{P}(|Y_1| \leq t)^N \leq (1 - 2\varphi(t))^N \leq \exp(-2\varphi(t)N).$$

Note that  $\lim_{N \rightarrow \infty} \varphi(c \sqrt{\ln(N)}) \cdot N = \infty$  for every  $c \in (0, \sqrt{2})$ . Hence

$$(41) \quad \exists N_2 \geq 2 \quad \forall N \geq N_2: \mathbb{P}\left(\max_{i=1, \dots, N} |Y_i| \leq \sqrt{\ln(N)}\right) \leq \varepsilon.$$

Put  $c_2 = \varepsilon / \sqrt{\ln(N_2)} \in (0, \infty)$ . Employing Lemma 17 we obtain

$$(42) \quad \forall N < N_2: \mathbb{P}\left(\max_{i=1, \dots, N} |Y_i| \leq c_2 \sqrt{\ln(N)}\right) \leq \mathbb{P}(|Y_1| \leq c_2 \sqrt{\ln(N_2)}) \leq \varepsilon.$$

Finally, let  $(\mu_i)_{i \in \mathbb{N}} \subset \mathbb{R}$ ,  $(\sigma_i)_{i \in \mathbb{N}} \subset [\delta, \infty)$  and let  $(Z_i)_{i \in \mathbb{N}}$  be a sequence of independent, real-valued random variables with  $Z_i \sim N(\mu_i, \sigma_i^2)$  for every  $n \in \mathbb{N}$ . By the Anderson inequality, see Anderson [2] or Lifshits [32, Corollary 7.1, p. 47], and the assumption on the variances  $(\sigma_i)_{i \in \mathbb{N}}$  we have for every  $p \in [1, \infty]$  that

$$(43) \quad \begin{aligned} \forall N \in \mathbb{N}, z \in [0, \infty): \mathbb{P}(|(Z_i)_{i=1, \dots, N}|_p \leq z) &\leq \mathbb{P}(|(Z_i - \mu_i)_{i=1, \dots, N}|_p \leq z) \\ &\leq \mathbb{P}\left(|((Z_i - \mu_i)/\sigma_i)_{i=1, \dots, N}|_p \leq z/\delta\right). \end{aligned}$$

Let  $c = \min(1/2, c_1, c_2) \cdot \delta \in (0, \infty)$ . Combining (39), (40), and (43) for  $p = 1$  we obtain

$$\forall N \in \mathbb{N}: \mathbb{P}\left(\sum_{i=1}^N |Z_i| \leq cN\right) \leq \mathbb{P}\left(\sum_{i=1}^N |Y_i| \leq \min(1/2, c_1) \cdot N\right) \leq \varepsilon,$$

which yields the statement in (i). Combining (41), (42), and (43) for  $p = \infty$  we obtain

$$\forall N \in \mathbb{N}: \mathbb{P}\left(\max_{i=1, \dots, N} |Z_i| \leq c \sqrt{\ln(N)}\right) \leq \mathbb{P}\left(\max_{i=1, \dots, N} |Y_i| \leq \min(1, c_2) \sqrt{\ln(N)}\right) \leq \varepsilon,$$

which yields the statement in (ii) and completes the proof.  $\square$

**Lemma 19.** *Let  $\varepsilon, \delta \in (0, \infty)$ . Then there exists a constant  $c \in (0, \infty)$  with the following property. If  $k \in 2\mathbb{N}$ ,  $B_1, \dots, B_{k/2}$  are independent Brownian bridges (from 0 to 0) on  $[0, 1/k]$ ,  $f_1, \dots, f_{k/2} \in L_1([0, 1/k])$ , and  $a_1, \dots, a_{k/2} \in \mathbb{R}$  with  $|a_1|, \dots, |a_{k/2}| \geq \delta$ , then*

$$\mathbb{P}\left(\sum_{i=1}^{k/2} \|a_i \cdot B_i - f_i\|_1 \geq c/\sqrt{k/2}\right) \geq 1 - \varepsilon.$$

*Proof.* Let  $\varepsilon, \delta \in (0, \infty)$  and choose  $c \in (0, \infty)$  according to Lemma 18. For  $i = 1, \dots, k/2$  define

$$Z_i = \sqrt{12} k^{3/2} \cdot \int_0^{1/k} (a_i \cdot B_i(t) - f_i(t)) dt.$$

Then  $Z_1, \dots, Z_{k/2}$  are independent normal random variables with

$$\text{Var}(Z_i) = 12k^3 a_i^2 \cdot \mathbb{E}\left[\left(\int_0^{1/k} B_i(t) dt\right)^2\right] = a_i^2 \geq \delta^2$$

for  $i = 1, \dots, k/2$ , see, e.g., Müller-Gronbach [38, Equation (16)]. Thus, by Lemma 18(i),

$$\mathbb{P}\left(\sum_{i=1}^{k/2} \|a_i \cdot B_i - f_i\|_1 \geq c/\sqrt{48k}\right) \geq \mathbb{P}\left(\sum_{i=1}^{k/2} |Z_i| \geq c \cdot k/2\right) \geq 1 - \varepsilon,$$

which completes the proof.  $\square$

#### APPENDIX B. LOCALIZATION TECHNIQUE: COMPARISON RESULTS

We relate the solution  $X$  of the integral equation (3) in Section 3 to the solution  $\tilde{X}$  of an integral equation with the same initial value  $\tilde{X}(0) = X(0)$  and coefficients  $\tilde{a}$  and  $\tilde{b}$  that coincide with  $a$  and  $b$  on a stripe  $[0, T] \times I \subseteq [0, T] \times \mathbb{R}$ , respectively.

**Lemma 20.** *Assume the setting in Section 3. Let  $\tilde{a}: [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  and  $\tilde{b}: [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  be Borel-measurable functions, let  $\emptyset \neq I \subseteq \mathbb{R}$  be an open interval and let  $C \in [0, \infty)$  be a constant such that for all  $t \in [0, T]$  and  $x \in \mathbb{R}$  it holds that*

$$|\tilde{a}(t, x)| + |\tilde{b}(t, x)| \leq C(1 + |x|),$$

for all  $t \in [0, T]$  and  $x, y \in I$  it holds that

$$(44) \quad |\tilde{a}(t, x) - \tilde{a}(t, y)| + |\tilde{b}(t, x) - \tilde{b}(t, y)| \leq C|x - y|,$$

and for all  $t \in [0, T]$  and  $x \in I$  it holds that

$$(45) \quad a(t, x) = \tilde{a}(t, x), \quad b(t, x) = \tilde{b}(t, x).$$

Assume further that  $\tilde{X}: [0, T] \times \Omega \rightarrow \mathbb{R}$  is an  $(\mathcal{F}_t)_{t \in [0, T]}$ -adapted stochastic process with continuous paths such that for all  $t \in [0, T]$  it holds  $\mathbb{P}$ -a.s. that

$$(46) \quad \tilde{X}(t) = X(0) + \int_0^t \tilde{a}(s, \tilde{X}(s)) ds + \int_0^t \tilde{b}(s, \tilde{X}(s)) dW(s).$$

Define the stopping time

$$(47) \quad \tau = \inf\{t \in [0, T]: X(t) \notin I\} \wedge \inf\{t \in [0, T]: \tilde{X}(t) \notin I\} \wedge T.$$

Then  $\mathbb{P}$ -a.s. for every  $t \in [0, T]$

$$X(t \wedge \tau) = \tilde{X}(t \wedge \tau).$$

Moreover,

$$\mathbb{P}(\{\forall t \in [0, T]: \tilde{X}(t) = X(t)\} \cap \{\forall t \in [0, T]: \tilde{X}(t) \in I\}) = \mathbb{P}(\forall t \in [0, T]: \tilde{X}(t) \in I).$$

*Proof.* For every  $n \in \mathbb{N}$  we define a stopping time by

$$\sigma_n = \inf\{t \in [0, T]: \max(|X(t)|, |\tilde{X}(t)|) \geq n\} \wedge T.$$

Then

$$(48) \quad X(t \wedge \tau) = \lim_{n \rightarrow \infty} X(t \wedge \tau \wedge \sigma_n), \quad \tilde{X}(t \wedge \tau) = \lim_{n \rightarrow \infty} \tilde{X}(t \wedge \tau \wedge \sigma_n)$$

for every  $t \in [0, T]$ . In view of (48) and the pathwise continuity of  $X$  and  $\tilde{X}$  it remains to prove that for every  $t \in [0, T]$  and every  $n \in \mathbb{N}$  it holds  $\mathbb{P}$ -a.s.

$$X(t \wedge \tau \wedge \sigma_n) = \tilde{X}(t \wedge \tau \wedge \sigma_n).$$

To this end we fix  $n \in \mathbb{N}$  and put  $\rho = \tau \wedge \sigma_n$ . Moreover, we put

$$A(t) = a(t, X(t)), \quad \tilde{A}(t) = \tilde{a}(t, \tilde{X}(t)), \quad \bar{A}(t) = \bar{a}(t, X(t))$$

as well as

$$B(t) = b(t, X(t)), \quad \tilde{B}(t) = \tilde{b}(t, \tilde{X}(t)), \quad \bar{B}(t) = \bar{b}(t, X(t))$$

for every  $t \in [0, T]$ . By (45) we have

$$\bar{A} = A \text{ and } \bar{B} = B \text{ on } \{(t, \omega) \in [0, T] \times \Omega: 0 \leq t < \rho(\omega)\}.$$

It follows that for all  $t \in [0, T]$  we have  $\mathbb{P}$ -a.s.

$$\begin{aligned} X(t \wedge \rho) &= X(0) + \int_0^{t \wedge \rho} A(s) ds + \int_0^{t \wedge \rho} B(s) dW(s) \\ &= X(0) + \int_0^{t \wedge \rho} \bar{A}(s) ds + \int_0^{t \wedge \rho} \bar{B}(s) dW(s), \end{aligned}$$

and therefore  $\mathbb{P}$ -a.s.

$$(49) \quad X(t \wedge \rho) - \tilde{X}(t \wedge \rho) = \int_0^{t \wedge \rho} (\bar{A}(s) - \tilde{A}(s)) ds + \int_0^{t \wedge \rho} (\bar{B}(s) - \tilde{B}(s)) dW(s).$$

By the Hölder inequality and (44) we get

$$(50) \quad \begin{aligned} \mathbb{E} \left[ \left( \int_0^{t \wedge \rho} (\bar{A}(s) - \tilde{A}(s)) ds \right)^2 \right] &\leq T \mathbb{E} \left[ \int_0^{t \wedge \rho} |\bar{A}(s) - \tilde{A}(s)|^2 ds \right] \\ &\leq TC^2 \int_0^t \mathbb{E} \left[ |X(s \wedge \rho) - \tilde{X}(s \wedge \rho)|^2 \right] ds. \end{aligned}$$

By the Itô isometry and (44) we get

$$(51) \quad \begin{aligned} \mathbb{E} \left[ \left( \int_0^{t \wedge \rho} (\bar{B}(s) - \tilde{B}(s)) dW(s) \right)^2 \right] &= \mathbb{E} \left[ \int_0^{t \wedge \rho} |\bar{B}(s) - \tilde{B}(s)|^2 ds \right] \\ &\leq C^2 \int_0^t \mathbb{E} \left[ |X(s \wedge \rho) - \tilde{X}(s \wedge \rho)|^2 \right] ds. \end{aligned}$$

Define a bounded and Borel-measurable function  $f: [0, T] \rightarrow [0, \infty)$  by  $f(t) = \mathbb{E}[|X(t \wedge \rho) - \tilde{X}(t \wedge \rho)|^2]$ . Using (49), (50), and (51) we conclude that for all  $t \in [0, T]$ ,

$$f(t) \leq 2(TC^2 + C^2) \int_0^t f(s) ds.$$

Hence  $f = 0$  by the Gronwall inequality. This proves the first part of the lemma.

For the second part, let  $f, g: [0, T] \rightarrow \mathbb{R}$  be continuous functions, let

$$\tau_0 = \inf\{t \in [0, T]: f(t) \notin I\} \wedge \inf\{t \in [0, T]: g(t) \notin I\} \wedge T,$$

and assume that for all  $t \in [0, T]$

$$f(t \wedge \tau_0) = g(t \wedge \tau_0) \quad \text{and} \quad f(t) \in I.$$

If  $\tau_0 < T$ , then the continuity of  $g$  and openness of  $I$  imply  $g(\tau_0) \notin I$  and hence  $f(\tau_0) = g(\tau_0) \notin I$ , which is a contradiction. This shows  $\tau_0 = T$  and hence  $f = g$ .  $\square$

Next, we provide sufficient conditions for an Itô process to stay (over time) in an open interval with positive probability. The following lemma is a consequence of the Girsanov theorem and the method of time change.

**Lemma 21.** *Assume the setting in Section 3. Let  $\alpha: [0, \infty) \times \Omega \rightarrow \mathbb{R}$  and  $\beta: [0, \infty) \times \Omega \rightarrow \mathbb{R}$  be bounded processes that are progressively measurable with respect to  $(\mathcal{F}_t)_{t \in [0, \infty)}$  such that  $\beta$  is bounded away from zero. Let  $I \subseteq \mathbb{R}$  be an open interval such that  $\mathbb{P}(X(0) \in I) > 0$ . Then we have for all  $S \in (0, \infty)$  that*

$$\mathbb{P}\left(\forall t \in [0, S]: \left[X(0) + \int_0^t \alpha(s) ds + \int_0^t \beta(s) dW(s)\right] \in I\right) > 0.$$

*Proof.* By the properties of  $\beta$ , the process  $Z: [0, \infty) \times \Omega \rightarrow [0, \infty)$  defined by

$$Z(t) = \int_0^t \beta^2(s) ds$$

is strictly increasing as well as (Lipschitz) continuous and satisfies  $Z(0) = 0$  as well as  $\lim_{t \rightarrow \infty} Z(t) = \infty$ . Let  $Z^{-1}$  denote the inverse of  $Z$ . Then  $Z^{-1}(t)$  is a stopping time for every  $t \in [0, \infty)$  and by the Dambis/Dubins-Schwarz theorem, see, e.g., Karatzas and Shreve [29, Theorem 3.4.6], we get that the process  $B: [0, \infty) \times \Omega \rightarrow \mathbb{R}$  defined by

$$(52) \quad B(t) = \int_0^{Z^{-1}(t)} \beta(s) dW(s)$$

is a standard Brownian motion with respect to the normal filtration  $(\mathcal{G}_t)_{t \in [0, \infty)} = (\mathcal{F}_{Z^{-1}(t)})_{t \in [0, \infty)}$ .

Let  $\omega \in \Omega$ . Then  $Z(\omega, \cdot): [0, \infty) \rightarrow [0, \infty)$  is absolutely continuous with  $Z(\omega, \cdot)' = \beta^2(\omega, \cdot) > 0$  Lebesgue almost everywhere. It follows that the inverse  $Z^{-1}(\omega, \cdot): [0, \infty) \rightarrow [0, \infty)$  is absolutely continuous with  $(Z^{-1}(\omega, \cdot))' = 1/(\beta^2(\omega, \cdot) \circ Z^{-1}(\omega, \cdot))$  Lebesgue almost everywhere, see, e.g., Bogachev [5, Exercises 5.8.51, 5.8.52]. Hence, by the change of variable formula we have for every  $t \in [0, \infty)$  that

$$(53) \quad \int_0^t \alpha(\omega, s) ds = \int_0^{Z(t)} \frac{\alpha(\omega, Z^{-1}(\omega, s))}{\beta^2(\omega, Z^{-1}(\omega, s))} ds.$$

Combining (52) and (53) yields that for every  $t \in [0, \infty)$  we have

$$(54) \quad X(0) + \int_0^{Z^{-1}(t)} \alpha(s) ds + \int_0^{Z^{-1}(t)} \beta(s) dW(s) = X(0) + \int_0^t \frac{\alpha(Z^{-1}(s))}{\beta^2(Z^{-1}(s))} ds + B(t).$$

Since the processes  $\alpha$  and  $\beta$  are progressively measurable with respect to  $(\mathcal{F}_t)_{t \in [0, \infty)}$  and the process  $Z^{-1}$  is measurable, we get that the process  $\gamma = (-\alpha(Z^{-1}(s))/\beta^2(Z^{-1}(s)))_{s \in [0, \infty)}$  is measurable and adapted with respect to  $(\mathcal{G}_t)_{t \in [0, \infty)}$ . Moreover, by the boundedness properties of  $\alpha$  and  $\beta$ , the process  $\gamma$  satisfies the Novikov condition, see, e.g., Karatzas and Shreve [29, Corollary 3.5.13]. Fix  $S \in (0, \infty)$ . Then, by the Girsanov theorem, see, e.g., Karatzas and Shreve [29, Theorem 3.5.1], the mapping

$$\tilde{\mathbb{P}}: \mathcal{G}_S \rightarrow [0, 1], \quad A \mapsto \mathbb{E}\left[\mathbf{1}_A \cdot \exp\left(\int_0^S \gamma(t) dB(t) - \frac{1}{2} \int_0^S \gamma^2(t) dt\right)\right]$$

is a probability measure on  $\mathcal{G}_S$ , which is equivalent to  $\mathbb{P}|_{\mathcal{G}_S}$ , and the process  $\tilde{B}: [0, S] \times \Omega \rightarrow \mathbb{R}$  given by

$$\tilde{B}(t) = B(t) - \int_0^t \gamma(s) ds$$

is a  $(\mathcal{G}_t)_{t \in [0, S]}$ -adapted Brownian motion on  $[0, S]$  with respect to  $\tilde{\mathbb{P}}$ . By (54) we then obtain that

$$\begin{aligned} 0 &< \tilde{\mathbb{P}}\left(\forall t \in [0, S]: \left[X(0) + \tilde{B}(t)\right] \in I\right) \\ &= \tilde{\mathbb{P}}\left(\forall t \in [0, S]: \left[X(0) + \int_0^{Z^{-1}(t)} \alpha(s) ds + \int_0^{Z^{-1}(t)} \beta(s) dW(s)\right] \in I\right). \end{aligned}$$

Hence it holds that

$$0 < \mathbb{P}\left(\forall t \in [0, S]: \left[X(0) + \int_0^{Z^{-1}(t)} \alpha(s) ds + \int_0^{Z^{-1}(t)} \beta(s) dW(s)\right] \in I\right).$$

Combining this with the fact that there exists  $c \in (0, \infty)$  such that  $Z^{-1}(t) \geq ct$  for every  $t \in [0, \infty)$  completes the proof of the lemma.  $\square$

The following proposition establishes a comparison result for SDEs. It is a key tool to relax global assumptions on the coefficients in order to obtain lower error bounds for SDEs under non-standard assumptions on the coefficients.

**Proposition 22** (Comparison result). *Assume the setting in Section 3. Let  $\tilde{a}: [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  and  $\tilde{b}: [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  be bounded and Borel-measurable functions such that*

$$(55) \quad \inf_{(t, x) \in [0, T] \times \mathbb{R}} |\tilde{b}(t, x)| > 0$$

and let  $C \in [0, \infty)$  be a constant such that for all  $t \in [0, T]$  and  $x, y \in \mathbb{R}$  it holds that

$$|\tilde{a}(t, x) - \tilde{a}(t, y)| + |\tilde{b}(t, x) - \tilde{b}(t, y)| \leq C|x - y|.$$

Moreover, let  $I \subseteq \mathbb{R}$  be an open interval such that

$$\mathbb{P}(X(0) \in I) > 0$$

and for all  $t \in [0, T]$  and  $x \in I$  it holds that

$$a(t, x) = \tilde{a}(t, x), \quad b(t, x) = \tilde{b}(t, x).$$

Assume further that  $\bar{X}: [0, T] \times \Omega \rightarrow \mathbb{R}$  is an  $(\mathcal{F}_t)_{t \in [0, T]}$ -adapted stochastic process with continuous sample paths such that

$$(56) \quad \mathbb{P}(\{X(0) = \bar{X}(0)\} \cap \{X(0) \in I\}) = \mathbb{P}(X(0) \in I)$$

and for all  $t \in [0, T]$  it holds  $\mathbb{P}$ -a.s. that

$$\bar{X}(t) = \bar{X}(0) + \int_0^t \tilde{a}(s, \bar{X}(s)) ds + \int_0^t \tilde{b}(s, \bar{X}(s)) dW(s).$$

Then we have

$$\mathbb{P}(\forall t \in [0, T]: \bar{X}(t) = X(t)) > 0.$$

*Proof.* Let  $\tilde{X}: [0, T] \times \Omega \rightarrow \mathbb{R}$  be an  $(\mathcal{F}_t)_{t \in [0, T]}$ -adapted stochastic process with continuous sample paths such that for all  $t \in [0, T]$  it holds  $\mathbb{P}$ -a.s. that

$$\tilde{X}(t) = X(0) + \int_0^t \tilde{a}(s, \tilde{X}(s)) ds + \int_0^t \tilde{b}(s, \tilde{X}(s)) dW(s).$$

Define  $\hat{a}: [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$  and  $\hat{b}: [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$  by

$$\hat{a}(t, x) = \begin{cases} \tilde{a}(t, x), & \text{if } t \in [0, T], \\ 0, & \text{else,} \end{cases} \quad \hat{b}(t, x) = \begin{cases} \tilde{b}(t, x), & \text{if } t \in [0, T], \\ 1, & \text{else.} \end{cases}$$

Observe that  $\hat{a}$  and  $\hat{b}$  are Borel-measurable and bounded. Moreover, (55) yields

$$\inf_{(t,x) \in [0,\infty) \times \mathbb{R}} |\hat{b}(t,x)| > 0.$$

Define the process  $\hat{X}: [0, \infty) \times \Omega \rightarrow \mathbb{R}$  by

$$\hat{X}(t) = \begin{cases} \tilde{X}(t), & \text{if } t \in [0, T], \\ W(t) - W(T) + \tilde{X}(T), & \text{else.} \end{cases}$$

Observe that  $\hat{X}$  is an  $(\mathcal{F}_t)_{t \in [0, \infty)}$ -adapted stochastic process with continuous sample paths and for all  $t \in [0, \infty)$  it holds  $\mathbb{P}$ -a.s. that

$$\hat{X}(t) = X(0) + \int_0^t \hat{a}(s, \hat{X}(s)) ds + \int_0^t \hat{b}(s, \hat{X}(s)) dW(s).$$

We may thus apply Lemma 21 with  $\alpha(t) = \hat{a}(t, \hat{X}(t))$ ,  $\beta(t) = \hat{b}(t, \hat{X}(t))$  and  $S = T$  to obtain

$$\mathbb{P}(\forall t \in [0, T]: \hat{X}(t) \in I) > 0,$$

and hence

$$(57) \quad \mathbb{P}(\forall t \in [0, T]: \tilde{X}(t) \in I) > 0.$$

Combining (56) with the facts  $\mathbb{P}(\tilde{X}(0) = X(0)) = 1$  and

$$\mathbb{P}(\forall t \in [0, T]: \tilde{X}(t) = \bar{X}(t)) = \mathbb{P}(\tilde{X}(0) = \bar{X}(0))$$

yields

$$(58) \quad \mathbb{P}(\{\forall t \in [0, T]: \tilde{X}(t) = \bar{X}(t)\} \cap \{\forall t \in [0, T]: \tilde{X}(t) \in I\}) = \mathbb{P}(\forall t \in [0, T]: \tilde{X}(t) \in I).$$

Lemma 20 shows that

$$\mathbb{P}(\{\forall t \in [0, T]: \tilde{X}(t) = X(t)\} \cap \{\forall t \in [0, T]: \tilde{X}(t) \in I\}) = \mathbb{P}(\forall t \in [0, T]: \tilde{X}(t) \in I).$$

Combining this with (58) yields

$$\mathbb{P}(\{\forall t \in [0, T]: \bar{X}(t) = X(t)\} \cap \{\forall t \in [0, T]: \tilde{X}(t) \in I\}) = \mathbb{P}(\forall t \in [0, T]: \tilde{X}(t) \in I).$$

Combining this with (57) yields the claim.  $\square$

#### ACKNOWLEDGMENT

Mario Hefter is supported by the Austrian Science Fund (FWF), Project F5506-N26, which is part of the Special Research Program ‘‘Quasi-Monte Carlo Methods: Theory and Applications’’.

#### REFERENCES

- [1] Aurélien Alfonsi. Strong order one convergence of a drift implicit Euler scheme: Application to the CIR process. *Statist. Probab. Lett.*, 83(2):602–607, 2013.
- [2] Theodore W. Anderson. The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Am. Math. Soc.*, 6:170–176, 1955.
- [3] Wolf-Jürgen Beyn, Elena Isaak, and Raphael Kruse. Stochastic C-stability and B-consistency of explicit and implicit Euler-type schemes. *J. Sci. Comput.*, 67(3):955–987, 2016.
- [4] Wolf-Jürgen Beyn, Elena Isaak, and Raphael Kruse. Stochastic C-stability and B-consistency of explicit and implicit Milstein-type schemes. *J. Sci. Comput.*, 70(3):1042–1077, 2017.
- [5] V. I. Bogachev. *Measure theory. Vol. I, II.* Springer-Verlag, Berlin, 2007.

- [6] Vladimir I. Bogachev. *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998.
- [7] Pamela M. Burrage, R. Herdiana, and Kevin Burrage. Adaptive stepsize based on control theory for stochastic differential equations. *J. Comput. Appl. Math.*, 170(2): 317–336, 2004.
- [8] James M. Calvin, Mario Hefter, and André Herzwurm. Adaptive approximation of the minimum of Brownian motion. *J. Complexity*, 39:17–37, 2017.
- [9] Stamatis Cambanis and Yaozhong Hu. Exact convergence rate of the Euler-Maruyama scheme, with application to sampling design. *Stochastics Stochastics Rep.*, 59(3-4): 211–240, 1996.
- [10] John M. C. Clark and R. J. Cameron. The maximum rate of convergence of discrete approximations for stochastic differential equations. In *Stochastic differential systems (Proc. IFIP-WG 7/1 Working Conf., Vilnius, 1978)*, volume 25 of *Lecture Notes in Control and Information Sci.*, pages 162–171. Springer, Berlin-New York, 1980.
- [11] Steffen Dereich, Andreas Neuenkirch, and Lukasz Szpruch. An Euler-type method for the strong approximation of the Cox-Ingersoll-Ross process. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 468(2140):1105–1115, 2012.
- [12] Jessica G. Gaines and Terence J. Lyons. Variable step size control in the numerical solution of stochastic differential equations. *SIAM J. Appl. Math.*, 57(5):1455–1484, 1997.
- [13] Máté Gerencsér, Arnulf Jentzen, and Diyora Salimova. On stochastic differential equations with arbitrarily slow convergence rates for strong approximation in two space dimensions. *To appear in Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*; *arXiv:1702.03229*, 2017.
- [14] Anja Göing-Jaeschke and Marc Yor. A survey and some generalizations of Bessel processes. *Bernoulli*, 9(2):313–349, 2003.
- [15] Simone Göttlich, Kerstin Lux, and Andreas Neuenkirch. The Euler scheme for stochastic differential equations with discontinuous drift coefficient: A numerical study of the convergence rate. *arXiv:1705.04562*, 2017.
- [16] Martin Hairer, Martin Hutzenthaler, and Arnulf Jentzen. Loss of regularity for Kolmogorov equations. *Ann. Probab.*, 43(2):468–527, 2015.
- [17] Mario Hefter and André Herzwurm. Optimal strong approximation of the one-dimensional squared Bessel process. *To appear in Commun. Math. Sci.*; *arXiv:1601.01455*, 2016.
- [18] Mario Hefter and André Herzwurm. Strong convergence rates for Cox-Ingersoll-Ross processes – full parameter range. *ArXiv e-prints*, 2016.
- [19] Mario Hefter and Arnulf Jentzen. On arbitrarily slow convergence rates for strong numerical approximations of Cox-Ingersoll-Ross processes and squared Bessel processes. *arXiv:1702.08761*, 2017.
- [20] Norbert Hofmann and Thomas Müller-Gronbach. On the global error of Itô-Taylor schemes for strong approximation of scalar stochastic differential equations. *J. Complexity*, 20:732–752, 2004.
- [21] Norbert Hofmann and Thomas Müller-Gronbach. A modified Milstein scheme for approximation of stochastic delay differential equations with constant time lag. *J. Comput. Appl. Math.*, 197(1):89–121, 2006.
- [22] Norbert Hofmann, Thomas Müller-Gronbach, and Klaus Ritter. Optimal approximation of stochastic differential equations by adaptive step-size control. *Math. Comp.*, 69(231): 1017–1034, 2000.
- [23] Norbert Hofmann, Thomas Müller-Gronbach, and Klaus Ritter. Step size control for the uniform approximation of systems of stochastic differential equations with additive noise. *Ann. Appl. Probab.*, 10(2):616–633, 2000.



- [24] Norbert Hofmann, Thomas Müller-Gronbach, and Klaus Ritter. The optimal discretization of stochastic differential equations. *J. Complexity*, 17(1):117–153, 2001.
- [25] Norbert Hofmann, Thomas Müller-Gronbach, and Klaus Ritter. Linear vs standard information for scalar stochastic differential equations. *J. Complexity*, 18:394–414, 2002.
- [26] Martin Hutzenthaler, Arnulf Jentzen, and Peter E. Kloeden. Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. *Ann. Appl. Probab.*, 22(4):1611–1641, 2012.
- [27] Martin Hutzenthaler, Arnulf Jentzen, and Peter E. Kloeden. Divergence of the multi-level Monte Carlo Euler method for nonlinear stochastic differential equations. *Ann. Appl. Probab.*, 23(5):1913–1966, 2013.
- [28] Arnulf Jentzen, Thomas Müller-Gronbach, and Larisa Yaroslavtseva. On stochastic differential equations with arbitrary slow convergence rates for strong approximation. *Commun. Math. Sci.*, 14(6):1477–1500, 2016.
- [29] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.
- [30] Chaman Kumar and Sotirios Sabanis. On Milstein approximations with varying coefficients: the case of super-linear diffusion coefficients. *arXiv:1601.02695*, 2016.
- [31] Gunther Leobacher and Michaela Szölgyenyi. A numerical method for SDEs with discontinuous drift. *BIT*, 56(1):151–162, 2016.
- [32] Mikhail Lifshits. *Lectures on Gaussian processes*. SpringerBriefs in Mathematics. Springer, Heidelberg, 2012.
- [33] Xuerong Mao. *Stochastic differential equations and applications*. Horwood Publishing Limited, Chichester, second edition, 2008.
- [34] Susanne Mauthner. Step size control in the numerical solution of stochastic differential equations. *J. Comput. Appl. Math.*, 100(1):93–109, 1998.
- [35] Kyoung-Sook Moon, Anders Szepessy, Raúl Tempone, and Georgios E. Zouraris. Convergence rates for adaptive weak approximation of stochastic differential equations. *Stoch. Anal. Appl.*, 23(3):511–558, 2005.
- [36] Thomas Müller-Gronbach. The optimal uniform approximation of systems of stochastic differential equations. *Ann. Appl. Probab.*, 12(2):664–690, 2002.
- [37] Thomas Müller-Gronbach. Strong approximation of systems of stochastic differential equations. *Habilitation thesis, TU Darmstadt*, pages iv+161, 2002.
- [38] Thomas Müller-Gronbach. Optimal pointwise approximation of SDEs based on Brownian motion at discrete points. *Ann. Appl. Probab.*, 14(4):1605–1642, 2004.
- [39] Thomas Müller-Gronbach and Larisa Yaroslavtseva. A note on strong approximation of SDEs with smooth coefficients that have at most linearly growing derivatives. *arXiv:1707.08818*, 2017.
- [40] Andreas Neuenkirch. Optimal approximation of SDE’s with additive fractional noise. *J. Complexity*, 22(4):459–474, 2006.
- [41] Andreas Neuenkirch. Optimal pointwise approximation of stochastic differential equations driven by fractional Brownian motion. *Stochastic Process. Appl.*, 118(12):2294–2333, 2008.
- [42] Andreas Neuenkirch and Taras Shalaiko. The maximum rate of convergence for the approximation of the fractional Lévy area at a single point. *J. Complexity*, 33:107–117, 2016.
- [43] Pawel Przybyłowicz. Optimality of Euler-type algorithms for approximation of stochastic differential equations with discontinuous coefficients. *Int. J. Comput. Math.*, 91(7):1461–1479, 2014.
- [44] Pawel Przybyłowicz. Optimal global approximation of SDEs with time-irregular coefficients in asymptotic setting. *Appl. Math. Comput.*, 270:441–457, 2015.

- [45] Pawel Przybyłowicz. Minimal asymptotic error for one-point approximation of SDEs with time-irregular coefficients. *J. Comput. Appl. Math.*, 282:98–110, 2015.
- [46] Werner Rümelin. Numerical treatment of stochastic differential equations. *SIAM J. Numer. Anal.*, 19(3):604–613, 1982.
- [47] Sotirios Sabanis. Euler approximations with varying coefficients: The case of superlinearly growing diffusion coefficients. *Ann. Appl. Probab.*, 26(4):2083–2105, 2016.
- [48] Xiaojie Wang and Siqing Gan. The tamed Milstein method for commutative stochastic differential equations with non-globally Lipschitz continuous coefficients. *Journal of Difference Equations and Applications*, 19(3):466–490, 2013.
- [49] Yuiti Yamato. Stochastic differential equations and nilpotent Lie algebras. *Z. Wahrsch. Verw. Gebiete*, 47(2):213–229, 1979.
- [50] Larisa Yaroslavtseva. On non-polynomial lower error bounds for adaptive strong approximation of SDEs. *J. Complexity*, 42:1–18, 2017.
- [51] Alexander K. Zvonkin. A transformation of the phase space of a diffusion process that removes the drift. *Mat. Sb. (N.S.)*, 22(1):129–149, 1974.

MARIO HEFTER, JOHANN RADON INSTITUTE FOR COMPUTATIONAL AND APPLIED MATHEMATICS, AUSTRIAN ACADEMY OF SCIENCES, ALTENBERGERSTRASSE 69, 4040 LINZ, AUSTRIA  
*E-mail address:* `mario.hefter@ricam.oeaw.ac.at`

ANDRÉ HERZWURM, FACHBEREICH MATHEMATIK, TECHNISCHE UNIVERSITÄT KAISERSLAUTERN, POSTFACH 3049, 67653 KAISERSLAUTERN, GERMANY  
*E-mail address:* `herzwurm@mathematik.uni-kl.de`

THOMAS MÜLLER-GRONBACH, FAKULTÄT FÜR INFORMATIK UND MATHEMATIK, UNIVERSITÄT PASSAU, INNSTRASSE 33, 94032 PASSAU, GERMANY  
*E-mail address:* `thomas.mueller-gronbach@uni-passau.de`