SEMI-IMPLICIT TAYLOR SCHEMES FOR STIFF ROUGH DIFFERENTIAL EQUATIONS

SEBASTIAN RIEDEL AND YUE WU

ABSTRACT. We study a class of semi-implicit Taylor-type numerical methods that are easy to implement and designed to solve multidimensional stochastic differential equations driven by a general rough noise, e.g. a fractional Brownian motion. In the multiplicative noise case, the equation is understood as a rough differential equation in the sense of T. Lyons. We focus on equations for which the drift coefficient may be unbounded and satisfies a one-sided Lipschitz condition only. We prove well-posedness of the methods, provide a full analysis, and deduce their convergence rate. Numerical experiments show that our schemes are particularly useful in the case of stiff rough stochastic differential equations driven by a fractional Brownian motion.

1. INTRODUCTION

Stiff differential equations, frequently encountered in practice, pose a challenging problem for the numerical simulation, both for deterministic and for stochastic systems. For stiff ordinary differential equations, it is well known that implicit methods typically perform better than explicit ones [HW96] and they are usually also the method of choice for stiff stochastic differential equations [KP92] (although one should not blindly follow this rule, cf. [LAE08]). A typical stiff equation possesses one or more coefficients that are unbounded, e.g. linear, with linear growth or even satisfy one-sided growth conditions. In this paper, we will concentrate on drift coefficients which satisfy a one-sided Lipschitz condition.

Stochastic differential equations are usually modeled with white noise, justified by its universality property. However, many recent works indicate that processes with fractional noise might be more appropriate for certain models. For instance, fractional noise has been successfully used in mathematical finance [Gua06, CS17, GJR18, BFG16, EER19] and in models for electricity markets [Ben17]. From a mathematical point of view, the analysis of these equations is usually more challenging due to the memory in the model and, consequently, the lack of the Markov property. In the multiplicative noise case, it is even not clear how to properly define the equations since Itô's theory of stochastic integration does not apply for non-semimartingales such as the fractional Brownian motion. However, this issue can be overcome using Lyons' theory of rough paths [Lyo98] which provides a deterministic theory powerful enough to deal with such equations.

Our aim in the present paper is to define and study numerical schemes that are suitable to solve stiff stochastic differential equations driven by a general rough noise. Inspired by [KPS91, KP92, Hig00], we will concentrate on methods where the implicit parameter appears in the drift component only, usually called *semi-implicit methods*. They are conceptually easier than fully-implicit methods and are known to perform well for stiff stochastic differential equations with additive noise and when the noise parameter is not too large. The most studied numerical schemes in the context of rough

differential equations are Taylor-type schemes [Dav07, FV10a, DNT12, BFRS16] (see, however, the recent works [HHW18, RR20] for an approach to Runge-Kutta methods), and we will study such methods in the present work, too. On the technical level, the biggest challenge is the presence of an unbounded drift in the equation which satisfies a one-sided Lipschitz condition only. In the case of additive noise, we can use a direct calculation to show that the numerical scheme remains bounded under this condition, cf. Section 3. Interestingly, a one-sided growth assumption (which follows by the one-sided Lipschitz property) is not sufficient to guarantee this, cf. [CHJ13, page 43] for a counterexample. In the multiplicative noise case, things are getting much more complicated. In the work [RS17], the authors can show that a rough differential equation with an unbounded drift has a global solution provided that the drift satisfies a further growth assumption in the normal directions, cf. (4.16). Our strategy in the present article is to impose that the continuous equation has a global solution and to derive under this assumption the boundedness and convergence of the numerical scheme. The advantage of this approach is that our results can be applied to any continuous equation, regardless of the precise assumptions on the vector fields, as long as the global solution exists.

The paper is structured as follows. In Section 3, we study the case of an ordinary differential equation perturbed by additive noise. The implicit Euler scheme is defined in (3.9). We show the convergence of the numerical solution to the true solution with a precise rate in Theorem 3.9. In Section 4, we study equations driven by multiplicative noise interpreted as rough differential equations. We define semi-implicit Euler (4.10), Milstein (4.11) and 3rd-order Milstein schemes (4.12) and analyze their convergence rates in Theorem 4.13. Inspired by [DNT12], we also propose *simplified* versions of the respective schemes, i.e. we replace the iterated (stochastic) integrals by a product of increments. These schemes are much easier to implement in practice. The convergence rate for these schemes are studied in Theorem 4.17 for a driver being a general Gaussian process. In Section 5, we illustrate our theoretical results through several numerical experiments for equations driven by a fractional Brownian motion. For both additive and multiplicative noise, the divergence of the forward Euler scheme with coarser step size is also discussed to illustrate the drawback of the forward schemes, while the semi-implicit schemes always return reliable simulations regardless of stepsize. This observation is somehow crucial to applications in the real world which may require less computational costs.

2. PRELIMINARIES AND NOTATION

This section introduces basic notations and useful mathematical results for both Section 3 and 4. Notations which are used exclusively in Section 4 will be postponed to the beginning of Section 4.

Let $T \in (0, \infty)$. Let $(\mathbb{R}^d, |\cdot|)$ be the Euclidean space equipped with the Euclidean distance. By $\langle \cdot, \cdot \rangle$ we denote the scalar product in $\mathbb{R}^d$.

First, let us recall the notion of α-Hölder and p-variation regularity from [FV10a, Chapter 5]:

Definition 2.1. A path $x \colon [0, T] \to \mathbb{R}^d$ is said to be
a. *α-Hölder continuous* with $\alpha \in (0, 1]$ if

$$(2.1) \qquad \|x\|_{\alpha;[0,T]} = \sup_{[s,t] \subset [0,T]} \frac{|x(t) - x(s)|}{|t - s|^\alpha} < \infty;$$

b. of *finite p-variation* for some $p \in [1, \infty)$ if

$$(2.2) \qquad \|x\|_{p\text{-var};[0,T]} = \sup_{(t_i) \subset [0,T]} \left(\sum_i |x(t_{i+1}) - x(t_i)|^p \right)^{\frac{1}{p}} < \infty,$$

where the supremum is taken over all finite partitions of the interval $[0, T]$.

The notation $C^{\alpha\text{-Höl}}([0,T], \mathbb{R}^d)$ is used for the set of α-Hölder paths x, which can be shown to be a Banach space with norm $x \mapsto |x(0)| + \|x\|_{\alpha;[0,T]}$. The notation $C^{p\text{-var}}([0,T], \mathbb{R}^d)$ will be used for the set of continuous $x \colon [0,T] \to \mathbb{R}^d$ of finite p-variation. Indeed, $C^{p\text{-var}}([0,T], \mathbb{R}^d)$ can be shown to be a Banach space with norm $x \mapsto |x(0)| + \|x\|_{p\text{-var};[0,T]}$.

Remark 2.2. Note that $C^{\alpha\text{-Höl}}([0,T], \mathbb{R}^d)$ and $C^{p\text{-var}}([0,T], \mathbb{R}^d)$ are subsets of $C([0,T], \mathbb{R}^d)$, the collection of all continuous path x, with norm $\|x\|_{\infty;[0,T]} := \sup_{t \in [0,T]} |x(t)|$. It is not hard to deduce that

$$(2.3) \qquad \|x\|_{\infty;[0,T]} \leq |x(0)| + \|x\|_{\alpha;[0,T]} T^\alpha$$

if $x \in C^{\alpha\text{-Höl}}([0,T], \mathbb{R}^d)$ and

$$(2.4) \qquad \|x\|_{\infty;[0,T]} \leq |x(0)| + \|x\|_{p\text{-var};[0,T]}$$

if $x \in C^{p\text{-var}}([0,T], \mathbb{R}^d)$.

Furthermore, one easily verifies that $C^{\alpha\text{-Höl}}([0,T], \mathbb{R}^d) \subset C^{p\text{-var}}([0,T], \mathbb{R}^d)$ if $\alpha \geq \frac{1}{p}$ as well as $C^{p\text{-var}}([0,T], \mathbb{R}^d) \subset C^{p'\text{-var}}([0,T], \mathbb{R}^d)$ if $1 \leq p \leq p' < \infty$. A more detailed discussion can be found in [FV10a, Chapter 5].

In the discussion of p-variation regularity, the concept of a *control function* is very useful:

Definition 2.3. Let $\Delta_T := \{(s,t) \in [0,T] \times [0,T] : 0 \leq s \leq t \leq T\}$ denote the simplex. A continuous non-negative function $\omega \colon \Delta_T \to [0, \infty)$ with $\omega(t,t) = 0$ for all $t \in [0,T]$ is called a *control function* if it is *superadditive*, i.e.,

$$\omega(s,t) + \omega(t,u) \leq \omega(s,u) \quad \text{for all } s \leq t \leq u \quad \text{in } [0,T].$$

If there exists a positive constant C such that

$$(2.5) \qquad |x(t) - x(s)|^p \leq C\omega(s,t) \text{ for every } s \leq t,$$

we say that ω *controls the p-variation of x*.

Note that (2.5) immediately implies that $\|x\|_{p\text{-var};[0,T]} \leq \omega(0,T)^{\frac{1}{p}}$. On the other hand, for $x \in C^{p\text{-var}}([0,T], \mathbb{R}^d)$, the function $\omega(s,t) = \|x\|_{p\text{-var};[s,t]}^p$ is a control function which controls the p-variation of x ([FV10a, Proposition 5.8]).

Let us finally recall two versions of the Gronwall inequality [Gro19]. The first version states the Gronwall inequality in a differential form. For a proof we refer, for instance, to [Emm04, Lemma 7.3.2].

Lemma 2.4 (Differential version of Gronwall's inequality). *Assume that $a \colon [0,T] \to \mathbb{R}$ is absolutely continuous and $g, \lambda \colon [0,T] \to \mathbb{R}$ are integrable, i.e., $g, \lambda \in L^1([0,T], \mathbb{R})$. If it holds*

$$\dot{a}(t) \leq g(t) + \lambda(t)a(t), \quad \text{for almost every } t \in [0,T].$$

Then, it follows that

$$a(t) \leq e^{\Lambda(t)}a(0) + \int_0^t e^{\Lambda(t) - \Lambda(s)} g(s) \, ds, \quad \text{for all } t \in [0,T],$$

where $\Lambda(t) := \int_0^t \lambda(s)\,\mathrm{d}s$.

In addition, we also rely on the discrete version of the Gronwall inequality. A proof of this can be found in [Cla87]. For its formulation we use the convention that a sum over an empty index set is equal to zero.

Lemma 2.5 (Discrete version of Gronwall's inequality). *Let $(u_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be two nonnegative sequences which satisfy, for given $a \in [0,\infty)$ and $N \in \mathbb{N}$, that*

$$u_n \le a + \sum_{i=1}^{n-1} b_i u_i, \quad n \in \{1,\dots,N\}.$$

Then, it follows that

$$u_n \le a \exp\Big(\sum_{i=1}^{n-1} b_i\Big), \quad n \in \{1,\dots,N\}.$$

Throughout this paper, the drift term $b\colon \mathbb{R}^d \to \mathbb{R}^d$ in the equation considered is spatially dependent only and assumed to satisfy the following assumption:

Assumption 2.6. The vector field $b\colon \mathbb{R}^d \to \mathbb{R}^d$ is continuous and satisfies a *one-sided Lipschitz condition*, i.e. there exists a constant $C_b \in [0,\infty)$ such that

$$(2.6) \qquad\qquad \langle b(\varsigma) - b(\zeta), \varsigma - \zeta \rangle \le C_b|\varsigma - \zeta|^2 \quad \forall \varsigma, \zeta \in \mathbb{R}^d.$$

Assumption 2.7. The vector field $b\colon \mathbb{R}^d \to \mathbb{R}^d$ is *locally Lipschitz continuous*. To be more precise, there exists a non-decreasing function $\Upsilon_b\colon [0,\infty) \to [0,\infty)$ such that for every $r \ge 0$,

$$(2.7) \qquad\qquad |b(\varsigma) - b(\zeta)| \le \Upsilon_b(r)|\varsigma - \zeta| \quad \forall \varsigma, \zeta \in \overline{B_r(0)},$$

where $\overline{B_r(0)}$ denotes the closed ball in $\mathbb{R}^d$ centered at 0 with radius r.

3. Additive noise

In this section, we consider rough differential equations of the form

$$(3.1) \qquad\qquad \begin{cases} \mathrm{d}y(t) = b(y(t))\,\mathrm{d}t + \mathrm{d}x(t), & t \in (0,T], \\ y(0) = \xi, \end{cases}$$

where $\xi \in \mathbb{R}^d$ is the initial condition, $b\colon \mathbb{R}^d \to \mathbb{R}^d$ is a vector field and $x\colon [0,T] \to \mathbb{R}^d$ is a given path. The equation is understood as an integral equation, i.e. y is the solution to (3.1) if and only if

$$(3.2) \qquad\qquad y(t) = \xi + \int_0^t b(y(s))\,\mathrm{d}s + x(t) - x(0), \quad t \in (0,T].$$

3.1. **Global existence and uniqueness.** To solve (3.1), we will use the following transformation: For $t \in [0,T]$ and $\zeta \in \mathbb{R}^d$ set

$$(3.3) \qquad\qquad f(t,\zeta) := b\big(\zeta + x(t)\big).$$

Then we see that, formally, the solution y to (3.1) is given by $y(t) = z(t) + x(t)$, where z is a solution to the initial value problem

$$(3.4) \qquad\qquad \begin{cases} \dot{z}(t) = f(t,z(t)), & t \in (0,T], \\ z(0) = \xi. \end{cases}$$

Therefore, the problem of solving (3.1) reduces to solving (3.4).

In the following lemma, we collect some properties of the non-homogeneous vector field f.

Lemma 3.1. *Let $x\colon [0,T] \to \mathbb{R}^d$ be bounded. Consider the mapping f defined by (3.3).*

(i) *If Assumption 2.6 is satisfied then f is one-sided Lipschitz continuous in space, uniformly in time, i.e.,*

$$\langle f(t,\varsigma) - f(t,\zeta), \varsigma - \zeta \rangle \leq C_b |\varsigma - \zeta|^2 \quad \forall \varsigma, \zeta \in \mathbb{R}^d \text{ and } t \in [0,T]. \tag{3.5}$$

(ii) *If Assumption 2.7 is satisfied then for every $r \in (0,\infty)$, $\varsigma, \zeta \in \overline{B_r(0)}$ and $t \in [0,T]$, it holds*

$$|f(t,\varsigma) - f(t,\zeta)| \leq \Upsilon_b\big(\rho_r(t)\big)|\varsigma - \zeta|, \tag{3.6}$$

where $\rho_r(t) = r + \|x\|_{\infty;[0,t]}$ is a non-decreasing function.

Proof. Straightforward. $\qquad\square$

Theorem 3.2. *Let $x\colon [0,T] \to \mathbb{R}^d$ be continuous and assume that Assumptions 2.6 and 2.7 are satisfied. Then there exists a unique, global solution z to (3.4) with*

$$\|z\|_{\infty;[0,T]} \leq \mathrm{e}^{C_b T}|\xi| + \int_0^T \mathrm{e}^{C_b(T-s)}\big|b(x(s))\big|\,\mathrm{d}s =: r_z.$$

Moreover, z is Lipschitz continuous on $[0,T]$ with Lipschitz constant bounded by

$$\|z\|_{1;[0,T]} \leq r_z \Upsilon_b\big(r_z + \|x\|_{\infty;[0,T]}\big) + \|b \circ x\|_{\infty;[0,T]}.$$

Proof. The existence of a unique local solution to (3.4) follows directly form the local Lipschitz continuity of f. For instance, we refer to [Hal80]. The norm estimates are also derived from the Gronwall inequality Lemma 2.4 in a standard way. For further details we refer to [EKKL19, Section 3]. $\qquad\square$

Due to $y = z + x$ we also immediately obtain the existence of a unique solution to (3.1). The result summarizes some additional properties which also follow from Theorem 3.2.

Corollary 3.3. *Let $x\colon [0,T] \to \mathbb{R}^d$ be continuous and assume that Assumptions 2.6 and 2.7 are satisfied. Then there exists a unique, global and continuous solution y to (3.1).*

In addition, let $\omega\colon \Delta_T \to [0,\infty)$ be a control function for the p-variation of x, $p \in [1,\infty)$. Then the p-variation of y is also controlled by a control function $\hat{\omega}$ which is given by

$$\hat{\omega}(s,t) = \|y\|^p_{p-\mathrm{var};[s,t]} \leq 2^{p-1}\|z\|^p_{1;[0,T]}|t-s|^p + 2^{p-1}\omega(s,t), \quad (s,t) \in \Delta_T.$$

If x is $\frac{1}{p}$-Hölder continuous, then also y is $\frac{1}{p}$-Hölder continuous with Hölder constant bounded by

$$\|y\|_{\frac{1}{p};[0,T]} \leq \|x\|_{\frac{1}{p};[0,T]} + \|z\|_{1;[0,T]}T^{1-\frac{1}{p}}.$$

Remark 3.4. The above theorem shows that the global, one-sided Lipschitz condition (2.6) implies non-explosion of the solution in the additive noise case. At first sight, this seems not very surprising since (2.6) clearly implies the one-sided growth condition

$$\langle b(\zeta), \zeta \rangle \leq C_1 + C_2|\zeta|^2 \quad \forall \zeta \in \mathbb{R}^d \tag{3.7}$$

which is known to prevent explosion in the case of classical ODEs, i.e., $x \equiv 0$, or in the case of stochastic differential equation driven by a Brownian motion. However, if the equation is driven by a general path, non-explosion can *not* be deduced by simply assuming (3.7). We refer to [CHJ13, p. 43] for a counterexample.

3.2. **Discrete approximations and error analysis.** Let us fix an equidistant partition $\mathcal{T}^h$ of $[0, T]$ of the form

$$\mathcal{T}^h = \{t_0 = 0 < t_1 < \cdots < t_j < \cdots < t_{N_h} = T\} \quad \text{with } t_j = jh. \tag{3.8}$$

Hereby, the step size $h \in (0, T]$ is determined by $h = \frac{T}{N_h}$ with $N_h \in \mathbb{N}$.

We will investigate the implicit Euler scheme given by

$$\begin{cases} y_{j+1} = y_j + hb(y_{j+1}) + x_{j+1} - x_j & \text{for } j \in \{0, \ldots, N_h - 1\}, \\ y_0 = \xi, \end{cases} \tag{3.9}$$

where, for each $i \in \{0, \ldots, N_h\}$, y_i denotes the numerical approximation of the exact solution $y(t_i)$ at time point t_i and x_i is short for $x(t_i)$.

The implementation of (3.9) requires the solution of a nonlinear equation in each time step. The following result ensures the existence of a unique $\mathbb{R}^d$-valued sequence $(y_j)_{0 \le j \le N_h}$ satisfying the difference equation (3.9). Proposition 3.5 is a standard result in nonlinear analysis and often called *Uniform Monotonicity Theorem*. For a proof we refer to [OR00, Chap. 6.4] and [SH96, Theorem C.2].

Proposition 3.5. *Let $G \colon \mathbb{R}^d \to \mathbb{R}^d$ be a continuous mapping such that there exists a constant $L_G \in (0, \infty)$ with*

$$\langle G(\varsigma) - G(\zeta), \varsigma - \zeta \rangle \ge L_G |\varsigma - \zeta|^2, \quad \text{for all } \varsigma, \zeta \in \mathbb{R}^d.$$

Then G is a homeomorphism with Lipschitz continuous inverse. In particular, it holds

$$|G^{-1}(\varsigma) - G^{-1}(\zeta)| \le \frac{1}{L_G} |\varsigma - \zeta|, \quad \text{for all } \varsigma, \zeta \in \mathbb{R}^d.$$

An application of Proposition 3.5 immediately gives the well-posedness of the numerical method.

Theorem 3.6 (Well-posedness)**.** *Let $(x_j)_{0 \le j \le N_h}$ be an $\mathbb{R}^d$-valued sequence. Let Assumption 2.6 be satisfied with one-sided Lipschitz constant C_b. If $C_b h < 1$ then there exists a unique $\mathbb{R}^d$-valued sequence $(y_j)_{0 \le j \le N_h}$ satisfying the difference equation (3.9).*

Proof. Let $C_b h < 1$ and define $G \colon \mathbb{R}^d \to \mathbb{R}^d$ by $G(\zeta) = \zeta - hb(\zeta)$ for all $\zeta \in \mathbb{R}^d$. Then it holds

$$\langle G(\varsigma) - G(\zeta), \varsigma - \zeta \rangle = |\varsigma - \zeta|^2 - h\langle b(\varsigma) - b(\zeta), \varsigma - \zeta \rangle \ge (1 - C_b h)|\varsigma - \zeta|^2.$$

Due to $C_b h < 1$ we have $L_G := 1 - C_b h > 0$. Hence, Proposition 3.5 is applicable. In particular, the sequence $(y_j)_{0 \le j \le N_h}$ defined by

$$y_{j+1} := G^{-1}(y_j + x_{j+1} - x_j)$$

for every $j \in \{0, 1, \ldots, N_h - 1\}$ satisfies (3.9). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We derive a bound for the numerical solution (3.9) uniformly with respect to the step size h.

Proposition 3.7. *Under the same assumptions as in Theorem 3.6, the solution $(y_j)_{0 \le j \le N_h}$ to the difference equation (3.9) satisfies*

$$\max_{n \in \{0, \ldots, N_h\}} |y_n| \le \mathrm{e}^{2C_b T} \Big(|\xi| + \max_{n \in \{1, \ldots, N_h\}} |b(x_n)|T \Big) + \max_{n \in \{0, \ldots, N_h\}} |x_n|,$$

provided $2C_b h \le 1$.

Proof. Define $\hat{y}_j := y_j - x_j$. Then, the recursion (3.9) can be rewritten in terms of $\hat{y}_j$ by

$$(3.10) \qquad \hat{y}_{j+1} = \hat{y}_j + hb(\hat{y}_{j+1} + x_{j+1}), \quad \text{for all } j \in \{0, \dots, N_h - 1\}.$$

To prove the boundedness, we use the following estimation:

$$(3.11) \qquad \begin{aligned} \langle \hat{y}_{j+1} - \hat{y}_j, \hat{y}_{j+1} \rangle &= h\langle b(\hat{y}_{j+1} + x_{j+1}), \hat{y}_{j+1} \rangle \\ &= h\langle b(\hat{y}_{j+1} + x_{j+1}) - b(x_{j+1}), \hat{y}_{j+1} \rangle + h\langle b(x_{j+1}), \hat{y}_{j+1} \rangle \\ &\leq hC_b |\hat{y}_{j+1}|^2 + h|b(x_{j+1})||\hat{y}_{j+1}|, \end{aligned}$$

where the inequality follows from the one-sided Lipschitz condition and the Cauchy–Schwarz inequality. On the other hand, note that

$$(3.12) \qquad \langle u_2 - u_1, u_2 \rangle = |u_2|^2 - \langle u_1, u_2 \rangle \geq |u_2|(|u_2| - |u_1|)$$

for all $u_1, u_2 \in \mathbb{R}^d$. From (3.12) and estimate (3.11), we can conclude that

$$\begin{aligned} |\hat{y}_{j+1}|(|\hat{y}_{j+1}| - |\hat{y}_j|) &\leq \langle \hat{y}_{j+1} - \hat{y}_j, \hat{y}_{j+1} \rangle \\ &\leq hC_b |\hat{y}_{j+1}|^2 + h|b(x_{j+1})||\hat{y}_{j+1}| \end{aligned}$$

After canceling one time $|\hat{y}_{j+1}|$ from both sides of the inequality we arrive at

$$|\hat{y}_{j+1}| - |\hat{y}_j| \leq C_b h |\hat{y}_{j+1}| + h|b(x_{j+1})|$$

for every $j \in \{0, \dots, N_h - 1\}$. Summing both sides up to arbitrary $n \in \{1, \dots, N_h\}$ yields

$$\begin{aligned} |\hat{y}_n| &\leq |\hat{y}_0| + C_b h \sum_{j=0}^{n-1} |\hat{y}_{j+1}| + h \sum_{j=0}^{n-1} |b(x_{j+1})| \\ &\leq |\hat{y}_0| + C_b h \sum_{j=1}^{n-1} |\hat{y}_j| + \frac{1}{2} |\hat{y}_n| + h \sum_{j=0}^{n-1} |b(x_{j+1})| \end{aligned}$$

due to $2C_b h \leq 1$. Then, an application of the discrete Gronwall inequality, i.e., Lemma 2.5, leads to

$$|\hat{y}_n| \leq 2e^{2C_b t_n}\left(|\hat{y}_0| + \max_{j \in \{1, \dots, n\}} |b(x_j)| t_n\right)$$

for every $n \in \{0, \dots, N_h\}$. Finally, after recalling the relationship between y and $\hat{y}$, we arrive at

$$\begin{aligned} \max_{n \in \{0, \dots, N_h\}} |y_n| &\leq \max_{n \in \{0, \dots, N_h\}} |\hat{y}_n| + \max_{n \in \{0, \dots, N_h\}} |x_n| \\ &\leq e^{2C_b T}\left(|\xi| + \max_{n \in \{1, \dots, N_h\}} |b(x_n)| T\right) + \max_{n \in \{0, \dots, N_h\}} |x_n|. \end{aligned}$$

This completes the proof. $\qquad \square$

Based on similar arguments as in the proof of Proposition 3.7 we can derive error estimates for the implicit Euler method. In particular, we show that the solution $(y_j)_{0 \leq j \leq N_h}$ to the difference equation (3.9) converges to the exact solution with an order $\frac{1}{p}$ depending on the regularity of the driving path x. We first investigate the case if x is of finite p-variation.

Theorem 3.8 (*p-variation case*)**.** *Let $x \colon [0, T] \to \mathbb{R}^d$ be continuous and of finite p-variation for some $p \in [1, \infty)$. Suppose that Assumption 2.6 is fulfilled and that the initial value problem (3.1) has a unique solution y of finite p-variation satisfying*

$$\|b \circ y\|_{p\text{-var};[0,T]} < \infty.$$

Then the implicit Euler scheme (3.9) *converges to the solution of* (3.1) *with order* $\frac{1}{p}$. *To be more precise, it holds*

$$\max_{n \in \{0,\ldots,N_h\}} |y(t_n) - y_n| \leq 2T^{1-\frac{1}{p}} \mathrm{e}^{2C_b T} \|b \circ y\|_{p-\mathrm{var};[0,T]} h^{\frac{1}{p}},$$

for all $h \leq \frac{1}{2C_b}$.

Proof. We denote by $e_j := y(t_j) - y_j$ the difference between the solutions of (3.1) and (3.9) at each time step $j \in \{0,\ldots,N_h\}$. Then observe that

$$\langle e_{j+1} - e_j, e_{j+1} \rangle = \Big\langle \int_{t_j}^{t_{j+1}} \big(b(y(s)) - b(y_{j+1})\big)\,\mathrm{d}s, e_{j+1} \Big\rangle$$

$$(3.13) \qquad = \Big\langle \int_{t_j}^{t_{j+1}} \big(b(y(s)) - b(y(t_{j+1}))\big)\,\mathrm{d}s, e_{j+1} \Big\rangle + \Big\langle \int_{t_j}^{t_{j+1}} \big(b(y(t_{j+1})) - b(y_{j+1})\big)\,\mathrm{d}s, e_{j+1} \Big\rangle$$

$$\leq \int_{t_j}^{t_{j+1}} \big\langle b(y(s)) - b(y(t_{j+1})), e_{j+1} \big\rangle\,\mathrm{d}s + C_b h |e_{j+1}|^2,$$

by applying (2.6). Next, the mean value theorem for integrals yields the existence of some $\xi_j \in (t_j, t_{j+1})$ with

$$\int_{t_j}^{t_{j+1}} \big\langle b(y(s)) - b(y(t_{j+1})), e_{j+1} \big\rangle\,\mathrm{d}s = h\big\langle b(y(\xi_j)) - b(y(t_{j+1})), e_{j+1} \big\rangle.$$

From (3.12) it follows that

$$|e_{j+1}|\big(|e_{j+1}| - |e_j|\big) \leq \big\langle e_{j+1} - e_j, e_{j+1} \big\rangle.$$

After inserting this into (3.13) an application of the Cauchy–Schwarz inequality shows that

$$|e_{j+1}|\big(|e_{j+1}| - |e_j|\big) \leq h\big|b(y(\xi_j)) - b(y(t_{j+1}))\big||e_{j+1}| + C_b h |e_{j+1}|^2$$

for every $j \in \{0,\ldots,N_h - 1\}$. Hence, after canceling $|e_{j+1}|$ from both sides of the inequality and summing up to n we arrive at

$$|e_n| \leq |e_0| + h\sum_{j=0}^{n-1} \big|b(y(\xi_j)) - b(y(t_{j+1}))\big| + C_b h \sum_{j=0}^{n-1} |e_{j+1}|$$

$$\leq |e_0| + h\sum_{j=0}^{n-1} \big|b(y(\xi_j)) - b(y(t_{j+1}))\big| + C_b h \sum_{j=1}^{n-1} |e_j| + \frac{1}{2}|e_n|,$$

where the last step follows from $2C_b h \leq 1$. Therefore, we have shown that

$$|e_n| \leq 2|e_0| + 2h\sum_{j=0}^{n-1} \big|b(y(\xi_j)) - b(y(t_{j+1}))\big| + 2C_b h \sum_{j=1}^{n-1} |e_j|$$

for every $n \in \{1,\ldots,N_h\}$. Hence, the discrete Gronwall inequality, Lemma 2.5, is applicable and yields

$$\max_{n \in \{1,\ldots,N_h\}} |e_n| \leq 2\mathrm{e}^{2C_b T}\Big(|e_0| + h\sum_{j=0}^{n-1} \big|b(y(\xi_j)) - b(y(t_{j+1}))\big|\Big).$$

Finally, an application of the Hölder inequality with $1 = \frac{1}{p} + \frac{1}{q}$ gives

$$h \sum_{j=0}^{n-1} \big|b(y(\xi_j)) - b(y(t_{j+1}))\big| \leq \Big(\sum_{j=0}^{N_h-1} h^q \Big)^{\frac{1}{q}} \Big(\sum_{j=0}^{N_h-1} \big|b(y(\xi_j)) - b(y(t_{j+1}))\big|^p \Big)^{\frac{1}{p}}$$

$$\leq T^{\frac{1}{q}} h^{1-\frac{1}{q}} \|b \circ y\|_{p-\mathrm{var};[0,T]}.$$

Since $e_0 = 0$ and $h^{1-\frac{1}{q}} = h^{\frac{1}{p}}$ this completes the proof. $\square$

Observe that the proof of Theorem 3.8 only requires Assumption 2.6. In the next theorem we additionally assume that x is $\frac{1}{p}$-Hölder continuous and b locally Lipschitz continuous. Then we obtain a more explicit error estimate.

Theorem 3.9 (Hölder case). *Let $x\colon [0,T] \to \mathbb{R}^d$ be $\frac{1}{p}$-Hölder continuous for some $p \in [1,\infty)$. Suppose that Assumption 2.6 and Assumption 2.7 are satisfied. Then the implicit Euler scheme (3.9) converges to the solution of (3.1) with order $\frac{1}{p}$. To be more precise, it holds*

$$\max_{n \in \{0,\dots,N_h\}} \big|y(t_n) - y_n\big| \leq 2T\mathrm{e}^{2C_bT} \Upsilon_b\big(\|y\|_{\infty;[0,T]}\big) \|y\|_{\frac{1}{p};[0,T]} h^{\frac{1}{p}},$$

for all step sizes satisfying $2C_b h \leq 1$.

Proof. In light of Theorem 3.8 it remains to estimate

$$\|b \circ y\|^p_{p-\mathrm{var};[0,T]} = \sup_{(t_i) \subset [0,T]} \sum_i |b(y(t_{i+1})) - b(y(t_i))|^p,$$

where the supremum is taken over all finite partitions of the interval $[0,T]$. To this end, let $(t_i)_{i \in \{0,\dots,N\}}$ be an arbitrary partition. From Corollary 3.3 it follows that the exact solution y is bounded. Then, it follows from Assumption 2.7 and the Hölder continuity of y that

$$\sum_i |b(y(t_{i+1})) - b(y(t_i))|^p \leq \Upsilon_b\big(\|y\|_{\infty;[0,T]}\big)^p \sum_i |y(t_{i+1}) - y(t_i)|^p$$

$$\leq \Upsilon_b\big(\|y\|_{\infty;[0,T]}\big)^p \|y\|^p_{\frac{1}{p};[0,T]} \sum_i (t_{i+1} - t_i)$$

$$= \Upsilon_b\big(\|y\|_{\infty;[0,T]}\big)^p \|y\|^p_{\frac{1}{p};[0,T]} T.$$

Inserting this into the error estimate in Theorem 3.8 then yields the assertion. $\square$

Remark 3.10. Under additional assumptions on b it is possible to obtain better convergence rates. For instance, in the case of a Brownian driver the method (3.9) coincides with an implicit version of the Milstein scheme. Provided the vector field b is sufficiently smooth, say $b \in \mathcal{C}_b^2$, it is known that the Milstein scheme converges pathwise with an order close to 1. For instance, we refer to the standard monographs [KP92, Mil95, MT04].

For deterministic and $\frac{1}{p}$-Hölder continuous drivers x the order of convergence $\frac{1}{p}$ is in general optimal. One way to improve the convergence rates is to artificially randomize the numerical method. We point the reader to the more detailed discussions in [EKKL19, KW17] and the references therein.

4. Multiplicative noise: the rough path case

The multiplicative noise version of the equation (3.1) would (naively) take the form

$$(4.1) \qquad \qquad \mathrm{d}y = b(y)\,\mathrm{d}t + \sigma(y)\,\mathrm{d}x, \quad y(0) = \xi$$

where x is an $\mathbb{R}^m$-valued path and $\sigma \colon \mathbb{R}^d \to L(\mathbb{R}^m, \mathbb{R}^d)$. However, this equation is ill-posed in the case when x has low (Hölder-) regularity, and we have to use rough path theory to make sense of it. Thus, we look at the rough differential equation

$$(4.2) \qquad \qquad \mathrm{d}y = b(y)\,\mathrm{d}t + \sigma(y)\,\mathrm{d}\mathbf{x}, \quad y(0) = \xi$$

instead where $\mathbf{x}$ will be a suitable rough path. Let us first recall the basic notions from rough path theory which can be found e.g. in [FH14].

Definition 4.1. Let $p \in [2, 3)$. A $\frac{1}{p}$-*Hölder rough path* is a pair $\mathbf{x} = (x, \mathbb{X}) \colon [0, T]^2 \to \mathbb{R}^m \oplus (\mathbb{R}^m \otimes \mathbb{R}^m)$ for which the algebraic identy

$$(4.3) \qquad \qquad \mathbb{X}_{s,t} - \mathbb{X}_{s,u} - \mathbb{X}_{u,t} = x_{s,u} \otimes x_{u,t}$$

holds for every $s, u, t \in [0, T]$, using the notation $x_{s,t} := x(t) - x(s)$, and for which

$$\|\mathbf{x}\|_{1/p;[0,T]} := \sup_{0 \le s < t \le T} \frac{|x_{s,t}|}{|t - s|^{\frac{1}{p}}} + \sqrt{\|\mathbb{X}\|_{2/p;[0,T]}} < \infty,$$

where

$$\|\mathbb{X}\|_{2/p;[0,T]} := \sup_{0 \le s < t \le T} \frac{|\mathbb{X}_{s,t}|}{|t - s|^{\frac{2}{p}}}.$$

Here we write as $\| \cdot \|_{1/p;[0,T]}$ to distinguish from $\| \cdot \|_{\frac{1}{p};[0,T]}$ introduced in Section 2. $\mathbf{x}$ is called *geometric* if

$$(4.4) \qquad \qquad \mathrm{Sym}(\mathbb{X}_{s,t}) = \frac{1}{2} x_{s,t} \otimes x_{s,t}$$

holds for every $s, t \in [0, T]$. If $\mathbf{x} = (x, \mathbb{X})$ and $\mathbf{y} = (y, \mathbb{Y})$ are two rough paths, their distance will be measured via the metric

$$\varrho_{1/p}(\mathbf{x}, \mathbf{y}) := \sup_{0 \le s < t \le T} \frac{|x_{s,t} - y_{s,t}|}{|t - s|^{\frac{1}{p}}} + \sup_{0 \le s < t \le T} \frac{|\mathbb{X}_{s,t} - \mathbb{Y}_{s,t}|}{|t - s|^{\frac{2}{p}}}.$$

It is possible to give a definition of a p-variation rough path too, but we will only consider the Hölder case here for simplicity. The object $\mathbb{X}_{s,t}$ for $s < t$ should be thought of the the second iterated integral $\int_{s<u<v<t} \mathrm{d}x_u \otimes \mathrm{d}x_v$. Indeed, if x is smooth (e.g. $\frac{1}{p}$-Hölder for $p \in [1, 2)$), we can define $\mathbb{X}$ as the second iterated Young integral [You36] and one can show that $(x, \mathbb{X})$ satisfies the conditions stated in Definition 4.1 and defines a geometric rough path. However, for $p \ge 2$, we are not able to use Young's theory anymore, and we have to *assume* that $\mathbb{X}$ exists.

Definition 4.2. Let $x \colon [0, T] \to \mathbb{R}^m$ be a $\frac{1}{p}$-Hölder path. If W is a finite dimensional vector space, a path $y \colon [0, T] \to W$ is called *controlled by* x if there exists a $\frac{1}{p}$-Hölder path $y' \colon [0, T] \to L(\mathbb{R}^m, W)$ such that for $R_{s,t} := y_{s,t} - y'(s)x_{s,t}$, one has

$$\|R\|_{2/p;[0,T]} := \sup_{0 \le s < t \le T} \frac{|R_{s,t}|}{|t - s|^{\frac{2}{p}}} < \infty.$$

The path y' is called a *Gubinelli derivative* of y.

One can show that the space of controlled paths with the norm

$$\|y, y'\|_{x, \frac{2}{p}; [0,T]} := \|y'\|_{\frac{1}{p}; [0,T]} + \|R\|_{2/p; [0,T]}$$

is a Banach space. Moreover, if (y, y') is controlled by x and σ is a sufficiently smooth function, the path $t \mapsto \sigma(y(t))$ is again controlled by x with Gubinelli derivative $\sigma(y(t))' = D\sigma(y(t))y'(t)$ [FH14, Lemma 7.3].

Theorem 4.3. *Let* $\mathbf{x} = (x, \mathbb{X})$ *be a* $\frac{1}{p}$*-Hölder rough path and* (y, y') *be controlled by* x. *Then the rough integral*

$$\int_0^T y(t) \, \mathrm{d}\mathbf{x}(t)$$

exists as the limit of Riemann sums of the form $\sum y(t_i) x_{t_i, t_{i+1}} + y'(t_i) \mathbb{X}_{t_i, t_{i+1}}$.

Proof. [FH14, Theorem 4.10]. □

Definition 4.4. A path $y \colon [0, T] \to \mathbb{R}^d$ is called a *solution to the rough differential equation*

$$(4.5) \qquad\qquad \mathrm{d}y = b(y) \, \mathrm{d}t + \sigma(y) \, \mathrm{d}\mathbf{x}, \quad y(0) = \xi$$

if

$$y(t) = \xi + \int_0^t b(y(s)) \, \mathrm{d}s + \int_0^t \sigma(y(s)) \, \mathrm{d}\mathbf{x}(s)$$

holds for all $t \in [0, T]$ where the second integral makes sense as a rough integral.

The question whether (4.2) possesses a solution will obviously depend on the coefficients. Next, we will define a space of functions σ which will be important for us.

Definition 4.5. For a function $\sigma \colon \mathbb{R}^d \to L(\mathbb{R}^m, \mathbb{R}^d)$, we denote by $D^k \sigma$ the k-th Fréchet derivative, $k \in \mathbb{N}_0$. The space of all k-times Fréchet differentiable functions will be denoted by $\mathcal{C}^k_{\mathrm{loc}}$. The space $\mathcal{C}^k_b$ consists of all functions $\sigma \in \mathcal{C}^k_{\mathrm{loc}}$ for which all derivatives and the function itself are bounded. Set

$$\mathcal{C}^\infty_{\mathrm{loc}} := \bigcap_{k \in \mathbb{N}} \mathcal{C}^k_{\mathrm{loc}} \quad \text{and} \quad \mathcal{C}^\infty_b := \bigcap_{k \in \mathbb{N}} \mathcal{C}^k_b.$$

We will also use the notation

$$\|\sigma\|_{\mathcal{C}^k} := \sum_{i=0}^k \|D^i \sigma\|_{\infty; \mathbb{R}^d}$$

and

$$\|\sigma\|_{\mathcal{C}^k; B} := \sum_{i=0}^k \|D^i \sigma\|_{\infty; B}$$

for any subset $B \subset \mathbb{R}^d$.

In many situations, (4.2) defines a continuous flow or at least a semiflow. We recall the definition below.

Definition 4.6. Let $I \subset \mathbb{R}$ be an interval and $\mathcal{X}$ be a set. A *semiflow* is a map

$$\phi \colon \{(s, t) \, : \, s \le t, \ s, t, \in I\} \times \mathcal{X} \to \mathcal{X}$$

which satisfies the following two properties:

(i) $\phi(t,t,\cdot) = \mathrm{Id}_{\mathcal{X}}$ for all $t \in I$, where $\mathrm{Id}_{\mathcal{X}} : \mathcal{X} \to \mathcal{X}$ is the identity operator.

(ii) $\phi(u,t,\cdot) \circ \phi(s,u,\cdot) = \phi(s,t,\cdot)$ for all $s \le u \le t$, $s,u,t \in I$.

If $\phi \colon I \times I \times \mathcal{X} \to \mathcal{X}$ and property (ii) holds for all $s,t,u \in I$, we call ϕ a *flow*. If $\mathcal{X}$ is a topological space and ϕ is in addition continuous in all its parameters with respect to the product topology, we call it a *continuous (semi)flow*.

Proposition 4.7. *Let $\mathbf{x} = (x, \mathbb{X})$ be a $\frac{1}{p}$-Hölder rough path, $p \in [2,3)$, and assume $\|\mathbf{x}\|_{\frac{1}{p};[0,T]} \le C_0$. Let $b \colon \mathbb{R}^d \to \mathbb{R}^d$ be bounded and Lipschitz continuous and let $\sigma \in \mathcal{C}_b^3$. Then (4.2) has a unique solution (y, y') in the space of controlled paths with $y'(t) = \sigma(y(t))$. Moreover, the equation induces a continuous flow ϕ and there are constants $\delta > 0$ and $L > 0$ depending on p, C_0, T, b and σ such that*

$$(4.6) \qquad |\phi(s,t,\zeta_1) - \phi(s,t,\zeta_2)| \le (1 + L|t-s|^{\frac{1}{p}})|\zeta_1 - \zeta_2|$$

holds for all $s < t \in [0,T]$ with $|t-s| \le \delta$ and all $\zeta_1, \zeta_2 \in \mathbb{R}^d$.

Proof. Proving that (4.2) has a unique solution for every initial condition ξ is very similar to [FH14, Theorem 8.4]: We first define

$$\mathcal{M}(y, y') := \left(\xi + \int_0^{\cdot} b(y(s))\,\mathrm{d}s + \int_0^{\cdot} \sigma(y(s))\,\mathrm{d}\mathbf{x}(s), \sigma(y(\cdot)) \right)$$

which is a mapping from the space of controlled paths to itself. Next, one can prove that this mapping is a contraction on a sufficiently small time interval $[0, T_0]$ which yields a fixed point, i.e. a solution. We can repeat the argument on the interval $[T_0, T_1]$, $[T_1, T_2]$ and so on. Using boundedness of σ and b, the length of these intervals can be bounded from below, therefore we can eventually construct solutions on every given time interval $[0,T]$ by gluing these solutions together.

It remains to prove the estimate (4.6). We start with a bound for the solution to (4.2). Set

$$(4.7) \qquad R_{u,v}^y := y_{u,v} - y'(u)x_{u,v} = \int_u^v b(y(w))\,\mathrm{d}w + \int_u^v \sigma(y(w))\,\mathrm{d}\mathbf{x}(w) - \sigma(y(u))x_{u,v}.$$

We claim that there is a constant C_1 depending on the parameters above such that

$$(4.8) \qquad \|y, y'\|_{x, \frac{2}{p};[s,t]} = \|y'\|_{\frac{1}{p};[s,t]} + \|R^y\|_{2/p;[s,t]} \le C_1$$

for all $s < t$ with $|t-s| \le \delta$. To see this, note first that

$$\|y'\|_{\frac{1}{p};[s,t]} = \|\sigma(y)\|_{\frac{1}{p};[s,t]} \le \|\sigma\|_{\mathcal{C}^1}\|y\|_{\frac{1}{p};[s,t]}.$$

For the remainder, we have

$$\|R^y\|_{2/p;[s,t]} \le |t-s|^{1-\frac{2}{p}}\|b\|_{\infty} + \sup_{s \le u < v \le t} \frac{\left| \int_u^v \sigma(y(w))\,\mathrm{d}\mathbf{x}(w) - \sigma(y(u))x_{u,v} \right|}{|v-u|^{\frac{2}{p}}}.$$

We can estimate the rough integral using [FH14, Theorem 4.10]: For $s \le u \le v \le t$,

$$\left| \int_u^v \sigma(y(w))\,\mathrm{d}\mathbf{x}(w) - \sigma(y(u))x_{u,v} \right| \le \left| \int_u^v \sigma(y(w))\,\mathrm{d}\mathbf{x}(w) - \sigma(y(u))x_{u,v} - (\sigma(y(u)))'\mathbb{X}_{u,v} \right| + |(\sigma(y(u)))'\mathbb{X}_{u,v}|$$

$$\le C\left(\|x\|_{\frac{1}{p};[u,v]}\|R^{\sigma(y)}\|_{2/p;[u,v]} + \|\mathbb{X}\|_{2/p;[u,v]}\|\sigma(y)'\|_{\frac{1}{p};[u,v]} \right)|v-u|^{\frac{3}{p}}$$

$$+ \|\sigma\|_{\mathcal{C}^1}\|\mathbb{X}\|_{2/p;[u,v]}|v-u|^{\frac{2}{p}}$$

where C depends on p and T. We used here that $\sigma(y)' = D\sigma(y)y' = D\sigma(y)\sigma(y)$. Using this identity again, we also obtain

$$\|\sigma(y)'\|_{\frac{1}{p};[u,v]} \leq \|\sigma\|_{C^3}\|y\|_{\frac{1}{p};[u,v]}.$$

As in [FH14, Lemma 7.3], one can show that

$$\|R^{\sigma(y)}\|_{2/p;[u,v]} \leq \frac{1}{2}\|\sigma\|_{C^2}\|y\|_{\frac{1}{p};[u,v]}^2 + \|\sigma\|_{C^1}\|R^y\|_{2/p;[u,v]}.$$

Putting these estimates together, we see that there is a constant C depending on the claimed parameters such that

$$\|R^y\|_{2/p;[s,t]} \leq C + C|t-s|^{\frac{1}{p}}\|R^y\|_{2/p;[s,t]}.$$

Choosing $\delta = \frac{1}{(2C)^p}$ yields a uniform bound for $\|R^y\|_{2/p;[s,t]}$ and thus also for $\|y,y'\|_{x,\frac{2}{p};[s,t]}$ as claimed. We proceed with proving (4.6). Let $s \in [0,T]$, $\zeta_1, \zeta_2 \in \mathbb{R}^d$ and

$$y^i(t) = \zeta_i + \int_s^t b(y^i(u))\,\mathrm{d}u + \int_s^t \sigma(y^i(u))\,\mathrm{d}\mathbf{x}(u); \quad t \in [s,T]$$

for $i = 1,2$. We will first give an estimate for the Hölder norm of $R^{y^1} - R^{y^2}$ where

$$R_{u,v}^{y^i} := y_{u,v}^i - (y^i)'(u)x_{u,v} = \int_u^v b(y^i(w))\,\mathrm{d}w + \int_u^v \sigma(y^i(w))\,\mathrm{d}\mathbf{x}(w) - \sigma(y^i(u))x_{u,v}, \quad i = 1,2.$$

Using the estimate [FH14, Theorem 4.10] for the rough integral, it is straightforward to show that

$$\begin{aligned}
\|R^{y^1} - R^{y^2}\|_{2/p;[s,t]} &\leq \|b\|_{\mathrm{Lip}}\|y^1 - y^2\|_{\infty;[s,t]}|t-s|^{1-\frac{2}{p}} \\
&\quad + C\left(\|R^{\sigma(y^1)} - R^{\sigma(y^2)}\|_{2/p;[s,t]} + \|\sigma(y^1)' - \sigma(y^2)'\|_{\frac{1}{p};[s,t]}\right)|t-s|^{\frac{1}{p}} \\
&\quad + C\|D\sigma(y^1)\sigma(y^1) - D\sigma(y^2)\sigma(y^2)\|_{\infty;[s,t]}
\end{aligned}$$

(4.9)

where C is a constant depending on the parameters stated above. Clearly,

$$\|y^1 - y^2\|_{\infty;[s,t]} \leq C|\zeta_1 - \zeta_2| + C|t-s|^{\frac{1}{p}}\|y^1 - y^2\|_{\frac{1}{p};[s,t]}.$$

From [FH14, Theorem 7.5],

$$\|R^{\sigma(y^1)} - R^{\sigma(y^2)}\|_{2/p;[s,t]} + \|\sigma(y^1)' - \sigma(y^2)'\|_{\frac{1}{p};[s,t]}$$

$$\leq C\left(\|R^{y^1} - R^{y^2}\|_{2/p;[s,t]} + \|(y^1)' - (y^2)'\|_{\frac{1}{p};[s,t]} + |(y^1)'(s) - (y^2)'(s)| + |\zeta_1 - \zeta_2|\right)$$

where we used the uniform bounds obtained in (4.8). Next,

$$|(y^1)'(s) - (y^2)'(s)| = |\sigma(y^1(s)) - \sigma(y^2(s))| = |\sigma(\zeta_1) - \sigma(\zeta_2)| \leq C|\zeta_1 - \zeta_2|$$

and

$$\|(y^1)' - (y^2)'\|_{\frac{1}{p};[s,t]} \leq C\|y^1 - y^2\|_{\frac{1}{p};[s,t]}.$$

For the last term in (4.9), we see that

$$\|D\sigma(y^1)\sigma(y^1) - D\sigma(y^2)\sigma(y^2)\|_{\infty;[s,t]} \leq C\|y^1 - y^2\|_{\infty;[s,t]} \leq C|\zeta_1 - \zeta_2| + C|t-s|^{\frac{1}{p}}\|y^1 - y^2\|_{\frac{1}{p};[s,t]}.$$

From $y_{s,t}^i = \sigma(\zeta_i)x_{s,t} + R_{s,t}^{y^i}$, we also have

$$\|y^1 - y^2\|_{\frac{1}{p};[s,t]} \leq C|\zeta_1 - \zeta_2| + |t - s|^{\frac{1}{p}}\|R^{y^1} - R^{y^2}\|_{2/p;[s,t]}.$$

Putting all these pieces together, we arrive at an estimate of the form

$$\|y^1 - y^2\|_{\frac{1}{p};[s,t]} + \|R^{y^1} - R^{y^2}\|_{2/p;[s,t]}$$
$$\leq C|t-s|^{1-\frac{1}{p}}\|y^1 - y^2\|_{\frac{1}{p};[s,t]} + C|t-s|^{1-\frac{2}{p}}|\zeta_1 - \zeta_2| + C|\zeta_1 - \zeta_2|$$
$$+ C|t-s|^{\frac{1}{p}}\left(\|R^{y^1} - R^{y^2}\|_{2/p;[s,t]} + \|y^1 - y^2\|_{\frac{1}{p};[s,t]} + |\zeta_1 - \zeta_2|\right).$$

Choosing $\delta > 0$ smaller if necessary, we obtain

$$\|y^1 - y^2\|_{\frac{1}{p};[s,t]} + \|R^{y^1} - R^{y^2}\|_{2/p;[s,t]} \leq C|\zeta_1 - \zeta_2|.$$

Now we have for $t - s \leq \delta$,

$$|y_t^1 - y_t^2| \leq |\zeta_1 - \zeta_2| + |t-s|^{\frac{1}{p}}\|y^1 - y^2\|_{\frac{1}{p};[s,t]} \leq (1 + C|t-s|^{\frac{1}{p}})|\zeta_1 - \zeta_2|$$

which was our claim. $\square$

Next, we define the schemes we will be interested in. Fix a $\frac{1}{p}$-Hölder rough path $(x, \mathbb{X})$. By Lyons' Extension theorem [LCL07, Theorem 3.7], there exists a unique element $\mathbb{X}^3 \colon [0,T]^2 \to (\mathbb{R}^m)^{\otimes 3}$ which satisfies

$$\|\mathbb{X}^3\|_{3/p;[0,T]} = \sup_{0 \leq s < t \leq T} \frac{|\mathbb{X}_{s,t}^3|}{|t-s|^{\frac{3}{p}}} < \infty$$

and

$$\mathbb{X}_{s,t}^3 - \mathbb{X}_{s,u}^3 - \mathbb{X}_{u,t}^3 = \mathbb{X}_{s,u} \otimes x_{u,t} + x_{s,u} \otimes \mathbb{X}_{u,t}$$

for every $s, u, t \in [0,T]$. In the sequel, we will often just speak of a $\frac{1}{p}$-Hölder rough path $\mathbf{x} = (x, \mathbb{X}^2, \mathbb{X}^3)$ where we set $\mathbb{X}^2 := \mathbb{X}$.

Note that we can view $\sigma \colon \mathbb{R}^d \to L(\mathbb{R}^m, \mathbb{R}^d)$ as a collection of vector fields $\sigma = (\sigma_1, \ldots, \sigma_m)$ where $\sigma_i \colon \mathbb{R}^d \to \mathbb{R}^d$ for any $i = 1, \ldots, m$. Recall that vector fields are in one-to-one correspondence with first order differential operators: if $V = (V^1, \ldots, V^d)$ is a vector field, the corresponding first order differential operator is defined by

$$V\varphi(\zeta) = \sum_{i=1}^d V^i(\zeta)\partial_i\varphi(\zeta)$$

for a differentiable function φ. If V and W are vector fields, VW denotes the second order differential operator obtained by applying W and V consecutively.

Definition 4.8. We fix an equidistant partition $\mathcal{T}^h$ of $[0,T]$ of the form

$$\mathcal{T}^h = \{t_0 = 0 < t_1 < \cdots < t_r < \cdots < t_{N_h} = T\} \text{ with } t_r = rh$$

where the step size $h \in (0, T]$ is determined by $h = \frac{T}{N_h}$, $N_h \in \mathbb{N}$. Let $\mathbf{x} = (x, \mathbb{X}^2, \mathbb{X}^3)$ be a $\frac{1}{p}$-Hölder rough path. We define three numerical approximations (y_r^l), $l = 1, 2, 3$, as follows:

$$(4.10) \qquad y_{r+1}^1 = y_r^1 + hb(y_{r+1}^1) + \sum_{i=1}^m \sigma_i(y_r^1) x_{t_r, t_{r+1}}^i,$$

$$(4.11) \qquad y_{r+1}^2 = y_r^2 + hb(y_{r+1}^2) + \sum_{i=1}^m \sigma_i(y_r^2) x_{t_r, t_{r+1}}^i + \sum_{i,j=1}^m \sigma_i \sigma_j \operatorname{Id}(y_r^2) \mathbb{X}_{t_r, t_{r+1}}^{2; i, j}$$

resp.

$$
\begin{aligned}
y_{r+1}^3 &= y_r^3 + hb(y_{r+1}^3) + \sum_{i=1}^m \sigma_i(y_r^3) x_{t_r, t_{r+1}}^i + \sum_{i,j=1}^m \sigma_i \sigma_j \operatorname{Id}(y_r^3) \mathbb{X}_{t_r, t_{r+1}}^{2; i, j} \\
(4.12) \\
&\quad + \sum_{i,j,k=1}^m \sigma_i \sigma_j \sigma_k \operatorname{Id}(y_r^3) \mathbb{X}_{t_r, t_{r+1}}^{3; i, j, k}
\end{aligned}
$$

for $r \in \{0, \dots, N_h - 1\}$ with initial condition $y_0^1 = y_0^2 = y_0^3 = \xi$, provided solutions to these equations exist and are unique.

Remark 4.9. For the readers convenience, we spell out the short-hand notation used above in coordinates: for $\xi \in \mathbb{R}^d$,

$$\sigma_i \sigma_j \operatorname{Id}(\xi) = \sum_{\alpha=1}^d \sigma_i^\alpha(\xi) \partial_\alpha \sigma_j(\xi),$$

$$\sigma_i \sigma_j \sigma_k \operatorname{Id}(\xi) = \sum_{\beta=1}^d \sum_{\alpha=1}^d \sigma_i^\beta(\xi) \partial_\beta \sigma_j^\alpha(\xi) \partial_\alpha \sigma_k(\xi) + \sigma_i^\beta(\xi) \sigma_j^\alpha(\xi) \partial_\beta \partial_\alpha \sigma_k(\xi)$$

where we used the product rule in line 2.

Theorem 4.10. *Let Assumption 2.6 be satisfied with one-sided Lipschitz constant C_b. If $C_b h < 1$ then there exist unique $\mathbb{R}^d$-valued sequences $(y_r^l)_{0 \le r \le N_h}$, $l = 1, 2, 3$, satisfying the difference equations (4.10), (4.11) and (4.12).*

Proof. As for Theorem 3.6, this is just an application of Proposition 3.5.   □

In the next proposition, we calculate the local error of the schemes defined above.

Proposition 4.11. *Let $\mathbf{x} = (x, \mathbb{X}^2, \mathbb{X}^3)$ be a $\frac{1}{p}$-Hölder rough path, $p \in [1, 3)$, and choose C_0 such that $\|\mathbf{x}\|_{1/p; [0, T]} \le C_0$. Let $b \colon \mathbb{R}^d \to \mathbb{R}^d$ be bounded and Lipschitz continuous with Lipschitz constant L and let $\sigma \in \mathcal{C}_b^3$. Consider the solution y to the rough differential equation*

$$\mathrm{d}y = b(y)\, \mathrm{d}t + \sigma(y)\, \mathrm{d}\mathbf{x}, \quad y(t_0) = y_0$$

with $t_0 \in [0, T]$. For $0 < h < \frac{1}{L}$, set

$$y_1^1 = y_0 + hb(y_1^1) + \sum_{i=1}^m \sigma_i(y_0) x_{t_0, t_0 + h}^i,$$

$$y_1^2 = y_0 + hb(y_1^2) + \sum_{i=1}^{m} \sigma_i(y_0) x_{t_0,t_0+h}^i + \sum_{i,j=1}^{m} \sigma_i \sigma_j \, \mathrm{Id}(y_0) \mathbb{X}_{t_0,t_0+h}^{2;i,j}$$

resp.

$$y_1^3 = y_0 + hb(y_1^3) + \sum_{i=1}^{m} \sigma_i(y_0) x_{t_0,t_0+h}^i + \sum_{i,j=1}^{m} \sigma_i \sigma_j \, \mathrm{Id}(y_0) \mathbb{X}_{t_0,t_0+h}^{2;i,j}$$

$$+ \sum_{i,j,k=1}^{m} \sigma_i \sigma_j \sigma_k \, \mathrm{Id}(y_0) \mathbb{X}_{t_0,t_0+h}^{3;i,j,k}.$$

Then there exists a constant $C > 0$ depending on p, C_0, T, b and σ such that

(4.13)
$$|y(t_0 + h) - y_1^1| \leq Ch^{\frac{2}{p}}$$

in the case $p \in [1,2)$ and

(4.14)
$$|y(t_0 + h) - y_1^2| \leq Ch^{\left(1 + \frac{1}{p}\right) \wedge \frac{3}{p}}$$

resp.

(4.15)
$$|y(t_0 + h) - y_1^3| \leq Ch^{1 + \frac{1}{p}}$$

in the case $p \in [1,3)$.

Proof. To prove (4.13), note that

$$|y(t_0 + h) - y_1^1| \leq \left| \int_{t_0}^{t_0+h} b(y(s)) - b(y(t_0)) \, \mathrm{d}s \right| + h|b(y_1^1) - b(y_0)|$$

$$+ \left| \int_{t_0}^{t_0+h} (\sigma(y(s)) - \sigma(y(t_0))) \, \mathrm{d}x(s) \right|.$$

For the first integral, we have

$$\left| \int_{t_0}^{t_0+h} b(y(s)) - b(y(t_0)) \, \mathrm{d}s \right| \leq \|b\|_{\mathrm{Lip}} h^{1+\frac{1}{p}} \|y\|_{\frac{1}{p};[t_0,t_0+h]} \leq Ch^{1+\frac{1}{p}}$$

where we used that $\|y\|_{\frac{1}{p}}$ can be bounded by a constant C depending on the stated parameters which can be deduced from (4.8). For the second term, we use the bound

$$|b(y_1^1) - b(y_0)| \leq \|b\|_{\mathrm{Lip}} |y_1^1 - y_0| \leq \|b\|_{\mathrm{Lip}} |b(y_1^1)| h + \|b\|_{\mathrm{Lip}} |\sigma(y_0)| |x(t_0 + h) - x(t_0)| \leq C(h + h^{\frac{1}{p}}).$$

We can use the standard estimate for Young integrals [You36] for the third term to obtain

$$\left| \int_{t_0}^{t_0+h} (\sigma(y(s)) - \sigma(y(t_0))) \, \mathrm{d}x(s) \right| \leq Ch^{\frac{2}{p}} \|x\|_{\frac{1}{p};[t_0,t_0+h]} \|\sigma(y)\|_{\frac{1}{p};[t_0,t_0+h]} \leq Ch^{\frac{2}{p}}.$$

Hence for a constant C,

$$|y(t_0 + h) - y_1^1| \leq C \left(h^{\frac{2}{p}} + h^{1+\frac{1}{p}} + h^2 \right) \leq Ch^{\frac{2}{p}}$$

and (4.13) is shown. We proceed with (4.14). By definition,

$$D\sigma(y_0)\sigma(y_0)\mathbb{X}_{t_0,t_0+h} = \sum_{i,j=1}^{m} \sigma_i \sigma_j \operatorname{Id}(y_0)(\mathbb{X}_{t_0,t_0+h}^{2;i,j}).$$

We have

$$|y(t_0+h) - y_1^2| \le \left| \int_{t_0}^{t_0+h} b(y(s)) - b(y(t_0)) \, \mathrm{d}s \right| + h|b(y_2^1) - b(y_0)|$$

$$+ \left| \int_{t_0}^{t_0+h} \sigma(y(s)) \, \mathrm{d}\mathbf{x}(s) - \sigma(y(t_0))(x(t_0+h) - x(t_0)) - D\sigma(y(t_0))\sigma(y(t_0))\mathbb{X}_{t_0,t_0+h} \right|$$

where

$$\left| \int_{t_0}^{t_0+h} b(y(s)) - b(y(t_0)) \, \mathrm{d}s \right| \le \|b\|_{\mathrm{Lip}} h^{1+\frac{1}{p}} \|y\|_{\frac{1}{p};[t_0,t_0+h]}$$

and

$$|b(y_2^1) - b(y_0)| \le \|b\|_{\mathrm{Lip}} \|b\|_{\infty} h + \|b\|_{\mathrm{Lip}} \|\sigma\|_{\infty} \|x\|_{\frac{1}{p};[t_0,t_0+h]} h^{\frac{1}{p}} + \|b\|_{\mathrm{Lip}} \|\sigma\|_{\mathcal{C}^1}^2 \|\mathbb{X}\|_{2/p;[t_0,t_0+h]} h^{\frac{2}{p}}$$

$$\le C(h + h^{\frac{1}{p}} + h^{\frac{2}{p}}).$$

It remains to estimate the rough integral. We already saw that y is controlled by x with Gubinelli derivative $y' = \sigma(y)$, and $\sigma(y)$ is controlled by x with Gubinelli derivative $\sigma(y)' = D\sigma(y)\sigma(y)$. By [FH14, Theorem 4.10],

$$\left| \int_{t_0}^{t_0+h} \sigma(y(s)) \, \mathrm{d}\mathbf{x}(s) - \sigma(y(t_0))(x(t_0+h) - x(t_0)) - D\sigma(y(t_0))\sigma(y(t_0))\mathbb{X}_{t_0,t_0+h} \right|$$

$$\le Ch^{\frac{3}{p}} \left(\|x\|_{\frac{1}{p};[t_0,t_0+h]} \|R^{\sigma(y)}\|_{2/p;[t_0,t_0+h]} + \|\mathbb{X}\|_{2/p;[t_0,t_0+h]} \|\sigma(y)\|_{\frac{1}{p};[t_0,t_0+h]} \right).$$

For $R^{\sigma(y)}$, note that

$$R_{t_0,t_0+h}^{\sigma(y)} = \sigma(y(t_0+h)) - \sigma(y(t_0)) - D\sigma(y(t_0))y'(t_0)x_{t_0,t_0+h}$$

$$= \sigma(y(t_0+h)) - \sigma(y(t_0)) - D\sigma(y(t_0))(y(t_0+h) - y(t_0)) + D\sigma(y(t_0))R_{t_0,t_0+h}^y$$

$$= \frac{D^2\sigma(\xi)}{2}(y(t_0+h) - y(t_0))^2 + D\sigma(y(t_0))R_{t_0,t_0+h}^y$$

for some $\xi \in \mathbb{R}^d$ on the line segment between $y(t_0+h)$ and $y(t_0)$. Therefore,

$$\|R^{\sigma(y)}\|_{2/p;[t_0,t_0+h]} \le \frac{C}{2}\|y\|_{\frac{1}{p};[t_0,t_0+h]}^2 + C\|R^y\|_{2/p;[t_0,t_0+h]}.$$

For R^y, we have

$$R_{t_0,t_0+h}^y = y(t_0+h) - y(t_0) - \sigma(y(t_0))x_{t_0,t_0+h} = \int_{t_0}^{t_0+h} b(y(s)) \, \mathrm{d}s + \int_{t_0}^{t_0+h} \sigma(y(s)) \, \mathrm{d}\mathbf{x}(s) - \sigma(y(t_0))x_{t_0,t_0+h}.$$

Setting

$$I_{s,t} := \int_s^t \sigma(y(u)) \, \mathrm{d}\mathbf{x}(u) - \sigma(y(s))x_{s,t} - D\sigma(y(s))\sigma(y(s))\mathbb{X}_{s,t},$$

we obtain

$$\|R^y\|_{2/p;[t_0,t_0+h]} \le Ch^{1-\frac{2}{p}} + h^{\frac{1}{p}}\|I\|_{3/p;[t_0,t_0+h]} + C^2\|\mathbb{X}\|_{2/p;[t_0,t_0+h]}.$$

To summarize, we have seen that there is a constant C depending on the stated parameters such that

$$\|I\|_{3/p;[t_0,t_0+h]} \le C\|x\|_{\frac{1}{p};[t_0,t_0+h]}h^{\frac{1}{p}}\|I\|_{3/p;[s,t]} + C.$$

Therefore, if $h \le \delta$ with $\delta = \frac{1}{2^p C^p \|x\|_{\frac{1}{p}}^p}$, we obtain the bound $\|I\|_{3/p;[t_0,t_0+h]} \le 2C$. Using this, we see that for a constant C,

$$|y(t_0+h) - y_1^2| \le C\left(h^{\frac{3}{p}} + h^{1+\frac{1}{p}} + h^{1+\frac{2}{p}} + h^2\right) \le Ch^{(1+\frac{1}{p})\wedge\frac{3}{p}}$$

provided $h \le \delta$, and (4.14) is shown. The proof for (4.14) is conceptually the same. The additional ingredient is a uniform bound for the $\frac{4}{p}$-Hölder norm of

$$(s,t) \mapsto \int_s^t \sigma(y(u))\,\mathrm{d}\mathbf{x}(u) - \sigma(y(s))x_{s,t} - D\sigma(y(s))\sigma(y(s))\mathbb{X}_{s,t} - \sum_{i,j,k=1}^m \sigma_i\sigma_j\sigma_k\,\mathrm{Id}(y(s))(\mathbb{X}_{s,t}^{3;i,j,k}).$$

This can be achieved by using *second order Gubinelli derivatives*. A path $\hat{y}\colon [0,T] \to L(\mathbb{R}^m, \mathbb{R}^d)$ is called controlled by the geometric rough path $\mathbf{x}$ with first and second Gubinelli derivatives

$$\hat{y}^{(1)}\colon [0,T] \to L(\mathbb{R}^m \otimes \mathbb{R}^m, \mathbb{R}^d) \cong L(\mathbb{R}^m L(\mathbb{R}^m, \mathbb{R}^d)),$$
$$\hat{y}^{(2)}\colon [0,T] \to L((\mathbb{R}^m)^{\otimes 3}, \mathbb{R}^d) \cong L(\mathbb{R}^m, L(\mathbb{R}^m \otimes \mathbb{R}^m, \mathbb{R}^d))$$

if $\hat{y}$, $\hat{y}^{(1)}$ and $\hat{y}^{(2)}$ are $\frac{1}{p}$-Hölder continuous and

$$\hat{y}(t) = \hat{y}(s) + \hat{y}^{(1)}(s)x_{s,t} + \hat{y}^{(2)}(s)\mathbb{X}_{s,t}^2 + R_{s,t}^3,$$
$$\hat{y}^{(1)}(t) = \hat{y}^{(1)}(s) + \hat{y}^{(2)}(s)x_{s,t} + R_{s,t}^2$$

where R^k is $\frac{k}{p}$-Hölder continuous, $k = 2,3$. This is a special case of the general concept introduced in [Gub10], see also [FH14, Section 7.6]. If we set

$$J_{s,t} := \hat{y}(s)x_{s,t} + \hat{y}^{(1)}(s)\mathbb{X}_{s,t}^2 + \hat{y}^{(2)}(s)\mathbb{X}_{s,t}^3,$$

we have

$$
\begin{aligned}
J_{s,t} - J_{s,u} - J_{u,t} &= \left(\hat{y}(s) - \hat{y}(u) + \hat{y}^{(1)}(s)x_{s,u} + \hat{y}^{(2)}(s)\mathbb{X}_{s,u}^2\right)x_{u,t} \\
&\quad + \left(\hat{y}^{(1)}(s) - \hat{y}^{(1)}(u) + \hat{y}^{(2)}(s)x_{s,u}\right)\mathbb{X}_{u,t}^2 + (\hat{y}^{(2)}(s) - \hat{y}^{(2)}(u))\mathbb{X}_{u,t}^3 \\
&= -R_{s,u}^3 x_{u,t} - R_{s,u}^2 \mathbb{X}_{u,t}^2 + (\hat{y}^{(2)}(s) - \hat{y}^{(2)}(u))\mathbb{X}_{u,t}^3.
\end{aligned}
$$

Applying the Sewing lemma [FH14, Lemma 4.2], we obtain

$$\left|\int_s^t \hat{y}(u)\,\mathrm{d}\mathbf{x}(u) - \hat{y}(s)x_{s,t} - \hat{y}^{(1)}(s)\mathbb{X}_{s,t}^{(2)} - \hat{y}^{(2)}(s)\mathbb{X}_{s,t}^{(3)}\right|$$
$$\le C|t-s|^{\frac{4}{p}}\left(\|x\|_{\frac{1}{p};[s,t]}\|R^3\|_{3/p;[s,t]} + \|\mathbb{X}^2\|_{2/p;[s,t]}\|R^2\|_{2/p;[s,t]} + \|\mathbb{X}^3\|_{3/p;[s,t]}\|\hat{y}^{(2)}\|_{\frac{1}{p};[s,t]}\right),$$

similar to [FH14, Theorem 4.10]. We can apply this estimate to

$$\hat{y}(u) = \sigma(y(u)),$$
$$\hat{y}^{(1)}(u) = D\sigma(y(u))\sigma(y(u)) \qquad \text{and}$$
$$\hat{y}^{(2)}(u) = D^2\sigma(y(u))(\sigma(y(u)) \otimes \sigma(y(u))) + D\sigma(y(u))D\sigma(y(u))\sigma(y(u))$$

and proceed as above to obtain the estimate

$$|y(t_0 + h) - y_1^2| \leq C \left(h^{\frac{4}{p}} + h^{1+\frac{1}{p}} + h^{1+\frac{2}{p}} + h^{1+\frac{3}{p}} + h^2 \right) \leq C h^{1+\frac{1}{p}}$$

where we used $1 + \frac{1}{p} < \frac{4}{p}$ for $p < 3$. Details are left to the reader.

$\square$

Remark 4.12. Assuming higher regularity of b, one can easily define a modification of the scheme (y_r^3) which has a local error of $\frac{4}{p}$. However, the order of the implementable schemes which we will define below (cf. Definition 4.15) will not increase for this modification because the rate will be dictated by the rate of the Wong-Zakai approximation, cf. the proof of the forthcoming Theorem 4.17, which is already smaller than the rate obtained for the scheme (y_r^3).

Theorem 4.13. *Let* $\mathbf{x}$ *be a* $\frac{1}{p}$-*Hölder rough path for some* $p \in [1,3)$. *Let* $b \colon \mathbb{R}^d \to \mathbb{R}^d$ *satisfy Assumption 2.6 and 2.7 and let* $\sigma \in \mathcal{C}_{\mathrm{loc}}^3$. *Assume that the rough differential equation* (4.2) *induces a continuous semiflow* ϕ *on the time interval* $[0, T]$ *For* $h > 0$, *consider the numerical approximation* $(y_n^l)_{n \in \{0,...,N_h\}}$, $l = 1, 2, 3$, *defined in* (4.10), (4.11) *resp.* (4.12). *Then there exist constants* $\delta > 0$ *and* $C > 0$ *such that*

$$\max_{n \in \{0,...,N_h\}} \left| y(t_n) - y_n^1 \right| \leq C h^{\frac{2}{p}-1}$$

for $p \in [1, 2)$ *and*

$$\max_{n \in \{0,...,N_h\}} \left| y(t_n) - y_n^2 \right| \leq C h^{\frac{1}{p} \wedge (\frac{3}{p}-1)},$$
$$\max_{n \in \{0,...,N_h\}} \left| y(t_n) - y_n^3 \right| \leq C h^{\frac{1}{p}}$$

for $p \in [1, 3)$ *and all step sizes satisfying* $h < \delta$.

Proof. Since ϕ is continuous, we can find a number $M > 0$ such that

$$\{\phi(0, t, \xi) : t \in [0, T]\} \subset B(0, M).$$

Since b is locally Lipschitz continuous, there is a bounded, Lipschitz continuous function $\bar{b} \colon \mathbb{R}^d \to \mathbb{R}^d$ which coincides with b on $B(0, M)$. Moreover, we can find a $\bar{\sigma} \in \mathcal{C}_b^3$ which coincides with σ on $B(0, M)$. Let $\bar{\phi}$ be the flow induced by the rough differential equation

$$\mathrm{d}y = \bar{b}(y)\,\mathrm{d}t + \bar{\sigma}(y)\,\mathrm{d}\mathbf{x}.$$

Let $(\bar{y}_n^l)_{n \in \{0,...,N_h\}}$, $l = 1, 2, 3$, denote the numerical approximations defined in (4.10), (4.11) resp. (4.12) where we replace b by $\bar{b}$ and σ by $\bar{\sigma}$. Using the local error obtained in Proposition 4.11 and the Lipschitz property of the flow map $\bar{\phi}$ deduced in Proposition 4.7 , it is straightforward, cf. e.g. [FV10a, Section 10.3.5] or [RR20, Proposition 4.1], to deduce the global error bounds

$$\max_{n \in \{0,...,N_h\}} \left| \bar{\phi}(0, t_n, \xi) - \bar{y}_n^1 \right| \leq C h^{\frac{2}{p}-1}$$

for $p \in [1, 2)$ and

$$\max_{n \in \{0, \ldots, N_h\}} \left| \bar{\phi}(0, t_n, \xi) - \bar{y}_n^2 \right| \leq C h^{\frac{1}{p} \wedge (\frac{3}{p} - 1)},$$

$$\max_{n \in \{0, \ldots, N_h\}} \left| \bar{\phi}(0, t_n, \xi) - \bar{y}_n^3 \right| \leq C h^{\frac{1}{p}}$$

for $p \in [1, 3)$ for sufficiently small $h > 0$. Since $\bar{\phi}(0, t, \xi) = \phi(0, t, \xi)$ for all $t \in [0, T]$, we can choose $\delta > 0$ sufficiently small to obtain that $\bar{y}_n^l \in B(0, M)$ for every $n \in \{0, \ldots, N_h\}$, $h < \delta$ and $l = 1, 2, 3$. From the uniqueness statement in Theorem 4.10, it follows that $\bar{y}_n^l = y_n^l$ for every $n \in \{0, \ldots, N_h\}$, $h < \delta$ and $l = 1, 2, 3$ which shows the claim.

$\square$

Remark 4.14. At the current stage, we do not know whether Assumption 2.6 and 2.7 on b alone imply the existence of a semiflow for a generic rough path $\mathbf{x}$, even for σ being bounded. In [RS17], one of us together with M. Scheutzow formulated a further condition: We assumed that there exists a constant $C > 0$ such that

(4.16)
$$\left| b(\xi) - \frac{\langle b(\xi), \xi \rangle \xi}{|\xi|^2} \right| \leq C(1 + |\xi|) \quad \text{for all } \xi \in \mathbb{R}^d \setminus \{0\}.$$

Assuming this assumption in addition to 2.6 and 2.7, [RS17, Theorem 4.3] implies the existence of a semiflow provided $\sigma \in \mathcal{C}_b^4$, therefore Theorem 4.13 applies in this case. The subtle case of unbounded diffusion vector fields was discussed by Lejay in the two works [Lej09, Lej12].

We want to apply numerical schemes in the stochastic case now, i.e. when the driving rough path is random. In this context, the higher order objects (i.e. the iterated integrals) are usually not explicitly known and hard to simulate. To overcome this issue, Deya-Neuenkirch-Tindel propose in [DNT12] a numerical scheme in which they replace the higher order objects by products of increments of the path. The same idea motivates us to look at the following schemes:

Definition 4.15. Let $\mathcal{T}^h$ denote the partition

$$\mathcal{T}^h = \{t_0 = 0 < t_1 < \cdots < t_r < \cdots < t_{N_h} = T\} \quad \text{with } t_r = rh$$

with step size $h = \frac{T}{N_h}$, $N_h \in \mathbb{N}$. Let $x \colon [0, T] \to \mathbb{R}^m$ be a path. Then we define two numerical schemes $(y_r^{\mathfrak{s}, l})$, $l = 2, 3$, as follows:

(4.17)
$$y_{r+1}^{\mathfrak{s}, 2} = y_r^{\mathfrak{s}, 2} + h b(y_{r+1}^{\mathfrak{s}, 2}) + \sum_{i=1}^m \sigma_i(y_r^{\mathfrak{s}, 2}) x_{t_r, t_{r+1}}^i + \frac{1}{2} \sum_{i,j=1}^m \sigma_i \sigma_j \operatorname{Id}(y_r^{\mathfrak{s}, 2}) x_{t_r, t_{r+1}}^i x_{t_r, t_{r+1}}^j$$

resp.

(4.18)
$$y_{r+1}^{\mathfrak{s}, 3} = y_r^{\mathfrak{s}, 3} + h b(y_{r+1}^{\mathfrak{s}, 3}) + \sum_{i=1}^m \sigma_i(y_r^{\mathfrak{s}, 3}) x_{t_r, t_{r+1}}^i + \frac{1}{2} \sum_{i,j=1}^m \sigma_i \sigma_j \operatorname{Id}(y_r^{\mathfrak{s}, 3}) x_{t_r, t_{r+1}}^i x_{t_r, t_{r+1}}^j$$
$$+ \frac{1}{6} \sum_{i,j,k=1}^m \sigma_i \sigma_j \sigma_k \operatorname{Id}(y_r^{\mathfrak{s}, 3}) x_{t_r, t_{r+1}}^i x_{t_r, t_{r+1}}^j x_{t_r, t_{r+1}}^k$$

for $r \in \{0, \ldots, N_h - 1\}$ with initial condition $y_0^{\mathfrak{s}, 2} = y_0^{\mathfrak{s}, 3} = \xi$, provided solutions to these equations exist and are unique.

We have already seen that Assumption 2.6 for be b implies that these schemes are well-defined provided $C_b h < 1$.

We will apply the schemes to rough differential equations driven by Gaussian processes in the sense of Friz-Victoir [FV10b]. Next, we recall a basic existence theorem.

Theorem 4.16. *Let $X = (X^1, \ldots, X^m)$ be a continuous, centered Gaussian process with independent components. Assume that each component has stationary increments and that the function ϱ^2 given by*

$$\varrho^2(|t - s|) = \mathbb{E}(|X_t - X_s|^2)$$

is concave with $\varrho(\tau) = \mathcal{O}(\tau^{\frac{1}{\rho}})$ for $\tau \to 0$ and some $\rho \in \left[1, \frac{3}{2}\right)$. Then there exists a lift of X to an enhanced Gaussian process $\mathbf{X} = (X, \mathbb{X})$ on a set of full measure, i.e. $\mathbf{X}$ is almost surely a $\frac{1}{p}$-Hölder rough path for any $2\rho < p < 3$. The second order process $\mathbb{X}$ is given as a limit in probability of usual Riemann sums.

Proof. Cf. [FGGR16] or [FH14, Theorem 10.9]. $\qquad\square$

Theorem 4.17. *Let $X = (X^1, \ldots, X^m)$ be as in Theorem 4.16 with corresponding lift $\mathbf{X}$. Assume that b satisfies Assumption 2.6 and 2.7 and that $\sigma \in \mathcal{C}^\infty_{loc}$. Assume that for every given $\frac{1}{p}$-Hölder rough path $\mathbf{x}$, the rough differential equation (4.2) induces a continuous semiflow $\phi^{\mathbf{x}}$ for which*

$$(4.19) \qquad \sup_{\mathbf{x}\,:\,\|\mathbf{x}\|_{1/p} \leq C_0} \sup_{t \in [0,T]} |\phi^{\mathbf{x}}(0, t, \xi)| < \infty$$

for any given $C_0 > 0$. Let Y denote the solution to the random rough differential equation (4.2) where we replace $\mathbf{x}$ by $\mathbf{X}$ and let $Y_n^{\mathfrak{s},l}$, $l = 2, 3$, denote the corresponding numerical approximation defined in (4.17) and (4.18).

Then for every $2\rho < p < 3$, there are almost surely finite random variables δ, C_2 and C_3 such that

$$\max_{n \in \{0, \ldots, N_h\}} \left| Y(t_n) - Y_n^{\mathfrak{s},2} \right| \leq C_2 h^{\frac{3}{p} - 1},$$

$$\max_{n \in \{0, \ldots, N_h\}} \left| Y(t_n) - Y_n^{\mathfrak{s},3} \right| \leq C_3 h^{\frac{2}{p} - \frac{1}{2}}$$

for all step sizes satisfying $h < \delta$.

Remark 4.18. The assumption (4.19) is very natural in the context of rough differential equations. It is satisfied, for instance, if the vector fields are bounded [FH14, Proposition 8.3] or if $\sigma \in \mathcal{C}^\infty_b$ and b satisfies (4.16) [RS17].

Proof of Theorem 4.17. The idea of the proof is from [DNT12] and was also used in [FR14]. First, it is easily seen that the schemes defined in (4.17) resp. (4.18) coincide with the ones defined in (4.11) resp. (4.12) when $\mathbf{X}$ is replaced by the canonical lift of the process X^h which is defined as the piecewise linear approximation of X at the points given by $\mathcal{T}^h$. Thus,

$$(4.20) \qquad \max_{n \in \{0, \ldots, N_h\}} \left| Y(t_n) - Y_n^{\mathfrak{s},l} \right| \leq \sup_{t \in [0,T]} |Y(t) - Y^h(t)| + \max_{n \in \{0, \ldots, N_h\}} |Y^h(t_n) - Y_n^{h;l}|$$

where Y^h is the solution to

$$\mathrm{d}Y^h = b(Y^h)\,\mathrm{d}t + \sigma(Y^h)\,\mathrm{d}X^h(\omega); \quad Y_0 = \xi$$

and $(Y_n^{h;l})$, $l = 1, 2$, is defined as in Definition 4.8 when the rough path is the canonical lift of X^h. With Theorem 4.13, we can give an estimate for the second term on the right hand side of

the inequality (4.20) and obtain the rates $\frac{3}{p} - 1$ resp. $\frac{1}{p}$. These estimates are indeed uniform due to assumption (4.19). The first term in (4.20) is the Wong-Zakai error. In the case of bounded vector fields, the solution map of a rough differential equation is locally Lipschitz continuous in the rough path topology [FH14, Theorem 8.5]. Using a localization argument as in Theorem 4.13 together with assumption (4.19), we may assume that the map is locally Lipschitz continuous also in our case. We can thus apply the results in [FR14] to obtain a rate of $\frac{2}{p} - \frac{1}{2}$ for the Wong-Zakai approximation. Since $\frac{3}{p} - 1 < \frac{2}{p} - \frac{1}{p}$ and $\frac{2}{p} - \frac{1}{2} < \frac{1}{p}$ for $p > 2$, the claim follows.

$\square$

5. Numerical experiments

In this section we perform several numerical experiments with the numerical methods discussed in this paper. In our examples we focus on rough differential equations where the driver is generated by a fractional Brownian motion.

Example 5.1. In the following we investigate the performance of the implicit Euler method (3.9) applied to a scalar rough differential equation driven by an additive fractional Brownian motion with different regularities. To be more precise, we consider

$$(5.1) \qquad \begin{cases} \mathrm{d}y(t) = (y(t) - y^3(t))\,\mathrm{d}t + \mathrm{d}B^H(t), & t \in (0,1], \\ y(0) = -3.0, \end{cases}$$

where B^H is a real-valued fractional Brownian motion with Hurst parameter $H \in (0,1)$. Note that the function $b\colon \mathbb{R} \to \mathbb{R}$ defined by $b(y) := y - y^3$ for $y \in \mathbb{R}$ satisfies a one-sided Lipschitz condition with constant $C_b = 1$. In particular, Assumptions 2.6 and 2.7 are satisfied. In the experiment, we choose H to be 0.75, 0.5, 0.25 and 0.10 respectively. For the simulation of the numerical scheme (3.9) we consider the step sizes $h \in \{2^{-7}, 2^{-8}, 2^{-9}, 2^{-10}, 2^{-11}, 2^{-12}\}$ and compare them to a reference solution obtained via a finer step size of $h_{\mathrm{ref}} = 2^{-14}$.

The fractional Brownian motion, as a Gaussian noise, is fully characterized by its mean and co-variance function. For our numerical experiment we first simulate a path of the fractional Brownian motion on the time grid

$$\mathcal{T}^{h_{\mathrm{ref}}} = \{t_0 = 0 < t_1 < \ldots < t_j < \ldots < t_{N_{h_{\mathrm{ref}}}} = 1\} \text{ with } t_j = jh_{\mathrm{ref}},$$

i.e., with the reference step size h_{ref}. Since the increments of a fractional Brownian motion are, in general, not mutually independent we generate the full path at once. To this end we first compute the $N_{h_{\mathrm{ref}}} \times N_{h_{\mathrm{ref}}}$-dimensional covariance matrix C^H whose (i,j)-th entry is defined by

$$C^H_{(i,j)} = \mathbb{E}[B^H(t_i)B^H(t_j)] = \frac{1}{2}\big(|t_i|^{2H} + |t_j|^{2H} - |t_i - t_j|^{2H}\big), \text{ for } i,j \in \{1, \ldots, N_{h_{\mathrm{ref}}}\}.$$

Note that the covariance matrix C^H is positive definite and symmetric. Thus, by an application of the Cholesky decomposition we obtain a lower-triangular matrix $L \in \mathbb{R}^{N_{h_{\mathrm{ref}}} \times N_{h_{\mathrm{ref}}}}$ with $LL^\top = C^H$. Then we draw from the distribution of a sample path of the fractional Brownian motion by taking note of

$$\big(B^H(t_1), \ldots, B^H(t_{N_{h_{\mathrm{ref}}}})\big)^\top \sim LV,$$

where $V = (V_1, \ldots, V_{N_{h_{\mathrm{ref}}}})^\top$ is an $N_{h_{\mathrm{ref}}}$-dimensional standard normally distributed vector. For the simulation with larger step sizes we simply restrict the generated trajectory of B^H to the coarser time grid.
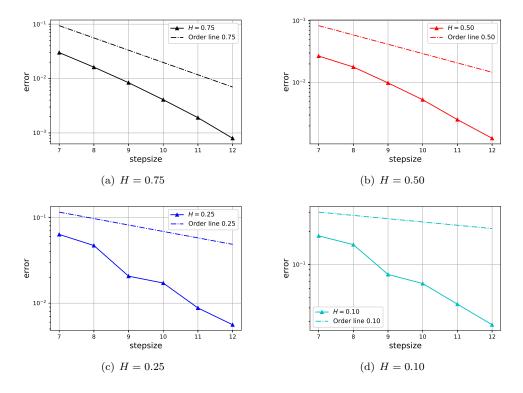
FIGURE 1. Numerical experiments for the implicit Euler method for the RDE (5.1) with $H \in \left\{ \frac{1}{10}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4} \right\}$: step sizes versus pathwise errors.

Once the trajectory of the fractional Brownian motion is simulated we can directly implement the implicit Euler method (3.9) for the approximation of the rough differential equation (5.1). In each step of the method we have to solve a nonlinear equation. In our experiment we accomplished this by an application of Newton's method.

In Figure 1 we show the experimental pathwise errors for the different values of the Hurst parameter H. The plots show the errors against the underlying step size, i.e., the number n on the x-axis indicates the corresponding simulation is based on the step size $h = 2^{-n}$.

First, we observe that all four curves become seemingly less smoother when the value of the Hurst parameter decreases. This is expected from the decreasing smoothness of the driving path B^H. In order to compare the convergence behaviour of the implicit Euler method (3.9) in our experiments with the theoretical result in Theorem 3.9 recall that the path of a fractional Brownian motion is α-Hölder continuous for every $\alpha \in (0, H)$. Thus, the theoretical order of convergence obtained in Theorem 3.9 is essentially equal to H. The respective theoretical orders of convergence are indicated by the order lines in each plot in Figure 1. Comparing this with the actually observed errors in our experiment we conclude that the performance of the implicit Euler method is apparently better in this example than predicted by Theorem 3.9. We also mention that, although Figure 1 only shows the result for just one particular sample path, one essentially obtains the same qualitative behavior of the numerical error for other typical paths of the driving fractional Brownian motion.

TABLE 1.   Numerical values of the pathwise errors and experimental order of convergence (EOC) for implicit Euler method for the RDE (5.1) simulations.

h	$H = 0.75$ error	EOC	$H = 0.50$ error	EOC	$H = 0.25$ error	EOC	$H = 0.10$ error	EOC
0.007813	0.029995		0.026809		0.063482		0.182718	
0.003906	0.016201	0.88	0.017836	0.60	0.047149	0.43	0.151680	0.27
0.001953	0.008391	0.95	0.009873	0.85	0.020681	1.18	0.080940	0.90
0.000977	0.004081	1.04	0.005284	0.90	0.017187	0.27	0.066802	0.28
0.000488	0.001907	1.10	0.002523	1.06	0.008794	0.97	0.043243	0.63
0.000244	0.000798	1.25	0.001261	1.00	0.005601	0.65	0.027993	0.63
Average		1.04		0.88		0.70		0.54

Table 1 contains the numerical values of the computed errors displayed in Figure 1. In addition, we computed the corresponding experimental orders of convergence defined by

$$\text{EOC} = \frac{\log(\text{error}(2^{-i})) - \log(\text{error}(2^{-i+1}))}{\log(2^{-i}) - \log(2^{-i+1})}$$

for $i \in \{7, 8, 9, 10, 11, 12\}$, where the term $\text{error}(2^{-i})$ denotes the error with step size 2^{-i}. Although the experimental orders of convergence are better then predicted by Theorem 3.9 the absolute values of the errors are visibly influenced by the Hurst parameter.

Finally, let us remark that for $H = 0.5$ (standard Brownian motion) the order of convergence observed in our experiment is close to 1. This is in line with standard results for stochastic differential equations with additive noise since in this case the implicit Euler method coincides with a Milstein-type method. Regarding the optimality of the convergence rates we also refer to the discussion in Remark 3.10.

Example 5.2. In the second example we consider the following scalar rough differential equation driven by an additive fractional Brownian motion B^H with Hurst parameter $H = 0.75$,

$$(5.2) \qquad \begin{cases} \mathrm{d}y(t) = -70y(t)\,\mathrm{d}t + \mathrm{d}B^H(t), & t \in (0, 1], \\ y(0) = 2.7. \end{cases}$$
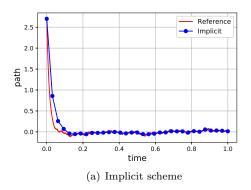
It can be verified that $b \colon \mathbb{R} \to \mathbb{R}$ defined by $\mathbb{R} \ni z \mapsto -70z \in \mathbb{R}$ is a one-sided Lipschitz function with constant -70, while it is globally Lipschitz continuous with constant 70. This discrepancy renders the problem *stiff*. This usually has the effect that the implicit Euler method has a much less restrictive upper step size bound compared to its explicit counterpart. For a more formal introduction of *stiffness* for numerical methods we refer to [HW96].

We can easily illustrate the difference in the stability behavior between the explicit and the implicit Euler method in light of the equation (5.2). First observe that the upper step size limits for the implicit Euler method (3.9) in Theorem 3.6 and Theorem 3.8 are void for every non-positive one-sided Lipschitz constants, since $C_b h < 1$ holds then true for any step size $h = \frac{T}{N_h}$.

Next, let us recall that explicit Euler method is given by

$$(5.3) \qquad y_{j+1}^e = y_j^e + hb(y_j^e) + B_{j+1}^H - B_j^H \quad \text{for } j \in \{0, \cdots, N_h - 1\},$$

with $y_0^e = 2.7$. Observe that the drift function b in (5.2) gives a strong push towards the origin. However, this behavior is only reproduced by the explicit Euler method if the step size is sufficiently

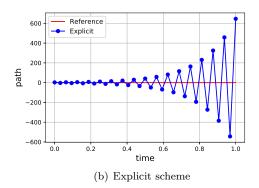(a) Implicit scheme   (b) Explicit scheme

FIGURE 2. Comparison between implicit and explicit schemes of RDE (5.2).

small. To be more precise, the one-step map of the explicit Euler method is estimated by

$$\left| y_{j+1}^e \right| \leq \left| y_j^e + hb(y_j^e) \right| + \left| B_{j+1}^H - B_j^H \right|$$

Thus, the drift part of the explicit Euler method is a contraction if and only if

$$(5.4) \qquad \left| 1 + hb'(y_s) \right| = \left| 1 - h70 \right| \leq 1.$$

Compare further with the linear asymptotic stability of dynamical systems in discrete time, for instance, in [Str18, Chapter 10]. One easily verifies that (5.4) leads to the step size bound $h \leq h_o := \frac{1}{35} < 2^{-5}$. If this bound is violated then the drift part of the explicit Euler method is too negative and typical trajectories of the explicit method start to oscillate.

We illustrate this behavior in Figure 2, where we plot a trajectory with both schemes, the explicit and the implicit Euler methods. Both subfigures (a) and (b) show a reference solution of (5.2) with step size $h_{\text{ref}} = 2^{-14}$ and a further trajectory for each method with the coarser step size $h = 2^{-5}$. Observe that this step size violates the condition (5.4) for the explicit Euler method.

In part (a) we see that the implicit Euler method already gives a rather good approximation of the reference solution for this step size. On the other hand, we observe in part (b) that the trajectory of the explicit method exhibits strong oscillations which are purely artificially induced and, therefore, undesirable.

In particular, this observation is of particular importance if the numerical method is embedded, for instance, in a multilevel Monte Carlo algorithm. Here the effectiveness of the multilevel algorithm depends on the availability of a one-step method which also behaves stable for rather coarse step sizes. For an analysis of the multilevel Monte Carlo algorithm in the context of rough differential equations we refer to [BFRS16].

Example 5.3. We consider the following 2-dim RDE driven by 2d fractional Brownian motion

$$(5.5) \qquad \begin{cases} \mathrm{d}y = (y - |y|^2 y)\,\mathrm{d}y + \sigma_1(y)\,\mathrm{d}B_{H_1}^1 + \sigma_2(y)\,\mathrm{d}B_{H_2}^2, & t \in (0,1], \\ y(0) = [10.0, -10.0]^T, \end{cases}$$

where $[B^{H_1}, B^{H_2}]^T$ is a 2-dim fractional Brownian motion with Hurst parameter $H_1, H_2 \in (0,1)$ in each direction, and $\sigma_1(y) = [\cos(y_2), -0.9 - 10\cos(y_1)]^T$ and $\sigma_2(y) = [\cos(|y|), 0]^T$. From [RS17,

(a) Pathwise errors   (b) Path simulation with $h = 2^{-7}$
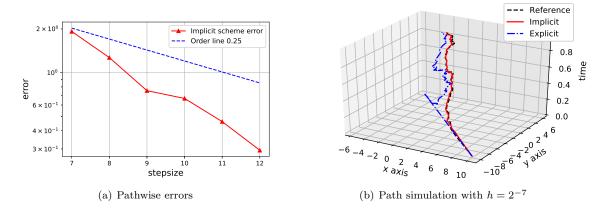
FIGURE 3. Numerical experiment for 2-dim RDE (5.5).

Theorem 4.3], it follows that the equation defines a semiflow and that the assumptions of Theorem 4.17 are satisfied.

We are free to choose the Hurst parameter $H \in (\frac{1}{3}, \frac{1}{2}]$. We pick $H_1 = H_2 = \frac{5}{12}$ in the experiment. We then simulate the solutions via the implicit Milstein scheme with stepsizes $h \in \{2^{-7}, 2^{-8}, 2^{-9}, 2^{-10}, 2^{-11}, 2^{-12}\}$ and compare them to the reference solution obtained via a finer step size of $h_{\text{ref}} = 2^{-14}$.

In Figure 3 (a), we consider the RDE (5.5) and plot the pathwise errors against the underlying step size, i.e., the number n on the x-axis indicates the corresponding simulation is based on the step size $h = 2^{-n}$. The finest step size here is 2^{-12}. In addition, average EOC obtained is 0.38, is larger compared to the expected order of convergence $\frac{3}{q} - 1 < \frac{3 \times 5}{12} - 1 = 0.25$ from Theorem 4.17. In Figure 3 (b), two path simulations is also demonstrated for the performance of explicit and implicit Milstein scheme with stepsize $h = 2^{-7}$. Clearly numerical solutions from the implicit scheme gives a better approximation.

In addition, we also test the corresponding forward and backward Euler scheme of RDE (5.5) with even coarser stepsize $h_o = 2^{-6}$. Forward Euler scheme returns an overflow error, which indicates an explosion of the solution of forward Euler scheme. Though eventually both forward and backward Euler schemes may give the same order of convergence, backward Euler outperforms backward one with coarser stepsizes. This observation is particularly crucial for easing computational burden of large-scaled simulations and thus has practical impact on computational cost.

REFERENCES

[Ben17]   Mikkel Bennedsen. A rough multi-factor model of electricity spot prices. *Energy Economics*, 63:301–313, 2017.

[BFG16]   Christian Bayer, Peter Friz, and Jim Gatheral. Pricing under rough volatility. *Quant. Finance*, 16(6):887–904, 2016.

[BFRS16] Christian Bayer, Peter K. Friz, Sebastian Riedel, and John Schoenmakers. From rough path estimates to multilevel Monte Carlo. *SIAM J. Numer. Anal.*, 54(3):1449–1483, 2016.

[CHJ13]   S. G. Cox, M. Hutzenthaler, and A. Jentzen. Local Lipschitz continuity in the initial value and strong completeness for nonlinear stochastic differential equations. *arXiv:1309.5595*, pages 1–84, 2013.

[Cla87]   D. S. Clark. Short proof of a discrete Gronwall inequality. *Discrete Appl. Math.*, 16(3):279–281, 1987.

[CS17]   Christoph Czichowsky and Walter Schachermayer. Portfolio optimisation beyond semimartingales: shadow prices and fractional Brownian motion. *Ann. Appl. Probab.*, 27(3):1414–1451, 2017.

[Dav07]   Alexander M. Davie. Differential equations driven by rough paths: an approach via discrete approximation. *Appl. Math. Res. Express. AMRX*, (2):Art. ID abm009, 40, 2007.

[DNT12]   A. Deya, A. Neuenkirch, and S. Tindel. A Milstein-type scheme without Lévy area terms for SDEs driven by fractional Brownian motion. *Ann. Inst. Henri Poincaré Probab. Stat.*, 48(2):518–550, 2012.

[EER19]   Omar El Euch and Mathieu Rosenbaum. The characteristic function of rough Heston models. *Math. Finance*, 29(1):3–38, 2019.

[EKKL19] M. Eisenmann, M. Kovács, R. Kruse, and S. Larsson. On a randomized backward Euler method for nonlinear evolution equations with time-irregular coefficients. *Found. Comput. Math.*, 2019. (Online first).

[Emm04]   E. Emmrich. *Gewöhnliche und Operator-Differentialgleichungen*. Vieweg, Wiesbaden, 2004.

[FGGR16] P. K. Friz, B. Gess, A. Gulisashvili, and S. Riedel. The Jain-Monrad criterion for rough paths and applications to random Fourier series and non-Markovian Hörmander theory. *Ann. Probab.*, 44(1):684–738, 2016.

[FH14]   P. K. Friz and M. Hairer. *A Course on Rough Paths: With an Introduction to Regularity Structures*, volume XIV of *Universitext*. Springer, Berlin, 2014.

[FR14]   P. K. Friz and S. Riedel. Convergence rates for the full Gaussian rough paths. *Ann. Inst. Henri Poincaré Probab. Stat.*, 50(1):154–194, 2014.

[FV10a]   P. K. Friz and N. B. Victoir. *Multidimensional Stochastic Processes as Rough Paths*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Theory and applications.

[FV10b]   Peter K. Friz and Nicolas B. Victoir. Differential equations driven by Gaussian signals. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(2):369–413, 2010.

[GJR18]   Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quant. Finance*, 18(6):933–949, 2018.

[Gro19]   T. H. Gronwall. Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Ann. of Math. Second Series*, 20(4):292–296, 1919.

[Gua06]   Paolo Guasoni. No arbitrage under transaction costs, with fractional Brownian motion and beyond. *Math. Finance*, 16(3):569–582, 2006.

[Gub10]   M. Gubinelli. Ramification of rough paths. *J. Differential Equations*, 248(4):693–721, 2010.

[Hal80]   J. K. Hale. *Ordinary Differential Equations*. Robert E. Krieger Publishing Company, Inc., Malabar, Florida, 1980.

[HHW18]   Jialin Hong, Chuying Huang, and Xu Wang. Symplectic Runge-Kutta methods for Hamiltonian systems driven by Gaussian rough paths. *Appl. Numer. Math.*, 129:120–136, 2018.

[Hig00]   Desmond J. Higham. Mean-square and asymptotic stability of the stochastic theta method. *SIAM J. Numer. Anal.*, 38(3):753–769, 2000.

[HW96]   E. Hairer and G. Wanner. *Solving ordinary differential equations. II*, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 1996. Stiff and differential-algebraic problems.

[KP92]   P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[KPS91]   P. E. Kloeden, E. Platen, and H. Schurz. The numerical solution of nonlinear stochastic dynamical systems: a brief introduction. *Internat. J. Bifur. Chaos Appl. Sci. Engrg.*, 1(2):277–286, 1991.

[KW17]   R. Kruse and Y. Wu. Error analysis of randomized Runge–Kutta methods for differential equations with time-irregular coefficients. *Comput. Methods Appl. Math.*, 17(3):479–498, 2017.

[LAE08]   Tiejun Li, Assyr Abdulle, and Weinan E. Effectiveness of implicit methods for stiff stochastic differential equations. *Commun. Comput. Phys.*, 3(2):295–307, 2008.

[LCL07]   Terry J. Lyons, Michael Caruana, and Thierry Lévy. *Differential equations driven by rough paths*, volume 1908 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6–24, 2004, With an introduction concerning the Summer School by Jean Picard.

[Lej09]   Antoine Lejay. On rough differential equations. *Electron. J. Probab.*, 14:no. 12, 341–364, 2009.

[Lej12]   Antoine Lejay. Global solutions to rough differential equations with unbounded vector fields. In *Séminaire de Probabilités XLIV*, volume 2046 of *Lecture Notes in Math.*, pages 215–246. Springer, Heidelberg, 2012.

[Lyo98]   Terry J. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana*, 14(2):215–310, 1998.

[Mil95]   G. N. Milstein. *Numerical Integration of Stochastic Differential Equations*, volume 313 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1995. Translated and revised from the 1988 Russian original.

[MT04]   G. N. Milstein and M. V. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Scientific Computation. Springer-Verlag, Berlin, 2004.

[OR00]   J. M. Ortega and W. C. Rheinboldt. *Iterative solution of nonlinear equations in several variables*, volume 30 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000. Reprint of the 1970 original.

[RR20]   Martin Redmann and Sebastian Riedel. Runge–kutta methods for rough differential equations. *arXiv:2003.12626*, 2020.

[RS17]   S. Riedel and M. Scheutzow. Rough differential equations with unbounded drift term. *J. Differential Equations*, 262(1):283–312, 2017.

[SH96]   A. M. Stuart and A. R. Humphries. *Dynamical Systems and Numerical Analysis*, volume 2 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 1996.

[Str18]   S. H. Strogatz. *Nonlinear Dynamics and Chaos: with Applications to Physics, Biology, Chemistry, and Engineering.*, volume 68. CRC Press, 2018.

[You36]   Laurence C. Young. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Math.*, 67(1):251–282, 1936.

Sebastian Riedel, Institut für Mathematik, Technische Universität Berlin, Germany and Weierstrass-Institut, Berlin, Germany

   E-mail address: riedel@math.tu-berlin.de

Yue Wu, Mathematical Institute, University of Oxford, Oxford, UK and Alan Turing Institute, London, UK

   E-mail address: yue.wu@maths.ox.ac.uk