Chapter 1

State-of-the-art numerical schemes for solving rough differential equations

M. Redmann and J. Werner

In this chapter, we give an introduction to rough paths. Subsequently, we consider various approaches to solve rough differential equations numerically, discuss advantages as well as drawbacks of each individual scheme and compare their performance in numerical experiments.

Contents

1. Introduction

The use of stochastic differential equations (SDEs) as a sophisticated mathematical tool for modeling real life applications has grown in popularity since involving uncertainties is more realistic. However, often it is not possible to determine SDE solutions in an analytically closed form. Therefore, numerical schemes are important tools to approximate their solutions. There are different classes of numerical methods to solve SDEs, the most popular ones are Taylor schemes namely the Euler-Maruyama and the Milstein scheme but also Runge-Kutta methods, see [1–5], and even some approaches from Lie theory see [6–8] have been proposed. In practise the implementation of all these techniques are limited to low strong approximation orders, since high order schemes suffer from the inability to efficiently calculate the involved iterated integrals. This makes the implementation of high order schemes efficient only in special cases.

A deterministic (or pathwise) approach to SDEs is given by the rough path theory invented by Lyons. In this case, the driver is a more general Hölder continuous function instead of a continuous semimartingale. Rough differential equations (RDEs) are driven by objects which can be interpreted as vectors of iterated integrals and therefore make higher order numerical methods interesting again. While the most studied numerical schemes available for RDEs are of Taylor-type [9, 10], there is few literature on Runge-Kutta-type schemes [11]. Many of these approaches rely on the availability of iterated integrals which are usually not known analytically. Therefore, methods using only path information of the driver have been considered [11–13] for a simpler implementation. However, they are restricted to low orders of convergence. Also the Lie theory approach, namely the Log-ODE method [14–17] is studied extensively which builds upon Taylor schemes involving iterated integrals. Unfortunately, Taylor methods rely on nested derivatives, which are expensive to compute making them inefficient in, e.g., large-scale settings. We aim to modify the Lie theory ansatz and exploit the advantage of Runge-Kutta schemes in this context which have lower computation cost since they are derivative free. The contribution of this chapter is to introduce a new Runge-Kutta-Log-ODE method which is based on a new Runge-Kutta scheme that is an extension of a Runge-Kutta ansatz from SDE theory [1]. Both new techniques have a considerably reduced computational effort compared to their well-known Taylor counterparts and explicitly depend on second order iterated integrals. In particular, this explicit dependence is an advantage compared to the class of methods used in [11]. In this work, the focus is rather on applying the new approaches and testing their theoretical properties in numerical experiments. Therefore, it can be seen as a good starting point for further theoretical studies of new methods with high potential.

This chapter is structured as follows. In Section 2, we introduce the basics of rough path theory and describe RDEs. Section 3 contains an overview on various numerical schemes including our new proposed approach. We conclude with Section 4 conducting numerical experiments for high-dimensional RDEs and providing a discussion on the performance of each method.

2. Foundation of rough path theory

First we introduce essentials of the rough path theory like the tensor algebra and signatures.

Definition 1 ([16] Definition A.1). We say that $T(\mathbb{R}^k) := \bigoplus_{i=0}^{\infty}(\mathbb{R}^k)^{\otimes i}$ is the tensor algebra of $\mathbb{R}^k$ and $T((\mathbb{R}^k)) = \{\mathbf{a} = (a_0, a_1, \dots) : a_n \in (\mathbb{R}^k)^{\otimes n}\ \forall n \geq 0\}$ is the set of formal series of tensors of $\mathbb{R}^k$. Similarly, we define the truncated tensor algebra $T^N(\mathbb{R}^k) := \mathbb{R} \oplus \mathbb{R}^k \oplus (\mathbb{R}^k)^{\otimes 2} \oplus \cdots \oplus (\mathbb{R}^k)^{\otimes N}$ for $N \in \mathbb{N}$. Moreover, $T(\mathbb{R}^k)$, $T((\mathbb{R}^k))$ and $T^N(\mathbb{R}^k)$ can be endowed with the operations of addition and multiplication.
Given $\mathbf{a} = (a_0, a_1, \dots)$ and $\mathbf{b} = (b_0, b_1, \dots)$, we have

$$\begin{aligned} \mathbf{a} + \mathbf{b} &= (a_0 + b_0, a_1 + b_1, \dots), \\ \mathbf{a} \otimes \mathbf{b} &= (c_0, c_1, \dots), \end{aligned} \tag{1}$$

where for $n \geq 0$ the n-th term $c_n \in (\mathbb{R}^k)^{\otimes n}$ can be written using the usual tensor product as

$$c_n := \sum_{i=0}^{n} a_i \otimes b_{n-i}.$$

In rough path theory, we consider two parameter functions with values in $T^N(\mathbb{R}^k)$. Instead of choosing two arbitrary parameters in $[0, T]$ it is often useful to order the parameters via the simplex $\Delta_T := \{(s,t) \in [0,T]^2 : s < t\}$. We follow this approach notation-wise.

Definition 2 ([16] Definition A.2). The signature $S_{\cdot,\cdot}(X) : \Delta_T \to T((\mathbb{R}^k))$ of a path $X \colon [0,T] \to \mathbb{R}^k$ of bounded variation over the interval $[s,t]$ is defined as the following collection of iterated (Riemann-Stieltjes) integrals:

$$S_{s,t}(X) := \left(1, X^{(1)}_{s,t}, X^{(2)}_{s,t}, \cdots\right) \in T((\mathbb{R}^k)),$$

where for $n \geq 1$,

$$X^{(n)}_{s,t} := \underset{s<u_1<\cdots<u_n<t}{\int \cdots \int} \mathrm{d}X_{u_1} \otimes \cdots \otimes \mathrm{d}X_{u_n} \in (\mathbb{R}^k)^{\otimes n}.$$

Similarly, we can define the depth-N (or truncated) signature of the path X on $[s,t]$ as

$$S^N_{s,t}(X) := \left(1, X^{(1)}_{s,t}, \cdots, X^{(N)}_{s,t}\right) \in T^N(\mathbb{R}^k),$$

Definition 3. We define $\pi_n \colon T^N(\mathbb{R}^k) \to (\mathbb{R}^k)^{\otimes n}$ as the projection map onto $(\mathbb{R}^k)^{\otimes n}$ for $n = 0, 1, \dots, N$.

Next we introduce a suitable norm for signatures and later also rough paths.

Definition 4. Let $\mathbf{X}\colon \Delta_T \to T^{\lfloor p \rfloor}(\mathbb{R}^k)$ where $\lfloor\cdot\rfloor$ denotes the floor function. For $p \geq 1$ we introduce the p-variation norm

$$\|\mathbf{X}\|_{p\text{-var}} := \max_{1\leq n\leq \lfloor p\rfloor} \sup_{\mathcal{D}} \Big(\sum_{t_i\in\mathcal{D}} \|\pi_n(\mathbf{X}_{t_i,t_{i+1}})\|^{\frac{p}{n}} \Big)^{\frac{n}{p}}$$

and the induced p-variation metric between to continuous paths $\mathbf{Z}^1$ and $\mathbf{Z}^2$ with values in $T^{\lfloor p \rfloor}(\mathbb{R}^k)$ as

$$d_p(\mathbf{Z}^1,\mathbf{Z}^2) := \max_{1\leq n\leq \lfloor p\rfloor} \sup_{\mathcal{D}} \Big(\sum_{t_i\in\mathcal{D}} \big\|\pi_n(\mathbf{Z}^1_{t_i,t_{i+1}}) - \pi_n(\mathbf{Z}^2_{t_i,t_{i+1}})\big\|^{\frac{p}{n}} \Big)^{\frac{n}{p}},$$

where the supremum is taken over all partitions $\mathcal{D}$ of $[0,T]$ and the norms $\|\cdot\|$ must satisfy

$$\|a\otimes b\| \leq C\|a\|\|b\|,$$

for $a \in (\mathbb{R}^k)^{\otimes n}$, $b \in (\mathbb{R}^k)^{\otimes m}$ and a constant $C \geq 0$. For example, we can take $\|\cdot\|$ to be the projective or injective tensor norms (see Propositions 2.1 and 3.1 in [18]). Additionally, we define a metric for the two cases $p = 0$ and $p = \infty$ the following way:

$$d_{\infty;[0,T]}(\mathbf{Z}^1,\mathbf{Z}^2) := \sup_{t\in[0,T]} d(\mathbf{Z}^1_{0,t},\mathbf{Z}^2_{0,t})$$
$$d_{0;[0,T]}(\mathbf{Z}^1,\mathbf{Z}^2) := \sup_{0\leq s<t\leq T} d(\mathbf{Z}^1_{s,t},\mathbf{Z}^2_{s,t}),$$

where d is the so-called Carnot–Caratheodory metric (see Theorem 7.32 [10]).

Next, we consider important objects of rough path theory, namely geometric p-rough paths.

Definition 5 ([16] Theorem B.1). For $p \geq 1$, we say $\mathbf{X}\colon \Delta_T \to T^{\lfloor p \rfloor}(\mathbb{R}^k)$ is a geometric p-rough path if $\mathbf{X}$ is a continuous path in the tensor algebra $T^{\lfloor p \rfloor}(\mathbb{R}^n)$ and there exists a sequence (x^n) of continuous finite variation paths $x^n : [0,T] \to \mathbb{R}^k$ whose truncated signatures converge to $\mathbf{X}$ in the p variation metric:

$$d_p\big(S^{\lfloor p \rfloor}(x^n),\mathbf{X}\big) \to 0, \tag{2}$$

as $n \to \infty$.

The following identity, known as Chen's relation, tells us precisely how to "patch together" rough paths over adjacent intervals.

Lemma 1. *If* $\mathbf{X}$ *is a geometric* p*-rough path, then*

$$\mathbf{X}_{s,t} = \mathbf{X}_{s,u} \otimes \mathbf{X}_{u,t} \tag{3}$$

holds for $0 \leq s < u < t \leq T$.

Proof. In [10, Theorem 7.11] this statement is proven for $S^{\lfloor p \rfloor}(x^n)$ instead of $\mathbf{X}$ (see (2)). A careful passage to the limit concludes the proof. □

If not stated otherwise from now on we will always assume $p \geq 1$.

Example 1. Assume that $X = W_{\cdot}(\omega)\colon [0,T] \to \mathbb{R}^k$ is a path of a standard Brownian motion W. Since the paths of W are almost surely of finite p-variation for $p > 2$, we need to make sense of $\mathbb{W}$ in $\mathbf{X} = (W, \mathbb{W})$ taking values in $T^2(\mathbb{R}^k)$. The two most common approaches are to define $\mathbb{W}$ via Itô or Stratonovich integrals. Notice that the construction of this second order information is not path-wise, although rough analysis is a path-wise ansatz. While the approach with Itô integrals leads to a non geometric rough path, the approach via Stratonovich integrals indeed results in a random geometric p-rough path $\mathbf{X} = (W, \mathbb{W}^{Strat})$, where for a path $W_{\cdot}(\omega)\colon [0,T] \to \mathbb{R}^k$

$$\mathbb{W}^{Strat}_{s,t}(\omega) = \left(\int_s^t W_{s,u} \otimes \circ \mathrm{d}W_u \right)(\omega)$$

is based on a Stratonovich integral. A potential smooth approximation in (2) is the well-known Wong-Zakai approximation [19].

Since we now have the definition of a rough path as a potential driver, the aim is to consider rough differential equations (RDEs) and to construct a solution concept.
We start by introducing a smoothness notion for the vector fields.

Definition 6 ([10] Definition 10.2). A map $f\colon E \to F$ between two normed spaces E, F is called γ-Lipschitz, in symbols $f \in \mathrm{Lip}(\gamma)$, if f is bounded with $\lfloor \gamma \rfloor$ bounded Fréchet derivatives, where the last Fréchet derivative $D^{\lfloor \gamma \rfloor} f$ is Hölder continuous with exponent $\gamma - \lfloor \gamma \rfloor$. Then, the following norm is finite

$$\| \cdot \|_{\mathrm{Lip}(\gamma)} := \max_{0 \leq k \leq \lfloor \gamma \rfloor} \|D^k f\|_\infty \vee \|D^{\lfloor \gamma \rfloor} f\|_{(\gamma - \lfloor \gamma \rfloor)\text{-Höl}}.$$

There are at least two concepts to define solutions of RDEs, which in our framework are both equivalent. We start with the classical approach of Lyons [20], which is similar to the SDE case. The idea is to define a rough integral and subsequently introduce a differential equation involving such an integral.

Definition 7 ([10] Definition 10.44). Let $\mathbf{X}\colon \Delta_T \to T^{\lfloor p \rfloor}(\mathbb{R}^k)$ be a geometric p-rough path [a], and $f = (f_1, \dots f_k)$ be a collection of maps $f_j\colon \mathbb{R}^{n_1} \to \mathbb{R}^k$. We say that $\mathbf{Z}\colon \Delta_T \to T^{\lfloor p \rfloor}(\mathbb{R}^{n_1})$ is a rough integral of f along $\mathbf{X}$, if there exists a sequence $(x^n)_n$ with $x^n\colon [0,T] \to \mathbb{R}^k$ and x^n is of bounded variation such that

$$\forall n : x_0^n = X_0$$

$$\lim_{n\to\infty} d_{0;[0,T]}(S^{\lfloor p \rfloor}(x^n), \mathbf{X}) = 0$$

$$\sup_n \|S^{\lfloor p \rfloor}(x^n)\|_{p\text{-var}} < \infty$$

where $X = \pi_1(\mathbf{X})$ and

$$\lim_{n\to\infty} d_\infty\big(S^{\lfloor p \rfloor}\big(\int f(x_u^n)\,\mathrm{d}x_u^n\big), \mathbf{Z}\big) = 0.$$

The following results gives conditions under which the rough integral is well defined.

Theorem 1 ([10] Theorem 10.47). *Assume that*

- *for* $f = (f_1, \dots, f_k)$*,* $f_i \in \mathrm{Lip}(\gamma - 1)$ *with* $i = 1, ..., k$ *and* $\gamma > p \geq 1$*;*
- $\mathbf{X}\colon \Delta_T \to T^{\lfloor p \rfloor}(\mathbb{R}^k)$ *is a geometric* p*-rough path.*

Then, for all $s < t \in [0,T]$*, there exists a unique rough-path integral of* f *along* $\mathbf{X}$*. The indefinite integral* $\int f(X)\,\mathrm{d}\mathbf{X}$ *is a geometric rough path: there exists a constant* C *depending only on* p *and* γ *such that for all* $s < t$ *in* $[0,T]$*,*

$$\Big\| \int f(X)\,\mathrm{d}\mathbf{X} \Big\|_{p\text{-}var;[s,t]} \leq C\|f\|_{\mathrm{Lip}(\gamma-1)}\big(\|\mathbf{X}\|_{p\text{-}var;[s,t]} \vee \|\mathbf{X}\|^p_{p\text{-}var;[s,t]}\big)$$

with $X = \pi_1(\mathbf{X})$*.*

Below, we write $\mathbf{X}_t$ instead of $\mathbf{X}_{0,t}$ for simplicity of the notation. Now, a solution to the RDE

$$\begin{aligned} \mathrm{d}Y_t &= f(Y_t)\,\mathrm{d}\mathbf{X}_t, \\ Y_0 &= y_0, \end{aligned} \tag{4}$$

[a] This definition also holds in the case of weak geometric p-rough paths being a more general concept.

can then be defined via the integral equation

$$Y_t = y_0 + \int_0^t f(Y_s)\,\mathrm{d}\mathbf{X}_s,$$

where it turns out that $\int_0^t f(Y_s)\,\mathrm{d}\mathbf{X}_s$ is well-defined as a rough integral under suitable regularity conditions on f. Notice that we actually need a more general notion of integrals than in Definition 7 in order to make sense of RDEs. However, we omit this extension for a better readability of this chapter. The other approach, which we will use in the rest of the paper, is to define a solution as an appropriate limit of solutions to Stieltjes differential equations.

Definition 8 ([10] Definition 10.17). Let $\mathbf{X}\colon \Delta_T \to T^{\lfloor p \rfloor}(\mathbb{R}^k)$ be a geometric p-rough path. We say that $Y \in C([0,T], \mathbb{R}^{n_1})$ is a solution to the RDE

$$\begin{aligned} \mathrm{d}Y_t &= f(Y_t)\,\mathrm{d}\mathbf{X}_t, \\ Y_0 &= y_0, \end{aligned}$$

if there exists a sequence of $(x^n)_n$ of bounded variations functions such that $\lim_{n\to\infty} d_{0;[0,T]}(S_{\lfloor p \rfloor}(x^n), \mathbf{X}) = 0$ and $\sup_n \|S_{\lfloor p \rfloor}(x^n)\|_{p\text{-var}} < \infty$ hold as well as solutions Y^n to the Stieltjes differential equations

$$\begin{aligned} \mathrm{d}Y_t^n &= f(Y_t^n)\,\mathrm{d}x_t^n, \\ Y_0^n &= y_0, \end{aligned}$$

exist such that

$$Y^n \to Y \text{ uniformly on } [0,T] \text{ as } n \to \infty.$$

Remark 1. Both concepts of solutions are equivalent in our framework (see [10] Remark 10.19). While the second concept might be more intuitive the approach of Lyons can immediately be generalized for more general rough paths $\mathbf{X}$.

At last we give an existence and uniqueness result on the solution of RDEs.

Theorem 2 ([10] Theorem 10.26). *Assume that*

- $f = (f_1, ..., f_k)$ *is a collection of* $\mathrm{Lip}(\gamma)$*-vector fields* $f_j\colon \mathbb{R}^n \to \mathbb{R}^k$ *for* $\gamma > p \geq 1$*;*
- $\mathbf{X}\colon \Delta_T \to T^{\lfloor p \rfloor}(\mathbb{R}^k)$ *is a geometric* p*-rough path,*
- $y_0 \in \mathbb{R}^n$ *thought of as initial conditions at time zero.*

Then, there exists a unique solution $Y\colon [0,T] \to \mathbb{R}^n$ *to the RDE*

$$\begin{aligned} \mathrm{d}Y_t &= f(Y_t)\,\mathrm{d}\mathbf{X}_t, \\ Y_0 &= y_0. \end{aligned} \tag{5}$$

A main advantage of rough path theory is that $Y(y_0, f, \mathbf{X})$ under the assumptions is locally Lipschitz continuous in the initial value, the vector field and the driver. This fact is used to prove the uniqueness of a solution of (5).

Remark 2. It is possible to recover the solution of (2) as rough path $\mathbf{Y}$ itself. In general we only consider the first level $\pi_1(\mathbf{Y}) = Y$ of the solution.

3. Overview of numerical schemes

In this section, we present schemes that can be used in order to compute numerical solutions to RDEs. Let $\mathbf{X}\colon \Delta_T \to T^{\lfloor p \rfloor}(\mathbb{R}^k)$ be a geometric p-rough path and $f = (f_1, \ldots, f_k)\colon \mathbb{R}^n \to \mathbb{R}^{n\times k}$. We aim to approximate a function $Y\colon [0,T] \to \mathbb{R}^n$ satisfying

$$\begin{aligned} \mathrm{d}Y_t &= f(Y_t)\,\mathrm{d}\mathbf{X}_t = \sum_{i=1}^{k} f_i(Y_t)\mathrm{d}\mathbf{X}_t^i, \\ Y_0 &= y_0. \end{aligned} \tag{6}$$

Additionally, we assume that $f_i \in \mathrm{Lip}(\gamma)$ for $\gamma > p$ and $i = 1, \ldots, k$ to ensure the existence of a unique solution according to Theorem 2.

3.1. *Taylor schemes*

A special case of this class of methods was first introduced in [9] and is extensively studied in [10] in full generality[b]. Notation-wise we follow the approach of [16].

We start by motivating the origin of the Taylor schemes. Let $X\colon [0,T] \to \mathbb{R}$ be continuously differentiable and assume that $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is smooth. We denote D_f as the Jacobian of f. Now, we exploit a first order Taylor expansion of f, i.e., $f(Y_s) \approx f(Y_a) + D_f(Y_a)(Y_s - Y_a)$ in order to find a Taylor approximation of the solution of (6) around the point a with $0 < t - a << 1$.

[b]What we call Taylor schemes is there studied under the name of (step-N) Euler scheme.

In this context, we neglect terms of higher order and get

$$\begin{aligned}
Y_t &= Y_a + \int_a^t f(Y_s)\mathrm{d}X_s \\
&\approx Y_a + \int_a^t f(Y_a) + D_f(Y_a)(Y_s - Y_a)\mathrm{d}X_s \\
&= Y_a + \int_a^t \left(f(Y_a) + D_f(Y_a)\int_a^s f(Y_u)\mathrm{d}X_u\right)\mathrm{d}X_s \\
&\approx Y_a + \int_a^t \left(f(Y_a) + D_f(Y_a)f(Y_a)\int_a^s \mathrm{d}X_u\right)\mathrm{d}X_s \\
&= Y_a + f(Y_a)\int_a^t \mathrm{d}X_s + D_f(Y_a)f(Y_a)\int_a^t\int_a^s \mathrm{d}X_u\mathrm{d}X_s \\
&= Y_a + f(Y_a)\pi_1(S_{a,t}(X)) + D_f(Y_a)f(Y_a)\pi_2(S_{a,t}(X)).
\end{aligned} \tag{7}$$

This gives a second order Taylor like expansion of the solution which in the case of SDEs is known as the Milstein scheme. The levels of the signature play a crucial role as polynomials on paths. Higher order Taylor expansions result in expressions using higher order signature terms.

Indeed this connection can be made rigorous and leads to a precise definition of the Taylor schemes. We start with describing the role of the vector fields which are usually known as elementary differentials.

Definition 9 ([16] Definition A.6). For $f\colon \mathbb{R}^n \to \mathbb{R}^{n\times k}$, we define $f^{\circ m}\colon \mathbb{R}^n \to L((\mathbb{R}^k)^{\otimes m}, \mathbb{R}^n)$ recursively for $m \in \mathbb{N}$ by

$$\begin{aligned}
f^{\circ 0}(y) &:= y, \\
f^{\circ 1}(y) &:= f(y), \\
f^{\circ m+1}(y) &:= D(f^{\circ m})(y)f(y),
\end{aligned}$$

for $y \in \mathbb{R}^n$, where $D(f^{\circ m})$ denotes the Fréchet derivative of $f^{\circ m}$.

Now, we can describe the Taylor schemes for more general drivers and orders covering the motivation in (7).

Definition 10 ([16] Definition A.7). Let $\mathbf{X}\colon \Delta_T \to T^{\lfloor p\rfloor}(\mathbb{R}^k)$ be a geometric p-rough path, $f \in C^{N-1}(\mathbb{R}^n)$ and $N = \lfloor p\rfloor$. The Taylor operator and the associated RDE approximation are given by

$$\mathrm{Taylor}(Y_s, f, \mathbf{X}_{s,t}) := \sum_{k=0}^{N} f^{\circ k}(Y_s)\pi_k(\mathbf{X}_{s,t}) \approx Y_t. \tag{8}$$

This local approximation is repeated m times on a partition $D = \{0 = t_0 < t_1 < \cdots < t_m = T\}$. This motivates to define the numerical solution of the N-th level Taylor method via

$$y_i^{N\text{-Tay}} = \text{Taylor}(y_{i-1}^{N\text{-Tay}}, f, \mathbf{X}_{t_i,t_{i+1}}), \quad (i = 1, \ldots, m),$$

where $y_i^{N\text{-Tay}} \approx Y_{t_i}$ and $y_0^{N\text{-Tay}} = y_0$.

Example 2. The reader might be familiar with these type of Taylor schemes from SDE theory. The strong Taylor approximations in [2] are very similar. Especially the case $N = 1$ is the Euler-Maruyama scheme for rough paths and the case $N = 2$ is the Milstein scheme for rough paths.

We continue with results on convergence and corresponding rates. Below, we make use of the Euclidean norm which we denote by $|\cdot|$ from now on.

Theorem 3 ([10] Theorem 10.30). *Assume that*

- $\mathbf{X}\colon \Delta_T \to T^{\lfloor p \rfloor}(\mathbb{R}^k)$ *is a geometric p-rough path,*
- $f = (f_1, ..., f_k)$ *is a collection of* Lip(γ)*-vector fields on* $\mathbb{R}^n$ *for* $\gamma > p \geq 1$,
- $D = \{0 = t_0 < t_1 < \cdots < t_m = T\}$ *is a fixed partition of* $[0, T]$, *where* $h = \max_i t_{i+1} - t_i$.

Set $N := \lfloor \gamma \rfloor \geq \lfloor p \rfloor$. *Then, there exists a constant* $C = C(p, \gamma) > 0$, *so that*

$$|Y_T - y_m^{N\text{-}Tay}| \leq Ch^\alpha,$$

where $\alpha = \frac{(N+1)}{p} - 1$ *and h is sufficiently small.*

Remark 3. These convergence rates are worst case rates and in some special cases better rates can arise. For example choose the random rough path $\mathbf{X} = (W, \mathbb{W}^{Strat})$ being the Stratonovich lift of a standard Brownian motion W. In this case, Theorem 3 gives a convergence rate of $\alpha = \frac{2+1}{2+\varepsilon_1} - 1 = 0.5 - \varepsilon_2$ for some arbitrary small $\varepsilon_1, \varepsilon_2 > 0$ and almost all paths of W but from [21] we know that the Milstein scheme converges a.s. pathwise with order $1 - \varepsilon$.

3.2. *Runge-Kutta schemes*

In the theory of ordinary differential equations Runge-Kutta schemes are preferred over the usual Taylor schemes. The reason for this is that the

elementary differentials (see Definition 9) are expensive to compute from a numerical perspective. The advantage of Runge-Kutta methods is that they are derivative free.
In [11], a Runge-Kutta approach is presented that discretizes very general RDEs with non-geometric drivers and it is shown that one can achieve an arbitrary good rate analogous to Theorem 3 using a proper choice of the coefficients within the numerical scheme (and given enough regularity of the vector fields). Although this Runge-Kutta ansatz works perfectly in theory, it has the disadvantage that the associated coefficients depend implicitly on iterated integrals and these dependencies are hard to identify in general. In order to overcome that the method is hard to implement, [11] presents another approach in case the driver is geometric. In detail, the geometric p-rough path is discretized using the lift of a piecewise linear approximation and Runge-Kutta schemes are used to solve the arising ODEs. This leads to schemes involving path information only. In contrast to the approach in [11], we want to introduce Runge-Kutta schemes for RDEs with explicit second order information. To the best knowledge of the authors, there exist no Runge-Kutta scheme for RDEs showing the dependence on higher levels explicitly. This means that we establish a higher-order scheme that is implementable. To do so, we exploit a Runge-Kutta scheme with explicit second order information designed for SDEs [1] and transfer it to the much more general case of RDEs. We investigate the scheme numerically including numerical evidence for the convergence of the scheme and an experiment on the order of convergence. The theoretical analysis is beyond the scope of this paper.

We start with defining a Runge-Kutta operator which shall replace the Taylor operator in (8). The definition of the Runge-Kutta operator is based on Section 6.1 of [1]. We specify the method by choosing fixed coefficients for the Runge-Kutta operator, which fulfill certain order conditions in the SDE case.

Definition 11. Let $\mathbf{X}\colon \Delta_T \to T^{\lfloor p \rfloor}(\mathbb{R}^k)$ be a geometric p-rough path with $p < 3$. Then, the Runge-Kutta operator and the associated RDE approximation are defined by

$$\begin{aligned} Y_t &\approx \text{Runge-Kutta}(Y_s, f, \mathbf{X}_{s,t}) \qquad (9) \\ &:= Y_s + \sum_{l=1}^{k} f_l(Y_s)\pi_1(\mathbf{X}_{s,t})_l + \frac{\sqrt{t-s}}{2}\sum_{l=1}^{k}\Big(f_l(Z_l^1) - f_l(Z_l^2)\Big), \end{aligned}$$

where $\pi_1(\mathbf{X}_{s,t})_l$ is the l-th element of $\pi_1(\mathbf{X}_{s,t})$ and

$$Z_l^1 = Y_s + \frac{1}{\sqrt{t-s}} \sum_{j=1}^{k} f_j(Y_s)\pi_2(\mathbf{X}_{s,t})_{j,l},$$

$$Z_l^2 = Y_s - \frac{1}{\sqrt{t-s}} \sum_{j=1}^{k} f_j(Y_s)\pi_2(\mathbf{X}_{s,t})_{j,l}$$

with $\pi_2(\mathbf{X}) \in (\mathbb{R}^k)^{\otimes 2} \simeq \mathbb{R}^{k\times k}$ and $\pi_2(\mathbf{X})_{j,l}$ is the element in the j-th row and in the l-th column.
This local approximation is repeated m times on a partition $D = \{0 = t_0 < t_1 < \cdots < t_m = T\}$. We define the numerical solution of the Runge-Kutta method via

$$y_i^{\mathrm{RK}} = \text{Runge-Kutta}(y_{i-1}^{\mathrm{RK}}, f, \mathbf{X}_{t_i,t_{i+1}}), \quad (i = 1, \ldots, m),$$

where $y_0^{\mathrm{RK}} = y_0$. The intuition is that $y_i^{\mathrm{RK}} \approx Y_{t_i}$.

Remark 4. While the Taylor scheme is a class of numerical methods depending on $N \in \mathbb{N}$ in Definition 11 we give the Runge-Kutta method for the concrete case $N = 2$. In [1] the Runge-Kutta methods for SDEs are given for $N \in \mathbb{N}$. In future publications the authors plan to investigate this case for RDEs too. The $N = 1$ Runge-Kutta scheme coincides with the 1-level Taylor method.

Remark 5. The motivation of the Taylor method (7) is based on the Taylor series. Similar one can motivate the Runge-Kutta operator by an expansion of the solution by the so-called B-series or Butcher series.

3.3. *Log-ODE method*

The method was introduced in [14], analyzed in [15] and got recent attention in [17]. Again we mainly follow the notation of [16]. The Log-ODE method is an approach from Lie theory and therefore has the advantage of respecting the geometry of the problem. If the solution of (6) lies in a certain manifold than it is possible to construct an approximation via the Log-ODE method which lies in the same manifold.

We start with introducing the basic Lie theory of rough paths.

Definition 12 ([16] Definition A.3). For $\mathbf{a} = (a_0, a_1, \ldots) \in T((\mathbb{R}^k))$ with $a_0 > 0$, define $\log(\mathbf{a})$ to be the element of $T((\mathbb{R}^k))$ given by the following series:

$$\log(\mathbf{a}) := \log(a_0) + \sum_{l=1}^{\infty} \frac{(-1)^l}{l} \big(\mathbf{1} - \frac{\mathbf{a}}{a_0}\big)^{\otimes l}, \tag{10}$$

where $\mathbf{1} = (1, 0, 0, \ldots)$ is the unit element of $T((\mathbb{R}^k))$ and $\log(a_0)$ is viewed as $\log(a_0)\mathbf{1}$.

Next, we define the truncated logarithm map.

Definition 13 ([16] Definition A.4). For $\mathbf{a} = (a_0, a_1, \ldots) \in T((\mathbb{R}^k))$ with $a_0 > 0$, define $\log^N(\mathbf{a})$ to be the element of $T^N(\mathbb{R}^k)$ defined from the logarithm map (10) as

$$\log^N(\mathbf{a}) := P_N(\log(\tilde{a})),$$

where $\tilde{a} = (a_0, a_1, \ldots, a_N, 0, 0, \ldots) \in T((\mathbb{R}^k))$ and P_N denotes the orthogonal projection map from $T((\mathbb{R}^k))$ onto $T^N(\mathbb{R}^k)$.

Finally, we introduce the log-signature.

Definition 14 ([16] Definition A.5). The log-signature of a path $X\colon [0,T] \to \mathbb{R}^n$ of bounded variation over the interval $[s,t]$ is defined as $\mathrm{LogSig}_{s,t}(X) := \log(S_{s,t}(X))$, where $S_{s,t}(X)$ denotes the signature of X. Likewise the depth-N log-signature of X is defined for each $N \in \mathbb{N}$ as $\mathrm{LogSig}^N_{s,t}(X) := \log^N(S^N_{s,t}(X))$

Remark 6. While for a path X of bounded variation the signature $S_{s,t}(X)$ (and likewise $S^N_{s,t}(X)$) is element of the Lie group, the log-signature $\mathrm{LogSig}_{s,t}(X)$ (and likewise $\mathrm{LogSig}^N_{s,t}(X)$) is element of the Lie algebra. For further information on the Lie theory of rough paths the authors suggest Chapter 7 of [10].

With this definitions we are ready to introduce the Log-ODE method.

Definition 15 ([16] Definition A.8). Let $\mathbf{X}\colon \Delta_T \to T^{\lfloor p \rfloor}(\mathbb{R}^k)$ be a geometric p-rough path. We consider the ODE

$$\begin{aligned} \frac{\mathrm{d}z}{\mathrm{d}u} &= \mathrm{Taylor}(z(u), f, \mathrm{LogSig}^N_{s,t}(\mathbf{X})), \\ z(0) &= Y_s, \end{aligned} \tag{11}$$

where $u \in [0,1]$. Now, (11) is constructed for the purpose of approximating the RDE solution Y in (6) at time point t in the sense that

$$Y_t \approx z(1).$$

Remark 7. The construction of the vector fields in (11) is similar to the Taylor methods in Definition 10. The difference is the use of the log-signature instead of the signature.

Remark 8. The procedure of Definition 15 usually is repeated m times on a partition $D = \{0 = t_0 < t_1 < \cdots < t_m = T\}$. We define the numerical solution of the N-th level Log-ODE method by

$$y_{i+1}^{N\text{-Log-ODE}} := z(1),$$

where z is the solution of

$$\frac{\mathrm{d}z}{\mathrm{d}u} = \mathrm{Taylor}(z(u), f, \mathrm{LogSig}^N_{t_i,t_{i+1}}(\mathbf{X})),$$
$$z(0) = y_i^{N\text{-Log-ODE}}$$

and $y_0^{N\text{-Log-ODE}} = y_0$. Then, $y_i^{N\text{-Log-ODE}} \approx Y_{t_i}$.

We continue with a result on the convergence order of the Log-ODE method.

Theorem 4 ([16] Theorem B.1). *Assume that*

- $\mathbf{X}\colon \Delta_T \to T^{\lfloor p \rfloor}(\mathbb{R}^k)$ *be a geometric p-rough path,*
- $f = (f_1, ..., f_k)$ *is a collection of* $\mathrm{Lip}(\gamma)$*-vector fields on* $\mathbb{R}^n$ *for* $\gamma > p \geq 1$*,*
- $D = \{0 = t_0 < t_1 < \cdots < t_m = T\}$ *is a fixed partition of* $[0, T]$*, where* $h = \max_i t_{i+1} - t_i$*.*

Set $N := \lfloor \gamma \rfloor \geq \lfloor p \rfloor$*. Then, there exists a constant* $C = C(p, \gamma, \|f\|_{Lip(\gamma)}) > 0$*, so that*

$$|Y_T - y_m^{N\text{-}Log\text{-}ODE}| \leq C h^{\alpha},$$

where $\alpha = \frac{N+1}{p} - 1$ *and* h *is sufficiently small.*

We see that the order of convergence for fixed N for the Log-ODE method and the Taylor method are the same.

Remark 9. The assumptions of the above theorem ensure that the right hand side of the ODE (11) is globally bounded and Lipschitz continuous. The above error estimate also holds when the vector field f is linear ([16], Remark B.8).

3.4. *RK-Log-ODE method*

On the foundation of the introduced Runge-Kutta operator the authors propose a Log-ODE method which from now on is mentioned as the RK-Log-ODE method. Since we only introduced the Runge-Kutta method for $N = 2$ we will also only have a look at the RK-Log-ODE method for $N = 2$.

Definition 16. Let $\mathbf{X}\colon \Delta_T \to T^{\lfloor p \rfloor}(\mathbb{R}^k)$ be a geometric p-rough path and $p < 3$, $N = \lfloor p \rfloor$. We consider the ODE

$$\begin{aligned} \frac{\mathrm{d}z}{\mathrm{d}u} &= \text{Runge-Kutta}(z(u), f, \text{LogSig}^N_{s,t}(\mathbf{X})), \\ z(0) &= Y_s, \end{aligned} \tag{12}$$

where $u \in [0,1]$.
Now (12) is constructed for the purpose of approximating the RDE solution Y in (6) at time point t in the sense that

$$Y_t \approx z(1).$$

Let $D = \{0 = t_0 < t_1 < \cdots < t_m = T\}$ be a partition of $[0,T]$. We define the numerical solution of the RK-Log-ODE method by

$$y^{\text{RK-Log-ODE}}_{i+1} := \tilde{z}(1),$$

where $\tilde{z}$ is the solution of

$$\begin{aligned} \frac{\mathrm{d}\tilde{z}}{\mathrm{d}u} &= \text{Runge-Kutta}(\tilde{z}(u), f, \text{LogSig}^N_{t_i,t_{i+1}}(\mathbf{X})), \\ \tilde{z}(0) &= y^{\text{RK-Log-ODE}}_i \end{aligned}$$

and $y^{\text{RK-Log-ODE}}_0 = y_0$. Then, $y^{\text{RK-Log-ODE}}_i \approx Y_{t_i}$.

4. Numerical examples

Before we present our results we start to outline the experiments. Let $\mathbf{X}\colon \Delta_T \to T^{\lfloor p \rfloor}(\mathbb{R}^k)$ be a geometric p-rough path and $f = (f_1, \dots, f_k)\colon \mathbb{R}^n \to \mathbb{R}^{n\times k}$. We aim to approximate the solution $Y\colon [0,T] \to \mathbb{R}^n$ of the RDE

$$\begin{aligned} \mathrm{d}Y_t &= f(Y_t)\,\mathrm{d}\mathbf{X}_t, \\ Y_0 &= y_0, \end{aligned} \tag{13}$$

where $t \in [0,T]$ with $T = 1$ and

$$(f(Y_t))_{ij} = \begin{cases} a_i \cdot \cos(j \cdot Y^i_t), & j \text{ even}, \\ a_i \cdot \sin(j \cdot Y^i_t), & j \text{ odd}. \end{cases}$$

The coefficients $a_i \in \mathbb{R}$ are sampled independently from a uniform distribution $U([-1,1])$. We solve this RDE for random initial values y_0 (also uniformly distributed on $[-1,1]^n$) and varying, growing dimensions n and k.
At first we comment on the geometric p-rough path $\mathbf{X}$. As the underlying function we choose a path of a fractional Brownian motion (fBm) B^H with Hurst index $H = 0.4$, i.e., $X = \pi_1(\mathbf{X}) = B^H(\omega)$. From analysis it is known that a function of Hölder regularity α is of finite p-variation with $p > \frac{1}{\alpha}$. Since the paths of the fBm are almost surely Hölder continuous with $\alpha < H$, X is of finite p-variation for all $p > 2.5$ (see [22]).
We approximate $\mathbf{X}_{s,t}$ via a sequence of $(x^n)_n$ of finite variation. This is possible by Definition 5. In [23] it is proven that the Wong-Zakai approximation, i.e., piecewise linear approximations is a possible choice in Definition 5 for a fBm with $\frac{1}{4} < H < \frac{1}{2}$. Since X is of finite p-variation for all $p > 2.5$ we need to compute the first $N = \lfloor p \rfloor = 2$ levels of the signature or the log-signature, respectively. While the first level $\pi_1(\mathbf{X}_{s,t})$ is nothing more than the increment $X_{s,t}$, the second level cannot be computed directly and is approximated via $\pi_2(S_{s,t}(x^n))$ which we computed directly using the following lemma.

Lemma 2. *Let x^n be piecewise linear for a partition $[s = \tau_0 < \cdots < \tau_n = t]$. Then,*

$$\pi_2(S_{s,t}(x^n)) = \frac{1}{2}\sum_{l=1}^{n} \pi_1(S_{\tau_{l-1},\tau_l}(x^n)) \otimes \pi_1(S_{\tau_{l-1},\tau_l}(x^n)) + \sum_{l=1}^{n}\sum_{i=1}^{l} \pi_1(S_{\tau_{i-1},\tau_i}(x^n)) \otimes \pi_1(S_{\tau_{l-1},\tau_l}(x^n)).$$

Proof. The statements holds true in the case $n = 1$, since

$$\int_{\tau_{l-1}}^{\tau_l}\int_{\tau_{l-1}}^{u} \mathrm{d}x_i^n(r)\,\mathrm{d}x_j^n(u) = \frac{1}{2}(x_i^n(\tau_l) - x_i^n(\tau_{l-1}))(x_j^n(\tau_l) - x_j^n(\tau_{l-1}))$$

for $i,j = (1,\ldots,k)$.
The general case follows via induction over n assuming that the statement holds true for $n-1$, with Chen's identity (3) and the arithmetic of the

tensor algebra (1) follows that

$$\begin{aligned}\pi_2(S_{s,t}(x^n)) &= \pi_2(S_{\tau_0,\tau_{n-1}}(x^n)) + \pi_2(S_{\tau_{n-1},\tau_n}(x^n)) \\ &\quad + \pi_1(S_{\tau_0,\tau_{n-1}}(x^n)) \otimes \pi_1(S_{\tau_{n-1},\tau_n}(x^n)) \\ &= \frac{1}{2}\sum_{l=1}^{n-1} \pi_1(S_{\tau_{l-1},\tau_l}(x^n)) \otimes \pi_1(S_{\tau_{l-1},\tau_l}(x^n)) \\ &\quad + \sum_{l=1}^{n-1}\sum_{i=1}^{l} \pi_1(S_{\tau_{i-1},\tau_i}(x^n)) \otimes \pi_1(S_{\tau_{l-1},\tau_l}(x^n)) \\ &\quad + \frac{1}{2}\pi_1(S_{\tau_{n-1},\tau_n}(x^n)) \otimes \pi_1(S_{\tau_{n-1},\tau_n}(x^n)) \\ &\quad + \pi_1(S_{\tau_0,\tau_{n-1}}(x^n)) \otimes \pi_1(S_{\tau_{n-1},\tau_n}(x^n)) \\ &= \frac{1}{2}\sum_{l=1}^{n} \pi_1(S_{\tau_{l-1},\tau_l}(x^n)) \otimes \pi_1(S_{\tau_{l-1},\tau_l}(x^n)) \\ &\quad + \sum_{l=1}^{n}\sum_{i=1}^{l} \pi_1(S_{\tau_{i-1},\tau_i}(x^n)) \otimes \pi_1(S_{\tau_{l-1},\tau_l}(x^n)).\end{aligned}$$

This concludes the proof. □

Then, we compute the second level of the log-signature $\mathrm{LogSig}^N_{s,t}(\mathbf{X})$ via

$$\pi_2(\mathrm{LogSig}^N_{s,t}(\mathbf{X}_{s,t})) = \pi_2(\mathbf{X}_{s,t}) - \frac{1}{2}(\pi_1(\mathbf{X}_{s,t}) \otimes \pi_1(\mathbf{X}_{s,t})).$$

Note that we choose a representation of $\mathrm{LogSig}^N_{s,t}(\mathbf{X})$ in terms of a basis of the truncated tensor algebra $T^N(\mathbb{R}^k)$ instead of a basis of the Lie algebra. All implementations are made in Python. There are some more implementation details to clarify. The derivatives in Definitions 10 and 15 are computed using forward automatic differentiation, where we use the function torch.func.jvp. The ODEs in Definitions 15 and 16 are solved using a single step of the Runge-Kutta 23 scheme from scipy library with the solve_ivp function. The underlying fBm is sampled on a partition of $[0, T]$ with twice as many grid points as the partition of the numerical solutions. For the computation of a reference solution of (13) we use Heun's third-order method from [11] with $m = 2^{15}$ grid points.

Now, we apply the proposed methods. Therefore, let TM denote the 2-nd level Taylor method (Definition 10), let Log-ODE denote the 2-nd level Log-ODE method (Definition 15). Finally, let RK and RK-Log-ODE denote the Runge-Kutta (Definition 11) resp. the Runge-Kutta-Log-ODE method (Definition 16). We start by an investigation of the error

$$\mathcal{E} = |Y_T - y_m|$$

for a partition $0 = t_0 < \cdots < t_m = T$. We analyze the error versus the number of gridpoints in log-log diagrams. The stepsize of each method equals $\frac{1}{m}$. Figure 1 suggests that RK and RK-Log-ODE indeed converge. From

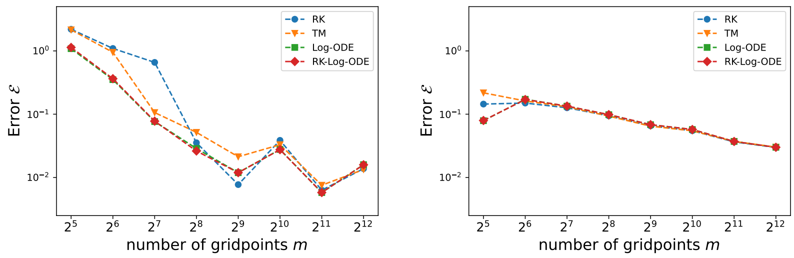

Figure 1. Error vs. grid points for RDE (13) with $n = 2$, $k = 2$ for 2 different paths

the convergence analysis of TM (Theorem 3) and Log-ODE (Theorem 4) we expect to see at least an order of convergence of $\alpha = \frac{N+1}{p} - 1 = \frac{3}{2.5} - 1 = 0.2$. The expected order of convergence (EOC) should theoretically neither be perturbed by the use of ODE solvers for the ODEs (11) and (12), since they have a way higher convergence rate of 3, nor the numerical approximation of $\mathbf{X}$ because this rate, according to [23], is $2H - 0.5 = 0.3$, which is also greater than $\alpha = 0.2$.

Since the development of the error suffers from the roughness of the fBm it is not always possible to observe a clear EOC. Therefore, in Figure 2, we average the error over 200 independent trajectories of the fBm in order to smooth out perturbations in the EOC. Figure 2 suggests a similar convergence behaviour for TM, RK, Log-ODE and RK-Log-ODE. The actual EOC seems to be higher than the theoretical predicted worst-case rate of 0.2. The authors expect the EOC to be $\alpha = 0.3$. This is natural as the total error of the approximation would then be determined by the error of the discretization of the rough path, i.e., $\mathbf{X}_{t_i,t_{i+1}} \approx S^2_{t_i,t_{i+1}}(x^n)$, where this approximation converges with a rate of 0.3. We refer to Remark 3 once more, where a similar effect was pointed out.

We complete the section with an investigation of the effort of each method for solving the RDE (13). In Table 1 we explore the running times for different dimensions and fixed grid size. In Table 2 we fix the dimensions and vary over the grid size. There are two important observations to make from the

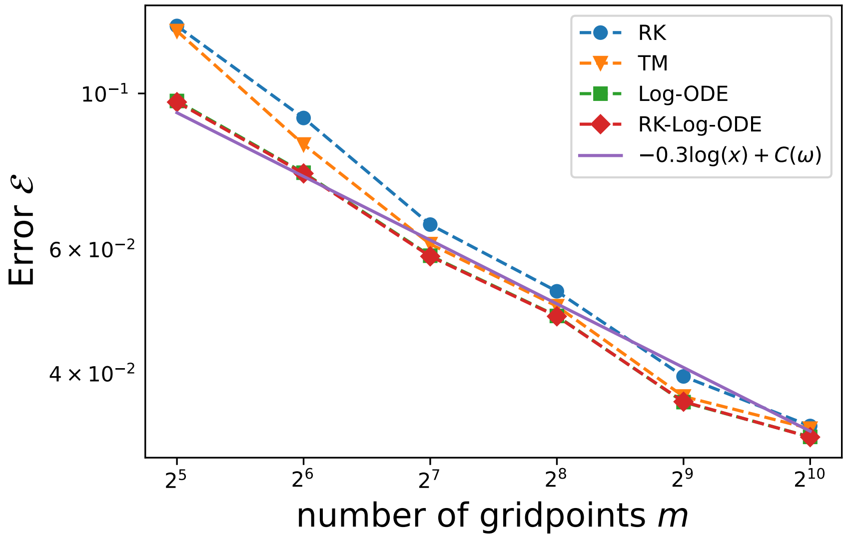

Figure 2. Error vs. grid points for RDE (13) with $n = k = 2$

Table 1. Running times of numerical schemes for RDE (13) for varying dimensions n, k and fixed $m = 2^8$

	$n = 2$ $k = 2$	$n = 10$ $k = 10$	$n = 100$ $k = 20$	$n = 1000$ $k = 20$
TM	0.25s	1.2s	3s	19s
RK	0.01s	0.1s	0.5s	1.6s
Log-ODE	1s	19s	160s	789s
RK-Log-ODE	0.08s	1.9s	12s	45s

experiments on the running time. At first the methods based on the Runge-Kutta operator heavily outperform their Taylor counterparts, namely RK is faster than TM and RK-Log-ODE performs better than Log-ODE. The second observation, one could conclude the TM and RK outperform their Log-ODE counterpart by just looking at the running time. This is not the intention of the experiment. Usually Log-ODE (and RK-Log-ODE) is used on a coarse grid and saves time this way, while TM (and RK) work on a really fine partition. Since we compared all methods for the same partition size this outcome is expected.

Table 2. Running times of numerical schemes for RDE (13) for varying grid points m and fixed dimensions $n = 2$, $k = 2$

	$m = 2^{10}$	$m = 2^{12}$	$m = 2^{14}$	$m = 2^{16}$
TM	1.57s	6.3s	25s	101s
RK	0.09s	0.35s	1.4s	5.7s
Log-ODE	6.6s	29s	103s	415s
RK-Log-ODE	0.45s	3.6s	7.1s	28s

5. Conclusion

In the present paper two efficient Runge-Kutta approaches for potentially high-dimensional RDEs driven by geometric p-rough paths with $p < 3$ are proposed. Compared to well-known methods such as the Taylor schemes (Definition 10) and the Log-ODE method (Definition 15) numerical evidence (see Tables 1 and 2) suggests significantly reduced computational cost for the proposed Runge-Kutta scheme (Definition 11) and the RK-Log-ODE method (Definition 16) especially in the case of high dimensions. While again numerical evidence (see Figures 1 and 2) indicates the same order of convergence for all methods. A numerical analysis of the proposed schemes is an open question as well as the generalization of the proposed schemes to general geometric p-rough paths for $p \geq 3$, which the authors intend to tackle in the future.

Acknowledgements

The work on this paper started while MR and JW were supported by the DFG individual grant “Low-order approximations for large-scale problems arising in the context of high-dimensional PDEs and spatially discretized SPDEs” – project number 499366908. Both authors would like to thank Wilfried Grecksch for his helpful suggestions on the manuscript and Christian Bayer for discussions on related topics.

References

[1] A. Rößler, Runge–Kutta Methods for the Strong Approximation of Solutions of Stochastic Differential Equations, *SIAM Journal on Numerical Analysis.* **48**(3), 922–952 (2010). doi: 10.1137/09076636X. URL `https://doi.org/10.1137/09076636X`.

[2] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*. Stochastic Modelling and Applied Probability, Springer Berlin

Heidelberg (2011). ISBN 9783540540625. URL https://books.google.de/books?id=BCvtssom1CMC.

[3] K. Burrage and P. M. Burrage, Order Conditions of Stochastic Runge–Kutta Methods by B-Series, *SIAM Journal on Numerical Analysis*. **38**(5), 1626–1646 (2000). doi: 10.1137/S0036142999363206. URL https://doi.org/10.1137/S0036142999363206.

[4] K. Debrabant and A. Kværnø, B–Series Analysis of Stochastic Runge–Kutta Methods That Use an Iterative Scheme to Compute Their Internal Stage Values, *SIAM Journal on Numerical Analysis*. **47**, 181–203 (01, 2008). doi: 10.1137/070704307.

[5] G. Milstein and M. Tretyakov, *Stochastic Numerics for Mathematical Physics*. Springer Berlin (01, 2004). ISBN 978-3-642-05930-8. doi: 10.1007/978-3-662-10063-9.

[6] F. Castell, Asymptotic expansion of stochastic flows, *Probab. Th. Rel. Fields*. **96**, 225–239 (1993). doi: https://doi.org/10.1007/BF01192134.

[7] J. Castell, Fabienne; Gaines. The ordinary differential equation approach to asymptotically efficient schemes for solution of stochastic differential equations. URL http://www.numdam.org/item/AIHPB_1996__32_2_231_0/ (1996).

[8] F. Castell and J. Gaines, An efficient approximation method for stochastic differential equations by means of the exponential Lie series, *Mathematics and Computers in Simulation*. **38**(1), 13–19 (1995). ISSN 0378-4754. doi: https://doi.org/10.1016/0378-4754(93)E0062-A. URL https://www.sciencedirect.com/science/article/pii/0378475493E0062A.

[9] A. M. Davie. Differential equations driven by rough paths: an approach via discrete approximation. URL https://arxiv.org/abs/0710.0772 (2007).

[10] P. K. Friz and N. B. Victoir, *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*. Cambridge Studies in Advanced Mathematics, Cambridge University Press (2010).

[11] M. Redmann and S. Riedel, Runge-Kutta methods for rough differential equations, *Journal of Stochastic Analysis*. **3**(4) (2022). doi: 10.31390/josa.3.4.06. URL https://repository.lsu.edu/josa/vol3/iss4/6.

[12] 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). ISSN 0246-0203. doi: 10.1214/10-AIHP392. URL http://dx.doi.org/10.1214/10-AIHP392.

[13] J. Hong, C. Huang, and X. Wang, Symplectic Runge-Kutta methods for Hamiltonian systems driven by Gaussian rough paths, *Appl. Numer. Math.* **129**, 120–136 (2018). ISSN 0168-9274. doi: 10.1016/j.apnum.2018.03.006. URL https://doi.org/10.1016/j.apnum.2018.03.006.

[14] T. Lyons. Rough paths, Signatures and the modelling of functions on streams. URL https://arxiv.org/abs/1405.4537 (2014).

[15] Y. Boutaib, L. G. Gyurkó, T. Lyons, and D. Yang. Dimension-free Euler estimates of rough differential equations. URL https://arxiv.org/abs/1307.4708 (2013).

[16] J. Morrill, P. Kidger, C. Salvi, J. Foster, and T. J. Lyons, Neural CDEs

for Long Time Series via the Log-ODE Method, *CoRR*. **abs/2009.08295** (2020). URL https://arxiv.org/abs/2009.08295.

[17] C. Bayer, S. Breneis, and T. Lyons. An Adaptive Algorithm for Rough Differential Equations. URL https://arxiv.org/abs/2307.12590 (2023).

[18] R. A. Ryan, *Introduction to Tensor Products of Banach Spaces*. Springer Monographs in Mathematics, Springer London (2002).

[19] E. Wong and M. Zakai, On the Convergence of Ordinary Integrals to Stochastic Integrals, *The Annals of Mathematical Statistics*. **36**(5), 1560–1564 (1965). ISSN 00034851. URL http://www.jstor.org/stable/2238444.

[20] T. J. Lyons, Differential equations driven by rough signals., *Revista Matemática Iberoamericana*. **14**(2), 215–310 (1998). URL http://eudml.org/doc/39555.

[21] P. Kloeden and A. Neuenkirch, The Pathwise Convergence of Approximation Schemes for Stochastic Differential Equations, *LMS Journal of Computation and Mathematics*. **10** (01, 2007). doi: 10.1112/S1461157000001388.

[22] G. Shevchenko. Fractional Brownian motion in a nutshell. URL https://arxiv.org/abs/1406.1956 (2014).

[23] P. Friz and S. Riedel. Convergence rates for the full Gaussian rough paths. URL https://arxiv.org/abs/1108.1099 (2012).