Journal of Stochastic Analysis

Volume 3 | Number 4 Article 6

December 2022

Runge-Kutta Methods for Rough Differential Equations

Martin Redmann
Martin Luther University Halle-Wittenberg, Institute of Mathematics, Theodor-Lieser-Str. 5, 06120 Halle (Saale), Germany, martin.redmann@mathematik.uni-halle.de

Sebastian Riedel
FernUniversität in Hagen, Faculty of Mathematics and Computer Science, 58084 Hagen, Germany, sebastian.riedel@fernuni-hagen.de

Recommended Citation
Redmann, Martin and Riedel, Sebastian (2022) "Runge-Kutta Methods for Rough Differential Equations," *Journal of Stochastic Analysis*: Vol. 3: No. 4, Article 6.
DOI: 10.31390/josa.3.4.06
Available at: https://digitalcommons.lsu.edu/josa/vol3/iss4/6

Journal of Stochastic Analysis
Vol. 3, No. 4 (2022) Article 6 (24 pages)
DOI: 10.31390/josa.3.4.06

RUNGE-KUTTA METHODS FOR ROUGH DIFFERENTIAL EQUATIONS

M. REDMANN* AND S. RIEDEL

Abstract. We study Runge-Kutta methods for rough differential equations which can be used to calculate solutions to stochastic differential equations driven by processes that are rougher than a Brownian motion. We use a Taylor series representation (B-series) for both the numerical scheme and the solution of the rough differential equation in order to determine conditions that guarantee the desired order of the local error for the underlying Runge-Kutta method. Subsequently, we prove the order of the global error given the local rate. In addition, we simplify the numerical approximation by introducing a Runge-Kutta scheme that is based on the increments of the driver of the rough differential equation. This simplified method can be easily implemented and is computational cheap since it is derivative-free. We provide a full characterization of this implementable Runge-Kutta method meaning that we provide necessary and sufficient algebraic conditions for an optimal order of convergence in case that the driver, e.g., is a fractional Brownian motion with Hurst index $\frac{1}{4} < H \leq \frac{1}{2}$. We conclude this paper by conducting numerical experiments verifying the theoretical rate of convergence.

Introduction

Ordinary differential equations (ODEs) have many real life applications. They, e.g., describe chemical, physiological and ecological processes or they appear as spatially discretized partial differential equations like the heat equation. Often analytic solutions to ODEs do not exist which requires numerical approximations in order to solve these equations. An important class of such schemes are Runge-Kutta methods [6, 21, 22] which can be of arbitrary order of convergence. These are often preferred in practice since they are derivative-free in contrast to Taylor methods. Computing derivatives of the right hand side function f_0 of an ODE can either be very costly or closed form expressions might not be available. However, in many applications uncertainties need to be taken into account. Therefore, for a more accurate modeling in such cases, a noise term can be added to an ODE leading to stochastic differential equations (SDEs). Runge-Kutta schemes for SDEs driven by a Brownian motion have already been established, see, e.g., [3, 10, 23, 28, 30].

Lyons' rough paths theory provides an alternative way to SDEs which goes far beyond the scope of usual Itō equations. In this paper, we are interested in studying numerical schemes to solve rough differential equations (RDEs) of the form

$$dy(t) = f_0(y(t))\,dt + f(y(t))\,d\mathbf{X}(t), \quad y(t_0) = y_0 \in \mathbb{R}^m, \tag{0.1}$$

where $\mathbf{X}$ is a suitable rough path above some α-Hölder path $X = (X^1, \dots, X^d)\colon [0,T] \to \mathbb{R}^d$, $f = (f_1, \dots, f_d)$ and $f_i\colon \mathbb{R}^m \to \mathbb{R}^m$ are vector fields for every $i = 0, \dots, d$. Such equations represent SDEs driven by stochastic processes that are potentially rougher than a Brownian motion if $\mathbf{X}$ is a random rough path, i.e. a stochastic process with sample paths lying in a rough paths space. One

Received 2022-6-9; Accepted 2022-12-10; Communicated by the editors.
2020 *Mathematics Subject Classification.* 60H10, 60H35, 60L20, 60L70, 65C30.
Key words and phrases. B-series, rough paths, Runge-Kutta methods.
* Corresponding author.

benefit of rough paths theory compared to Itō's theory is that one is not restricted to the martingale framework. In fact, there is a large class of stochastic processes which have "natural extensions" to rough paths valued processes, cf. [16]. For instance, many Gaussian processes possess such a "natural lift" including the fractional Brownian motion with Hurst parameter $H > 1/4$, but even more general processes like the bifractional Brownian motion, Volterra processes or processes which can be represented by random Fourier series, cf. [12] for a discussion.

In the context of rough paths theory, numerical schemes are indispensable when simulating the solution to an RDE driven by a random rough path or when discretizing rough stochastic partial differential equations [4]. In fact, numerical schemes played a fundamental role in rough paths theory from the very beginning. This is probably most visible in the work of Davie [9], where the Milstein scheme is used to solve rough differential equations theoretically. This approach was generalized to higher order Taylor-type schemes by Friz and Victoir [16]. However, in a stochastic context, these schemes are of little use in practice since they contain iterated stochastic integrals whose distribution is unknown in general. To overcome this difficulty, Deya, Neuenkirch and Tindel introduced so-called *simplified* schemes in [11] in which the iterated stochastic integrals are replaced by products of increments of the driving process. These schemes were successfully used in different contexts, cf. e.g. [5, 4]. However, as Taylor methods, these numerical approximations suffer from the need to calculate and evaluate derivatives of the vector fields f_k. E.g., the scheme studied in [5] involves all terms of the form $(f_i f_j)(y) := Df_j(y)f_i(y)$ and $(f_i f_j f_k)(y)$ with $i, j, k \in \{1, \dots, d\}$ leading to an enormous blow up in the number of vector fields already for small parameters m and d making it very expensive due to the required function calculations and evaluations. In fact, even if all desired derivatives are available, such a Taylor-based discretization method is not feasible in moderate high dimensions. This is particularly true in a large scale setting if (0.1), for instance, represents a spatially discretized rough partial differential equation. Therefore, we see the need of studying Runge-Kutta methods for rough differential equations that can easily be implemented and only involve suitable linear combinations of the vector fields instead of (higher-order) derivatives. These Runge-Kutta schemes are very benefitial as they are feasible for large m and d.

Our approach to establish Runge-Kutta methods is classical, both in the deterministic and the stochastic context: First, we define a class of equations which can be expanded in a B-series. Second, we have to find a B-series representation of the equation (0.1). Comparing both series, we can, in principle, deduce the order conditions of the Runge-Kutta method by matching their coefficients. A B-series representation of an ODE contains combinations of products and derivatives of the defining vector field which can be described in the language of trees. For SDEs, integrated products of iterated stochastic integrals have to be considered in addition which can be described in the same language. We call such objects *tree-iterated (stochastic) integrals* in the sequel. A rough path in the sense of Lyons [26, 25, 24] is a collection of objects which "mimic" the iterated integrals of the underlying path. Lyons' theory is able to solve differential equations driven by *geometric* rough paths, i.e. those which obey the usual chain rule. Gubinelli realized in [17] that one can even solve rough differential equations driven by non-geometric rough paths if one additionally assumes that all tree-iterated integrals are known. He calls such objects *branched rough paths*. Thinking of B-series representations of SDEs, this is a very natural approach to solve equations of the form (0.1). For us, it is therefore reasonable to use his theory and to interpret the equation (0.1) as a rough differential equation driven by a branched rough path. Doing this, we are able to deduce the B-series expansion of (0.1) in Theorem 2.10. Comparing both B-series and matching their coefficients up to a given order for an arbitrary multidimensional driving process X and its tree-iterated integrals can be very hard, cf. [3, Section 4] for a 2-dimensional example. However, we

already pointed out that in practice, it is hard to simulate the tree-iterated integrals anyway. We therefore make the same ansatz as in [11] and replace the tree-iterated integrals by products of increments. This simplifies the task of matching the coefficients a lot, and one is able to deduce the order conditions, in principle, up to any order, cf. Theorem 3.3 and the subsequent remark. We call such schemes *simplified Runge-Kutta methods*. As in [11], the Wong-Zakai error plays a fundamental role in their convergence analysis. Loosely speaking, our main result is the following:

Theorem. *Let* $\mathbf{X}$ *be an* α*-Hölder rough path (branched or geometric) and assume that* f_0 *and* f *are sufficiently smooth and bounded with bounded derivatives. If the Wong-Zakai error to approximate* (0.1) *is of order* r_0*, a simplified Runge-Kutta method* (3.5) *of order* p *converges with rate* $\min\{(p+1)\alpha - 1, r_0\}$.

As an application, we can study the scheme when the driving process is a fractional Brownian motion with Hurst parameter $H \in (1/4, 1/2]$. In this case, the Wong-Zakai error is arbitrarily close to $2H - 1/2$ almost surely, cf. [14]. We therefore obtain:

Corollary. *For a fractional Brownian motion with Hurst parameter* $H \in (1/4, 1/2]$*, a simplified Runge-Kutta scheme of order* 3 *converges with rate arbitrarily close to* $2H - 1/2$ *almost surely.*

We already pointed out that numerical schemes studied in the context of rough paths theory are mostly of Taylor-type. To our knowledge, the only exception is the article by Hong, Huang and Wang [18] where a class of symplectic Runge-Kutta methods is considered to solve Hamiltonian equations driven by Gaussian processes. Our article differs from [18] in several regards. On the technical level, no B-series are used in [18], the authors have to prove all necessary estimates "by hand" in the framework of geometric rough paths. Consequently, they do not provide general order conditions. For instance, no explicit Runge-Kutta methods are deduced in [18]. Moreover, their approach is probably hard to generalize to schemes of arbitrary order, whereas our approach does not have any limitations in this regard.

The article is structured as follows. In Section 1, we define the equations which can be expanded to obtain the desired B-series. Section 2 explains the concept of branched rough paths, deduces the B-series representation for equation (0.1) and discusses the local error of full Runge-Kutta methods. Simplified Runge-Kutta methods are defined in Section 3, where the necessary order conditions are derived to obtain the local error of the numerical scheme. In Section 4, we deduce the global error for our methods. The article closes with numerical experiments presented in Section 5.

Let us finally mention that in the whole article, we will discard the drift in (0.1) and consider equations of the form

$$dy(t) = f(y(t))\, d\mathbf{X}(t), \quad y(t_0) = y_0, \tag{0.2}$$

only which simplifies the exposition a lot. Furthermore, this is not a real limitation if we assume that the first component of X is just the path $t \mapsto t$.

Notation and Basic Definitions

General notation. Let I be an interval in $\mathbb{R}$ and V be a linear space. We call a function $X\colon I \to V$ a *path* and $X_{s,t} := X(t) - X(s)$ an *increment*. For a general two-parameter function $X\colon I \times I \to V$, we will often write $X_{s,t}$ instead of $X(s,t)$. If $|\cdot|$ is a norm on V and $X\colon I \times I \to V$, we set

$$\|X\|_\alpha := \|X\|_{\alpha;I} := \sup_{\substack{s,t\in I \\ s\neq t}} \frac{|X_{s,t}|}{|t-s|^\alpha}$$

for $\alpha \in (0,1]$ and call it the *α-Hölder (semi-)norm* of X. For $x \in \mathbb{R}$, we use the notation $\lfloor x \rfloor := \max\{k \in \mathbb{Z} : k \leq x\}$. Let $\gamma > 0$ and $\gamma = [\gamma] + \{\gamma\}$ where $[\gamma]$ is an integer and $\{\gamma\} \in (0,1]$. We will say that a vector field $f\colon \mathbb{R}^m \to \mathbb{R}^m$ belongs to the class Lip^γ if f is $[\gamma]$-times continuously differentiable and the $[\gamma]$-th derivative is locally $\{\gamma\}$-Hölder continuous. f is of class Lip^γ_b if, in addition, f and all its derivatives are bounded and if the $[\gamma]$-th derivative is globally $\{\gamma\}$-Hölder continuous. More generally, a collection of vector fields $f = (f_1, \dots, f_d)$ is of class Lip^γ resp. Lip^γ_b if every f_i, $i = 1, \dots, d$, is of class Lip^γ resp. Lip^γ_b.

Trees and the Connes-Kreimer Hopf algebra. In the following section, we will recall some basic notions on trees and define the Connes-Kreimer Hopf algebra. For a general account on Hopf algebras, cf. [31] or [1]. The Connes-Kreimer Hopf algebra is discussed e.g. in [27].

We define $\mathcal{T}$ to be the set of all rooted, labeled trees with vertex decorations from the set $\{1, \dots, d\}$. We will use a recursive definition to construct trees. The empty tree will be denoted by 1. We use the convention that $1 \notin \mathcal{T}$ and set $\mathcal{T}^0 := \mathcal{T} \cup \{1\}$. If $\tau_1, \dots \tau_m \in \mathcal{T}^0$, $[\tau_1 \cdots \tau_m]_a$ denotes the tree obtained by attaching all trees $\tau_1, \dots \tau_m$ to a new vertex which we label by $a \in \{1, \dots, d\}$. If $d = 1$, we can suppress the vertex decorations and consider unlabeled trees instead. We use the notation $\bullet_a = [1]_a$ for the single vertex tree with label a and $\bullet$ for the unlabeled single vertex tree. For example,

$$[[[1]_a]_b]_c = \overset{a}{\underset{c}{\overset{|}{\underset{|}{b}}}}, \quad [[1]_a[1]_b]_c = \overset{a}{\underset{c}{\curlyvee}}\overset{b}{}, \qquad [[[1]_a[1]_b]_c]_d = \overset{a}{\underset{d}{\overset{|}{\underset{|}{c}}}}\overset{b}{},$$

are labeled trees, and

$$[[[1]]] = \overset{|}{\underset{|}{\bullet}}, \quad [[1][1]] = \curlyvee, \qquad [[[1][1]]] = \curlyvee,$$

are the corresponding unlabeled versions. The order of the branches of the tree does not matter, i.e. $[\tau_{\sigma(1)} \cdots \tau_{\sigma(m)}]_a = [\tau_1 \cdots \tau_m]_a$ holds for every permutation σ. That means, for example,

$$\overset{a}{\underset{c}{\curlyvee}}\overset{b}{} = \overset{b}{\underset{c}{\curlyvee}}\overset{a}{} \quad \text{and} \quad \curlyvee = \curlyvee.$$

If $\tau \in \mathcal{T}^0$ is a tree, $|\tau|$ denotes the number of vertices. The set $\mathcal{T}_N$ consists of all trees $\tau \in \mathcal{T}$ such that $|\tau| \leq N$ and we set $\mathcal{T}^0_N := \mathcal{T}_N \cup \{1\}$. In the unlabeled case, we have, for example,

$$\mathcal{T}_4 = \left\{\bullet, \overset{|}{\bullet}, \overset{|}{\overset{|}{\bullet}}, \curlyvee, \overset{|}{\overset{|}{\overset{|}{\bullet}}}, \curlyvee\overset{|}{\bullet}, \curlyvee\bullet\right\}.$$

We let

$$\mathcal{F} := \{\tau_1 \cdots \tau_m : \tau_i \in \mathcal{T},\ m \in \mathbb{N}\}$$

denote the set of unordered forests and define $\mathcal{F}^0 := \mathcal{F} \cup \{1\}$. The map $|\cdot|$ is extended to $\mathcal{F}^0$ by setting

$$|\tau_1 \cdots \tau_m| := |\tau_1| + \ldots + |\tau_m|.$$

As before, $\mathcal{F}_N$ contains all $\mathfrak{h} \in \mathcal{F}$ with $|\mathfrak{h}| \leq N$ and $\mathcal{F}^0_N := \mathcal{F}_N \cup \{1\}$. For example, again in the unlabeled case,

$$\mathcal{F}_3 = \left\{\bullet, \bullet\bullet, \overset{|}{\bullet}, \bullet\bullet\bullet, \bullet\overset{|}{\bullet}, \overset{|}{\overset{|}{\bullet}}, \curlyvee\right\}.$$

We define $(\mathcal{H}, \cdot)$ to be the commutative polynomial algebra with coefficients in $\mathbb{R}$ generated by the variables $\mathcal{T}$. Alternatively, we can view $\mathcal{H}$ as the real vector space spanned by the elements in $\mathcal{F}^0$.

Typical elements in $\mathcal{H}$ are thus

$$1 + 3\bullet\bullet - \sqrt{2}\,\vee\!\bullet, \quad \bullet\!\!\downarrow - \frac{3}{2}\,\vdots, \quad \text{etc.}$$

and an example for the multiplication is

$$(2\,\vdots - 3\bullet\bullet) \cdot (1 + 4\,\vee\!\bullet) = 2\,\vdots - 3\bullet\bullet + 8\,\vdots\!\vee\!\bullet - 12\bullet\bullet\!\vee\!\bullet.$$

Let $\mathcal{H}\otimes\mathcal{H}$ denote the (algebraic) tensor product of $\mathcal{H}$ with itself. We define a linear map $M\colon \mathcal{H}\otimes\mathcal{H} \to \mathcal{H}$ by setting $M(x \otimes y) := x \cdot y$ for $x, y \in \mathcal{H}$. Also, we define the *unit* $\eta\colon \mathbb{R} \to \mathcal{H}$ by $\eta(x) := x \cdot 1$ where 1 denotes the empty tree. $(\mathcal{H}, M, \eta)$ is an algebra.

Next, we define a coalgebra structure on $\mathcal{H}$ (cf. [31, p. 4] for a precise definition). We define a linear map $\Delta\colon \mathcal{H} \to \mathcal{H}\otimes\mathcal{H}$, the *co-multiplication* or *co-product*, recursively by setting $\Delta 1 := 1\otimes 1$ and

$$\Delta[\tau_1 \cdots \tau_m]_a := [\tau_1 \cdots \tau_m]_a \otimes 1 + (\mathrm{id} \otimes B_+^a)(\Delta\tau_1 \cdots \Delta\tau_m)$$

for a tree $[\tau_1 \cdots \tau_m]_a$ where B_+^a is the operator defined by $B_+^a(\tau_1 \cdots \tau_n) := [\tau_1 \cdots \tau_n]_a$ on the forest $\tau_1 \cdots \tau_n$. We then extend the definition to forests by setting $\Delta(\tau_1 \cdots \tau_m) := \Delta\tau_1 \cdots \Delta\tau_m$ and eventually define Δ on $\mathcal{H}$ by linear extension. We will use Sweedler's notation

$$\Delta\mathfrak{h} = \sum_{(\mathfrak{h})} \mathfrak{h}^{(1)} \otimes \mathfrak{h}^{(2)}.$$

The *co-unit* is the linear map $\epsilon\colon \mathcal{H} \to \mathbb{R}$ which sends the empty tree to 1 and any nonempty forest to 0. This turns $(\mathcal{H}, \Delta, \epsilon)$ into a coalgebra. Furthermore, M and Δ are compatible in such a way that $(\mathcal{H}, M, \eta, \Delta, \epsilon)$ is a *bialgebra* [31, p. 53]. In order to obtain a *Hopf algebra* [31, p. 71], we need to define an *antipode* $S\colon \mathcal{H} \to \mathcal{H}$, i.e. a map which satisfies

$$M(\mathrm{id} \otimes S)\Delta\mathfrak{h} = M(S \otimes \mathrm{id})\Delta\mathfrak{h} = \mathfrak{h}$$

for every $\mathfrak{h} \in \mathcal{H}$. For our purposes, it suffices to state that the unique antipode exists, cf. [27, p. 44]. The Hopf algebra $(\mathcal{H}, M, \eta, \Delta, \epsilon, S)$ is called the *Connes-Kreimer Hopf algebra* [7], cf. also [19, Chapter 2]. Since the unit and co-unit are not important for us, we will just call $(\mathcal{H}, M, \Delta, S)$ the Connes-Kreimer Hopf algebra. There is also the notion of the *dual* of a Hopf algebra, cf. [1, p. 87]. The dual of the Connes-Kreimer Hopf algebra will be denoted by $(\mathcal{H}^*, \star, \delta, S^*)$. While $\mathcal{H}$ consists of all polynomials with real coefficients and unknown taken from the set $\mathcal{T}$, $\mathcal{H}^*$ can be identified with the set of the respective formal power series. We will identify elements in the basis $\mathcal{F}$ of $\mathcal{H}$ with elements in $\mathcal{H}^*$ by the natural pairing $\langle \mathfrak{h}_1, \mathfrak{h}_2 \rangle = \delta_{\mathfrak{h}_1, \mathfrak{h}_2}$ for every $\mathfrak{h}_1, \mathfrak{h}_2 \in \mathcal{F}$. The map $\star\colon \mathcal{H}^* \otimes \mathcal{H}^* \to \mathcal{H}^*$ is also called *convolution*. It is dual to Δ which means that

$$\langle \mathfrak{f} \star \mathfrak{g}, \mathfrak{h} \rangle = \langle \mathfrak{f} \otimes \mathfrak{g}, \Delta\mathfrak{h} \rangle$$

holds for every $\mathfrak{f}, \mathfrak{g} \in \mathcal{H}^*$ and $\mathfrak{h} \in \mathcal{H}$.

1. The Full Runge-Kutta Method

In this section, we will define s-stage Runge-Kutta methods. We follow the approach developed by Burrage and Burrage in [3]. Let $Z^{(1)}, \ldots, Z^{(d)}$ be given $s \times s$-matrices and $z^{(1)}, \ldots, z^{(d)}$ vectors

in $\mathbb{R}^s$. For given $y_n \in \mathbb{R}^m$, consider the equations

$$\begin{aligned} Y_i &= y_n + \sum_{k=1}^{d} \sum_{j=1}^{s} Z_{ij}^{(k)} f_k(Y_j) \\ y_{n+1} &= y_n + \sum_{k=1}^{d} \sum_{i=1}^{s} z_i^{(k)} f_k(Y_i). \end{aligned} \tag{1.1}$$

In what follows, Z and z will depend on increments of a rough path $\mathbf{X}$ with regularity α that drives the rough differential equation (0.2). Consequently, both values will also depend on the step size $h > 0$ of the numerical scheme (1.1). In applications, we will replace the fixed rough path by a stochastic process with sample paths being α-rough paths almost surely. In that case, Z and z will be random variables. Note that the first equation can be implicit in which case the existence of a solution is not guaranteed. In fact, this question will depend on the properties of the vector fields f_i. For instance, it can be shown that solutions exist in case that all vector fields are bounded, cf. [18, Proposition 4.1]. However, we will not address this question here and just assume that solutions exist.

Set $\Phi(1)(h) := \mathbf{1}_s := (1, \cdots, 1)^T \in \mathbb{R}^s$ for $i = 1, \dots, d$ and for a tree $\tau = [\tau_1 \cdots \tau_n]_i$,

$$\Phi(\tau)(h) := \Pi_{j=1}^{n}(Z^{(i)}\Phi(\tau_j)(h)), \quad a(\tau)(h) := \langle z^{(i)}, \Pi_{k=1}^{n}\Phi(\tau_k)(h)\rangle. \tag{1.2}$$

Notice that above, the product of two vectors has to be understood component-wise.

Definition 1.1. Let $f = (f_1, \dots, f_d)$ be sufficiently smooth such that all derivatives below exist. For a tree $\tau \in \mathcal{T}^0$, we define the *elementary differentials* $F(\tau)\colon \mathbb{R}^m \to \mathbb{R}^m$ recursively by setting

(i) $F(1)(y) := y$,
(ii) $F(\bullet_i)(y) := f_i(y)$ and
(iii) $F(\tau)(y) := f_i^{(n)}(y)(F(\tau_1)(y), \dots, F(\tau_n)(y))$ for a tree $\tau = [\tau_1 \cdots \tau_n]_i$ where $f_i^{(n)}$ denotes the n-th total derivative of f_i

for $y \in \mathbb{R}^m$.

Next,we define some combinatoric quantities. For unlabeled trees, we set

$$\gamma(1) = 1, \quad \gamma(\bullet) = 1, \quad \gamma([\tau_1, \dots, \tau_k]) = |[\tau_1, \dots, \tau_k]| \prod_{i=1}^{k} \gamma(\tau_i),$$

and

$$\beta(1) = 1, \quad \beta(\bullet) = 1, \quad \beta(\tau) := \binom{|\tau| - 1}{|\tau_1|, \dots, |\tau_k|} \frac{1}{r_1! \cdots r_q!} \prod_{j=1}^{k} \beta(\tau_j)$$

where $\tau = [\tau_1, \dots, \tau_k] = [(\tau_1)^{r_1}, \dots, (\tau_q)^{r_q}]$, $\tau_1, \dots, \tau_q$ being pairwise distinct trees. For labeled trees, we use the same definition. The main result in [3] we are going to use is the following:

Theorem 1.2. *The Taylor series expansion of* (1.1) *is*

$$y_1 = y_0 + \sum_{\tau \in \mathcal{T}} \frac{\gamma(\tau)}{|\tau|!} \beta(\tau) a(\tau)(h) F(\tau)(y_0). \tag{1.3}$$

Proof. [3, Theorem 2.5]. □

The coefficients in (1.2) are sometimes noted in a different form which we recall now. Following [17], we define the *symmetry factor* σ for unlabeled trees as

$$\sigma(1) = 1, \quad \sigma(\bullet) = 1, \quad \sigma(\tau) = r_1! \cdots r_q! \prod_{j=1}^{k} \sigma(\tau_j)$$

where $\tau = [\tau_1, \dots, \tau_k] = [(\tau_1)^{r_1}, \dots, (\tau_q)^{r_q}]$ with $\tau_1, \dots, \tau_q$ being pairwise distinct trees. For labeled trees, the same definition is used.

Lemma 1.3. *For every tree $\tau \in \mathcal{T}^0$,*

$$\frac{1}{\sigma(\tau)} = \frac{\gamma(\tau)}{|\tau|!}\beta(\tau).$$

Proof. We prove this lemma by induction on the height of τ for unlabeled trees. The equality is true for $\tau = 1$ and $\tau = \bullet$. Let us assume that the claim is true for each sub-tree of $\tau = [\tau_1, \dots, \tau_k] = [(\tau_1)^{r_1}, \dots, (\tau_q)^{r_q}]$. Then, we have

$$\begin{aligned}
\beta(\tau) &= \frac{(|\tau|-1)!}{|\tau_1|! \cdots |\tau_k|!} \frac{1}{r_1! \cdots r_q!} \prod_{j=1}^{k} \beta(\tau_j) \\
&= \frac{(|\tau|-1)!}{|\tau_1|! \cdots |\tau_k|!} \frac{1}{r_1! \cdots r_q!} \prod_{j=1}^{k} \frac{|\tau_j|!}{\gamma(\tau_j)} \frac{1}{\sigma(\tau_j)} \\
&= \frac{(|\tau|-1)!}{r_1! \cdots r_q!} \frac{|\tau|}{\gamma(\tau)} \prod_{j=1}^{k} \frac{1}{\sigma(\tau_j)} \\
&= \frac{|\tau|!}{\gamma(\tau)} \frac{1}{\sigma(\tau)}.
\end{aligned}$$

□

2. Branched Rough Paths and B-series Expansion for Rough Differential Equations

In this section, we recall the concept of a *branched rough paths* introduced by Gubinelli in [17]. We use a similar approach and notation as Hairer and Kelly in [19]. Our main goal is to deduce the B-series expansion of (0.2) which we will eventually achieve in Theorem 2.10. Note that Hairer and Kelly state a similar result in [19, Proposition 3.8]. However, we can not use their result for two reasons: First, it is not quantitative, i.e. the order of the truncation error is not specified. Second, it is not entirely correct since the expansion in [19, Proposition 3.8] lacks the symmetry factor. We also want to mention that the correct B-series expansion is deduced, even in an infinite dimensional framework, in [8] and [2]. However, the authors do not specify the order of the truncation error there either.

Definition 2.1. Let $\alpha \in (0, 1]$. A *α-branched rough path* is a map $\mathbf{X}\colon [0,T] \times [0,T] \to \mathcal{H}^*$ such that

(1) for all $s, t \in [0,T]$ and all $\mathfrak{h}_1, \mathfrak{h}_2 \in \mathcal{H}$,

$$\langle \mathbf{X}_{s,t}, \mathfrak{h}_1 \rangle \langle \mathbf{X}_{s,t}, \mathfrak{h}_2 \rangle = \langle \mathbf{X}_{s,t}, \mathfrak{h}_1 \cdot \mathfrak{h}_2 \rangle$$

(2) for all $s, u, t \in [0,T]$,

$$\mathbf{X}_{s,t} = \mathbf{X}_{s,u} \star \mathbf{X}_{u,t}$$

(3) for every $\tau \in \mathcal{T}$,

$$\sup_{s \neq t} \frac{|\langle \mathbf{X}_{s,t}, \tau \rangle|}{|t-s|^{\alpha|\tau|}} < \infty.$$

The space of α-branched rough paths will be denoted by $\mathscr{C}^\alpha([0,T], \mathbb{R}^d)$. It is a complete metric space with metric

$$\varrho_\alpha(\mathbf{X}, \mathbf{Y}) := \sum_{\tau \in \mathcal{T}_N} \sup_{s \neq t} \frac{|\langle \mathbf{X}_{s,t} - \mathbf{Y}_{s,t}, \tau \rangle|}{|t-s|^{\alpha|\tau|}}$$

where $N = \lfloor 1/\alpha \rfloor$.

Example 2.2. Let $X = (X^1, \dots, X^d) \colon [0,T] \to \mathbb{R}^d$ be a piece-wise C^1-path. We define $\mathbf{X}$ by setting

$$\langle \mathbf{X}_{s,t}, \bullet_i \rangle = X^i(t) - X^i(s) \quad \text{and} \quad \langle \mathbf{X}_{s,t}, [\tau_1 \cdots \tau_n]_i \rangle = \int_s^t \langle \mathbf{X}_{s,u}, \tau_1 \rangle \cdots \langle \mathbf{X}_{s,u}, \tau_n \rangle \, dX^i(u)$$

for any tree $\tau = [\tau_1 \cdots \tau_n]_i \in \mathcal{T}$. For example,

$$\langle \mathbf{X}_{s,t}, \begin{array}{c} \bullet^k \\ \bullet_i \quad \bullet^l \end{array} \rangle = \int_s^t \left(\int_s^u \int_s^v dX^k(w) \, dX^j(v) \right) \left(\int_s^u dX^l(v) \right) dX^i(u).$$

We then set

$$\langle \mathbf{X}_{s,t}, \tau_1 \cdots \tau_n \rangle = \langle \mathbf{X}_{s,t}, \tau_1 \rangle \cdots \langle \mathbf{X}_{s,t}, \tau_n \rangle$$

for any forest $\tau_1 \cdots \tau_n$. We can now extend $\mathbf{X}$ linearly to a map on $\mathcal{H}$ and therefore obtain a map $\mathbf{X} \colon [0,T] \times [0,T] \to \mathcal{H}^*$. It can be shown that this map defines a α-branched rough path for every $\alpha \in (0,1]$. If X is α-Hölder continuous for some $\alpha \in (1/2, 1]$, we can use the Young integral to define $\mathbf{X}$ as above. In this case, it can be shown that $\mathbf{X}$ defines a α'-branched rough path for every $\alpha' \in (0, \alpha]$. In this case, $\mathbf{X}$ is called the *natural lift* of the (smooth) path X to a branched rough path.

Next, we define a class of paths which we can integrate against a branched rough path.

Definition 2.3. Let $\mathbf{X}$ be a α-branched rough path and $N = \lfloor 1/\alpha \rfloor$. A path $\mathbf{Z} \colon [0,T] \to \mathcal{H}_{N-1}$ satisfying

$$\langle \mathfrak{h}, \mathbf{Z}(t) \rangle = \langle \mathbf{X}_{s,t} \star \mathfrak{h}, \mathbf{Z}(s) \rangle + R^{\mathfrak{h}}_{s,t} \tag{2.1}$$

for each $\mathfrak{h} \in \mathcal{F}^0_{N-1}$ where $|R^{\mathfrak{h}}_{s,t}| \leq C|t-s|^{(N-|\mathfrak{h}|)\alpha}$ is called *controlled by* $\mathbf{X}$. We will also say that $\mathbf{Z}$ is a *controlled path above the path* $t \mapsto Z(t) := \langle 1, \mathbf{Z}(t) \rangle$. More generally, we call a path $\mathbf{Z} \colon [0,T] \to (\mathcal{H}_{N-1})^m$ *controlled by* $\mathbf{X}$ if (2.1) holds, understood as an equation in $\mathbb{R}^m$. The space of controlled paths $\mathbf{Z} \colon [0,T] \to (\mathcal{H}_{N-1})^m$ will be denoted by $\mathcal{Q}_{\mathbf{X}}(\mathbb{R}^m)$ which is a Banach space with the norm

$$\|\mathbf{Z}\|_{\mathcal{Q}_{\mathbf{X}}(\mathbb{R}^m)} := |\mathbf{Z}(0)| + \sum_{\mathfrak{h} \in \mathcal{F}^0_{N-1}} \|R^{\mathfrak{h}}\|_{(N-|\mathfrak{h}|)\alpha}.$$

Lemma 2.4. *Let $\mathbf{X}$ be a α-branched rough path and $\mathbf{Z}$ be controlled by $\mathbf{X}$. Fix $i \in \{1, \dots, d\}$ and set*

$$\tilde{Z}_{s,t} := \sum_{\mathfrak{h} \in \mathcal{F}^0_{N-1}} \langle \mathfrak{h}, \mathbf{Z}(s) \rangle \langle \mathbf{X}_{s,t}, [\mathfrak{h}]_i \rangle.$$

Then,

$$\tilde{Z}_{s,t} - \tilde{Z}_{s,u} - \tilde{Z}_{u,t} = -\sum_{\mathfrak{h}\in\mathcal{F}^0_{N-1}} \langle \mathbf{X}_{u,t}, [\mathfrak{h}]_i\rangle R^{\mathfrak{h}}_{s,u}.$$

Proof. Let $\tilde{\mathfrak{h}} \in \mathcal{F}^0_{N-1}$. Then,

$$\begin{aligned}
\langle \mathbf{X}_{s,t} \star \tilde{\mathfrak{h}}, \mathbf{Z}(s)\rangle &= \sum_{\mathfrak{h}\in\mathcal{F}^0_{N-1}} \langle \mathbf{X}_{s,t} \star \tilde{\mathfrak{h}}, \mathfrak{h}\rangle\langle \mathfrak{h}, \mathbf{Z}(s)\rangle \\
&= \sum_{\mathfrak{h}\in\mathcal{F}^0_{N-1}} \langle \mathbf{X}_{s,t} \otimes \tilde{\mathfrak{h}}, \Delta\mathfrak{h}\rangle\langle \mathfrak{h}, \mathbf{Z}(s)\rangle \\
&= \sum_{\mathfrak{h}\in\mathcal{F}^0_{N-1}} \sum_{(\mathfrak{h})} \langle \mathbf{X}_{s,t} \otimes \tilde{\mathfrak{h}}, \mathfrak{h}^{(1)}\otimes\mathfrak{h}^{(2)}\rangle\langle \mathfrak{h}, \mathbf{Z}(s)\rangle \\
&= \sum_{\mathfrak{h}\in\mathcal{F}^0_{N-1}} \sum_{(\mathfrak{h})} \mathbf{1}_{\{\mathfrak{h}^{(2)}=\tilde{\mathfrak{h}}\}} \langle \mathbf{X}_{s,t}, \mathfrak{h}^{(1)}\rangle\langle \mathfrak{h}, \mathbf{Z}(s)\rangle.
\end{aligned}$$

For $\mathfrak{h} \in \mathcal{F}^0_{N-1}$,

$$\begin{aligned}
\langle \mathbf{X}_{s,t}, [\mathfrak{h}]_i\rangle &= \langle \mathbf{X}_{s,u} \star \mathbf{X}_{u,t}, [\mathfrak{h}]_i\rangle \\
&= \langle \mathbf{X}_{s,u} \otimes \mathbf{X}_{u,t}, \Delta[\mathfrak{h}]_i\rangle \\
&= \langle \mathbf{X}_{s,u} \otimes \mathbf{X}_{u,t}, [\mathfrak{h}]_i \otimes 1\rangle + \langle \mathbf{X}_{s,u} \otimes \mathbf{X}_{u,t}, (\mathrm{id}\otimes B^i_+)\Delta\mathfrak{h}\rangle \\
&= \langle \mathbf{X}_{s,u}, [\mathfrak{h}]_i\rangle + \sum_{(\mathfrak{h})} \langle \mathbf{X}_{s,u}, \mathfrak{h}^{(1)}\rangle\langle \mathbf{X}_{u,t}, [\mathfrak{h}^{(2)}]_i\rangle.
\end{aligned}$$

It follows that

$$\begin{aligned}
\tilde{Z}_{s,t} - \tilde{Z}_{s,u} - \tilde{Z}_{u,t} &= \sum_{\mathfrak{h}\in\mathcal{F}^0_{N-1}} \langle \mathfrak{h}, \mathbf{Z}(s)\rangle \left(\langle \mathbf{X}_{s,t}, [\mathfrak{h}]_i\rangle - \langle \mathbf{X}_{s,u}, [\mathfrak{h}]_i\rangle\right) - \langle \mathfrak{h}, \mathbf{Z}(u)\rangle\langle \mathbf{X}_{u,t}, [\mathfrak{h}]_i\rangle \\
&= \sum_{\mathfrak{h}\in\mathcal{F}^0_{N-1}} \sum_{(\mathfrak{h})} \langle \mathfrak{h}, \mathbf{Z}(s)\rangle\langle \mathbf{X}_{s,u}, \mathfrak{h}^{(1)}\rangle\langle \mathbf{X}_{u,t}, [\mathfrak{h}^{(2)}]_i\rangle - \langle \mathfrak{h}, \mathbf{Z}(u)\rangle\langle \mathbf{X}_{u,t}, [\mathfrak{h}]_i\rangle \\
&= \sum_{\mathfrak{h}\in\mathcal{F}^0_{N-1}} \sum_{(\mathfrak{h})} \sum_{\tilde{\mathfrak{h}}\in\mathcal{F}^0_{N-1}} \mathbf{1}_{\{\mathfrak{h}^{(2)}=\tilde{\mathfrak{h}}\}} \langle \mathfrak{h}, \mathbf{Z}(s)\rangle\langle \mathbf{X}_{s,u}, \mathfrak{h}^{(1)}\rangle\langle \mathbf{X}_{u,t}, [\tilde{\mathfrak{h}}]_i\rangle - \langle \mathfrak{h}, \mathbf{Z}(u)\rangle\langle \mathbf{X}_{u,t}, [\mathfrak{h}]_i\rangle \\
&= \sum_{\tilde{\mathfrak{h}}\in\mathcal{F}^0_{N-1}} \sum_{\mathfrak{h}\in\mathcal{F}^0_{N-1}} \sum_{(\mathfrak{h})} \mathbf{1}_{\{\mathfrak{h}^{(2)}=\tilde{\mathfrak{h}}\}} \langle \mathfrak{h}, \mathbf{Z}(s)\rangle\langle \mathbf{X}_{s,u}, \mathfrak{h}^{(1)}\rangle\langle \mathbf{X}_{u,t}, [\tilde{\mathfrak{h}}]_i\rangle - \langle \tilde{\mathfrak{h}}, \mathbf{Z}(u)\rangle\langle \mathbf{X}_{u,t}, [\tilde{\mathfrak{h}}]_i\rangle \\
&= \sum_{\tilde{\mathfrak{h}}\in\mathcal{F}^0_{N-1}} \langle \mathbf{X}_{u,t}, [\tilde{\mathfrak{h}}]_i\rangle \left(\langle \mathbf{X}_{s,u} \star \tilde{\mathfrak{h}}, \mathbf{Z}(s)\rangle - \langle \tilde{\mathfrak{h}}, \mathbf{Z}(u)\rangle\right) \\
&= -\sum_{\tilde{\mathfrak{h}}\in\mathcal{F}^0_{N-1}} \langle \mathbf{X}_{u,t}, [\tilde{\mathfrak{h}}]_i\rangle R^{\tilde{\mathfrak{h}}}_{s,u}.
\end{aligned}$$

□

Theorem 2.5 (Gubinelli)**.** *Let $T > 0$, $\mathbf{X}$ be a α-branched rough path and $\mathbf{Z}$ be controlled by $\mathbf{X}$. Then,*

$$\int_s^t Z(r)\, d\mathbf{X}^i(r) := \lim_{|\mathcal{P}|\to 0} \sum_{[u,v]\in\mathcal{P}} \tilde{Z}_{u,v}$$

exists for every $i \in \{1,\dots,d\}$ and $[s,t] \subseteq [0,T]$ where

$$\tilde{Z}_{u,v} := \sum_{\mathfrak{h}\in\mathcal{F}^0_{N-1}} \langle \mathfrak{h}, \mathbf{Z}(u)\rangle \langle \mathbf{X}_{u,v}, [\mathfrak{h}]_i\rangle$$

and $\mathcal{P}$ denotes a partition of $[s,t]$ with mesh size $|\mathcal{P}|$. Moreover, there exists a constant C depending only on α and T such that

$$\left|\int_s^t Z(r)\, d\mathbf{X}^i(r) - \tilde{Z}_{s,t}\right| \le C|t-s|^{(N+1)\alpha} \sum_{\mathfrak{h}\in\mathcal{F}^0_{N-1}} \|\langle \mathbf{X}_{\cdot,\cdot}, [\mathfrak{h}]_i\rangle\|_{(|\mathfrak{h}|+1)\alpha;[s,t]} \|R^{\mathfrak{h}}\|_{(N-|\mathfrak{h}|)\alpha;[s,t]}.$$

Proof. This is a consequence of the sewing lemma [13, Lemma 4.2] and Lemma 2.4. □

The above theorem defines a map which sends a controlled path $\mathbf{Z}$ to a path $t \mapsto \int_0^t Z(r)\, d\mathbf{X}^i(r) \in \mathbb{R}^m$. In fact, this map can be naturally extended to a map $\mathbf{Z} \mapsto \int_0^\cdot \mathbf{Z}(r)\, d\mathbf{X}^i(r)$ where $t \mapsto \int_0^t \mathbf{Z}(r)\, d\mathbf{X}^i(r)$ is a controlled path above $t \mapsto \int_0^t Z(r)\, d\mathbf{X}^i(r)$. To do this, we have to specify $\langle \mathfrak{h}, \int_0^t \mathbf{Z}(r)\, d\mathbf{X}^i(r)\rangle$ for every dual element $\mathfrak{h} \in \mathcal{F}^*_{N-1} \cup \{1\}$. We set

$$\langle 1, \int_0^t \mathbf{Z}(r)\, d\mathbf{X}^i(r)\rangle := \int_0^t Z(r)\, d\mathbf{X}^i(r)$$

and

$$\langle [\tau_1\cdots\tau_n]_i, \int_0^t \mathbf{Z}(r)\, d\mathbf{X}^i(r)\rangle := \langle \tau_1\cdots\tau_n, \mathbf{Z}(t)\rangle.$$

In all other cases, we define

$$\langle \tau_1\cdots\tau_n, \int_0^t \mathbf{Z}(r)\, d\mathbf{X}^i(r)\rangle := 0.$$

More generally, if $\mathbf{Z} = (\mathbf{Z}^1,\dots,\mathbf{Z}^d)$ and every $\mathbf{Z}^i$ is controlled by $\mathbf{X}$, we define a controlled path $t \mapsto \int_0^t \mathbf{Z}(r)\cdot d\mathbf{X}(r)$ by setting

$$\langle 1, \int_0^t \mathbf{Z}(r)\cdot d\mathbf{X}(r)\rangle := \sum_{i=1}^d \langle 1, \int_0^t \mathbf{Z}^i(r)\, d\mathbf{X}^i(r)\rangle,$$

$$\langle [\tau_1\cdots\tau_n]_i, \int_0^t \mathbf{Z}(r)\cdot d\mathbf{X}(r)\rangle := \langle \tau_1\cdots\tau_n, \mathbf{Z}^i(t)\rangle$$

for a tree $[\tau_1\cdots\tau_n]_i \in \mathcal{T}_{N-1}$, $i \in \{1,\dots,d\}$ and

$$\langle \tau_1\cdots\tau_n, \int_0^t \mathbf{Z}(r)\cdot d\mathbf{X}(r)\rangle := 0$$

otherwise.

Theorem 2.6. *The map*

$$I\colon \mathcal{Q}_{\mathbf{X}}(\mathbb{R}^m)^d \to \mathcal{Q}_{\mathbf{X}}(\mathbb{R}^m)$$
$$\mathbf{Z} \mapsto \int_0^{\cdot} \mathbf{Z}(r)\cdot d\mathbf{X}(r)$$

is well-defined and continuous.

Proof. [17, Theorem 8.5]. □

The next lemma shows that controlled paths composed with sufficiently smooth functions are again controlled.

Lemma 2.7. *Let $\phi\colon \mathbb{R}^m \to \mathbb{R}^m$ be sufficiently smooth such that all derivatives below exist. For $\mathbf{Z} \in \mathcal{Q}_{\mathbf{X}}(\mathbb{R}^m)$, we define $\langle 1, \phi(\mathbf{Z}(t))\rangle := \phi(Z(t))$ and*

$$\langle \mathfrak{h}, \phi(\mathbf{Z}(t))\rangle := \sum_{n=1}^{N-1} \sum_{\mathfrak{h}_1\cdots\mathfrak{h}_n=\mathfrak{h}} \frac{1}{n!}\phi^{(n)}(Z(t))\left(\langle \mathfrak{h}_1, \mathbf{Z}(t)\rangle, \dots, \langle \mathfrak{h}_n, \mathbf{Z}(t)\rangle\right)$$

for $\mathfrak{h} \in \mathcal{F}^*_{N-1}$. Then, $\phi(\mathbf{Z}) \in \mathcal{Q}_{\mathbf{X}}(\mathbb{R}^m)$.

Proof. [17, Lemma 8.4]. □

We are now able to say what a solution to (0.2) actually means.

Definition 2.8. A path $y\colon [0,T] \to \mathbb{R}^m$ is a *solution to* (0.2) if $y(t_0) = y_0$ and if there exists a controlled path $\mathbf{Y} \in \mathcal{Q}_{\mathbf{X}}(\mathbb{R}^m)$ above y such that

$$\mathbf{Y}(t) - \mathbf{Y}(s) = \int_s^t f(\mathbf{Y}(r))\cdot d\mathbf{X}(r) \tag{2.2}$$

holds for every $s \le t$, $s,t \in [t_0,T]$, where we set $f(\mathbf{Y}(r)) := (f_1(\mathbf{Y}(r)), \dots f_d(\mathbf{Y}(r)))$.

Proving that (2.2) admits a (unique) solution is done by a standard fixed-point argument [17, Theorem 8.8]. If f is of class $\mathrm{Lip}^{\gamma-1}$ for some $\gamma > \frac{1}{\alpha}$, a local solution to (2.2) exists. If f is $\mathrm{Lip}_b^{\gamma-1}$, the solution exists on every time interval. For f being of class Lip^γ resp. Lip_b^γ, the local resp. global solution is unique. Moreover, in the second case, the solution map is continuous.

Recall the definition of the elementary differential $F(\tau)$ for $\tau \in \mathcal{T}^0$ given in Definition 1.1.

Lemma 2.9. *Let $\mathbf{Y}\colon [0,T] \to \mathcal{H}_{N-1}$ with $y(t) = \langle 1, \mathbf{Y}(t)\rangle$ be a solution to* (2.2)*. Then, the coefficients of $\mathbf{Y}$ are given by*

$$\langle \tau, \mathbf{Y}(t)\rangle = \frac{1}{\sigma(\tau)}F(\tau)(y(t))$$

for $\tau \in \mathcal{T}^*_{N-1} \cup \{1\}$ and $\langle \tau_1\cdots\tau_n, \mathbf{Y}(t)\rangle = 0$ for $\tau_1\cdots\tau_n \in \mathcal{F}^*_{N-1} \setminus \mathcal{T}^*_{N-1}$.

Proof. Being a solution to (2.2) means that

$$y(t) - y(s) = \langle 1, \int_s^t f(\mathbf{Y}(r))\cdot d\mathbf{X}(r)\rangle$$

and

$$\langle [\tau_1\cdots\tau_n]_i, \mathbf{Y}(t)\rangle = \langle [\tau_1\cdots\tau_n]_i, \int_0^t f(\mathbf{Y}(r))\cdot d\mathbf{X}(r)\rangle \tag{2.3}$$

for all $[\tau_1 \cdots \tau_n]_i \in \mathcal{T}^*_{N-1}$, $i \in \{1, \dots, d\}$, and

$$\langle \tau_1 \cdots \tau_n, \mathbf{Y}(t) \rangle = 0$$

for all $\tau_1 \cdots \tau_n \in \mathcal{F}^*_{N-1} \setminus \mathcal{T}^*_{N-1}$. We prove the assertion for all trees $\tau \in \mathcal{T}^*_{N-1}$ by induction on the height of τ. For $\tau = 1$, the claim follows by definition. Now let $\tau = [\tau_1 \cdots \tau_n]_i = [(\tau_1)^{r_1} \cdots (\tau_q)^{r_q}]_i \in \mathcal{T}^*_{N-1}$ for some $i \in \{1, \dots, d\}$ and pairwise distinct trees $\tau_1, \dots, \tau_q$. From (2.3), we have

$$\begin{aligned}
\langle \tau, \mathbf{Y}(t) \rangle &= \langle [\tau_1 \cdots \tau_n]_i, \int_0^t f(\mathbf{Y}(r)) \cdot d\mathbf{X}(r) \rangle \\
&= \langle \tau_1 \cdots \tau_n, f_i(\mathbf{Y}(t)) \rangle \\
&= \sum_{\substack{\lambda_1, \cdots, \lambda_n \in \mathcal{T}^*_{N-2} \\ \lambda_1 \cdots \lambda_n = \tau_1 \cdots \tau_n}} \frac{1}{n!} f_i^{(n)}(y(t)) \left(\langle \lambda_1, \mathbf{Y}(t) \rangle, \dots, \langle \lambda_n, \mathbf{Y}(t) \rangle \right) \\
&= \frac{1}{r_1! \cdots r_q!} \frac{1}{n!} \sum_{\sigma \in \mathrm{sym}(n)} f_i^{(n)}(y(t)) \left(\langle \tau_{\sigma(1)}, \mathbf{Y}(t) \rangle, \dots, \langle \tau_{\sigma(n)}, \mathbf{Y}(t) \rangle \right) \\
&= \frac{1}{r_1! \cdots r_q!} f_i^{(n)}(y(t)) \left(\langle \tau_1, \mathbf{Y}(t) \rangle, \dots, \langle \tau_n, \mathbf{Y}(t) \rangle \right) \\
&= \frac{1}{r_1! \cdots r_q!} \frac{1}{\sigma(\tau_1) \cdots \sigma(\tau_n)} f_i^{(n)}(y(t)) \left(F(\tau_1)(y(t)), \dots, F(\tau_n)(y(t)) \right) \\
&= \frac{1}{\sigma(\tau)} F(\tau)(y(t))
\end{aligned}$$

by induction hypothesis.

□

Theorem 2.10. *Let* $\mathbf{X}$ *be an* α*-branched rough path and* $h > 0$*. Then,* (0.2) *has the expansion*

$$y(t_0 + h) = y_0 + \sum_{\tau \in \mathcal{T}_p} \frac{1}{\sigma(\tau)} F(\tau)(y_0) \langle \mathbf{X}_{t_0, t_0 + h}, \tau \rangle + \mathcal{O}(h^{(p+1)\alpha})$$

for every $p \geq \lfloor 1/\alpha \rfloor$*.*

Proof. Let $s < t$. Note that

$$y(t) - y(s) = \sum_{i=1}^{d} \int_s^t f_i(y(r)) \, d\mathbf{X}^i(r).$$

For $i \in \{1, \dots, d\}$, set

$$\tilde{Z}^i_{s,t} := \sum_{\mathfrak{h} \in \mathcal{F}^0_{p-1}} \langle \mathfrak{h}, f_i(\mathbf{Y}(s)) \rangle \langle \mathbf{X}_{s,t}, [\mathfrak{h}]_i \rangle = \sum_{\mathfrak{h} \in \mathcal{F}^0_{p-1}} \frac{1}{\sigma([\mathfrak{h}]_i)} F([\mathfrak{h}]_i)(y(s)) \langle \mathbf{X}_{s,t}, [\mathfrak{h}]_i \rangle$$

where we use Lemma 2.9 for the equality. We therefore obtain

$$y(t) - y(s) = \sum_{i=1}^{d} \tilde{Z}^i_{s,t} + R_{s,t} = \sum_{\tau \in \mathcal{T}_p} \frac{1}{\sigma(\tau)} F(\tau)(y(s)) \langle \mathbf{X}_{s,t}, \tau \rangle + R_{s,t}$$

where

$$R_{s,t} = \sum_{i=1}^{d} \int_s^t f_i(y(r))\, d\mathbf{X}^i(r) - \tilde{Z}^i_{s,t}.$$

Using Lemma 2.4 and the sewing Lemma [13, Lemma 4.2], we conclude that $R_{s,t}$ is of order $\mathcal{O}((t-s)^{(p+1)\alpha})$.

□

We introduce the local error by $le(t_0, y_0; h) := y(t_0 + h) - y_1$ which is the error of one step with the iterative scheme (1.1) starting in the exact value y_0. Comparing Theorems 1.2 and 2.10 and exploiting Lemma 1.3, we see that the local error is

$$|le(t_0, y_0; h)| = \mathcal{O}(h^{(p+1)\alpha}) \tag{2.4}$$

for sufficiently small $h > 0$ if and only if

$$\langle \mathbf{X}_{t_0,t_0+h}, \tau \rangle = a(\tau)(h) \quad \forall \tau \in \mathcal{T} \text{ with } |\tau| \leq p. \tag{2.5}$$

3. Simplified Runge Kutta Methods

In the following, $\mathbf{X}$ denotes an α-branched rough path for some $\alpha \in (0, 1]$. Assume that there is a smooth path X^h such that its natural lift $\mathbf{X}^h$ to a branched rough path, cf. Example 2.2, approximates $\mathbf{X}$, i.e., $\varrho_\alpha(\mathbf{X}^h, \mathbf{X}) \to 0$ for $h \to 0$. This implies that $\mathbf{X}$ is a *geometric* rough path [19, Section 4] and that $\varrho^g_\alpha(\mathbf{X}^h, \mathbf{X}) \to 0$ where ϱ^g_α denotes the inhomogeneous rough paths metric for geometric rough paths [16]. We introduce the equation associated to the smooth driver by

$$dy^h(t) = f(y^h(t))\, d\mathbf{X}^h(t), \quad y^h(t_0) = y_0. \tag{3.1}$$

This equation can be solved by considering $\mathbf{X}$ as a branched rough path or a geometric rough path, the solution is the same in both cases. It also coincides with the solution to the corresponding Riemann-Stieltjes equation which is well-defined since X^h is smooth by assumption. Since the solution to (0.2) is a locally Lipschitz continuous function of $\mathbf{X}$, cf. [16] in the case of geometric rough paths or [17, Theorem 8.8] for branched rough paths, i.e.,

$$\sup_{t \in [t_0, T]} |y(t) - y^h(t)| \lesssim \varrho^g_\alpha(\mathbf{X}^h, \mathbf{X}), \tag{3.2}$$

we find that y^h is close to y for sufficiently small h. In this section, we restrict ourselves to branched rough paths $\mathbf{X}$ that are limits of lifted piece-wise linear approximations X^h of X. When we replace $\mathbf{X}$ by the lift of a stochastic process, this property holds almost surely, e.g., for semi-martingales, fractional Brownian motions with Hurst index $H > \frac{1}{4}$ and other Gaussian processes [16]. This piece-wise linear approximation to X on some grid $t_0 < t_1 < \ldots < t_N = T$ is constructed as follows:

$$X^h(t) = X(t_k) + \frac{t - t_k}{h_k} \left[X(t_{k+1}) - X(t_k) \right], \quad t \in (t_k, t_{k+1}], \tag{3.3}$$

where $h_k = t_{k+1} - t_k$ and $k = 0, 1, \ldots, N-1$. We assume that this piece-wise linear approximation converges with rate $r_0 > 0$, meaning that

$$\varrho^g_\alpha(\mathbf{X}^h, \mathbf{X}) = \mathcal{O}(h^{r_0})$$

for sufficiently small h, where $h = \max_{k=0,\ldots,N-1} |t_{k+1} - t_k|$.

Example 3.1. In [14], the almost sure convergence rate of $\mathbf{X}^h$ to $\mathbf{X}$ is calculated for the natural lift $\mathbf{X}$ (in the sense of [15]) of a large class of Gaussian processes X in the metric ϱ_α^g. In particular, for the lift of a fractional Brownian motion with Hurst parameter $H \in (1/4, 1, 2]$, one can show that the rate r_0 is arbitrarily close to $2H - 1/2$ almost surely provided one chooses α sufficiently small.

Below, we analyze the order of the local error of some simplified Runge-Kutta scheme if the underlying driver is $\mathbf{X}^h$, considered as α-branched rough path. This scheme is obtained by setting

$$Z_{ij}^{(k)} = a_{ij} X^k_{t_n,t_{n+1}} \quad \text{and} \quad z_i^{(k)} = b_i X^k_{t_n,t_{n+1}} \tag{3.4}$$

in (1.1), where $X^k_{t_n,t_{n+1}}$ denotes the increment of the kth component of X on $[t_n, t_{n+1}]$. Method (1.1) then becomes

$$\begin{aligned} Y_i^h &= y_n^h + \sum_{k=1}^d \sum_{j=1}^s a_{ij} f_k(Y_j^h) X^k_{t_n,t_{n+1}} = y_n^h + \sum_{j=1}^s a_{ij} f(Y_j^h) X_{t_n,t_{n+1}} \\ y_{n+1}^h &= y_n^h + \sum_{k=1}^d \sum_{i=1}^s b_i f_k(Y_i^h) X^k_{t_n,t_{n+1}} = y_n^h + \sum_{i=1}^s b_i f(Y_i^h) X_{t_n,t_{n+1}}, \end{aligned} \tag{3.5}$$

where $\mathcal{A} = (a_{ij})$ is a deterministic matrix and $b = (b_i)$ a deterministic vector. This Runge-Kutta method based on the increments of X was considered in [18] in the context of implicit schemes for equations driven by a certain class of Gaussian processes. We aim to find general conditions on the coefficients b and $\mathcal{A}$ that guarantee the desired order of the local error when approximating (3.1). We begin with a result characterizing the branched rough path if the underlying path is given by (3.3).

Proposition 3.2. *Let $\tau \in \mathcal{T}$ be a tree of order p, i.e, $|\tau| = p$. Then, for the branched rough path associated to the piece-wise linear approximation in* (3.3)*, we have*

$$\langle \mathbf{X}^h_{t_k,t_{k+1}}, \tau \rangle = \frac{1}{\gamma(\tau)} X^{i_1}_{t_k,t_{k+1}} X^{i_2}_{t_k,t_{k+1}} \dots X^{i_p}_{t_k,t_{k+1}}$$

where the $i_\ell \in \{1, 2, \dots, d\}$ are the labels of the tree τ and $X^i_{t_k,t_{k+1}}$ is the increment of the ith component of X on $[t_k, t_{k+1}]$.

Proof. We prove by induction on the height of τ that

$$\langle \mathbf{X}^h_{t_k,t}, \tau \rangle = \frac{1}{\gamma(\tau)} \left(\frac{t - t_k}{h_k} \right)^p X^{i_1}_{t_k,t_{k+1}} X^{i_2}_{t_k,t_{k+1}} \dots X^{i_p}_{t_k,t_{k+1}} \tag{3.6}$$

for $t \in (t_k, t_{k+1}]$. Setting $t = t_{k+1}$ then yields the claim. The identity is true for $\tau = 1$ and $\tau = \bullet_{i_1}$. Let us assume that (3.6) is true for all sub-trees of $\tau = [\tau_1, \dots, \tau_n]_{i_p}$. Then, according to Example 2.2, we have

$$\begin{aligned} \langle \mathbf{X}^h_{t_k,t}, \tau \rangle &= \int_{t_k}^t \langle \mathbf{X}^h_{t_k,u}, \tau_1 \rangle \cdots \langle \mathbf{X}^h_{t_k,u}, \tau_n \rangle \, dX^{h,i_p}(u) = \int_{t_k}^t \langle \mathbf{X}^h_{t_k,u}, \tau_1 \rangle \cdots \langle \mathbf{X}^h_{t_k,u}, \tau_n \rangle \frac{X^{i_p}_{t_k,t_{k+1}}}{h_k} \, du \\ &= \int_{t_k}^t \frac{1}{\gamma(\tau_1)} \left(\frac{u - t_k}{h_k} \right)^{p_1} X^{\tilde{i}_1}_{t_k,t_{k+1}} \cdots X^{\tilde{i}_{p_1}}_{t_k,t_{k+1}} \cdots \frac{1}{\gamma(\tau_n)} \left(\frac{u - t_k}{h_k} \right)^{p_n} X^{\bar{i}_1}_{t_k,t_{k+1}} \cdots X^{\bar{i}_{p_n}}_{t_k,t_{k+1}} \frac{X^{i_p}_{t_k,t_{k+1}}}{h_k} \, du, \end{aligned}$$

where $p_i := |\tau_i|$ $(i = 1, \dots, n)$. Since $\sum_{i=1}^n p_i = p - 1$ and $\Pi_{i=1}^n \frac{1}{\gamma(\tau_i)} = \frac{p}{\gamma(\tau)}$, we obtain

$$\langle \mathbf{X}^h_{t_k,t}, \tau \rangle = \int_{t_k}^t \frac{p}{\gamma(\tau)} \left(\frac{u - t_k}{h_k} \right)^{p-1} \frac{1}{h_k} X^{i_1}_{t_k,t_{k+1}} \cdots X^{i_p}_{t_k,t_{k+1}} \, du = \frac{1}{\gamma(\tau)} \left(\frac{t - t_k}{h_k} \right)^p X^{i_1}_{t_k,t_{k+1}} \cdots X^{i_p}_{t_k,t_{k+1}}$$

which concludes the proof of this proposition. □

The local error of the simplified Runge-Kutta scheme applied to (3.1) is defined as $le^h(t_0, y_0; h) := y^h(t_0 + h) - y^h_1$. We can now rewrite (2.4) and (2.5) using Proposition 3.2. Moreover, within the series representation given in Theorem 1.2, $a(\tau)(h)$ is replaced by $a^h(\tau)(h)$ if the simplifying ansatz (3.4) is used. Now, the order of the local error of (3.5) is

$$|le^h(t_0, y_0; h)| = \mathcal{O}(h^{(p+1)\alpha}) \tag{3.7}$$

for sufficiently small $h > 0$ if and only if

$$\frac{1}{\gamma(\tau)} X^{i_1}_{t_0,t_0+h} X^{i_2}_{t_0,t_0+h} \dots X^{i_{|\tau|}}_{t_0,t_0+h} = a^h(\tau)(h) \quad \forall \tau \in \mathcal{T} \text{ with } |\tau| \le p, \tag{3.8}$$

where α is the Hölder regularity of X. Based on (3.8), we aim to find proper choices of $\mathcal{A}$ and b in (3.5) that provide the desired local rate in (3.7). In order to simplify the notation in the result below, we introduce

$$c_i := \sum_{j=1}^s a_{ij}. \tag{3.9}$$

We now formulated conditions for the order of the local error associated to the simplified Runge-Kutta scheme.

Theorem 3.3. *The simplified Runge-Kutta method* (3.5) *approximating* (3.1) *has a local error of order* $(p + 1)\alpha$*, i.e.,*

$$|le^h(t_0, y_0; h)| = \mathcal{O}(h^{(p+1)\alpha})$$

if and only if the following conditions are satisfied for all $\ell = 1, \dots, p$*:*

ℓ	
1	$\sum_{i=1}^s b_i = 1$
2	$\sum_{i=1}^s b_i c_i = \frac{1}{2}$
3	$\sum_{i=1}^s b_i c_i^2 = \frac{1}{3}$, $\quad \sum_{i=1}^s \sum_{j=1}^s b_i a_{ij} c_j = \frac{1}{6}$

TABLE 1. Algebraic conditions for the local error of the simplified Runge-Kutta method.

Proof. Let $i_1, i_2, i_3 \in \{1, \dots, d\}$. We start analyzing (3.8) for all trees of order one, i.e., $\tau = \bullet_{i_1}$. Using the definition of $a^h(\tau)(h)$, i.e., we plug in (3.4) in the definition of $a(\tau)(h)$ (see (1.2)), we obtain

$$a^h(\bullet_{i_1})(h) := \langle z^{(i_1)}, \Phi(1)(h)\rangle = \sum_{i=1}^{s} z_i^{(i_1)} = \sum_{i=1}^{s} b_i X^{i_1}_{t_0,t_0+h}.$$

Inserting this into (3.8), we find

$$\frac{1}{\gamma(\bullet_{i_1})} X^{i_1}_{t_0,t_0+h} = \sum_{i=1}^{s} b_i X^{i_1}_{t_0,t_0+h}$$

which is equivalent to $\sum_{i=1}^{s} b_i = 1$. We continue with the trees of order two. These are of the form $\tau = [\bullet_{i_2}]_{i_1}$. Again, we determine $a^h(\tau)(h)$ which is

$$a^h(\tau)(h) := \langle z^{(i_1)}, \Phi(\bullet_{i_2})(h)\rangle = \langle z^{(i_1)}, Z^{(i_2)}\Phi(1)(h)\rangle = \langle b, \mathcal{A}\mathbf{1}_s\rangle X^{i_1}_{t_0,t_0+h} X^{i_2}_{t_0,t_0+h},$$

using the representations in (3.4). With this expression for $a^h(\tau)(h)$, (3.8) becomes

$$\frac{1}{2} X^{i_1}_{t_0,t_0+h} X^{i_2}_{t_0,t_0+h} = \langle b, \mathcal{A}\mathbf{1}_s\rangle X^{i_1}_{t_0,t_0+h} X^{i_2}_{t_0,t_0+h}$$

exploiting that $\gamma(\tau) = 2$. This is equivalent to $\sum_{i=1}^{s} b_i c_i = \frac{1}{2}$. We conclude this proof by considering the order three trees. We start with trees of the form $\tau = [[\bullet_{i_3}]_{i_2}]_{i_1}$. Then, $a^h(\tau)(h)$ is

$$\begin{aligned} a^h(\tau)(h) &= \langle z^{(i_1)}, \Phi([\bullet_{i_3}]_{i_2})(h)\rangle = \langle z^{(i_1)}, Z^{(i_2)}\Phi(\bullet_{i_3})(h)\rangle = \langle z^{(i_1)}, Z^{(i_2)}Z^{(i_3)}\mathbf{1}_s\rangle \\ &= \langle b, \mathcal{A}(\mathcal{A}\mathbf{1}_s)\rangle X^{i_1}_{t_0,t_0+h} X^{i_2}_{t_0,t_0+h} X^{i_3}_{t_0,t_0+h}. \end{aligned}$$

Moreover, we see that $\gamma(\tau) = 6$. Using the above, (3.8) for $\tau = [[\bullet_{i_3}]_{i_2}]_{i_1}$ is equivalent to

$$\frac{1}{6} = \langle b, \mathcal{A}(\mathcal{A}\mathbf{1}_s)\rangle = \sum_{i=1}^{s}\sum_{j=1}^{s} b_i a_{ij} c_j.$$

Now, the only type of tree left is the branched tree $\tau = [\bullet_{i_2}, \bullet_{i_3}]_{i_1}$. The corresponding $a^h(\tau)(h)$ is

$$\begin{aligned} a^h(\tau)(h) &= \langle z^{(i_1)}, \Phi(\bullet_{i_2})(h)\Phi(\bullet_{i_3})(h)\rangle = \langle z^{(i_1)}, Z^{(i_2)}\mathbf{1}_s Z^{(i_3)}\mathbf{1}_s\rangle \\ &= \langle b, (\mathcal{A}\mathbf{1}_s)(\mathcal{A}\mathbf{1}_s)\rangle X^{i_1}_{t_0,t_0+h} X^{i_2}_{t_0,t_0+h} X^{i_3}_{t_0,t_0+h}. \end{aligned}$$

Notice that the product of two vectors is meant component-wise. For this tree, (3.8) therefore is equivalent to

$$\frac{1}{3} = \frac{1}{\gamma(\tau)} = \langle b, (\mathcal{A}\mathbf{1}_s)(\mathcal{A}\mathbf{1}_s)\rangle = \sum_{i=1}^{s} b_i c_i^2$$

which finally proves the claim. □

Remark 3.4. In fact, we can easily find algebraic conditions for any $\ell > 3$ in Table 1 by considering trees $\tau \in \mathcal{T}$ with $|\tau| > 3$ in (3.8). This means that we can achieve a local rate of $(p+1)\alpha$ for the simplified Runge-Kutta method for arbitrary $p \in \mathbb{N}$.

The conditions given in Table 1 are nothing but the consistency conditions known for the case of $f \equiv 0$ in (0.1), see, e.g., [20]. The consistency order is the order in the step size h of the expression $\frac{le(t_0,x_0;h)}{h}$. If all conditions in Table 1 are fulfilled, then one has a scheme of consistency order 3 assuming $f \equiv 0$ in (0.1). Such 3rd order schemes are well-studied in the ordinary differential equation scenario. Below, we provide just a few examples that satisfy these conditions.

Example 3.5. We introduce the general Butcher-scheme:

$$\mathcal{BS} := \begin{array}{c|ccc} a_{11} & a_{12} & \dots & a_{s1} \\ a_{21} & a_{22} & \dots & a_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ a_{s1} & a_{s2} & \cdots & a_{ss} \\ \hline b_1 & b_2 & \cdots & b_s \end{array}.$$

Notice that the node values c_i, defined in (3.9), are only used to formulate the order conditions in Theorem 3.3. However, they are not required to characterize Runge-Kutta methods for autonomous systems considered in this paper. Therefore, we omit them in the Butcher tableau.

(i) An explicit Runge-Kutta scheme satisfying all conditions in Table 1 is Heun's third-order method:

$$\mathcal{BS} = \begin{array}{c|ccc} 0 & 0 & 0 \\ 1/3 & 0 & 0 \\ 0 & 2/3 & 0 \\ \hline 1/4 & 0 & 3/4 \end{array}.$$

Hence, the iterative scheme (3.5) is

$$y^h_{n+1} = y^h_n + \frac{1}{4}[f(y^h_n) + 3f(Y^h_3)]X_{t_n,t_{n+1}}$$

where Y^h_3 is given by

$$Y^h_3 = y^h_n + \frac{2}{3}f(Y^h_2)X_{t_n,t_{n+1}} \quad \text{with} \quad Y^h_2 = y^h_n + \frac{1}{3}f(y^h_n)X_{t_n,t_{n+1}}.$$

(ii) Another explicit method fulfilling the conditions in Table 1 is Kutta's third order scheme:

$$\mathcal{BS} = \begin{array}{c|ccc} 0 & 0 & 0 \\ 1/2 & 0 & 0 \\ -1 & 2 & 0 \\ \hline 1/6 & 2/3 & 1/6 \end{array}.$$

Consequently, the simplified Runge-Kutta method is

$$y^h_{n+1} = y^h_n + \frac{1}{6}[f(y^h_n) + 4f(Y^h_2) + f(Y^h_3)]X_{t_n,t_{n+1}},$$

where Y^h_2 and Y^h_3 are computed by

$$Y^h_2 = y^h_n + \frac{1}{2}f(y^h_n)X_{t_n,t_{n+1}} \quad \text{and} \quad Y^h_3 = y^h_n + [-f(y^h_n) + 2f(Y^h_2)]X_{t_n,t_{n+1}}.$$

Notice that there are many other schemes satisfying the above conditions, e.g., [18, Corollary 5.1] provide two implicit Runge-Kutta methods (for stochastic differential equations driven by a certain class of Gaussian processes) that satisfy the requirements in Table 1.

4. Global Rates

4.1. Global rate of the full Runge-Kutta scheme. Let y^{s,y_0}_t denote the solution to (0.2) at time t starting in y_0 at s, i.e., $y^{s,y_0}_s = y_0$. Let a numerical scheme be given as the following one step method:

$$y_{n+1} = y_n + \Phi(y_n, \mathbf{X}_{t_n,t_{n+1}}),$$

where $t_0 < t_1 < \ldots < t_N = T$ is a partition of $[t_0, T]$. Below, we analyze the order of convergence of the numerical method (1.1). The next proposition shows that we loose one order from the local to the global error.

Proposition 4.1. *If there is a constant $C_1 > 0$ such that*

$$|y_t^{s,y_0} - y_0 - \Phi(y_0, \mathbf{X}_{s,t})| \leq C_1 |t-s|^{1+r} \tag{4.1}$$

for $|t-s|$ being sufficiently small and if

$$|y_t^{s,y_0} - y_t^{s,\tilde{y}_0}| \leq C_2 |y_0 - \tilde{y}_0| \tag{4.2}$$

for some constant $C_2 > 0$, where $y_0, \tilde{y}_0 \in \mathbb{R}^m$ and $0 \leq s \leq t \leq T$. Then, there is some $C > 0$ such that

$$\max_{k=0,\ldots,N} |y(t_k) - y_k| \leq C h^r$$

for $r > 0$, where $h = \max_{k=0,\ldots,N-1} |t_{k+1} - t_k|$.

Proof. We write the global error as follows

$$y_k - y(t_k) = \sum_{j=0}^{k-1} \left(y_{t_k}^{t_{j+1}, y_{j+1}} - y_{t_k}^{t_j, y_j} \right) \tag{4.3}$$

using that $y_{t_k}^{t_0,y_0} = y(t_k)$ and $y_{t_k}^{t_k,y_k} = y_k$. We combine

$$y_{t_k}^{t_j,y_j} = y_{t_k}^{t_{j+1}, y_{t_{j+1}}^{t_j,y_j}} \quad \text{and} \quad y_{j+1} = y_j + \Phi(y_j, \mathbf{X}_{t_j,t_{j+1}})$$

with (4.3) which yields

$$\begin{aligned} |y(t_k) - y_k| &= \sum_{j=0}^{k-1} |y_{t_k}^{t_{j+1}, y_j + \Phi(y_j, \mathbf{X}_{t_j,t_{j+1}})} - y_{t_k}^{t_{j+1}, y_{t_{j+1}}^{t_j,y_j}}| \leq C_2 \sum_{j=0}^{k-1} |y_j + \Phi(y_j, \mathbf{X}_{t_j,t_{j+1}}) - y_{t_{j+1}}^{t_j,y_j}| \\ &\leq C_1 C_2 \sum_{j=0}^{k-1} |t_{j+1} - t_j|^{r+1} \leq C_1 C_2 h^r \sum_{j=0}^{k-1} (t_{j+1} - t_j) \leq C_1 C_2 (T - t_0) h^r \end{aligned}$$

exploiting assumptions (4.1) and (4.2). This concludes the proof of this proposition. □

4.2. Global rate of the simplified Runge-Kutta scheme. In this section, we study a particular case of $\mathbf{X}$ being an α-Hölder geometric rough path, $0 < \alpha \leq 1$, that can be approximated by the lift of its piece-wise linear approximated underlying path X. For such driver, the simplified Runge-Kutta method (3.5) converges. Its order is shown in the following theorem.

Theorem 4.2. *Let $\mathbf{X}$ be an α-Hölder geometric rough path in (0.2), $0 < \alpha \leq 1$, and let its piece-wise linear approximation X^h be given by (3.3). We assume that the Wong-Zakai approximation converges with rate $r_0 > 0$, meaning that*

$$\sup_{t \in [t_0,T]} |y(t) - y^h(t)| = \mathcal{O}(h^{r_0}) \tag{4.4}$$

for sufficiently small h, where $h = \max_{k=0,\ldots,N-1} |t_{k+1} - t_k|$ is the maximal step size of the underlying grid, y and y^h are the solutions to (0.2) and (3.1), respectively. If all conditions in Table 1

are satisfied and the right hand side f is of class Lip_b^γ for some $\gamma > \frac{1}{\alpha}$, then the simplified Runge-Kutta method (3.5) *converges with rate $\eta = \min\{r_0, 4\alpha - 1\}$ to the solution of* (0.2)*, i.e., there is a constant $C > 0$ such that*

$$\max_{k=0,\dots,N} |y(t_k) - y_k^h| \le Ch^\eta$$

for sufficiently small h.

Proof. It holds that

$$|y(t_k) - y_k^h| \le \sup_{t\in[t_0,T]} |y(t) - y^h(t)| + \max_{k=0,\dots,N} |y^h(t_k) - y_k^h|.$$

Theorem 3.3 gives us a rate of 4α for the local error of simplified Runge-Kutta method. Proposition 4.1 now provides that

$$\max_{k=0,\dots,N} |y^h(t_k) - y_k^h| = \mathcal{O}(h^{4\alpha-1})$$

if assumption (4.2) holds true. Let y_t^{h,s,y_0} denote the solution to (3.1) with initial time s and initial state y_0. Since $\mathbf{X}$ is α Hölder and since $\mathbf{X}^h$ is convergent and hence bounded in h, there is a constant $K > 0$ independent of h such that

$$|y_t^{h,s,y_0} - y_t^{h,s,\tilde{y}_0}| \le K\, \mathrm{e}^{K|t-s|^\alpha}\, |y_0 - \tilde{y}_0|,$$

cf. [16]. This implies (4.2) and concludes the proof. □

Remark 4.3.
(i) From (3.2), a sufficient condition for (4.4) is that $\varrho_{\alpha'}^g(\mathbf{X}^h, \mathbf{X}) = \mathcal{O}(h^{r_0})$ for some $0 < \alpha' \le \alpha$ in which case one has to assume $f \in \mathrm{Lip}_b^\gamma$ for some $\gamma > \frac{1}{\alpha'}$.
(ii) Theorem 4.2 is formulated for any roughness parameter $\alpha > 0$. For a fractional Brownian motion with Hurst parameter $H \in (1/4, 1)$, it gives an optimal rate of convergence in the case when $H \in (1/4, 1/2]$. Indeed, from [14], we know that r_0 can be chosen arbitrarily close to $2H - 1/2$. Since $2H - 1/2 < 4H - 1$, the convergence rate of the simplified Runge-Kutta scheme is arbitrarily close to $2H - 1/2$ almost surely. This rate is the same as for the simplified Milstein scheme introduced in [11], cf. [14], which is believed to be optimal due to the results obtained in [29].

5. Numerical Experiments

We illustrate the rate of convergence of Heun's method (presented in Example 3.5) applied to (0.2), where $\mathbf{X}$ is the geometric lift of a path of a two-dimensional fractional Brownian motion X with independent components and Hurst index H. Notice that we also conducted all numerical experiments for Kutta's third order scheme but these results are omitted since they are basically the same as for the Heun method. This is because the error of both schemes is mainly determined by the same Wong-Zakai approximation. In the first example, $f_1(y) = \cos(y)$, $f_2(y) = \sin(y)$ and $y(t) \in \mathbb{R}$ are chosen resulting in

$$dy(t) = \cos(y(t))\, d\mathbf{X}^1(t) + \sin(y(t))\, d\mathbf{X}^2(t), \quad y(0) = 1, \quad t \in [0, T]. \tag{5.1}$$

Equation (5.1) was considered before by Deya, Neuenkirch and Tindel in [11] in the context of rates of convergence for a Milstein scheme. We now use equidistant grid points, i.e., $h = \frac{T}{N}$ and fix $T = 0.25$. The only reason for considering such a terminal time T is that we aim for a very

small h within the numerical investigations of the rate of convergence. We determine the maximal discretization error

$$\mathcal{E}(h) := \max_{k=0,\dots,N} |y(t_k) - y_k^h|$$

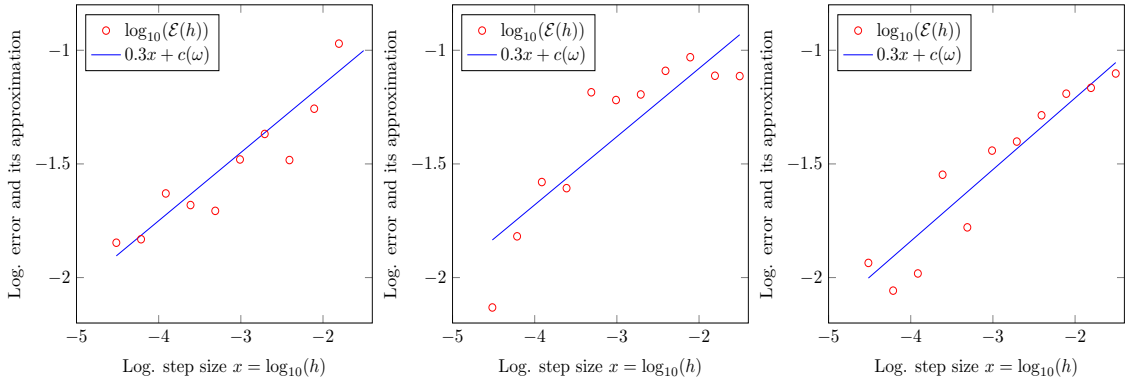
for different h, where y_k^h is the kth iterate of the simplified scheme in (3.5) with coefficients defined in Example 3.5 (i). There is no explicit representation for the solution to (5.1). Therefore, we create a reference solution based on Heun's method with step size $h = 2^{-16}$. Here, the reference solution is calculated with the same numerical scheme since there is no well-established other method for such settings that can be used instead. In Figure 1, the red circles show the logarithmic discretization error $\log_{10}(\mathcal{E}(h))$ in dependence of the logarithmic step size $\log_{10}(h)$ for paths X of fractional Brownian motions with different Hurst parameters.The straight blue lines with slope $2H - 0.5$ represent the expected order of convergence. These rates are confirmed in Figure 1 up to an acceptable deviation between the red circles and the lines. This deviation depends on the rates, e.g., we have a larger gap for $H = 0.4$ (Figure 1a). This can be explained by the roughness of the problem and the associated low order of convergence requiring a very small h. Therefore, the asymptotic behaviour cannot be observed very well with a reasonable underlying step size. This effect can, e.g., also be seen in the numerics section of [11]. Increasing the order of convergence by enlarging H (Figures 1b and 1c), a good line fitting is obtained and the asymptotics are visible. It is worth noticing that the case $H = 0.7$ in Figure 1c was not addressed in Remark 4.3. Still, the numerical experiment shows that the rate of the Heun method remains $2H - 0.5$.

In order to give some more insights why the error asymptotics of the Heun method are hard to observe especially for small orders of convergence, let us study another example with the same type of drivers but with a convergence rate that it is not influenced by the Wong-Zakai approximation. This means that we numerical investigate the rate of Heun's method applied to (3.1) with a driver as in (3.3). We set $f_1(y) = A_1 y$, $f_2(y) = A_2 y$, $A_1, A_2 \in \mathbb{R}^{2\times 2}$, $y(t) \in \mathbb{R}^2$ in (0.2) and obtain

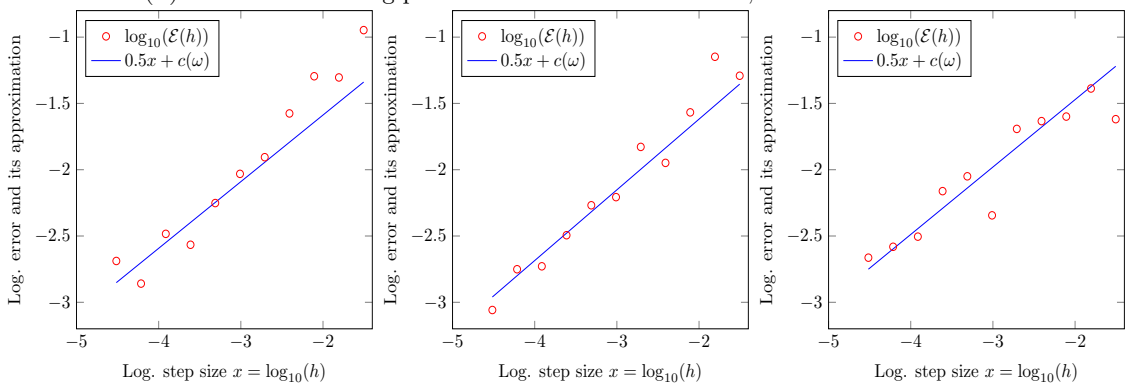
$$dy(t) = A_1 y(t)\, d\mathbf{X}^1(t) + A_2 y(t)\, d\mathbf{X}^2(t), \quad y(0) = \begin{pmatrix} 0.5 & 1 \end{pmatrix}^\top, \quad t \in [0,T], \tag{5.2}$$

where $A_1 = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ and $A_2 = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ are commuting matrices. This choice provides the existence of an explicit solution $y(t) = \exp\{A_1 X^1(t) + A_2 X^2(t)\} y(0)$ which can be checked by applying the chain rule. Since this function does not depend on iterated integrals, we obtain a higher order of convergence than in the previous example given the same type of drivers. In fact, the lines in Figure 2 represent a rate of $4H - 1$ which is the one of the Heun scheme applied to (3.1). Due to a low line fitting error for sufficiently small h we can well observe the error asymptotics for each $H = 0.4, 0.5, 0.7$. In particular, we have a regularity of $H = 0.4$ and a rate of 0.6 in Figure 2a which is significantly lower than in Figure 1c ($H = 0.7$, rate 0.9). Still the line fitting in Figure 2a is better for $h < 10^{-3}$. Therefore, it seems that not solely the rates of convergence (or the roughness of the problem) but the asymptotic behaviour of iterated integral approximations cause a limited observability of the discretization scheme asymptotics.

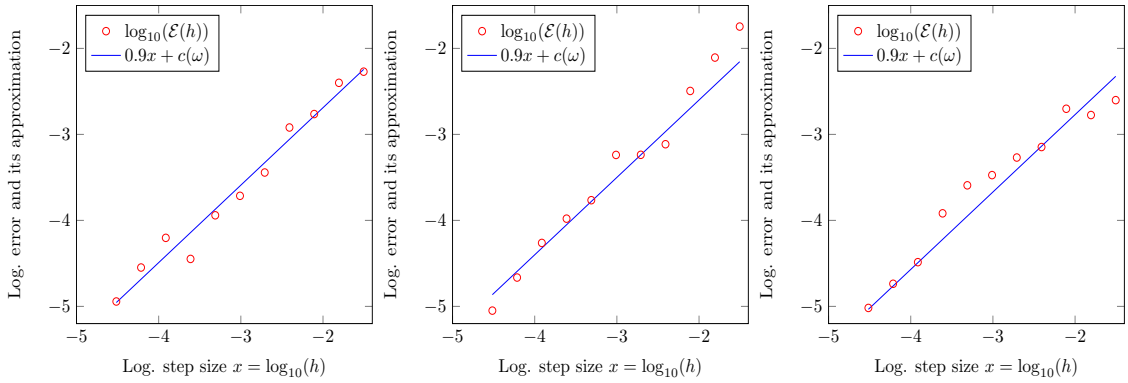
Acknowledgments. SR is supported by the MATH+ project AA4-2 *Optimal control in energy markets using rough analysis and deep networks.* Work on this paper was started while MR and SR were supported by the DFG via Research Unit FOR 2402. Both authors would like to thank Rosa Preiß for related discussions and for providing us with her Master thesis in which she corrected the



(A) Three fixed driving paths with Hurst index $H = 0.4$; rate $2H - 0.5 = 0.3$.

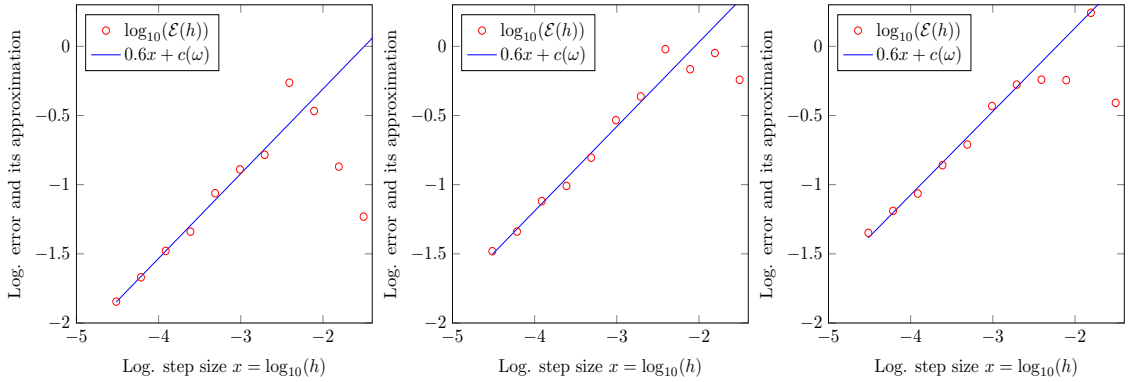


(B) Three fixed driving paths with Hurst index $H = 0.5$; rate $2H - 0.5 = 0.5$.

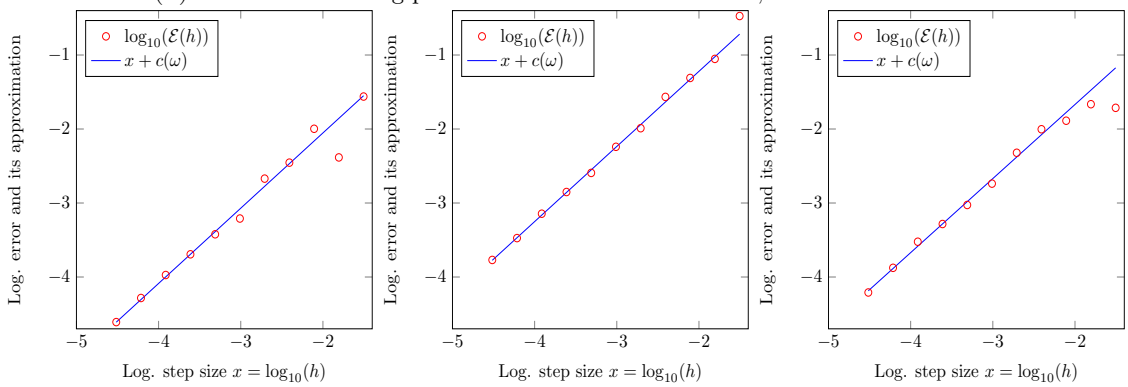


(C) Three fixed driving paths with Hurst index $H = 0.7$; rate $2H - 0.5 = 0.9$.

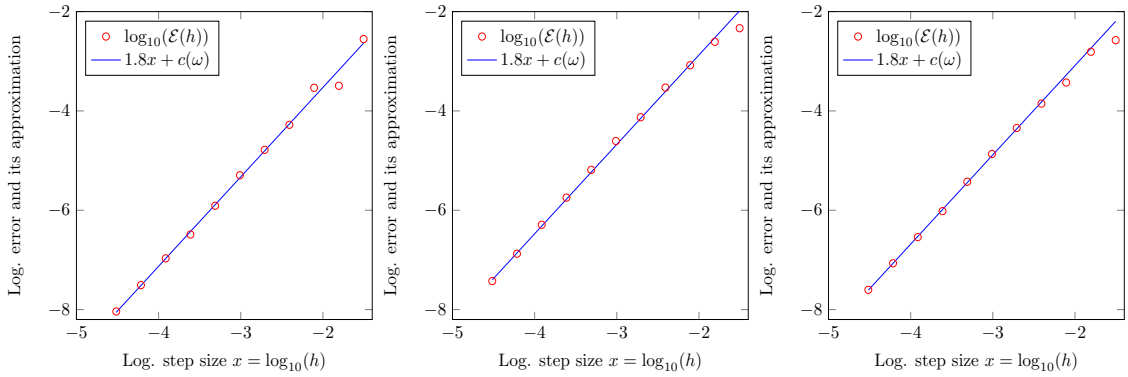
FIGURE 1. Maximum discretization error of Heun method applied to (5.1) driven by paths of fractional Brownian motions with different Hurst indices.



(A) Three fixed driving paths with Hurst index $H = 0.4$; rate $4H - 1 = 0.6$.



(B) Three fixed driving paths with Hurst index $H = 0.5$; rate $4H - 1 = 1$.



(C) Three fixed driving paths with Hurst index $H = 0.7$; and rate $4H - 1 = 1.8$.

FIGURE 2. Maximum discretization error of Heun method applied to (5.2) driven by paths of fractional Brownian motions with different Hurst indices.

proof of [19, Proposition 3.8]. Finally, MR is supported by the DFG via the individual grant "Low-order approximations for large-scale problems arising in the context of high-dimensional PDEs and spatially discretized SPDEs".

References

1. Eiichi Abe. *Hopf algebras*, volume 74 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge-New York, 1980. Translated from the Japanese by Hisae Kinoshita and Hiroko Tanaka.
2. Ismaël Bailleul. On the definition of a solution to a rough differential equation. *arXiv:1803.06479v3*, 2019.
3. Kevin Burrage and Pamela M. Burrage. Order Conditions of Stochastic Runge–Kutta Methods by B-Series. *SIAM Journal on Numerical Analysis*, 38(5):1626–1646, 2000.
4. Christian Bayer, Denis Belomestny, Martin Redmann, Sebastian Riedel, and John Schoenmakers. Solving linear parabolic rough partial differential equations. *Journal of Mathematical Analysis and Applications*, 490(1), 2020.
5. 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.
6. John C. Butcher. *The Numerical Analysis of Ordinary Differential Equations: Runge-Kutta and General Linear Methods*. Wiley-Interscience, USA, 1987.
7. Alain Connes and Dirk Kreimer. Hopf algebras, renormalization and noncommutative geometry. *Comm. Math. Phys.*, 199(1):203–242, 1998.
8. Thomas Cass and Martin P. Weidner. Tree algebras over topological vector spaces in rough path theory. *arXiv:1604.07352v2*, 2017.
9. 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.
10. Kristian Debrabant and Anne Kværnø. B-series analysis of stochastic Runge-Kutta methods that use an iterative scheme to compute their internal stage values. *SIAM J. Numer. Anal.*, 47(1):181–203, 2008/09.
11. Aurélien Deya, Andreas Neuenkirch, and Samy 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.
12. Peter K. Friz, Benjamin Gess, Archil Gulisashvili, and Sebastian 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.
13. Peter K. Friz and Martin Hairer. *A Course on Rough Paths with an introduction to regularity structures*, volume XIV of *Universitext*. Springer, Berlin, 2014.
14. Peter K. Friz and Sebastian Riedel. Convergence rates for the full Gaussian rough paths. *Ann. Inst. Henri Poincaré Probab. Stat.*, 50(1):154–194, 2014.
15. Peter K. Friz and Nicolas B. Victoir. Differential equations driven by Gaussian signals. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(2):369–413, 2010.
16. Peter K. Friz and Nicolas B. Victoir. *Multidimensional stochastic processes as rough paths*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Theory and applications.
17. Massimiliano Gubinelli. Ramification of rough paths. *J. Differential Equations*, 248(4):693–721, 2010.
18. 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.
19. Martin Hairer and David Kelly. Geometric versus non-geometric rough paths. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(1):207–251, 2015.
20. Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations; 2nd ed.*, volume 31. Springer, 2006.
21. Ernst Hairer, Syvert Paul Nørsett, and Gerhard Wanner. *Solving ordinary differential equations. I: Nonstiff problems. 2nd revised ed., 3rd corrected printing*. Springer, Berlin, 2010.
22. Ernst Hairer and Gerhard Wanner. *Solving ordinary differential equations. II: Stiff and differential-algebraic problems. Reprint of the 1996 2nd revised ed.* Springer, Berlin, 2010.
23. Peter Eris Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 21 of *Applications of Mathematics*. Springer-Verlag, Berlin, 2nd edition, 1999.
24. 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.
25. Terry J. Lyons and Zhongmin Qian. *System control and rough paths*. Oxford Mathematical Monographs. Oxford University Press, Oxford, 2002. Oxford Science Publications.
26. Terry J. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana*, 14(2):215–310, 1998.
27. Dominique Manchon. Hopf algebras, from basics to applications to renormalization. *arXiv:math/0408405v2*, 2006.
28. Grigori N. Milstein and Michael V. Tretyakov. *Stochastic numerics for mathematical physics*. Scientific Computation. Berlin: Springer. ixx, 594 p., 2004.
29. Andreas Neuenkirch, Samy Tindel, and Jérémie Unterberger. Discretizing the fractional Lévy area. *Stochastic Process. Appl.*, 120(2):223–254, 2010.
30. Andreas Rößler. Runge-Kutta methods for the strong approximation of solutions of stochastic differential equations. *SIAM J. Numer. Anal.*, 48(3):922–952, 2010.
31. Moss E. Sweedler. *Hopf algebras*. Mathematics Lecture Note Series. W. A. Benjamin, Inc., New York, 1969.

Martin Redmann: Martin Luther University Halle-Wittenberg, Institute of Mathematics, Theodor-Lieser-Str. 5, 06120 Halle (Saale), Germany
Email address: martin.redmann@mathematik.uni-halle.de

Sebastian Riedel: FernUniversität in Hagen, Faculty of Mathematics and Computer Science, Chair of Applied Stochastics, 58084 Hagen, Germany
Email address: sebastian.riedel@fernuni-hagen.de