HAL open science

Constructing general rough differential equations through flow approximations

Antoine Lejay

To cite this version:

Antoine Lejay. Constructing general rough differential equations through flow approximations. 2020. hal-02871886v1

HAL Id: hal-02871886

https://inria.hal.science/hal-02871886v1

Preprint submitted on 17 Jun 2020 (v1), last revised 13 Jan 2022 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Constructing General Rough Differential Equations through Flow Approximations

Antoine Lejay*

June 18, 2020

The non-linear sewing lemma constructs flows of rough differential equations from a braod class of approximations called almost flows. We consider a class of almost flows that could be approximated by solutions of ordinary differential equations, in the spirit of the backward error analysis. Mixing algebra and analysis, a Taylor formula with remainder and a composition formula are central in the expansion analysis. With a suitable algebraic structure on the non-smooth vector fields to be integrated, we recover in a single framework several results regarding high-order expansions for various kind of driving paths. We also extend the notion of driving rough path. We also introduce as an example a new family of branched rough paths, called aromatic rough paths modeled after aromatic Butcher series.

Keywords: Rough differential equations; Branched rough paths; Aromatic Butcher series.

1. Introduction

Introduced at the end of the 1990s by T. Lyons [36], the theory of rough paths defines pathwise solutions to stochastic differential equations driven by Brownian

*Université de Lorraine, CNRS, Inria, IECL, F-54000 Nancy, France, antoine.lejay@univ-lorraine.fr

paths and more generally by a large class irregular paths, random or deterministic. The core idea is that the Brownian paths need to be lifted as a path living in a non-commutative truncated tensor algebra. For general, irregular path, the order of truncation to consider depends on the regularity of the path.

Rough differential equations (RDE) generalize the notion of controlled differential equations in which the controls, or driving paths, are irregular in space, *e.g.,* α-Hölder. Such paths are called *rough paths*. Since the seminal work [36], several concurrent constructions have been given to the existence of solutions of RDE: by solving a fixed point [36, 35, 33], as limit of discrete approximations by A. M. Davie [22], by approximating the driving paths with geodesics, by P. Friz and N. Victoir [25, 26], or by approximating the flows by I. Bailleul [2, 4].

In the recent series of works [15, 14, 16], we have studied the various properties of the non-linear sewing lemma which consider an abstract family of approximation of flows. This extend the results in [2, 4] and also encompass the approaches from A.M. Davie and P. Friz & N. Victoir [22, 26]. The main point of the non-linear sewing lemma is to construct approximations of flows which are "rectified" as flows.

A rough path of order n is a path living in a truncated tensor algebra $\mathrm{T}_{\leqslant n}(\mathrm{U}) := \bigoplus_{k=0}^{n} \mathrm{U}^{\otimes k}$. A broad family of rough paths may be constructed from approximating smooth paths through by their iterated integrals. It shares the same algebraic properties as the iterated integrals that K.T. Chen pointed out in the 1950s [20]. Some controls of the regularity are also needed in accordance with the grading of the tensor space. In view of performing integration, several alternative and extension of rough paths have been given. The first one is that of *controlled rough paths* by M. Gubinelli [28]. Later, M. Gubinelli introduced the notion of *branched rough paths* [29] which are encoded through trees.

The notion of branched rough paths and corresponding RDE has subsequently been studied in [31, 44, 8, 9, 19]. In [19], a solution to RDE driven by branched rough paths have been given that extends the one of A.M. Davie: a solution is a path $y : [0, T] \to \mathrm{V}$ characterized by

$$|y_t - \phi_{t,s}(y_s)| \leqslant C|t - s|^{\theta} \text{ for some } \theta > 1 \tag{1}$$

for a proper approximation $\phi_{t,s}$ constructed using the *elementary differentials* already used in the context of the studies of B-series and Runge-Kutta scheme [30]. The article [41] study a Runge-Kutta scheme for RDE, while [5, 17] use algebraic results on Taylor series, B-series and the Faà di Bruno formula as a core. A variant to branched rough paths consists in using planar trees [21]. While these articles focus mainly on the properties of the driving paths, the algebraic properties of the vector fields play a large role there.

In a generic way, we can see a rough path as a path $\mathbf{x} : [0,T] \to \mathscr{W}$ taking its values in a group of an algebra $\mathscr{W}$ (either a tensor algebra or an algebra of trees). Equivalently, we can see $\mathbf{x}$ as a $\mathscr{W}$-valued family $\{\mathbf{x}_{s,t}\}_{s\leqslant t}$ satisfying the *Chen relation*

$$\mathbf{x}_{r,s}\mathbf{x}_{s,t} = \mathbf{x}_{r,t} \text{ for any } r \leqslant s \leqslant t. \tag{2}$$

Let us now consider either an algebra homomorphism F from $\mathscr{W}$ to the space $\mathscr{C}$ of continuous functions from a Banach space V into itself with compositions We assume that $F[1] = \mathrm{id}$, the identity map. Such a F is called a *character*. The Chen relation (2) translates into

$$F[\mathbf{x}_{s,t}] = F[\mathbf{x}_{r,s}] \circ F[\mathbf{x}_{s,t}] \tag{3}$$

This relation implies that $F[\mathbf{x}_{s,t}](a)$ is a *flow*.

The driving path is not necessarily constructed as satisfying the Chen relation in $\mathscr{W}$, but for example in a quotient space $\widetilde{\mathscr{W}} := \mathscr{W}/\mathscr{I}$ for an ideal $\mathscr{I}$. Truncated tensor algebras are of this type. In such a case, (3) is no longer true. However, there are situations in which it is "close" to be true in the sense that for some $\theta > 1$,

$$|F[\mathbf{x}_{s,t}](a) - F[\mathbf{x}_{r,s}] \circ F[\mathbf{x}_{s,t}](a)| \leqslant C|t-s|^{\theta} \text{ for any } s \leqslant t,\ a \in \mathrm{V}. \tag{4}$$

With a few additional assumptions on $F[\mathbf{x}_{r,s}]$, the non-linear sewing lemma [2, 15, 14] shows that there exists a flow $\{\phi_{t,s}\}_{s\leqslant t}$ (*i.e.*, $\phi_{t,s} \circ \phi_{s,t} = \phi_{t,r}$ for any $r \leqslant s \leqslant t$) close to $\{F[\mathbf{x}_{s,t}]\}_{s\leqslant t}$. Hence, the path $y : [0,T] \to \mathrm{V}$ defined by $y_t = \phi_{t,s}(a)$ actually satisfies (1). Thus, y is a suitable object for being a solution to a generalized notion of differential equation.

Given a $\mathrm{T}_{\leqslant n}(\mathrm{U})$-valued rough path $\mathbf{x}$ with components $\mathbf{x}^{(i)} \in \mathrm{U}^{\otimes i}$, $i = 0, \ldots, n$, as well as a function $f : \mathrm{V} \to L(\mathrm{U}, \mathrm{V})$, a Taylor expansion leads us to naturally consider

$$F[\mathbf{x}_{s,t}](a) = \sum_{k=0}^{n} f^{\{k\}}(a)\mathbf{x}^{(k)}_{s,t} \text{ for any } s \leqslant t, \tag{5}$$

where $f^{\{k+1\}} = \mathrm{D}f^{\{k\}} \cdot f$ and $f^{(0)} = \mathrm{id}$. Such an approximation was considered first in [22] for $n = 2$, and then in [25, 5] for arbitrary orders n. In particular, in [25], it is shown that thanks to an astute use of the Davie lemma and sub-Riemannian geodesic approximations, (4) holds with $\theta = (n+1)/p$ for a $(1/p)$-Hölder rough path with $n \geqslant \lfloor p \rfloor$. The effect of the approximation on the constant C in (1) was studied in [10, 7]. Generalization to branched rough paths with trees of order n also lead to similar controls [8, 19].

Another approximation is given by setting $F[\mathbf{x}_{s,t}](a) = z_1^{s,t}$ where $z^{s,t}$ is the solution to the ODE

$$z_r^{s,t} = a + \int_0^r F[\boldsymbol{\lambda}_{s,t}](z_r^{s,t})\,\mathrm{d}r \text{ with } \boldsymbol{\lambda}_{s,t} = \log(\mathbf{x}_{s,t}) \text{ in } \widetilde{\mathscr{W}}. \tag{6}$$

The approach is developed in [2, 3, 4].

When $\phi_{t,s}$ is seen as an approximation, the kind of control in (4) is related to *consistency* of the scheme. Under extra hypothesis, the rate of the numerical Euler scheme obtained by composing the $\phi_{t,s}$ over small time intervals, may be deduced from the knowledge of θ [26, 16, 13].

It is important to point out that naive computations do not lead to a proper assessment of θ when $n \geqslant 3$. They actually lead to $\theta = 3/p$ whatever the order n of the expansion. To reach a rate that depends on n, one really needs computations of *algebraic nature*. In particular, for $n \geqslant 3$, we are limited to use *weak geometric rough paths*, which are obtained as the limits of smooth rough paths naturally lifted through their iterated integrals. However, it was shown first in [33] that for the order of regularity $p \in [2, 3)$, and then in [31] and studied subsequently in [44, 9] using branched rough paths that one could reduced non geometric rough paths to geometric ones through suitable isomorphisms.

Beyond the algebraic aspect, another issue raised to get proper controls on (4) is to manage the fact that in an expression like (5) involves functions of different order of regularity when f is not smooth. One should really take care of the regularity of the functions when expanding then through Taylor approximations to keep only the relevant terms.

To overcome these difficulties, we consider an approach based on a "backward error analysis" [42]. We consider solving an ordinary differential equation (ODE) of type

$$y_t[\alpha](a) = a + \int_0^t F[\alpha](y_s[\alpha])\,\mathrm{d}s, \tag{7}$$

for a suitable $\alpha \in \widetilde{\mathscr{W}}$ which we related to some $\beta \in \widetilde{\mathscr{W}}$ through

$$F[\beta](a) = y_1[\alpha](a) + \text{ remainder}. \tag{8}$$

Or course, β and α are linked by a transform of type exponential/logarithmic. We actually prove a stronger statement:

$$F[\gamma](y_1[\alpha](a)) = F[\beta(\alpha,\gamma)](a) + \text{ remainder}$$

for some $\beta(\alpha,\gamma) \in \widetilde{\mathscr{W}}$. This is the key to control over composition of type $F[\alpha] \circ F[\beta]$, and the to rectify $\{F[\mathbf{x}_{s,t}]\}_{s \leqslant t}$ as a flow using the non-linear sewing lemma.

An expression of type (8) is obtained by iterating the use of the Newton formula. This connect (6) and (5).

In our context, we then consider that F from $\widetilde{\mathscr{W}}$ to the space of continuous functions not as a character in the space of continuous functions with composition, but as satisfying a relation of type

$$\mathrm{D}F[\beta] \cdot F[\alpha] = F[\alpha\beta]$$

for a suitable α, where D is the Fréchet derivative. We call such a map a Newtonian map.

Alternatively, we may consider F as an algebra morphism from $\widetilde{\mathscr{W}}$ to the algebra of differential operators. To use the Newton formula, $F[\alpha]$ shall be a first-order differential operator. This imposes some limitations on the kind of α to consider, which typically shall belong to a Lie algebra. This explains why weak rough paths [34, 26] shall be considered when dealing with high-order expansions (however, the correspondence between geometric and non-geometric rough paths stated above may be used [31]).

In this article,

- We give a Taylor formula with remainder for approximating flows of ODE with non-smooth vector fields defined as algebra homomorphisms from $\mathscr{W}$ to an algebra of functions or of differential operators.
- We also provide a composition formula from which we derive inequalities of type (4). The composition formula is a key statement to use the Davie lemma as a substitute to the Gronwall lemma and therefore construct to almost flows from flows.
- We extend the notion of a driving rough paths as a path with values in the truncated algebra $\mathscr{W}_{\leqslant n}$ derived from quotienting graded algebra $\mathscr{W}$ up to a given order n. The latter result may easily be generalized by replacing $\mathscr{W}_{\leqslant n}$ to a quotient of $\mathscr{W}$. This is useful when $\mathscr{W}$ is the superposition of algebras.
- We show how our results encompass some classical results about ODE, mainly the Lie-Trotter formula as well as controls of type (4) for rough differential.
- We obtain decaying (in n) bounds similar to the ones of [10] on the constant in (4) even for driving path living in an infinite dimensional Banach space as we do not rely on using geodesic approximations.
- Finally, we introduce a novel notion of *aromatic rough paths*. This class contains the class branched rough paths which itself contains the class of tensor-valued rough paths. This is a natural expansion to the context of rough path of the results of Munthe-Kaas *et al.* [11, 39, 37] on aromatic Butcher series introduced to construct affine equivariant numerical integrators.

Here, we focus mostly on constructing suitable vector fields, from which we deduce the properties a driving path shall satisfy. The literature mentioned above focus first on a suitable structure for the driving path. RDE are defined afterwards.

Beyond aromatic rough paths, other rough paths of similar nature and encoded by trees could be considered. The Grossmann-Larson (or Cayley) algebra [27] encodes the way differential operators acts on themselves and leads to a slightly different structure than the one used for the aromatic B-series. However, one may expect that all the structures lead to similar results. The prevalence of trees in all these representations is understood by the fact that trees are convenient ways to encode both operations on iterated integrals and on composition of differential operators. The relationship between these different constructions will be subject to future work.

Outline. In Section 2, we present the Taylor and the composition formula with remainder as well as the suitable algebraic setup related to Newtonian maps and operators. In Section 3, we gives a general construction of flows from rough paths, and gives the suitable controls. We show how to apply these results to ODE in Section 4 and to RDE in Section 5. Section 6 is devoted to branched rough paths, and the construction of aromatic branched rough paths.

2. The Taylor and the composition formula with remainder

2.1. The dual nature of ODEs

Let V be a Banach space. The Fréchet derivative operator is denote by D. As usual, we denote by $\mathscr{C}^k(\mathrm{V},\mathrm{V})$ the space of functions which are continuously differentiable from V to V with derivatives of order $k \geqslant 0$.

Let $\mathscr{L}$ be a set of elements, considered as letters. Let F be a map from $\mathscr{L}$ to the class of Lipschitz functions from V to V. The ODE

$$y_t[\alpha](a) = a + \int_0^t F[\alpha](y_s[\alpha](a))\,\mathrm{d}s \text{ for } t \geqslant 0 \tag{9}$$

has a unique V-valued continuous solution $y[\alpha](a)$ for any $\alpha \in \mathrm{V}$.

The Newton formula applied on a function $g \in \mathscr{C}^1(\mathrm{V},\mathrm{V})$ yields that

$$g(y_t[\alpha](a)) = g(a) + \int_0^t (\mathrm{D}g(y_s[\alpha](a)) \cdot F[\alpha](y_s[\alpha](a))\,\mathrm{d}s \text{ for } t \geqslant 0. \tag{10}$$

We define the first order differential operator $F^\dagger[\alpha]$ by

$$F^\dagger[\alpha]g(a) := \mathrm{D}g(a) \cdot F[\alpha](a) \text{ for } a \in \mathrm{V},\ g \in \mathscr{C}^1(\mathrm{V},\mathrm{V}), \tag{11}$$

as well as the linear differential operator

$$Y_t[\alpha]g(a) := g(y_t[\alpha](a)) \text{ for } a \in \mathrm{V},\ g \in \mathscr{C}^0(\mathrm{V},\mathrm{V}). \tag{12}$$

The Newton formula (10) is then rewritten

$$Y_t[\alpha]g(a) = g(a) + \int_0^t Y_s[\alpha]F^\dagger[\alpha]g(a)\,\mathrm{d}s \text{ for } a \in \mathrm{V},\ g \in \mathscr{C}^1(\mathrm{V},\mathrm{V}).$$

In the functional form, this means the linear equation

$$Y_t[\alpha] = 1 + \int_0^t Y_s[\alpha]F^\dagger[\alpha]\,\mathrm{d}s, \tag{13}$$

where 1 is the neutral operator $1g = g$ for $g \in \mathscr{C}^1(\mathrm{V},\mathrm{V})$.

Conversely, given a differential operator $F^\dagger[\alpha]$ with smooth coefficients, a Picard approximation applied to (13) leads to

$$Y_t[\alpha] = 1 + F^\dagger[\alpha] + \frac{t}{2!}F^\dagger[\alpha]^2 + \frac{t}{3!}F^\dagger[\alpha]^3 + \cdots \text{ for } t \geqslant 0, \tag{14}$$

where $F^\dagger[\alpha]^k := F^\dagger[\alpha](F^\dagger[\alpha]^{k-1})$ for $k \geqslant 2$. For convenience, we set $F^\dagger[\alpha]^0 = 1$.

We now assume that $\alpha \mapsto F^\dagger[\alpha]$ is a homomorphism from an algebra $\mathscr{W} \supset \mathscr{L}$ to the algebra $\mathscr{D}$ of differential operators with smooth coefficients. We rewrite the expansion (14) as

$$Y_t[\alpha] = F^\dagger[\exp(t\alpha)] \text{ for } t \geqslant 0 \text{ with } \exp(\alpha) = \sum_{k\geqslant 0} \frac{t^k}{k!}\alpha^k. \tag{15}$$

Let $\boldsymbol{i}$ be the identity map from V to V. When $F^\dagger[\alpha]$ and $F[\alpha]$ are related by (11), then $Y_t[\alpha]g = g(y_t[\alpha])$, $t \geqslant 0$, where $y[\alpha]$ solves (9) for any starting point $a \in \mathrm{V}$. As soon as $k \geqslant 2$, $F^\dagger[\alpha^k]$ is in general a differential operator of degree k.

Alternatively, when dealing with functions $F[\alpha]$ instead of the differential form $F^\dagger[\alpha]$, the condition $F^\dagger[\alpha\beta] = F^\dagger[\alpha]F^\dagger[\beta]$ is replaced by the following condition on the Lie derivative:

$$F[\alpha\beta](a) = \mathrm{D}F[\beta](a) \cdot F[\alpha](a) \text{ for } a \in \mathrm{V},\ \beta \in \mathscr{W},\ \alpha \in \mathscr{L} \cap \mathscr{W}. \tag{16}$$

Comparing (15) with (12) means that $Y_t[\alpha]$ may be understood as either a non-linear function acting on functions by composition or as a differential operator, hence the dual nature. In the context of numerical approximations, this duality is expressed through the notion of S-series, which generalizes the notion of B-series [40]. Considering functions circumvent the constraint that $F^\dagger[\alpha]$ shall be first order differential operator.

2.2. Setup for vector fields with non-smooth coefficients

The expansion in (15) is only valid if the function $F[\alpha] = F^\dagger[\alpha]\imath$ is smooth. This is not the case in many practical situations. We present a setup to deal with regular yet not smooth functions.

Hypothesis 1. The algebra $\mathscr{W}$ is a graded, unital algebra $\mathscr{W} = \bigoplus_{k\geqslant 0}\mathscr{W}_k$ with $\mathscr{W}_0 \simeq \mathbb{R}$ and $ab \in \mathscr{W}_{i+j}$ whenever $a \in \mathscr{W}_i$ and $b \in \mathscr{W}_j$. The neutral element 1 belongs to $\mathscr{W}_0$.

Notation 1. Fix $n \geqslant 0$. We define $\mathscr{W}_{\leqslant n} := \bigoplus_{k=0}^{n}\mathscr{W}_k$ and $\mathscr{W}_{>n} := \bigoplus_{k\geqslant n+1}\mathscr{W}_k$. Since $\mathscr{W}_{>n}$ is a two-sided ideal with respect to the multiplication of $\mathscr{W}$, $\mathscr{W}_{\leqslant n}$ is identified with the quotient space $\mathscr{W}/\mathscr{W}_{>n}$.

Notation 2. We decompose $\alpha \in \mathscr{W}$ as $\alpha = \sum_{i=0}^{+\infty} a^{(i)}$ with $\alpha^{(i)} \in \mathscr{W}_i$, $i \in \mathbb{N}$.

Notation 3 (Relations $\triangleright$ and $\blacktriangleright$)**.** For $\alpha, \beta \in \mathscr{W}$, we define

$$\alpha \triangleright \beta := \sum_{0\leqslant i+j\leqslant n} \alpha^{(i)}\beta^{(j)} \in \mathscr{W}_{\leqslant n} \tag{17}$$

$$\text{and } \alpha \blacktriangleright \beta := \sum_{i+j\geqslant n+1} \alpha^{(i)}\beta^{(j)} \in \mathscr{W}_{>n}. \tag{18}$$

Convention 1. For non-associative algebras, these products are always applied from right to left, for example $\alpha \triangleright \beta \triangleright \gamma$ means $\alpha \triangleright (\beta \triangleright \gamma)$.

The product $\triangleright$ is the projection of $\alpha\beta$ onto $\mathscr{W}_{\leqslant n}$. Furthermore, for $\alpha, \beta, \gamma \in \mathscr{W}$,

$$\alpha \triangleright \gamma \triangleright \beta = \alpha \triangleright (\gamma\beta), \tag{19}$$

$$\alpha\beta = \alpha \triangleright \beta + \alpha \blacktriangleright \beta \tag{20}$$

$$\text{and } \alpha \blacktriangleright \beta = \sum_{i+j=n+1}^{2n} \alpha^{(i)}\beta^{(j)} \text{ when } \alpha, \beta \in \mathscr{W}_n. \tag{21}$$

Notation 4 (Exponential and logarithm)**.** We write $\gamma^{\triangleright k}$ for $\gamma \in \mathscr{W}$. The exponential and the logarithm related to $\triangleright$ are defined by

$$\exp_\triangleright(\alpha) := \sum_{k\geqslant 0} \frac{1}{k!}\alpha^{\triangleright k}$$

$$\text{and } \log_\triangleright(1+\alpha) := \sum_{k\geqslant 0}(-1)^{k+1}\alpha^{\triangleright k},$$

for $\alpha \in \mathscr{W}$ with $\alpha^{(0)} = 0$.

Hypothesis 2. There exists a $\mathbb{R}$-submodule $\mathscr{L}$ of $\mathscr{W}$ which is stable under the projection $\alpha \mapsto \alpha^{(i)}$, for $i = 0, \ldots, n$. We set $\mathscr{L}_{\leqslant n} := \mathscr{L} \cap \mathscr{W}_{\leqslant n}$.

2.3. Newtonian map

Our first definition concerns parametric families of functions.

Definition 1 (Newtonian map)**.** Assume Hypotheses 1 and 2. A *Newtonian map* is a linear map F from $\mathscr{W}_{\leqslant(n+1)}$ to $\mathscr{C}^0(\mathrm{V},\mathrm{V})$ such that

- $F[1] = \imath$.
- $F[\alpha]$ is $\mathscr{C}^1$ for any $\alpha \in \mathscr{W}_{\leqslant n}$.
- $F[\alpha]$ is Lipschitz continuous for any $\alpha \in \mathscr{L}_{\leqslant n}$.
- for any $(\alpha,\beta) \in \mathscr{L}_{\leqslant n} \times \mathscr{W}_n$,

$$F[\alpha\beta] := \mathrm{D}F[\beta] \cdot F[\alpha].$$

Remark 1. When we combine (1) with (20) and (21), we obtain for $(\alpha,\beta) \in \mathscr{L}_{\leqslant n} \times \mathscr{W}_n$,

$$F[\alpha\beta] = F[\alpha \triangleright \beta] + F[\alpha \triangleright \beta] = F[\alpha \triangleright \beta] + R$$

where $\alpha \triangleright \beta \in \mathscr{W}_{\leqslant n}$ and

$$R := \sum_{i+j=n+1}^{2n} \mathrm{D}F[\beta^{(j)}] \cdot F[\alpha^{(i)}]$$

since $\alpha^{(i)} \in \mathscr{L}_{\leqslant n}$ by Hypothesis 2. Thus, $F[\alpha \triangleright \beta] \in \mathscr{C}^1$ and $R \in \mathscr{C}^0$. Thus justifies to use both $\triangleright$ and $\blacktriangleright$ to keep everything well defined.

Let us give a few examples. Elementary differentials, initially related to Runge-Kutta numerical schemes [30], provides us with another natural family of Newtonian maps which we develop later in Section 6.

Example 1. Let $\mathscr{F}$ be the vector space of functions of class $\mathscr{C}^\infty(\mathrm{V},\mathrm{V})$ which we equip with the product $f * g := (a \mapsto \mathrm{D}g(a) \cdot f(a))$. Thus, $(\mathscr{F},+,*)$ becomes a non-associative algebra with a right unit $\imath : \mathrm{V} \to \mathrm{V}$, the identify map on V. Any algebrea homomorphism F from $(\mathscr{W},+,\cdot)$ to $(\mathscr{F},+,*)$ is a Newtonian map.

An instance of the Example 1 is given by analytic functions.

Example 2. Let us consider as $\mathscr{W}$ the algebra of $\mathbb{R}$-valued sequences $\alpha := \{\alpha_i\}_{i\in\mathbb{N}}$ with the product

$$\alpha\beta := \left\{ \sum_{i+j=k} j\beta_{j+1}\alpha_i \right\}_{k\in\mathbb{N}}.$$

We set $\mathscr{L} := \mathscr{W}_{>0}$. For $\mathrm{V} = \mathbb{R}$, the function $F[\alpha](a) = \sum_{i\in\mathbb{N}} \alpha_i a^i$ is then a Newtonian map, as $\mathrm{D}F[\beta](a) = \sum_{i\in\mathbb{N}} i\beta_{i+1}a^i$ and $\mathrm{D}F[\beta](a) \cdot F[\alpha](a) = F[\alpha\beta](a)$.

Linear operators also give another instance of Example 1.

Example 3. We consider $\mathscr{W}$ is a Banach algebra and we let $\mathrm{V} := \mathscr{W}$. For $\alpha \in \mathscr{W}$, we define $F[\alpha](a) := a\alpha$ for $a \in \mathrm{V}$. We note that $\mathrm{D}F[\alpha](a) \cdot h = h\alpha$ for any $h \in \mathrm{V}$. Thus,

$$F[\beta] * F[\alpha](a) := \mathrm{D}F[\beta](a) \cdot F[\alpha](a) = a\alpha\beta = F[\alpha\beta](a).$$

Then F is a Newtonian map with $\mathscr{L} = \mathscr{W}$.

Remark 2. In general, the unit element 1 of $\mathscr{W}$ cannot belongs to $\mathscr{L}$, as $F[1 \cdot \beta](a) = \mathrm{D}F[\beta](a) \cdot a$ and this should be equal to $F[\beta]$ while $F[1] = \imath$. This is however possible in Example 3.

Example 4 (Derivatives of Newtonian maps). Let F be a Newtonian map. Then $G[\alpha]$ defined by

$$G[\alpha](a, b) = \begin{bmatrix} F[\alpha](a) \\ \mathrm{D}F[\alpha](a) \cdot b \end{bmatrix}$$

is a Newtonian map whenever the regularity is suitable. This statement implies that one may construct flows and their derivatives at the same time.

2.4. Newtonian operators

Definition 2 (Differential operator)**.** Fix $m \geqslant 0$. A *m-order differential operator* H is a linear operator of type

$$Hg(a) = \sum_{k=0}^{m} \mathrm{D}^k g(a) \cdot H_k(a) \text{ for any } g \in \mathscr{C}^m(\mathrm{V}, \mathrm{V}),$$

where H_k is a continuous function with values in the symmetrization of $\mathrm{V}^{\otimes k}$, and H_m is non-vanishing. We use the convention that $\mathrm{D}^0 g = g$.

Definition 3 (Suitable function)**.** Given a differential operator H, we say that g is a *suitable function* when the regularity of g is greater or equal than the order of H so that Hg is a continuous function.

Definition 4 (Newtonian operator)**.** Assume Hypotheses 1 and 2. A *Newtonian operator* is a linear map $F^\dagger$ from $\mathscr{W}_{\leqslant(n+1)}$ to the class of differential operator such that

- $F^\dagger[1] = 1$, where 1 is the neutral operator $1g = g$ for any $g \in \mathscr{C}^0$.
- $F^\dagger[\beta]$ has coefficients of class $\mathscr{C}^1$ for any $\beta \in \mathscr{W}_{\leqslant n}$.
- $F^\dagger[\alpha]g = \mathrm{D}g \cdot F[\alpha]$ for Lipschitz continuous function $F[\alpha]$ for any $\alpha \in \mathscr{L}_{\leqslant n}$.
- for any $(\alpha, \beta) \in \mathscr{L}_{\leqslant n} \times \mathscr{W}_n$,

  $$F^\dagger[\alpha\beta]g := \mathrm{D}F^\dagger[\beta]g \cdot F[\alpha] = F^\dagger[\alpha]F^\dagger[\beta]g,$$

  for any suitable function $g : \mathrm{V} \to \mathrm{V}$.

When $F^\dagger[\alpha]$ and $F^\dagger[\beta]$ are two first order differential operators, their product is not a first order differential operator in general. The proof of the next lemma is classical [30, Section III.5.2, p. 85] (it relies on the fact that $\mathrm{D}^2 g \cdot a \otimes \beta = \mathrm{D}^2 g \cdot b \otimes a$ for any $a, b \in \mathrm{V}$ and any function $g : \mathrm{V} \to \mathrm{V}$ of class $\mathscr{C}^2$).

Lemma 1. *Set $[\alpha, \beta]_\triangleright := \alpha \triangleright \beta - \beta \triangleright \alpha$. When $\alpha, \beta \in \mathscr{L}_{\leqslant n}$, then for any suitable function g,*

$$F^\dagger[[\alpha, \beta]_\triangleright]g = \mathrm{D}g \cdot (F^\dagger[\alpha]F[\beta] - F^\dagger[\beta]F[\alpha])$$

so that $F^\dagger[[\alpha, \beta]_\triangleright]$ is a first order differential operator without zero order term.

The proof of the next three lemmas are immediate.

Definition 5 (Related Newtonian operators and map)**.** A Newtonian map F and a Newtonian operator $F^\dagger$ are related when $F^\dagger$ is the Lie derivative with respect to F, *i.e.*,

$$F^\dagger[\alpha]g(a) = \mathrm{D}g(a) \cdot F[\alpha](a) \text{ for any } \alpha \in \mathscr{L}_{\leqslant n},\ a \in \mathrm{V} \text{ and } g \in \mathscr{C}^1(\mathrm{V}, \mathrm{V}). \tag{22}$$

Lemma 2. *Assume that the Newtonian map F and the Newtonian operator $F^\dagger$ are related. Then $g(y_t[\alpha]) = Y_t[\alpha]g$, where $y[\alpha]$ solves* (9) *and $Y[\alpha]$ solves* (13)*.*

Lemma 3. *Let $F^\dagger$ be a Newtonian operator. Then $F[\alpha] := F^\dagger[\alpha]\imath$ is a Newtonian map and* (22) *is true.*

Lemma 4. *Let $F[\alpha]$ be defined as a $\mathscr{C}^n$ function which is Lipschitz on V for any α in a set $\mathscr{L}$. Let $\mathscr{W}$ be the algebra freely generated by the elements of $\mathscr{L}$. We assume that $\mathscr{W}$ is a graded algebra (the grading is not necessarily the one that arises from the number of terms in the monomials). We define $F^\dagger$ by $F^\dagger[\alpha]g(a) := \mathrm{D}g(a) \cdot F[\alpha]$ for $\alpha \in \mathscr{L}$ and recursively $F^\dagger[\alpha\gamma]g(a) := F^\dagger[\alpha]F^\dagger[\gamma]g(a)$ for any $\gamma = \alpha_1 \cdots \alpha_m$ with $\gamma \in \mathscr{W}_{\leqslant n}$. Then $F^\dagger$ defines a Newtonian operator.*

2.5. The Taylor formula with remainder

The now state the Taylor formula.

Lemma 5 (Taylor formula with remainder, functional form)**.** *Let F be a Newtonian map. For any $\alpha \in \mathscr{L}_{\leqslant n}$ and $\beta \in \mathscr{W}_{\leqslant n}$,*

$$F[\beta](y_t[\alpha](a)) = F[\exp_\triangleright(t\alpha) \triangleright \beta](a) + R_t[\alpha, \beta](a) \textit{ for } t \geqslant 0, \tag{23}$$

where $y[\alpha]$ solves (9) *and*

$$R_t[\alpha, \beta](a) = \int_0^t F[\alpha \blacktriangleright (\exp_\triangleright((t-s)\alpha) \triangleright \beta)](y_s[\alpha](a))\,\mathrm{d}s \textit{ for } t \geqslant 0.$$

Proof. Using the Newton formula applied to (9), since $F[\beta]$ is of class $\mathscr{C}^1$,

$$\begin{aligned} F[\beta](y[\alpha]_t(a)) &= F[\beta](a) + \int_0^t \mathrm{D}F[\beta]F[\alpha](y_s[\alpha](a))\,\mathrm{d}s \\ &= F[\beta](a) + tF[\alpha \triangleright \beta](a) + \int_0^t F[\alpha \blacktriangleright \beta](y_s[\alpha](a))\,\mathrm{d}s \\ &\qquad + \int_0^t (F[\alpha \triangleright \beta](y_s[\alpha](a)) - F[\alpha \triangleright \beta](a))\,\mathrm{d}s. \end{aligned}$$

Iterating this procedure,

$$\begin{aligned} &F[\beta](y[\alpha]_t(a)) \\ &\quad = \sum_{k=0}^{+\infty} \frac{t^k}{k!} F[\alpha^{\triangleright k} \triangleright \beta](a) + \sum_{k=1}^{+\infty} \int_0^t \int_0^{t_1} \cdots \int_0^{t_{k-1}} F[\alpha \blacktriangleright \alpha^{\triangleright(k-1)} \triangleright \beta](y_{t_k}[\alpha](a))\,\mathrm{d}s. \end{aligned}$$

Using for example [23, Exercise 2.81, p. 253],

$$\begin{aligned} &\int_0^t \int_0^{t_1} \cdots \int_0^{t_{k-1}} F[\alpha \blacktriangleright \alpha^{\triangleright(k-1)} \triangleright \beta](y_{t_k}[\alpha](a))\,\mathrm{d}t_k \cdots \mathrm{d}t_1 \\ &\qquad\qquad = \int_0^t \frac{(t-s)^{k-1}}{(k-1)!} F[\alpha \blacktriangleright \alpha^{\triangleright(k-1)}\beta](y[\alpha]_s(a))\,\mathrm{d}s. \end{aligned}$$

With the definition of the exponential $\exp_\triangleright$, this gives the desired formula. □

The proof of the operational form is similar so that we skip it.

Lemma 6 (Taylor formula with remainder, operational form)**.** *Let $F^\dagger$ be a Newtonian operator. For any $\alpha \in \mathscr{L}_{\leqslant n}$ and $\beta \in \mathscr{W}_{\leqslant n}$,*

$$Y_t[\alpha]F^\dagger[\beta] = F^\dagger[\exp_\triangleright(t\alpha) \triangleright \beta] + R_t^\dagger[\alpha, \beta], \tag{24}$$

where $Y[\alpha]$ is defined by (12) *and*

$$R_t^\dagger[\alpha, \beta]g(a) := \int_0^t Y_s[\alpha]F^\dagger[\alpha \blacktriangleright (\exp_\triangleright((t-s)\alpha) \triangleright \beta)]g(a)\,\mathrm{d}s,$$

for any suitable function g.

Remark 3. For $b, b' \in \mathrm{V}$, we consider $\mathscr{L} = \{\alpha\}$ for a symbol α and $\mathscr{W}$ the word algebra generated, with $\mathscr{W}_k$ the span of α^k. We set $F^\dagger[\alpha]g(a) := \mathrm{D}g(a) \cdot (b' - b)$. We define for $g \in \mathscr{C}^k$ and $a \in \mathrm{V}$,

$$F^\dagger[\alpha^k](a) := F^\dagger[\alpha]F^\dagger[\alpha^{k-1}]g(a) = \mathrm{D}^k g(a) \cdot (b' - b)^{\otimes k} \text{ for } a \in \mathrm{V}.$$

The Taylor formula (24) contains the usual Taylor formula with remainder for any $g \in \mathscr{C}^{n+1}$:

$$g(b') = \sum_{k=0}^{n} \frac{1}{k!} \mathrm{D}^k g(b) \cdot (b' - b) + \int_0^1 \frac{(1-s)^n}{n!} \mathrm{D}^{n+1} g(tb + (1-t)b') \cdot (b' - b)^{\otimes k} \, \mathrm{d}s$$

since $\alpha \triangleright \exp_{\triangleright}((t-s)\alpha) = (t-s)^n \alpha^{n+1}/n!$.

2.6. The composition formula

The subspace $\mathscr{L}$ is not necessarily a subalgebra. We start by defining a binary relation between elements of $\mathscr{L}$.

Notation 5. We write

$$\alpha \circledast_{\triangleright} \beta := \log_{\triangleright}(\exp_{\triangleright}(\alpha) \triangleright \exp_{\triangleright}(\beta)) \text{ for } \alpha, \beta \in \mathscr{L}. \tag{25}$$

By the properties of $\triangleright$, $\alpha \circledast_{\triangleright} \beta \in \mathscr{L}_{\leqslant n}$.

Usually, expressions of $\alpha \circledast_{\triangleright} \beta$ is given by the truncated Baker-Campbell-Hausdorff-Dynkin (BCHD) formula [12]. BCHD type formula can be extended to non-associative algebras [38].

Hypothesis 3. The submodule $\mathscr{L}$ is such that $\alpha \circledast_{\triangleright} \beta \in \mathscr{L}_{\leqslant n}$ for any $\alpha, \beta \in \mathscr{L}$.

Example 5. When $\mathscr{W}$ is a graded Lie algebra and $\mathscr{L}$ is a Lie algebra generated by $[\alpha, \beta] = \alpha\beta - \beta\alpha$ containing $\mathscr{W}_1$, then Hypothesis 3 holds true.

Lemma 7 (Composition formula, functional form)**.** *Let F be a Newtonian map. For $\alpha, \beta \in \mathscr{L}_{\leqslant n}$, under Hypotheses 1, 2 and 3,*

$$\begin{aligned} y_1[\beta] \circ y_1[\alpha](a) &= y_1[\alpha \circledast_{\triangleright} \beta](a) \\ &\quad + R_1[\alpha, \exp_{\triangleright}(\beta)](a) + R_1[\beta, 1](\Phi[\alpha](a)) - R_1[\alpha \circledast_{\triangleright} \beta, 1](a). \end{aligned} \tag{26}$$

Proof. Using the Taylor formula on $y_1[\beta]$ with $F[1]$,

$$y_1[\beta] \circ y_1[\alpha](a) = F[\exp_{\triangleright}(\beta)](y_1[\alpha](a)) + R_1[\beta, 1](y_1[\alpha](a)).$$

Applying again the Taylor formula on $y_1[\alpha]$ with $F[\exp_{\triangleright}(\beta)]$,

$$y_1[\beta] \circ y_1[\alpha](a) = F[\exp_{\triangleright}(\alpha) \triangleright \exp_{\triangleright}(\beta)](a) + R_1[\alpha, \exp_{\triangleright}(\beta)](a) + R_1[\beta, 1](y_1[\alpha](a)).$$

Finally, using the Taylor formula on $y_1[\alpha \circledast_{\triangleright} \beta]$ with $f[1]$ leads to (26). □

We now state the operational counterpart of the composition formula.

Lemma 8 (Composition formula, operational form)**.** *Let $F^\dagger$ be a Newtonian operator. For $\alpha, \beta \in \mathscr{L}_{\leqslant n}$, under Hypotheses 1, 2 and 3,*

$$Y_1[\alpha]Y_1[\beta]g = Y_1[\alpha \circledast_\triangleright \beta]g \\ + R_1^\dagger[\alpha, \exp_\triangleright(\beta)]g + R_1^\dagger[\beta, 1]g \circ Y_1[\alpha] - R_1^\dagger[\alpha \circledast_\triangleright \beta, 1]g \quad (27)$$

for any suitable function g.

3. Flows of differential equations controlled by rough paths

3.1. Building $\mathscr{W}$-valued rough paths

As introduced by T. Lyons in [36], a p-rough path is a path with values in the truncated tensor algebra $\mathrm{T}_n(\mathrm{U})$ of a Banach space U and of finite p-variation. Later, M. Gubinelli constructed a rough path with values in the algebra of rooted trees [29]. These constructions extend the notion of differential equations.

The notion of rough paths itself is easily extended to a general graded Banach algebra.

Definition 6 (Control)**.** A *control* is a map $\omega : \mathbb{T}_+^2 \to \mathbb{R}_+$ which is super-additive, that is $\omega_{r,s} + \omega_{s,t} \leqslant \omega_{r,t}$ and continuous close to its diagonal.

Hypothesis 4. The graded algebra $\mathscr{W}$ is a Banach algebra with a metric $|\cdot|$ which satisfies $|\alpha\beta| \leqslant |\alpha| \cdot |\beta|$ for any $\alpha, \beta \in \mathscr{W}$ and

$$|\alpha| = \sum_{i \geqslant 0} |\alpha^{(i)}| \text{ where } \alpha^{(i)} \in \mathscr{W}_i,\ i \geqslant 0 \text{ and } \alpha = \sum_{i \geqslant 0} \alpha^{(i)}.$$

Definition 7 ($\mathscr{W}$-valued rough paths)**.** A $\mathscr{W}$-valued family $\mathbf{x} := (\mathbf{x}_{s,t})_{(s,t)\in\mathbb{T}_+^2}$ is a *$\mathscr{W}$-valued p-rough path* if there exists a $\mathbb{R}_+$-valued family $\{\mu_i\}_{i\in\mathbb{N}}$ such that

- For $i \geqslant 0$, $\mathbf{x}_{s,t}^{(i)} \in \mathscr{W}_i$ with $\mathbf{x}_{s,t}^{(0)} = 1$ and

$$|\mathbf{x}_{s,t}^{(i)}| \leqslant \mu_i \omega_{s,t}^{i/p} \text{ for any } i \geqslant 1,\ (s,t) \in \mathbb{T}_+^2. \quad (28)$$

- The family $\mathbf{x}$ is *multiplicative*, that is

$$\mathbf{x}_{r,s}\mathbf{x}_{s,t} = \mathbf{x}_{r,t} \text{ for any } (r,s,t) \in \mathbb{T}_+^3. \quad (29)$$

- For any $(s,t) \in \mathbb{T}_+^2$, $|\mathbf{x}_{s,t}|$ is finite.

Since $\mathscr{W}_{>n}$ is a two-sided ideal, $\mathscr{W}_n$ is isomorphic to the quotient space $\mathscr{W}/\mathscr{W}_{>n}$. The Lyons extension theorem is also easily generalized,

Proposition 1 (Lyons extension theorem)**.** *Let* $\mathbf{x}$ *be a* $\mathscr{W}_n$*-valued rough path with* $n > \lfloor p \rfloor$*. Then there exists a unique* $\mathscr{W}$*-valued rough path* $\mathbf{y}_{s,t}$ *such that* $\mathbf{x}_{s,t} = \mathbf{y}_{s,t}^{(\leqslant n)}$*, where* $\alpha^{(\leqslant n)} := \sum_{i=0}^{n} \alpha^{(i)}$*.*

Proof. The $\mathscr{W}_n$-rough path $\mathbf{x}$ is almost multiplicative in $\mathscr{W}$, that is

$$\mathbf{x}_{r,s,t} := \mathbf{x}_{r,s}\mathbf{x}_{s,t} - \mathbf{x}_{r,t} = \mathbf{x}_{r,s} \triangleright \mathbf{x}_{s,t} \in \mathscr{W}_{>n} \text{ for any } (r,s,t) \in \mathbb{T}_+^3.$$

Therefore,

$$|\mathbf{x}_{r,s,t}| \leqslant \omega_{r,t}^{(n+1)/p}(n-1)|\mathbf{x}_{r,t}|^2.$$

The multiplicative sewing lemma [36, 24] is sufficient to conclude. □

3.2. Rough Differential Equations driven by $\mathscr{W}$-valued rough paths

Hypothesis 5. Fix $n > 0$. There exist

- a submodule $\mathscr{L}$ of $\mathscr{W}$ satisfying Hypothesis (2).
- for some $p \geqslant 1$ and $n \geqslant \lfloor p \rfloor$, a $\mathscr{W}$-valued p-rough path $\mathbf{x}$ with $\boldsymbol{\lambda}_{s,t} := \log_{\triangleright}(\mathbf{x}_{s,t}) \in \mathscr{L}_{\leqslant n}$.
- a Newtonian map F.

Hypothesis 6. There exists a $\nu_i \in \mathbb{R}_+$, $i = 1, \dots, n$ such that for any $\alpha \in \mathscr{W}_n$,

$$\max\{\|F[\alpha^{(i)}]\|_{\mathrm{Lip}}, \|F[\alpha^{(i)}]\|_{\infty}\} \leqslant \nu_i |\alpha^{(i)}| \text{ for } i = 0, \dots, n.$$

Notation 6. We define a family ϕ of maps from V to V by

$$\phi_{t,s}(a) := y[\boldsymbol{\lambda}_{s,t}]_1(a) \text{ for } a \in \mathrm{V},\ (s,t) \in \mathbb{T}_+^2.$$

Proposition 2. *Under Hypotheses 5 and 6, if* $\mathscr{W}$ *is finite dimensional, then for any* $a \in \mathrm{V}$*, there exists a path* $y \in \mathscr{C}^0(\mathbb{T}, \mathrm{V})$*, not necessarily unique, such that* $y_0(a) = a$ *and*

$$|y_t - \phi_{t,s}(y_s)| \leqslant C\omega_{s,t}^{(n+1)/p} \textit{ for } (s,t) \in \mathbb{T}_+^2. \tag{30}$$

Such a path is called a D-solution *to*

$$y_t = a + \int_0^t F[\,\mathrm{d}\mathbf{x}_r](y_s)\,\mathrm{d}s \textit{ for } t \geqslant 0. \tag{31}$$

Remark 4. Although it is possible to give an explicit bound for C from the following proof, this constant is improved in Section 3.3.

Notation 7. We denote by $\mathscr{C}^n_{\mathrm{b}}$ the set of functions of class $\mathscr{C}^n$ with are bounded with bounded, continuous derivatives up to order n. For $g \in \mathscr{C}^n_{\mathrm{b}}$, we set $\|g\|_{\leqslant n} := \sum_{k=1}^n \|\mathrm{D}^k g\|_\infty$.

Corollary 1 (Newton's formula)**.** *Assume Hypotheses 5 and 6. Let $F^\dagger$ be the Newtonian map related to F through* (22)*. Then for any function g of class $\mathscr{C}^n_{\mathrm{b}}$,*

$$\left| g(y_t) - F^\dagger\left[\mathbf{x}^{(\leqslant n)}_{s,t}\right] g(y_s) \right| \leqslant C \|g\|_{\leqslant n} \omega_{s,t}^{(n+1)/p} \text{ for } (s,t) \in \mathbb{T}^2_+$$

for a constant C whenever $\|F^\dagger[\alpha]g\|_\infty \leqslant K|\alpha| \cdot \|g\|_{\leqslant n}$ for some constant $K \geqslant 0$ and any $\alpha \in \mathscr{L}_{\leqslant n}$.

We recall the definition of an *almost flow* (or *approximate flow* in [2]), here in the context of a Banach space. The definition of [15] is slightly weaker. For the sake of simplicity, we restrict ourselves to bounded functions/vector fields: see *e.g.* [16] for controls allowing to remove such an assumption. Also, we may weaken the control over the Lipschitz norm of $F[\alpha]$ to a control over its Hölder norm in Hypothesis 6.

Definition 8 (Almost flow)**.** A family $\phi := \{\phi_{s,t}\}_{(s,t)\in\mathbb{T}^2_+}$ of maps from V to V is an *almost flow* if for some non-decreasing, continuous function $\delta : \mathbb{R}_+ \to \mathbb{R}_+$ with $\delta(0) = 0$,

- $\phi_{t,t} = \imath$ for any $t \in \mathrm{V}$.
- $\|\phi_{t,s} - \imath\|_\infty \leqslant \delta_T$ for any $(s,t) \in \mathbb{T}^2_+$.
- $\|\phi_{t,s}\|_{\mathrm{Lip}} \leqslant 1 + \delta_T$ for any $(s,t) \in \mathbb{T}^2_+$.
- For some $\theta > 1$ and some constant $L \geqslant 0$,

$$\|\phi_{t,s} \circ \phi_{s,r} - \phi_{t,r}\|_\infty \leqslant L\omega_{r,t}^\theta \text{ for any } (r,s,t) \in \mathbb{T}^3_+. \tag{32}$$

Besides, ϕ is a *flow* if $L = 0$ in (32).

The proof of the next results is proved using an homogeneity argument. We skip its proof.

Lemma 9. *There exist some constants $\lambda_i, L_i \geqslant 0$, $i = 1, \dots, n$ such that*

$$|\boldsymbol{\lambda}^{(i)}_{s,t}| \leqslant \lambda_i \omega_{s,t}^{i/p} \text{ and } |\exp(\theta \boldsymbol{\lambda}_{s,t})^{(i)}| \leqslant L_i \omega_{s,t}^{i/p} \text{ for any } (s,t) \in \mathbb{T}^2_+ \text{ and } 0 \leqslant \theta \leqslant 1 \tag{33}$$

for $i = 1, \dots, n$.

Proof of Proposition 2. We prove first that ϕ is an almost flow. Since for any $s \in \mathbb{T}$, $\boldsymbol{\lambda}_{s,s} = 0$, $\phi_{s,s}(a) = a$.

From applications of the Gronwall lemma,

$$\|\phi_{t,s} - \imath\|_\infty \leqslant N_{s,t}\|\phi_{t,s}\|_{\mathrm{Lip}} \leqslant N_{s,t}$$

with

$$N_{s,t} := \sum_{i=1}^{n} \nu_i |\boldsymbol{\lambda}_{s,t}^{(i)}| \leqslant \sum_{i=1}^{n} \nu_i \lambda_i \omega_{s,t}^{i/p},$$

for any $(s,t) \in \mathbb{T}_+^2$. We then set $\delta_T := N_{0,T}$.

Since, $\mathbf{x}_{r,s} \triangleright \mathbf{x}_{s,t} = \mathbf{x}_{r,t}$ in $\mathscr{W}_{\leqslant n}$, it follows from (25) that $\lambda_{r,s} \circledast_\triangleright \lambda_{s,t} = \lambda_{r,t}$ for any $(r,s,t) \in \mathbb{T}_+^2$. The composition formula yields

$$\phi_{t,s} \circ \phi_{s,r}(a) - \phi_{t,r}(a) = R[\boldsymbol{\lambda}_{r,s}, \mathbf{x}_{s,t}^{(\leqslant n)}](a) + R[\boldsymbol{\lambda}_{s,t}, 1](\phi_{r,s}(a)) - R[\boldsymbol{\lambda}_{r,t}, 1](a).$$

With (33), (28) and Hypothesis (6), for some constant C',

$$\|R[\boldsymbol{\lambda}_{r,s}, \mathbf{x}_{s,t}^{(\leqslant n)}]\|_\infty \leqslant \sum_{k=n+1}^{2n} \nu_k \sum_{i+j=k} \left(\sum_{\ell+\ell'=i} \lambda_\ell \Lambda_{\ell'} \right) \mu_i \omega_{r,t}^{i/p} \omega_{s,t}^{j/p} \leqslant C' \omega_{r,t}^{(n+1)/p}$$

for any $(r,s,t) \in \mathbb{T}_+^3$. With $s = t$, this inequality applies to $\|R[\boldsymbol{\lambda}_{s,t}, 1]\|_\infty$. Hence,

$$\|\phi_{t,s} \circ \phi_{s,r} - \phi_{t,r}\|_\infty \leqslant C \omega_{r,t}^{(n+1)/p}.$$

This proves that ϕ is an almost flow. The results follows from [15, Theorem 2]. □

Proof of Corollary 1. For any $(s,t) \in \mathbb{T}_+^2$,

$$\begin{aligned}\left| g(y_t) - F^\dagger\left[\mathbf{x}_{s,t}^{(\leqslant n)}\right] g(y_s) \right| &\leqslant |g(y_t) - \phi_{t,s} g(y_s)| + \left|\phi_{t,s} g(y_s) - F^\dagger[\mathbf{x}_{s,t}] g(y_s)\right| \\ &\leqslant \|g\|_{\mathrm{Lip}} \cdot |y_t - \phi_{t,s}(y_s)| + \|R^\dagger[\boldsymbol{\lambda}_{s,t}, 1] g(y_s)\|_\infty \\ &\leqslant C\|g\|_{\mathrm{Lip}} \omega_{s,t}^{(n+1)/p} + C'\|g\|_{\leqslant n} \omega_{s,t}^{(n+1)/p}.\end{aligned}$$

The last inequality follows from Lemma 9. □

Remark 5. In the proof of Proposition 2, when F and $F^\dagger$ are related, we have proved that

$$F^\dagger[\mathbf{x}_{r,s} \circledast_\triangleright \mathbf{x}_{s,t}] = y_1[\boldsymbol{\lambda}_{s,t}](y_1[\boldsymbol{\lambda}_{r,s}](a)) + \text{remainder}$$

as well as

$$F^\dagger[\mathbf{x}_{r,s} \circledast_\triangleright \mathbf{x}_{s,t}] = F[\mathbf{x}_{s,t}](F[\mathbf{x}_{r,s}](a)) + \text{remainder}.$$

As the coefficients of $F^\dagger[\mathbf{x}_{s,t}]$ may only be of class $\mathscr{C}_{\mathrm{b}}^1$, $F^\dagger[\mathbf{x}_{s,t}] F^\dagger[\mathbf{x}_{r,s}]$ is not necessarily well defined, while $F[\mathbf{x}_{s,t}](F[\mathbf{x}_{r,s}(a))$ is.

We now give a sufficient condition to ensure uniqueness. Actually, much more results can be given, such as generic properties, rate of convergence of numerical schemes and so on as for ODE with Lipschitz vector fields [15, 14, 16].

Definition 9 (4-points control)**.** A function $g : \mathrm{V} \to \mathrm{V}$ satisfies a *4-points control* whenever

$$|g(a) - g(b) - g(c) + g(d)| \leqslant \widehat{g}(|a-b| \vee |c-d|) \times (|a-c| \vee |b-d|) + g^{\circledast}|a-b-c+d|$$

for a non-decreasing, continuous function $\widehat{g} : \mathbb{R}_+ \to \mathbb{R}_+$ and $C \geqslant 1$.

The proof of the next lemma is a direct consequence of the Gronwall lemma.

Lemma 10. *Let g be a function that satisfies a 4-points control. Let y be the solution to the ODE $y_\tau(a) = a + \int_0^t g(y_\sigma(a))\,\mathrm{d}\sigma$ (which is necessarily unique since g is Lipschitz continuous). Then $h := a \mapsto y_1(a)$ satisfies a 4-points control with $\widehat{h} := \exp(g^{\circledast})\widehat{g}(\|h\|_{\mathrm{Lip}})\|h\|_{\mathrm{Lip}}$ and $h^{\circledast} := \exp(g^{\circledast})$, where $\|h\|_{\mathrm{Lip}} \leqslant \exp(\|g\|_{\mathrm{Lip}})$.*

Proposition 3. *Assume that $F[\boldsymbol{\lambda}_{s,t}]$ satisfies a 4-points control for any $(s,t) \in \mathbb{T}^2_+$ with $F[\boldsymbol{\lambda}_{s,t}]^{\circledast} \leqslant \delta_T$ and for some $\theta > 1$,*

$$\|R[\boldsymbol{\lambda}_{s,r}, \mathbf{x}^{(\leqslant n)}_{s,t}]\|_{\mathrm{Lip}} + \|R[\mathbf{x}^{(\leqslant n)}_{s,t}, 1]\|_{\mathrm{Lip}} \leqslant C\omega^{\theta}_{r,t} \text{ for any } (r,s,t) \in \mathbb{T}^3_+. \tag{34}$$

Then $\{\phi_{s,t}\}$ is a stable almost flow and the D-solution to (31) is unique. The D-solution also exists in an infinite Banach space.

Alternatively, we may consider working on $F[\mathbf{x}^{(\leqslant n)}_{s,t}]$ using perturbation results [15, 14].

Proposition 4. *The family ψ defined by $\psi_{t,s}(a) := F[\mathbf{x}^{(\leqslant n)}_{s,t}]$, $(s,t) \in \mathbb{T}^2_+$ is an almost flow with $\|\psi_{s,t} - \phi_{s,t}\|_\infty \leqslant C\omega^{\theta \wedge (n+1)/p}_{s,t}$ for any $(s,t) \in \mathbb{T}^2_+$. If in addition (34) holds and ϕ is a stable almost flow, then ψ is also a stable almost flow.*

3.3. Decay estimate

One could see the $\phi_{t,s}$ as numerical integrators and to use the almost flow to construct approximation schemes. This is why (30), which corresponds to consistency, actually determines the rate of convergence from the knowledge.

An inequality of type (30) only state consistency. When $\|\phi_{t,s}\|_{\mathrm{Lip}}$ is of order $t - s$, for example for ODE, then the Gronwall lemma gives the rate of convergence. This is no longer true when $\phi_{s,t}$ is not of order $t - s$. In Proposition 3, we give a

sufficient condition to get uniqueness and rate of convergence. The Davie lemma (See Appendix A) is a substitute to the Gronwall lemma. Its requires to already have an estimate on $\Gamma_{s,t} := y_t - \phi_{t,s}(y_s)$ of type $|\Gamma_{s,t}| \leqslant C\omega_{s,t}^{\theta}$ for some $\theta > 1$. Given that $|\Gamma_{r,s,t}| \leqslant M\omega_{r,t}^{\theta}$ with $\Gamma_{r,s,t} = \Gamma_{r,s} + \Gamma_{s,t} - \Gamma_{r,t}$, then C may be replaced by KM for a constant K depending only on $\omega_{0,T}$ and $\theta > 1$. This was the strategy proposed first by P. Friz and N. Victoir [25, 26] to study high-order Euler schemes. They however use a geodesic approximation to get first $|\Gamma_{s,t}| \leqslant C\omega_{s,t}^{\theta}$ with a constant C that depends on the length of an approximation by a smooth path. They hence obtain a constant M in $|\Gamma_{r,s,t}| \leqslant M\omega_{r,t}^{\theta}$ that does not depend on the regularity of the path.

It is possible to get the constant C explicitly. For this, it relies on getting bounds on the λ_i in (33). However, this is useless as the constant may be strongly improved. For this, we follow the general idea from [10] (see also [8, 8] for related works). The restriction to consider only a rough paths with values in a finite-dimensional space is removed as we do not rely on using a sequence of smooth approximations of the path.

Here, we consider a Newtonian operator $F^{\dagger}$, as we saw in Lemma 4 how to transform a Newtonian map into a Newtonian operator. From now, we denote by y a D-solution to $y_t = a + \int_0^t F[\,\mathrm{d}\mathbf{x}_r](y_r)\,\mathrm{d}r$.

Notation 8. We define

$$R_j^{s,t}(g) := g(y_t) - F^{\dagger}\Big[\mathbf{x}_{s,t}^{(\leqslant j)}\Big]g(y_s) \tag{35}$$

for $j = 0, 1, 2, \ldots, n$, $g \in \mathscr{C}^j$ and any $(s,t) \in \mathbb{T}_+^2$.

Hypothesis 7. Hypothesis 6 holds. Besides, there exists a constant $\gamma \geqslant 0$ such that $n - 1 + \gamma > p$ and for each $j \in \{1, \ldots, n-1\}$ a constant $k_j \geqslant 0$ such that

$$\|R_j^{s,t}(F[\alpha]g)\|_{\infty} \leqslant k_j|\alpha| \cdot \|g\|_{\leqslant j}\omega_{s,t}^{(\gamma+j)/p} \tag{36}$$

for any $g \in \mathscr{C}_{\mathrm{b}}^j$ and any $\alpha \in \mathscr{W}_{\leqslant j}$.

Proposition 5 (Propagation of decay)**.** *Consider Hypothesis 7. We assume that for a constant A,*

$$\|R_n^{s,t}(g)\|_{\infty} \leqslant A\|g\|_{\leqslant n}\omega_{s,t}^{(n+\gamma)/p}, \text{ for any } 0 \leqslant s \leqslant t \leqslant T, \tag{37}$$

for $g \in \mathscr{C}_{\mathrm{b}}^n$. Then (37) holds true with A replaced by

$$k_n := \sup_{\substack{(r,s,t)\in\mathbb{T}_+^3\\ r<s<t}} |\mathbf{x}| \frac{(1 - 2^{(p-n-\gamma)/p})^{-1}\sum_{j=1}^n k_{n-j}\nu_j\omega_{r,s}^{(\gamma+n-j)/p}\omega_{s,t}^{j/p}}{\omega_{r,t}^{(n+\gamma)/p}}. \tag{38}$$

Proof. Set for $(r,s,t) \in \mathbb{T}_+^3$ and a suitable function $g : \mathrm{V} \to \mathrm{V}$,

$$V_{r,s,t} := F^\dagger[\mathbf{x}_{s,t}^{(\leqslant n)} - 1]g(y_s) - F^\dagger[\mathbf{x}_{s,t}^{(\leqslant n)} - 1]g(y_r). \tag{39}$$

Recall that $F^\dagger[1]g = g$ so that $F^\dagger[\mathbf{x}_{s,t}^{(\leqslant n)}]g(a) = F^\dagger[\mathbf{x}_{s,t}^{(\leqslant n)} - 1]g(a) + g(a)$.
On the one hand,

$$R_n^{r,s}[g](y_r) + R_n^{s,t}[g](y_s) = g(y_t) - F^\dagger[\mathbf{x}^{(\leqslant n)} x_{r,s}]g(y_r) - F^\dagger[\mathbf{x}_{s,t}^{(\leqslant n)} - 1]g(y_r) + V_{r,s,t}. \tag{40}$$

On the other hand,

$$\begin{aligned} R_n^{r,t}[g](y_r) &= g(y_t) - \sum_{\substack{0 \leqslant i+j \leqslant n \\ i,j \geqslant 0}} F^\dagger[\mathbf{x}_{r,s}^{(i)} \mathbf{x}_{s,t}^{(j)}]g(y_r) \\ &= g(y_t) - \sum_{j=1}^{n} \sum_{i-1}^{n-j} F^\dagger[\mathbf{x}_{r,s}^{(i)} \mathbf{x}_{s,t}^{(j)}]g(y_r) - F^\dagger[\mathbf{x}_{s,t}^{(\leqslant n)}]g(y_r) - F^\dagger[\mathbf{x}_{r,s}^{(\leqslant n)} - 1]g(y_r). \end{aligned} \tag{41}$$

With (35),

$$\begin{aligned} F^\dagger[\mathbf{x}_{s,t}^{(j)}]g(y_s) - F^\dagger[\mathbf{x}_{s,t}^{(j)}]g(y_r) &= F^\dagger[\mathbf{x}_{r,s}^{(\leqslant n-j)}]F^\dagger[\mathbf{x}_{s,t}^{(j)}]g(y_r) + R_{n-j}^{r,s}(F^\dagger[\mathbf{x}_{s,t}^{(j)}]g)(y_r) \\ &= \sum_{i-1}^{n-1} F^\dagger[\mathbf{x}_{r,s}^{(j)} \mathbf{x}_{s,t}^{(i)}]g(y_r) + R_{n-j}^{r,s}(F^\dagger[\mathbf{x}_{s,t}^{(j)}]g)(y_r). \end{aligned}$$

With $V_{s,t}$ introduced in (39),

$$-V_{s,t} = \sum_{j=1}^{n} \sum_{i=1}^{n-j} F^\dagger[\mathbf{x}_{r,s}^{(j)} \mathbf{x}_{s,t}^{(i)}]g(y_r) + \sum_{j=1}^{n} R_{n-j}^{r,s}(F^\dagger[\mathbf{x}_{s,t}^{(j)}]g)(y_r). \tag{42}$$

Combining (40) with (41) and (42),

$$R_n^{r,s}[g](y_r) + R_n^{s,t}[g](y_s) - R_n^{r,t}[g](y_r) = -\sum_{j=1}^{n} R_{n-j}^{r,s}(F^\dagger[\mathbf{x}_{s,t}^{(j)}]g)(y_r).$$

Using Hypothesis 7,

$$|R_n^{r,t}[g](y_r)| \leqslant |R_n^{r,s}[g](y_r)| + |R_n^{s,t}[g](y_s)| + \sum_{j=1}^{n} |||g||| k_{n-j} |\mathbf{x}| \nu_j \omega_{r,s}^{(\gamma+n-j)/p} \omega_{s,t}^{j/p}.$$

The Davie lemma (Lemma 15 in Appendix) allows one to conclude. □

An immediate consequence of Proposition 5 combined with the neo-classical inequality [32, Theorem 1.2] is the following control similar to the one in [10] without being restricted to finite dimensional rough paths as no smooth approximmation of the path is used. In [13], it is proved that the consistency holds at order $(n+1)/\gamma$ even in presence of an infinite dimensional rough path. Yet the constant is not identified. Thanks to the Lyons extension theorem (See [36, Theorem 2.2.1, p. 242] or [34, Theorem 3.1.2]), (43) and (44) are satisfied for rough paths by using a proper choice of the control ω. Actually, with Example 3, Proposition 5 may be used to prove the Lyons extension theorem.

Corollary 2 (Propagation of factorial decay)**.** *Let us fix $n \geqslant 1$ and $\gamma > 0$. Assume that*

$$k_j \leqslant \frac{B}{(j/p)!} \text{ for } j = 0, \dots, m \text{ with } m + \gamma > p \tag{43}$$

$$\text{and } \nu_j \leqslant \frac{KB}{(j/p)!} \text{ for } i = 0, \dots, n \tag{44}$$

for some $K \geqslant 0$ with

$$B \leqslant \sqrt{p(1 - 2^{(p-m-\gamma)/p})}.$$

Then (43) *is also true for $j = m+1, \dots, n$.*

4. Application to ODEs

Let us consider a set of letters I. The algebra $\mathscr{W}$ is the algebra of words (*i.e.*, monomials) freely generated by I. The unit element of $\mathscr{W}$ is the empty word $\varnothing$. The length of a word w is denoted by $|w|$. By convention, $|\varnothing| = 0$. We define $\mathscr{W}_k := \{w \in \mathscr{W} \mid |w| = k\}$, so that $\mathscr{W}$ is graded by $\{\mathscr{W}_k\}_{k \geqslant 0}$.

We transform any $\mathbb{R}$-valued family $\{m_i\}_{i \in I}$ into a homomorphism from $\mathscr{W}$ to $\mathbb{R}$ by $m_{i_1 \cdots i_k} = m_{i_1} \cdots m_{i_k}$ for any word $i_1 \cdots i_k$ of $\mathscr{W}$.

Let us fix $n > 0$. Each $i \in I$ is associated with a function $f_i : \mathrm{V} \to \mathrm{V}$ of class $\mathscr{C}^n$ and globally Lipschitz. We define a Newtonian map F through

$$F[\varnothing] = \imath, \ F[i] = f_i \text{ for } i \in I \text{ and } F[iw](a) = \mathrm{D}F[w](a) \cdot F[i](a)$$

for any $w \in \mathscr{W}_{\leqslant n}$. For $w \in \mathscr{W}_{\leqslant n}$, $F[w] \in \mathscr{C}^{n+1-|w|}$. In particular, $F[iw]$ is continuous when $|w| = n$ and $i \in I$.

Proposition 6. *We set $\alpha = \sum_{i\in I} m_i i \in \mathscr{A}$ for $m_i \in \mathbb{R}$. Let y be the unique solution to the ODE*

$$y_t = a + \int_0^t \sum_{i\in I} m_i f_i(y_s)\,\mathrm{d}s = a + \int_0^t F[\alpha](y_s)\,\mathrm{d}s,\ \ t \geqslant 0 \text{ for } a \in \mathrm{V}.$$

For $w \in \mathscr{W}_{\leqslant n}$,

$$F[w](y_t) = \sum_{k=0}^{n-|w|} \frac{t^k}{k!} \sum_{i_1\cdots i_k \in I} F[m_{i_1\cdots i_k} i_1 \cdots i_k w](a) + R(t,a) \tag{45}$$

with

$$R(t,a) = \int_0^t \sum_{i_1\cdots i_{n-|w|}\in I} \frac{(t-s)^{n-|w|}}{(n-|w|)!} F[m_{i_1\cdots i_{n+1-|w|}} i_1 \cdots i_{n+1-|w|} w](y_s)\,\mathrm{d}s.$$

Proof. For $\alpha = \sum_{i\in I} m_i i$, and $t \geqslant 0$,

$$\exp(t\alpha) = \sum_{k\geqslant 0} \frac{t^k}{k!} \sum_{i_1\cdots i_k \in I} m_{i_1} \cdots m_{i_k} i_1 \cdots i_k = \sum_{k\geqslant 0} \frac{t^k}{k!} \sum_{i_1\cdots i_k \in I} m_{i_1\cdots i_k} i_1 \cdots i_k.$$

Since $\mathscr{W}_{\leqslant n}$ contains only words of size smaller than n,

$$\exp(t\alpha) \triangleright w = \sum_{k=0}^{n-|w|} \frac{t^k}{k!} \sum_{i_1\cdots i_k \in I} m_{i_1\cdots i_k} i_1 \cdots i_k w$$

$$\text{and } \alpha \blacktriangleright \exp((t-s)\alpha) \triangleright w = \frac{(t-s)^{n-|w|}}{(n-|w|)!} \sum_{i_1\cdots i_{n+1-|w|}\in I} m_{i_1\cdots i_{n+1-|w|}} i_1 \cdots i_{n+1-|w|} w.$$

This concludes the proof. □

We now show that it is also useful to recover some standard results on ODE.

Corollary 3 ([30, Lemma 5.4, p. 85])**.** *For i, j such that $f_i, f_j \in \mathscr{C}^1(\mathrm{V},\mathrm{V})$, we assume that the vector fields $f_i\mathrm{D}$ and $f_j\mathrm{D}$ commutes, that is $\mathrm{D}f_j \cdot f_i - \mathrm{D}f_i \cdot f_j$ vanishes. Then $\Phi[i] \circ \Phi[j] = \Phi[j] \circ \Phi[i]$.*

Proof. We set $n = 2$. For $\alpha, \beta \in \mathscr{A}$ and $t \geqslant 0$, a classical computation shows that

$$\exp_{\triangleright}(t\alpha) \triangleright \exp_{\triangleright}(t\beta) = \exp_{\triangleright}\left(t\alpha + t\beta + \frac{t^2}{2}[\alpha,\beta]\right)$$

with $[\alpha, \beta] := \alpha\beta - \beta\alpha$. We then define $(\alpha \star \beta)(t) := t\alpha + t\beta - \frac{t^2}{2}[\alpha, \beta]$. The composition formula yields

$$\Phi[\gamma(t)] = \Phi[t\beta] \circ \Phi[t\alpha] + \epsilon(t) \text{ with } |\epsilon(t)| \leqslant Ct^3$$

as an explicit computation on (45) shows it. If $F[ij] = F[ji]$, then $\Phi[(i \star j)(t)] = \Phi[(i + j)t]$. Hence, $|\Phi[t\beta] \circ \Phi[t\alpha] - \Phi[t\alpha] \circ \Phi[t\beta]| \leqslant Ct^3$. We conclude using the technique proposed in [30, Lemma III.5.4, p. 85]. □

Similarly, results such as convergence of the Strang splitting and the Lie-Trotter formula [30] may be recovered by this approach.

Remark 6. A variant consists in assuming that the weights m_i are piecewise continuous functions from $\mathbb{T}$ to $\mathbb{R}$, so that the ODE is a controlled one, where the m_i are the time derivatives of the piecewise continuous control $x : \mathbb{T} \to \mathbb{R}^{\dim(I)}$. This leads to similar algebraic computations.

Remark 7. When one considers the non-autonomous differential equation $\dot{y}_t = f(t, y_t)$, one may fix some integer $n \geqslant 1$ and set $m_i(t) = 1_{t \in [iT/n, (i+1)T/n}$ as well as $f_i(a) = f(iT/n, a)$. Using Remark 6, we then obtain discrete time approximations of the chronological expansion developed in [1] as well as the Magnus formula [6, 43, 30]. This also explains why the Magnus formula is sometimes called a *continuous-time Baker-Campbell-Hausdorff-Dynkin formula.*

5. High-order expansion of rough differential equations

We now consider the situation where $\mathscr{W}$ is a tensor algebra. In this case, our notion of RDE is the same as the one by A. M. Davie [22], P. Friz & N. Victoir [25, 26], I. Bailleul [2] as we saw in [15, 14] that the framework of almost flows encompasses all these constructions.

Here, we focus on studying high-order expansions of the numerical schemes. Such a study was initiated in [25]. It was then followed by [2, 4, 3, 10] and more recently by [41] where Runge-Kutta schemes were studied or in [5] for studying rough transport equation.

Notation 9 (Tensor algebra)**.** We consider a separable Banach space U with basis $\{e_i\}_{i \in I}$. We denote by T(U) its tensor algebra

$$\mathscr{W} := \mathrm{T}(\mathrm{U}) := \mathbb{R} \oplus \mathrm{U} \oplus \mathrm{U}^{\otimes 2} \oplus \cdots .$$

The tensor algebra T(U) is isomorphic to the algebra of words generated by the alphabet I. It is also graded by the $\mathscr{W}_k = \mathrm{U}^{\otimes k}$, $k \geqslant 0$.

Let us fix $n \geqslant 1$. We associate with each e_i a function f_i of class at most $\mathscr{C}^n$ and globally Lipschitz. We consider the differential operators defined by

$$F^\dagger[1]g := g,\ F^\dagger[e_i]g := \mathrm{D}g \cdot f_i \text{ and } F^\dagger[\alpha \otimes \beta]g = F^\dagger[\alpha]gF^\dagger[\beta]g.$$

Thus, $F^\dagger$ defines a Newtonian operator, which is easily put in relation with the Newtonian map of Section 4.

Notation 10 (Lie algebra)**.** We denote by $\mathscr{L}$ the smallest Lie algebra generated U and close under the Lie brackets $[\alpha, \beta] := \alpha \otimes \beta - \beta \otimes \alpha$.

The proof of the following result is standard. It assert that if $\alpha \in \mathscr{L}_{\leqslant n}$, then $F[\alpha]$ is actually a first order differential operator. It is proved inductively on the levels of brackets in g(U).

Lemma 11. *If $\alpha \in \mathscr{L}$, then $F^\dagger[\alpha]g(a) = \mathrm{D}g(a) \cdot F^\dagger[\alpha]\imath(a)$.*

We refer to [36, 26, 34, 31] for an extensive discussion about the properties of weak geometric rough paths.

Definition 10 (Weak geometric rough path)**.** Let $\mathbf{x}$ be a $\mathscr{W}_n$-valued p-rough path (possibly $n = +\infty$ with $\mathscr{W}_\infty = \mathscr{W}$). We say that $\mathbf{x}$ is a *weak geometric rough path* if $\log(\mathbf{x}_{s,t})$ belongs to the Lie algebra.

The Baker-Campbell-Hausdorff-Dynkin formula [12] states for for any $\alpha, \beta \in \mathscr{L}$, there exists a $\gamma \in \mathscr{L}$, which is explicitly computable from α and β such that

$$\exp(\alpha) \otimes \exp(\beta) = \exp(\gamma) \tag{46}$$

for some $\gamma \in \mathscr{L}$, where exp is the exponential in the tensor product. Since $\mathscr{W}_{>n}$ is a two-sided ideal, by identifying $\mathscr{W}_{\leqslant n}$ with the quotient space $\mathscr{W}/\mathscr{W}_{>n}$, there exists $\gamma_n \in \mathscr{L}_{\leqslant n}$ such that

$$\exp_\triangleright(\alpha) \triangleright \exp_\triangleright(\beta) = \exp_\triangleright(\gamma_n), \tag{47}$$

as $\triangleright$ is identified with the truncated tensor product $\otimes$. Therefore, Hypothesis 3 holds.

Moreover, when $\alpha \in \mathscr{L}_{\leqslant n}$, then $F^\dagger[\alpha]$ is a first-order differential operator by Lemma 11 as $\mathscr{L}$ is stable by projection onto $\mathscr{W}_{\leqslant n}$. This proves that $F^\dagger$ is a Newtonian operator with respect to the choice of the tensor algebra T(U) for $\mathscr{W}$ and its Lie algebra for $\mathscr{L}$.

We then recover the high-order expansion given in [25, 10] without relying on the use of sub-Riemannian geodesics. Thus, these results also hold true in an infinite dimensional setting, as already noted in [4, 13]. However, with respect to the two last works, we benefits from the decay estimate of [10].

Proposition 7 (Consistency)**.** *Under the above hypotheses, when* $\mathbf{x}$ *is a finite* p*-rough paths and* $f_i \in \mathscr{C}^n$ *with* $n+1 > p$*, any* D*-solution* z *satisfies*

$$|z_t - F^\dagger[\mathbf{x}_{s,t}^{(\leqslant n)}]\imath(z_s)| \leqslant C\omega_{s,t}^{(n+1)/p} \text{ and } |z_t - \phi_{t,s}(z_s)| \leqslant C\omega_{s,t}^{(n+1)/p}$$

where $\phi_{t,s}(a) = y[\log_\triangleright(\mathbf{x}_{s,t}^{(\leqslant n)})](a)$*.*

Coupled with the results in [14, 16], we also obtain a rate of convergence of $(n+1)/p - 1$ of the numerical scheme. This is the same rate as in [26, Theorem 10.3.3].

6. Branched rough paths

As in Sections 4 and 5, we consider an alphabet I whose letters i are associated with of function f_i. We introduce first a few notions on trees.

Definition 11 (Decorated tree)**.** A *I-decorated tree* is a tree whose vertices are associated with a letter i in I. We denote by $\mathcal{T}$ the set of decorated trees.

Notation 11 (Forest)**.** The formal, commutative and associative product of trees is called a *forest*. A forest can be seen as a monomial of the $\mathbb{R}$-module $\mathbb{R}\langle\mathcal{T}\rangle$ freely generated by the set $\mathcal{T}$. The neutral element of $\mathbb{R}\langle\mathcal{T}\rangle$, called the *empty tree*, is denoted by $\varnothing$.

Definition 12 (Grafting)**.** Let us conider m trees $\tau_1 \cdots \tau_m \in \mathcal{T}$ as well as a decoration $i \in I$. We denote by $\langle\tau_1, \ldots, \tau_m\rangle_i$ the tree obtained by grafting the roots of $\tau_1, \ldots, \tau_m$ to a new vertex which we decorate with i. The tree with a single root decorated by i is written $\bullet_i$. The number of nodes of $\tau \in \mathcal{T}_\star$ is denoted by $|\tau|$, with $|\varnothing| = 0$. It holds that $|\langle\tau_1, \ldots, \tau_m\rangle| = |\tau_1| + \cdots + |\tau_m| + 1$. The gratfting operation $(\tau_1, \ldots, \tau_m) \mapsto \langle\tau_1, \ldots, \tau_m\rangle$ is extended to $\mathbb{R}\langle\mathcal{T}\rangle^m$ by multilinearity.

Decorated rooted trees are defined recursively through the grafting operation from the family of single root trees $\{\bullet_i\}_{i\in I}$.

6.1. B-series

From the set $\{f_i\}_{i\in I}$ of functions, we define the elementary differentials which is a family of functions indiced by I-decorated rooted trees. The B in B-series stands for J.C. Butcher, who was the first to exhibit their algebraic features to study Runge-Kutta schemes in [18].

Definition 13 (Elementary differentials)**.** To each I-decorated rooted tree τ in $\mathscr{T}$, we associate the function $F[\tau] : \mathrm{V} \to \mathrm{V}$ by the following way

$$F[\varnothing](a) = a,\ F[\bullet_i](a) = f_i(a)$$
$$\text{and } F[\langle \tau_1, \dots, \tau_m \rangle_i](a) = \mathrm{D}^m f_i(a) \cdot F[\tau_1](a) \otimes \cdots \otimes F[\tau_m](a),$$

as long as the f_i have enough regularity to ensure that $F[\tau]$ is continuous. We extend F by linearity over $\mathbb{R}\langle \mathscr{T} \rangle$. Such an F is called an *elementary differential.*

To apply our construction, we define a new algebra $(\mathscr{W}, +, \cdot)$ from $\mathbb{R}\langle \mathscr{T} \rangle$ with the product defined as a bilinear relation recursively from

$$\sigma \cdot \bullet_i = \langle \sigma \rangle_i$$
$$\text{and } \sigma \cdot \langle \tau_1, \dots, \tau_m \rangle_i = \langle \sigma, \tau_1, \dots, \tau_m \rangle_i + \sum_{j=1}^{m} \langle \tau_1, \dots, \sigma \cdot \tau_j, \dots, \tau_m \rangle_i.$$

This algebra $\mathscr{W}$ is graded by the number of nodes.

Remark 8. The algebra $(\mathscr{W}, +, \cdot)$ is non-associative so that $\tau_1 \cdots \tau_m$ means $\tau_1 \cdot (\cdots (\tau_{m-1} \cdot \tau_m) \cdots)$.

Hypothesis 8. Let us fix $n \geqslant 1$ such that for each $\tau \in \mathscr{W}_{\leqslant n}$, $F[\tau] \in \mathscr{C}^1$.

Lemma 12. *Let σ, τ be a trees in $\mathscr{W}$. Then F is a Newtonian map.*

Example 6. Assume that I contains only a single letter, so that we drop any mention to it. Let y be the solution to the ODE $y_t = a + \int_0^t f(y_s)\,\mathrm{d}s = a + \int_0^t F[\bullet](y_s)\,\mathrm{d}s$. The Taylor formula at order n implies that

$$y_t = a + F[\exp_{\triangleright}(t\bullet)](a) + \mathrm{O}(t^{n+1}),$$

which in turns means that $y^{(m)}_{|t=0}(a) = F[\bullet^{\cdot m}](a)$, where $y^{(m)}$ is the m-th order derivative of y (in time).

For the reason given in Example 6, $F[\gamma]$ is called a *B-series.*

Proposition 8. *Let $\mathbf{x} := \{\mathbf{x}_{s,t}\}_{(s,t)\in\mathbb{T}^2_+}$ be a $\mathscr{W}_{\leqslant n}$-valued p-rough path with $n \geqslant \lfloor p \rfloor$. Then there exists a D-solution y associated with the almost flow ϕ defined by $\phi_{t,s}(a) := F[\mathbf{x}_{s,t}](a)$ and*

$$|y_t - F[\mathbf{x}_{s,t}](y_s)| \leqslant C\omega_{s,t}^{(n+1)/p}, \quad \textit{for } (s,t) \in \mathbb{T}^2_+.$$

We consider now a Banach space U with a basis $\{e_i\}_{i\in I}$ whose elements are then indiced by I. Now, let $\mathbf{x}$ be a $\mathrm{T}_{\leqslant 2}(\mathrm{U})$-valued p-rough path with $2 \leqslant p < 3$ We consider a family $\{f_i\}_{i\in I}$ of functions which are regular enough.

The *Davie approximation* is the almost flow (considered first by A.M. Davie [22]) defined by

$$\phi_{t,s}(a) := a + \sum_{i\in I} f_i(a)\mathbf{x}^{(1),i}_{s,t} + \sum_{i,j\in I} \mathbf{x}^{(2),ij}_{s,t} \mathrm{D}f_j(a)\cdot f_i(a) = F[\mathbf{x}_{s,t}](a). \tag{48}$$

which serves as an approximation to the RDE usually written $z_t = a + \int_0^t f(z_s)\,\mathrm{d}\mathbf{x}_s$, which actually means $z_t = a + \int_0^t F[\,\mathrm{d}\mathbf{x}_s](y_s)$.

The proof of the following theorem is only a rewriting using the formalism of trees. This is a result similar to the one of [19]. With [31, Proposition 3.8, p. 228], this leads to the same construction (only at the first level) of the RDE driven by branched rough paths introduced by [29].

Proposition 9. *The $\mathscr{W}_{\leqslant 2}$-valued family τ defined by*

$$\boldsymbol{\tau}_{s,t} := 1 + \sum_{i\in I} \mathbf{x}^{(1),i}_{s,t}\, \bullet_i + \sum_{i,j\in I} \mathbf{x}^{(2),ij}_{s,t} \langle \bullet_i \rangle_j.$$

is a $\mathscr{W}_{\leqslant 2}$-valued p-rough path and the Davie approximation defined by (48) *and rewritten*

$$\phi_{t,s}(a) = F[\boldsymbol{\tau}_{s,t}](a) \text{ for } a \in \mathrm{V},\ (s,t) \in \mathbb{T}^2_+$$

defines an almost flow.

Proof. In $\mathscr{W}_{\leqslant 2}$ (recall that $\triangleright$ is nothing more than $\cdot$ in the quotient space $\mathscr{W}_{\leqslant 2}$),

$$\boldsymbol{\tau}_{r,s} \triangleright \boldsymbol{\tau}_{s,t} = 1 + \sum_{i\in I} \bullet_i \mathbf{x}^{(1),i}_{r,t} + \sum_{i,j\in I} \langle \bullet_j \rangle_i (\mathbf{x}^{(2),ij}_{r,s} + \mathbf{x}^{(2),ij}_{s,t}) + \sum_{i,j\in I} \bullet_i \cdot \bullet_j \mathbf{x}^{(1),i}_{r,s} \mathbf{x}^{(2),j}_{s,t},$$

since $\mathbf{x}^{(1)}_{r,s} + \mathbf{x}^{(1)}_{s,t} = x^1_{r,t}$. Regarding the term of order 2, $\mathbf{x}^{(2)}_{r,t} = \mathbf{x}^{(2)}_{r,s} + \mathbf{x}^{(2)}_{s,t} + \mathbf{x}^{(1)}_{r,s} \otimes \mathbf{x}^{(1)}_{s,t}$ in $\mathrm{T}_{\leqslant 2}(\mathrm{U})$. We check that $\bullet_i \cdot \bullet_j = \langle \bullet_i \rangle_j$. Thus, $\boldsymbol{\tau}$ is multiplicative in $\mathscr{W}_{\leqslant 2}$.

The control over $|\boldsymbol{\tau}^{(i)}_{s,t}|$ are easily obtained. □

The main implication of Proposition 9 is the convergence of higher-order numerical schemes constructed from B-series, also called Runge-Kutta schemes.

Corollary 4. *Let $n \geqslant 3$. We consider $\boldsymbol{\tau}$ given in Proposition 9. Any $\mathscr{W}_{\leqslant n}$-valued p-rough path $\boldsymbol{\sigma}$ with $\boldsymbol{\sigma}^{(\leqslant 2)}_{s,t} = \boldsymbol{\tau}_{s,t}$ for any $(s,t) \in \mathbb{T}^2_+$ gives rise to the sams flow as the one that stems from the approxiation $\{F[\boldsymbol{\tau}_{s,t}]\}_{(s,t)\in\mathbb{T}^2_+}$*

Finally, in $\mathscr{W}_{\leqslant 2}$,

$$\log(\boldsymbol{\tau}_{s,t}) = \sum_{i\in I} \bullet_i \mathbf{x}_{r,t}^{(1),i} + \sum_{i,j\in I} \langle \bullet_j \rangle_i (\mathbf{x}_{r,s}^{(2),ij} + \mathbf{x}_{s,t}^{(2),ij}) - \frac{1}{2} \sum_{i,j\in I} \mathbf{x}^{(1),i} \mathbf{x}^{(1),j} \langle \bullet_i \rangle_j.$$

The next lemma is a rewritting in the context of branched trees on the definition of weak geometric rough paths [34, 26, 33].

Lemma 13. *A weak geometric p-rough path is a rough path $\mathbf{x}$ such that $\mathbf{x}_{s,t} = \exp(\boldsymbol{\lambda}_{s,t})$ where $\boldsymbol{\lambda}_{s,t}$ belongs to the Lie algebra $\mathscr{L}$ for any $(s,t) \in \mathbb{T}_+^2$. In $\mathscr{W}_{\leqslant 2}$,*

$$\mathbf{x}_{s,t} = 1 + \mathbf{x}_{s,t}^{(1)} + \frac{1}{2} \mathbf{x}_{s,t}^{(1)} \otimes \mathbf{x}_{s,t}^{(1)} + \mathbf{a}_{s,t} \tag{49}$$

where $\mathbf{a}_{s,t} \in \mathrm{U} \otimes \mathrm{U}$ is anti-symmetric, that is $\mathbf{a}_{s,t}^{ij} = -\mathbf{a}_{s,t}^{ji}$ for any $i, j \in I$.

It follows from Lemma 13 that if $\boldsymbol{\tau}_{s,t}$ proceed from a weak geometric $\mathrm{T}_{\leqslant 2}(\mathrm{U})$-valued rough path as in (49), then

$$\boldsymbol{\lambda}_{s,t} := \log_{\triangleright}(\boldsymbol{\tau}_{s,t}) = \mathbf{x}_{s,t}^{(1)} + \sum_{i,j\in I} \langle \bullet j \rangle_i \mathbf{a}_{s,t}^{ij}.$$

One recovers the construction from I. Bailleul [2, 4] when one considers as an almost flow $\{\Phi[\boldsymbol{\lambda}_{s,t}]\}_{(s,t)\in\mathbb{T}_+^2}$.

Let us end this section by noting that for any $\mathscr{W}_{\leqslant n}$-valued p-rough path $\boldsymbol{\tau}$, it is always possible to compute $\boldsymbol{\lambda}_{s,t} = \log_{\triangleright}(\boldsymbol{\tau}_{s,t})$, whether or not the path is a weak geometric rough path.

6.2. Aromatic rough paths

The notion of aromatic tree extends the one of rooted trees. Aromatic trees were introduced in [39, 37] as an extension of B-series to study equivarient numerical schemes.

Definition 14 (Directed graph)**.** A directed graph g is defined as a finite set of vertices $V(g)$ and edges $E(g) \subset V(g) \times V(g)$. We allow the empty graph $\varnothing$ with no vertices. Two graphs g and h are *equivalent* is there exists a one-to-one map $f : V(g) \to V(h)$ such that $(f \times f)(E(g)) = E(h)$.

Definition 15 (Aromatic forest)**.** An *aromatic forest* is an equivalence class of directed graphs where each node has at most one outgoing edge. A *root* of an aromatic forest is a node with no outgoing edge.

We now consider that $\mathrm{V} := \mathbb{R}^m$. Let I be a set of letters. To each letter $\alpha \in I$, we associate a function $f[\alpha] : \mathrm{V} \to \mathrm{V}$ regular enough. The i-th component of $f[\alpha]$ is written $f^i[\alpha]$.

Definition 16 (Elementary differential operator)**.** We consider a I-decorated aromatic forest τ. We associate with τ a differential operator $F^\dagger[\tau]$ constructed the following way:

- Each vertex v of τ is also associated with $i_v \in \{1, \dots, m\}$ and thus decorated by $(\tau_v, i_v) \in \mathscr{L} \times \{1, \dots, m\}$.
- Let v be a vertex of τ with s ingoing edges from the vertices labelled by $q_1, \dots, q_s$ regarding their components in $\{1, \dots, m\}$. We then consider the factor
$$f(v, i_v, i_1, \dots, i_s) := \frac{\partial f^{i_v}[\tau]}{\partial x_{i_1} \cdots \partial x_{i_s}},$$
where $f^j[\tau_v]$ is the j-component of $f[\tau_v]$. If the vertex has no ingoing edge, then $f(v, i_1) := f[\tau_v]^{i_v}$.
- To each root s of τ, we also associate the factor $\frac{\partial}{\partial x_s}$.
- We multiply the factors and we sum according the Einstein summation convention, *i.e.*, we sum over each factor appearing twice, lower (differentiation) and upper.

Example 7. The tree $\bullet_\alpha$ which gives rise to the vector field $F^\dagger[\bullet_\alpha] := \sum_i f[\alpha]^i \partial_{x_i}$. while
$$\langle \bullet_\alpha \rangle_\beta \text{ gives rise to } F^\dagger[\langle \bullet_\alpha \rangle_\beta] := \sum_{i,j} f^i[\alpha] \frac{\partial f^j[\beta]}{\partial x_i} \frac{\partial}{\partial x_j}$$
and
$$\langle \bullet_\alpha \bullet_\gamma \rangle_\beta \text{ gives rise to } F^\dagger[\langle \bullet_\alpha \bullet_\gamma \rangle_\beta] := \sum_{i,j,k} f^i[\alpha] f^k[\gamma] \frac{\partial^2 f^j[\beta]}{\partial x_i \partial_{x_k}} \frac{\partial}{\partial x_j}.$$
On the other hand, an aroma with a single vertex, which is simply a loop with a single edge both ingoing and outgoing is associated with the divergence $\sum_i f_i$.

Remark 9. Elementary differentials which are differential operators are first order are encoded by aromatic forest with a single root.

Definition 17 (Composition of I-decorated aromatic forest)**.** Given two aromatic I-decorated aromatic forests, we define their products $\sigma\tau$ as the sum of the graphs γ obtained the following way:

- we superpose σ and τ, that is we construct a graph whose edges (resp. vertices) is obtained by the union of the edges (resp. vertices) of σ and τ.
- for each root s of σ and each node t of τ, we form a graph as above to which we add en edge outgoing from s and ingoing from t.

This product rule (see [11, Definition 3.7, p. 205]) is the one that corresponds to the product of two elementary differentials. The next lemma follows from [11, Lemma 3.8, p. 205] and Remark 9 above.

Lemma 14. *Let σ and τ be two I-decorated aromatic forests be such that σ has only one root. Then $\sigma\tau$ has only one root. Let $\mathscr{L}$ be the $\mathbb{R}$-submodule freely generated by I-decorated rooted trees with one root. Assume that $F^{\dagger}[\alpha]\imath$ is Lipschitz continuous for any $\alpha \in \mathscr{L}_{\leqslant n}$. Then $F^{\dagger}$ is a Newtonian operator.*

It is then possible to extend the notion of RDE driven by branched rough paths to RDE driven by aromatic rough paths.

A. A Davie lemma

There are several forms of the Davie lemma. The one presented here is a simplified form of the one in [14] and of [26, Lemma 10.59, p. 270] for the special case $\varpi(x) = x^{\theta}$, $\theta > 1$.

Lemma 15 (A Davie's lemma)**.** *Let ϖ be such that $\varpi(0) = 0$ and $2\varpi(x) \leqslant \kappa\varpi(2x)$ for some $\kappa < 1$. Let $\{U_{s,t}\}_{0\leqslant s\leqslant t\leqslant T}$ be a $\mathbb{R}_+$-valued family as well as $E, M \geqslant 0$ such that*

$$U_{s,t} \leqslant E\varpi(\omega_{s,t}) \text{ and } U_{r,t} \leqslant U_{r,s} + U_{s,t} + M\varpi(\omega_{r,t})$$

for any $0 \leqslant r \leqslant s \leqslant t \leqslant T$. Then $U_{r,t} \leqslant M/(1-\varkappa)\varpi(\omega_{r,t})$.

Acknowledgement. The author thanks Laure Coutin and Antoine Brault for interesting discussions related to this article.

References

[1] A. A. Agračev and R. V. Gamkrelidze. Exponential representation of flows and a chronological enumeration. *Math. USSR Sb.*, 35:727–785, 1979.

[2] I. Bailleul. Flows driven by rough paths. *Rev. Mat. Iberoamericana*, 31(3):901–934, 2015.

[3] I. Bailleul and S. Riedel. Rough flows. *J. Math. Soc. Japan.*, 71(3):915–978, 2019.

[4] Ismaël Bailleul. Flows driven by Banach space-valued rough paths. In *Séminaire de Probabilités XLVI*, volume 2123 of *Lecture Notes in Math.*, pages 195–205. Springer, Cham, 2014.

[5] Carlo Bellingeri, Ana Djurdjevac, Peter K. Friz, and Nikolas Tapia. Transport and continuity equations with (very) rough noise, 2020.

[6] S. Blanes, F. Casas, J. A. Oteo, and J. Ros. The Magnus expansion and some of its applications. *Phys. Rep.*, 470(5-6):151–238, 2009.

[7] Horatio Boedihardjo. Factorial decay of iterated rough integrals, 2016.

[8] Horatio Boedihardjo. Decay rate of iterated integrals of branched rough paths. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 35(4):945–969, 2018.

[9] Horatio Boedihardjo and Ilya Chevyrev. An isomorphism between branched and geometric rough paths. *Ann. Inst. H. Poincaré Probab. Statist.*, 55(2):1131–1148, 2019.

[10] Horatio Boedihardjo, Terry Lyons, and Danyu Yang. Uniform factorial decay estimates for controlled differential equations. *Electron. Commun. Probab.*, 20:no. 94, 11, 2015.

[11] Geir Bogfjellmo. Algebraic structure of aromatic B-series. *J. Comput. Dyn.*, 6(2):199–222, 2019.

[12] Andrea Bonfiglioli and Roberta Fulci. *Topics in noncommutative algebra*, volume 2034 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2012. The theorem of Campbell, Baker, Hausdorff and Dynkin.

[13] Youness Boutaib, Lajos Gergely Gyurkó, Terry Lyons, and Danyu Yang. Dimension-free Euler estimates of rough differential equations. *Rev. Roumaine Math. Pures Appl.*, 59(1):25–53, 2014.

[14] A. Brault and A. Lejay. The non-linear sewing lemma II: Lipschitz continuous formulation, 2018.

[15] A. Brault and A. Lejay. The non-linear sewing lemma I: weak formulation. *Electronic Journal of Probability*, 24(59):1–24, 2019.

[16] A. Brault and A. Lejay. The non-linear sewing lemma III: Stability and generic properties, *Forum Mathematicum*, to appear, 2020.

[17] Yvain Bruned, Charles Curry, and Kurusch Ebrahimi-Fard. Quasi-shuffle algebras and renormalisation of rough differential equations. *Bull. London Math. Soc.*, 52:43–63, 2020.

[18] J. C. Butcher. An algebraic theory of integration methods. *Mathematics of Computation*, 26(117):79–79, January 1972.

[19] Thomas Cass and Martin P. Weidner. Tree algebras over topological vector spaces in rough path theory, 2016.

[20] Kuo-Tsai Chen. Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula. *Ann. of Math. (2)*, 65:163–178, 1957.

[21] Charles Curry, Kurusch Ebrahimi-Fard, Dominique Manchon, and Hans Z. Munthe-Kaas. Planarly branched rough paths and rough differential equations on homogeneous spaces, 2018.

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

[23] J. J. Duistermaat and J. A. C. Kolk. *Multidimensional real analysis. I. Differentiation*, volume 86 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2004. Translated from the Dutch by J. P. van Braam Houckgeest.

[24] Denis Feyel, Arnaud de La Pradelle, and Gabriel Mokobodzki. A non-commutative sewing lemma. *Electron. Commun. Probab.*, 13:24–34, 2008.

[25] Peter Friz and Nicolas Victoir. Euler estimates for rough differential equations. *J. Differential Equations*, 244(2):388–412, 2008.

[26] 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.

[27] Robert L. Grossman and Richard G. Larson. Differential algebra structures on families of trees. *Adv. in Appl. Math.*, 35(1):97–119, 2005.

[28] Massimiliano Gubinelli. Controlling rough paths. *Journal of Functional Analysis*, 216(1):86–140, 2004.

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

[30] Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric numerical integration*, volume 31 of *Springer Series in Computational Mathematics*. Springer, Heidelberg, 2010. Structure-preserving algorithms for ordinary differential equations, Reprint of the second (2006) edition.

[31] Martin Hairer and David Kelly. Geometric versus non-geometric rough paths. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(1):207–251, 2015.

[32] Keisuke Hara and Masanori Hino. Fractional order Taylor's series and the neo-classical inequality. *Bull. Lond. Math. Soc.*, 42(3):467–477, 2010.

[33] Antoine Lejay and Nicolas Victoir. On (p, q)-rough paths. *J. Differential Equations*, 225(1):103–133, 2006.

[34] Terry Lyons and Zhongmin Qian. *System control and rough paths*. Oxford Mathematical Monographs. Oxford University Press, Oxford, 2002.

[35] Terry Lyons and Nicolas Victoir. An extension theorem to rough paths. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 24(5):835–847, 2007.

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

[37] Robert I. McLachlan, Klas Modin, Hans Munthe-Kaas, and Olivier Verdier. B-series methods are exactly the affine equivariant methods. *Numer. Math.*, 133(3):599–622, 2016.

[38] J. Mostovoy, J. M. Pérez-Izquierdo, and I. P. Shestakov. A non-associative Baker-Campbell-Hausdorff formula. *Proc. Amer. Math. Soc.*, 145(12):5109–5122, 2017.

[39] Hans Munthe-Kaas and Olivier Verdier. Aromatic Butcher series. *Found. Comput. Math.*, 16(1):183–215, 2016.

[40] A. Murua. Formal series and numerical integrators. I. Systems of ODEs and symplectic integrators. *Appl. Numer. Math.*, 29(2):221–251, 1999.

[41] Martin Redmann and Sebastian Riedel. Runge-kutta methods for rough differential equations, 2020.

[42] Sebastian Reich. Backward error analysis for numerical integrators. *SIAM J. Numer. Anal.*, 36(5):1549–1570, 1999.

[43] Robert S. Strichartz. The Campbell-Baker-Hausdorff-Dynkin formula and solutions of differential equations. *J. Funct. Anal.*, 72(2):320–345, 1987.

[44] Nikolas Tapia and Lorenzo Zambotti. The geometry of the space of branched rough paths, 2018.