A geometric view on rough differential equations

a thesis presented for the degree of

Doctor of Philosophy of Imperial College London

and the

Diploma of Imperial College

by

Martin Philipp Weidner

Department of Mathematics
Imperial College
London, SW7 2AZ
United Kingdom

July 2018

Declaration of originality

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Martin Weidner

Copyright declaration

Abstract

Our contribution to the theory of rough paths is twofold. On the one hand we introduce tree algebras over topological vector spaces which yields a framework for branched rough path in infinite dimensional spaces. In particular, we derive a change-of-variables formula in this settings from a purely combinatorial argument and we revisit the relationship between geometric, branched and general rough paths and we discuss to what extent each of them can be used to drive differential equations on manifolds.

On the other hand we discuss global solutions of rough differential equations on manifolds. We show that instead of considering bounds for the full higher order derivatives of the vector fields it is enough to look at Lie brackets. In our proofs we extend methods from Riemannian geometry that have been used to obtain characterisations of complete vector fields on manifolds. This leads to a unified result for rough differential equations that captures examples such as bounded vector fields, vector fields of linear growth or complete commuting vector fields all at once. In particular, our result holds for vector spaces and it improves existing non-explosion criteria in this case. We present an example of an equation which has global solutions for all initial values but where existing methods were not sufficient to give a proof.

Finally, we give an outlook as to where the new results may lead to further developments.

Acknowledgements

First and foremost I would like to thank my supervisor Thomas Cass for the great support he provided while I was working on this thesis. He always provided valuable feedback that guided me along my way while he also fully believed in me so that I could follow my own ideas.

Many fruitful thoughts and ideas that went into this work have originated from the regular Berlin-Oxford young researchers meetings and therefore I would like to thank Peter Friz and Terry Lyons as well as their co-organisers for their efforts and initiative.

I learned about most of the algebraic aspects that went into Chapter 4 during the 2015 CIMPA summer school on interactions of combinatorics and probability in São Carlos and I would like to thank the organisers and speakers for those helpful two weeks.

Furthermore I want to thank Anderson Santos and Marta Guzzon for their constant help in overcoming all kinds of administrative hurdles.

My time at Imperial College would have been a lot more boring without all my fellow doctoral students, especially those in office 6M09. Our lunch breaks, football in Hyde park, and mathematical discussions have always been a pleasure.

Finally, I thank my parents for their support, be it financially or otherwise, and most of all I thank my girlfriend Theresa for being as wonderful as she is.

Contents

1. Introduction

Consider the rough differential equation

$$\mathrm{d}y_t = V(y_t)\mathrm{d}\mathbf{x}_t.$$

It can be understood as a system which transforms an input signal $\mathbf{x}$, which is given in the form of a rough path, into a response y, which is given by the solution of the equation. It is therefore an important first step to fix the categories to which the input space and the response space should belong. Originally [Lyo98], the theory has been developed using Banach spaces as both input and response spaces and the linear structure of both spaces has been exploited extensively. Since then, various alternative approaches to rough differential equations [Dav08, Gub04, FV10, Bai15] have been introduced, some of which rely considerably less on the linear structure of the response space. The main motif of the present work is the hypothesis that using manifolds instead of linear spaces on the response side not only allows for a wider range of applications, but also leads to a better understanding of the theory as one avoids constraints that stem from overemphasising the linear structure. Ultimately, our main new result that follows from these considerations is Theorem 5.16, which sheds a new light on the question of global solvability of rough differential equations. It improves existing results for vector spaces even though it is formulated for general manifolds.

The idea to get rid of the linear structure of the input space or the response space (or both) is of course not new. We would like to mention the works [CLL11,Bai14,BL15,CDL15,DS17], which have in common that they establish various notions of rough paths on manifolds. In the present work we do not need these concepts since the input space will be a linear space throughout and the response will be an ordinary continuous path on a manifold without any additional terms of higher order.

When we replace the response space by a manifold, then the behaviour of the rough differential equation under a change of coordinates becomes relevant. As long as the input space is finite-dimensional, one can approximate geometric rough paths by smooth paths (see [FV06] for a discussion of such approximation results) in order to show that the behaviour of differential equations under a change of coordinates carries over from smooth paths to rough paths. In contrast, approximation arguments can be delicate in infinite-dimensional spaces. A second problem with the aforementioned approximations arises when branched rough paths [Gub10] come into play. It has been shown that branched rough paths can be treated as geometric rough paths [HK15, BC17] and therefore branched rough paths are in principle accessible by all arguments that can be used for geometric rough paths. However, we believe that this indirect reasoning hides some crucial insights and it turns out that in our proof of a chain rule in Theorem 4.31 branched rough paths are in fact the more natural objects to work with rather than geometric rough paths. Furthermore, our proof of Theorem 4.31 overcomes the restrictive assumption that the input-space is finite-dimensional since we do not use smooth approximations. Instead, our result follows from purely combinatorial considerations that are based on duality relations in the underlying algebraic structure. In particular, we provide a systematic treatment of branched rough paths over infinite-dimensional spaces in contrast to previous works such as [Gub10, HK15], where the authors have only discussed the finite-dimensional case. In fact, even our underlying construction of tree algebras over infinite-dimensional topological vector spaces seems to be new.

The content of Chapter 4 is also available in [CW17] and has been submitted for publication. A publication based on Chapter 5 is currently under preparation.

The following chapters are organised as follows. In Chapter 2 we present basic results about bialgebras (and in particular the tensor algebra over a vector space), smooth manifolds, Riemannian metrics and ordinary differential equations. Those results form the basis upon which the subsequent chapters are built. Chapter 3 contains a brief introduction to the concept of rough paths and rough differential equations. In light of the various ways to treat rough differential equations that have been introduced so far, our principal aim is to identify a setting that is well-suited to be generalised to manifolds. In Chapter 4 we develop our first main contribution, namely a framework for branched rough paths in infinite-dimensional spaces along with a chain rule for rough differential equations. Furthermore we revisit the relationship between general, geometric and branched rough paths in the context of differential geometry. Finally, Chapter 5 contains our second main contribution, which is a new approach towards global solvability of rough differential equations. Open questions and an outlook towards future research are then discussed in Chapter 6.

2. Preliminaries

In this chapter we introduce notions and results from several areas which will be relevant in the rest of this work. Most of the content of this chapter can be found in the standard literature about the respective topics. In particular, we would like to mention [LCL07] as a general resource for the algebraic background of rough path theory as well as [DNR00,Reu93] for more details on Hopf algebras and related topics. We also give some proofs of well-established results. Partly this is for the convenience of the reader and partly to present arguments which we will extend and modify in later chapters.

2.1. The tensor algebra over a vector space

The aim of this section is to introduce the tensor algebra over a vector space along with some of its fundamental properties to the extent that is necessary to introduce rough paths. The content of this section also serves as a basis for our discussion of tree algebras over vector spaces in Chapter 4.

Let E be a vector space. E can be infinite-dimensional in general, even though we will occasionally assume $E = \mathbf{R}^d$ later on, especially in Chapter 5. For $k \in \mathbf{N}$ we denote by $E^{\otimes k}$ the k-th tensor power of E. Later on we will consider tensor products of the tensor algebra with itself when we discuss coproducts. To make sure that this does not lead to confusion, we will not write the $\otimes$-operator explicitly when working with elements of $E^{\otimes k}$ for $k \in \mathbf{N}$. E.g., we write $e_1 e_2$ instead of $e_1 \otimes e_2$ for $e_1, e_2 \in E$. We say that an element $a \in E^{\otimes k}$ is an *elementary tensor* if there exist vectors $e_1, \dots, e_k \in E$ such that $a = e_1 \cdots e_k$.

Definition 2.1 The *tensor algebra* over E is the set

$$T(E) := \bigoplus_{k=0}^{\infty} E^{\otimes k} = \left\{ \sum_{k=0}^{\infty} \lambda_k a^k : a^k \in E^{\otimes k}, \lambda_k = 0 \text{ for almost all } k \in \mathbf{N} \right\}$$

together with tensor multiplication $\otimes$. Equivalently, $T(E)$ is the free algebra generated by any basis of E.

Let us also note that $T(E)$ carries a natural grading with the homogeneous subspaces given by the tensor powers of E. For $k \in \mathbf{N}$ we denote by π^k the projection to $E^{\otimes k}$. The product $\otimes$ is compatible with this grading in the sense that for $k, l \in \mathbf{N}$ and $a \in E^{\otimes k}, b \in E^{\otimes l}$ we have $ab \in E^{\otimes(k+l)}$.

2.1.1. The bialgebra structure of the tensor algebra

In addition to the tensor product we can equip the set of tensors with a coproduct which turns it into a bialgebra. Let us start by introducing the necessary terminology.

Definition 2.2 Let H be a vector space, assume that there are linear functions

$$\begin{aligned} \star &: H \otimes H \to H \ (\textit{product}) & \Delta &: H \to H \otimes H \ (\textit{coproduct}) \\ \eta &: \mathbf{R} \to H \ (\textit{unit}) & \varepsilon &: H \to \mathbf{R} \ (\textit{counit}) \\ S &: H \to H \ (\textit{antipode}) \end{aligned}$$

and consider the statements

(i) $(a \star b) \star c = a \star (b \star c)$ (*associativity*)
(ii) $\eta(1) \star a = a \star \eta(1) = a$ (*unitality*)
(iii) $(\mathrm{id} \otimes \Delta) \circ \Delta = (\Delta \otimes \mathrm{id}) \circ \Delta$ (*coassociativity*)
(iv) $(\varepsilon \otimes \mathrm{id}) \circ \Delta = (\mathrm{id} \otimes \varepsilon) \circ \Delta = \mathrm{id}$ (*counitality*)

$$\left.\begin{array}{ll} \text{(v)} & \Delta(a \star b) = \Delta(a) \star \Delta(b) \\ \text{(vi)} & \Delta \circ \eta = \eta \otimes \eta \\ \text{(vii)} & \varepsilon(a \star b) = \varepsilon(a)\varepsilon(b) \\ \text{(viii)} & \varepsilon \circ \eta = \mathrm{id} \end{array}\right\} (\textit{compatibility})$$

(ix) $\star \circ (S \otimes \mathrm{id}) \circ \Delta = \star \circ (\mathrm{id} \otimes S) \circ \Delta = \eta \circ \varepsilon$

for all $a, b, c \in H$.

Then

- $(H, \star, \eta)$ is called an *algebra* if (i) and (ii) hold,
- (H, Δ, ε) is called a *coalgebra* if (iii) and (iv) hold,
- $(H, \star, \eta, \Delta, \varepsilon)$ is called a *bialgebra* if (i)–(viii) hold and
- $(H, \star, \eta, \Delta, \varepsilon, S)$ is called a *Hopf algebra* if (i)–(ix) hold.

Remark 2.3 Both the unit and the counit are uniquely determined by the product and the coproduct respectively if they exist. In the examples that we will cover they will always exist and it will be obvious how to choose them. Therefore we will never mention them explicitly. Also, (ii) implies that η is injective and thus we can and will identify $\eta(\mathbf{R})$ with

R. Furthermore, a given bialgebra can have at most one antipode that turns it into a Hopf algebra [DNR00, Rem. 4.2.3]. Even though all of the bialgebras that we will be dealing with admit an antipode, we will never exploit this fact and therefore we will not give explicit descriptions of the antipodes. See Appendix A for more details.

Before we proceed, let us introduce a useful notation for coproducts, called the *Sweedler notation.* If (C, Δ) is a coalgebra we define

$$c_{(1)} \otimes c_{(2)} := \Delta(c) \tag{2.1}$$

for $c \in C$. This notation hides the fact that $\Delta(c)$ is in general a sum of tensors rather than an elementary tensor of $C \otimes C$ and therefore some care has to be taken. To illustrate the usefulness of this notation, we can write (ix) of Definition 2.2 as

$$S(h_{(1)}) \star h_{(2)} = h_{(1)} \star S(h_{(2)}) = \eta(\varepsilon(h))$$

for $h \in H$. Furthermore coassociativity (iii) of Δ reads as

$$c_{(1)} \otimes c_{(2)(1)} \otimes c_{(2)(2)} = c_{(1)(1)} \otimes c_{(1)(2)} \otimes c_{(2)}$$

and thus

$$c_{(1)} \otimes c_{(2)} \otimes c_{(3)} := (\Delta \otimes \mathrm{id})(\Delta(c)) = (\mathrm{id} \otimes \Delta)(\Delta(c))$$

is well-defined.

Let us now see how Definition 2.2 can be applied to the case $H = T(E)$. As a product we will use the standard tensor product $\otimes$ (even though we will usually omit the operator $\otimes$ in order to avoid confusion as described above). As a coproduct we use the *shuffle coproduct* $\Delta_{\sqcup\!\sqcup}$, which is given recursively by

$$\Delta_{\sqcup\!\sqcup}(e) = e \otimes 1 + 1 \otimes e$$

for $e \in E$ and by

$$\Delta_{\sqcup\!\sqcup}(ab) = \Delta_{\sqcup\!\sqcup}(a)\Delta_{\sqcup\!\sqcup}(b)$$

for $a, b \in T(E)$. Since E generates $T(E)$ this defines $\Delta_{\sqcup\!\sqcup}$ for all elements of $T(E)$. In order to given an explicit, non-recursive formula for $\Delta_{\sqcup\!\sqcup}$ we have to introduce the set of *shuffle permutations.* For $n \in \mathbf{N}$ let $\mathfrak{S}_n$ be the symmetric group over n elements. For $k, l \in \mathbf{N}$ the subset $\mathrm{Sh}(k,l) \subset \mathfrak{S}_{k+l}$ is given by

$$\mathrm{Sh}(k,l) := \{\sigma \in \mathfrak{S}_{k+l} : \sigma(1) < \cdots < \sigma(k), \sigma(k+1) < \cdots < \sigma(k+l)\} .$$

Proposition 2.4 *The shuffle-coproduct $\Delta_{\sqcup\sqcup}$ is given by*

$$\Delta_{\sqcup\sqcup}(a) := \sum_{k=0}^{n} \sum_{\sigma \in \mathrm{Sh}(k,n-k)} (e_{\sigma(1)} \cdots e_{\sigma(k)}) \otimes (e_{\sigma(k+1)} \cdots e_{\sigma(n)}) \in T(E) \otimes T(E)$$

for an elementary tensor $a = e_1 \cdots e_n \in E^{\otimes n}$.

Proof. For $n = 1$ this is clear. Hence we assume that the claim holds for some $n \in \mathbf{N}$. Then we have for $a = e_1 \cdots e_n \in E^{\otimes n}$ and $e_{n+1} \in E$

$$\begin{aligned}
&\Delta_{\sqcup\sqcup}(ae_{n+1}) = \Delta_{\sqcup\sqcup}(a)\Delta_{\sqcup\sqcup}(e_{n+1}) \\
&= \left(\sum_{k=0}^{n} \sum_{\sigma \in \mathrm{Sh}(k,n-k)} (e_{\sigma(1)} \cdots e_{\sigma(k)}) \otimes (e_{\sigma(k+1)} \cdots e_{\sigma(n)}) \right) (e_{n+1} \otimes 1 + 1 \otimes e_{n+1}) \\
&= \sum_{k=0}^{n} \sum_{\sigma \in \mathrm{Sh}(k,n-k)} \big[(e_{\sigma(1)} \cdots e_{\sigma(k)} e_{n+1}) \otimes (e_{\sigma(k+1)} \cdots e_{\sigma(n)}) \\
&\qquad\qquad + (e_{\sigma(1)} \cdots e_{\sigma(k)}) \otimes (e_{\sigma(k+1)} \cdots e_{\sigma(n)} e_{n+1}) \big] \\
&= \sum_{k=0}^{n+1} \sum_{\sigma \in \mathrm{Sh}(k,n+1-k)} (e_{\sigma(1)} \cdots e_{\sigma(k)}) \otimes (e_{\sigma(k+1)} \cdots e_{\sigma(n+1)})
\end{aligned}$$

and thus the claim follows. ■

Proposition 2.5 *The tuple*

$$(T(E), \otimes, \Delta_{\sqcup\sqcup})$$

is a bialgebra.

Proof. We have to verify (i)–(viii) of Definition 2.2. The algebra properties (i) and (ii), i.e. the associativity of $\otimes$ and the unitality are clear.

Let us show that (iii) holds, i.e. that $\Delta_{\sqcup\sqcup}$ is coassociative. For $e \in E$ we have

$$(\Delta_{\sqcup\sqcup} \otimes \mathrm{id})(\Delta_{\sqcup\sqcup}(e)) = (\mathrm{id} \otimes \Delta_{\sqcup\sqcup})(\Delta_{\sqcup\sqcup}(e)) = 1 \otimes 1 \otimes e + 1 \otimes e \otimes 1 + e \otimes 1 \otimes 1.$$

Now assume that (iii) holds for all $a \in E^{\otimes n}$ for some $n \in \mathbf{N}$. Then, for $a \in E^{\otimes n}$ and $e \in E$ we have

$$
\begin{aligned}
&(\Delta_{\sqcup\sqcup} \otimes \mathrm{id})(\Delta_{\sqcup\sqcup}(ae)) \\
=&(\Delta_{\sqcup\sqcup} \otimes \mathrm{id})(\Delta_{\sqcup\sqcup}(a))(\Delta_{\sqcup\sqcup} \otimes \mathrm{id})(\Delta_{\sqcup\sqcup}(e)) = (\mathrm{id} \otimes \Delta_{\sqcup\sqcup})(\Delta_{\sqcup\sqcup}(a))(\mathrm{id} \otimes \Delta_{\sqcup\sqcup})(\Delta_{\sqcup\sqcup}(e)) \\
=&(\mathrm{id} \otimes \Delta_{\sqcup\sqcup})(\Delta_{\sqcup\sqcup}(ae)),
\end{aligned}
$$

which shows that (iii) holds for all $a \in E^{\otimes(n+1)}$ and thus for all $a \in T(E)$. The counitality (iv) follows from a similar inductive argument.

The crucial part of the compatibility conditions is (v), which follows directly from our recursive definition of $\Delta_{\sqcup\sqcup}$, since we have defined $\Delta_{\sqcup\sqcup}$ in such a way that

$$\Delta_{\sqcup\sqcup}(ab) = \Delta_{\sqcup\sqcup}(a)\Delta_{\sqcup\sqcup}(b) \tag{2.2}$$

holds for all $a, b \in T(E)$. The remaining compatibility conditions (vi)–(viii) can also be verified easily.

■

We will see in Chapter 4 that a notion of duality between bialgebras is a key ingredient for one of our main results. Therefore we introduce the concept of dual pairs of vector spaces and bialgebras.

Definition 2.6 Let E and E' be vector spaces and let

$$\langle \cdot, \cdot \rangle : E \times E' \to \mathbf{R}$$

be a bilinear map. We say that $(E, E', \langle \cdot, \cdot \rangle)$ is a *dual pair (of vector spaces)* if

(i) $\langle e, e' \rangle = 0$ for all $e \in E$ implies $e' = 0$ and

(ii) $\langle e, e' \rangle = 0$ for all $e' \in E'$ implies $e = 0$.

The most important examples of dual pairs are given by a vector space E together with its algebraic or topological dual and the natural pairing between them.

Definition 2.7 Let $(G, \star_G, \Delta_G)$ and $(H, \star_H, \Delta_H)$ be two bialgebras. Let $\langle \cdot, \cdot \rangle : G \times H \to \mathbf{R}$ be a bilinear map such that $(G, H, \langle \cdot, \cdot \rangle)$ is a dual pair of vector spaces. Then $(G, H, \langle \cdot, \cdot \rangle)$ is called a *dual pair of bialgebras* if we have

$$\langle g_1 \otimes g_2, \Delta_H(h) \rangle = \langle g_1 \star_G g_2, h \rangle \quad \text{and} \quad \langle \Delta_G(g), h_1 \otimes h_2 \rangle = \langle g, h_1 \star_H h_2 \rangle \tag{2.3}$$

for all $g, g_1, g_2 \in G$ and all $h, h_1, h_2 \in H$.

Remark 2.8 Let $B_G \subset G$ and $B_H \subset H$ be algebra generators of G and H respectively. Then it is sufficient to verify (2.3) for all $g, g_2 \in G$, $g_1 \in B_G$, $h, h_2 \in H$ and $h_1 \in B_H$.

Proposition 2.9 *Let $(G, \star_G, \Delta_G)$ be a bialgebra and let $(H, \star_H, \Delta_H)$ be such that there exists a bilinear map $\langle \cdot, \cdot \rangle$ as in Definition 2.7. Then $(H, \star_H, \Delta_H)$ is also a bialgebra.*

Proof. We have to verifty that $(H, \star_H, \Delta_H)$ satisfies (i)–(viii) of Definition 2.2. We will give the details for associativity (i), coassociativity (iii) and compatibility (v) of Δ_H and $\star_H$.

In order to show that $\star_H$ is associative, let $h_1, h_2, h_3 \in H$. Then we have

$$\begin{aligned}
&\langle g, h_1 \star_H (h_2 \star_H h_3)\rangle = \langle g_{(1)} \otimes g_{(2)}, h_1 \otimes (h_2 \star_H h_3)\rangle \\
=&\langle g_{(1)} \otimes g_{(2)} \otimes g_{(3)}, h_1 \otimes (h_2 \otimes h_3)\rangle = \langle g_{(1)} \otimes g_{(2)} \otimes g_{(3)}, (h_1 \otimes h_2) \otimes h_3\rangle \\
=&\langle g, (h_1 \star_H h_2) \star_H h_3\rangle
\end{aligned}$$

for all $g \in G$. Since $\langle \cdot, \cdot \rangle$ separates H, this shows that $\star_H$ is associative.

Next, let $h \in H$. Then we have

$$\begin{aligned}
&\langle g_1 \otimes g_2 \otimes g_3, h_{(1)} \otimes h_{(2)(1)} \otimes h_{(2)(2)}\rangle = \langle g_1 \otimes (g_2 \star_g g_3), h_{(1)} \otimes h_{(2)}\rangle \\
=&\langle g_1 \star_G g_2 \star_G g_3, h\rangle = \langle (g_1 \star_g g_2) \otimes g_2, h_{(1)} \otimes h_{(2)}\rangle = \langle (g_1 \otimes g_2 \otimes g_3, h_{(1)(1)} \otimes h_{(1)(2)} \otimes h_{(2)}\rangle
\end{aligned}$$

for all $g_1, g_2, g_3 \in G$ and hence the coassociativiy of Δ_H follows.

Finally, for $h_1, h_2 \in H$ we have

$$\begin{aligned}
&\langle g_1 \otimes g_2, \Delta_H(h_1 \star_H h_2)\rangle = \langle g_1 \star_G g_2, h_1 \star_H h_2\rangle = \langle \Delta_G(g_1 \star_G g_2), h_1 \otimes h_2\rangle \\
=&\langle \Delta_G(g_1) \star_G \Delta_G(g_2), h_1 \otimes h_2\rangle \\
=&\langle (g_1)_{(1)} \otimes (g_2)_{(1)} \otimes (g_1)_{(2)} \otimes (g_2)_{(2)}, (h_1)_{(1)} \otimes (h_1)_{(2)} \otimes (h_2)_{(1)} \otimes (h_2)_{(2)}\rangle \\
=&\langle (g_1)_{(1)} \otimes (g_1)_{(2)} \otimes (g_2)_{(1)} \otimes (g_2)_{(2)}, (h_1)_{(1)} \otimes (h_2)_{(1)} \otimes (h_1)_{(2)} \otimes (h_2)_{(2)}\rangle \\
=&\langle g_1 \otimes g_2, \Delta_H(h_1) \star_H \Delta_H(h_2)\rangle
\end{aligned}$$

for all $g_1, g_2 \in G$ which shows that Δ_H and $\star_H$ are compatible. ■

Remark 2.10 Let $(E, E', \langle \cdot, \cdot \rangle)$ be a dual pair of vector spaces. Then for each $k \in \mathbf{N}$ there is a natural pairing between $E^{\otimes k}$ and $E'^{\otimes k}$ given by

$$E^{\otimes k} \times E'^{\otimes k} \to \mathbf{R} : (e_1 \cdots e_k, e'_1 \cdots e'_k) \mapsto \langle e_1, e'_1\rangle \cdots \langle e_k, e'_k\rangle.$$

We abuse notation and denote this pairing also by $\langle\cdot,\cdot\rangle$. Then $(E^{\otimes k}, E'^{\otimes k}, \langle\cdot,\cdot\rangle)$ is also a dual pair [RR80, VII, Lem. 1]. The pairing can be extended further to the whole tensor algebra and we find that $(T(E), T(E'), \langle\cdot,\cdot\rangle)$ is also a dual pair.

For a given dual pair $(E, E', \langle\cdot,\cdot\rangle)$ Remark 2.10 leads to a dual pair between $T(E)$ an $T(E')$. We will now equip $T(E')$ with a product $\sqcup\!\sqcup$ and a coproduct $\Delta_\otimes$ that turns it into a bialgebra in such a way that $(T(E), \otimes, \Delta_{\sqcup\!\sqcup})$ and $(T(E'), \sqcup\!\sqcup, \Delta_\otimes)$ form a dual pair of bialgebras.

For elementary tensors $b_1 = e'_1 \cdots e'_k \in E'^{\otimes k}$ and $b_2 = e'_{k+1} \cdots e'_{k+l} \in E'^{\otimes l}$ we define the *shuffle product* of b_1 and b_2 as

$$b_1 \sqcup\!\sqcup\, b_2 := \sum_{\sigma \in \mathrm{Sh}(k,l)} e'_{\sigma^{-1}(1)} \cdots e'_{\sigma^{-1}(k+l)}.$$

The shuffle product is then extended linearly to all elements of $T(E')$.

Furthermore, for an elementary tensor $b = e'_1 \cdots e'_n \in E'^{\otimes n}$ the coproduct $\Delta_\otimes(b)$ is given by

$$\Delta_\otimes(b) = \sum_{k=0}^{n} (e'_1 \cdots e'_k) \otimes (e'_{k+1} \cdots e'_n)$$

and it is then extended linearly.

Proposition 2.11 *Let $(E, E', \langle\cdot,\cdot\rangle)$ be a dual pair of vector spaces. Then $(T(E'), \sqcup\!\sqcup, \Delta_\otimes)$ is a bialgebra and the natural pairing between*

$$(T(E), \otimes, \Delta_{\sqcup\!\sqcup}) \quad \textit{and} \quad (T(E'), \sqcup\!\sqcup, \Delta_\otimes)$$

yields a dual pair of bialgebras.

Proof. We only have to show that the conditions of Definition 2.7 hold. Then the fact that the tuple $(T(E'), \sqcup\!\sqcup, \Delta_\otimes)$ is a bialgebra follows via Proposition 2.9. In Remark 2.10 we have already mentioned that $T(E)$ and $T(E')$ separate each other.

We will prove the duality between $\Delta_{\sqcup\!\sqcup}$ and $\sqcup\!\sqcup$ in detail. Let $b_1 \in E'^{\otimes k}$, $b_2 \in E'^{\otimes l}$ and $a = e_1 \cdots e_n \in E^{\otimes n}$. If $k + l \neq n$, then we have

$$\langle \Delta_{\sqcup\!\sqcup}(a), b_1 \otimes b_2 \rangle = \langle a, b_1 \sqcup\!\sqcup\, b_2 \rangle = 0.$$

Thus assume $k + l = n$. Then we have

$$\langle \Delta_{\sqcup\!\sqcup}(a), b_1 \otimes b_2 \rangle = \sum_{i=0}^{n} \sum_{\sigma \in \mathrm{Sh}(i,n-i)} \langle e_{\sigma(1)} \cdots e_{\sigma(i)}, b_1 \rangle \langle e_{\sigma(i+1)} \cdots e_{\sigma(n)}, b_2 \rangle$$

$$= \sum_{\sigma \in \mathrm{Sh}(k,l)} \langle e_{\sigma(1)} \cdots e_{\sigma(k)}, b_1 \rangle \langle e_{\sigma(k+1)} \cdots e_{\sigma(n)}, b_2 \rangle = \langle a, b_1 \sqcup\!\sqcup\, b_2 \rangle.$$

The dual relation between $\Delta_\otimes$ and $\otimes$ is shown in the same way. ■

2.1.2. Tensor norms and the topological tensor algebra

So far our considerations about the tensor algebra are purely algebraic. This is insufficient for two reasons. First, as we will see in Chapter 3, rough path theory has both an algebraic and an analytic component. It is therefore necessary that we introduce a way to carry over a given topological structure on E to $T(E)$. Second, we will see that the set $T(E)$ is too small to contain the so-called *signatures* of rough paths. Therefore we have to consider its completion under a suitable topology. The main tool that we will employ in this section are tensor norms as developed by Grothendieck [Gro55]. In addition, [RR80, Bou07] are helpful.

Assume that E is equipped with a locally convex topology which is generated by a family $(\xi_i)_{i\in\mathcal{I}}$ of semi-norms for some index set $\mathcal{I}$. Unfortunately, there is no canonical way to construct semi-norms on $E^{\otimes k}$ out of $(\xi_i)_{i\in\mathcal{I}}$. Instead, we have to make a sensible choice. Two possible choices that are used very frequently are the following.

Definition 2.12 Let $a \in E^{\otimes k}$ and let $\xi_1, \ldots, \xi_k$ be semi-norms on E. The *projective tensor semi-norm* $\xi_1 \otimes_{\mathrm{pr}} \cdots \otimes_{\mathrm{pr}} \xi_k$ is given by

$$(\xi_1 \otimes_{\mathrm{pr}} \cdots \otimes_{\mathrm{pr}} \xi_k)(a) := \inf \left\{ \sum_{i=1}^m \xi_1\left(e_1^{(i)}\right) \cdots \xi_k\left(e_k^{(i)}\right) : a = \sum_{i=1}^m e_1^{(i)} \cdots e_1^{(i)} \right\}$$

and the *injective tensor norm* $\xi_1 \otimes_{\mathrm{in}} \cdots \otimes_{\mathrm{in}} \xi_k$ is given by

$$(\xi_1 \otimes_{\mathrm{in}} \cdots \otimes_{\mathrm{in}} \xi_k)(a) := \sup\{|(\varphi_1 \cdots \varphi_k)(a)| : \varphi_1, \ldots, \varphi_k \in E', \xi_1'(\varphi_1) = \cdots = \xi_k'(\varphi_k) = 1\},$$

where E' is the continuous dual of E.

Both of these ways to extend semi-norms lead to systems of *cross semi-norms*. This means that for $a \in E^{\otimes k}$ and $b \in E^{\otimes l}$ we have

$$(\xi_1 \otimes \cdots \otimes \xi_{k+l})(ab) = (\xi_1 \otimes \cdots \otimes \xi_k)(a) \cdot (\xi_{k+1} \otimes \cdots \otimes \xi_{k+l})(b). \tag{2.4}$$

Among all topological tensor products the projective tensor product is especially useful because it generalises the (algebraic) universal property of tensor products to the topological setting in a natural way.

Proposition 2.13 *[Gro55, Chap. 1, Prop. 1] Let E and F be semi-normed spaces with semi-norms ξ and η respectively. Then $\xi \otimes_{\mathrm{pr}} \eta$ is the unique semi-norm on $E \otimes F$ which turns the algebraic isomorphism between the bounded bilinear maps on $E \times F$ and the bounded linear maps on $E \otimes F$ into an isometry.*

From now on we will assume that the tensor powers $E^{\otimes n}$ for $n \in \mathbf{N}$ are equipped with a system of cross semi-norms. Note that $E^{\otimes n}$ is in general not complete, even if E is. We denote its completion by $E^{\widehat{\otimes} n}$.

We equip $T(E)$ with the product topology and it is easy to see that (2.4) guarantees that the products and coproducts that we have defined above are continuous with respect to this topology.

Let us now consider the completion $\widehat{T}(E)$ of $T(E)$. While it is straightforward to extend the products continuously to $\widehat{T}(E)$, the situation is more subtle for the coproducts. The problem is that the tensor product $\widehat{T}(E) \otimes \widehat{T}(E)$ is not complete. Therefore the coproducts $\Delta_\otimes$ and $\Delta_\sqcup$ take values in $\widehat{T}(E) \widehat{\otimes} \widehat{T}(E)$ rather than $\widehat{T}(E) \otimes \widehat{T}(E)$ when we extend them to $\widehat{T}(E)$. Thus the tuples $(\widehat{T}(E), \otimes, \Delta_\sqcup)$ and $(\widehat{T}(E), \sqcup, \Delta_\otimes)$ are not bialgebras in the strict sense of Definition 2.2. Nevertheless we will refer to them as bialgebras rather than introduce a new term since this difference will be of no concern for us.

Remark 2.14 There are many sensible choices for a topology on $\bigoplus_{k=0}^{\infty} E^{\otimes k}$. For instance, one could consider the much stronger topology that is used in [CL16] or, if E is Banach, considering the l^p direct sum for $p \in [1, \infty]$ might be a useful choice. We have chosen the product topology because it is the weakest sensible choice (in the sense that it renders the canonical projections continuous) and hence it leads to the largest possible closure (which turns out to be the direct product $\prod_{k=0}^{\infty} E^{\otimes k}$). Therefore the following separation results hold automatically for any other reasonable choice of topology.

Definition 2.15 Let $(E, E', \langle \cdot, \cdot \rangle)$ be a dual pair. The *weak topology* $\sigma(E, E')$ on E is defined as the coarsest topology which makes $\langle \cdot, e' \rangle$ continuous for all $e' \in E'$.

Lemma 2.16 *Let $(E, E', \langle \cdot, \cdot \rangle)$ be a dual pair. Then the weak topology $\sigma(E, E')$ on E is the weakest topology on E such that E' is the continuous dual of E.*

Proof. This is one part of the Mackey-Arens Theorem [Bou07, Chap. IV, §1, Thm. 1] ∎

Lemma 2.17 *Let $(E, E', \langle \cdot, \cdot \rangle)$ be a dual pair of vector spaces and let $\widehat{E}$ be the completion of E with respect to the weak topology $\sigma(E, E')$. Then $\langle \cdot, e' \rangle$ can be extended continuously to $\widehat{E}$ for all $e' \in E'$ and $(\widehat{E}, E', \langle \cdot, \cdot \rangle)$ is also a dual pair.*

Proof. We know from Lemma 2.16 that E' is the continuous dual of E under the weak topology. Thus E' is also the continuous dual of the completion $\widehat{E}$ under the weak topology and thus it separates $\widehat{E}$. ∎

Proposition 2.18 *Let E be a locally convex vector space and let E' be its continuous dual. Then $((\widehat{T}(E), \otimes, \Delta_{\sqcup\!\sqcup}), (T(E'), \sqcup\!\sqcup, \Delta_{\otimes}), \langle\cdot,\cdot\rangle)$ is a dual pair of bialgebras, where $\langle\cdot,\cdot\rangle$ is the extension of the canonical pairing between E and E' as in Remark 2.10.*

Proof. In light of Proposition 2.11 we only have to show that $T(E')$ separates $\widehat{T}(E)$. It is clear that any topology on $T(E)$ which is induced by cross semi-norms as defined above is stronger than the weak topology $\sigma(T(E), T(E'))$. Thus the complete tensor algebra $\widehat{T}(E)$ with respect to such a cross semi-norm topology is contained in the completion of $T(E)$ with respect to the weak topology. An application of Lemma 2.17 in combination with Remark 2.10 therefore yields that $T(E')$ separates $\widehat{T}(E)$. ∎

2.1.3. Primitive and group-like elements

Definition 2.19 Let $(H, \star, \Delta)$ be a bialgebra. We say that an element $h \in H$ is

- *primitive* if $\Delta(h) = 1 \otimes h + h \otimes 1$,
- *group-like* if $\Delta(h) = h \otimes h$.

Proposition 2.20 *Let $(H, G, \langle\cdot,\cdot\rangle)$ be a dual pair of bialgebras. An element $h \in H$ is group-like if and only if we have*

$$\langle h, g_1 \star_G \cdots \star_G g_k\rangle = \langle h, g_1\rangle \cdots \langle h, g_k\rangle \tag{2.5}$$

for all $g_1, \ldots, g_k \in G$.

Proof. Assume that h is group-like. Using Sweedler notation and a straightforward induction over $k \in \mathbf{N}$, we obtain

$$\langle h, g_1 \star_G \cdots \star_G g_k\rangle = \langle h_{(1)} \otimes \cdots \otimes h_{(k)}, g_1 \otimes \cdots \otimes g_k\rangle = \langle h \otimes \cdots \otimes h, g_1 \otimes \cdots \otimes g_k\rangle = \langle h, g_1\rangle \cdots \langle h, g_k\rangle$$

for all $g_1, \ldots g_k \in G$.

On the other hand, if we assume that (2.5) holds for all $g_1, \ldots g_k \in G$, then this is in particular true for $k = 2$ and we have

$$\langle h_{(1)} \otimes h_{(2)}, g_1 \otimes g_2\rangle = \langle h, g_1 \star_G g_2\rangle = \langle h, g_1\rangle\langle h, g_2\rangle = \langle h \otimes h, g_1 \otimes g_2\rangle$$

whence the claim follows, since the elements of G separate H. ∎

From the counitality we have $\varepsilon(a) + a = a$ if a is primitive and $\varepsilon(a)a = a$ if a is group-like. Note that for $T(E)$ the counit ε is simply the projection π^0 to $E^{\otimes 0} \simeq \mathbf{R}$. Thus we have $\pi^0(a) = \varepsilon(a) = 0$ if a is primitive and $\pi^0(a) = \varepsilon(a) = 1$ if a is group-like. Therefore, if $a \in \widehat{T}(E)$ is primitive, then the exponential of a, which is given by

$$\exp_{\otimes}(a) := \sum_{k=0}^{\infty} \frac{a^{\otimes k}}{k!}$$

converges because only finitely many terms of this series yield terms of a given degree. For the same reason, we find that if $g \in \widehat{T}(E)$ is group-like, then its logarithm, which is given by

$$\log_{\otimes}(g) := \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(g-1)^{\otimes k}}{k},$$

is well-defined.

Proposition 2.21 *An element $g \in (\widehat{T}(E), \otimes, \Delta_{\sqcup\!\sqcup})$ is group-like if and only if* $\log_{\otimes}(g)$ *is primitive. Equivalently, an element $a \in (\widehat{T}(E), \otimes, \Delta_{\sqcup\!\sqcup})$ is primitive if and only if* $\exp_{\otimes}(a)$ *is group-like.*

Proof. If a is primitive then we have

$$\Delta_{\sqcup\!\sqcup}(\exp_{\otimes}(a)) = \exp_{\otimes}(\Delta_{\sqcup\!\sqcup}(a)) = \exp_{\otimes}(1 \otimes a) \exp_{\otimes}(a \otimes 1) = \exp_{\otimes}(a) \otimes \exp(a)$$

and thus $\exp_{\otimes}(a)$ is group-like. Likewise, if g is group-like we have

$$\Delta_{\sqcup\!\sqcup}(\log_{\otimes}(g)) = \log_{\otimes}(\Delta_{\sqcup\!\sqcup}(g)) = \log_{\otimes}(g \otimes g) = \log_{\otimes}(1 \otimes g) + \log_{\otimes}(g \otimes 1) = 1 \otimes \log_{\otimes}(g) + \log_{\otimes}(g) \otimes 1$$

and thus $\log_{\otimes}(g)$ is primitive. ■

In order to characterise the primitive elements of the tensor algebra, let us introduce the set $\mathcal{L}(E) \subset T(E)$ of *Lie polynomials* over E. For two elements $a, b \in T(E)$ we define $[a, b] := ab - ba$. Then we proceed inductively

$$\begin{aligned} \mathcal{L}^1(E) &:= E \\ \mathcal{L}^{k+1}(E) &:= \left\{ [e, a] : e \in E, a \in \mathcal{L}^k(E) \right\} \end{aligned}$$

and finally we define

$$\mathcal{L}(E) := \bigoplus_{k=1}^{\infty} \mathcal{L}^k(E).$$

Denote the completion of $\mathcal{L}(E)$ under the trace topology by $\widehat{\mathcal{L}}(E)$.

Proposition 2.22 *The primitive elements of $(\widehat{T}(E), \otimes, \Delta_{\sqcup\!\sqcup})$ are given by $\widehat{\mathcal{L}}(E)$.*

Proof. This follows from [Reu93, Theorem 1.4]. ■

2.1.4. The truncated tensor algebra

Definition 2.23 The *truncated tensor algebra of degree* $n \in \mathbf{N}$ over E is defined as the quotient algebra

$$T^{(n)}(E) := T(E) / \bigoplus_{k=n+1}^{\infty} E^{\otimes k}$$

with respect to the tensor product $\otimes$. Its completion is denoted by $\widehat{T}^{(n)}(E)$.

Note that $T^{(n)}(E)$ can be canonically embedded into $\widehat{T}(E)$ as a subspace but not as a subalgebra. We denote the canonical projection from $\widehat{T}(E)$ to $T^{(n)}(E)$ by $\pi^{(n)}$. Whenever E is finite-dimensional, then so is $T^{(n)}(E)$ and furthermore we have $\widehat{T}^{(n)}(E) = T^{(n)}(E)$ in this case.

An important subspace of $T^{(n)}(E)$ is the space

$$\mathcal{L}^{(n)}(E) := \bigoplus_{k=1}^{n} \mathcal{L}^{k}(E),$$

which is the Lie-subalgebra of $T^{(n)}(E)$ that is generated by E. We will also need the associated Lie-group

$$G^{(n)}(E) := \exp_{\otimes}\left(\mathcal{L}^{(n)}(E)\right) \in T^{(n)}(E).$$

As usual, we denote by $\widehat{\mathcal{L}}^{(n)}(E)$ and $\widehat{G}^{(n)}(E)$ the respective completions.

Proposition 2.24 *An element $g \in \widehat{T}(E)$ is group-like if and only if its projection to $\widehat{T}^{(n)}(E)$ lies in $\widehat{G}^{(n)}(E)$ for all $n \in \mathbf{N}$.*

Proof. This is an immediate consequence of Propositions 2.21 and 2.22. ■

2.2. Smooth manifolds

In this section we will introduce some basic notions from differential geometry such that we can refer to them in Chapter 5. Our main reference for this section is [dC92].

Definition 2.25 Let $\mathcal{M}$ be a connected second countable Hausdorff space. Assume that there exists $n \in \mathbf{N}$ such that for every $y \in \mathcal{M}$ there exists an open neighbourhood U of y and a homeomorphism $\varphi_U : U \to B^n$ such that $\varphi(y) = 0$. Here we have written B^n for the open unit ball around the origin in $\mathbf{R}^n$. Assume furthermore that for any two such neighbourhoods U, V with $U \cap V \neq \emptyset$ the transition map

$$\varphi_V \circ \varphi_U^{-1} : \varphi_U(U \cap V) \to \varphi_V(U \cap V)$$

is a smooth diffeomorphism. Then $\mathcal{M}$ is called a *smooth manifold of dimension n*. Furthermore, the tuples (U, φ_U) are called *coordinate charts* and the family consisting of all coordinate charts is called *atlas*.

For a given smooth manifold $\mathcal{M}$ we will always assume that its atlas $\mathcal{A}$ is maximal. By this we mean that there is no strictly larger atlas $\tilde{\mathcal{A}} \supset \mathcal{A}$ whose transition functions are all smooth.

Let $\mathcal{M}$ be a smooth manifold and fix $y \in \mathcal{M}$ along with a coordinate neighbourhood (U, φ_U) of y. We say that two curves $\gamma^1, \gamma^2 : (-1, 1) \to \mathcal{M}$ with $\gamma_0^1 = \gamma_0^2 = y$ are equivalent, if we have $(\varphi \circ \gamma^1)'(0) = (\varphi \circ \gamma^2)'(0)$.

Definition 2.26 For $y \in \mathcal{M}$ define the *tangent space at y* to be the set

$$T_y\mathcal{M} := \{[\gamma] \mid \gamma : (-1, 1) \to \mathcal{M}, \gamma_0 = y\},$$

where $[\gamma]$ is the equivalence class of γ under the equivalence relation described above. The elements of $T_y\mathcal{M}$ are called *tangent vectors*. The set

$$T\mathcal{M} := \bigcup_{y \in \mathcal{M}} \{y\} \times T_y\mathcal{M}$$

is called the *tangent bundle* of $\mathcal{M}$.

Whenever we use notation for $T\mathcal{M}$ that is usually reserved for vector spaces, such as the dual $T\mathcal{M}^*$ or the tensor power $T\mathcal{M}^{\otimes k}$, these constructions are to be understood fibre-wise.

An alternative way is to view tangent vectors as *derivations* in the following way. A map $f : \mathcal{M} \to \mathbf{R}$ is said to be smooth if $f \circ \varphi_U^{-1}$ is smooth for every coordinate chart U. The set of all smooth maps from $\mathcal{M}$ to $\mathbf{R}$ is denoted by $\mathbf{C}^\infty(\mathcal{M})$. For $f \in C^\infty(\mathcal{M})$ and $y \in \mathcal{M}$ we define the derivative of f in the direction of $v \in T_y\mathcal{M}$ as

$$vf := (f \circ \gamma)'(0) = \lim_{h \to 0} \frac{f(\gamma_h) - f(y)}{h}, \tag{2.6}$$

where γ is an arbitrary representative of v.

Definition 2.27 A *vector field* on $\mathcal{M}$ is a map $V : \mathcal{M} \to T\mathcal{M}$ which assigns to each $y \in \mathcal{M}$ a vector in $T_y\mathcal{M}$ such that the map

$$\mathcal{M} \to \mathbf{R} : y \mapsto V(y)f$$

is smooth for all $f \in C^\infty(\mathcal{M})$. The set of all smooth vector fields on $\mathcal{M}$ is denoted by $\Gamma^\infty(\mathcal{M})$.

One way to view vector fields is to interpret them as linear maps from $C^\infty(\mathcal{M})$ to itself, i.e. we denote by Vf the map that is given by $Vf(y) := V(y)f$. In particular, the vector fields on $\mathcal{M}$ span a subalgebra of $\mathrm{End}(C^\infty(\mathcal{M}))$. Thus we can define the *Lie bracket* of two vector fields by putting

$$[V, W]f := VWf - WVf.$$

It is a direct consequence of Schwarz' theorem that we have $[V, W] \in \Gamma^\infty(\mathcal{M})$ for all vector fields $V, W \in \Gamma^\infty(\mathcal{M})$.

Remark 2.28 A given collection $V = (V_1, \ldots, V_d)$ of smooth vector fields on $\mathcal{M}$, can be interpreted as a linear map

$$V : \mathbf{R}^d \to \Gamma^\infty(\mathcal{M}) : e_i \mapsto V_i$$

for $i \in \{1, \ldots, d\}$,where $e_1, \ldots, e_d \in \mathbf{R}^d$ are the standard basis vectors of $\mathbf{R}^d$. This map can then be extended to an algebra morphism defined on $T(\mathbf{R}^d)$ which assigns to every $a \in T(V)$ an element $V_a \in \mathrm{End}(C^\infty(\mathcal{M}))$. We will frequently use this extension implicitly from now on.

2.2.1. Riemannian metrics

In order to talk about lengths of tangent vectors and angels between them, we require an additional structure on $\mathcal{M}$ which is compatible with the smooth structure of $\mathcal{M}$.

Definition 2.29 A *Riemannian metric* on $\mathcal{M}$ is a map $g : \mathcal{M} \to T\mathcal{M}^* \otimes T\mathcal{M}^*$ which assigns to each $y \in \mathcal{M}$ a scalar product on $T_y\mathcal{M}$ such that the map

$$\mathcal{M} \to \mathbf{R} : y \mapsto g_y(V(y), W(y))$$

is smooth for all $V, W \in \Gamma^\infty(\mathcal{M})$.

We will often write $\langle v, w\rangle_g$ instead of $g_y(v, w)$ for $v, w \in T_y\mathcal{M}$ in order to simplify the notation. Furthermore we define $|v|_g := \sqrt{\langle v, v\rangle_g}$.

As soon as we fix a Riemannian metric on $\mathcal{M}$ this induces two important additional structures. The first one is a distance function on $\mathcal{M}$.

Definition 2.30 Let $\gamma : [0, T] \to \mathcal{M}$ be a smooth curve. Its *length* $L(\gamma)$ is given by

$$L(\gamma) := \int_0^T |\dot{\gamma}_t|_g \mathrm{d}t.$$

Definition 2.31 Let $y, z \in \mathcal{M}$ be two points of $\mathcal{M}$. The *Riemannian distance* $d_g(y, z)$ between y and z is given by

$$d_g(y, z) := \inf\{L(\gamma) \mid \gamma : [0, 1] \to \mathcal{M}, \gamma_0 = y, \gamma_1 = z\},$$

where the infimum ranges over all smooth paths that connect y and z.

The second important structure that comes with a Riemannian metric is an *affine connection.* The connection provides a way to talk about derivatives of vector fields (and higher order tensor fields) on $\mathcal{M}$.

Definition 2.32 An *affine connection* ∇ on $\mathcal{M}$ is a bilinear map

$$\nabla : \Gamma^\infty(\mathcal{M}) \times \Gamma^\infty(\mathcal{M}) \to \Gamma^\infty(\mathcal{M}) : (V, W) \mapsto \nabla_V W$$

such that
(i) $\nabla_{fV} W = f \nabla_V W$ and
(ii) $\nabla_V (fW) = Vf \cdot W + f \nabla_V W$
hold for all $f \in C^\infty(\mathcal{M})$.

Note that $(\nabla_V W)(y)$ depends only on the value of $V(y)$ and not on the values that V takes elsewhere. This allows us to consider the map

$$\nabla W : \mathcal{M} \to T\mathcal{M}^* \otimes T\mathcal{M} : y \mapsto (T_y\mathcal{M} \ni v \mapsto \nabla_v W).$$

We say that ∇W is the *covariant derivative* of W.

Proposition 2.33 (Levi-Civita Theorem [dC92, Chap. 2, Thm. 3.6]) *Given a Riemannian metric g there is a unique affine connection ∇ on $\mathcal{M}$ which satisfies*
(i) $X\langle V, W\rangle_g = \langle \nabla_X V, W\rangle_g + \langle V, \nabla_X W\rangle_g$ *and*
(ii) $\nabla_V W - \nabla_W V = [V, W]$
for all $X, V, W \in \Gamma^\infty(\mathcal{M})$. *The connection* ∇ *is called the* Levi-Civita connection.

Proof. Let us prove the uniqueness of ∇. From (i) and (ii) we obtain

$$X\langle V, W\rangle_g + V\langle W, X\rangle_g - W\langle X, V\rangle_g$$

$$
\begin{aligned}
&=\langle \nabla_X V, W\rangle_g + \langle V, \nabla_X W\rangle_g + \langle \nabla_V W, X\rangle_g + \langle W, \nabla_V X\rangle_g - \langle \nabla_W X, V\rangle_g + \langle X, \nabla_W V\rangle_g \\
&=\langle \nabla_X V + \nabla_V X, W\rangle_g + \langle \nabla_V W - \nabla_W V, X\rangle_g + \langle \nabla_X W - \nabla_W X, V\rangle_g \\
&=2\langle \nabla_X V, W\rangle_g + \langle [V,X], W\rangle_g + \langle [V,W], X\rangle_g + \langle [X,W], V\rangle_g
\end{aligned}
$$

which implies

$$
\begin{aligned}
&\langle \nabla_X V, W\rangle_g \\
=&\frac{1}{2}\left(X\langle V, W\rangle_g + V\langle W, X\rangle_g - W\langle X, V\rangle_g - \langle [V,X], W\rangle_g - \langle [V,W], X\rangle_g - \langle [X,W], V\rangle_g\right).
\end{aligned}
\tag{2.7}
$$

Hence the value of $\nabla_X V$ is fully determined by the right-hand side of (2.7), hence there can be at most one ∇ satisfying (i) and (ii).

In order to prove existence of ∇ it suffices to check that the vector field $\nabla_X V$ which is define by (2.7) satisfies Definition 2.32. See [dC92, Chap. 2, Thm. 3.6] for the details. ∎

Therefore, if we have a Riemannian metric g we will always denote by ∇ (or ∇^g if we want to distinguish between several metrics) the associated Levi-Civita connection.

Let $\gamma : [0,1] \to \mathcal{M}$ be a shortest path between $y \in \mathcal{M}$ and $z \in \mathcal{M}$, i.e. assume $\gamma_0 = y$, $\gamma_1 = z$ and $L(\gamma) = d_g(y,z)$. One can verify [dC92] that in this case γ satisfies

$$
\nabla_{\dot\gamma_t}\dot\gamma_t = 0. \tag{2.8}
$$

Technically, we have to extend the tangent vector field $\dot\gamma$ to a vector field in a neighbourhood of γ in order to compute its covariant derivative. However, it turns out that $\nabla \dot\gamma_t$ does not depend on the choice of this extension. We call any curve $\gamma : (a,b) \to \mathcal{M}$ with $a, b \in \mathbf{R} \cup \{\pm\infty\}$ which satisfies (2.8) a *geodesic*.

Since (2.8) is a second-order equation it has a unique local solution if we fix the initial value and velocity. Thus for $y \in \mathcal{M}$ we can define the *exponential map*

$$
\exp_y : T_y\mathcal{M} \supset U \to \mathcal{M} : v \mapsto \gamma_1
$$

where γ solves (2.8) with $\gamma_0 = y$ and $\dot\gamma_0 = v$. A priori, the exponential map is only defined on some subset U of the tangent space since the existence of a global solution for (2.8) cannot be guaranteed. However, we are only interested in metrics for which the exponential map is globally defined for all $y \in \mathcal{M}$. The metrics with this property are called *complete* and they are characterised by the following famous result.

Proposition 2.34 (Hopf-Rinow Theorem [dC92, Chap. 7, Thm. 2.8]) *Let g be a Riemannian metric. Then the following are equivalent.*

(i) $(\mathcal{M}, d_g)$ *is a complete metric space.*

(ii) The bounded sets of $(\mathcal{M}, d_g)$ *are relatively compact.*

(iii) For each $y \in \mathcal{M}$ *the exponential map* $\exp_y$ *is defined on the whole tangent space* $T_y\mathcal{M}$.

Furthermore, if (i)–(iii) hold, then there exists a length-minimising curve between any two points $y, z \in \mathcal{M}$.

In order to get an intuition for what it means for a Riemannian metric to be complete, we consider the simplest non-trivial manifold, namely $\mathbf{R}$.

Example 2.35 As is well known, $\mathbf{R}$ together with its canonical distance function forms a complete metric space. Thus the metric given by $g_y^{(1)}(v, w) := vw$ for $y, v, w \in \mathbf{R}$ is complete.

Next, we consider the metric given by $g_y^{(2)}(v, w) := \frac{vw}{1+y^2}$. The associated distance function can be computed explicitly as

$$d_{g^{(2)}}(x, y) = \int_x^y \frac{\mathrm{d}t}{\sqrt{1+t^2}} = \sinh^{-1}(y) - \sinh^{-1}(x)$$

for $x, y \in \mathbf{R}$. Since $\sinh^{-1}$ is a proper function it is clear that bounded sets are relatively compact. Thus $g^{(2)}$ is also complete.

Finally, let $g^{(3)}$ be the metric which is given by $g_y^{(3)}(v, w) := \frac{vw}{(1+y^2)^2}$. This leads to the distance function

$$d_{g^{(3)}}(x, y) = \int_x^y \frac{\mathrm{d}t}{1+t^2} = \tan^{-1}(y) - \tan^{-1}(x)$$

for $x, y \in \mathbf{R}$. Thus we have $d_{g^{(3)}}(-\infty, \infty) = \pi < \infty$ and therefore $g^{(3)}$ is not complete.

Example 2.36 Another example for a non-complete Riemannian manifold is $\mathbf{R}^2 \backslash \{0\}$ with the induced canonical metric from $\mathbf{R}^2$. One can easily check that there is no length-minimising curve between, say, $(-1, 0)$ and $(1, 0)$.

We conclude this section with a discussion of the differentiability of the Riemannian distance d_g. For a fixed point $z \in \mathcal{M}$ we define the map

$$f_z : \mathcal{M} \to \mathbf{R} : y \mapsto d_g(y, z).$$

This map is proper (i.e. pre-images of compact sets are compact) if and only if g is complete. Furthermore, it is never smooth since one can check that it is not differentiable at $y = z$.

Nevertheless, for every $y \in \mathcal{M}$ and every tangent vector $v \in T_y\mathcal{M}$ we can define the (one-sided) directional derivative

$$vf_z := \lim_{h \downarrow 0} \frac{f_z(\gamma_h) - f_z(y)}{h} \tag{2.9}$$

for a representative γ of v, cf. (2.6). The existence of those limits is not obvious and we give a proof in Appendix B, where we also show that the derivative satisfies

$$|vf_z| \leq |v|_g.$$

Also, the following result allows us to replace f_z with a smooth function in some situations.

Lemma 2.37 *For every $\varepsilon > 0$ there exists a map $f_z^\varepsilon \in C^\infty(\mathcal{M})$ such tat*

(i) $|f_z^\varepsilon(y) - f_z(y)| \leq \varepsilon$ *and*

(ii) $|vf_z^\varepsilon| \leq (1+\varepsilon)|v|_g$

for all $y \in \mathcal{M}$ and all $v \in T\mathcal{M}$.

Proof. This is a simple application of [AFLMR07, Thm. 1]. ∎

2.2.2. Ordinary differential equations

Let $\mathcal{M}$ be a smooth manifold and let $V \in \Gamma^\infty(\mathcal{M})$ be a vector field and consider the equation

$$\dot{y}_t = V(y_t). \tag{ODE}$$

It is well-known that if $y_0 \in \mathcal{M}$ is fixed and $\varepsilon > 0$ is sufficiently small then there exists a unique $y : (-\varepsilon, \varepsilon) \to \mathcal{M}$ which solves (ODE) with initial value y_0. Furthermore, one can either extend y to the whole real line or one can find $a, b \in \mathbf{R}$ such that y can be extended to the interval (a, b) and such that the image of (a, b) under y is not relatively compact. In this section we would like to focus on the question, under which conditions on V the solution y can be extended to the whole real line. The main motivation for this section is to provide a benchmark for Chapter 5 in which we will discuss the analogous question for rough differential equations. On the other hand, Chapter 5 will also partly rely on these ODE results.

Definition 2.38 A vector field $V \in \Gamma^\infty(\mathcal{M})$ is called *complete* if (ODE) has a global solution for all initial values $y_0 \in \mathcal{M}$.

It is not a coincidence that the term *complete* is used for both vector fields and Riemannian metrics, since the two notions of completeness are closely related as we will see below. Before we state the main result in this direction, recall that a function between topological spaces is called *proper* if its preimages of compact sets are compact.

Proposition 2.39 ([Gli97, Thm. 1.1]) *Let $V \in \Gamma^\infty(\mathcal{M})$ and denote by $V^+ \in \Gamma^\infty(\mathcal{M} \times \mathbf{R})$ the vector field which is given by $V^+(y,t) = (V(y), 1)$. Then the following are equivalent.*

(i) V is complete.

(ii) There exists a proper function $f \in C^\infty(\mathcal{M} \times \mathbf{R})$ such that $V^+ f$ is bounded.

(iii) There exists a complete Riemannian metric g on $\mathcal{M} \times \mathbf{R}$ such that $|V^+|_g$ is bounded.

Proof. Let us first show that (iii) implies (ii). Let thus g be a complete metric such that $|V^+|_g$ is bounded and fix an arbitrary point $z \in \mathcal{M}$. We claim that the map f_z^ε from Lemma 2.37 satisfies (ii). On the one hand, it is clearly proper since g is complete. On the other hand, we have

$$|V^+ f_z^\varepsilon| \leq (1+\varepsilon)|V^+|_g$$

and thus (ii) holds.

Now assume that (ii) holds for some $f \in C^\infty(\mathcal{M})$ and that V^+ is not complete. In this case we can find $a, b \in \mathbf{R}$ such that $y : (a,b) \longrightarrow \mathcal{M}$ satisfies (ODE) and $y((a,b))$ is not relatively compact. On the other hand, we have

$$|f(y_b) - f(y_a)| = \left| \int_a^b V^+ f(y_s) \mathrm{d}s \right| \leq (b-a)\|V^+ f\|_\infty < \infty.$$

Since f is assumed to be proper, this is a contradiction and hence V^+, and therefore V, must be complete.

We postpone the proof that (i) implies (iii) until later, where we show a stronger and more general result in Proposition 5.4 ∎

If we are only interested in one direction of Propostion 2.39 we can ignore the extension to $\mathcal{M} \times \mathbf{R}$ and obtain the following corollary.

Corollary 2.40 *A vector field $V \in \Gamma^\infty(\mathcal{M})$ is complete if either*

(i) there exists a proper function $f \in C^\infty(\mathcal{M})$ such that Vf is bounded or

(ii) there exists a complete Riemannian metric g on $\mathcal{M}$ such that $|V|_g$ is bounded.

Arguably the most relevant manifold is $\mathbf{R}^n$. In this case one can derive a criterion from Corollary 2.40 which is formulated in terms of the growth of the vector field.

Corollary 2.41 *Let $V \in \Gamma^\infty(\mathbf{R}^n)$ be a vector field and assume that there exists a non-decreasing smooth function $h : \mathbf{R}_{\geq 0} \longrightarrow \mathbf{R}_{\geq 1}$ such that*

$$|V(y)| \leq h(|y|) \quad \textit{and} \quad \int_0^\infty \frac{\mathrm{d}t}{h(t)} = \infty$$

holds for all $y \in \mathbf{R}^n$. Then V is complete.

Proof. Check that

$$f : \mathcal{M} \longrightarrow \mathbf{R} : y \mapsto \int_0^{\sqrt{|y|^2+1}} \frac{\mathrm{d}t}{h(t)}$$

satisfies (i) of Corollary 2.40. ∎

However, one should note that Corollary 2.41 is really less general than Corollary 2.40 as the following example illustrates.

Example 2.42 Consider the vector field $V : \mathbf{R}^2 \longrightarrow \mathbf{R}^2 : (x, y) \mapsto (y^2, 0)$. The condition of Corollary 2.41 is clearly not satisfied. Nevertheless we can define

$$f : \mathbf{R}^2 \longrightarrow \mathbf{R} : (x, y) \mapsto \frac{\sqrt{x^2+1}}{y^2+1} + y^2,$$

which satisfies (i) of Corollary 2.40. Alternatively we can consider the Riemannian metric on $\mathbf{R}^2$ which is given by

$$g : \mathbf{R}^2 \longrightarrow \mathbf{R}^{2\times 2} : (x, y) \mapsto \begin{pmatrix} \frac{1}{1+y^4} & 0 \\ 0 & 1 \end{pmatrix}$$

and which satisfies (ii) of Corollary 2.40.

Finally, in view of Chapter 5 it is an important observation that the completeness of vector fields is not preserved by taking sums or Lie brackets. The following examples show what can go wrong.

Example 2.43 Let $V_1, V_2 \in \Gamma^\infty(\mathbf{R})$ be the vector fields in $\mathbf{R}$ which are given by

$$V_1(y) = e^y \sin(y)^2 \quad \text{and} \quad V_2(y) = e^y \cos(y)^2$$

respectively. V_1 is complete since any solution y with $y_0 \in [k\pi, (k+1)\pi]$ for $k \in \mathbf{N}$ stays in this interval for all times. An analogous argument yields completeness of V_2. However, we have

$$(V_1 + V_2)(y) = e^y,$$

and thus $V_1 + V_2$ is not complete.

Example 2.44 Let $V_1, V_2 \in \Gamma^\infty(\mathbf{R})$ be the vector fields in $\mathbf{R}$ which are given by $V_1(y) = \cos(e^y)$ and $V_2(y) = \sin(e^y)$ respectively. Both vector fields are bounded and therefore they are complete. However, the bracket $[V_1, V_2]$ is not complete since we have

$$[V_1, V_2](y) = e^y.$$

3. Rough paths and rough differential equations

A wide range of works has been written about rough paths and rough differential equation and many different points of view have been developed since the publication of the seminal work [Lyo98]. In this chapter we would like to summarise the various approaches and see how they compare. We will in particular focus on the question to which extent each approach generalises to a manifold setting. This will enable us to understand how the new contributions in Chapters 4 and 5 fit into to the picture.

3.1. Rough paths

Throughout this section E is a Banach space with norm $\|\cdot\|$. Furthermore, assume that $T(E)$ is equipped with a system of cross-norms and denote the corresponding norm on $E^{\otimes k}$ by $\|\cdot\|^{\otimes k}$. Let furthermore $T > 0$ be our (finite) time horizon and assume $p \geq 1$. By $[p]$ we denote the largest integer which is less than or equal to p. Furthermore we introduce the set

$$\Delta_T := \{(s,t) \in [0,T]^2 : s \leq t\}.$$

Definition 3.1 A continuous map $\omega : \Delta_T \to \mathbf{R}_{\geq 0}$ is called a *control function* if it satisfies

$$\omega(s,t) + \omega(t,u) \leq \omega(s,u)$$

for all $0 \leq s \leq t \leq u \leq T$.

Definition 3.2 Let ω be a control function. A continuous map $\mathbf{x} : \Delta_T \to T^{([p])}(E)$ is called a *p-rough path controlled by ω* if it satisfies

(i) $\mathbf{x}_{s,t}\mathbf{x}_{t,u} = \mathbf{x}_{s,u}$ for all $0 \leq s \leq t \leq u \leq T$ and

(ii) $\|\pi^k(\mathbf{x}_{s,t})\|^{\otimes k} \leq C\omega(s,t)^{k/p}$ for some $C > 0$ and all $(s,t) \in \Delta_T$ and all $k \in \{1, \ldots, [p]\}$.

Property (i) in the previous definition is called *multiplicativity* and if (ii) holds we say that $\mathbf{x}$ has *finite p-variation controlled by ω*. It is well known that the concept of finite p-variation for continuous paths in the classical sense is closely related to Hölder continuity. This remains true in the context of rough paths, i.e. it is possible to formulate (ii) in terms of Hölder continuity. This is equivalent to limiting the choice of ω to the map $(s,t) \mapsto |t-s|$. Therefore it leads to a smaller class of rough paths and causes the set of rough paths to be no longer invariant under reparametrisations. Thus working with Hölder norms becomes more natural in situations where the parametrisation of given functions matters, e.g. when working with stochastic partial differential equations as in [FH14].

An important qualitative takeaway from Definition 3.2 is that the less regular $\mathbf{x}$ is the more terms we need in addition to the underlying path increments $(s,t) \mapsto \pi^1(\mathbf{x}_{s,t})$. It is natural to expect that a p-rough path can be canonically treated as a q-rough path if $p \leq q$. This is indeed true and the following fundamental result guarantees the existence and uniqueness of the missing higher order terms.

Proposition 3.3 ([LCL07, Thm. 3.7]) *Let $\mathbf{x}$ be a p-rough path controlled by some control ω. Then there exists a unique map*

$$S(\mathbf{x}) : \Delta_T \to \widehat{T}(E)$$

such that

(i) $S(\mathbf{x})_{s,t} S(\mathbf{x})_{t,u} = S(\mathbf{x})_{s,u}$ *for all* $0 \leq s \leq t \leq u \leq T$,

(ii) $\|\pi^k(S(\mathbf{x})_{s,t})\|^{\otimes k} \leq C\omega(s,t)^{k/p}$ *for some* $C > 0$ *and all* $(s,t) \in \Delta_T$ *and all* $k \in \mathbf{N}$ *and*

(iii) the projection of $S(\mathbf{x})_{s,t}$ *to* $T^{(\lfloor p \rfloor)}(E)$ *coincides with* $\mathbf{x}_{s,t}$ *for all* $(s,t) \in \Delta_T$.

The tensor series $S(\mathbf{x})_{0,T}$ is called the *signature* of $\mathbf{x}$ and it has a number of remarkable features. Most importantly, is been shown in [HL10] (for the case $p = 1$) and in [BGLY16] (for arbitrary $p \geq 1$) that $S(\mathbf{x})_{0,T} \in \widehat{T}(E)$ fully determines $\mathbf{x}_{s,t}$ for all values $(s,t) \in \Delta_T$ up to so-called tree-like paths. Furthermore, if $\mathbf{x}$ is a random rough path which satisfies certain moment estimates, then the expectation of $S(\mathbf{x})_{0,T}$ fully determines its distribution, at least if $\mathbf{x}$ is a geometric rough path [CL16] (see below for a definition of geometric rough paths).

Thus Proposition 3.3 is even useful in the case $p = 1$, i.e. when the "rough" path in question happens to be an ordinary continuous path of bounded variation. It turns out that the first few terms of the signature $S(\mathbf{x})_{0,T}$ already contain a significant amount of information about $\mathbf{x}$ and this idea can be used to develop numerical schemes. See [Lyo14] for a survey of possible applications.

3.1.1. Geometric, branched and general rough paths

Unfortunately, Definition 3.2 is too general to lead to a sensible theory of rough differential equations. The original solution to this problem from [Lyo98] is to use the following subset of rough paths.

Definition 3.4 A rough path $\mathbf{x} : \Delta_T \to T^{([p])}(E)$ is called *geometric rough path* if it satisfies

$$\mathbf{x}_{s,t} \in G^{[p]}(E)$$

for all $(s,t) \in \Delta_T$.

Remark 3.5 In some of the literature the term *weakly geometric rough path* is used and *geometric rough path* is reserved for a subset of what we have defined in Definition 3.4, see [FV06] for details. We will however not need this distinction and the term *geometric rough path* is always to be understood in the sense of Definition 3.4.

In the next section below we will look at examples of rough paths and we will see that in the context of stochastic analysis geometric rough paths can be thought of as a generalisation of Stratonovich integration. In Chapters 4 and 5 we will show how geometric rough paths lead to a well defined notion of rough differential equations on manifolds that does not require any additional structure (in particular we do not have to choose a connection), which supports the analogy with Stratonovich integration.

By contrast, Itō equations only make sense on manifolds if we interpret them as Itō-corrected Stratonovich equations, which only works if the manifold is equipped with a connection. This suggests that there is a notion of rough paths that is less general than Definition 3.2 but more general than Definition 3.4. Those rough paths have been introduced in [Gub10] for finite-dimensional E and they are called *branched rough paths*. Instead of the truncated tensor algebra they take their values in the truncated tree algebra over E. We postpone the details to Chapter 4 where we develop a new framework that makes it possible to deal with branched rough paths when E is infinite dimensional.

3.1.2. Deterministic and stochastic examples

Example 3.6 Let $x : [0,T] \to E$ be a smooth path. Then $\mathbf{x} : \Delta_T \to T^{(1)}(E) : (s,t) \mapsto (1, x_t - x_s)$ is a geometric 1-rough path controlled by $\omega : \Delta_T \to \mathbf{R}_{\geq 0} : (s,t) \mapsto |t-s|$. Indeed, (i) of Definition 3.2 holds because we have

$$\mathbf{x}_{s,t}\mathbf{x}_{t,u} = (1, \pi^1(\mathbf{x}_{s,t}) + \pi^1(\mathbf{x}_{t,u})) = (1, x_t - x_s + x_u - x_t) = (1, x_u - x_s) = \mathbf{x}_{s,u}$$

Furthermore (ii) of Definition 3.2 certainly holds because a smooth path is Lipschitz continuous on a compact interval. Finally, Definition 3.4 is satisfied because we have $(1, v) \in G^{(1)}(E)$ for all $v \in E$.

Example 3.7 The map

$$\mathbf{x} : \Delta_T \to T^{(2)}(\mathbf{R}^2) : (s,t) \to \left(1, 0, (t-s)\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\right)$$

is a geometric 2-rough path controlled by $\omega : \Delta_T \to \mathbf{R}_{\geq 0} : (s,t) \mapsto |t-s|$. Indeed, we have

$$\begin{aligned}\mathbf{x}_{s,t}\mathbf{x}_{t,u} &= (1, \pi^1(\mathbf{x}_{s,t}) + \pi^1(\mathbf{x}_{t,u}), \pi^2(\mathbf{x}_{s,t}) + \pi^1(\mathbf{x}_{s,t})\pi^1(\mathbf{x}_{t,u}) + \pi^2(\mathbf{x}_{t,u})) \\ &= \left(1, 0, (u-s)\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\right) = \mathbf{x}_{s,u}\end{aligned}$$

and (ii) of Definition 3.2 holds as well. Furthermore we have

$$\log_\otimes(\mathbf{x}_{s,t}) = \left(0, 0, (t-s)\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\right) = \left(0, 0, (t-s)\left[\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right]\right) \in \mathcal{L}^{(2)}(\mathbf{R}^2)$$

and thus $\mathbf{x}$ is geometric. The rough path $\mathbf{x}$ is called *pure area rough path.*

Example 3.8 Let X be Brownian motion in $\mathbf{R}^d$ and define

$$\mathbf{X} : \Delta_T \to T^{(2)}(\mathbf{R}^d) : (s,t) \mapsto \left(1, X_t - X_s, \int_s^t (X_r - X_s) \circ \mathrm{d}X_r\right),$$

where $\int \cdot \circ \mathrm{d}X_r$ denotes the Stratonovich integral. Then we have

$$\begin{aligned}&\mathbf{X}_{s,t}\mathbf{X}_{t,u} \\ &= \left(1, X_t - X_s + X_u - X_t, \int_s^t (X_r - X_s) \circ \mathrm{d}X_r + (X_t - X_s)(X_u - X_t) + \int_t^u (X_r - X_t) \circ \mathrm{d}X_r\right) \\ &= \left(1, X_u - X_s, \int_s^t (X_r - X_s) \circ \mathrm{d}X_r + (X_t - X_s)(X_u - X_t) - X_t(X_u - X_t) + \int_t^u X_r \circ \mathrm{d}X_r\right) \\ &= \left(1, X_u - X_s, \int_s^t (X_r - X_s) \circ \mathrm{d}X_r - X_s(X_u - X_t) + \int_t^u X_r \circ \mathrm{d}X_r\right) \\ &= \left(1, X_u - X_s, \int_s^t (X_r - X_s) \circ \mathrm{d}X_r + \int_t^u (X_r - X_s) \circ \mathrm{d}X_r\right) = \mathbf{X}_{s,u}.\end{aligned}$$

Furthermore, it is well known that Brownian motion is α-Hölder continuous for all $\alpha < 1/2$.

Therefore we have

$$|X_t - X_s| \le C|t-s|^{1/p}$$

for all $(s,t) \in \Delta_T$ and all $p > 2$ and some $C > 0$. Showing the corresponding Hölder estimate for the second-order term of $\mathbf{X}$ is slightly more difficult and we refer to [FH14, Prop. 3.4]. More importantly, we have

$$\begin{aligned}\log_\otimes(\mathbf{X}_{s,t}) &= \left(0, X_t - X_s, \int_s^t (X_r - X_s) \circ \mathrm{d}X_r - \frac{1}{2}(X_t - X_s)(X_t - X_s)\right) \\ &= \left(0, X_t - X_s, \int_s^t X_r \circ \mathrm{d}X_r - \frac{1}{2}(X_t X_t - X_s X_s) + \frac{1}{2}(X_s X_t - X_t X_s)\right).\end{aligned}$$

Since the Stratonovich integral satisfies the usual integration-by-parts rule, the symmetric part of the matrix $\int_s^t X_r \circ \mathrm{d}X_r$ is given by $\frac{1}{2}(X_t X_t - X_s X_s)$. Thus we see that $\log_\otimes(\mathbf{X}_{s,t})$ is indeed a Lie polynomial and therefore $\mathbf{X}$ is a geometric p-rough path for any $p \in (2,3)$.

Example 3.9 Let X be Brownian motion in $\mathbf{R}^d$ and define

$$\mathbf{X} : \Delta_T \to T^{(2)}(\mathbf{R}^d) : (s,t) \mapsto \left(1, X_t - X_s, \int_s^t (X_r - X_s)\mathrm{d}X_r\right).$$

The only difference in comparison with the previous example is that we use the Itō integral instead of the Stratonovich integral. It is clear that $\mathbf{X}$ is still a p-rough path for every $p \in (2,3)$. However, the symmetric part of $\log_\otimes(\mathbf{X}_{s,t})$ no longer vanishes and is instead given by

$$-\frac{t-s}{2}\sum_{i=1}^d e_i e_i.$$

This is a consequence of Itō's formula. Hence $\mathbf{X}$ is not geometric.

3.2. Rough differential equations

Throughout this section E and F are Banach spaces, $\mathbf{x}$ is a geometric rough path over E and $V : E \times F \to F$ is a continuous map which is linear in the first argument and smooth (in the Fréchet sense) in the second argument. One can view such a map either as a linear map from E into the space $\Gamma^\infty(F)$ of smooth vector fields in F or as a smooth map from F into the space $\mathrm{Hom}(E,F)$ of linear maps from E to F. When we want to emphasize the former point of view we write $V_e(y)$ instead of $V(e,y)$ and when we want to emphasize the latter, then we write $V(y)e$. Note that the definition of a vector field in an infinite dimensional space is more involved than in the finite-dimensional case. In the language of [KM97] the vector fields V_e for $e \in E$ are *kinetic vector fields.*

The objective of this section is to give meaning to the equation

$$\mathrm{d}y_t = V(y_t)\mathrm{d}\mathbf{x}_t, \quad y_0 = \xi \in F. \tag{RDE}$$

for a geometric rough path $\mathbf{x}$ and to discuss existence of local and global solutions.

In addition to Lyons' original work [Lyo98], various different ways to treat rough differential equations have been introduced. They can be split into methods that define solutions as fixed points of Picard iterations [Lyo98, LCL07, Gub04, FH14] and methods that define solutions as the limit of solutions of certain ordinary differential equations [FV10, Bai15] or discrete approximations [Dav08].

We will discuss the various approaches to handle RDEs informally and we will not give precise definitions. Instead we focus on the key differences and on how well each approach generalises to a non-linear setting.

3.2.1. Solutions as fixed points of Picard iterations

The basic idea of [Lyo98, LQ02, LCL07] is to transform (RDE) into an integral equation of the form

$$\mathbf{z} = \int h(\mathbf{z})\mathrm{d}\mathbf{z}, \tag{3.1}$$

where $\mathbf{z}$ is a rough path over the space $E \oplus F$. The map h is given by

$$h(x_1, y_1)(x_2, y_2) = (x_2, V(y_1 + \xi)x_2).$$

One can then use the theory of rough integration to give meaning to the expression

$$\int h(\mathbf{z})\mathrm{d}\mathbf{z}$$

for a rough path $\mathbf{z}$ over $E \oplus F$. A solution of (RDE) is then defined to be a rough path $\mathbf{z}$ which satisfies (3.1) and whose projection to a rough path over E coincides with $\mathbf{x}$. In particular, the solution is itself a rough path. Nevertheless one can consider the projection of $\mathbf{z}$ to F in order to obtain an ordinary solution path in F. The existence of solutions is shown using a suitably adapted form of Picard iterations.

Remark 3.10 The above approach makes heavy use of the fact that the solution of (RDE) takes its values in the linear space F. Therefore, it is difficult to generalise these ideas to situations where F is replace by a general manifold.

In the more recent work [Gub04] the concept of *controlled rough paths* is introduced. We follow [FH14] and sketch this concept in the case $p \in [2, 3)$. A crucial property of controlled rough paths is that they are defined with respect to a reference rough path. In the context of rough differential equations this reference path is given by the driving noise $\mathbf{x}$. Let us thus assume that $\mathbf{x}$ is a p-rough path for $p \in [2, 3)$. Then a $1/p$-Hölder continuous path $y : [0, T] \to F$ is said to be *controlled by* $\mathbf{x}$ if there exists a $1/p$-Hölder continuous path $y' : [0, T] \to \mathrm{Hom}(E, F)$, a remainder $R^y : \Delta_T \to F$ and $C > 0$ such that

$$y_t - y_s = y'_s \pi^1(\mathbf{x}_{s,t}) + R^y_{s,t} \quad \text{and} \quad \|R^y_{s,t}\| \leq C|t-s|^{2/p} \tag{3.2}$$

hold for all $(s, t) \in \Delta_T$. One can interpret (3.2) as a generalisation of Taylor's theorem where the path y is not approximated by polynomials but by the homogeneous terms of the rough path $\mathbf{x}$. One can then define integrals of controlled rough paths (and functions thereof) with respect to the reference path $\mathbf{x}$. Thus one can rewrite (RDE) as the integral equation

$$y_t = \xi + \int_0^t V(y_s) \mathrm{d}\mathbf{x}_s \tag{3.3}$$

and treat this equation as a fixed point equation in the space of controlled rough paths and apply a Picard iteration to obtain a solution. It turns out [FH14, Thm. 8.4] that if y solves (3.3), then y' is given by $y'_s = V(y_s)$. The concept of controlled rough paths has been developed further to *regularity structures* [Hai14], which is a theory that applies to a large class of singular partial differential equations.

Remark 3.11 Recently, a theory of controlled rough paths on manifolds has been developed in [DS17]. However, it requires additional structure on the manifold. Therefore we have chosen to use a different approach in Chapter 5, where we introduce rough differential equations on manifolds, in order to keep the definition as elementary as possible.

3.2.2. Solutions as limits of approximations

Before we discuss the content of [FV10], we would like to note that the authors work with the standing assumption that E is finite dimensional and some of their arguments are not valid for general Banach spaces. The main idea of [FV10] is to approximate $\mathbf{x}$ by rough paths lifts of smooth paths (in the sense of Proposition 3.3). When $\mathbf{x}$ is given by the lift of a path

$\tilde{x} : [0, T] \to E$ of bounded variation, then (RDE) can be defined to be equivalent to the ODE

$$\dot{y}_t = V(y_t)\dot{\tilde{x}}_t.$$

Hence one defines the solution of the original RDE as the uniform limit of the ODE solutions as the smooth approximation $S^{[p]}(\tilde{x})$ converge to $\mathbf{x}$ in a suitable sense. In particular, the solution of (RDE) is defined to be an ordinary continuous path in F without any terms of higher order. It is shown in [FV10] that a solution $y : [0, T] \to F$ of (RDE) satisfies (recall the notation from Remark 2.28)

$$\|y_t - V_{\mathbf{x}_{s,t}} \operatorname{id}_F(y_s)\| \leq C\omega(s, t)^{([p]+1)/p} \tag{3.4}$$

for some $C > 0$ and all $(s, t) \in \Delta_T$.

The inequality (3.4) originally appeared in [Dav08] where it serves as the defining property of y. Davie then treats the expression $V_{\mathbf{x}_{s,t}} \operatorname{id}_F(y_s)$ as a generalised Euler approximation for y_t and shows that by iterating it in the usual way the discrete approximations converge to a path y which satisfies (3.4).

Remark 3.12 Below we will use (3.4) as the defining property for solutions of (RDE). In Chapter 5 we will use a slightly modified version which works on general manifolds. The main reason for our choice of definition is that the inequality can be formulated without introducing any additional structure (like a metric and/or a connection) on the manifold. Furthermore it defines the solution to be an ordinary, manifold-valued continuous path, which is arguably the most natural object to expect in this context.

Let us finally discuss the approach of [Bai15], where the central objects of interest are flows of diffeomorphisms. Thus instead of solving (RDE) for one fixed initial value $\xi \in F$ one obtains a smooth map

$$\phi^{V,\mathbf{x}}_{t \leftarrow s} : F \to F$$

for each $(s, t) \in \Delta_T$ which maps a given initial value at time s to the solution at time t. Furthermore those solution maps satisfy

$$\phi^{V,\mathbf{x}}_{u \leftarrow t} \circ \phi^{V,\mathbf{x}}_{t \leftarrow s} = \phi^{V,\mathbf{x}}_{u \leftarrow s}.$$

The solution flows are obtained by approximating them by solution flows of ODEs in the following way. We have seen above that $\log_\otimes(\mathbf{x}_{s,t})$ is a Lie polynomial for all $(s, t) \in \Delta_T$. Thus $V_{\log_\otimes(\mathbf{x}_{s,t})}$ is a vector field and hence we can consider its time-1 flow $\phi_1^{V_{\log_\otimes(\mathbf{x}_{s,t})}}$. One can then show that an appropriate Trotter-type approximation converges and the limit is then defined to be the solution flow of (RDE).

Remark 3.13 The Trotter-type approximation of the RDE solution flow by ODE solution flows along iterated Lie brackets will be an important tool in Chapter 5, where we describe it in detail and use it to show our new result concerning the global existence of solutions.

3.2.3. Local and global existence of solutions

After this superficial survey we will now fix a rigorous definition for solutions of (RDE). For the time being F is a Banach space and we will extend the following definition to manifolds in Chapter 5.

Definition 3.14 We say that a continuous path $y : [0, T] \to F$ solves (RDE) if there exists a function $\theta : \mathbf{R} \to \mathbf{R}$ with $\theta(h)/h \to 0$ for $h \to 0$ such that

$$\|y_t - V_{\mathbf{x}_{s,t}} \operatorname{id}_F(y_s)\| \leq \theta(\omega(s,t)) \tag{3.5}$$

holds for all $(s, t) \in \Delta_T$.

Recall that we assume V_e to be a smooth vector field in F for all $e \in E$. While there are interesting results about the minimal regularity of the vector fields that is necessary to have local existence and/or uniqueness of solutions as in [FV10], we will not discuss them in the present work, i.e. we will always assume that the vector fields are sufficiently regular (e.g. smooth) and focus on their global properties.

Proposition 3.15 *For each $\xi \in F$*

- *there either exists a continuous map $y : [0, T] \to F$ such that y solves* (RDE)
- *or there exists $\tau \in (0, T]$ and a continuous map $y : [0, \tau) \to F$ such that*
  - *(i) the restriction of y to $[0, t]$ solves* (RDE) *for all $t \in (0, \tau)$ and*
  - *(ii) the image of $[0, \tau)$ under y is not relatively compact.*

In either case the solution y is unique.

In the first case, i.e. if y is defined on the whole interval $[0, T]$, we say that (RDE) has a *global solution*. Let us now review existing results about the existence of global solutions. For this purpose we assume that $E = \mathbf{R}^d$ and $F = \mathbf{R}^n$ are finite-dimensional vector spaces until the end of this chapter.

In its original formulation [Lyo98, Thm. 4.1.1] provides a combined existence and uniqueness result which, when translated into our smooth setting, reads as follows.

Proposition 3.16 *Assume that V_e is a bounded vector field with bounded derivatives up to order $[p]+1$ for all $e \in \mathbf{R}^d$. Then* (RDE) *has a unique global solution for all initial values $\xi \in \mathbf{R}^n$.*

This can be seen as a generalisation of Picard's existence (and uniqueness) theorem. It has first been observed in [Dav08, Thm. 2.2] that one can relax the assumptions on the order up to which the derivatives have to be bounded and thus obtain a result which is more akin to Peano's existence theorem. See also [FV10, Thm. 10.14] for a more general formulation. With our assumptions the result can be formulated as follows.

Proposition 3.17 *Assume that V_e is a bounded vector field with bounded derivatives up to order $[p]$ for all $e \in \mathbf{R}^d$. Then* (RDE) *has a global solution for all initial values $\xi \in \mathbf{R}^n$.*

Note that neither of the two above results captures linear vector fields since they are not bounded. However, one can show separately that linear rough differential equations always have global solutions. [FV10, Ex. 10.56] provides a result which extends Proposition 3.17 in such a way that it includes linear vector fields. However, it has been pointed out in [BC18] that the sketched proof of this result is not correct. Accordingly, [BC18] offers a corrected proof which also leads to a more general result than [FV10, Ex. 10.56]. We will come back to [BC18] after Theorem 5.16. A similar line of thought has been followed in [Lej12] with a focus on the special case $p \in (2,3)$. However, all of the above results have in common that they do not address the situation where $\mathbf{R}^n$ is replaced by a general manifold and, as a consequence, they formulate the conditions on the global behaviour of the vector fields in terms of the canonical Euclidean metric in $\mathbf{R}^n$. Therefore those results do not apply to certain interesting cases, see e.g. Example 5.29. These observations serve as a motivation for Chapter 5, where we obtain Theorem 5.16, which captures all previous global existence results as well as previously unknown cases.

4. Branched rough paths in infinite-dimensional spaces

In order to introduce branched rough paths in the sense of [Gub10] we have to give a precise definition of the relevant tree algebra which will play the role that the tensor algebra plays in the definition of a geometric rough path. So far, branched rough paths have only been studied in a finite-dimensional setting [Gub10, HK15]. Thus the first new contribution within this chapter is to develop a theory of tree-algebras over arbitrary locally convex vector spaces. We will then derive a change-of-variable formula for geometric and branched rough paths following a purely algebraic reasoning without referring to any approximation argument. This is the second main contribution of this chapter and it will allow us to clarify, when the difference between branched and geometric rough paths becomes relevant, complementing the results of [HK15, BC17].

4.1. Bialgebras of trees

In Chapter 2 we have introduced two different bialgebras on the set of tensors, which form a dual pair of bialgebras. We have started with the tensor product $\otimes$, which is easy to define. We have then defined the compatible coproduct $\Delta_{\sqcup\!\sqcup}$ recursively, exploiting that the tensor algebra is generated by the underlying vector space. Finally the shuffle product $\sqcup\!\sqcup$ and the coproduct $\Delta_{\otimes}$ are given as the duals of $\Delta_{\sqcup\!\sqcup}$ and $\otimes$ respectively. In the present chapter we will follow an analogous strategy in order to obtain two bialgebra structure on sets of trees with minimal effort. We will start with the Connes-Kreimer product $\circ$ and then we define the associated coproduct Δ_{CK} recursively. The Grossman-Larson product $\star$ and coproduct Δ_{GL} are then given as the respective dual operations.

4.1.1. The set of non-planar rooted trees

There exist various equivalent definitions for the set of non-planar rooted trees and here we choose to follow [Hof03] and work with equivalence classes of finite partially ordered sets.

Two partially ordered sets are defined to be *equivalent* if there exists an order preserving bijection between them.

Definition 4.1 A *Non-planar rooted tree* $\mathfrak{t}$ is an equivalence class of finite partially ordered sets, where each underlying set has a unique minimal element and for every $v \in S$ the set $\{w \in S : w < v\}$ is totally ordered. The set of all (non-planar rooted) trees is denoted by $\mathfrak{Tr}$.

We abuse notation in that by a tree $\mathfrak{t} \in \mathfrak{Tr}$ we usually mean a generic representative of the equivalence class and we will address the problem of well-definedness where we deem it necessary.

The elements of a tree $\mathfrak{t} \in \mathfrak{Tr}$ are called *vertices* and its unique minimal element is called its *root* and it is denoted by $\mathrm{r}(\mathfrak{t})$. By $\mathrm{v}(\mathfrak{t}) := \mathfrak{t} \backslash \mathrm{r}(\mathfrak{t})$ we denote the set of all vertices of $\mathfrak{t}$ without the root. Finally we define $|\mathfrak{t}|$ to be the number of elements of $\mathrm{v}(\mathfrak{t})$ and we call this number the *degree* of $\mathfrak{t}$. There exists a unique $\mathfrak{t} \in \mathfrak{Tr}$ with $|\mathfrak{t}| = 0$ and we denote this tree by $\mathbf{1}$.

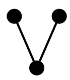
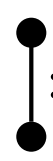
The *symmetry group* $\mathrm{SG}(\mathfrak{t})$ of $\mathfrak{t} \in \mathfrak{Tr}$ is the group of all order-preserving bijections from $\mathrm{v}(\mathfrak{t})$ to $\mathrm{v}(\mathfrak{t})$.

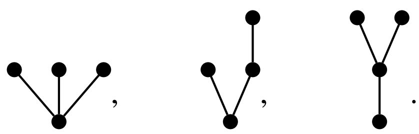
For $\mathfrak{t} \in \mathfrak{Tr}$ and $v \in \mathfrak{t}$, denote by $\mathrm{c}(v)$ the set of *children* of v, i.e. all vertices $w \in \mathfrak{t}$ such that $w > v$ and such that there exists no $x \in \mathfrak{t}$ with $w > x > v$.

Trees can be canonically interpreted as graphs where the vertices of the graph correspond to the vertices of the tree and the edges of the graph correspond to the statement 'is a child of'.

An important operation is that of *grafting* two trees together. Let $\mathfrak{t}, \mathfrak{s} \in \mathfrak{Tr}$ be two trees. A grafting map from $\mathfrak{t}$ to $\mathfrak{s}$ is a map $d : \mathrm{c}(\mathrm{r}(\mathfrak{t})) \to \mathfrak{s}$, i.e. it assigns a vertex of $\mathfrak{s}$ to every child of the root of $\mathfrak{t}$. Denote by $\mathrm{Gr}(\mathfrak{t}, \mathfrak{s})$ the set of all such maps. Given $d \in \mathrm{Gr}(\mathfrak{t}, \mathfrak{s})$ we define $\mathfrak{t} \overset{d}{\rightharpoonup} \mathfrak{s} := \mathrm{v}(\mathfrak{t}) \sqcup \mathfrak{s}$ with the additional relation $v > d(v)$ for every $v \in \mathrm{c}(\mathrm{r}(\mathfrak{t}))$ (plus all necessary additional relations so that transitivity holds). Here $\sqcup$ denotes the disjoint union.

In the special case where $d(v) = \mathrm{r}(\mathfrak{s})$ for all $v \in \mathrm{c}(\mathrm{r}(\mathfrak{t}))$ we write $\mathfrak{t} \circ \mathfrak{s}$ instead of $\mathfrak{t} \overset{d}{\rightharpoonup} \mathfrak{s}$. Observe that the map $\mathfrak{Tr} \times \mathfrak{Tr} \to \mathfrak{Tr} : (\mathfrak{t}, \mathfrak{s}) \mapsto \mathfrak{t} \circ \mathfrak{s}$ is both associative and commutative since the operation $\circ$ simply identifies the roots of $\mathfrak{t}$ and $\mathfrak{s}$.

Example 4.2 These are all possible ways of grafting $\mathfrak{t} =$  to $\mathfrak{s} =$ :



Note however that $|\mathrm{Gr}(\mathfrak{t}, \mathfrak{s})| = 4$ since there are two possibilities to obtain the middle one of the above three trees. Furthermore the first tree corresponds to $\mathfrak{t} \circ \mathfrak{s}$.

Let $\mathsf{t} \in \mathfrak{Tr}$ be a tree. We call a subset $C \subset \mathrm{v}(\mathsf{t})$ of its vertices an *admissible cut* if none of its elements are comparable, i.e. if for all $v, w \in C$ we have neither $v \prec w$ nor $w \prec v$. The set of all admissible cuts of t is denoted by $\mathrm{Cut}(\mathsf{t})$. For a given cut $C \in \mathrm{Cut}(\mathsf{t})$ we define the trees

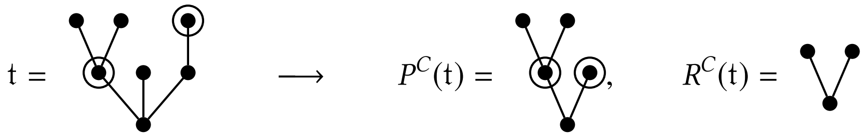
$$R^C(\mathsf{t}) := \mathsf{t} \setminus \{v \in \mathrm{v}(\mathsf{t}) : \text{there exists } w \in C \text{ such that } w \preceq v\}$$

and

$$P^C(\mathsf{t}) := \{r\} \sqcup \{v \in \mathrm{v}(\mathsf{t}) : \text{there exists } w \in C \text{ such that } w \preceq v\}$$

with the additional relations $r \prec w$ for all $w \in C$ (plus all necessary additional relations so that transitivity holds). Intuitively, $R^C(\mathsf{t})$ contains the tree that remains after cutting away the elements of C and $P^C(\mathsf{t})$ is obtained by grafting all the subtrees that have been removed to a new root.

Example 4.3 The result of the cutting operation when C is given by the highlighted vertices.



In order to keep track of the vertices of a tree $\mathsf{t} \in \mathfrak{Tr}$, we would like to consider bijections $\rho : \mathrm{v}(\mathsf{t}) \to \{1, \dots, |\mathsf{t}|\}$. However, since we are working with equivalence classes of trees (even though we usually abuse notation and work with representatives as mentioned above), we have to consider equivalence classes of such bijections. Thus we assume that every representative T of t comes with a bijective map $\rho_T : \mathrm{v}(T) \to \{1, \dots, |\mathsf{t}|\}$ which we call the *enumeration of* T. Furthermore we assume that for any two representatives T and $\tilde{T}$ of t the enumerations ρ_T and $\rho_{\tilde{T}}$ are equivalent in the sense that there exists an order preserving bijection $\varphi : T \to \tilde{T}$ with $\rho_T = \rho_{\tilde{T}} \circ \varphi$.

For a tree $\mathsf{t} \in \mathfrak{Tr}$ we obtain an embedding of $\mathrm{SG}(\mathsf{t})$ into the symmetric group $\mathfrak{S}_{|\mathsf{t}|}$ by its enumeration ρ_{t}, i.e. we define the group

$$\{\rho_{\mathsf{t}} \pi \rho_{\mathsf{t}}^{-1} : \pi \in \mathrm{SG}(\mathsf{t})\} \subseteq \mathfrak{S}_{|\mathsf{t}|}.$$

We will abuse notation and denote this embedding by $\mathrm{SG}(\mathsf{t})$ as well.

For $k \in \mathbf{N}$ we denote by $\mathfrak{Tr}_k \subset \mathfrak{Tr}$ the set of all trees whose root has exactly k children and by $\mathfrak{Tr}^k \subset \mathfrak{Tr}$ the set of all trees with degree k. We will also use the notation $\mathfrak{Tr}^k_n := \mathfrak{Tr}^k \cap \mathfrak{Tr}_n$. The set $\mathfrak{Tr}_1$ is especially important as the following trivial yet important lemma shows.

Lemma 4.4 *Let* $\mathrm{t} \in \mathfrak{Tr}_k$*. Then there exist unique trees* $\mathrm{t}_1, \ldots, \mathrm{t}_k \in \mathfrak{Tr}_1$ *such that we have the decomposition* $\mathrm{t} = \mathrm{t}_1 \circ \cdots \circ \mathrm{t}_k$*.*

Let us finally define the bijection

$$\mathfrak{B}^+ : \mathfrak{Tr} \to \mathfrak{Tr}_1, \tag{4.1}$$

which adds a new root to a given tree $\mathrm{t} \in \mathfrak{Tr}$ and thus the old root of t becomes the unique child of the new root.

4.1.2. Algebras of labelled trees over a vector space

Let M be a set. The most straightforward way to define a labelling of the vertices of a tree $\mathrm{t} \in \mathfrak{Tr}$ with labels from M is to consider a map $l : \mathrm{v}(\mathrm{t}) \to M$. More precisely, we have to consider equivalence classes of such maps where two maps l and $\tilde{l}$ are equivalent if there exists $\pi \in \mathrm{SG}(\mathrm{t})$ such that $l = \tilde{l}\pi$. We can then consider the (free) vector space spanned by all of those equivalence classes of labellings for a fixed tree $\mathrm{t} \in \mathfrak{Tr}$. Let us denote this vector space by $\mathcal{L}_{\mathrm{t}}(M)$.

This definition is perfectly fine and especially if M is finite there is not much of a difference between labelled and unlabelled trees since in that case the space $\mathcal{L}_{\mathrm{t}}(M)$ is finite dimensional for every $\mathrm{t} \in \mathfrak{Tr}$. However, if M is infinite, then the vector space $\mathcal{L}_{\mathrm{t}}(M)$ is infinite dimensional for every $\mathrm{t} \in \mathfrak{Tr}\backslash\{\mathbf{1}\}$. It is therefore desirable to equip $\mathcal{L}_{\mathrm{t}}(M)$ with a topology since topological vector spaces are usually nicer to work with than infinite dimensional vector spaces without a topology, for example when it comes to dual spaces.

Hence we would like to establish a different point of view which takes the (free) vector space E spanned by M as the label set rather than the set M itself. In fact, it is usually E that is given in the first place rather than M. Thus our aim is to describe the spaces $\mathcal{L}_{\mathrm{t}}(M)$ in terms of E rather than M so that we can then use a given topology on E to construct a topology on $\mathcal{L}_{\mathrm{t}}(M)$.

In terms of the application to rough paths that we will present in Section 4.2.1 the vector space E will take the role of the noise space. The topological tree algebra generated by it – in the way given below – will then be the space in which the rough path lift of the noise takes its values.

Let now E be a given vector space. We will perform a purely algebraic construction in this section and then consider topological vector spaces in the following one. A permutation $\sigma \in \mathfrak{S}_k$ with $k \in \mathbf{N}$ can be canonically interpreted as a linear map on $E^{\otimes k}$ which is given by $\sigma(e_1 \otimes \cdots \otimes e_k) = e_{\sigma(1)} \otimes \cdots \otimes e_{\sigma(k)}$ for elementary tensors. Given a subgroup $G \subset \mathfrak{S}_k$ we define the projection

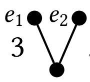
$$\mathrm{sym}_G : E^{\otimes k} \to E^{\otimes k} : a \mapsto \frac{1}{|G|} \sum_{\sigma \in G} \sigma(a).$$

Definition 4.5 Let E be a vector space, let $\mathfrak{t} \in \mathfrak{Tr}$ be a tree and let $\mathrm{SG}(\mathfrak{t})$ be its symmetry group. Then the $\mathfrak{t}$-*tensor power of* E is defined as the quotient space

$$E^{\otimes \mathfrak{t}} := E^{\otimes |\mathfrak{t}|} \Big/ \ker\left(\mathrm{sym}_{\mathrm{SG}(\mathfrak{t})}\right)$$

The space $E^{\otimes \mathfrak{t}}$ is isomorphic to the image of $\mathrm{sym}_{\mathrm{SG}(\mathfrak{t})}$, which provides a more direct description of $E^{\otimes \mathfrak{t}}$. We say that an element $\mathbf{t} \in E^{\otimes \mathfrak{t}}$ is an *elementary labelled tree* if its equivalence class contains an elementary tensor. The set of all elementary labelled trees in $E^{\otimes \mathfrak{t}}$ spans $E^{\otimes \mathfrak{t}}$.

To illustrate an elementary labelled tree we choose one of its elementary representatives as in the following example. E.g., for $\mathfrak{t} =$  we can visualise $\mathbf{t} = [2e_1 \otimes e_2 + e_2 \otimes e_1] \in E^{\otimes \mathfrak{t}}$ as

$\mathbf{t} = 3$  .

Let us now define the following spaces.

$$\mathfrak{Tr}^k(E) := \bigoplus_{\mathfrak{t} \in \mathfrak{Tr}^k} E^{\otimes \mathfrak{t}} \text{ for } k \in \mathbf{N}$$
$$\mathfrak{Tr}_k(E) := \bigoplus_{\mathfrak{t} \in \mathfrak{Tr}_k} E^{\otimes \mathfrak{t}} \text{ for } k \in \mathbf{N}$$
$$\mathfrak{Tr}^k_n(E) := \bigoplus_{\mathfrak{t} \in \mathfrak{Tr}^k_n} E^{\otimes \mathfrak{t}} = \mathfrak{Tr}^k(E) \cap \mathfrak{Tr}_n(E) \text{ for } k, n \in \mathbf{N}$$
$$\mathfrak{Tr}(E) := \bigoplus_{\mathfrak{t} \in \mathfrak{Tr}} E^{\otimes \mathfrak{t}} = \bigoplus_{k=0}^{\infty} \mathfrak{Tr}^k(E) = \bigoplus_{k=0}^{\infty} \mathfrak{Tr}_k(E) = \bigoplus_{k=0}^{\infty} \bigoplus_{n=0}^{\infty} \mathfrak{Tr}^k_n(E)$$

The definition of $\mathfrak{Tr}(E)$ provides canonical projections onto $\mathfrak{Tr}^k(E)$, $\mathfrak{Tr}_k(E)$ and $\mathfrak{Tr}^k_n(E)$ for $k, n \in \mathbf{N}$ and we denote these projections by π^k, π_k and π^k_n respectively.

Remark 4.6 We will be working with two different types of equivalence classes. On the one hand, the elements of $\mathfrak{Tr}$ are equivalence classes of partially ordered sets, on the other hand the elements of $\mathfrak{Tr}(E)$ are equivalence classes of tensors. In order to keep the confusion that arises from this to a minimum, we will adhere to the following notation throughout this chapter. Fraktur letters $\mathfrak{t}, \mathfrak{s} \in \mathfrak{Tr}$ denote trees and capital letters T, S are representatives of $\mathfrak{t}$ and $\mathfrak{s}$ respectively if we need to refer to them explicitly. Bold face letters $\mathbf{t}, \mathbf{s} \in \mathfrak{Tr}(E)$ denote labelled trees and lower case letters t, s denote representatives of $\mathbf{t}$ and $\mathbf{s}$ respectively.

We will now follow the same procedure that we used for the tensor algebra in Chapter 2 in order to introduce a bialgebra structure. We start by defining an associative product on $\mathfrak{Tr}(E)$.

Definition 4.7 Let $\mathrm{t}, \mathfrak{s} \in \mathfrak{Tr}$, let $\mathbf{t} \in E^{\otimes \mathrm{t}}$ and $\mathbf{s} \in E^{\otimes \mathfrak{s}}$ and let $t \in E^{\otimes |\mathrm{t}|}$ and $s \in E^{\otimes |\mathfrak{s}|}$ be representatives of $\mathbf{t}$ and $\mathbf{s}$ respectively. Then we define

$$\mathbf{t} \circ \mathbf{s} := \left[\left(\rho_{\mathrm{t},\mathfrak{s}} \circ \rho^{-1}_{\mathrm{t}\circ\mathfrak{s}} \right) (t \otimes s) \right] \in E^{\mathrm{t}\circ\mathfrak{s}}, \tag{4.2}$$

where

$$\rho_{\mathrm{t},\mathfrak{s}} : \mathrm{v}(\mathrm{t}) \sqcup \mathrm{v}(\mathfrak{s}) \longrightarrow \{1, \ldots, |\mathrm{t}| + |\mathfrak{s}|\} : v \mapsto \begin{cases} \rho_{\mathrm{t}}(v) & v \in \mathrm{v}(\mathrm{t}), \\ \rho_{\mathfrak{s}}(v) + |\mathrm{t}| & v \in \mathrm{v}(\mathfrak{s}). \end{cases}$$

The operation $\circ$ is then extended to a bilinear map on $\mathfrak{Tr}(E) \times \mathfrak{Tr}(E)$.

Proposition 4.8 *The expression* $\mathbf{t} \circ \mathbf{s}$ *is well-defined in that it depends neither on the choice of* t *and* s *nor on the choice of representatives of* t *and* $\mathfrak{s}$*. Furthermore the map*

$$\mathfrak{Tr}(E) \times \mathfrak{Tr}(E) \longrightarrow \mathfrak{Tr}(E) : (\mathbf{t}, \mathbf{s}) \mapsto \mathbf{t} \circ \mathbf{s}$$

is associative.

Proof. Choosing a different representative of t and/or a different choice of t amounts to replacing t by $\sigma(t)$ for some $\sigma \in \mathrm{SG}(\mathrm{t})$. Furthermore, note that $\mathrm{SG}(\mathrm{t})$ is a subgroup of $\mathrm{SG}(\mathrm{t} \circ \mathfrak{s})$ since $\circ$ simply identifies the roots of t and $\mathfrak{s}$. Thus there exists $\widehat{\sigma} \in \mathrm{SG}(\mathrm{t} \circ \mathfrak{s})$ with the property that $\widehat{\sigma}|_{\mathrm{t}} = \sigma$ and $\widehat{\sigma}|_{\mathfrak{s}} = \mathrm{id}_{\mathfrak{s}}$ Hence we have

$$\mathbf{t} \circ \mathbf{s} = \left[\left(\rho_{\mathrm{t},\mathfrak{s}} \circ \rho^{-1}_{\mathrm{t}\circ\mathfrak{s}} \right) (t \otimes s) \right] = \left[\widehat{\sigma} \left(\left(\rho_{\mathrm{t},\mathfrak{s}} \circ \rho^{-1}_{\mathrm{t}\circ\mathfrak{s}} \right) (t \otimes s) \right) \right] = \left[\left(\rho_{\mathrm{t},\mathfrak{s}} \circ \rho^{-1}_{\mathrm{t}\circ\mathfrak{s}} \right) (\sigma(t) \otimes s) \right]$$

and therefore the value of $\mathbf{t} \circ \mathbf{s}$ does not change if we replace t by $\sigma(t)$. The same argument shows that choosing a different representative of $\mathfrak{s}$ and/or a different choice of s does not alter $\mathbf{t} \circ \mathbf{s}$.

To prove associativity, let $\mathrm{t}_1, \mathrm{t}_2, \mathrm{t}_3 \in \mathfrak{Tr}$ and $\mathbf{t}_1 \in E^{\otimes \mathrm{t}_1}$, $\mathbf{t}_2 \in E^{\otimes \mathrm{t}_2}$, $\mathbf{t}_3 \in E^{\otimes \mathrm{t}_3}$ and let $t_1 \in E^{|\mathrm{t}_1|}$, $t_2 \in E^{|\mathrm{t}_2|}$ and $t_3 \in E^{|\mathrm{t}_3|}$ be representatives of $\mathbf{t}_1$, $\mathbf{t}_2$ and $\mathbf{t}_3$ respectively. Then we have

$$(\mathbf{t}_1 \circ \mathbf{t}_2) \circ \mathbf{t}_3 = \left[\left(\rho_{\mathrm{t}_1\circ\mathrm{t}_2,\mathrm{t}_3} \circ \rho^{-1}_{\mathrm{t}_1\circ\mathrm{t}_2\circ\mathrm{t}_3} \right) \left(\left(\rho_{\mathrm{t}_1,\mathrm{t}_2} \circ \rho^{-1}_{\mathrm{t}_1\circ\mathrm{t}_2} \right) (t_1 \otimes t_2) \otimes t_3 \right) \right]$$

and

$$\mathbf{t}_1 \circ (\mathbf{t}_2 \circ \mathbf{t}_3) = \left[\left(\rho_{\mathrm{t}_1,\mathrm{t}_2\circ\mathrm{t}_3} \circ \rho^{-1}_{\mathrm{t}_1\circ\mathrm{t}_2\circ\mathrm{t}_3} \right) \left(t_1 \otimes \left(\rho_{\mathrm{t}_2,\mathrm{t}_3} \circ \rho^{-1}_{\mathrm{t}_2\circ\mathrm{t}_3} \right) (t_2 \otimes t_3) \right) \right]$$

Both of these expressions are equal to

$$\left[\left(\rho_{\mathrm{t}_1,\mathrm{t}_2,\mathrm{t}_3} \circ \rho^{-1}_{\mathrm{t}_1\circ\mathrm{t}_2\circ\mathrm{t}_3}\right)(t_1 \otimes t_2 \otimes t_3)\right],$$

where

$$\rho_{\mathrm{t}_1,\mathrm{t}_2,\mathrm{t}_3} : \mathrm{v}(\mathrm{t}_1) \sqcup \mathrm{v}(\mathrm{t}_2) \sqcup \mathrm{v}(\mathrm{t}_3) \to \{1,\dots,|\mathrm{t}_1|+|\mathrm{t}_2|+|\mathrm{t}_3|\} : v \mapsto \begin{cases} \rho_{\mathrm{t}_1}(v) & v \in \mathrm{v}(\mathrm{t}_1), \\ \rho_{\mathrm{t}_2}(v) + |\mathrm{t}_1| & v \in \mathrm{v}(\mathrm{t}_2), \\ \rho_{\mathrm{t}_3}(v) + |\mathrm{t}_1| + |\mathrm{t}_2| & v \in \mathrm{v}(\mathrm{t}_3). \end{cases}$$

In particular we have $\mathbf{t}_1 \circ (\mathbf{t}_2 \circ \mathbf{t}_3) = (\mathbf{t}_1 \circ \mathbf{t}_2) \circ \mathbf{t}_3$, which finishes the proof. ■

If $\mathbf{t} \in E^{\otimes \mathrm{t}}$ is an elementary labelled tree as defined above, then the unique decomposition $\mathrm{t} = \mathrm{t}_1 \circ \cdots \circ \mathrm{t}_k$ from Lemma 4.4 induces a decomposition

$$\mathbf{t} = \mathbf{t}_1 \circ \cdots \circ \mathbf{t}_k \text{ with } \mathbf{t}_1 \in E^{\otimes \mathrm{t}_1}, \dots, \mathbf{t}_k \in E^{\otimes \mathrm{t}_k}. \tag{4.3}$$

Now we introduce a coproduct on $\mathfrak{Tr}(E)$ in a recursive way, such that the compatibility with $\circ$ is immediately apparent from the definition. To do so, we have to extend the map $\mathfrak{B}^+$ as in (4.1) from $\mathfrak{Tr}$ to $\mathfrak{Tr}(E)$. For a vector $e \in E$ the map $\mathfrak{B}^+_e : \mathfrak{Tr}(E) \to \mathfrak{Tr}_1(E)$ is given by

$$\mathfrak{B}^+_e(\mathbf{t}) = [(\rho^+_{\mathrm{t}} \circ \rho^{-1}_{\mathfrak{B}^+(\mathrm{t})})(t \otimes e)],$$

where $\mathbf{t} \in E^{\otimes \mathrm{t}}$, $\mathrm{t} \in \mathfrak{Tr}$ and $t \in E^{\otimes |\mathrm{t}|}$ is a representative of $\mathbf{t}$. The map

$$\rho^+_{\mathrm{t}} : \mathrm{v}(\mathfrak{B}^+(\mathrm{t})) \to \{1,\dots,|\mathrm{t}|+1\}$$

is given by $\rho^+_{\mathrm{t}}(v) = \rho_{\mathrm{t}}(v)$ for $v \notin \mathrm{c}(\mathrm{r}(\mathfrak{B}^+(\mathrm{t})))$ and $\rho^+_{\mathrm{t}}(v) = |\mathrm{t}|+1$ for $v \in \mathrm{c}(\mathrm{r}(\mathfrak{B}^+(\mathrm{t})))$. Note that this definition makes sense since the root of $\mathfrak{B}^+(\mathrm{t})$ has only one child. The well-definedness follows analogously to the proof of Proposition 4.8. The crucial observation for the proof is that $\mathrm{SG}(\mathfrak{B}^+(\mathrm{t}))$ and $\mathrm{SG}(\mathrm{t})$ are isomorphic.

For fixed $e \in E$ the map $\mathfrak{B}^+_e$ is certainly not surjective as a map from $\mathfrak{Tr}(E)$ to itself. Nevertheless, every $\mathbf{t} \in \mathfrak{Tr}_1(E)$ is contained in the image of $\mathfrak{B}^+_e$ for some $e \in E$ and e is unique up to scalar multiples.

Definition 4.9 Define the linear map $\Delta_{\mathrm{CK}} : \mathfrak{Tr}(E) \to \mathfrak{Tr}(E) \otimes \mathfrak{Tr}(E)$ by

- $\Delta_{\mathrm{CK}}(\mathbf{1}) = \mathbf{1} \otimes \mathbf{1}$,

- $\Delta_{\mathrm{CK}}(\mathfrak{B}_e^+(\mathbf{t})) = \mathfrak{B}_e^+(\mathbf{t}) \otimes \mathbf{1} + (\mathrm{id}_{\mathfrak{Tr}(E)} \otimes \mathfrak{B}_e^+)(\Delta_{\mathrm{CK}}(\mathbf{t}))$ for all trees $\mathbf{t} \in \mathfrak{Tr}(E)$ and all $e \in E$,
- $\Delta_{\mathrm{CK}}(\mathbf{t}_1 \circ \cdots \circ \mathbf{t}_k) = \Delta_{\mathrm{CK}}(\mathbf{t}_1) \circ \cdots \circ \Delta_{\mathrm{CK}}(\mathbf{t}_k)$ for all trees $\mathbf{t}_1, \ldots, \mathbf{t}_k \in \mathfrak{Tr}_1(E)$.

We will give an explicit, non-recursive description of Δ_{CK} below.

Proposition 4.10 *The tuple*

$$\mathcal{H}_{\mathrm{CK}}(E) := (\mathfrak{Tr}(E), \circ, \Delta_{\mathrm{CK}})$$

is a bialgebra.

Proof. The only non-obvious property that remains to be shown is the coassociativity of Δ_{CK}. Let $\mathfrak{t} = \overset{\bullet}{\underset{\bullet}{|}}$ and $\mathbf{t} \in E^{\otimes \mathfrak{t}}$. Then we have

$$(\Delta_{\mathrm{CK}} \otimes \mathrm{id}_{\mathfrak{Tr}(E)})(\Delta_{\mathrm{CK}}(\mathbf{t})) = (\mathrm{id}_{\mathfrak{Tr}(E)} \otimes \Delta_{\mathrm{CK}})(\Delta_{\mathrm{CK}}(\mathbf{t})) = \mathbf{t} \otimes \mathbf{1} \otimes \mathbf{1} + \mathbf{1} \otimes \mathbf{t} \otimes \mathbf{1} + \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{t}.$$

Thus assume that coassociativity holds for all trees of degree n or less for some $n \in \mathbf{N}$. Then for $\mathbf{t} \in \mathfrak{Tr}^n(E)$ and $e \in E$ we have

$$\begin{aligned}
&(\Delta_{\mathrm{CK}} \otimes \mathrm{id}_{\mathfrak{Tr}(E)})(\Delta_{\mathrm{CK}}(\mathfrak{B}_e^+\mathbf{t})) \\
=&(\Delta_{\mathrm{CK}} \otimes \mathrm{id}_{\mathfrak{Tr}(E)}) \left(\mathfrak{B}_e^+(\mathbf{t}) \otimes \mathbf{1} + (\mathrm{id}_{\mathfrak{Tr}(E)} \otimes \mathfrak{B}_e^+) (\Delta_{\mathrm{CK}} (\mathbf{t}))\right) && (4.4a)\\
=&\mathfrak{B}_e^+(\mathbf{t}) \otimes \mathbf{1} \otimes \mathbf{1} + (\mathrm{id}_{\mathfrak{Tr}(E)} \otimes \mathfrak{B}_e^+) (\Delta_{\mathrm{CK}} (\mathbf{t})) \otimes \mathbf{1} \\
&+ (\mathrm{id}_{\mathfrak{Tr}(E)} \otimes \mathrm{id}_{\mathfrak{Tr}(E)} \otimes \mathfrak{B}_e^+) \left((\Delta_{\mathrm{CK}} \otimes \mathrm{id}_{\mathfrak{Tr}(E)}) (\Delta_{\mathrm{CK}} (\mathbf{t}))\right) && (4.4b)\\
=&\mathfrak{B}_e^+(\mathbf{t}) \otimes \mathbf{1} \otimes \mathbf{1} + (\mathrm{id}_{\mathfrak{Tr}(E)} \otimes \mathfrak{B}_e^+) (\Delta_{\mathrm{CK}} (\mathbf{t})) \otimes \mathbf{1} \\
&+ (\mathrm{id}_{\mathfrak{Tr}(E)} \otimes \mathrm{id}_{\mathfrak{Tr}(E)} \otimes \mathfrak{B}_e^+) \left((\mathrm{id}_{\mathfrak{Tr}(E)} \otimes \Delta_{\mathrm{CK}}) (\Delta_{\mathrm{CK}} (\mathbf{t}))\right) && (4.4c)\\
=&\mathfrak{B}_e^+(\mathbf{t}) \otimes \mathbf{1} \otimes \mathbf{1} \\
&+ (\mathrm{id}_{\mathfrak{Tr}(E)} \otimes \Delta_{\mathrm{CK}}) \left((\mathrm{id}_{\mathfrak{Tr}(E)} \otimes \mathfrak{B}_e^+) (\Delta_{\mathrm{CK}} (\mathbf{t}))\right) && (4.4d)\\
=&(\mathrm{id}_{\mathfrak{Tr}(E)} \otimes \Delta_{\mathrm{CK}}) \left(\mathfrak{B}_e^+(\mathbf{t}) \otimes \mathbf{1} + (\mathrm{id}_{\mathfrak{Tr}(E)} \otimes \mathfrak{B}_e^+) (\Delta_{\mathrm{CK}} (\mathbf{t}))\right) && (4.4e)\\
=&(\mathrm{id}_{\mathfrak{Tr}(E)} \otimes \Delta_{\mathrm{CK}}) \left(\Delta_{\mathrm{CK}}(\mathfrak{B}_e^+(\mathbf{t}))\right). && (4.4f)
\end{aligned}$$

The crucial step is (4.4c), where we have used the induction hypothesis. The other steps are obtained by applying the recursive relation from Definition 4.9. Thus we have shown that coassociativity holds for all $\mathbf{t} \in \mathfrak{Tr}_1^{n+1}(E)$. Since Δ_{CK} is an algebra morphism and since all elements of $\mathfrak{Tr}^{n+1}(E) \backslash \mathfrak{Tr}_1^{n+1}(E)$ can be written as products of elements of degree n or less, we find that coassociativity holds for all $\mathbf{t} \in \mathfrak{Tr}^{n+1}(E)$ and thus for all $\mathbf{t} \in \mathfrak{Tr}(E)$ by induction.

■

In order to give a non-recursive description of Δ_{CK} and its dual, we have to extend two auxiliary constructions from the previous section, namely grafting and cutting, to tensors.

Definition 4.11 Let $\mathfrak{t}, \mathfrak{s} \in \mathfrak{Tr}$ and let $t \in E^{\otimes|\mathfrak{t}|}$ and $s \in E^{\otimes|\mathfrak{s}|}$ be two elementary tensors. Let furthermore $d \in \mathrm{Gr}(\mathfrak{t}, \mathfrak{s})$ be a grafting map. Then we define

$$t \overset{d}{\rightharpoonup} s := \left(\rho_{\mathfrak{t},\mathfrak{s}} \circ \rho^{-1}_{\mathfrak{t} \overset{d}{\rightharpoonup} \mathfrak{s}}\right)(t \otimes s) \in E^{\otimes(|\mathfrak{t}|+|\mathfrak{s}|)}, \tag{4.5}$$

where

$$\rho_{\mathfrak{t},\mathfrak{s}} : \mathrm{v}(\mathfrak{t}) \sqcup \mathrm{v}(\mathfrak{s}) \to \{1, \ldots, |\mathfrak{t}| + |\mathfrak{s}|\} : v \mapsto \begin{cases} \rho_{\mathfrak{t}}(v) & v \in \mathrm{v}(\mathfrak{t}), \\ \rho_{\mathfrak{s}}(v) + |\mathfrak{t}| & v \in \mathrm{v}(\mathfrak{s}). \end{cases} \tag{4.6}$$

Remark 4.12 This definition has to be handled with care since the value of $t \overset{d}{\rightharpoonup} s$ may depend on the choice of representatives of $\mathfrak{t}$ and $\mathfrak{s}$. However, one can easily check that if T and $\tilde{T}$ represent $\mathfrak{t}$, and S and $\tilde{S}$ represent $\mathfrak{s}$, then there exist $\sigma_1 \in \mathrm{SG}(\mathfrak{t})$ and $\sigma_2 \in \mathrm{SG}(\mathfrak{s})$ so that $t \overset{d}{\rightharpoonup} s$, computed with respect to T and S, and $\sigma_1(t) \overset{d}{\rightharpoonup} \sigma_2(s)$, computed with respect to $\tilde{T}$ and $\tilde{S}$, coincide. This allows us to fix one representative of each $\mathfrak{t} \in \mathfrak{Tr}$, once and for all, because $t \overset{d}{\rightharpoonup} s$ is only an auxiliary construction for Definition 4.16, which will yield an object that is invariant under the action of $\mathrm{SG}(\mathfrak{t})$ and $\mathrm{SG}(\mathfrak{s})$.

Let us now turn to cutting. Let $\mathfrak{t} \in \mathfrak{Tr}$ and $C \in \mathrm{Cut}(\mathfrak{t})$. Then the vertices of $R^C(\mathfrak{t})$ and $P^C(\mathfrak{t})$ are defined as subsets of the vertices of $\mathfrak{t}$ and hence there are canonical injections

$$\phi_{C,R} : \mathrm{v}\left(R^C(\mathfrak{t})\right) \to \mathrm{v}(\mathfrak{t}) \qquad \text{and} \qquad \phi_{C,P} : \mathrm{v}\left(P^C(\mathfrak{t})\right) \to \mathrm{v}(\mathfrak{t}).$$

Definition 4.13 Let $\mathfrak{t} \in \mathfrak{Tr}$ be a tree and let $C \in \mathrm{Cut}(\mathfrak{t})$ be an admissible cut. Then we define

$$\mathbf{R}^C : E^{\otimes|\mathfrak{t}|} \to E^{\otimes|R^C(\mathfrak{t})|} : e_1 \otimes \cdots \otimes e_{|\mathfrak{t}|} \mapsto e_{\varphi_{C,R}(1)} \otimes \cdots \otimes e_{\varphi_{C,R}(|R^C(\mathfrak{t})|)}$$

and

$$\mathbf{P}^C : E^{\otimes|\mathfrak{t}|} \to E^{\otimes|P^C(\mathfrak{t})|} : e_1 \otimes \cdots \otimes e_{|\mathfrak{t}|} \mapsto e_{\varphi_{C,P}(1)} \otimes \cdots \otimes e_{\varphi_{C,P}(|P^C(\mathfrak{t})|)},$$

where

$$\varphi_{C,R} := \rho_{\mathfrak{t}} \circ \phi_{C,R} \circ \rho^{-1}_{R^C(\mathfrak{t})} \in \mathfrak{S}_{|R^C(\mathfrak{t})|} \quad \text{and} \quad \varphi_{C,P} := \rho_{\mathfrak{t}} \circ \phi_{C,P} \circ \rho^{-1}_{P^C(\mathfrak{t})} \in \mathfrak{S}_{|P^C(\mathfrak{t})|}.$$

Remark 4.14 As in Remark 4.12 we emphasize that $\mathbf{R}^C$ and $\mathbf{P}^C$ are only well-defined if we use one fixed representative for each $\mathfrak{t} \in \mathfrak{Tr}$.

Proposition 4.15 *Let* $\mathfrak{t} \in \mathfrak{Tr}$ *and let* $\mathbf{t} \in E^{\otimes \mathfrak{t}}$ *be an elementary tree with representative* $t \in E^{\otimes |\mathfrak{t}|}$. *Then we have*

$$\Delta_{\mathrm{CK}}(\mathbf{t}) = \sum_{C \in \mathrm{Cut}(\mathfrak{t})} \left[\mathbf{P}^C(t)\right] \otimes \left[\mathbf{R}^C(t)\right] .$$

In particular, the right-hand side is well-defined and does not depend on the choice of t.

Proof. For $\mathbf{t} = \mathbf{1}$ the claim is certainly true. Assume that it holds for all trees of degree n or less for some $n \in \mathbf{N}$ and let $\mathfrak{t} \in \mathfrak{Tr}^n$, $\mathbf{t} \in E^{\otimes \mathfrak{t}}$ and $e \in E$. Then we have

$$\begin{aligned}
&\Delta_{CK}(\mathfrak{B}_e^+(\mathbf{t})) \\
=&\mathfrak{B}_e^+(\mathbf{t}) \otimes \mathbf{1} + (\mathrm{id}_{\mathfrak{Tr}(E)} \otimes \mathfrak{B}_e^+)(\Delta_{\mathrm{CK}}(\mathbf{t})) \\
=&\mathfrak{B}_e^+(\mathbf{t}) \otimes \mathbf{1} + \sum_{C \in \mathrm{Cut}(\mathfrak{t})} \left[\mathbf{P}^C(t)\right] \otimes \mathfrak{B}_e^+ \left(\left[\mathbf{R}^C(t)\right]\right) \\
=& \sum_{C \in \mathrm{Cut}(\mathfrak{B}(\mathfrak{t}))} \left[\mathbf{P}^C((\rho_{\mathfrak{t}}^+ \circ \rho_{\mathfrak{B}^+(\mathfrak{t})}^{-1})(t \otimes e))\right] \otimes \left[\mathbf{R}^C((\rho_{\mathfrak{t}}^+ \circ \rho_{\mathfrak{B}^+(\mathfrak{t})}^{-1})(t \otimes e))\right]
\end{aligned}$$

The final equality can be seen as follows. There is a canonical embedding of $\mathrm{Cut}(\mathfrak{t})$ into $\mathrm{Cut}(\mathfrak{B}^+(\mathfrak{t}))$. In fact, the difference of those two sets contains exactly one element, namely the cut which contains only the unique child of the root of $\mathfrak{B}^+(\mathfrak{t})$. This cut yields the term $\mathfrak{B}_e^+(\mathbf{t}) \otimes \mathbf{1}$. For all other cuts, the $\mathbf{P}$-term is unchanged and the $\mathbf{R}$-term obtains the new root. Since all representatives of $\mathfrak{B}_e^+(\mathbf{t})$ can be written in the form $(\rho_{\mathfrak{t}}^+ \circ \rho_{\mathfrak{B}^+(\mathfrak{t})}^{-1})(t \otimes e)$ for some representative t of $\mathbf{t}$, this shows that the claim holds for all $\mathbf{t} \in \mathfrak{Tr}_1(E)$. Furthermore, for two elementary trees $\mathbf{t} \in \mathfrak{Tr}(E)$ and $\mathbf{s} \in \mathfrak{Tr}(E)$ we have

$$\begin{aligned}
&\Delta_{\mathrm{CK}}(\mathbf{t} \circ \mathbf{s}) = \Delta_{\mathrm{CK}}(\mathbf{t}) \circ \Delta_{\mathrm{CK}}(\mathbf{s}) \\
=& \sum_{C_1 \in \mathrm{Cut}(\mathfrak{t})} \left[\mathbf{P}^{C_1}(t)\right] \otimes \left[\mathbf{R}^{C_1}(t)\right] \circ \sum_{C_2 \in \mathrm{Cut}(\mathfrak{s})} \left[\mathbf{P}^{C_2}(s)\right] \otimes \left[\mathbf{R}^{C_2}(s)\right] \\
=& \sum_{C_1 \in \mathrm{Cut}(\mathfrak{t})} \sum_{C_2 \in \mathrm{Cut}(\mathfrak{s})} \left[\mathbf{P}^{C_1}(t)\right] \circ \left[\mathbf{P}^{C_2}(s)\right] \otimes \left[\mathbf{R}^{C_1}(t)\right] \circ \left[\mathbf{R}^{C_2}(s)\right] \\
=& \sum_{C \in \mathrm{Cut}(\mathfrak{t} \circ \mathfrak{s})} \left[\mathbf{P}^C\left(\left(\rho_{\mathfrak{t},\mathfrak{s}} \circ \rho_{\mathfrak{t} \circ \mathfrak{s}}^{-1}\right)(t \otimes s)\right)\right] \otimes \left[\mathbf{R}^C\left(\left(\rho_{\mathfrak{t},\mathfrak{s}} \circ \rho_{\mathfrak{t} \circ \mathfrak{s}}^{-1}\right)(t \otimes s)\right)\right] ,
\end{aligned}$$

which implies that the claim holds for all $\mathbf{t} \in \mathfrak{Tr}(E)$. In the final step we have used the identification

$$\mathrm{Cut}(\mathfrak{t} \circ \mathfrak{s}) = \{C_1 \cup C_2 : C_1 \in \mathrm{Cut}(\mathfrak{t}), C_2 \in \mathrm{Cut}(\mathfrak{s})\} .$$

■

In the next step we introduce the product $\star$ and the coproduct Δ_{GL}, which will turn out to be the duals of Δ_{CK} and $\circ$ respectively.

Definition 4.16 Let E be a vector space, let $\mathfrak{t}, \mathfrak{s} \in \mathfrak{Tr}$ and let $\mathbf{t} \in E^{\otimes \mathfrak{t}}$ and $\mathbf{s} \in E^{\otimes \mathfrak{s}}$ be elementary labelled trees.

(i) For elementary representatives t and s of $\mathbf{t}$ and $\mathbf{s}$ we define

$$\mathbf{t} \star \mathbf{s} := \sum_{d \in \mathrm{Gr}(\mathfrak{t},\mathfrak{s})} \left[t \stackrel{d}{\rightharpoonup} s \right] \in \mathfrak{Tr}(E).$$

(ii) If $\mathbf{t}$ can be decomposed into $\mathbf{t} = \mathbf{t}_1 \circ \cdots \circ \mathbf{t}_k$ as in (4.3) we define

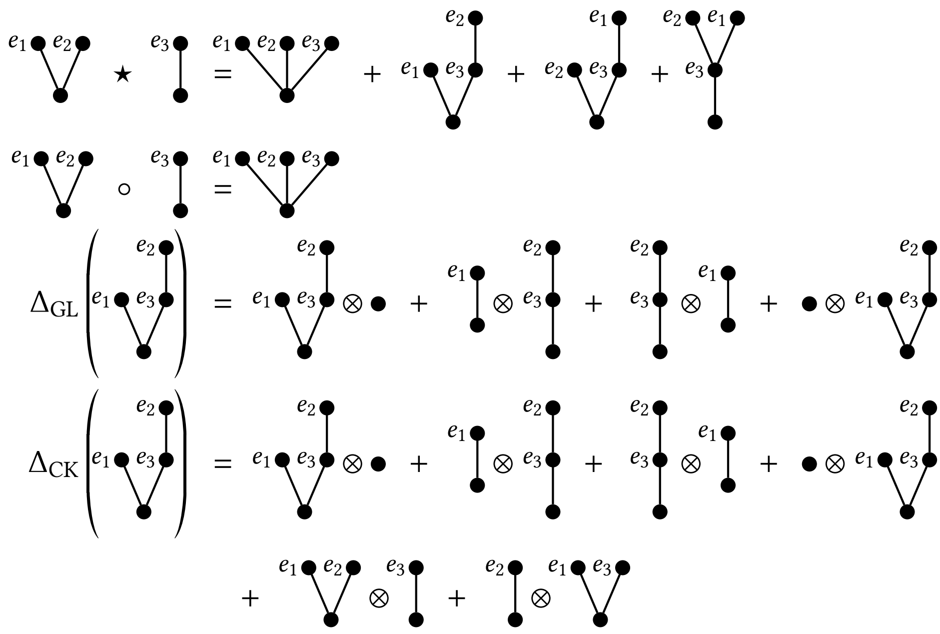
$$\Delta_{\mathrm{GL}}(\mathbf{t}) := \sum_{I \subset \{1,\ldots,k\}} \bigcirc_{i \in I} \mathbf{t}_i \otimes \bigcirc_{i \in I^c} \mathbf{t}_i \in \mathfrak{Tr}(E) \otimes \mathfrak{Tr}(E),$$

where the sum ranges over all subsets of $\{1, \ldots, k\}$ and we set $\bigcirc_{i \in \emptyset} \mathbf{t}_i = \mathbf{1}$.

Both maps are then extended linearly to all $\mathbf{t}, \mathbf{s} \in \mathfrak{Tr}(E)$.

Before we proceed, let us give examples for all four operations that we have introduced.

Example 4.17



The following lemma finally justifies Remark 4.12.

Lemma 4.18 *With the notation as in Definition 4.16,* $\mathbf{t} \star \mathbf{s}$ *does not depend on the choice of* t *and* s*.*

Proof. It is enough to cover the case where $\mathbf{t}$ and $\mathbf{s}$ are elementary labelled trees. Let thus $\tilde{t}$ and $\tilde{s}$ be different elementary representatives of $\mathbf{t}$ and $\mathbf{s}$ respectively. A simple yet crucial observation is that there exist $\sigma_1 \in \mathrm{SG}(\mathfrak{t})$ and $\sigma_2 \in \mathrm{SG}(\mathfrak{s})$ such that we have

$$\left[\tilde{t} \overset{d}{\rightharpoonup} \tilde{s}\right] = \left[\sigma_1(t) \overset{d}{\rightharpoonup} \sigma_2(s)\right] = \left[t \overset{\sigma_2 \circ d \circ \sigma_1}{\rightharpoonup} s\right],$$

where the second equality is a straightforward consequence of Definition 4.11. On the other hand, for fixed $\sigma_1 \in \mathrm{SG}(\mathfrak{t})$ and $\sigma_2 \in \mathrm{SG}(\mathfrak{s})$ we have

$$\mathrm{Gr}(\mathfrak{t}, \mathfrak{s}) = \{\sigma_2 \circ d \circ \sigma_1 : d \in \mathrm{Gr}(\mathfrak{t}, \mathfrak{s})\}.$$

Hence $\mathbf{t} \star \mathbf{s}$ is well-defined as long as the same representatives are chosen for each term in the defining sum. ■

Lemma 4.19 *Let* $\mathfrak{t} \in \mathfrak{Tr}$ *and let* $(E, E', \langle\cdot,\cdot\rangle)$ *be a dual pair of vector spaces as in Definition 2.6. Define*

$$\langle\cdot,\cdot\rangle_{\mathfrak{t}} : E^{\otimes\mathfrak{t}} \times E'^{\otimes\mathfrak{t}} \to \mathbf{R} : (\mathbf{t}_1, \mathbf{t}_2) \mapsto \sum_{\sigma\in\mathrm{SG}(\mathfrak{t})} \langle\sigma(t_1), t_2\rangle = \sum_{\sigma\in\mathrm{SG}(\mathfrak{t})} \langle t_1, \sigma(t_2)\rangle \tag{4.7}$$

where $t_1 \in E^{\otimes|\mathfrak{t}|}$ *and* $t_2 \in E'^{\otimes|\mathfrak{t}|}$ *are representatives of* $\mathbf{t}_1$ *and* $\mathbf{t}_2$ *respectively. Then*

$$\left(E^{\otimes\mathfrak{t}}, E'^{\otimes\mathfrak{t}}, \langle\cdot,\cdot\rangle_{\mathfrak{t}}\right)$$

is a dual pair of vector spaces.

Proof. Let us first show that (4.7) does not depend on the choice of t_1 or t_2. Thus, let $\tilde{t}_1$ be a different representative of $\mathbf{t}_1$. Then we have

$$\sum_{\sigma\in\mathrm{SG}(\mathfrak{t})} \langle\sigma(t_1), t_2\rangle - \sum_{\sigma\in\mathrm{SG}(\mathfrak{t})} \langle\sigma(\tilde{t}_1), t_2\rangle = \left\langle \sum_{\sigma\in\mathrm{SG}(\mathfrak{t})} \sigma(t_1 - \tilde{t}_1), t_2 \right\rangle = 0$$

and similarly for t_2.

Let us now prove that the pairing separates the spaces. Fix $\mathbf{t}_2 \in E'^{\otimes\mathfrak{t}}$ and assume that we have $\langle\mathbf{t}_1, \mathbf{t}_2\rangle_{\mathfrak{t}} = 0$ for all $\mathbf{t}_1 \in E^{\otimes\mathfrak{t}}$. Thus if we fix a representative t_2 of $\mathbf{t}_2$ we have

$$0 = \langle\mathbf{t}_1, \mathbf{t}_2\rangle_{\mathfrak{t}} = \sum_{\sigma\in\mathrm{SG}(\mathfrak{t})} \langle t_1, \sigma(t_2)\rangle = \left\langle t_1, \sum_{\sigma\in\mathrm{SG}(\mathfrak{t})} \sigma(t_2)\right\rangle$$

for all $t_1 \in E^{\otimes|\mathfrak{t}|}$ which implies $\sum_{\sigma\in\mathrm{SG}(\mathfrak{t})}\sigma(t_2) = 0$, i.e. $\mathbf{t}_2 = 0$. This last implication holds because $\left(E^{\otimes|\mathfrak{t}|}, E'^{\otimes|\mathfrak{t}|}, \langle\cdot,\cdot\rangle\right)$ is a dual pair, as we have seen in Remark 2.10. Therefore the maps $\langle \mathbf{t}_1, \cdot\rangle_{\mathfrak{t}}$ with $\mathbf{t}_1 \in E^{\otimes\mathfrak{t}}$ separate $E'^{\otimes\mathfrak{t}}$. The converse statement is shown in exactly the same way and hence the proof is finished. ■

Theorem 4.20 *Let $(E, E', \langle\cdot,\cdot\rangle)$ be a dual pair of vector spaces. Define*

$$\langle\cdot,\cdot\rangle_{\mathfrak{Tr}} : \mathfrak{Tr}(E) \times \mathfrak{Tr}(E') \to \mathbf{R} : (\mathbf{t}_1, \mathbf{t}_2) \mapsto \delta_{\mathfrak{t}_1,\mathfrak{t}_2}\langle \mathbf{t}_1, \mathbf{t}_2\rangle_{\mathfrak{t}_1}$$

for $\mathbf{t}_1 \in E^{\otimes\mathfrak{t}_1}$ and $\mathbf{t}_2 \in E'^{\otimes\mathfrak{t}_2}$. Then

$$\mathcal{H}_{\mathrm{GL}}(E) := (\mathfrak{Tr}(E), \star, \Delta_{\mathrm{GL}})$$

is a bialgebra and $(\mathcal{H}_{\mathrm{GL}}(E), \mathcal{H}_{\mathrm{CK}}(E'), \langle\cdot,\cdot\rangle_{\mathfrak{Tr}})$ is a dual pair of bialgebras.

Proof. We need to show two things. Firstly we have to make sure that $\mathfrak{Tr}(E)$ and $\mathfrak{Tr}(E')$ separate each other. This follows immediately from Lemma 4.19. Secondly we have to show that the pairing satisfies the conditions of Definition 2.7. The claim will then follow from Proposition 2.9.

The duality relation between the algebraic structures has been shown in [Hof03, Prop. 4.4] for the special case $E = \mathbf{R}$, i.e. for unlabelled trees. We will extend their arguments to the labelled case.

Let $\mathfrak{t}_1, \mathfrak{t}_2, \mathfrak{s} \in \mathfrak{Tr}$ be trees and let $\mathbf{t}_1 \in E^{\otimes\mathfrak{t}}$, $\mathbf{t}_2 \in E^{\otimes\mathfrak{t}_2}$ and $\mathbf{s} \in E'^{\otimes\mathfrak{s}}$ be elementary trees with elementary representatives t_1, t_2 and s respectively.

It is an easy exercise to show that both $\mathcal{H}_{\mathrm{GL}}(E)$ and $\mathcal{H}_{\mathrm{CK}}(E)$ are generated as an algebra by the elements of $\mathfrak{Tr}_1(E)$. In light of Remark 2.8 it is therefore sufficient to show

$$\langle \mathbf{t}_1 \star \mathbf{t}_2, \mathbf{s}\rangle_{\mathfrak{Tr}} = \langle \mathbf{t}_1 \otimes \mathbf{t}_2, \Delta_{\mathrm{CK}}(\mathbf{s})\rangle_{\mathfrak{Tr}} \quad \text{and} \quad \langle \mathbf{t}_1 \circ \mathbf{t}_2, \mathbf{s}\rangle_{\mathfrak{Tr}} = \langle \mathbf{t}_1 \otimes \mathbf{t}_2, \Delta_{\mathrm{GL}}(\mathbf{s})\rangle_{\mathfrak{Tr}}$$

for $\mathfrak{t}_1 \in \mathfrak{Tr}_1$. By definition the first equality is equivalent to

$$\begin{aligned} &\sum_{d\in\mathrm{Gr}(\mathfrak{t}_1,\mathfrak{t}_2;\mathfrak{s})}\ \sum_{\sigma\in\mathrm{SG}(\mathfrak{s})} \langle t_1 \overset{d}{\rightharpoonup} t_2, \sigma(s)\rangle \\ &= \sum_{C\in\mathrm{Cut}(\mathfrak{s};\mathfrak{t}_1,\mathfrak{t}_2)}\ \sum_{\sigma_1\in\mathrm{SG}(\mathfrak{t}_1)}\ \sum_{\sigma_2\in\mathrm{SG}(\mathfrak{t}_2)} \langle \sigma_1(t_1) \otimes \sigma_2(t_2), \mathbf{P}^C(s) \otimes \mathbf{R}^C(s)\rangle, \end{aligned} \tag{4.8}$$

where we denote by $\mathrm{Gr}(\mathfrak{t}_1, \mathfrak{t}_2; \mathfrak{s})$ the set of all grafting maps d such that $\mathfrak{t}_1 \overset{d}{\rightharpoonup} \mathfrak{t}_2 = \mathfrak{s}$ and by $\mathrm{Cut}(\mathfrak{s}; \mathfrak{t}_1, \mathfrak{t}_2)$ the set of all cuts C such that $P^C(\mathfrak{s}) = \mathfrak{t}_1$ and $R^C(\mathfrak{s}) = \mathfrak{t}_2$.

Denote by c the unique child of the root of $\mathrm{t}_1 \in \mathfrak{Tr}_1$. Any $d \in \mathrm{Gr}(\mathrm{t}_1, \mathrm{t}_2)$ is then uniquely determined by the vertex of t_2 to which c is attached. Let us now rewrite the left-hand side of (4.8). Instead of summing over all $d \in \mathrm{Gr}(\mathrm{t}_1, \mathrm{t}_2; \mathfrak{s})$ we can alternatively fix an arbitrary $\bar{d} \in \mathrm{Gr}(\mathrm{t}_1, \mathrm{t}_2; \mathfrak{s})$, define $v := \bar{d}(c)$ and rewrite the expression as

$$\frac{1}{|\mathrm{Fix}(v, \mathrm{t}_2)|} \sum_{\pi \in \mathrm{SG}(\mathrm{t}_2)} \sum_{\sigma \in \mathrm{SG}(\mathfrak{s})} \langle t_1 \overset{\bar{d}}{\rightharpoonup} \pi(t_2), \sigma(s) \rangle, \tag{4.9}$$

where $\mathrm{Fix}(v, \mathrm{t}_2) := \{\sigma \in \mathrm{SG}(\mathrm{t}_2) : \sigma(v) = v\}$. In order to verify that (4.9) is indeed equal to the left hand side of (4.8) we observe that

$$\sum_{\sigma \in \mathrm{SG}(\mathfrak{s})} \langle t_1 \overset{\bar{d}}{\rightharpoonup} \pi(t_2), \sigma(s) \rangle = \sum_{\sigma \in \mathrm{SG}(\mathfrak{s})} \langle t_1 \overset{\bar{d}}{\rightharpoonup} \tilde{\pi}(t_2), \sigma(s) \rangle$$

holds for $\pi, \tilde{\pi} \in \mathrm{SG}(\mathrm{t}_2)$ whenever $\pi^{-1}(v) = \tilde{\pi}^{-1}(v)$. For fixed $\pi \in \mathrm{SG}(\mathrm{t}_2)$ there are $|\mathrm{Fix}(v, \mathrm{t}_2)|$ ways to choose $\tilde{\pi}$ in such a way.

On the other hand, any cut $C \in \mathrm{Cut}(\mathfrak{s}; \mathrm{t}_1, \mathrm{t}_2)$ contains exactly one vertex and hence we can choose $\overline{C} \in \mathrm{Cut}(\mathfrak{s}; \mathrm{t}_1, \mathrm{t}_2)$ such that for the canonical injection $\phi_{\overline{C},P} : \mathrm{v}(\mathrm{t}_1) \to \mathrm{v}(\mathfrak{s})$ we have that $w := \phi_{\overline{C},R}(c)$ is the unique element of $\overline{C}$. Thus we can rewrite the right-hand side of (4.8) as

$$\frac{1}{|\mathrm{Fix}(w, \mathfrak{s})|} \sum_{\tau \in \mathrm{SG}(\mathfrak{s})} \sum_{\sigma_1 \in \mathrm{SG}(\mathrm{t}_1)} \sum_{\sigma_2 \in \mathrm{SG}(\mathrm{t}_2)} \left\langle \sigma_1(t_1) \otimes \sigma_2(t_2), \mathbf{P}^{\overline{C}}(\tau(s)) \otimes \mathbf{R}^{\overline{C}}(\tau(s)) \right\rangle, \tag{4.10}$$

where $\mathrm{Fix}(w, \mathfrak{s}) := \{\sigma \in \mathrm{SG}(\mathfrak{s}) : \sigma(w) = w\}$. This can verified as in the previous paragraph. Furthermore, since $\phi_{\overline{C},P}$ induces a canonical embedding of $\mathrm{SG}(\mathrm{t}_1)$ into $\mathrm{SG}(\mathfrak{s})$ as a subgroup we can rewrite (4.10) as

$$\begin{aligned} &\frac{|\mathrm{SG}(\mathrm{t}_1)|}{|\mathrm{Fix}(w, \mathfrak{s})|} \sum_{\tau \in \mathrm{SG}(\mathfrak{s})} \sum_{\sigma_2 \in \mathrm{SG}(\mathrm{t}_2)} \left\langle t_1 \otimes \sigma_2(t_2), \mathbf{P}^{\overline{C}}(\tau(s)) \otimes \mathbf{R}^{\overline{C}}(\tau(s)) \right\rangle \\ =&\frac{|\mathrm{SG}(\mathrm{t}_1)|}{|\mathrm{Fix}(w, \mathfrak{s})|} \sum_{\tau \in \mathrm{SG}(\mathfrak{s})} \sum_{\sigma_2 \in \mathrm{SG}(\mathrm{t}_2)} \left\langle t_1 \overset{\bar{d}}{\rightharpoonup} \sigma_2(t_2), \tau(s) \right\rangle \end{aligned}$$

One verifies that $|\mathrm{Fix}(w, \mathfrak{s})| = |\mathrm{Fix}(v, \mathrm{t}_2)||\mathrm{SG}(\mathrm{t}_1)|$ holds, which shows (4.8).

The second condition in (2.3), i.e. the relationship between $\circ$ and Δ_{GL}, is shown in a similar way, albeit the argument is simpler because $\circ$ and Δ_{GL} are more elementary than $\star$ and Δ_{CK}. ∎

Let us conclude this section with a characterisation of the primitive elements of $\mathcal{H}_{\mathrm{GL}}(E)$.

Proposition 4.21 *An element* $\mathfrak{t} \in \mathcal{H}_{\mathrm{GL}}(E)$ *is primitive if and only if* $\mathfrak{t} \in \mathfrak{Tr}_1(E)$.

Proof. This is obvious from the definition of Δ_{GL}. ■

4.1.3. Algebras of labelled trees over a topological vector space

Let E be a complete locally convex vector space. Our aim is to understand how the topology of E can induce a topology on $\mathfrak{Tr}(E)$. This question can be split into two problems. First we have to introduce a topology on each of the spaces $E^{\otimes \mathfrak{t}}$ for $\mathfrak{t} \in \mathfrak{Tr}$ and second we have to choose an appropriate topology on their direct sum.

Let $\mathfrak{t} \in \mathfrak{Tr}$ be a tree. The space $E^{\otimes \mathfrak{t}}$ is (isomorphic to) a quotient of a tensor power of E and therefore we can use the well-developed theory of topological tensor products that we have already introduced in Chapter 2. Both the projective and the injective seminorms are symmetric in the sense of the following definition.

Definition 4.22 Let E be a locally convex vector space and let $(\xi_i)_{i \in I}$ be a family of seminorms that generates the topology of E. A system of cross seminorms on the tensor powers of E is called *symmetric* if for all $k \in \mathbf{N}$, all $i_1, \ldots, i_k \in I$, all $a \in E^{\otimes k}$ and all $\sigma \in \mathfrak{S}_k$ we have

$$\xi_1 \otimes \cdots \otimes \xi_k(a) = \xi_{\sigma(1)} \otimes \cdots \otimes \xi_{\sigma(k)}(\sigma(a)).$$

From now on we will always assume that the tensor powers of E are equipped with a symmetric system of cross seminorms This assumption implies that for all trees $\mathfrak{t} \in \mathfrak{Tr}$ the projection $\mathrm{sym}_{\mathrm{SG}(\mathfrak{t})}$ defined above is continuous. Therefore it can be uniquely extended to a continuous map on $E^{\widehat{\otimes}|\mathfrak{t}|}$ and we define

$$E^{\widehat{\otimes}\mathfrak{t}} := E^{\widehat{\otimes}|\mathfrak{t}|} \Big/ \ker\left(\mathrm{sym}_{\mathrm{SG}(\mathfrak{t}))}\right) .$$

Equivalently, $E^{\widehat{\otimes}\mathfrak{t}}$ is the completion of $E^{\otimes \mathfrak{t}}$ under the quotient topology. Now we define

$$\widehat{\mathfrak{Tr}}(E) := \mathrm{cl}\left(\bigoplus_{\mathfrak{t} \in \mathfrak{Tr}} E^{\widehat{\otimes}\mathfrak{t}}\right),$$

where the closure is taken with respect to the product topology. Furthermore, for $\mathfrak{t}, \mathfrak{s} \in \mathfrak{Tr}$ we have an isomorphism

$$E^{\otimes \mathfrak{t}} \otimes E^{\otimes \mathfrak{s}} \simeq \left(E^{\otimes|\mathfrak{t}|} \otimes E^{\otimes|\mathfrak{s}|}\right) \Big/ \left(E^{\otimes|\mathfrak{t}|} \otimes \ker\left(\mathrm{sym}_{\mathrm{SG}(\mathfrak{s})}\right) + \ker\left(\mathrm{sym}_{\mathrm{SG}(\mathfrak{t})}\right) \otimes E^{\otimes|\mathfrak{s}|}\right)$$

and therefore the system of cross-seminorms that we have chosen for the $E^{\otimes k}$ gives a locally convex topology on $E^{\otimes \mathfrak{t}} \otimes E^{\otimes \mathfrak{s}}$. Thus, using distributivity of the tensor product and direct sums, we define the completion of $\widehat{\mathfrak{Tr}}(E) \otimes \widehat{\mathfrak{Tr}}(E)$ via

$$\widehat{\mathfrak{Tr}}(E)\widehat{\otimes}\widehat{\mathfrak{Tr}}(E) = \mathrm{cl}\left(\bigoplus_{\mathfrak{t},\mathfrak{s}\in\mathfrak{Tr}} \left(E^{\widehat{\otimes}\mathfrak{t}}\widehat{\otimes}E^{\widehat{\otimes}\mathfrak{s}}\right)\right).$$

Furthermore we extend the operations of $\mathcal{H}_{\mathrm{CK}}$ and $\mathcal{H}_{\mathrm{GL}}$ continuously and we define

$$\widehat{\mathcal{H}}_{\mathrm{CK}} := (\widehat{\mathfrak{Tr}}(E), \circ, \Delta_{\mathrm{CK}}) \quad \text{and} \quad \widehat{\mathcal{H}}_{\mathrm{GL}} := (\widehat{\mathfrak{Tr}}(E), \star, \Delta_{\mathrm{GL}}).$$

Proposition 4.23 *Let E be a locally convex vector space and let E' be its continuous dual. Then $(\widehat{\mathcal{H}}_{\mathrm{GL}}(E), \mathcal{H}_{\mathrm{CK}}(E'), \langle\cdot,\cdot\rangle_{\mathfrak{Tr}})$ is a dual pair of bialgebras.*

Proof. This follows from Theorem 4.20 in combination with the arguments that we have used in the proof of Proposition 2.18. ■

Corollary 4.24 *Let E be a locally convex vector space and let $\mathbf{t} \in \widehat{\mathcal{H}}_{\mathrm{GL}}(E)$ be group-like i.e. such that it satisfies $\Delta_{\mathrm{GL}}(\mathbf{t}) = \mathbf{t} \otimes \mathbf{t}$. Then we have*

$$\mathbf{t} = \exp_{\circ}(\pi_1(\mathbf{t})) := \sum_{k=0}^{\infty} \frac{(\pi_1(\mathbf{t}))^{\circ k}}{k!},$$

where $\pi_1 : \mathfrak{Tr}(E) \to \mathfrak{Tr}_1(E)$ is the canonical projection as before.

Proof. Let E' be the continuous dual of E and let $\mathbf{s} \in \mathfrak{Tr}(E')$ be an elementary labelled tree with decomposition $\mathbf{s} = \mathbf{s}_1 \circ \cdots \circ \mathbf{s}_k$ as in (4.3). Then we have

$$\begin{aligned}
\langle \mathbf{t}, \mathbf{s}\rangle_{\mathfrak{Tr}} &= \langle \mathbf{t}, \mathbf{s}_1 \circ \cdots \circ \mathbf{s}_k\rangle_{\mathfrak{Tr}} \\
&= \langle \mathbf{t}, \mathbf{s}_1\rangle_{\mathfrak{Tr}} \cdots \langle \mathbf{t}, \mathbf{s}_k\rangle_{\mathfrak{Tr}} \\
&= \langle \pi_1(\mathbf{t}), \mathbf{s}_1\rangle_{\mathfrak{Tr}} \cdots \langle \pi_1(\mathbf{t}), \mathbf{s}_k\rangle_{\mathfrak{Tr}},
\end{aligned}$$

where we have first used Proposition 2.20 in combination with Proposition 4.23 and then the fact that $\mathbf{s}_i \in \mathfrak{Tr}_1(E)$ for all $i \in \{1, \ldots, k\}$. On the other hand we have

$$\begin{aligned}
\left\langle (\pi_1(\mathbf{t}))^{\circ k}, \mathbf{s}\right\rangle_{\mathfrak{Tr}} &= \left\langle (\pi_1(\mathbf{t}))^{\circ k}, \mathbf{s}_1 \circ \cdots \circ \mathbf{s}_k\right\rangle_{\mathfrak{Tr}} \\
&= k! \langle \pi_1(\mathbf{t}), \mathbf{s}_1\rangle_{\mathfrak{Tr}} \cdots \langle \pi_1(\mathbf{t}), \mathbf{s}_k\rangle_{\mathfrak{Tr}},
\end{aligned}$$

which follows from (4.7) and the multinomial formula.

Since we have chosen an arbitrary $\mathbf{s} \in \mathfrak{Tr}(E')$ and since $\mathfrak{Tr}(E')$ separates the points of $\widehat{\mathfrak{Tr}}(E)$ this shows that we have $\pi_k(\mathbf{t}) = \frac{1}{k!}(\pi_1(\mathbf{t}))^{\circ k}$ for all $k \in \mathbf{N}$. Hence the claim follows. ■

Remark 4.25 Corollary 4.24 tells us that we can reconstruct any group-like element in $\mathcal{H}_{\mathrm{GL}}(E)$ easily from its projection to $\mathfrak{Tr}_1(E)$ by only using the $\circ$-product. Note that even though $\pi_1(\mathbf{t})$ is a primitive element for all $\mathbf{t} \in \mathcal{H}_{\mathrm{GL}}(E)$, it is in general not equal to $\log_\star(\mathbf{t})$ if $\mathbf{t}$ is group-like. In particular, it is important to not confuse Corollary 4.24 with the trivial statement

$$\mathbf{t} = \exp_\star(\log_\star(\mathbf{t}))$$

for a group-like element $\mathbf{t}$.

4.1.4. The truncated tree algebra

Let us finally consider truncations of the space $\mathfrak{Tr}(E)$. For $k \in \mathbf{N}$ define the subspace

$$I^{(k)} := \bigoplus_{\mathsf{t}\in\mathfrak{Tr}:|\mathsf{t}|>k} E^{\otimes \mathsf{t}}$$

and the quotient space

$$\mathfrak{Tr}^{(k)}(E) := \mathfrak{Tr}(E) \Big/ I^{(k)} .$$

The canonical projection $\mathfrak{Tr}(E) \to \mathfrak{Tr}^{(k)}(E)$ is denoted by $\pi^{(k)}$. We will furthermore need the projection $\pi_n^{(k)} := \pi^{(k)} \circ \pi_n = \pi_n \circ \pi^{(k)}$. Observe that $I^{(k)}$ is an ideal for both $\star$ and $\circ$. Hence $\mathfrak{Tr}^{(k)}$ can be interpreted as a quotient algebra of both $(\mathfrak{Tr}(E), \star)$ and $(\mathfrak{Tr}(E), \circ)$. On the other hand $I^{(k)}$ is a coideal for neither Δ_{GK} nor Δ_{CL} and hence the quotient is not a quotient bialgebra. Furthermore $\mathfrak{Tr}^{(k)}(E)$ is isomorphic (as a vector space) to the closed subspace

$$\bigoplus_{\mathsf{t}\in\mathfrak{Tr}:|\mathsf{t}|\le k} E^{\otimes \mathsf{t}}$$

of $\mathfrak{Tr}(E)$, which justifies the name *truncated tree algebra*. We will use this isomorphism implicitly without mentioning it. Note however that this vector space isomorphism is not an algebra morphism since $\bigoplus_{\mathsf{t}\in\mathfrak{Tr}:|\mathsf{t}|\le k} E^{\otimes \mathsf{t}}$ is not even a subalgebra of $\mathfrak{Tr}(E)$.

The completion $\widehat{G}^{(n)}_{\mathrm{GL}}(E)$ of the group

$$G^{(n)}_{\mathrm{GL}}(E) := \exp_\star\left(\mathfrak{Tr}_1^{(n)}(E)\right) \subset \mathfrak{Tr}^{(n)}(E)$$

will serve as the space in which branched rough paths take their values in the next section.

In analogy to Proposition 2.24 it is easy to see that an element $\mathbf{t} \in \widehat{\mathcal{H}}_{\mathrm{GL}}(E)$ is group-like if and only if we have $\pi^{(n)}(\mathbf{t}) \in \widehat{G}^{(n)}_{\mathrm{GL}}(E)$ for all $n \in \mathbf{N}$.

Proposition 4.26 *Let E be a locally convex vector space and let $\mathbf{t} \in \widehat{\mathcal{H}}_{\mathrm{GL}}(E)$ be group-like. Then we have*

$$\pi_k^{(n)}(\mathbf{t}) = \frac{\left(\pi_1^{(n)}(\mathbf{t})\right)^{\circ k}}{k!} + \mathbf{r},$$

where $\mathbf{r}$ is such that $\pi^{(n)}(\mathbf{r}) = 0$.

Proof. This follows immediately from Corollary 4.24. ∎

4.2. Differential equations driven by branched rough paths

Everything that we need to introduce branched rough paths is now set up. We begin with the definition. We will discuss the relation between Definitions 3.2, 3.4 and 4.27 at the end of this chapter. Throughout this section we assume that E is a Banach space, which implies that $\widehat{\mathfrak{Tr}}^{(k)}(E)$ is also Banach for all $k \in \mathbf{N}$. Furthermore, we assume that the projective tensor product is chosen in the construction of $\widehat{\mathfrak{Tr}}(E)$.

Definition 4.27 Let $T > 0$, let $p \geq 1$ and let ω be a control function. A continuous map $\mathbf{x} : \Delta_T \to \widehat{G}^{([p])}_{\mathrm{GL}}(E)$ is called a *branched p-rough path controlled by ω* if it satisfies

(i) $\mathbf{x}_{s,t} \star \mathbf{x}_{t,u} = \mathbf{x}_{s,u}$ for all $0 \leq s \leq t \leq u \leq T$ and

(ii) $\left\|\pi^k(\mathbf{x}_{s,t})\right\| \leq C\omega(s,t)^{\frac{k}{p}}$ for some $C > 0$ and all $(s,t) \in \Delta_T$ and all $k \in \{1, \dots, [p]\}$.

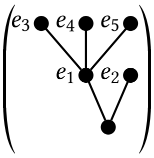
Let F be another Banach space and let $V : E \times F \to F$ be a continuous map which is linear in the first argument and smooth (in the Fréchet sense) in the second argument. Let furthermore $f : F \to G$ be a smooth map into some Banach space G. Then for $y \in F$ and fixed $k \in \mathbf{N}$ we construct a linear map

$$\Psi_{V,f,y} : \widehat{\mathfrak{Tr}}^{(k)}(E) \to G$$

in the following way. It maps an elementary tree $\mathbf{t}$ to a product of derivatives of f and V, evaluated at y. The root of $\mathbf{t}$ corresponds to f and a vertex labelled by $e \in E$ corresponds to the vector field V_e. The number of children of a vertex indicates how often the corresponding term has to be differentiated. The terms corresponding to the children are then plugged into the derivative.

Example 4.28

$$\Psi_{V,f,y}\left(\right) = f''(y)[V'''_{e_1}(y)[V_{e_3}(y), V_{e_4}(y), V_{e_5}(y)], V_{e_2}(y)]$$



The generalisation of $\Psi_{V,f,y}$ from elementary labelled trees to general elements of $\widehat{\mathfrak{Tr}}^{(k)}(E)$ requires some care and it is the reason for why we have insisted on using the projective tensor product at the beginning of this section. In fact, what we have defined so far is essentially a bounded multilinear map from $E \times \cdots \times E$ to G. Therefore the universal property of the projective tensor norm from Proposition 2.13 allows us to extend this map to a bounded linear map on $E^{\hat{\otimes}\mathfrak{t}}$ for $\mathfrak{t} \in \mathfrak{Tr}$ and hence to a bounded linear map on $\widehat{\mathfrak{Tr}}^{(k)}(E)$.

Remark 4.29 Note that the map $\Psi_{V,f,y}$ is well-defined even though we work with non-planar trees since higher order derivatives of of the V and f are symmetric multilinear forms.

There is a particular choice of f which will play a crucial role in the next section. For any positive integer k we define the map

$$\mathrm{id}_F^k : F \to F^{\otimes k} : x \mapsto x^{\otimes k}.$$

The maps $\Psi_{V,\mathrm{id}_F^k,y}$ have two important properties. First they satisfy

$$\Psi_{V,f,y}(\mathbf{t}) = \frac{1}{k!} f^{(k)}(y)[\Psi_{V,\mathrm{id}_F^k,y}(\mathbf{t})] \tag{4.11}$$

for all $\mathbf{t} \in \widehat{\mathfrak{Tr}}_k(E)$. Second they are compatible with the $\circ$-product of labelled trees in the sense that for two trees $\mathbf{t}_1 \in \widehat{\mathfrak{Tr}}_k(E)$ and $\mathbf{t}_2 \in \widehat{\mathfrak{Tr}}_l(E)$ we have

$$\Psi_{V,\mathrm{id}_F^{k+l},y}(\mathbf{t}_1 \circ \mathbf{t}_2) = \frac{(k+l)!}{k!l!} \operatorname{sym}_{\mathfrak{S}_{k+l}} \left(\Psi_{V,\mathrm{id}_F^k,y}(\mathbf{t}_1) \otimes \Psi_{V,\mathrm{id}_F^l,y}(\mathbf{t}_2) \right). \tag{4.12}$$

Definition 4.30 Let $\mathbf{x}$ be a branched p-rough path controlled by ω. We say that a continuous path $y : [0,T] \to F$ solves the equation

$$\mathrm{d}y_t = V(y_t)\mathrm{d}\mathbf{x}_t \tag{RDEb}$$

if there exists a function $\theta : \mathbf{R} \to \mathbf{R}$ with $\theta(h)/h \to 0$ for $h \to 0$ such that

$$\|y_t - \Psi_{V,\mathrm{id}_F,y_s}(\mathbf{x}_{s,t})\| \le \theta(\omega(s,t))$$

holds for all $(s,t) \in \Delta_T$.

4.2.1. Change of variables for rough differential equations

Theorem 4.31 *Let $V : E \times F \to F$ be a continuous map which is linear in the first argument and smooth in the second argument. Let furthermore $\mathbf{x}$ be a branched p-rough path controlled by ω and let $y : [0,T] \to F$ be a continuous path. Then the following are equivalent.*

(i) The path y solves (RDEb) *in the sense of Definition 4.30.*

(ii) For every Banach space G and every smooth $f : F \to G$ there exists a function $\theta_f : \mathbf{R} \to \mathbf{R}$ with $\theta_f(h)/h \to 0$ for $h \to 0$ such that

$$\|f(y_t) - \Psi_{V,f,y_s}(\mathbf{x}_{s,t})\| \le \theta_f(\omega(s,t)) \tag{4.13}$$

holds for all $(s,t) \in \Delta_T$.

Proof. The implication $(ii) \Rightarrow (i)$ is trivial, simply choose $f = \mathrm{id}_F$. For the other direction, the idea is to split the term $\Psi_{V,f,y_s}(\mathbf{x}_{s,t})$ into terms that we can control. Using the shorthand $y_{s,t} := y_t - y_s$ we have (see explanations below)

$$\begin{aligned}
&\left\|f(y_t) - \Psi_{V,f,y_s}(\mathbf{x}_{s,t})\right\| \\
&\le \left\|\sum_{k=1}^{[p]} \frac{f^{(k)}(y_s)}{k!}\left[y_{s,t}^{\otimes k} - \Psi_{V,\mathrm{id}_F^k,y_s}(\pi_k(\mathbf{x}_{s,t}))\right]\right\| + C_1\|y_{s,t}\|^{[p]+1} &\text{(4.14a)}\\
&\le \left\|\sum_{k=1}^{[p]} \frac{f^{(k)}(y_s)}{k!}\left[y_{s,t}^{\otimes k} - \frac{1}{k!}\Psi_{V,\mathrm{id}_F^k,y_s}\left((\pi_1(\mathbf{x}_{s,t}))^{\circ k}\right)\right]\right\| + C_1\|y_{s,t}\|^{[p]+1} + C_2\omega(s,t)^{\frac{[p]+1}{p}} &\text{(4.14b)}\\
&= \left\|\sum_{k=1}^{[p]} \frac{f^{(k)}(y_s)}{k!}\left[y_{s,t}^{\otimes k} - \left(\Psi_{V,\mathrm{id}_F,y_s}\left(\pi_1(\mathbf{x}_{s,t})\right)\right)^{\otimes k}\right]\right\| + C_1\|y_{s,t}\|^{[p]+1} + C_2\omega(s,t)^{\frac{[p]+1}{p}} &\text{(4.14c)}\\
&= \left\|\sum_{k=1}^{[p]} \frac{f^{(k)}(y_s)}{k!}\left[\left(y_{s,t} - \Psi_{V,\mathrm{id}_F,y_s}\left(\pi_1(\mathbf{x}_{s,t})\right)\right) \otimes \sum_{i=0}^{k-1} y_{s,t}^{\otimes(k-1-i)}\Psi_{V,\mathrm{id}_F,y_s}\left(\pi_1(\mathbf{x}_{s,t})\right)^{\otimes i}\right]\right\| \\
&\qquad + C_1\|y_{s,t}\|^{[p]+1} + C_2\omega(s,t)^{\frac{[p]+1}{p}} &\text{(4.14d)}\\
&= \left\|\sum_{k=1}^{[p]} \frac{f^{(k)}(y_s)}{k!}\left[\left(y_t - \Psi_{V,\mathrm{id}_F,y_s}\left(\mathbf{x}_{s,t}\right)\right) \otimes \sum_{i=0}^{k-1} y_{s,t}^{\otimes(k-1-i)}\Psi_{V,\mathrm{id}_F,y_s}\left(\pi_1(\mathbf{x}_{s,t})\right)^{\otimes i}\right]\right\| \\
&\qquad + C_1\|y_{s,t}\|^{[p]+1} + C_2\omega(s,t)^{\frac{[p]+1}{p}}. &\text{(4.14e)}
\end{aligned}$$

In this computation we have used the following arguments.

(4.14a) Replace f by its Taylor approximation up to degree $[p]$ and make use of (4.11).

(4.14b) Apply Proposition 4.26, where the remainder $\mathbf{r}$ leads to the term that involves ω.

(4.14c) Use (4.12), where we can ignore the symmetrisation because $f^{(k)}(y_s)$ is a symmetric multilinear form anyway.

(4.14d) Use the elemetary identity $a^k - b^k = (a-b)(a^{k-1} + a^{k-2}b + \cdots + ab^{k-2} + b^{k-1})$, which holds in principle only for commuting variables a and b. It can be applied nevertheless, again because $f^{(k)}(y_s)$ is symmetric.

(4.14e) Observe that Ψ_{V,id,y_s} maps any tree which is neither in $\widehat{\mathfrak{Tr}}_1(E)$ nor in $E^{\otimes 1}$ to zero.

Now define the constant C_3 by

$$C_3 := \sup_{0\le s<t\le T} \sup_{a\in F:\|a\|=1} \left\| \sum_{k=1}^{[p]} \frac{f^{(k)}(y_s)}{k!} \left[a \otimes \sum_{i=0}^{k-1} y_{s,t}^{\otimes(k-1-i)} \Psi_{V,\mathrm{id}_F,y_s} \left(\pi_1(\mathbf{x}_{s,t})\right)^{\otimes i} \right] \right\|.$$

We have $C_3 < \infty$ because $f^{(k)}(y)$ is a continuous multilinear map for all $k \in \mathbf{N}$ and all $y \in F$. Thus (4.14e) can be estimated by

$$\begin{aligned} &C_3\|y_t - \Psi_{V,\mathrm{id}_F,y_s}\left(\mathbf{x}_{s,t}\right)\| + C_1\|y_{s,t}\|^{[p]+1} + C_2\omega(s,t)^{\frac{[p]+1}{p}} \\ \le& C_3\theta(\omega(s,t)) + C_4\omega(s,t)^{\frac{[p]+1}{p}}, \end{aligned}$$

for some $C_4 > 0$, where we have also used that y has finite p-variation controlled by (a scalar multiple of) ω. Defining

$$\theta_f(h) := C_3\theta(h) + C_4 h^{\frac{[p]+1}{p}}$$

finishes the proof. ∎

Remark 4.32 It has been pointed out in [Bai18] that there are alternative ways to obtain Theorem 4.31. However, we believe that our proof is the most elementary and direct way way to proceed. This becomes appearent by considering the case of a path x with bounded p-variation for $p \in [1,2)$. Then all of the combinatorial aspects vanish and (4.14) boils down to the following simple application of Taylor's theorem.

$$f(y_t) - \Psi_{V,f,y_s}(x_{s,t})$$

$$
\begin{aligned}
&= f(y_t) - f(y_s) - V_{x_{s,t}} f(y_s) \\
&= f'(y_s)(y_t - y_s) + o(|y_t - y_s|) - f'(y_s) V_{x_{s,t}}(y_s) \\
&= f'(y_s)\left(y_t - y_s - V_{x_{s,t}}(y_s)\right) + o(|y_t - y_s|),
\end{aligned}
$$

This shows the desired equivalence immediately.

4.2.2. Rough differential equations and differential geometry

Even though we have formulated Theorem 4.31 only for branched rough paths it is also valid for geometric rough path since the latter can be interpreted as branched rough paths in a canonical way. The relationship between branched and geometric rough paths has been discussed in detail in [HK15], at least for the finite-dimensional case. The crucial difference between branched and geometric rough paths is that geometric rough paths take their values in the tensor algebra over E rather than the Grossman-Larson algebra.

In order to embed the space of geometric rough paths into the space of branched rough paths, Define $\mathfrak{t} = \overset{\bullet}{\underset{\bullet}{|}}$ and note that $E^{\otimes \mathfrak{t}} \simeq E$. Hence we can define the continuous algebra morphism

$$
\Phi : \widehat{T}(E) \to \widehat{\mathcal{H}}_{\mathrm{GL}}(E)
$$

which is uniquely determined by $\Phi(e) = e \in E^{\otimes \mathfrak{t}}$ for $e \in E$.

Proposition 4.33 *The morphism Φ preserves the bialgebra structure.*

Proof. By construction of Φ, it satisfies $\Phi(ab) = \Phi(a) \star \Phi(b)$ for all $a, b \in \widehat{T}(E)$. It remains to show that we have $\Phi(\Delta_{\sqcup\!\sqcup}(a)) = \Delta_{\mathrm{GL}}(\Phi(a))$ for all $a \in \widehat{T}(E)$. It sufficies to show that this holds for $a \in E$ since the algebra generated by E is a dense subset of $\widehat{T}(E)$. However, all elements of $E \subset \widehat{T}(E)$ are primitive and thus the proof is finished since the elements of $\Phi(E) \subset \widehat{\mathcal{H}}_{\mathrm{GL}}(E)$ are also primitive, i.e. we have

$$
\Phi(\Delta_{\sqcup\!\sqcup}(a)) = \Phi(a) \otimes \mathbf{1} + \mathbf{1} + \Phi(a) = \Delta_{\mathrm{GL}}(\Phi(a))
$$

for all $a \in E$. ∎

Thus Φ lets us embed the space of geometric rough paths into the space of branched rough paths. In [HK15] it has in fact been shown that conversely, one can interpret every branched rough path as a geometric one, albeit this geometric rough path will then take values in the tensor algebra over an *augmented* version of the original vector space.

Let us now see what the 'non-geometricity' of general branched rough paths is about. Let $e_1, \ldots, e_k \in E$. Then we have

$$\Psi_{V,f,y}\left(\Phi(e_1 \otimes \cdots \otimes e_k)\right) = \Psi_{V,f,y}\left({}^{e_1}\mathbf{I} \star \cdots \star {}^{e_k}\mathbf{I}\right) = V_{e_1} \cdots V_{e_k} f(y),$$

because the $\star$-product simply encodes the product rule. Thus we see that for a geometric rough path $\mathbf{x}$ the object $\Psi_{V,f,y}(\mathbf{x}_{s,t})$ is given by vector fields acting on a function. An important implication from a geometric point of view is that $\Psi_{V,f,y}$ can be evaluated using arbitrary coordinates in F without changing its value. In particular, if we work with geometric rough paths, we can easily replace F by a smooth manifold $\mathcal{M}$ without changing anything.

In contrast, if $\mathbf{x}$ is a branched rough path, then $\Psi_{V,f,y}(\mathbf{x}_{s,t})$ in general involves derivatives of vector fields and higher order derivatives of f, which are delicate objects when it comes to a change of coordinates. To wit, if we naively compute the value of $\Psi_{V,f,y}(\mathbf{x}_{s,t})$ with respect to two different coordinate systems we will in general get two different results. To get the correct result, we have to express the canonical connection on F in the new coordinates and then compute the derivatives of V and f that appear in $\Psi_{V,f,y}(\mathbf{x}_{s,t})$ as covariant derivatives with respect to this connection. The following corollary of Theorem 4.31 captures this observation in a rigorous way. We write $\Psi_{V,f,y,\nabla}(\mathbf{x}_{s,t})$ instead of $\Psi_{V,f,y}(\mathbf{x}_{s,t})$ if it is computed with respect to some connection ∇ on F which is different from the canonical one.

Corollary 4.34 *Let the notation be as in Theorem 4.31 and assume that $y_t \in U$ for some open set $U \subset F$ and all $t \in [0,T]$. Let $W \subset F$ be another open set and let $\varphi : U \to W$ be a diffeomorphism. For each $e \in E$ define the vector field $\tilde{V}_e$ on W by*

$$\tilde{V}_e(p) = \varphi'(\varphi^{-1}(p))[V_e(\varphi^{-1}(p))].$$

Then y solves (RDEb) in the sense of Definition 4.30 if and only if $\tilde{y}$ given by $\tilde{y}_t := \varphi(y_t)$ solves

$$\mathrm{d}\tilde{y}_t = \tilde{V}(\tilde{y}_t)\mathrm{d}\mathbf{x}_t,$$

where $\Psi_{V,f,y}(\mathbf{x}_{s,t})$ in Definition 4.30 is replaced by $\Psi_{V,f,y,\tilde{\nabla}}(\mathbf{x}_{s,t})$ and the push-forward connection $\tilde{\nabla}$ is given by

$$\tilde{\nabla}_{\tilde{Y}}\tilde{X}(\tilde{p}) := \tilde{X}'(\tilde{p})[\tilde{Y}(\tilde{p})] - \varphi''(\varphi^{-1}(\tilde{p}))[(\varphi'(\varphi^{-1}(\tilde{p})))^{-1}[\tilde{X}(\tilde{p})], (\varphi'(\varphi^{-1}(\tilde{p})))^{-1}[\tilde{Y}(\tilde{p})]]$$

for vector fields $\tilde{X}, \tilde{Y}$ on F and $\tilde{p} \in W$.

Proof. Writing φ itself in the new coordinates on W gives $\tilde{\varphi}(\tilde{p}) = \varphi(\varphi^{-1}(\tilde{p})) = \tilde{p}$, i.e. $\tilde{\varphi} = \mathrm{id}_W$.

Thus we have

$$\tilde{y}_t - \Psi_{\tilde{V},\mathrm{id}_W,\tilde{y}_s,\tilde{\nabla}}(\mathbf{x}_{s,t}) = \varphi(y_t) - \Psi_{\tilde{V},\tilde{\varphi},\tilde{y}_s,\tilde{\nabla}}(\mathbf{x}_{s,t}) = \varphi(y_t) - \Psi_{V,\varphi,y_s}(\mathbf{x}_{s,t}),$$

where the last equality is obtained by writing everything in the original coordinates. The claim follows now immediately from Theorem 4.31. ■

Remark 4.35 For an arbitrary connection ∇, the expression $\Psi_{V,f,y,\nabla}(\mathbf{x}_{s,t})$ is not well-defined. To see this, consider Remark 4.29. What we have stated there is not true for arbitrary connections. In fact, the arguments of higher order covariant derivatives can be permuted without changing its value if and only if the connection has vanishing curvature. Thus this is not a problem for Corollary 4.34, since it only deals with the push-forward of the canonical, i.e. flat, connection which is also flat. In general, however, this observation reveals that branched rough paths as we have defined them (and as they are defined in most of the existing literature) are unsuited to drive rough differential equations on manifolds with arbitrary connections. To resolve this problem, one has to replace the non-planar trees by planar ones, i.e. one has to introduce an order on the children of each vertex. The relevant algebraic structure is developed in [MKW08]. Furthermore, the connection between planar trees and branched rough paths has recently been studied in [CEFMMK18], albeit in the very specific context of homogeneous spaces. See also [Foi02a, Foi02b] for more details on planar trees. All those ideas fit very well into our framework of tree algebras over vector spaces. Working with planar trees simplifies our construction since there is no need to factor out the symmetry groups of the trees in this case. It might be an interesting problem for future research to work these ideas out in detail.

Hence we see that it is the independence of a choice of a connection which makes the geometric rough paths special among all branched rough paths. Differential equations driven by geometric rough paths can therefore be interpreted as generalised Stratonovich equations whereas differential equations driven by branched rough paths should be seen as a generalisation of Itō-corrected Stratonovich equations. In order to justify this interpretation, let us consider the example of Brownian motion in $\mathbf{R}^d$.

Example 4.36 Let X be a standard Brownian motion in $\mathbf{R}^d$ and denote by $\mathbf{X}^{\mathrm{S}}$ its Stratonovich lift and by $\mathbf{X}^{\mathrm{I}}$ its Itō lift to a rough path as in Examples 3.8 and 3.9. For $(s,t) \in \Delta_T$ we have

$$\log_\otimes(\mathbf{X}^{\mathrm{I}}_{s,t}) = \log_\otimes(\mathbf{X}^{\mathrm{S}}_{s,t}) + (0, 0, -\frac{t-s}{2}\sum_{i=1}^{d} e_i \otimes e_i).$$

Thus if we consider the embedding of $\mathbf{X}^{\mathrm{I}}_{s,t}$ into $\mathcal{H}_{\mathrm{GL}}$ via the morphism Φ we obtain

$$\log_\star\left(\Phi\left(\mathbf{X}^{\mathrm{I}}_{s,t}\right)\right) = \log_\star\left(\Phi\left(\mathbf{X}^{\mathrm{S}}_{s,t}\right)\right) - \frac{t-s}{2}\sum_{i=1}^{d}\left(\begin{smallmatrix} {\scriptstyle e_i}\bullet & & {\scriptstyle e_i}\bullet \\ & \bigvee & \\ & \bullet & \end{smallmatrix} + \begin{smallmatrix} {\scriptstyle e_i}\bullet \\ | \\ {\scriptstyle e_i}\bullet \\ | \\ \bullet \end{smallmatrix}\right).$$

Since $\log_\star(\Phi(\mathbf{X}^{\mathrm{S}}_{s,t}))$ is primitive, $\Phi(\mathbf{X}^{\mathrm{I}})$ would be a branched rough path if it was not for the $\bigvee$-terms. Let us thus see what happens if we simply ignore those terms, i.e. if we consider the branched rough path $\overline{\mathbf{X}}^{\mathrm{I}}$ given by

$$\overline{\mathbf{X}}^{\mathrm{I}}_{s,t} := \Phi\left(\mathbf{X}^{\mathrm{I}}_{s,t}\right) + \frac{t-s}{2}\sum_{i=1}^{d}\begin{smallmatrix} {\scriptstyle e_i}\bullet & & {\scriptstyle e_i}\bullet \\ & \bigvee & \\ & \bullet & \end{smallmatrix} = \Phi\left(\mathbf{X}^{\mathrm{S}}_{s,t}\right) - \frac{t-s}{2}\sum_{i=1}^{d}\begin{smallmatrix} {\scriptstyle e_i}\bullet \\ | \\ {\scriptstyle e_i}\bullet \\ | \\ \bullet \end{smallmatrix}$$

One can check that this alteration does not interfer with conditions (i) and (ii) of Definition 4.27 and thus $\overline{\mathbf{X}}$ is indeed a branched rough path. Let $V = (V_1, \ldots, V_d)$ be a collection of smooth vector fields in some vector space F. The crucial observation is that Definition 3.14 applied to V and $\mathbf{X}^{\mathrm{I}}$ is equivalent to Definition 4.30 when applied to V and $\overline{\mathbf{X}}^{\mathrm{I}}$ because the $\bigvee$ -terms correspond to second-order derivatives of id_F and thus their contribution vanishes. Thus it follows from [FH14, Thm. 9.1] that the solution of a rough differential equation driven by $\overline{\mathbf{X}}^{\mathrm{I}}$ is precisely the solution of the corresponding Itō equation. As a conclusion, branched rough paths allow us to treat Itō equations in a coordinate-invariant way and hence they generalise the concept of Itō-corrected Stratonovich equations.

Remark 4.37 The above transformation of a general rough path into a branched rough path can be performed for any p-rough path with $p < 3$, see [FH14, Exercise 2.14]. This leads to the impression that for $p < 3$, there is not much of a difference between geometric and non-geometric rough paths since Definition 3.14 leads to the "correct" solution in any case, a point of view that has been advocated in [LQ02, FH14]. In a geometric context this point of view requires some care since Definition 3.14 effectively disguises the dependence of the solution on the chosen connection.

As a matter of fact the case $p < 3$ is also special in view of Remark 4.35 since trees of degree 2 or less are too small to reveal the difference between planar and non-planar trees. Indeed, the only tree of degree 2 which is actually branched is $\bigvee$ and it is only branched at the root, thus causing no problems because the second covariant derivative of a *function* is symmetric, whether the space is curved or not.

A careful reading of [CDL15] shows that the authors have used arguments similar to those that went into the proof of Theorem 4.31 in order to obtain well-definedness of RDEs on submanifolds of $\mathbf{R}^k$, albeit in the less general case $p < 3$.

Furthermore, Theorem 4.31 yields a way to *define* what we mean by the solution of an RDE on a manifold by testing (4.13) for a suitable set of real-valued test functions. If the manifold is finite-dimensional, then smooth compactly supported test functions are an obvious choice, see Chapter 5 for details. This definition has been suggested before, see e.g. [ABT15] and the references therein.

5. Rough differential equations on manifolds

In this chapter we will use the results from the previous chapter to introduce a well-defined notion of rough differential equations on manifolds driven by vector space valued rough paths. In contrast to the other chapters, we will only deal with finite-dimensional situations. Hence the space E, which serves as the underlying space for the driving signal, will be replaced by $\mathbf{R}^d$ and the space F, in which the solutions take their values, will be replace by a (finite-dimensional) smooth manifold $\mathcal{M}$. Our main concern is the existence of global solutions.

Our main result on the global existence of solutions relies on the fact that solutions of rough differential equations can be approximated by solutions of ordinary differential equations via Trotter-type approximations as in [Bai15]. This approach should also be seen in the context of earlier works such as [Str87]. The key to our result are precise representations of the approximation errors which will then lead to sharp estimates. Therefore we will first discuss ordinary differential equations and their flows in order to get the desired error representations.

5.1. Ordinary differential equations

In the context of rough differential equation, Lipschitz continuity is a key concept. Therefore we begin this section with a precise definition of Lipschitz continuity for vector fields on manifolds and we establish the relationship between Lipschitz continuity and completeness in this context.

Since our plan is to show the existence of global solutions of certain rough differential equations by approximating them with flows of ordinary differential equations, we need good global estimates on the approximation error. This is the content of the second part of this section.

5.1.1. Lipschitz continuity of vector fields

Definition 5.1 Let $V = (V_1, \dots, V_d)$ be a collection of smooth vector fields on $\mathcal{M}$ and let g be a complete Riemannian metric on $\mathcal{M}$. We say that V is *Lipschitz continuous with respect to g* if the operator norm $\|\nabla V_i\|_g$ is bounded for all $i \in \{1, \dots, d\}$. In this case we define

$$\|V\|_{g-\mathrm{Lip}} := \max_{i\in\{1,\dots,d\}} \sup_{y\in\mathcal{M}} \|\nabla V_i(y)\|_g.$$

In the context of rough differential equations, many estimates become considerably easier when the vector fields are bounded. Originally, it has been advocated in [Ste70] to include boundedness into the definition of Lipschitz continuity in a different context. If we work with respect to the canonical metric on $\mathbf{R}^n$ this is an additional constraint since Lipschitz continuity with respect to the canonical metric only implies linear or sublinear growth rather than boundedness. Luckily, our definition of Lipschitz continuity with respect to a metric gives us some leeway through the choice of the metric. The following result shows that it is not necessary to include boundedness in the conditions of Definition 5.1, i.e. the boundedness of the function itself comes for free if we have established the boundedness of its derivative.

Proposition 5.2 *Let $V = (V_1, \dots, V_d)$ be a collection of vector fields on $\mathcal{M}$ which is Lipschitz continuous with respect to some complete Riemannian metric g. Then g can be chosen such that we have $|V_i|_g \leq 1$ for all $i \in \{1, \dots, d\}$.*

Proof. Assume that V is Lipschitz continuous with respect to g. Define

$$\rho : \mathcal{M} \to \mathbf{R} : y \mapsto |V_1(y)|_g^2 + \cdots + |V_d(y)|_g^2 + 1$$

and for a given arc-length parametrised curve $\gamma : \mathbf{R} \to \mathcal{M}$ we introduce

$$h_\gamma : \mathbf{R} \to \mathbf{R} : t \mapsto \sqrt{\rho(\gamma_t)}.$$

Then we have

$$\left|\dot{h}_\gamma(t)\right| = \frac{1}{h_\gamma(t)} \left| \sum_{i=1}^d \langle \nabla_{\dot{\gamma}_t} V_i, V_i\rangle_g \right| \leq \frac{\|V\|_{g-\mathrm{Lip}}}{h_\gamma(t)} \sum_{i=1}^d |V_i(\gamma_t)|_g \leq \|V\|_{g-\mathrm{Lip}} \sqrt{d} \tag{5.1}$$

and thus h_γ grows at most linearly fast. In a similar way we obtain the inequality

$$|W\rho| = 2 \left| \sum_{i=1}^d \langle \nabla_W V_i, V_i \rangle_g \right| \leq 2\|V\|_{g-\mathrm{Lip}} |W|_g \sum_{i=1}^d |V_i|_g \leq 2\|V\|_{g-\mathrm{Lip}} \sqrt{d} |W|_g \sqrt{\rho}. \tag{5.2}$$

for every vector field W.

Let us now define the Riemannian metric $\tilde{g}$ on $\mathcal{M}$ which is given by

$$\tilde{g}(y) = \frac{1}{\rho(y)} g(y)$$

and denote by $\tilde{\nabla}$ the covariant derivative associated to $\tilde{g}$. It is obvious that we have $|V_i|_{\tilde{g}} \leq 1$ for $i \in \{1, \ldots, d\}$. Hence it remains to verify that V is Lipschitz continuous with respect to $\tilde{g}$ and that $\tilde{g}$ is complete.

In order to prove the boundedness of the covariant derivatives it suffices to show that the expression $\langle \tilde{\nabla}_X V_i, W \rangle_{\tilde{g}}$ is bounded for all vector fields W and X with $|W|_{\tilde{g}}, |X|_{\tilde{g}} \leq 1$. To do so, we use the explicit formula (2.7) for the Levi-Civita connection and compute

$$\begin{aligned}
&\langle \tilde{\nabla}_X V_i, W \rangle_{\tilde{g}} \\
=&\frac{1}{2} \left(X\langle V_i, W \rangle_{\tilde{g}} + V_i \langle W, X \rangle_{\tilde{g}} - W\langle X, V_i \rangle_{\tilde{g}} - \langle [V_i, X], W \rangle_{\tilde{g}} - \langle [V_i, W], X \rangle_{\tilde{g}} - \langle [X, W], V_i \rangle_{\tilde{g}} \right) \\
=&\frac{1}{\rho} \langle \nabla_X V_i, W \rangle_g - \frac{X\rho}{2\rho^2} \langle V_i, W \rangle_g - \frac{V_i \rho}{2\rho^2} \langle W, X \rangle_g + \frac{W\rho}{2\rho^2} \langle X, V_i \rangle_g .
\end{aligned}$$

From $|W|_g, |X|_g \leq \sqrt{\rho}$ we see that the first term in this expression is bounded. For the three remaining terms this follows from (5.2) in combination with $|V_i|_g \leq \sqrt{\rho}$.

Finally, assume that $\gamma : [0, \infty) \to \mathcal{M}$ is a diverging curve which is arc-length parametrised with respect to g. Then we have

$$\int_0^\infty |\dot{\gamma}_t|_{\tilde{g}} \mathrm{d}t = \int_0^\infty \frac{\mathrm{d}t}{h_\gamma(t)} = \infty,$$

where the last equality holds because (5.1) implies that h_γ grows at most linearly fast. Hence $\tilde{g}$ satisfies (ii) of Proposition 2.34 and is therefore complete and the proof is finished. ∎

From now on, whenever we have a Lipschitz continuous collection $V = (V_1, \ldots, V_d)$ of vector fields we will implicitly assume that it comes with a metric g as in Proposition 5.2.

Lemma 5.3 *Let $V = (V_1, \ldots, V_d)$ be a collection of vector fields which is Lipschitz continuous with respect to a metric g. Then we have*

$$|[V_i, V_j]|_g \leq 2 \|V\|_{g\text{-Lip}}$$

for all $i, j \in \{1, \ldots d\}$.

Proof. Since the Levi-Civita connection ∇ is torsion-free we have

$$|[V_i, V_j]|_g = |\nabla_{V_i} V_j - \nabla_{V_j} V_i|_g \leq \|\nabla V_j\|_g |V_i|_g + \|\nabla V_i\|_g |V_j|_g \leq 2\|V\|_{g\text{-Lip}},$$

whence the claim follows. ∎

5.1.2. Complete vector fields and their flows

If V is a complete vector field, then we can consider the map $\phi_t^V : \mathcal{M} \to \mathcal{M}$ which maps a point $y_0 \in \mathcal{M}$ to the solution of (ODE) at time t for all $t \in \mathbf{R}$. This map is called the *flow of* V. It is well known that ϕ_t^V is a smooth diffeomorphism if V is smooth. The existence of flows helps us to show the following relation between Lipschitz continuity and completeness of vector fields. The result is an extension of [Gli97, Thm. 1.1] and thus it completes the proof of Proposition 2.39.

Proposition 5.4 *Let $V = (V_1, \ldots, V_d)$ be a collection of smooth vector fields such that we have $[V_i, V_j] = 0$ for all $i, j \in \{1, \ldots, d\}$. For $i \in \{1, \ldots, d\}$ define the vector field V_i^+ on $\mathcal{M}^+ := \mathcal{M} \times \mathbf{R}^d$ by $V_i^+(y, t) = (V_i(y), e_i)$. Then there exists a complete Riemannian metric with $\nabla V_i^+ = 0$ for all $i \in \{1, \ldots, d\}$ if and only if all V_i are complete.*

Proof. Assume that a complete metric g is given such that $\nabla V_i^+ = 0$. Then it follows easily that V_i^+ is bounded. We have already shown in Proposition 2.39 that this implies the completeness of V_i.

In order to prove the converse direction we assume that all V_i are complete and then construct a complete Riemannian metric g such that $\nabla V_i^+ = 0$ holds for all $i \in \{1, \ldots, d\}$. Fix an arbitrary complete metric g_1 on $\mathcal{M}$, denote by g_2 the canonical Euclidean metric on $\mathbf{R}^d$ and define

$$g := g_1 \oplus g_2$$

on $\mathcal{M} \times \mathbf{R}^d$. Then g is certainly complete and it is easy to see that the vector fields W_i given by $W_i(y, t) = (0, e_i)$ for $i \in \{1, \ldots, d\}$ satisfy ∇W_i=0. Now consider the (global) diffeomorphism

$$\Phi : \mathcal{M} \times \mathbf{R}^d \to \mathcal{M} \times \mathbf{R}^d : (y, t) \mapsto \left(\phi_1^{t_1 V_1 + \cdots + t_d V_d}(y), t\right) = \left(\left(\phi_{t_1}^{V_1} \circ \cdots \circ \phi_{t_d}^{V_d}\right)(y), t\right).$$

A direct computation shows that the pushforward vector field $\Phi_* W_i$ of W_i under Φ is given by the vector field V_i^+ for all $i \in \{1, \ldots, d\}$. Indeed, we have

$$(\Phi_* W_i)(y, t) = \mathrm{D}\Phi(\Phi^{-1}(y, t)) W_i(\Phi^{-1}(y, t))$$

$$= \begin{pmatrix} \frac{\partial}{\partial \tilde{y}} \Phi(\tilde{y}, \tilde{t})|_{(\tilde{y},\tilde{t})=\Phi^{-1}(y,t)} & V(y) \\ & I_d \end{pmatrix} \begin{pmatrix} 0 \\ e_i \end{pmatrix} = \begin{pmatrix} V_i(y) \\ e_i \end{pmatrix} = V_i^+(y, t)$$

Thus the pullback metric $(\Phi^{-1})^* g$ of g under the inverse of Φ satisfies the desired properties.

■

The flow ϕ_t^V induces a linear operator on $C^\infty(\mathcal{M})$ which is given by

$$f \mapsto f \circ \phi_t^V.$$

We denote this operator by e^{tV}. The reason for this notation is that e^{tV} behaves a lot like an operator exponential as we shall see below. Note however that the exponential series with the operator V plugged in as the argument does not necessarily converge.

Our principal source of inspiration for the remainder of this section is [DT13], albeit their main result [DT13, Thm. 3.1] is not sufficient for our desired application to rough differential equations. We modify their arguments in order to obtain our own approximation error representation in Proposition 5.9.

Definition 5.5 Let $A(\tau)$ be a linear operator on $C^\infty(\mathcal{M})$ for all $\tau \in \mathbf{R}$. We say that the map $\tau \mapsto A(\tau)$ is *differentiable* if the map $\tau \mapsto (A(\tau)f)(y)$ is differentiable for all $f \in C^\infty(\mathcal{M})$ and all $y \in \mathcal{M}$. In this case, we define $\frac{\mathrm{d}}{\mathrm{d}\tau} A(\tau)$ to be the operator which is given by

$$\left(\left(\frac{\mathrm{d}}{\mathrm{d}\tau} A(\tau)\right) f\right)(y) = \frac{\mathrm{d}}{\mathrm{d}\tau}((A(\tau)f)(y))$$

We also say that A is a *differentiable time-dependent operator.*

With this definition we have that $t \mapsto e^{tV}$ solves the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t} e^{tV} = e^{tV} V, \tag{5.3}$$

which by definition is equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(f \circ \phi_t^V(y)\right) = Vf\left(\phi_t^V(y)\right)$$

for all $f \in C^\infty(\mathcal{M})$ and all $y \in \mathcal{M}$. Furthermore one easily verifies that two differentiable time-dependent operators A and B satisfy the product rule

$$\frac{\mathrm{d}}{\mathrm{d}\tau}(A(\tau)B(\tau)) = \frac{\mathrm{d}}{\mathrm{d}\tau} A(\tau)B(\tau) + A(\tau)\frac{\mathrm{d}}{\mathrm{d}\tau} B(\tau). \tag{5.4}$$

Combining (5.3) and (5.4) leads to

$$0 = e^{tV} \frac{\mathrm{d}}{\mathrm{d}\tau}(e^{-tV} e^{tV}) = -V e^{tV} + e^{tV} V = \left[e^{tV}, V\right].$$

Let us now turn to some more substantial auxiliary results. The first one is essentially a non-commutative chain rule.

Lemma 5.6 *Let V be a differentiable time-dependent operator such that $V(\tau)$ is a complete vector field on $\mathcal{M}$ for every $\tau \in \mathbf{R}$. Then we have*

$$\frac{\mathrm{d}}{\mathrm{d}\tau} e^{V(\tau)} = \int_0^1 e^{(1-\alpha)V(\tau)} \frac{\mathrm{d}}{\mathrm{d}\tau} V(\tau) e^{\alpha V(\tau)} \mathrm{d}\alpha = \int_0^1 e^{\alpha V(\tau)} \frac{\mathrm{d}}{\mathrm{d}\tau} V(\tau) e^{(1-\alpha)V(\tau)} \mathrm{d}\alpha.$$

Proof. We only have to show the first equation since the second one follows immediately from a variable substitution. We show

$$e^{-V(\tau)} \frac{\mathrm{d}}{\mathrm{d}\tau} e^{V(\tau)} = \int_0^1 e^{-\alpha V(\tau)} \frac{\mathrm{d}}{\mathrm{d}\tau} V(\tau) e^{\alpha V(\tau)} \mathrm{d}\alpha,$$

which is equivalent to the first equation. Clearly we have

$$e^{-V(\tau)} \frac{\mathrm{d}}{\mathrm{d}\tau} e^{V(\tau)} = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\alpha} \left(e^{-\alpha V(\tau)} \frac{\mathrm{d}}{\mathrm{d}\tau} e^{\alpha V(\tau)} \right) \mathrm{d}\alpha.$$

Therefore we compute

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\alpha} \left(e^{-\alpha V(\tau)} \frac{\mathrm{d}}{\mathrm{d}\tau} e^{\alpha V(\tau)} \right) \\
&= - e^{-\alpha V(\tau)} V(\tau) \frac{\mathrm{d}}{\mathrm{d}\tau} e^{\alpha V(\tau)} + e^{-\alpha V(\tau)} \frac{\mathrm{d}}{\mathrm{d}\alpha} \frac{\mathrm{d}}{\mathrm{d}\tau} e^{\alpha V(\tau)} \\
&= - e^{-\alpha V(\tau)} V(\tau) \frac{\mathrm{d}}{\mathrm{d}\tau} e^{\alpha V(\tau)} + e^{-\alpha V(\tau)} \frac{\mathrm{d}}{\mathrm{d}\tau} \frac{\mathrm{d}}{\mathrm{d}\alpha} e^{\alpha V(\tau)} \\
&= - e^{-\alpha V(\tau)} V(\tau) \frac{\mathrm{d}}{\mathrm{d}\tau} e^{\alpha V(\tau)} + e^{-\alpha V(\tau)} \frac{\mathrm{d}}{\mathrm{d}\tau} \left(V(\tau) e^{\alpha V(\tau)} \right) \\
&= - e^{-\alpha V(\tau)} V(\tau) \frac{\mathrm{d}}{\mathrm{d}\tau} e^{\alpha V(\tau)} + e^{-\alpha V(\tau)} \frac{\mathrm{d}}{\mathrm{d}\tau} V(\tau) e^{\alpha V(\tau)} + e^{-\alpha V(\tau)} V(\tau) \frac{\mathrm{d}}{\mathrm{d}\tau} e^{\alpha V(\tau)} \\
&= e^{-\alpha V(\tau)} \frac{\mathrm{d}}{\mathrm{d}\tau} V(\tau) e^{\alpha V(\tau)}
\end{aligned}$$

as desired. ∎

Lemma 5.7 *Let V and W be complete vector fields on $\mathcal{M}$. Then we have*

$$\left[e^V, W\right] = \int_0^1 e^{(1-\alpha)V}[V,W]e^{\alpha V}\mathrm{d}\alpha = \int_0^1 e^{\alpha V}[V,W]e^{(1-\alpha)V}\mathrm{d}\alpha.$$

Proof. The argument is the same as in the previous lemma. We show

$$e^{-V}\left[e^V, W\right] = \int_0^1 e^{-\alpha V}[V,W]e^{\alpha V}\mathrm{d}\alpha.$$

This is done exactly as above by computing

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}\alpha}\left(e^{-\alpha V}\left[e^{\alpha V}, W\right]\right) = \frac{\mathrm{d}}{\mathrm{d}\alpha}\left(W - e^{-\alpha V}We^{\alpha V}\right)\\ =&e^{-\alpha V}VWe^{\alpha V} - e^{-\alpha V}WVe^{\alpha V} = e^{-\alpha V}[V,W]e^{\alpha V},\end{aligned}$$

whence the claim follows. ∎

Lemma 5.8 *Let V and W be two complete smooth vector fields on $\mathcal{M}$. Then we have*

$$\begin{aligned}&\int_0^1 e^{-\alpha V}We^{\alpha V}\mathrm{d}\alpha\\ =&\sum_{n=1}^{m}\frac{(-1)^{n-1}}{n!}\mathrm{ad}_V^{n-1}W + \int_0^1 \frac{(\alpha-1)^m}{m!}e^{-\alpha V}\left(\mathrm{ad}_V^m W\right)e^{\alpha V}\mathrm{d}\alpha\end{aligned}$$

for every $m \in \mathbb{N}$, where $\mathrm{ad}_V^0 W = W$ and $\mathrm{ad}_V^n W = \left[V, \mathrm{ad}_V^{n-1}W\right]$.

Proof. We claim that the given formula is the Taylor expansion of the map

$$h : \tau \mapsto \int_0^\tau e^{-\alpha V}We^{\alpha V}\mathrm{d}\alpha$$

around $\tau_0 = 0$ and evaluated at $\tau = 1$ with the remainder given in integral form. In order to verify this we only have to show that the n-th derivative $h^{(n)}$ of h is given by

$$h^{(n)}(\tau) = (-1)^{n-1}e^{-\tau V}\left(\mathrm{ad}_V^{n-1}W\right)e^{\tau V}$$

for $n \geq 1$. For $n = 1$ this is obvious. Thus we assume that the claim holds for some $n \geq 1$. Then we have

$$h^{(n+1)}(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau}h^{(n)}(\tau) = (-1)^{n-1}\left(-e^{-\tau V}V\mathrm{ad}_V^{n-1}We^{\tau V} + e^{-\tau V}\mathrm{ad}_V^{n-1}WVe^{\tau V}\right)$$

$$= (-1)^n e^{-\tau V} \mathrm{ad}_V^n W e^{\tau V},$$

which finishes the proof. ■

Let us now introduce the Baker-Campbell-Hausdorff formula. In fact, we will only need its existence rather than its explicit formulation. To this end, let A and B be formal Lie series in d variables. Then one can show that there exists another Lie series, which we denote by $\mathrm{BCH}(A, B)$, such that the formal identity

$$e^A e^B = e^{\mathrm{BCH}(A,B)} \tag{5.5}$$

holds [Reu93, Cor. 3.4]. Since the space of formal Lie series carries a natural grading, we can define $\mathrm{BCH}_k(A, B)$ to be the truncation of $\mathrm{BCH}(A, B)$ such that only the terms of degree k or less are left.

Let now $V = (V_1, \ldots, V_d)$ be a collection of smooth vector fields. Define

$$\begin{aligned} \mathcal{L}^1(V) &:= \{V_1, \ldots, V_d\} \\ \mathcal{L}^k(V) &:= \left\{[V_i, W] : i \in \{1, \ldots, d\}, W \in \mathcal{L}^{k-1}(V)\right\} \\ \mathcal{L}^{(k)}(V) &:= \mathcal{L}^1(V) \cup \cdots \cup \mathcal{L}^k(V). \end{aligned} \tag{5.6}$$

In other words, $\mathrm{span}\left(\mathcal{L}^{(k)}(V)\right)$ is obtained by evaluating Lie polynomials of degree k or less for the variables $V_1, \ldots, V_d$.

Furthermore, for a finite subset $\{v_1, \ldots, v_d\}$ of some vector space, we define

$$\mathrm{span}_{\leq C}(\{v_1, \ldots, v_d\}) := \{\alpha_1 v_1 + \cdots + \alpha_d v_d : |\alpha_i| \leq C \text{ for } i \in \{1, \ldots, d\}\}.$$

Proposition 5.9 *Let $V = (V_1, \ldots, V_d)$ be a collection of smooth vector fields and assume that all elements of $\mathrm{span}\left(\mathcal{L}^{(k)}(V)\right)$ are complete. Then, for $C > 0$ and $W_1, W_2 \in \mathrm{span}_{\leq C}\left(\mathcal{L}^{(k)}(V)\right)$, we have*

$$\begin{aligned} &e^{W_1} e^{W_2} - e^{\mathrm{BCH}_k(W_1,W_2)} \\ &= \int_0^1 \int_0^1 e^{\tau W_1} e^{(1-\alpha)\tau W_2} \widehat{R}_1(\alpha, \tau) e^{\alpha\tau W_2} e^{-\mathrm{BCH}_k(\tau W_1, \tau W_2)} e^{\mathrm{BCH}_k(W_1,W_2)} \mathrm{d}\alpha \mathrm{d}\tau \\ &+ \int_0^1 \int_0^1 e^{\tau W_1} e^{\tau W_2} e^{-\alpha \mathrm{BCH}_k(\tau W_1, \tau W_2)} \widehat{R}_2(\alpha, \tau) e^{(\alpha-1)\mathrm{BCH}_k(\tau W_1, \tau W_2)} e^{\mathrm{BCH}_k(W_1,W_2)} \mathrm{d}\alpha \mathrm{d}\tau, \end{aligned}$$

such that the vector fields $\widehat{R}_1(\alpha,\tau)$ and $\widehat{R}_2(\alpha,\tau)$ are of the form

$$\widehat{R}_i(\alpha,\tau) = \sum_{j=1}^{m} p_i^j(\alpha,\tau)\left[R_i^j, \bar{R}_i^j\right] \tag{5.7}$$

for some $m \in \mathbf{N}$, where

(i) the $p_i^j : [0,1]^2 \to \mathbf{R}$ are polynomials which are bounded from above by a constant which depends only on C and k

(ii) and for fixed $i \in \{1,2\}$ and $j \in \{1,\dots,m\}$ there exist $l, \bar{l} \in \{1,\dots,k\}$ such that $l + \bar{l} > k$ and $R_i^j \in \mathcal{L}^l(V)$ and $\bar{R}_i^j \in \mathcal{L}^{\bar{l}}(V)$.

Proof. Instead of considering the term $e^{W_1}e^{W_2} - e^{\mathrm{BCH}_k(W_1,W_2)}$ directly we compute the expression $e^{W_1}e^{W_2}e^{-\mathrm{BCH}_k(W_1,W_2)} - \mathrm{id}_{C^\infty(\mathcal{M})}$, from which the result follows by multiplying with $e^{\mathrm{BCH}_k(W_1,W_2)}$ from the right. We obviously have

$$e^{W_1}e^{W_2}e^{-\mathrm{BCH}_k(W_1,W_2)} - \mathrm{id}_{C^\infty(\mathcal{M})} = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\tau}\left(e^{\tau W_1}e^{\tau W_2}e^{-\mathrm{BCH}_k(\tau W_1,\tau W_2)}\right)\mathrm{d}\tau.$$

Hence it remains to find a suitable expression for the derivative inside the integral. Therefore we compute

$$\begin{aligned}
&e^{-\tau W_2}e^{-\tau W_1}\frac{\mathrm{d}}{\mathrm{d}\tau}\left(e^{\tau W_1}e^{\tau W_2}e^{-\mathrm{BCH}_k(\tau W_1,\tau W_2)}\right)e^{\mathrm{BCH}_k(\tau W_1,\tau W_2)}\\
=&e^{-\tau W_2}We^{\tau W_2} + W_2 - \int_0^1 e^{-\alpha\mathrm{BCH}_k(\tau W_1,\tau W_2)}\frac{\mathrm{d}}{\mathrm{d}\tau}\mathrm{BCH}_k(\tau W_1,\tau W_2)e^{\alpha\mathrm{BCH}_k(\tau W_1,\tau W_2)}\mathrm{d}\alpha\\
=&W_1 + W_2\\
&+\tau\int_0^1 e^{-\alpha\tau W_2}[W_1,W_2]e^{\alpha\tau W_2}\mathrm{d}\alpha && (5.8a)\\
&-\int_0^1 e^{-\alpha\mathrm{BCH}_k(\tau W_1,\tau W_2)}\frac{\mathrm{d}}{\mathrm{d}\tau}\mathrm{BCH}_k(\tau W_1,\tau W_2)e^{\alpha\mathrm{BCH}_k(\tau W_1,\tau W_2)}\mathrm{d}\alpha, && (5.8b)
\end{aligned}$$

where we have used Lemma 5.6 for the first equation and Lemma 5.7 for the second one. We can split $[W_1,W_2]$ and $\frac{\mathrm{d}}{\mathrm{d}\tau}\mathrm{BCH}_k(\tau W_1,\tau W_2)$ into their homogeneous components by writing

$$\begin{aligned}
[W_1,W_2] &= A_2 + A_3 + \cdots + A_k + \bar{A}\\
\frac{\mathrm{d}}{\mathrm{d}\tau}\mathrm{BCH}_k(\tau W_1,\tau W_2) &= B_1 + B_2 + \cdots + B_k,
\end{aligned}$$

where A_i and B_i are of degree i (as Lie polynomials in the variables $V_1,\dots,V_d$) for all indices

$i \in \{1, \dots, k\}$ and $\bar{A}$ is already of the desired form for the remainder, i.e. it can be written in the form of (5.7). Thus we can write (5.8a) as

$$\sum_{i=2}^{k} \tau \int_0^1 e^{-\alpha\tau Z} A_i e^{\alpha\tau Z} \mathrm{d}\alpha + \tau \int_0^1 e^{-\alpha\tau Z} \bar{A} e^{\alpha\tau Z} \mathrm{d}\alpha$$

and apply Lemma 5.8 to each term in the sum separately such that we expand the term involving A_i up to degree $k - i$. All remainders that we obtain this way are of the desired form. An analogous procedure can be applied to (5.8b). It only remains to observe that all non-remainder terms are Lie polynomials of degree k or less and thus (5.5) implies that they get cancelled. ∎

Lemma 5.10 *Let V be Lipschitz continuous with respect to a metric g. Then we have*

$$d_g\left(\phi_t^V(y_1), \phi_t^V(y_2)\right) \le e^{t\|V\|_{g-\mathrm{Lip}}} d_g(y_1, y_2).$$

for all $y_1, y_2 \in \mathcal{M}$.

Proof. See Theorem 1.2 of [KSSV06]. ∎

Lemma 5.11 *Let $\varphi : \mathcal{M} \to \mathcal{M}$ be a Lipschitz continuous diffeomorphism with Lipschitz constant L with respect to the distance d_g associated to some metric g. Fix $z \in \mathcal{M}$ and let $f_z := d_g(z, \cdot)$. Let furthermore W be an arbitrary vector field. Then we have*

$$|W(f_z \circ \varphi)| \le L|W|_g.$$

Proof. We compute

$$\begin{aligned}|W(f_z \circ \varphi)(y)| &= \lim_{h\downarrow 0} \frac{\left|f_z\left(\varphi\left(\phi_h^W(y)\right)\right) - f_z(\varphi(y))\right|}{h} \le \lim_{h\downarrow 0} \frac{\left|d_g\left(\varphi\left(\phi_h^W(y)\right), \varphi(y)\right)\right|}{h} \\ &\le L \lim_{h\downarrow 0} \frac{1}{h}\left|d_g\left(\phi_h^W(y), y\right)\right| = L|W|_g.\end{aligned}$$

See also Appendix B for more details on the differentiability of f_z. ∎

Proposition 5.12 *Let $V = (V_1, \dots, V_d)$ be a collection of smooth vector fields such that the collection $\mathcal{L}^{(k)}(V)$ is Lipschitz continuous with respect to a complete metric g. Then there exists $C > 0$, which depends only on k and d, such that for all $\lambda \in (0, 1)$ and for all vector fields*

$W_1, W_2 \in \operatorname{span}_{\leq 1}(\mathcal{L}^{(k)}(\lambda V))$ *we have*

$$d_g\left(\phi_1^{W_2}\left(\phi_1^{W_1}(y_0)\right), \phi_1^{\mathrm{BCH}_k(W_1,W_2)}(y_0)\right) \leq C\lambda^{k+1}\|\mathcal{L}^{(k)}(V)\|_{g-\mathrm{Lip}}e^{C\lambda\|\mathcal{L}^{(k)}(V)\|_{g-\mathrm{Lip}}}$$

uniformly over all $y_0 \in \mathcal{M}$. *Here we have used the notation* $\lambda V := (\lambda V_1, \ldots, \lambda V_d)$.

Proof. Define $z := \phi_1^{W_2}\left(\phi_1^{W_1}(y_0)\right)$ and consider the function $f_z := d_g(z, \cdot)$ as before. Then we have

$$d_g\left(\phi_1^{W_2}\left(\phi_1^{W_1}(y_0)\right), \phi_1^{\mathrm{BCH}_k(W_1,W_2)}(y_0)\right) = f\left(\phi_1^{\mathrm{BCH}_k(W_1,W_2)}(y_0)\right) - f\left(\phi_1^{W_2}\left(\phi_1^{W_1}(y_0)\right)\right).$$

Thus the claim follows directly from Proposition 5.9, where the remainders can be estimated with Lemmas 5.3, 5.10 and 5.11.

■

5.2. Global solutions of rough differential equations

Definition 5.13 Let $\mathbf{x}$ be a geometric p-rough path over $\mathbf{R}^d$ controlled by ω. Furthermore, let $V = (V_1, \ldots, V_d)$ be a collection of smooth vector fields on $\mathcal{M}$. We say that $y : [0, T] \to \mathcal{M}$ solves the *rough differential equation*

$$\mathrm{d}y_t = V(y_t)\mathrm{d}\mathbf{x}_t \tag{RDEm}$$

if for each compactly supported smooth test function $f : \mathcal{M} \to \mathbf{R}$ there exists a function $\theta_f : \mathbf{R} \to \mathbf{R}$ with $\theta_f(h)/h \to 0$ for $h \to 0$ such that

$$\left|f(y_t) - V_{\mathbf{x}_{s,t}}f(y_s)\right| \leq \theta_f(\omega(s,t)) \tag{5.9}$$

holds for all $0 \leq s \leq t \leq T$.

The following local existence and uniqueness result is obtained by a standard argument, which can be found e.g. in [IW89] in the context of stochastic differential equations. A crucial role in the proof is played by Theorem 4.31 and Corollary 4.34.

Proposition 5.14 *For each initial value* $y_0 \in \mathcal{M}$

- *there either exists a continuous map* $y : [0, T] \to \mathcal{M}$ *which solves* (RDEm)
- *or there exists* $\tau \in (0, T]$ *and a continuous map* $y : [0, \tau) \to \mathcal{M}$ *such that*

(i) the restriction of y to $[0, t]$ solves (RDEm) for all $t \in (0, \tau)$ and

(ii) the image of $[0, \tau)$ under y is not relatively compact.

In either case the solution y is unique.

Proof. Fix a locally finite coordinate covering $(U_i, \psi_i)_{i \in \mathcal{I}}$ of $\mathcal{M}$ and assume that all coordinate neighbourhoods are relatively compact and that the closure of any coordinate neighbourhood U_i is contained in another coordinate neighbourhood $\widetilde{U}_i$.

Now fix $i \in \mathcal{I}$ such that $y_0 \in U_i$ and represent the vector fields V in the local coordinates of U_i and denote these representations by $\tilde{V}$. Since $U_i \Subset \mathcal{M}$ and consequently $\psi_i(U_i) \Subset \mathbf{R}^n$ are relatively compact, we can extend the vector fields $\tilde{V}$ to smooth and bounded vector fields on $\mathbf{R}^n$. Now consider the RDE

$$\mathrm{d}\tilde{y}_t = \tilde{V}(\tilde{y}_t)\mathrm{d}\mathbf{x}_t, \quad \tilde{y}_0 = \psi_i(y_0) \in \psi_i(U) \tag{5.10}$$

on $\mathbf{R}^n$. It follows from Proposition 3.16 that it has a unique global solution $\tilde{y}$ on $[0, T]$.

Define $\tau_1^i := \inf\{t : \tilde{y}_t \notin \psi_i(U_i)\} \wedge T$. If U_j is another neighbourhood of y_0, we obtain a solution $\hat{y}$ in $\psi_j(U_j)$. However, it follows from Corollary 4.34 that both solutions coincide on $\mathcal{M}$ up to time $\tau_1^i \wedge \tau_1^j$. Hence the solution is independent of the chosen coordinates.

Now define $\tau_1 := \max\{\tau_1^i : i \in \mathcal{I}, y_0 \in U_i\}$. If $\tau_1 = T$, then there exists a neighbourhood of y_0 which contains the whole solution. (Recall that we have assumed that the closure of every U_i is contained in another coordinate neighbourhood.) Otherwise we proceed iteratively and consider the solution which starts at y_{τ_1} at time τ_1. This gives us a time τ_2 and so forth. Now there are two cases. Either there exists $k \in \mathbf{N}$ with $\tau_k = T$, in which case the solution exists globally, i.e. on the whole interval $[0, T]$. Or we have $\tau_k < T$ for all $k \in \mathbf{N}$, in which case the solution exists up to every time $t < \tau := \lim_{k \to \infty} \tau_k$ but not on the compact interval $[0, \tau]$. In this case $\{y_t : 0 \le t < \tau\}$ cannot be relatively compact because it has, by construction, a locally finite cover consisting of infinitely many open sets.

To finish the proof, let us verify that y satisfies Definition 5.13 and that it is the only continuous path to do so. Fix a compact interval $[0, t]$ on which the solution exists. By the above construction this interval splits into subintervals $[0, t] = [0, \tau_1] \cup \cdots \cup [\tau_{k-1}, t]$, such that there exist coordinate neighbourhoods $U_1, \ldots, U_k$ with $y_s \in U_i$ for $s \in [\tau_{i-1}, \tau_i]$. It follows from Theorem 4.31 that (5.9) holds for each of the subintervals and hence it holds for the whole interval $[0, t]$. Conversely, Theorem 4.31 also implies that any continuous path $y : [\tau_{i-1}, \tau_i] \to \mathcal{M}$ which satisfies (5.9) solves (5.10) in the respective coordinate chart. Hence the uniqueness property of Proposition 3.16 carries over to equations manifolds. ∎

If (RDEm) has a global solution, i.e. if the solution exists on the whole interval $[0,T]$ for every initial value y_0, then we can associate a well-defined flow $\Phi^{\mathbf{x},V}$ to (RDEm) and we obtain the following approximation.

Lemma 5.15 *Let $\mathbf{x}$ be a geometric p-rough path over $\mathbf{R}^d$ which is controlled by a control ω and let $V = (V_1, \dots, V_d)$ be a collection of smooth vector fields on $\mathcal{M}$. Then for each compactly supported smooth test function $f : \mathcal{M} \to \mathbf{R}$ there exists $\bar{\theta}_f : \mathbf{R} \to \mathbf{R}$ with $\bar{\theta}_f(h)/h \to 0$ for $h \to 0$ such that*

$$\left\| f \circ \Phi^{\mathbf{x},V}_{t\leftarrow s} - f \circ \phi_1^{V_{\log(\mathbf{x}_{s,t})}} \right\|_\infty \leq \bar{\theta}_f(\omega(s,t))$$

holds for all $0 \leq s \leq t \leq T$.

Proof. Using the triangle inequality we obtain

$$\left\| f \circ \Phi^{\mathbf{x},V}_{t\leftarrow s} - f \circ \phi_1^{V_{\log(\mathbf{x}_{s,t})}} \right\|_\infty \leq \left\| f \circ \Phi^{\mathbf{x},V}_{t\leftarrow s} - V_{\mathbf{x}_{s,t}} f \right\|_\infty + \left\| V_{\mathbf{x}_{s,t}} f - f \circ \phi_1^{V_{\log(\mathbf{x}_{s,t})}} \right\|_\infty .$$

It follows directly from Definition 5.13 that the first term satisfies the desired bound. To obtain a bound for the second term, fix $y \in \mathcal{M}$ and consider the Taylor expansion of the map

$$\tau \mapsto f\left(\phi_\tau^{V_{\log(\mathbf{x}_{s,t})}}(y)\right)$$

around $\tau = 0$, which gives

$$f\left(\phi_1^{V_{\log(\mathbf{x}_{s,t})}}(y)\right) = \sum_{k=0}^{[p]} \frac{1}{k!} V^k_{\log(\mathbf{x}_{s,t})} f(y) + \int_0^1 \frac{\tau^{[p]}}{[p]!} V^{[p]+1}_{\log(\mathbf{x}_{s,t})} f\left(\phi_\tau^{V_{\log(\mathbf{x}_{s,t})}}(y)\right) \mathrm{d}\tau .$$

Furthermore we have

$$\sum_{k=0}^{[p]} \frac{1}{k!} V^k_{\log(\mathbf{x}_{s,t})} f(y) = V_{\mathbf{x}_{s,t}} f(y) + o(\omega(s,t)).$$

Now the claim follows since f is compactly supported and hence the bounds on the remainders are uniform in y. ∎

5.2.1. Main result on global existence

Theorem 5.16 *Let $V = (V_1, \dots, V_d)$ be a collection of smooth vector fields on $\mathcal{M}$ such that $\mathcal{L}^{(k)}(V)$ (as defined in (5.6)) is Lipschitz continuous with respect to a complete metric. Then (RDEm) has a global solution for every geometric p-rough path $\mathbf{x}$ with $[p] \leq k$ and every initial* *value* $y_0 \in \mathcal{M}$.

Remark 5.17 The recent work [BC18] also gives a criterion for the global existence of solutions for rough differential equations in terms of iterated Lie brackets. They only consider equations driven along vector fields in $\mathbf{R}^n$ and they formulate their condition with respect to the Euclidean norm. Thus they cannot assume boundedness of the vector fields and their Lie brackets as this would exclude e.g. linear vector fields. Therefore they have to introduce a notion of α-growth and both their conditions and their arguments in the proof become rather technical. This illustrates how useful it is to work with arbitrary Riemannian metrics since Proposition 5.2 always allows us to assume boundedness.

The following paragraphs follow [Bai15] in that we combine their arguments with our approximation error estimates in terms of the Lie brackets from above.

For a given control function ω we define a sequence of partitions on a given interval $[s,t]$ as follows. Define $\pi_0 := \{s,t\}$. In order to proceed inductively we assume that we have already defined

$$\pi_n = \{\tau_0, \tau_1, \ldots, \tau_{2^n-1}, \tau_{2^n}\} \tag{5.11}$$

for some $n \in \mathbf{N}$ and then we define

$$\pi_{n+1} := \{\tau_0, \tilde{\tau}_1, \tau_1, \ldots, \tilde{\tau}_{2^n}, \tau_{2^n}\} \tag{5.12}$$

such that $\tilde{\tau}_i \in (\tau_{i-1}, \tau_i)$ and $\omega(\tau_{i-1}, \tilde{\tau}_i) = \omega(\tilde{\tau}_i, \tau_i)$ for all $i \in \{1, \ldots, 2^n\}$.

Remark 5.18 If the map $[s,T] \to \mathbf{R} : t \mapsto \omega(s,t)$ is strictly increasing for every $s \in [0,T]$, then the values $\tilde{\tau}_i$ in the above construction are unique. We can always assume that this is the case by replacing ω with $\tilde{\omega}$ as appropriate, where $\tilde{\omega}(s,t) := \omega(s,t) + |t-s|$.

Let us now fix $T > 0$, a collection of vector fields $V = (V_1, \ldots, V_d)$ and a geometric rough path $\mathbf{x}$ controlled by ω. For $0 \le s < t \le T$ and a partition $\rho = \{s = \tau_0, \ldots, \tau_r = t\}$ of $[s,t]$ define

$$\varphi_{t \leftarrow s}^{\mathbf{x},V,\rho} := \phi_1^{V_{\log\left(\mathbf{x}_{\tau_{r-1},\tau_r}\right)}} \circ \cdots \circ \phi_1^{V_{\log\left(\mathbf{x}_{\tau_0,\tau_1}\right)}}. \tag{5.13}$$

Proposition 5.19 *Let* $\mathbf{x}$ *be a geometric* p*-rough path controlled by* ω *for some* $p \ge 1$. *Let* $V = (V_1, \ldots, V_d)$ *be a collection of smooth vector fields such that* $\mathcal{L}^{(\lfloor p \rfloor)}(V)$ *is Lipschitz continuous with respect to a complete metric* g. *Then there exists a constant* $C > 0$ *which depends neither on* n *nor on* y *such that*

$$d_g\left(\varphi_{T \leftarrow 0}^{\mathbf{x},V,\pi_n}(y), y\right) \le C$$

holds.

Proof. First of all, it follows from Corollary 2.41 that the maps $\varphi^{\mathbf{x},V,\pi_n}_{T\leftarrow 0}$ are well-defined. We follow the proof of [Bai15, Lem. 2.3] and proceed inductively over n. First of all, note that the multiplicative property of $\mathbf{x}$ implies that

$$V_{\log(\mathbf{x}_{s,u})} = \mathrm{BCH}_{[p]}\left(V_{\log(\mathbf{x}_{s,t})}, V_{\log(\mathbf{x}_{t,u})}\right).$$

Therefore Proposition 5.12 guarantees that we can choose $C_1 > 0$ and $\delta > 0$ such that for all $s, t \in [0, T]$ with $|t - s| < \delta$ we have

$$d_g\left(\varphi^{\mathbf{x},V,\pi_1}_{t\leftarrow s}(y), \varphi^{\mathbf{x},V,\pi_0}_{t\leftarrow s}(y)\right) \leq C_1\omega(s,t)^\theta,$$

where we define $\theta := ([p] + 1)/p > 1$ in order to simplify the notation. Furthermore, define

$$C_2 := \frac{2^\theta}{2^{\theta-1} - 1} C_1 > 0.$$

By decreasing δ, Lemma 5.10 allows us to assume that for all $s, t \in [0, T]$ with $|t - s| < \delta$ and for all $y_1, y_2 \in \mathcal{M}$ we have

$$d_g\left(\varphi^{\mathbf{x},V,\pi_0}_{t\leftarrow s}(y_1), \varphi^{\mathbf{x},V,\pi_0}_{t\leftarrow s}(y_2)\right) \leq 2^{\theta-1} d_g(y_1, y_2)$$

We will now prove that we have

$$d_g\left(\varphi^{\mathbf{x},V,\pi_n}_{u\leftarrow s}(y), \varphi^{\mathbf{x},V,\pi_0}_{u\leftarrow s}(y)\right) \leq C_2\omega(s,u)^\theta \tag{5.14}$$

for all $y \in \mathcal{M}$, all $n \in \mathbf{N}$ and all $s, u \in [0, T]$ with $|u - s| < \delta$. The claim holds trivially for $n = 0$. Hence assume that it holds for some $n \in \mathbf{N}$. By construction of the partitions π_n there exists $t \in (s, u)$ with $\varphi^{\mathbf{x},V,\pi_{n+1}}_{u\leftarrow s} = \varphi^{\mathbf{x},V,\pi_n}_{u\leftarrow t} \circ \varphi^{\mathbf{x},V,\pi_n}_{t\leftarrow s}$ and such that $\omega(s,t) = \omega(t,u) \leq \omega(s,u)/2$. Hence we have

$$\begin{aligned}
&d_g\left(\varphi^{\mathbf{x},V,\pi_{n+1}}_{u\leftarrow s}(y), \varphi^{\mathbf{x},V,\pi_0}_{u\leftarrow s}(y)\right)\\
\leq &d_g\left(\varphi^{\mathbf{x},V,\pi_n}_{u\leftarrow t}\left(\varphi^{\mathbf{x},V,\pi_n}_{t\leftarrow s}(y)\right), \varphi^{\mathbf{x},V,\pi_0}_{u\leftarrow t}\left(\varphi^{\mathbf{x},V,\pi_n}_{t\leftarrow s}(y)\right)\right)\\
&+ d_g\left(\varphi^{\mathbf{x},V,\pi_0}_{u\leftarrow t}\left(\varphi^{\mathbf{x},V,\pi_n}_{t\leftarrow s}(y)\right), \varphi^{\mathbf{x},V,\pi_0}_{u\leftarrow t}\left(\varphi^{\mathbf{x},V,\pi_0}_{t\leftarrow s}(y)\right)\right)\\
&+ d_g\left(\varphi^{\mathbf{x},V,\pi_0}_{u\leftarrow t}\left(\varphi^{\mathbf{x},V,\pi_0}_{t\leftarrow s}(y)\right), \varphi^{\mathbf{x},V,\pi_0}_{u\leftarrow s}(y)\right)\\
\leq &C_2\omega(t,u)^\theta + 2^{\theta-1} d_g\left(\varphi^{\mathbf{x},V,\pi_n}_{t\leftarrow s}(y), \varphi^{\mathbf{x},V,\pi_0}_{t\leftarrow s}(y)\right) + C_1\omega(s,u)^\theta\\
\leq &C_2\omega(t,u)^\theta + 2^{\theta-1} C_2\omega(s,t)^\theta + C_1\omega(s,u)^\theta
\end{aligned}$$

$$
\begin{aligned}
&\le C_2\omega(s,u)^\theta \left[\frac{1}{2^\theta} + \frac{2^{\theta-1}}{2^\theta} + \frac{2^{\theta-1}-1}{2^\theta}\right] \\
&= C_2\omega(s,u)^\theta,
\end{aligned}
$$

which finishes the induction. It follows that we have

$$
\begin{aligned}
d_g\left(\varphi^{\mathbf{x},V,\pi_n}_{t\leftarrow s}(y), y\right) &\le d_g\left(\varphi^{\mathbf{x},V,\pi_n}_{t\leftarrow s}(y), \varphi^{\mathbf{x},V,\pi_0}_{t\leftarrow s}(y)\right) + d_g\left(\varphi^{\mathbf{x},V,\pi_0}_{t\leftarrow s}(y), y\right) \\
&\le C_2\omega(s,t)^\theta + C_3\omega(s,t)^{1/p}
\end{aligned}
$$

for some constant $C_3 > 0$ as long as $|t-s| < \delta$. Because of the way we have constructed the partitions π_n, for sufficiently large $n \in \mathbb{N}$ we can find $\tilde{n} < n$ and $0 = \tau_0 < \cdots < \tau_m = T$ such that

$$
\varphi^{\mathbf{x},V,\pi_n}_{T\leftarrow 0} = \varphi^{\mathbf{x},V,\pi_{\tilde{n}}}_{\tau_m\leftarrow\tau_{m-1}} \circ \cdots \circ \varphi^{\mathbf{x},V,\pi_{\tilde{n}}}_{\tau_1\leftarrow\tau_0}
$$

and

$$
\sup_{i\in\{1,\dots,m\}} |\tau_i - \tau_{i-1}| \le \delta.
$$

Hence we obtain

$$
\begin{aligned}
& d_g\left(\varphi^{\mathbf{x},V,\pi_n}_{T\leftarrow 0}(y), y\right) \\
\le& \sum_{i=1}^m d_g\left(\varphi^{\mathbf{x},V,\pi_{\tilde{n}}}_{\tau_i\leftarrow\tau_{i-1}} \circ \cdots \circ \varphi^{\mathbf{x},V,\pi_{\tilde{n}}}_{\tau_1\leftarrow\tau_0}(y), \varphi^{\mathbf{x},V,\pi_{\tilde{n}}}_{\tau_{i-1}\leftarrow\tau_{i-2}} \circ \cdots \circ \varphi^{\mathbf{x},V,\pi_{\tilde{n}}}_{\tau_1\leftarrow\tau_0}(y)\right) \\
\le& \sum_{i=1}^m \left[C_2\omega(\tau_{i-1},\tau_i)^\theta + C_3\omega(\tau_{i-1},\tau_i)^{1/p}\right] < \infty,
\end{aligned}
$$

which finishes the proof since the last line depends neither on n nor on y. ■

Corollary 5.20 *If $V = (V_1, \dots, V_d)$ is Lipschitz continuous with respect to a complete metric g, then for each $y \in \mathcal{M}$ there exists a compact set $K(y)$ which contains $\varphi^{\mathbf{x},V,\pi_n}_{t\leftarrow 0}(y)$ for all $n \in \mathbb{N}$ and all $t \in [0,T]$.*

Proof. It follows from Proposition 5.19 that we can choose $K(y)$ to be a closed disc around y with respect to d_g and with sufficiently large radius. Since g is complete, this is indeed a compact set. ■

Proof of Theorem 5.16. Fix $y_0 \in \mathcal{M}$ and let $W = (W_1, \dots, W_d)$ be a collection of compactly supported smooth vector fields on $\mathcal{M}$ which satisfy

$$W_i|_{K(y_0)} = V_i|_{K(y_0)}$$

for all $i \in \{1, \dots, d\}$ and for $K(y_0)$ as in Corollary 5.20. Since the elements of W are smooth and compactly supported it follows that

$$\mathrm{d}y_t = W(y_t)\mathrm{d}\mathbf{x}_t$$

has a global solution for every geometric rough path $\mathbf{x}$ and hence we can consider the corresponding solution flow $\Phi^{\mathbf{x},W}$ on $\mathcal{M}$. We have to prove that $\Phi^{\mathbf{x},W}_{t\leftarrow 0}(y_0) \in K(y_0)$ for all $t \in [0,T]$ and we will do this by showing that $\varphi^{\mathbf{x},W,\rho}_{t\leftarrow 0}(y_0)$ converges to $\Phi^{\mathbf{x},W}_{t\leftarrow 0}(y_0)$ as $|\rho| \to 0$. The claim will then follow by considering the sequence of partitions π_n from above.

Fix $t \in [0,T]$ and let ρ be a partition of $[0,t]$ which is given by

$$\rho = \{0 = \tau_0 < \cdots < \tau_r = t\}.$$

Define ρ_i to be the partition on $[0, \tau_i]$ which is given by $\{\tau_0 < \cdots < \tau_i\}$. Finally, since $K(y_0)$ is compact, it can be covered by finitely many coordinate neighbourhoods and thus we can assume without loss of generality that it is contained in a single coordinate neighbourhood, i.e. we can assume that $\mathcal{M} = \mathbf{R}^k$. Thus we have

$$\begin{aligned}
&\left|\Phi^{\mathbf{x},W}_{t\leftarrow 0}(y_0) - \varphi^{\mathbf{x},W,\rho}_{t\leftarrow 0}(y_0)\right| \\
=&\left|\sum_{i=1}^{r}\left(\Phi^{\mathbf{x},W}_{t\leftarrow\tau_i}\left(\Phi^{\mathbf{x},W}_{\tau_i\leftarrow\tau_{i-1}}\left(\varphi^{\mathbf{x},W,\rho_{i-1}}_{\tau_{i-1}\leftarrow 0}(y_0)\right)\right) - \Phi^{\mathbf{x},W}_{t\leftarrow\tau_i}\left(\phi_1^{W_{\log(\mathbf{x}_{\tau_{i-1},\tau_i})}}\left(\varphi^{\mathbf{x},W,\rho_{i-1}}_{\tau_{i-1}\leftarrow 0}(y_0)\right)\right)\right)\right| \\
\leq& L\sum_{i=1}^{r}\left|\left(\Phi^{\mathbf{x},W}_{\tau_i\leftarrow\tau_{i-1}}\left(\varphi^{\mathbf{x},W,\rho_{i-1}}_{\tau_{i-1}\leftarrow 0}(y_0)\right) - \phi_1^{W_{\log(\mathbf{x}_{\tau_{i-1},\tau_i})}}\left(\varphi^{\mathbf{x},W,\rho_{i-1}}_{\tau_{i-1}\leftarrow 0}(y_0)\right)\right)\right| \\
\leq& L\sum_{i=1}^{r}\bar{\theta}_{\mathrm{id}_{\mathbf{R}^n}}(\omega(\tau_{i-1},\tau_i)) = L\sum_{i=1}^{r}\omega(\tau_{i-1},\tau_i)\frac{\bar{\theta}_{\mathrm{id}_{\mathbf{R}^n}}(\omega(\tau_{i-1},\tau_i))}{\omega(\tau_{i-1},\tau_i)} \\
\leq& L\omega(0,T)\sup_{i\in\{1,\dots,r\}}\frac{\bar{\theta}_{\mathrm{id}_{\mathbf{R}^n}}(\omega(\tau_{i-1},\tau_i))}{\omega(\tau_{i-1},\tau_i)}.
\end{aligned}$$

The first inequality holds because W is compactly supported and hence $\Phi^{\mathbf{x},W}_{t\leftarrow\tau_i}$ is Lipschitz continuous with some Lipschitz constant L. The second inequality follows from Lemma 5.15, which also implies that

$$\sup_{i\in\{1,\dots,r\}} \frac{\bar{\theta}_{\mathrm{id}_{\mathbf{R}^n}}(\omega(\tau_{i-1},\tau_i))}{\omega(\tau_{i-1},\tau_i)} \to 0$$

as $|\rho| \to 0$. (Note that Lemma 5.15 can be applied with $f = \mathrm{id}_{\mathbf{R}^n}$ even though $\mathrm{id}_{\mathbf{R}^n}$ is not compactly supported, because W itself is compactly supported.) In particular we have

$$\Phi^{\mathbf{x},W}_{t\leftarrow 0}(y_0) = \lim_{n\to\infty} \varphi^{\mathbf{x},W,\pi_n}_{t\leftarrow 0}(y_0) = \lim_{n\to\infty} \varphi^{\mathbf{x},V,\pi_n}_{t\leftarrow 0}(y_0) \in K(y_0)$$

and therefore $\Phi^{\mathbf{x},V}_{t\leftarrow 0}(y_0)$ is well-defined and we have $\Phi^{\mathbf{x},V} = \Phi^{\mathbf{x},W}$, which finishes the proof. ∎

5.2.2. Equations with drift term

For a collection of smooth vector fields $V = (V_1, \dots, V_d)$ and another smooth vector field V_0 consider the equation

$$\mathrm{d}y_t = V_0(y_t)\mathrm{d}t + V(y_t)\mathrm{d}\mathbf{x}_t \tag{RDEd}$$

In principle, there is no need to pay special attention to the drift term. The reason is that for any given rough path $\mathbf{x}$ over $\mathbf{R}^d$ we can consider its underlying $\mathbf{R}^d$-valued path x and construct a rough path lift $\hat{\mathbf{x}}$ over $\mathbf{R}^{d+1}$ of the path $t \mapsto (t, x_t)$, such that the canonical projections of $\hat{\mathbf{x}}$ to $\widehat{T}(\mathbf{R}^d)$ coincide with $\mathbf{x}$. The lift $\hat{\mathbf{x}}$ exists and is uniquely determined by $\mathbf{x}$ since the missing terms are obtained by integrating $\mathbf{x}$ repeatedly (with respect to the path $t \mapsto t$). Thus if we define $\hat{V} := (V_0, V_1, \dots, V_d)$, then (RDEd) is a rough differential equation driven by $\hat{\mathbf{x}}$ along $\hat{V}$ in the usual sense.

However, if we discuss the drift vector field separately, we can impose considerably weaker assumptions on it than if we consider it as one of the diffusion vector fields.

Lemma 5.21 *Let $\mathbf{x}$ be a geometric p-rough path over $\mathbf{R}^d$ controlled by ω and let $V = (V_1, \dots, V_d)$ be a collection of smooth vector fields on $\mathcal{M}$. Let V_0 be another smooth vector field on $\mathcal{M}$. Then $y : [0, T] \to \mathcal{M}$ solves (RDEd) if and only if for each compactly supported smooth test function $f : \mathcal{M} \to \mathbf{R}$ there exists a function $\theta_f : \mathbf{R} \to \mathbf{R}$ with $\theta_f(h)/h \to 0$ for $h \to 0$ such that*

$$\left| f(y_t) - (t-s)V_0 f(y_s) - V_{\mathbf{x}_{s,t}} f(y_s) \right| \le \theta_f(\omega(s,t))$$

holds for all $0 \le s \le t \le T$.

Proof. Assume that y solves (RDEd) and define $\hat{V} := (V_0, V_1, \dots, V_d)$. Then we have

$$\left| f(y_t) - (t-s)V_0 f(y_s) - V_{\mathbf{x}_{s,t}} f(y_s) \right|$$

$$\leq \left| f(y_t) - \hat{V}_{\hat{\mathbf{x}}_{s,t}} f(y_s) \right| + \left| \hat{V}_{\hat{\mathbf{x}}_{s,t}} f(y_s) - (t-s)V_0 f(y_s) - V_{\mathbf{x}_{s,t}} f(y_s) \right|.$$

and the first term satisfies the desired bound because of Definition 5.13 and the second term does so as well because it only consists of cross-terms between $\mathbf{x}$ and the smooth path since all other terms get cancelled.

On the other hand we have

$$\begin{aligned} &\left| f(y_t) - \hat{V}_{\hat{\mathbf{x}}_{s,t}} f(y_s) \right| \\ &\leq \left| f(y_t) - (t-s)V_0 f(y_s) - V_{\mathbf{x}_{s,t}} f(y_s) \right| + \left| (t-s)V_0 f(y_s) + V_{\mathbf{x}_{s,t}} f(y_s) - \hat{V}_{\hat{\mathbf{x}}_{s,t}} f(y_s) \right|. \end{aligned}$$

and thus the converse direction follow from the same argument. ■

Corollary 5.22 (cf. Lemma 5.15) *Let $\mathbf{x}$ be a geometric p-rough path over $\mathbf{R}^d$ which is controlled by a control ω and let $V = (V_1, \ldots, V_d)$ be a collection of smooth vector fields on $\mathcal{M}$. Let V_0 be another smooth vector field on $\mathcal{M}$. Then for each compactly supported smooth test function $f : \mathcal{M} \to \mathbf{R}$ there exists $\bar{\theta}_f : \mathbf{R} \to \mathbf{R}$ with $\bar{\theta}_f(h)/h \to 0$ for $h \to 0$ such that*

$$\left\| f \circ \Phi^{\hat{\mathbf{x}},\hat{V}}_{t \leftarrow s} - f \circ \phi_1^{(t-s)V_0 + V_{\log(\mathbf{x}_{s,t})}} \right\|_\infty \leq \bar{\theta}_f(\omega(s,t))$$

holds for all $0 \leq s \leq t \leq T$.

Proof. Using the triangle inequality we obtain

$$\begin{aligned} &\left\| f \circ \Phi^{\hat{\mathbf{x}},\hat{V}}_{t \leftarrow s} - f \circ \phi_1^{(t-s)V_0 + V_{\log(\mathbf{x}_{s,t})}} \right\|_\infty \\ &\leq \left\| f \circ \Phi^{\hat{\mathbf{x}},\hat{V}}_{t \leftarrow s} - (t-s)V_0 f - V_{\mathbf{x}_{s,t}} f \right\|_\infty + \left\| (t-s)V_0 f + V_{\mathbf{x}_{s,t}} f - f \circ \phi_1^{(t-s)V_0 + V_{\log(\mathbf{x}_{s,t})}} \right\|_\infty. \end{aligned}$$

It follows directly from Lemma 5.21 that the first term satisfies the desired bound. For the second term, consider the Taylor expansion of the map

$$\tau \mapsto \phi_\tau^{(t-s)V_0 + V_{\log(\mathbf{x}_{s,t})}}$$

around $\tau = 0$ as in the proof of Lemma 5.15. ■

Theorem 5.23 *Let V_0 be a smooth vector field on $\mathcal{M}$ and let $V = (V_1, \ldots, V_d)$ be a collection of smooth vector fields on $\mathcal{M}$ such that $\mathcal{L}^{(k)}(V) \cup \{V_0\}$ is Lipschitz continuous with respect to a complete metric g. Then the rough differential equation (RDEd) has a global solution for every geometric p-rough path $\mathbf{x}$ with $[p] \leq k$ and every initial value $y_0 \in \mathcal{M}$.*

Proof. With the appropriate amendments we can use the same arguments that have lead to Theorem 5.16. We sketch the necessary changes.

First we have to modify our Trotter-type approximations (5.13) so that they involve the drift vector field in the correct way. Thus define

$$\varphi_{t\leftarrow s}^{\mathbf{x},V,\rho} := \phi_1^{(\tau_r-\tau_{r-1})V_0+V_{\log(\mathbf{x}_{\tau_{r-1},\tau_r})}} \circ \cdots \circ \phi_1^{(\tau_1-\tau_0)V_0+V_{\log(\mathbf{x}_{\tau_0,\tau_1})}}.$$

for a partition $\rho = \{s = \tau_0, \ldots, \tau_r = t\}$ of $[s,t]$. Furthermore let $(\pi_n)_{n\in\mathbf{N}}$ be the sequence of partitions given by (5.11) and (5.12). We have to show that $d_g\left(\varphi_{T\leftarrow 0}^{\mathbf{x},V,\pi_n}(y), y\right)$ is bounded uniformly in $n \in \mathbf{N}$ and $y \in \mathcal{M}$. Therefore we modify Proposition 5.9 in the following way. Assume that A and B are Lie series in the variables $V_0, V_1, \ldots, V_d$. Then we denote by $\mathrm{BCH}_{1,k}(A,B)$ the Lie polynomial which is obtained by truncating all terms of degree $k+1$ or higher as well as all terms which involve V_0 and are of degree 2 or higher. Thus we can formulate and prove a version of Proposition 5.9 where BCH_k is replaced by $\mathrm{BCH}_{1,k}$ and which leads to additional remainder terms involving V_0. This allows us to adapt Proposition 5.12 accordingly and thus we obtain the boundedness of $d_g\left(\varphi_{T\leftarrow 0}^{\mathbf{x},V,\pi_n}(y), y\right)$ just as in Proposition 5.19. Finally we conclude as in the proof of Theorem 5.16, albeit by referring to Corollary 5.22 instead of Lemma 5.15. ∎

5.2.3. Examples

Example 5.24 (Additive noise) Consider the vector field

$$V : \mathbf{R}^2 \to \mathbf{R}^2 : y \mapsto |y|^\theta \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} y$$

for $\theta \geq 0$. Then the equation

$$\mathrm{d}y_t = V(y_t)\mathrm{d}t + \mathrm{d}x_t \tag{5.15}$$

has a global solution y for every initial value $y_0 \in \mathbf{R}^2$ and for every (rough) path x if and only if we have $\theta \leq 1$.

Remark 5.25 This type of example is also captured by [RS17, Thm. 4.3]. However, their result only shows that global solutions exist for $\theta = 0$ and hence our claim is more general.

Proof of Example 5.24. In [CHJ13] the authors construct an x which causes explosion in finite time under the assumption that $\theta = 2$. We present their construction in a generalised form.

Assume $\theta > 1$, fix $\delta \in (1, \theta)$ and consider the solution z of the ordinary differential equation

$$\dot{z}_t = V\left(z_t - \frac{1}{1 + |z_t|^\delta}\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix} z_t\right).$$

In polar coordinates this reads as

$$\dot{r}_t = \frac{r_t^{1+\theta}}{\left(1 + r_t^\delta\right)^{1+\theta}}\left(1 + \left(1 + r_t^\delta\right)^2\right)^{\theta/2} \tag{5.16}$$

$$\dot{\varphi}_t = \frac{r_t^{\theta}}{\left(1 + r_t^\delta\right)^{\theta}}\left(1 + \left(1 + r_t^\delta\right)^2\right)^{\theta/2}, \tag{5.17}$$

where the right-hand side of (5.16) grows asymptotically like $r_t^{1+\theta-\delta}$, i.e. the solution r_t explodes in finite time for every initial value different from 0. For a fixed initial value z_o, denote the explosion time by τ and define the continuous path

$$x : \mathbf{R} \to \mathbf{R}^2 : t \mapsto \begin{cases} -\frac{1}{1+|z_t|^\delta}\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix} z_t, & t \in [0, \tau)\\ 0, & t \in [\tau, \infty).\end{cases}$$

Then $y := z + x$ solves (5.15) with initial value $y_0 = z_0 + x_0$ and explodes at the finite time τ. The equation for y_0 can be solved for z_0 and thus we can obtain an example for every non-zero initial value y_0. Figure 5.1 shows the solution y and the driver x for an exemplary choice of parameters.

Let us show that x is Hölder continuous and can thus be lifted to a rough path (even though the actual choice of the lift is irrelevant because we are dealing with additive noise). From (5.16) it follows that the radial component r of z behaves asymptotically like $(t-\tau)^{-\frac{1}{\theta-\delta}}$ around τ. Hence (5.17) implies that the tangential component φ behaves like $(t - \tau)^{-\frac{\delta}{\theta-\delta}}$ around τ. Instead of analysing the Hölder continuity of x directly we can simplify the problem and consider the map

$$\tilde{x} : [0, 1] \to \mathbf{R}^2 : t \mapsto t^{\frac{\delta-1}{\theta-\delta}}\begin{pmatrix}\cos\left(t^{-\frac{\delta}{\theta-\delta}}\right)\\ \sin\left(t^{-\frac{\delta}{\theta-\delta}}\right)\end{pmatrix}$$

instead. Let $s, t \in [0, 1]$. Without loss of generality we can assume that there exists $n \in \mathbf{N}$ such that

$$s = (2\pi n + \bar{s})^{-\frac{\theta-\delta}{\delta}} \quad \text{and} \quad t = (2\pi n + \bar{t})^{-\frac{\theta-\delta}{\delta}}$$

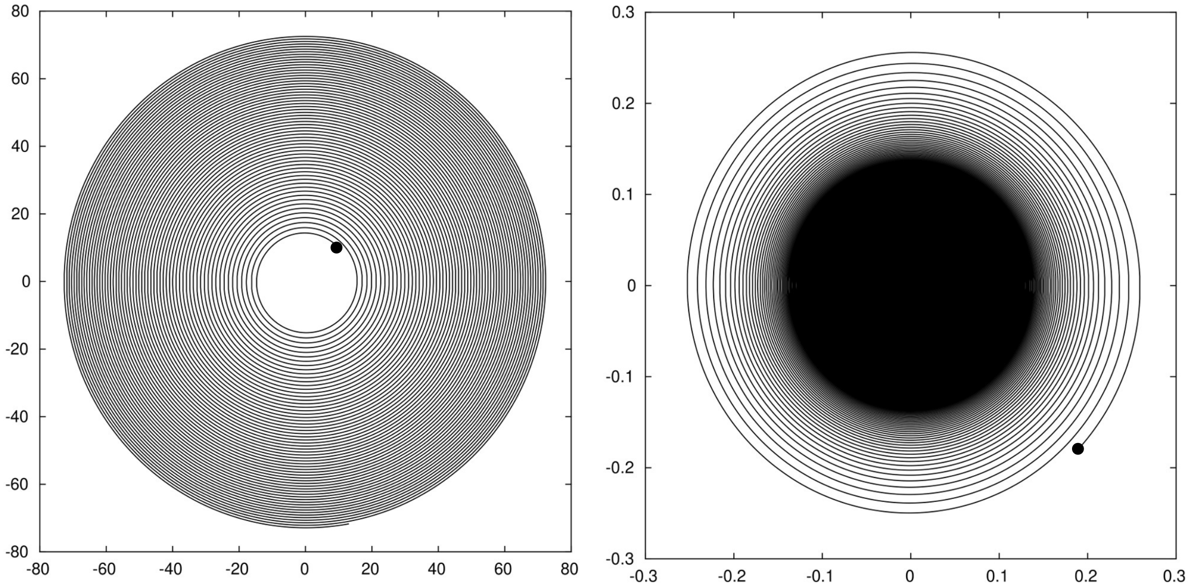


Figure 5.1.: The solution y for $t \in [0, 0.3]$ (left) and the driver x for $t \in [0, 0.5]$ (right) with $\theta = 2$, $\delta = 1.5$ and initial value $y_0 = (10, 10)$. The singularity occurs at $\tau \approx 0.534$. The highlighted points correspond to $t = 0$.

for $\bar{s}, \bar{t} \in [0, 2\pi]$. Then the mean value theorem implies that there exists $C_1 > 0$ which does not depend on s or t such that we have

$$|\tilde{x}_t - \tilde{x}_s| \leq C_1 n^{-\frac{\delta-1}{\delta}} |\bar{t} - \bar{s}| \leq 2\pi C_1 n^{-\frac{\delta-1}{\delta}} |\bar{t} - \bar{s}|^{\frac{\delta-1}{\theta}}. \tag{5.18}$$

On the other hand, again using the mean value theorem, we find that there exists $C_2 > 0$ such that we have

$$|t - s| \geq \frac{\left|(2\pi n + \bar{s})^{\frac{\theta-\delta}{\delta}} - (2\pi n + \bar{t})^{\frac{\theta-\delta}{\delta}}\right|}{(2\pi(n+1))^{\frac{2(\theta-\delta)}{\delta}}} \geq C_2 n^{-\frac{\theta}{\delta}} |\bar{s} - \bar{t}|$$

or equivalently

$$|t - s|^{\frac{\delta-1}{\theta}} \geq C_2^{\frac{\delta-1}{\theta}} n^{-\frac{\delta-1}{\delta}} |\bar{s} - \bar{t}|^{\frac{\delta-1}{\theta}}. \tag{5.19}$$

Thus combining (5.18) and (5.19) we find that $\tilde{x}$ (and therefore x) is Hölder continuous with exponent $\frac{\delta-1}{\theta} \in (0, 1)$. In particular, by choosing θ and δ to be large, we obtain examples which are arbitrarily close to having bounded variation.

Let us now discuss the case where $\theta \leq 1$. In polar coordinates the vector field V is given by $V(r, \varphi) = (0, r^\theta)$ and the standard Euclidean metric of $\mathbf{R}^2$ is given by $\mathrm{d}r + r^2 \mathrm{d}\varphi$. Choose a smooth function $\rho : \mathbf{R}_{\geq 0} \to \mathbf{R}_{\geq 0}$ such that $\rho(t) = t$ when t is close to 0 and $\rho(t) = 1$ for $t \geq 1$.

Then define the metric

$$\tilde{g}(r, \varphi) := \mathrm{d}r + \rho(r^2)\mathrm{d}\varphi,$$

which is easily seen to be complete. The covariant derivative of V (in polar coordinates) with respect to $\tilde{g}$ is given by

$$\nabla V(r, \varphi) = \begin{pmatrix} 0 & 0 \\ \theta r^{\theta-1} & 0 \end{pmatrix}$$

for $r \geq 1$, which is certainly bounded. If r is close to zero, then $\tilde{g}$ is equal to the standard Euclidean metric and therefore one can directly compute the covariant derivative in Cartesian coordinates to verify that it is bounded. Furthermore, the canonical coordinate vector fields in $\mathbf{R}^2$ transform into

$$\begin{aligned} (r, \varphi) &\mapsto \left(\cos(\varphi), -\frac{\sin(\varphi)}{r}\right) \\ \text{and } (r, \varphi) &\mapsto \left(\sin(\varphi), \frac{\cos(\varphi)}{r}\right), \end{aligned}$$

when we write them in polar coordinates. Thus their covariant derivatives with respect to $\tilde{g}$ are certainly bounded. Hence we have shown that the collection consisting of V and the coordinate vector fields of $\mathbf{R}^2$ is Lipschitz continuous with respect go $\tilde{g}$. Therefore it follows from Theorem 5.23 that (5.15) has a global solution for every driving rough path and every initial value if $\theta \leq 1$. ■

Remark 5.26 This example also shows that the Lipschitz assumption with respect to g in Theorems 5.16 and 5.23 cannot be replaced by a mere boundedness assumption since V can be bounded with respect to a complete metric by rescaling the canonical metric in the tangential direction, no matter how large θ is.

One can even consider an isometric embedding of $\mathbf{R}^2$ (equipped with a metric which bounds V) into $\mathbf{R}^k$ for some $k \in \mathbf{N}$. This way one obtains examples of RDEs driven along bounded vector fields by paths which are almost of bounded variation whose solution explodes in finite time.

Example 5.27 (Commuting vector fields) Let $V = (V_1, \ldots, V_d)$ be a collection of complete vector fields on some manifold $\mathcal{M}$ such that all pairwise Lie brackets $[V_i, V_j]$ vanish for $i, j \in \{1, \ldots, d\}$. Then it follows from Proposition 5.4 and Theorem 5.16 that a rough differential equation along the vector fields V_i^+ on $\mathcal{M}^+ = \mathcal{M} \times \mathbf{R}^d$ has a global solution for every driving rough path and every initial value. Thus we can consider the canonical projection from $\mathcal{M}^+$ onto $\mathcal{M}$ to obtain the solution of the equation driven along the original vector fields V.

One can of course construct a solution for commuting vector fields much easier by simply concatenating the flows along the involved vector fields, see e.g. [LQ02]. However, our point is that commuting vector fields are also captured by Theorem 5.16, thus underlining its universality.

Example 5.28 (Linear vector fields) Let $V = (V_1, \ldots, V_d)$ be a collection of linear vector fields on $\mathbf{R}^n$. Then the iterated Lie brackets of any order of V are also linear. Thus for all $k \in \mathbf{N}$ all elements of $\mathcal{L}^{(k)}(V)$ are Lipschitz continuous with respect to the canonical Euclidean metric on $\mathbf{R}^n$ (recall however that for our argument we have to use the rescaled metric from Proposition 5.2 so that the vector fields become bounded). Hence we can apply Theorem 5.16 and we find that rough differential equations along linear vector fields always have global solutions for every driving path and every initial value.

Example 5.29 Let $V = (V_1, V_2)$ be as in Example 2.44 and let $\mathbf{x}$ be the pure area rough path as in Example 3.7. Then the RDE

$$\mathrm{d}y_t = V(y_t)\mathrm{d}\mathbf{x}_t$$

is equivalent to the ODE

$$\dot{y}_t = [V_1, V_2](y_t) = e^{y_t},$$

which is does not have a global solution for any initial value y_0. Therefore the conditions which Theorem 5.16 imposes on the Lie brackets are necessary, even in this case where we have exactly $p = 2$.

Example 5.30 In [LS11] the authors construct a bounded vector field $V : \mathbf{R}^2 \to \mathbf{R}^2$ such that the solution of

$$\mathrm{d}y_t = V(y_t)\mathrm{d}\mathbf{x}_t$$

explodes in finite time (at least for certain initial values) if $\mathbf{x}$ is the Itō lift of some carefully chosen one-dimensional Brownian sample path. The set of sample paths with this property has positive measure. This illustrates that the situation for general branched rough paths is more difficult than in the geometric case. In contrast, [LS11] also contains a Stratonovich example, however one needs two vector fields in this case and accordingly the noise is two-dimensional.

6. Discussion and open problems

There are two important benchmarks that should be considered in order to gauge Theorems 5.16 and 5.23. On the one hand, there is the limit case $p = 1$ where the RDE becomes an ODE, on the other hand there is the case where the driving path is a Brownian sample path.

In the case $p = 1$ our result requires all vector fields to be (jointly) Lipschitz continuous with respect to some complete metric g. However, it is clear that boundedness is actually sufficient in this case. Comparing this with Remark 5.26 indicates that there is a qualitative difference between $p > 1$ and $p = 1$. A closer look at Example 5.29 reveals that a similar phenomenon happens at other integral values of p.

For stochastic differential equations there are two different notions of global solvability, namely weak and strong completeness. An SDE is said to be *weakly complete* if for all fixed initial vales there exists a global solution almost surely, albeit the set of measure 1 on which this happens may depend on the initial value. In contrast, an SDE is called *strongly complete* if the set of sample paths for which a global solution exists can be chosen independently of the initial value. Because of the pathwise nature of RDEs, it does not make sense to compare our results with criteria that guarantee weak completeness. Hence we focus on strong completeness. Arguably the most general condition for strong completeness has been given in [Li94], see also [Kun90]. Most importantly, it is sufficient to impose first order conditions, i.e. to control vector fields and their derivatives. In contrast, Thereom 5.16 requires a second order condition, namely controlling the Lipschitz constant of the Lie brackets, since for Brownian rough paths we have $p > 2$. This hints towards possible future improvements of Theorem 5.16. We will come back to this thought below.

Example 5.24 nicely illustrates how the existence of global solutions for rough differential equations is governed by the interaction of the vector fields rather than their individual behaviour. Indeed, even if we have $\theta > 1$ the drift vector field by itself is “harmless” and its integral curves are circles around the origin. On the other hand, we have seen that a carefully chosen additive (rough) path leads to explosion in finite time. By increasing θ we can even construct such a path with Hölder exponent arbitrarily close to 1. Therefore, even if p is very close to 1, the Lipschitz assumption in Theorems 5.16 and 5.23 cannot be relaxed in general.

However, some important remarks about the driving path x in Example 5.24 are in order. It differs from many paths that show up in stochastic applications in several (related) aspects. For instance it is not truly rough in the sense of [FS13] or θ-Hölder rough in the sense of [HP13]. In fact, the path has bounded variation on all closed intervals which do not contain the explosion time τ. Thus the Lévy area away from τ is given by a standard Riemann-Stieltjes integral. A direct computation shows that this Lévy area is unbounded as one approaches τ. This implies in particular that the path does not have a *natural lift* to a rough path in the sense that it cannot be obtained through piecewise linear or mollifier approximations as in [FV10, Thm. 15.33]. Note that similar comments apply to the pure-area rough path used in Example 5.29.

In light of the previous paragraphs an interesting problem is to find classes of rough paths for which Theorem 5.16 can be relaxed. For instance, *is it possible to relax the assumptions on the vector fields in Theorem 5.16 if we require that the driving path be truly rough?* Furthermore, in a stochastic context it may be fruitful to use more subtle properties of the sample paths in the spirit of [CLL13]. Also, the methods of [FV10, Sect. 10.5] could help to obtain a refined result. These approaches might lead to a better understanding of strong completeness of stochastic differential equations from a rough path perspective.

Another starting point for future research might be [RS17] which establishes the existence of global forward-solutions under a one-sided growth condition on the drift vector field. After introducing a sensible notion of "one-sided growth" on manifolds, it should be possible to generalise their results and embed them into our framework from Chapter 5.

Another open problem is the formulation of explicit global estimates for the solutions. In most applications the underlying manifold is actually given by $\mathbf{R}^n$ and one is interested in estimating the Euclidean norm of the solution. However, the generality of our approach leads to a difficulty here because Theorem 5.16 captures vector fields which lead to arbitrarily fast growing solutions (with respect to the Euclidean norm). Indeed, consider the ODEs $\dot{y}_t = y_t$, $\dot{y}_t = y_t \log(1+y_t^2)$, $\dot{y}_t = y_t \log(1+y_t^2)\log(\log(1+y_t^2))$, etc. They are all covered by Theorem 5.16 but their solutions grow asymptotically like $\exp(t)$, $\exp(\exp(t))$, $\exp(\exp(\exp(t)))$, etc. Hence there is no universal bound on the growth.

Let us finally see how our results might lead to a better understanding of the existence of smooth densities for equations driven by stochastic (usually Gaussian) rough paths in the spirit of (Malliavin's interpretation of) Hörmander's theorem. The main reference in this direction is [CHLT15]. Their central result [CHLT15, Thm. 3.5] assumes that the involved vector fields are bounded and have bounded derivatives of all order. This raises two questions, namely whether such a strong assumption is necessary and how it can be reformulated in order to obtain an analogous result on manifolds.

These questions are interesting even in the classical setting of stochastic differential equations driven by Brownian motion. In [Tan83] Malliavin's work is generalised to manifold-valued differential equations. The standing assumption is that the manifold can be embedded into $\mathbf{R}^k$ such that all derivatives of the embedded vector fields are bounded. This "brute force" condition is difficult to verify and it does not reveal the underlying geometric structure. For instance, it does not clarify the role of the global behaviour of the Lie brackets. In [Dri04] the author assumes that all manifolds under consideration are compact while suggesting the condition from [Tan83] as an alternative.

We conclude that a detailed analysis of the global requirements on the vector fields for the existence of densities (be it for SDEs or RDEs) is lacking so far and our results from Chapter 5 might help to resolve this problem.

A. Graded bialgebras are Hopf algebras

As we have seen in Definition 2.2 a Hopf algebra is a bialgebra with an additional map, the antipode. In light of Remark 2.3 we will give a general construction for an antipode that applies to all bialgebras that play a role in the present work.

Definition A.1 A bialgebra $(H, \star, \Delta)$ is called *graded* if the underlying vector space is graded and the grading

$$H = \bigoplus_{k=0}^{\infty} H^k$$

satisfies

$$H^0 = \mathbf{R}, \quad H^j \star H^k \subset H^{j+k} \quad \text{and} \quad \Delta\left(H^k\right) = \bigoplus_{i=0}^{k} H^i \otimes H^{k-i}$$

for all $j, k \in \mathbf{N}$.

If H is a graded bialgebra then the concatenation of the counit and the unit $\eta \circ \varepsilon : H \longrightarrow H$ is the orthogonal projection π^0 onto H^0. Thus we can write the defining property of the antipode as $S\left(h_{(1)}\right) \star h_{(2)} = h_{(1)} \star S\left(h_{(2)}\right) = \pi_0(h)$.

Lemma A.2 *Let $(H, \star, \Delta)$ be a graded bialgebra. Then there exists a unique antipode S such that $(H, \star, \Delta, S)$ is a Hopf algebra.*

Proof. We use the idea of [FGB01]. For two linear maps $f, g : H \longrightarrow H$ we define the *convolution* $f * g : H \longrightarrow H$ which is given by

$$(f * g)(h) = f\left(h_{(1)}\right) \star g\left(h_{(2)}\right).$$

It follows from the (co)associativity of $\star$ and Δ that $*$ is associative. Furthermore $\pi_0 : H \longrightarrow H$ is the identity element for $*$ and thus the convolution turns the space of endomorphisms on H into a unital algebra. The antipode S is the unique inverse id_H^{*-1} of the identity map if it exists.

It remains to show that id_H^{*-1} exists. To this end, expand id_H^{*-1} into the formal geometric series

$$\mathrm{id}_H^{*-1} = (\pi_0 - (\pi_0 - \mathrm{id}_H))^{*-1} = \sum_{k=0}^{\infty} (\pi_0 - \mathrm{id}_H)^{*k}.$$

The key observation is that $(\pi_0 - \mathrm{id}_H)\left(H^0\right) = \{0\}$ and hence for a fixed homogeneous element $h \in H^n$ with $n \in \mathbf{N}$ we have

$$\left(\mathrm{id}_H^{*-1}\right)(h) = \sum_{k=0}^{n} (\pi_0 - \mathrm{id}_H)^{*k}$$

because the remaining terms vanish. Thus we have shown that $S := \mathrm{id}_H^{*-1}$ is well-defined for homogeneous elements and therefore for all elements by linearity. ∎

B. Differentiability of the Riemannian distance

Let $\mathcal{M}$ be a smooth manifold with a complete Riemannian metric g. For two points $y, z \in \mathcal{M}$ with $y \neq z$ denote by $l := d_g(y, z)$ their distance and by $\gamma(z, y)$ the set of all unit-speed length-minimising curves $c : [0, l] \to \mathcal{M}$ with $c(0) = z$ and $c(l) = y$.

Proposition B.1 *Let $\mathcal{M}$ be a smooth manifold and let g be a Riemannian metric on $\mathcal{M}$. For fixed $z \in \mathcal{M}$ the map*

$$f_z : \mathcal{M} \to \mathbf{R} : y \mapsto d_g(y, z)$$

has one-sided directional derivatives at all points $y \in \mathcal{M}$. Furthermore, for $v \in T_z\mathcal{M}$ we have

$$vf_z = |v|_g$$

and for $y \in \mathcal{M}$ with $l := d_g(y, z) \neq 0$ and $v \in T_y\mathcal{M}$ we have

$$vf_z = \min\{\langle v, \dot{c}(l)\rangle_g : c \in \gamma(z, y)\}.$$

In particular, for all $v \in T\mathcal{M}$ the estimate

$$|vf_z| \leq |v|_g$$

holds.

Proof. We give a more detailed version of the sketched proof in [Iva10]. Let us first discuss differentiability at $y = z$. For a tangent vector $v \in T_z\mathcal{M}$ we have

$$vf_z = \lim_{t \downarrow 0} \frac{f_z(\exp_z(tv)) - f_z(z)}{t} = \lim_{t \downarrow 0} \frac{d_g(z, \exp_z(tv))}{t} = \lim_{t \downarrow 0} \frac{t|v|_g}{t} = |v|_g.$$

Now assume $y \neq z$ and set $l := d_g(y, z)$. Let furthermore $v \in T_y\mathcal{M}$. For $c \in \gamma(z, y)$ and $\varepsilon > 0$ define a smooth variation $c^v : [0, \varepsilon) \times [0, l] \to \mathcal{M}$ which satisfies $c^v(s, l) = \exp_y(sv)$

and $c^v(s,0) = z$ for all $s \in [0,\varepsilon)$. From [CE75, Chap. 1.1] we get for the length $L(c^v(s,\cdot))$ that

$$\frac{\partial}{\partial s} L(c^v(s,\cdot))|_{s=0+} = \langle v, \dot{c}(l)\rangle_g. \tag{B.1}$$

Since $\gamma(z,y)$ is compact we can choose $\bar{c} \in \gamma(z,y)$ such that

$$\langle v, \dot{\bar{c}}(l)\rangle_g = \min\{\langle v, \dot{c}(l)\rangle_g : c \in \gamma(z,y)\}.$$

Thus (B.1) implies that $L(\bar{c}^v(s,\cdot))$ decreases as fast as possible (or increases as slowly as possible) in s for small s for some variation of $\bar{c}$. Hence, for sufficiently small s, the variation $\bar{c}^v$ can be chosen such that $\bar{c}^v(s,\cdot)$ minimises the length between z and $\exp_y(sv)$, in other words such that $L(\bar{c}^v(s,\cdot)) = d_g(z, \exp_y(sv)) = f_z(\exp_y(sv))$. In that case we have

$$vf_z = \frac{\partial}{\partial s} f_z(\exp_y(sv))|_{s=0+} = \frac{\partial}{\partial s} L(\bar{c}^v(s,\cdot))|_{s=0+} = \langle v, \dot{\bar{c}}(l)\rangle_g,$$

which finishes the proof. ∎

References

[ABT15] J. Angst, I. Bailleul, and C. Tardif, *Kinetic Brownian motion on Riemannian manifolds*, Electron. J. Probab. **20** (2015).

[AFLMR07] D. Azagra, J. Ferrera, F. López-Mesas, and Y. Rangel, *Smooth approximation of Lipschitz functions on Riemannian manifolds*, J. Math. Anal. Appl. **326** (2007), no. 2, 1370–1378.

[Bai14] I. Bailleul, *Rough integrators on Banach manifolds*, arXiv:1403.3285 [math] (2014).

[Bai15] ———, *Flows driven by rough paths*, Rev. Mat. Iberoam. **31** (2015), no. 3, 901–934.

[Bai18] ———, *On the definition of a solution to a rough differential equation*, arXiv:1803.06479 [math] (2018).

[BC18] I. Bailleul and R. Catellier, *Non-explosion criteria for rough differential equations driven by unbounded vector fields*, arXiv:1802.04605 [math] (2018).

[BC17] H. Boedihardjo and I. Chevyrev, *An isomorphism between branched and geometric rough paths*, arXiv:1712.01965 [math] (2017).

[BGLY16] H. Boedihardjo, X. Geng, T. J. Lyons, and D. Yang, *The signature of a rough path: Uniqueness*, Adv. Math. **293** (2016), 720–737.

[Bou07] N. Bourbaki, *Éléments de mathématique. Espaces vectoriels topologiques. Chapitres 1 à 5*, reprint of the 1981 original ed., Springer, Berlin, 2007.

[BL15] Y. Boutaib and T. J. Lyons, *A new definition of rough paths on manifolds*, arXiv:1510.07833 [math] (2015).

[CDL15] T. Cass, B. K. Driver, and C. Litterer, *Constrained rough paths*, P. Lond. Math. Soc. **111** (2015), no. 6, 1471–1518.

[CHLT15] T. Cass, M. Hairer, C. Litterer, and S. Tindel, *Smoothness of the density for solutions to Gaussian rough differential equations*, Ann. Probab. **43** (2015), no. 1, 188–239.

[CLL11] T. Cass, C. Litterer, and T. J. Lyons, *Rough Paths on Manifolds*, New Trends in Stochastic Analysis and Related Topics, vol. Volume 12, Interdisciplinary Mathematical Sciences, no. Volume 12, WORLD SCIENTIFIC, 2011, pp. 33–88.

[CLL13] ———, *Integrability and tail estimates for Gaussian rough differential equations*, Ann. Probab. **41** (2013), no. 4, 3026–3050.

[CW17] T. Cass and M. P. Weidner, *Tree algebras over topological vector spaces in rough path theory*, arXiv:1604.07352 [math] (2017).

[CE75] J. Cheeger and D. G. Ebin, *Comparison theorems in Riemannian geometry*, North-Holland mathematical library, no. v. 9, North-Holland Pub. Co. ; American Elsevier Pub. Co, Amsterdam : New York, 1975.

[CL16] I. Chevyrev and T. J. Lyons, *Characteristic functions of measures on geometric rough paths*, Ann. Probab. **44** (2016), no. 6, 4049–4082.

[CHJ13] S. Cox, M. Hutzenthaler, and A. Jentzen, *Local Lipschitz continuity in the initial value and strong completeness for nonlinear stochastic differential equations*, arXiv:1309.5595 [math] (2013).

[CEFMMK18] C. Curry, K. Ebrahimi-Fard, D. Manchon, and H. Z. Munthe-Kaas, *Planarly branched rough paths and rough differential equations on homogeneous spaces*, arXiv:1804.08515 [math] (2018).

[DNR00] S. Dascalescu, C. Nastasescu, and S. Raianu, *Hopf Algebra: An Introduction*, CRC Press, 2000.

[Dav08] A. M. Davie, *Differential Equations Driven by Rough Paths: An Approach via Discrete Approximation*, Appl. Math. Res. Exp. (2008).

[DT13] S. Descombes and M. Thalhammer, *The Lie-Trotter splitting for nonlinear evolutionary problems with critical parameters: A compact local error representation and application to nonlinear Schrödinger equations in the semiclassical regime*, IMA J. Numer. Anal. **33** (2013), no. 2, 722–745.

[dC92] M. do Carmo, *Riemannian Geometry*, Birkhäuser, Boston, 1992.

[Dri04] B. K. Driver, *Curved Wiener Space Analysis*, Real and Stochastic Analysis, Trends in Mathematics, Birkhäuser Boston, 2004, pp. 43–198.

[DS17] B. K. Driver and J. S. Semko, *Controlled Rough Paths on Manifolds I*, Rev. Mat. Iberoam. **33** (2017), no. 3, 85–950.

[FGB01] H. Figueroa and J. M. Gracia-Bondía, *On the antipode of Kreimer's Hopf algebra*, Mod. Phys. Lett. A **16** (2001), no. 22, 1427–1434.

[Foi02a] L. Foissy, *Les algèbres de Hopf des arbres enracinés décorés, I*, B. Sci. Math. **126** (2002), no. 3, 193–239.

[Foi02b] ———, *Les algèbres de Hopf des arbres enracinés décorés, II*, B. Sci. Math. **126** (2002), no. 4, 249–288.

[FS13] P. Friz and A. Shekhar, *Doob-Meyer for rough paths*, Bull Inst. Math. Acad. Sin. (N.S.) **8** (2013), no. 1, 73–84.

[FV06] P. Friz and N. Victoir, *A note on the notion of geometric rough paths*, Probab. Theory Rel. **136** (2006), no. 3, 395–416.

[FH14] P. K. Friz and M. Hairer, *A course on rough paths*, Universitext, Springer, Cham, 2014.

[FV10] P. K. Friz and N. B. Victoir, *Multidimensional stochastic processes as rough paths*, Cambridge Studies in Advanced Mathematics, vol. 120, Cambridge University Press, Cambridge, 2010.

[Gli97] Y. Gliklikh, *Global Analysis in Mathematical Physics*, Applied Mathematical Sciences, vol. 122, Springer, New York, 1997.

[Gro55] A. Grothendieck, *Produits tensoriels topologiques et espaces nucléaires*, Memoirs of the American Mathematical Society, vol. 16, American Mathematical Society, Providence, 1955.

[Gub04] M. Gubinelli, *Controlling rough paths*, J. Funct. Anal. **216** (2004), no. 1, 86–140.

[Gub10] ———, *Ramification of rough paths*, J. Diff. Equations **248** (2010), no. 4, 693–721.

[Hai14] M. Hairer, *A theory of regularity structures*, Invent. Math. **198** (2014), no. 2, 269–504.

[HK15] M. Hairer and D. Kelly, *Geometric versus non-geometric rough paths*, Ann. I. H. Poincaré-PR **51** (2015), no. 1, 207–251.

[HP13] M. Hairer and N. S. Pillai, *Regularity of laws and ergodicity of hypoelliptic SDEs driven by rough paths*, Ann. Probab. **41** (2013), no. 4, 2544–2598.

[HL10] B. Hambly and T. J. Lyons, *Uniqueness for the signature of a path of bounded variation and the reduced path group*, Ann. Math. **171** (2010), no. 1, 109–167.

[Hof03] M. E. Hoffman, *Combinatorics of Rooted Trees and Hopf Algebras*, T. Am. Math. Soc. **355** (2003), no. 9, 3795–3811.

[IW89] N. Ikeda and S. Watanabe, *Stochastic differential equations and diffusion processes*, North-Holland, 1989.

[Iva10] S. Ivanov, *Smoothness of distance function in Riemannian Manifolds*, MathOverflow, https://mathoverflow.net/q/21316 (2010).

[KM97] A. Kriegl and P. W. Michor, *The Convenient Setting of Global Analysis*, Mathematical Surveys and Monographs, vol. 53, American Mathematical Society, Providence, 1997.

[Kun90] H. Kunita, *Stochastic flows and stochastic differential equations*, Cambridge Studies in Advanced Mathematics, vol. 24, Cambridge University Press, Cambridge, 1990.

[KSSV06] M. Kunzinger, H. Schichl, R. Steinbauer, and J. A. Vickers, *Global Gronwall estimates for integral curves on Riemannian Manifolds*, Rev. Mat. Complut. **19** (2006), no. 1, 133–137.

[Lej12] A. Lejay, *Global Solutions to Rough Differential Equations with Unbounded Vector Fields*, Séminaire de Probabilités XLIV, Lecture Notes in Mathematics, Springer, Berlin, Heidelberg, 2012, pp. 215–246.

[Li94] X.-M. Li, *Strong p-completeness of stochastic differential equations and the existence of smooth flows on noncompact manifolds*, Probab. Theory Rel. **100** (1994), no. 4, 485–511.

[LS11] X.-M. Li and M. Scheutzow, *Lack of strong completeness for stochastic flows*, Ann. Probab. **39** (2011), no. 4, 1407–1421.

[Lyo98] T. J. Lyons, *Differential equations driven by rough signals*, Rev. Mat. Iberoam. **14** (1998), no. 2, 215–310.

[Lyo14] ______, *Rough paths, signatures and the modelling of funtions on streams*, Proceedings of the International Congress of Mathematicians 2014 (Seoul), vol. IV, Kyung Moon, 2014, pp. 163–184.

[LCL07] T. J. Lyons, M. J. Caruana, and T. Lévy, *Differential Equations Driven by Rough Paths*, Lect. Notes Math., vol. 1908, Springer, Berlin, 2007.

[LQ02] T. J. Lyons and Z. Qian, *System Control and Rough Paths*, Oxford Mathematical Monographs, Oxford University Press, 2002.

[MKW08] H. Z. Munthe-Kaas and W. M. Wright, *On the Hopf Algebraic Structure of Lie Group Integrators*, Foundations of Computational Mathematics **8** (2008), no. 2, 227–257.

[Reu93] C. Reutenauer, *Free Lie algebras*, London Mathematical Society Monographs. New Series, vol. 7, Clarendon Press, Oxford, 1993.

[RS17] S. Riedel and M. Scheutzow, *Rough differential equations with unbounded drift term*, J. Diff. Equations **262** (2017), no. 1, 283–312.

[RR80] A. Robertson and W. Robertson, *Topological Vector Spaces*, Cambridge Tracts in Mathematics and Mathematical Physics, vol. 53, Cambridge University Press, Cambridge, 1980.

[Ste70] E. M. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, 1970.

[Str87] R. S. Strichartz, *The Campbell-Baker-Hausdorff-Dynkin formula and solutions of differential equations*, J. Funct. Anal. **72** (1987), no. 2, 320–345.

[Tan83] S. Taniguchi, *Malliavin's stochastic calculus of variations for manifold-valued Wiener functionals and its applications*, Z. Wahrscheinlichkeit **65** (1983), no. 2, 269–290.