Rough Path Perspectives on the Itô-Stratonovich Dilemma

a thesis presented for the degree of
Doctor of Philosophy of Imperial College London
by

Emilio Rossi Ferrucci

Department of Mathematics
Imperial College London

October 2021

Imperial College London

Candidate: Emilio Rossi Ferrucci

Title: *Rough Path Perspectives on the Itô-Stratonovich Dilemma*

Submitted for the degree of Ph.D.

Department of Mathematics, Imperial College London. 180 Queen's Gate, London SW7 2AZ

Supervisors: Prof. Damiano Brigo, Dr. Thomas Cass, Dr. John Armstrong (co-supervisor, KCL)

Submission dates: October 14, 2021 (first submission), January 4, 2022 (resubmission — this version)

Doctoral studies funded by the Centre for Financial Computing and Analytics (UCL)

Author emails: emilio.rossi-ferrucci16@imperial.ac.uk (current affiliation), emilioferrucci@gmail.com (personal)

Mathematics subject classification (2020): 60Lxx, 60L10, 60L20, 60L70, 60Hxx, 60H05, 60H07, 60H20, 60Gxx, 60G15, 60G22, 60E10, 58J65, 16T05, 16T30

This dissertation was typeset in LaTeX. Diagrams were produced in Ti*k*Z and plots in Wolfram Mathematica.

Statement of originality

I certify that this thesis is the product of my own work, and that any ideas or quotations from the work of others, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Emilio Rossi Ferrucci

Rough Path Perspectives on the Itô-Stratonovich Dilemma

Abstract

This thesis is comprised of six distinct research projects which share the theme of rough and stochastic integration theory.

Chapter 1 deals with the problem of approximating an SDE X in $\mathbb{R}^d$ with one Y defined on a specified submanifold, so as to minimise quantities such as $\mathbb{E}[|Y_t - X_t|^2]$ for small t: this is seen to be best performed when using Itô instead of Stratonovich calculus.

Chapter 2 develops the theory of not necessarily geometric $3 > p$-rough paths on manifolds. Drawing on [FH14, É89, É90] we define controlled rough integration and RDEs both in the local and extrinsic framework, with the latter generalising [CDL15]. Finally, we lay out the theory of parallel transport and Cartan development, for which non-geometricity results in second-order conditions and corrections to the classical formulae.

In Chapter 3 we treat the theory of geometric rough paths of arbitrary roughness in the framework of controlled paths of [Gub04], from an algebraic and combinatorial point of view, and avoiding the smooth approximation arguments used in [FV10b]. As an application, we show how our emphasis on functoriality allows for a simple transposition of the theory to the manifold setting.

The goal of Chapter 4 is to treat the theory of branched rough paths on manifolds. Drawing on [HK15, Kel12], we show how to lift a controlled path to a rough path. The "transfer principle", intended in the sense of Malliavin and Emery, refers to the expression of a connection-dependent "intrinsic differential" $\mathrm{d}_\nabla \boldsymbol{X}$ that defines integration in a coordinate-invariant manner, which we derive by combining Kelly's bracket corrections with certain higher-order Christoffel symbols. In reviewing branched rough paths, special attention is given to those that can be defined on Hoffman's quasi-shuffle algebra [Hof00], for which some of the relations simplify.

The final two chapters do not involve any differential geometry. Chapter 5 is a report on work in progress, the aim of which is to compute the Wiener chaos decomposition (and in particular the expectation) of the signature of certain multidimensional Gaussian processes such as $1/3 < H$-fractional Brownian motion (fBm). This generalises the results of [BC07], arrived at through a piecewise-linear approximation argument which fails when $1/4 < H \leq 1/2$. Furthermore, our calculation restricts to that of [Bau04] in the case of Brownian motion, and can be applied to other semimartingales, such as the Brownian bridge. Our novel approach makes use of Malliavin calculus and the recent rough-Skorokhod conversion formula of [CL19].

Finally, in Chapter 6 we combine the topics of the previous two to define a branched rough path above multidimensional $1/4 < H$-fBm, and compute its terms and correction terms. Our rough path is defined intrinsically and canonically in terms of the stochastic process, restricts to the Itô rough path when $H = 1/2$, has the property that its integrals of one-forms vanish in mean, and is not quasi-geometric when $H \in (1/4, 1/3]$.

Preface

The tale of stochastic integration begins with Kiyoshi Itô, who in [Itô44] defined the integral that now bears his name, with the relative notion of differential equation [Itô46] and his famous second-order change of variables formula [Itô51] following shortly thereafter. A little over a decade later it was discovered, by Donald Fisk in his PhD thesis [Fis63] and independently by Ruslan Stratonovich [Str66], that a different definition of stochastic integral was available. While Itô had defined his integral using Riemann-Stieltjes approximations with evaluations of the integrand at the initial point in each interval of the partition (which for the integral $\int H \mathrm{d}X$ would look like $\sum_i H_{t_i}(X_{t_{i+1}} - X_{t_i})$), the integrand in the Fisk-Stratonovich definition is evaluated at the midpoint ($\sum_i H_{\frac{t_i+t_{i+1}}{2}}(X_{t_{i+1}} - X_{t_i})$ for continuous H).

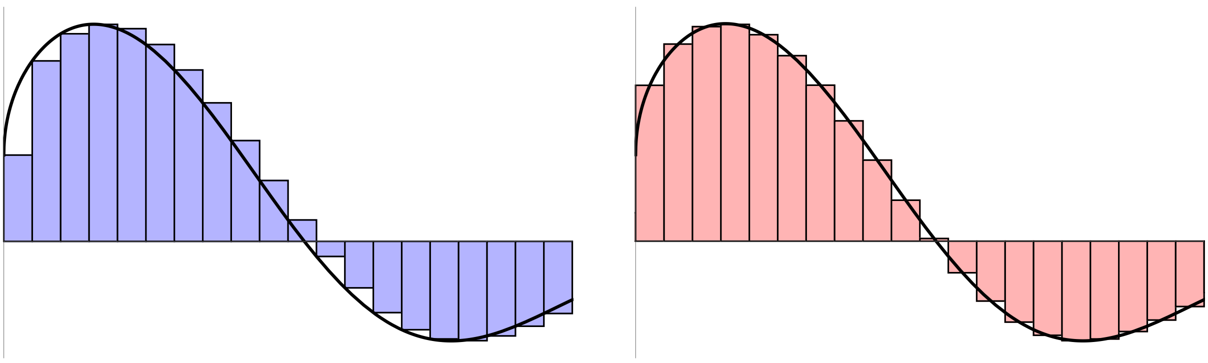


Figure 1: Riemann sums approximations for the integral of a smooth function with initial point evaluation (on the left), and with midpoint evaluation (on the right); since this is a classical Riemann integral, the two definitions of integral coincide.

The Itô integral has clear advantages from the standpoint of probability theory: it is (under reasonable assumptions) a martingale when the integrator is such, and its variance can be computed easily thanks to an isometry property. While Stratonovich and Fisk's integral does not satisfy these properties (since it "looks into the future"), it has the great benefit of satisfying the same laws as ordinary calculus, such as integration by parts and the fundamental theorem of calculus. Furthermore, around the same time Wong and Zakai realised that defining integration [WZ65a] and SDEs [WZ65b] by smoothly approximating the noise, integrating in the usual Riemann sense, and taking limits, the result coincided not with Itô's integral but with what would soon after be named the Stratonovich integral. These two competing definitions gave rise to a dilemma: which is the correct way to integrate?

The Itô-Stratonovich dilemma does not, of course, admit a clear resolution. Both theories are perfectly valid, can be converted into one another, and the choice of which to use became contingent on the branch of stochastic analysis under consideration: the Itô integral remained the weapon of choice of probabilists and researchers in the nascent discipline of mathematical finance, while physicists and differential geometers gravitated towards the Stratonovich integral, due to its good behaviour under change of coordinates. Indeed, it is this property that makes it possible to define geometric constructions involving stochastic processes just as one would define them for smooth curves, simply by replacing ordinary calculus with Stratonovich calculus: this is what Malliavin later called the "transfer principle". It is not true, however, that Itô calculus on manifolds is impossible to carry out. An early attempt was made by Itô himself [Itô50], who asked the very natural question of how to define SDEs on smooth manifolds. While this paper focused on existence, uniqueness and Markov property of the solution, Itô's description of SDEs had the drawback of depending on the chart (a problem that does not arise for Stratonovich SDEs). It was not until much later that Laurent Schwartz [Sch82] and Paul-André Meyer [Mey81, Mey82] independently worked out the correct framework needed to make Itô calculus coordinate-invariant. Their ideas were carefully compiled by Michel Emery, one of Meyer's students, in the very accessible [É89]. He followed this textbook up a year later with a paper [É90] in which he established that the "Itô differential" $\mathrm{d}X$ could be conceived of not only as a formal element of a second-order tangent space (as Schwartz had regarded it), but, alternatively, as an infinitesimal tangent vector (an idea already present in Meyer's work), the only caveat being that the resulting definitions would depend on a covariant derivative on the manifold. This, at least in theory, put Itô and Stratonovich on an equal footing as stochastic calculi on Riemannian manifolds, since they were now both capable of integrating one-forms and defining SDEs in terms of vector fields. In practice, in the geometric setting Stratonovich calculus continued to be used almost exclusively, and still today Itô calculus carries, rather unfairly, the reputation of being unsuitable for such purposes.

Around the same time that Schwartz and Meyer were developing their calculus of second-order vectors and forms, a different contribution to stochastic calculus came from Hans Föllmer. In fact, Föllmer's idea was not so much an addition to probability theory as it was a recognition of the fact that certain aspects of it did not actually have to be stochastic. In [Föl81] he showed that the Itô formula could be derived without reference to a probability space or measure, as long as one accepted that the quadratic variation would depend on a sequence of partitions. Many authors continued to investigate the delicate interplay between probability and analysis in the context of stochastic integration, which turned out to be related to the study of iterated path integrals valued in nilpotent Lie groups and to Kuo-Tsai Chen's earlier work on loop space cohomology [Che77]. This line of research led to a series of papers of Terry Lyons [Lyo94], some of which written with Zhongmin Qian [LQ98], culminating in the paper [Lyo98] (written at Imperial College London!), in which the theory of rough paths was established.

In a nutshell, a rough path consists of a structure "above" a path $X\colon [0,T] \to \mathbb{R}^d$ which emulates the first few iterated integrals on the simplex $\boldsymbol{X}_{st} = \int_{s<u_1<\ldots<u_n<t} \mathrm{d}X_{u_1} \otimes \cdots \otimes \mathrm{d}X_{u_n}$, the exact number required depending on the regularity of X. When the path X is too rough (as is the case for Brownian motion), these integrals are not defined in the sense of Riemann-Stieltjes, and must therefore be postulated through other methods, subject to certain analytic and algebraic constraints. The purpose of a rough path is to define an integration theory for X via Taylor-type expansions, which in turn gives meaning to multidimensional rough ODEs (or rough differential equations — RDEs) $\mathrm{d}Y = V(Y)\mathrm{d}\boldsymbol{X}$. When X is a stochastic process, $\boldsymbol{X}$ is usually constructed using probabilistic notions of convergence; however, once $\boldsymbol{X}$ is defined, all the central constructions of rough path theory follow from pathwise and deterministic analysis. This separation of probability and analysis is crucial for one of the main contributions of rough path theory: that of making the RDE solution map $\boldsymbol{X} \mapsto Y$ continuous under appropriate norms, thus conferring "robustness" to the theory. This is not achievable in ordinary stochastic analysis, since probability is involved in the definition of

integral: the map sending X to the solution to the Itô equation $\mathrm{d}Y = V(Y)\mathrm{d}X$ (or Stratonovich equation $\mathrm{d}Y = V(Y) \circ \mathrm{d}X$) cannot be made continuous. Another aspect that makes Lyons' theory so appealing is its generality, and the ease with which it can be used to generate new integration theories: not only is it capable of reproducing the Itô and Stratonovich integrals, but many other types of processes, such as certain Gaussian and Markov processes that are not semimartingales, admit lifts to rough paths too. Of course, when X is of bounded $2 > p$-variation, everything reduces to Young integration [You36].

A rough path canonically defines its signature, i.e. the full stack of iterated integrals, which is already of interest when the path is of bounded variation. Indeed, it was under this assumption that Lyons, together with Ben Hambly, proved the main result that establishes the importance of the signature: in [HL10] (later generalised to rough paths in [BGLY16]) it was shown that the signature of a path, evaluated at its fixed initial and terminal times, characterises the path up to "retracings". In dimensions greater than 1, a generic path can be expected to almost never retrace itself: this remarkable characterisation therefore yields a way of translating the entirety of the data contained in the path into a formal series of tensors. A probabilistic version of this result [CL16] states that, under certain hypotheses, the law of a stochastic process is determined by the process's expected signature, which therefore plays the same role that moments do for random variables. Recently, these ideas have found numerous applications in machine learning for sequential, multimodal data streams, e.g. [PAGG$^+$18].

Since its inception, the theory of rough paths has benefited from numerous additions and reformulations. In [Gub04] Massimiliano Gubinelli gave the definition of controlled rough paths: enriched paths which can be integrated against the reference rough path. Controlled paths play a role dual to that of rough paths and make it possible for much of the theory to be expressed in linear terms, which is not possible when dealing solely with rough paths, which are non-linear objects. In [Gub10] Gubinelli defined branched rough paths, the most general type possible: unlike the better-known geometric ones, the integration theories defined by branched rough paths need not satisfy the rules of ordinary calculus, such as integration by parts. His ideas were reformulated by David Kelly in his PhD thesis [Kel12], in part published jointly with his supervisor in [HK15], in which it is shown that branched rough paths can be reduced to geometric ones, albeit via a non-canonical procedure. Many different authors worked on Gaussian rough paths, and their relationship with Malliavin calculus [CQ02, LQ02, FV10a]; the best known result in this area, a Hörmander condition for Gaussian (non-Markovian) rough paths was proved by Thomas Cass and Peter Friz in [CF10]. Recently, the area of "rough analysis" has enjoyed a phenomenal boost in recognition and popularity due to its relationship with Martin Hairer's acclaimed theory of regularity structures [Hai14], and thanks to the textbook [FH14], aimed at a broad readership.

When I arrived in London to start my MRes+PhD I held a master's with a focus in algebraic and differential topology, and very little higher knowledge of probability theory. The idea for my PhD loosely consisted of leveraging my "non-standard background" as an aspiring probabilist, in particular my familiarity with smooth manifolds, to take further certain novel viewpoints on Itô calculus on manifolds [AB18] and ensuing approximation problems [AB16] that Damiano Brigo and John Armstrong (my supervisor and co-supervisor) had been working on, and to relate their ideas to the analysis of rough paths, of which my other supervisor, Tom Cass, was a specialist. As I began to learn about stochastic calculus, the distinction between Itô and Stratonovich integration immediately caught my eye, and I became curious about how these subtleties could be couched in the language of rough paths in a maximally general manner, especially in the setting of smooth manifolds. One of the very first people I met in London was Michel Emery, who had been invited to give a talk on the Schwartz-Meyer approach to Itô calculus, about which he also spoke to me at length in private. When I returned home to Italy for Christmas that year, I found that he had mailed me a copy of his textbook with a thoughtful inscription. Upon reading it, I developed a soft spot for Emery's treatment of stochastic calculus on manifolds, which went on to become one of the major sources of inspiration in my studies.

After concluding the taught portion of my MRes, I identified my foothold in graduate research as Damiano and John's work on projections of SDEs onto submanifolds, an approximation problem that points towards Itô, not Stratonovich, calculus in the setting of differential geometry. In my attempt to understand their paper, I found myself rewriting it by using ambient coordinates instead of local ones. This reformulation, which included several other improvements and additions, turned into a paper, and Chapter 1 of this thesis. Some of my contributions were included in the paper [ABRF19], now published in the *Proceedings of the London Mathematical Society*. In the meantime, my interest had shifted from stochastic calculus to rough paths. Having studied [FH14], which treats non-geometric rough paths as well as geometric ones, and read the recent paper [CDL15] about geometric rough paths on manifolds, I decided to merge my supervisors' areas of interest and study the behaviour of non-geometric rough paths on manifolds. This project, which begins by applying the ideas of [É89, É90], and goes on to investigate the more advanced topics of parallel transport and development, is contained in Chapter 2. While writing this paper, I had been taking some notes on how the structural aspects of rough paths (and their controlled paths), which are easy to describe in the case of bounded $3 < p$-variation, work for geometric rough paths of low regularity. Spelling these out would constitute an approach to the fundamentals of geometric rough path theory, which, within its scope, is alternative to the one of [FV10b], that relies on smooth approximation arguments and does not consider controlled paths. When I told Tom about my ideas, he invited me to join an ongoing project with Bruce Driver and Christian Litterer. This led to a visit to Christian in York, and a fruitful collaboration that resulted in the content of Chapter 3. This and the previous chapters are finished papers, both accepted for publication in the *Journal of the London Mathematical Society*.

The second part of the thesis concerns more advanced topics. I recalled Damiano telling me, at the very beginning of my PhD, that in my study of stochastic processes on manifolds I should try and "go as low as possible" in terms of regularity. This had already been done, in the previous chapter, for geometric rough paths. Damiano, however, had made his comment in the context of Itô calculus, which is not geometric. On Tom's advice, I began to read about branched rough paths, since these are able to describe non-geometric integration theories of arbitrary roughness. After a long and careful study of the main reference on this topic [HK15] and the lesser known PhD thesis [Kel12], which nevertheless contains results that proved essential to my goal, I was able to lay out what I believe can be called the natural generalisation of Meyer and Emery's work to branched rough paths. This project is the subject matter of Chapter 4.

While I was writing Chapters 2 and 3, I realised that my thesis did not contain much probability theory, something which did not seem aligned with the spirit of the original plan. I knew about the connection between rough paths and Malliavin calculus, and about Tom's recent formula [CL19, CL20] obtained jointly with his former student Nengli Lim, relating rough and Skorokhod integration, the natural (but not pathwise) extension of the Itô integral to more general Gaussian processes. Tom had told me that he saw a possible application of his work in computing the expected signature of a large class of Gaussian processes, a complicated task which had not be achieved via elementary methods. The preliminary results of this study are contained in

Chapter 5: while the formal calculation, leading to the desired result, is provided in full, there are still details missing, which are clearly highlighted. Of the six chapters, this is the only one of which the main results rest on unproved technical assertions. While writing the previous two chapters, I wondered whether it was possible to use Skorokhod integration to generate a non-trivial example of a stochastic branched rough path, something I could not find in the literature. While the naïve approach to this fails, there is a workaround that satisfies the needed regularity condition, described in Chapter 6.

Writing this thesis has given me the opportunity to immerse myself in a thriving area of modern mathematics, to take a glimpse at how higher knowledge is formed and evolves in a research community, and to offer a few of my own perspectives on a famous dilemma; one which, I suspect, will not be settled anytime soon.

How this thesis is structured. Each of the first five chapters of this thesis consists of a self-contained project, which introduces its own definitions and conventions; Chapter 6, instead, uses the notation and some of the results of the previous two. Since the chapters are on closely related topics, a little "code duplication" was inevitable: some definitions and proofs are revisited in different contexts, mostly across Chapters 2,3 and 4. This is partly necessary, since the theory requires minor (but non-trivial) modifications in each chapter, and in any case has the benefit of not requiring the reader to constantly refer to other parts of this manuscript. Moreover, despite having part of the setup in common, the main results in the chapters are distinct. In Chapter 2 these are on the extrinsic approach and the theory of non-geometric parallel transport and Cartan development, in Chapter 3 the main focus is on the combinatorics of shuffles and ordered shuffles applied to geometric (controlled) rough paths, and in Chapter 4 the goal is to define integration on manifolds. While this had been done for the other two chapters, it was not their main focus, since for $3 > p$-geometric rough paths the integral is structurally identical to the one defined by Emery, and for geometric rough paths there is no difficulty in integrating on manifolds. Similarly, while there is some overlap between Chapter 3 and Chapter 4 in terms of the combinatorics, the lemmata of the former is developed more systematically, which would have been much more difficult to do for the latter, due to the greater complexity of tree algebras.

At the very beginning of each chapter there is a "Project status" paragraph, in which it is stated whether the contents of the chapter are entirely or in part contained in an arXiv preprint or a submitted/published paper, and it is disclosed whether there are any co-authors, and if so, who they are. To summarise the situation very briefly here, Chapters 1,2 and 3 consist of finished papers; the first has contributed to a published article, and the other two have recently been accepted for publication. Chapter 4 can be considered a finished project, but has not yet been posted to arXiv or submitted for publication. Chapter 5 is a report on work in progress: the case for its inclusion in this thesis is made in the chapter's introduction and the precise details of what is still missing are stated precisely. Chapter 6 can be considered a finished short project, although there are no current plans for it to be posted or submitted. All chapters except for the fourth and sixth are projects written jointly with my supervisors and/or other senior academics, all of whom are aware of the work's inclusion in this thesis. These collaborations were carried out in the manner in which supervision of PhD students is customarily conducted.

Each chapter begins with an introduction that announces and motivates its contents. The body of each of the first five chapters begins with one or two background sections. These sections are meant to establish the notation, "import" and review the theory that is essential to the rest of the chapter, possibly with minor reformulations and accompanying examples; they generally do not contain, unless otherwise mentioned, what may be called original research. These are Section 1.1, Section 1.2; Section 2.1, Section 2.2; to some extent Section 3.1 (in which, however, the main two lemmas are original); Section 4.1; Section 5.1 and to some extent Section 5.2

(in which the result referenced at the beginning is reformulated). At the end of each chapter there is a "Conclusions and further directions" section, which contains reflections on the ideas that have been presented, on any aspects that might deserve further attention, and problems left open that could form the basis for future research.

A word on notation. While each chapter was written individually (and with different supervisors and collaborators), I have attempted to keep the notation as uniform as possible throughout. Any discrepancies should not cause ambiguity, since chapters do not use symbols introduced elsewhere without explicit mention. Since some of the formulae are quite long, I have preferred compact notation when possible. This involves using the Einstein summation convention, which is only implied when one of the indices is a subscript and the other a superscript. Another convention adopted for similar reasons involves only writing the evaluation of a product of functions once, e.g. $fgh(x)$ instead of $f(x)g(x)h(x)$. In general, I have attempted not to clutter equations with redundant information: for example, the partial derivative $\partial_{\alpha\beta\gamma} f$ is written without the superscript "3", since the fact that it is a third derivative is already clear from the presence of three indices. A few different fonts are used, which have the benefit of making global notation for recurring symbols symbols possible: for example, $\mathcal{N}$ is used in Chapter 4 to denote the number of symmetries of a forest and δ is used in the last two chapters for Skorokhod integration; neither is used for other purposes. Other global symbols include p, the rough path's variation regularity, related to the Hurst parameter H when considering fractional Brownian motions, and d, usually the dimension of the Euclidean space, which is different to d, used to denote differentials. When writing stochastic or rough differential equations, Greek indices are usually used for the driving signal, while latin letters index the solution.

Acknowledgements. First and foremost, I would like to thank my supervisors Damiano Brigo and Thomas Cass, and my co-supervisor John Armstrong. The idea for a PhD at the intersection of stochastics, geometry and rough paths came from Damiano: I could not have written this thesis without his vision. I am grateful to him and to John for convincing me that "Itô on manifolds" is not only possible, but also worth the trouble, and to John for encouraging me to check and corroborate my work by making plots in Mathematica, which add some colour to these pages. Tom supervised me for the second part of my studies, during which I worked on the projects that involve the most complex mathematics. His very clever idea to apply the rough-Skorokhod formula to the signature of a Gaussian process proved successful, and his advice that I learn about branched rough paths was very well thought out.

I would also like to thank Christian Litterer and Bruce Driver for collaborating with me (together with Tom) on the topics of Chapter 3. My studies have greatly benefited from numerous other interactions, both in person and by email, with other students and academics from around the world. I am grateful for their insight, and for making me feel accepted as part of their research community.

I am grateful to Martin Hairer and Zhongmin Qian for accepting to examine this thesis, and for the time and consideration that they dedicated to it. Their several helpful comments have improved its quality.

My studies were supported by the CDT in Financial Computing and Analytics, based at UCL and run by Philip Treleaven and Yonita Carter, who have always been very helpful in administrative matters.

Finally, I wish to thank my family for always encouraging me in whatever I decided to do.

London, October & December 2021,
Emilio Rossi Ferrucci

Contents

List of figures

1

Projections of SDEs onto submanifolds

Project status. The work presented in this chapter originated from my attempt to understand and improve certain aspects of a paper of my supervisors [AB16], a project that had begun before I started my studies in London. This resulted in the independent preprint [ABF18], of which this chapter is an edited version. Since some of my work was deemed relevant to the original project, it was included in the version that was submitted to, and eventually published in the *Proceedings of the London Mathematical Society*, [ABRF19]. My contribution to this publication consists of expressing and proving optimality without resorting to stochastic Taylor expansions, but directly in terms of the SDEs, and dealing with the resulting complications (namely that the solution can explode or exit a tubular neighbourhood of the manifold). Another difference between the two papers is that [AB16] uses local coordinates, while this chapter uses ambient ones, a choice that I viewed as more natural, given that the manifold M is always embedded, and indeed preferable, since the formulae for the projections do not depend on the chart and are therefore more easily interpretable. Finally, it should be mentioned that both papers cover material that the other does not: here we show how the projections emerge naturally from different versions of stochastic calculus before showing their optimality, while [AB16] additionally considers applications to non-linear filtering. Specific examples or observations present in [AB16] that are readapted to the different setting of this chapter are explicitly referenced; this is not done for the main theory, the core ideas of which are original to [AB16].

Introduction

Consider the following problem: we are given an autonomous ODE

$$\dot{X}_t = F(X_t) \tag{1.1}$$

in $\mathbb{R}^d$, and a smooth embedded manifold $M \hookrightarrow \mathbb{R}^d$. Let π be the metric projection of a tubular neighbourhood of M onto M (see (1.52) below). We seek an M-valued ODE, i.e. a vector field $\overline{F}$ on M, tangent at each point to M, with the property that the solution to

$$\dot{Y}_t = \overline{F}(Y_t) \tag{1.2}$$

is optimal in the sense that the first coefficient of the Taylor expansion in $t = 0$ of either

$$|Y_t - X_t|^2 \quad \text{or} \quad |Y_t - \pi(X_t)|^2 \tag{1.3}$$

is minimised for any initial condition $X_0 = Y_0 = y_0 \in M$. This requirement represents the slowest possible divergence of Y from the original solution X (resp. from its metric projection on M), subject to the constraint of Y arising as the solution of a closed form ODE on M. It is an easy exercise (using (1.55) below) to check that these optimisation problems both result in the same solution, which consists in $\overline{F}(y)$ being the orthogonal projection of the vector $F(y)$ onto the tangent space T_yM.

The paper [AB16], which is motivated by applications to non-linear filtering, explores an extension of this problem to the case of SDEs. The optimality criteria (1.3) do not carry over in a straightforward fashion, and are formulated through the machinery of weak and strong Itô-Taylor expansions. In this chapter we tackle the same problem through a different perspective, which we proceed to describe.

In Section 1.1 we begin with a survey of SDEs on manifolds. Here we introduce three ways of representing them: the Stratonovich, Schwartz-Meyer (or 2-jet) and Itô representations. The first and second have the advantage of not requiring a connection on the tangent bundle of the manifold, the second and third are defined in terms of the Itô integral, while the first and third have vector coefficients. Focusing on the diffusion case, we show how to pass from one representation to another. In Section 1.2 we prepare the framework for manifolds M embedded in $\mathbb{R}^d$. These are entirely general Riemannian manifolds, due to the Nash embedding theorem, and have the advantage of being describable using ambient coordinates. We use this framework to study the equations introduced in the previous section, on embedded manifolds. In Section 1.3 we associate to each manifold-valued SDE representation a natural projection, which gives rise to an SDE on a submanifold: the Stratonovich projection (defined by projecting the Stratonovich coefficients), the Itô-jet projection (defined by projecting the Schwartz morphism, or 2-jet, which defines the SDE), and the Itô-vector projection (defined by projecting the Itô coefficients, and interpreting the resulting equation w.r.t. the Riemannian connection on the embedded submanifold). These projections coincide with the ones introduced in [AB16], but are given a more solid theoretical underpinning, which sheds light on their analytic and probabilistic properties. We then derive formulae for the three projections, preferring ambient coordinates to local coordinates. In Section 1.4 we formulate the optimality criteria satisfied by the Itô-vector and Itô-jet projections using respectively an explicit weak and mean-square formulation, instead of invoking Itô-Taylor expansions as done in [AB16]. This has the advantage of representing a more tangible property of the solution, and is accompanied by an argument, based on martingale estimates, used to deal with the problem of the solution exiting the tubular neighbourhood of M. Our main theorems Theorem 1.16 and Theorem 1.19 replicate the findings [AB16, Theorem 2, Theorem 3] in this new setting. The fact that the Stratonovich projection does not satisfy either of these optimality criteria is a confirmation of the fact that Itô calculus on manifolds can be of great interest. In Section 1.5 we provide examples showing that the three projections are genuinely distinct, we prove the Itô projections are optimal also when formulating the optimality criteria using M's intrinsic geometry, and explore notions of optimality that are satisfied by the more naïve Stratonovich projection.

Although the material presented here overlaps to a significant degree with the ideas of [AB16], this chapter is entirely self-contained. Moreover, we believe the framework chosen here has a number of advantages of which we hope to make use in future work, as described in Conclusions and further directions.

1.1 SDEs on manifolds

We begin this chapter with a primer on manifold-valued SDEs. Since manifolds, unlike Euclidean space, do not come naturally equipped with coordinates, especially not global ones, the challenge is to express an SDE using intrinsic, coordinate-free notions. Equivalently, one can define an SDE locally in an arbitrary chart, and show that the property of a process of being a solution does not depend on the chart. The coordinate-free definition of a time-homogeneous ODE on a smooth, m-dimensional manifold M is well known: this consists of a tangent vector field, i.e. a section of the tangent bundle of M, $V \in \Gamma TM$. We will denote Γ the set of sections of a fibre bundle, i.e. the smooth right inverses to the bundle projection. A solution to the ODE defined by V is a smooth curve X, defined on some interval of $\mathbb{R}$, with the property that $\dot{X}_t = V_{X_t}$ for all t. This is a coordinate-free definition, and in a chart $\varphi\colon U \to \mathbb{R}^m$ (U open set in M) it corresponds to requiring that, writing $\varphi(X_t) = {}^{\varphi}X_t$ and $V_x = {}^{\varphi}V^k_x \partial_x\varphi_k$, we have ${}^{\varphi}\dot{X}^k_t = {}^{\varphi}V^k_{X_t}$ for all t for which both sides are defined. Notice the sum over k: this is the Einstein summation convention, which we will use throughout this thesis whenever possible; also, $\partial_x\varphi_k$ are the elements of the basis of T_xM defined by the chart φ:

$$\partial_x\varphi_k(f) \coloneqq \frac{\partial(f \circ \varphi^{-1})}{\partial x^k}(\varphi(x)) \quad \text{for } f \in C^\infty M \tag{1.4}$$

In this section we will give similar descriptions of Stratonovich and Itô (non path-dependent) SDEs on manifolds. From now on we will avoid the φ superscripts when no ambiguity occurs, e.g. the previous identity will be written $\dot{X}^k_t = V^k_{X_t}$.

We begin with the Stratonovich case, following mainly [É89, Ch. VII], although the topic is well known. As for the familiar $\mathbb{R}^d$-valued case we will also need a driving semimartingale, which, given the context we are working in can be taken to be valued in another manifold N, of dimension n. Given a stochastic setup $(\Omega, \mathcal{F}_\cdot, P)$ satisfying the usual conditions, a continuous adapted stochastic process $Z\colon \Omega \times \mathbb{R}_{\geq 0} \to N$ is said to be a *semimartingale* if, for all $f \in C^\infty N$, $f(Z)$ is a semimartingale. Just as for the ODE case, what is needed to define a Stratonovich SDE in M driven by Z is a section of some vector bundle: in this case, however, the bundle is no longer just TM, but $\mathrm{Hom}(TN, TM) \to M \times N$, i.e. the vector bundle of linear maps from TN to TM. An element $F \in \Gamma\mathrm{Hom}(TN, TM)$ corresponds to a smooth map $M \times N \ni (x, z) \mapsto F(x, z) \in \mathrm{Hom}(T_zN, T_xM)$. The Stratonovich SDE

$$\mathrm{d}X_t = F(X_t, Z_t) \circ \mathrm{d}Z_t \tag{1.5}$$

in local coordinates (this requires choosing a chart both on N and on M) as $\mathrm{d}X^k_t = F^k_\gamma(X_t, Z_t) \circ \mathrm{d}Z^\gamma_t$ on random intervals that make both sides of the expression well defined. We will always use Greek letters as indices for the driving process, and Latin letters as indices for the solution. The key property that allows one to prove that the coordinate formulation of Stratonovich SDEs holds for all other charts (on the intersection of their respective domains) is that Stratonovich equations satisfy the first order chain rule: clearly (1.5) would not be similarly well defined with Itô integration. One can also define a solution without invoking charts: this entails defining a Stratonovich integral taking as integrator an M-valued semimartingale X and as integrand a previsible process H with values in the cotangent bundle of M and relatively compact image (*locally bounded*), s.t. at each t, H_t is in the fibre at X_t: this yields an $\mathbb{R}$-valued semimartingale which we can write as

$$\int_0^\cdot \langle H_s, \circ \mathrm{d}X_s \rangle \tag{1.6}$$

The angle brackets refer to dual pairing of vectors and covectors. This integral is characterised as being the unique map satisfying the following three properties

Additivity. For all locally bounded previsible H, G above X

$$\int_0^\cdot \langle H_s + G_s, \circ \mathrm{d}X_s \rangle = \int_0^\cdot \langle H_s, \circ \mathrm{d}X_s \rangle + \int_0^\cdot \langle G_s, \circ \mathrm{d}X_s \rangle$$

Associativity. For a real-valued, locally bounded adapted process λ

$$\int_0^\cdot \langle \lambda_s H_s, \circ \mathrm{d}X_s \rangle = \int_0^\cdot \lambda_s \circ \mathrm{d} \textstyle\int_0^s \langle H_u, \circ \mathrm{d}X_u \rangle$$

Change of variable formula. For all $f \in C^\infty M$

$$\int_0^\cdot \langle \mathrm{d}_{X_s} f, \circ \mathrm{d}X_s \rangle = f(X) - f(X_0)$$

where $\mathrm{d}f$ is the one-form given by taking the differential of f. One can then use this integral to say that X solves (1.5) if for all admissible integrands H (even just those arising as the evaluation of a one-form at X)

$$\int_0^\cdot \langle H_s, \circ \mathrm{d}X_s \rangle = \int_0^\cdot \langle F(X_s, Z_s)^* H_s, \circ \mathrm{d}Z_s \rangle \tag{1.7}$$

where the $*$ denotes dualisation.

Remark 1.1 (Autonomousness and explicitness). If $N = \mathbb{R}^n$ we can call (1.5) *autonomous* if $F(z, x)$ does not depend on z, and if $M = \mathbb{R}^m$ we can call it *explicit* if $F(z, x)$ does not depend on x. However, in the general manifold setting these two concepts do not carry over, at least not unless N (resp. M) is parallelisable, with a chosen trivialisation of its tangent bundle. An analogous consideration applies to other flavours of SDEs introduced in this section.

Example 1.2 (Stratonovich diffusion)**.** An important example is the case where $N = \mathbb{R}_{\geq 0} \times \mathbb{R}^n$ and $Z_t = (t, W_t)$, W an n-dimensional Brownian motion, and F not depending explicitly on W. This means (1.5) becomes

$$\mathrm{d}X_t = \sigma_\gamma(X_t, t) \circ \mathrm{d}W_t^\gamma + b(X_t, t)\mathrm{d}t \tag{1.8}$$

for $\sigma_\gamma, b \in \Gamma\mathrm{Hom}(T\mathbb{R}_{\geq 0}, TM) = C^\infty(\mathbb{R}_{\geq 0}, \Gamma TM)$, $\gamma = 1, \ldots, n$. Stratonovich diffusions are sections of the vector bundle

$$\begin{aligned}\mathrm{Diff}^n_{\mathrm{Strat}} \coloneqq \{F \in \mathrm{Hom}(T(\mathbb{R}_{\geq 0} \oplus \mathbb{R}^n), TM) : \forall w_1, w_2 \in \mathbb{R}^n \; F(t, w_1; x) = F(t, w_2; x)\} \\ \to M \times \mathbb{R}_{\geq 0}\end{aligned} \tag{1.9}$$

i.e. elements of the vector space $\Gamma\mathrm{Diff}^n_{\mathrm{Strat}}$. Notice that the base space is not $M \times (\mathbb{R}_{\geq 0} \times \mathbb{R}^n)$, since independence of the Brownian motion allows us to forget the $\mathbb{R}^n$ component.

We note that no additional structure on N and M, apart from their smooth atlas, is needed to define Stratonovich equations. Stratonovich SDEs are the most used in stochastic differential geometry, as they behave well w.r.t. notions of first order calculus: for instance, if there exists an embedded submanifold M' of M such that $F(y, z)$ maps to $T_y M'$ for all $z \in N$ and all $y \in M'$, then the solution to the Stratonovich SDE defined by F started on M' will remain on M' for the duration of its lifetime. This is evident from our intrinsic approach, by considering $F|_{M' \times N}$, but some authors who develop Stratonovich calculus on manifolds extrinsically prove this by showing that the distance between the solution and the manifold (embedded in Euclidean space) is zero [Hsu02, Prop. 1.2.8]. The existence and uniqueness of solutions to Stratonovich SDEs can be treated by using the Whitney embedding theorem to embed N and M in Euclidean spaces of high enough dimension, and smoothly extending F so that it vanishes outside a compact set containing the manifolds. Invoking the usual existence and uniqueness theorem (e.g. [Pro05, Theorems 38-40]), and the good behaviour of Stratonovich SDEs w.r.t. submanifolds, immediately proves that a unique solution exists up to a positive stopping time, provided F is smooth. We will mostly not be concerned with global-in-time existence in this thesis, although sufficient conditions for such behaviour can usually be obtained by requiring global Lipschitz continuity w.r.t. complete Riemannian metrics.

We now pass to Itô theory on manifolds, as developed in [É89, Ch.VI]. The difficulty lies in the second order chain rule of the Itô integral. For this reason, we need to invoke structures of order higher than 1. Let the *second order tangent bundle* of M, $\mathbb{T}M$, denote the bundle of second order differential operators without a constant term, i.e. given a local chart φ containing x in its domain, an element of $L_x \in \mathbb{T}_x M$ consists of a map

$$L_x \colon C^\infty M \to \mathbb{R}, \quad L_x f = L^k_x \frac{\partial f}{\partial \varphi^k} + L^{ij}_x \frac{\partial^2 f}{\partial \varphi^i \partial \varphi^j} \tag{1.10}$$

The coefficients L^k_x, L^{ij}_x obviously depend on φ, but their existence does not; moreover, requiring $L^{ij}_x = L^{ji}_x$ ensures their uniqueness for the given chart φ. Note that if the L^{ij}_x's vanish $L_x \in T_x M$. $\mathbb{T}M$ is given the unique topology and smooth structure that makes the projection $\mathbb{T}M \to M$, $L_x \mapsto x$ a locally trivial surjective submersion. Just as for the first order case, there is an obvious notion of induced bundle map $\mathbb{T}f \colon \mathbb{T}N \to \mathbb{T}M$ for $f \in C^\infty(N, M)$. A chart φ containing x in its domain defines the basis

$$\{\partial_x \varphi_k, \partial^2_x \varphi_{ij} = \partial^2_x \varphi_{ji} \mid k, i, j = 1, \ldots, n\} \tag{1.11}$$

so the dimension of $\mathbb{T}M$ (as a vector bundle) is $m + m(m+1)/2$. The fundamental properties of $\mathbb{T}M$ are summarised the short exact sequence of vector bundles over M

$$0 \longrightarrow TM \xrightarrow{\ i\ } \mathbb{T}M \xrightarrow{\ p\ } TM \odot TM \longrightarrow 0 \tag{1.12}$$

with the third term denoting symmetric tensor product, the first map the obvious inclusion and the second map given by

$$L_x \mapsto \big(f, g \mapsto \frac{1}{2}(L_x(fg) - f(x)L_x g - g(x)L_x f)\big) \tag{1.13}$$

Roughly speaking, this means that $\mathbb{T}M$ is "noncanonically the direct sum of TM and $TM \odot TM$". This short exact sequence of course dualises to a short exact sequence of dual bundles. Elements of $\mathbb{T}^*_x M$ can always be represented as $\mathbb{d}_x f$, defined by

$$\langle \mathbb{d}_x f, L_x \rangle \coloneqq L_x(f) \tag{1.14}$$

for some $f \in C^\infty M$ (this is of course only true at a point: not all sections of $\mathbb{T}M$ are of the form $\mathbb{d}f$). We now wish to define an Itô-type equation using second order tangent bundles instead of ordinary tangent bundles. For this we need a notion of field of maps $\mathbb{F}(x,z) \colon \mathbb{T}_z N \to \mathbb{T}_x M$. Since the bundles in question are linear, it is tempting to allow $\mathbb{F}(x,z)$ to be an arbitrary linear map, but a more stringent condition is necessary to guarantee well-posedness: the correct requirement is that $\mathbb{F}(x,z)$ define a morphism of short exact sequences, i.e. a commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & T_z N & \longrightarrow & \mathbb{T}_z N & \longrightarrow & T_z N \odot T_z N & \longrightarrow & 0 \\ & & \Big\downarrow{\scriptstyle F(x,z)} & & \Big\downarrow{\scriptstyle \mathbb{F}(x,z)} & & \Big\downarrow{\scriptstyle F(x,z)\otimes F(x,z)} & & \\ 0 & \longrightarrow & T_x M & \longrightarrow & \mathbb{T}_x M & \longrightarrow & T_x M \odot T_x M & \longrightarrow & 0 \end{array} \tag{1.15}$$

with $F(x,z) = \mathbb{F}(x,z)|_{T_z N}$. $\mathbb{F}(x,z)$ is then called a *Schwartz morphism*, and we can then view $\mathbb{F}$ as being the section of a sub-fibre bundle $\mathrm{Sch}(N,M)$ of $\mathrm{Hom}(\mathbb{T}N, \mathbb{T}M)$ over $M \times N$ consisting of such maps, which we call the *Schwartz bundle*. Note that $\mathrm{Sch}(N,M)$ is not closed under sum and scalar multiplication taken in the vector bundle $\mathrm{Hom}(\mathbb{T}N, \mathbb{T}M)$, and thus can only be treated as a fibre bundle. Now, given $\mathbb{F} \in \Gamma\mathrm{Sch}(N,M)$, we will give a meaning to the SDE

$$\mathbb{d}X_t = \mathbb{F}(X_t, Z_t)\mathbb{d}Z_t \tag{1.16}$$

which we will call a *Schwartz-Meyer equation*. Heuristically, if X is an M-valued semimartingale the second order differential $\mathbb{d}X_t$ should be interpreted in local coordinates φ as

$$\mathbb{d}X_t = \mathrm{d}X^k_t \partial_{X_t}\varphi_k + \tfrac{1}{2}\mathrm{d}[X^i, X^j]_t \partial^2_{X_t}\varphi_{ij} \in \mathbb{T}_{X_t}M \tag{1.17}$$

where the first differential is an Itô differential; this expression is seen to be invariant under change of charts, thanks to the Itô formula. Then, given charts φ in M and ϑ on N, and writing

$$\begin{aligned} \mathbb{F}(x,z)\partial_z\vartheta_\gamma &= \mathbb{F}^k_\gamma(x,z)\partial_x\varphi_k + \mathbb{F}^{ij}_\gamma(x,z)\partial^2_x\varphi_{ij} \\ \mathbb{F}(x,z)\partial^2_z\vartheta_{\alpha\beta} &= \mathbb{F}^k_{\alpha\beta}(x,z)\partial_x\varphi_k + \mathbb{F}^{ij}_{\alpha\beta}(x,z)\partial^2_x\varphi_{ij} \end{aligned} \tag{1.18}$$

(1.16) becomes the system

$$\begin{cases} \mathrm{d}X^k_t = \mathbb{F}^k_\gamma(X_t, Z_t)\mathrm{d}Z^\gamma_t + \frac{1}{2}\mathbb{F}^k_{\alpha\beta}(X_t, Z_t)\mathrm{d}[Z^\alpha, Z^\beta]_t \\ \frac{1}{2}\mathrm{d}[X^i, X^j]_t = \mathbb{F}^{ij}_\gamma(X_t, Z_t)\mathrm{d}Z^\gamma_t + \frac{1}{2}\mathbb{F}^{ij}_{\alpha\beta}(X_t, Z_t)\mathrm{d}[Z^\alpha, Z^\beta]_t \end{cases} \tag{1.19}$$

Computing the quadratic covariation matrix of X from the first equation above, using the Kunita-Watanabe identity, and comparing with the second results in the requirement that

$$\mathbb{F}^{ij}_{\gamma} \equiv 0; \quad \mathbb{F}^{ij}_{\alpha\beta} \equiv \tfrac{1}{2}\big(\mathbb{F}^i_\alpha \mathbb{F}^j_\beta + \mathbb{F}^j_\alpha \mathbb{F}^i_\beta\big) \tag{1.20}$$

which correspond precisely to the Schwartz condition (1.15), and justifies this requirement. (1.19) now reduces to its first line, i.e. the Itô SDE

$$\mathrm{d}X^k_t = \mathbb{F}^k_\gamma(X_t, Z_t)\mathrm{d}Z^\gamma_t + \tfrac{1}{2}\mathbb{F}^k_{\alpha\beta}(X_t, Z_t)\mathrm{d}[Z^\alpha, Z^\beta]_t \tag{1.21}$$

on random intervals that make both sides of the expression well-defined.

Example 1.3 (Schwartz-Meyer diffusion)**.** Proceeding as in Example 1.2, but with Schwartz-Meyer equations, we can define the Schwartz-Meyer SDE

$$\begin{aligned} \mathbb{d}X_t &= \mathbb{F}(X_t, t)\mathbb{d}Z_t \\ &= \sigma_\gamma(X_t, t)\mathrm{d}W^\gamma_t + \left(\mathbb{F}_0 + \frac{1}{2}\sum_{\gamma=1}^{n} \mathbb{F}_{\gamma\gamma}\right)(X_t, t)\mathrm{d}t \end{aligned} \tag{1.22}$$

where we can call $\mathbb{F}_\gamma = \sigma_\gamma$ the diffusion coefficients, since they are elements of $C^\infty(\mathbb{R}_{\geq 0}, \Gamma TM)$; this also holds for $\gamma = 0$, but not for $\mathbb{F}_{\alpha\beta} \in C^\infty(\mathbb{R}_{\geq 0}, \Gamma\mathbb{T}M)$. Therefore the coefficient of $\mathrm{d}t$, the "drift", cannot be interpreted as a vector. Note that setting $\mathbb{F}_{\gamma\gamma} \equiv 0$ does not guarantee that such coefficients will vanish w.r.t. another chart, since the transformation rule for them involves the $\mathbb{F}^{ij}_{\alpha\beta}$'s which cannot vanish by the second Schwartz condition (1.20); in other words, there is no way to do away with the non vector-valued drift in (1.22). We can consider Schwartz Meyer diffusions as being sections of the fibre bundle

$$\begin{aligned} \mathrm{Diff}^n_{\mathrm{Sch}} M :=& \frac{\{\mathbb{F} \in \mathrm{Sch}(\mathbb{R}_{\geq 0} \times \mathbb{R}^n, M) : \forall w_1, w_2 \in \mathbb{R}^n\ \mathbb{F}(t, w_1; x) = \mathbb{F}(t, w_2; x)\}}{\mathbb{F} \sim \mathbb{G} \Leftrightarrow \mathbb{F}_{\gamma \geq 1} = \mathbb{G}_\gamma,\ \mathbb{F}_0 + \frac{1}{2}\sum_{\gamma=1}^n \mathbb{F}_{\gamma\gamma} = \mathbb{G}_0 + \frac{1}{2}\sum_{\gamma=1}^n \mathbb{G}_{\gamma\gamma}} \\ &\to M \times \mathbb{R}_{\geq 0} \end{aligned} \tag{1.23}$$

This means that, similarly to the case of (1.9) we are only considering $\mathbb{F}$'s that do not depend explicitly on the Brownian motion, and we are quotienting out the part that is not relevant for (1.22).

Just as for Stratonovich SDEs, Schwartz-Meyer equations can also be seen to come from an integral

$$\int_0^\cdot \langle \mathbb{H}_s, \mathbb{d}X_s \rangle \tag{1.24}$$

where the process $\mathbb{H}$ is now valued in $\mathbb{T}^*M$. The axioms for this *Schwartz-Meyer integral* are similar:

Additivity. For all locally bounded previsible $\mathbb{H}, \mathbb{G}$ above X

$$\int_0^\cdot \langle \mathbb{H}_s + \mathbb{G}_s, \mathbb{d}X_s \rangle = \int_0^\cdot \langle \mathbb{H}_s, \mathbb{d}X_s \rangle + \int_0^\cdot \langle \mathbb{G}_s, \mathbb{d}X_s \rangle$$

Associativity. For a real-valued, locally bounded adapted process λ

$$\int_0^\cdot \langle \lambda_s \mathbb{H}_s, \mathbb{d}X_s \rangle = \int_0^\cdot \lambda_s \mathrm{d} \textstyle\int_0^s \langle H_u, \mathbb{d}X_u \rangle$$

Change of variable formula. For all $f \in C^\infty M$

$$\int_0^\cdot \langle \mathbb{d}_{X_s} f, \mathbb{d}X_s \rangle = f(X) - f(X_0)$$

Notice how Itô integration is used in the associativity axiom. The property of a process of being a solution of (1.16) is then defined in complete analogy to (1.7).

The recent paper [AB18] treats SDEs on manifolds using a representation which is similar to that of (1.16), but which has a distinct advantage when it comes to numerical schemes. Here the authors focus on the autonomous diffusion case, without explicitly taking time as a driver ($N = \mathbb{R}^n$, $Z_t = W_t$), and take the field of Schwartz morphisms $\mathbb{F}$ to be *induced* by a *field of maps* i.e. a smooth function $f\colon \mathbb{R}^n \times M \to M$, $f_x := f(\cdot, x)$, s.t. for all $x \in M$, $f_x(0) = x$: this means

$$\mathbb{F}(x) = \mathbb{T}_0 f_x \tag{1.25}$$

In coordinates φ on M this amounts to

$$\sigma^k_\gamma(x) = \frac{\partial(\varphi^k \circ f_x)}{\partial w^\gamma}(0), \quad \mathbb{F}^k_{\alpha\beta}(x) = \frac{\partial^2(\varphi^k \circ f_x)}{\partial w^\alpha \partial w^\beta}(0) \tag{1.26}$$

with $\mathbb{F}_0 = 0$ (note how the drift comes from the quadratic variation of Brownian motion, without having to require time as a driving process). This particular form of $\mathbb{F}$ is useful because it automatically defines a numerical scheme for the solution of the SDE, similar to the Euler scheme, which cannot be defined in a coordinate-free way on a manifold: the linear structure lacked by M is replaced with iterative interpolations along the f_x's. This also has the advantage of guaranteeing that if the maps are valued in M, so are all the approximations.

"Itô-type" Diffusions on manifolds have also been investigated by other authors, most notably by [BD90, Ch.4] (although we refer to the more recent exposition [Gli11, §7.2]), who call the bundle $\mathrm{Diff}^n_{\mathrm{Sch}} M$ the *Itô bundle*, and give a local description of it. Although we will not need this formulation in the following sections, we include a description of it to establish the link with the other approach. There are (at least) two ways of describing a fibre bundle $\pi\colon E \to M$: one is by simply exhibiting the manifolds E, M and the surjective submersion π, and by checking local triviality; this is the approach taken here. The second approach involves declaring the base space M, the structure group G (a Lie group), the typical fibre F (a smooth manifold, carrying a left action of G by smooth maps) and a covering $\{U_\lambda\}_\lambda$ of M together with maps $g_{\nu\mu}\colon U_\mu \cap U_\nu \to G$ satisfying the cocycle conditions $\forall \lambda, \mu, \nu \; g_{\nu\mu} g_{\mu\lambda} = g_{\nu\lambda}$. Then the total space and bundle projection can be reconstructed by gluing all the $U_\lambda \times F$'s together according to the $g_{\nu\mu}$'s:

$$E := \frac{\bigcup_\lambda \{\lambda\} \times U_\lambda \times F}{(\mu, x, e) \sim (\nu, y, f) \Leftrightarrow x = y, \; f = g_{\nu\mu}(x).e} \xrightarrow{\pi} M, \quad [\mu, x, e] \mapsto x \tag{1.27}$$

Of course, the local description can be obtained from the ordinary one by fixing a local trivialisation, a model for the fibre, a Lie group capturing all transformations of the fibres, etc. Now, we define the candidate bundle of Schwartz-Meyer diffusions to have base space $M \times \mathbb{R}_{\geq 0}$ and typical fibre $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m) \oplus \mathbb{R}^m$. Recall that we observed that the Schwartz bundle is not linear: this should rule out the usual choices $G = GL(n, \mathbb{R}), O(n)$, valid for vector bundles. Indeed, the transformation laws for $\mathrm{Diff}^n_{\mathrm{Sch}} M$ are succinctly modelled by the *Itô group*

$$\mathfrak{I}^m := GL(m, \mathbb{R}) \times \mathrm{Hom}(\mathbb{R}^m \odot \mathbb{R}^m, \mathbb{R}^m) \tag{1.28}$$

$$(A, a)(B, b) \coloneqq (A \circ B, A \circ b + a \circ (B \otimes B)) \tag{1.29}$$

with identity $(I_m, 0)$, acting on $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m) \oplus \mathbb{R}^m$ from the left by

$$(A, a).(\sigma, \eta) \coloneqq \big(A \circ \sigma, A\eta + \tfrac{1}{2}\mathrm{tr}(a \circ (\sigma \otimes \sigma))\big) \tag{1.30}$$

where the trace is taken componentwise. Given an open covering $\{U_\lambda\}_\lambda$ (consisting of, say, open balls) of M, and charts $\varphi_\lambda \colon U_\lambda \to \mathbb{R}^m$, we define

$$g_{\nu\mu}(x \in U_\mu \cap U_\nu) \coloneqq \big(J(\varphi_\nu \circ \varphi_\mu^{-1})(x), H(\varphi_\nu \circ \varphi_\mu^{-1})(x)\big) \tag{1.31}$$

the Jacobian and Hessian of the change of coordinates. The isomorphism between the bundle that we have just described and $\mathrm{Diff}^n_{\mathrm{Sch}} M$ is given by (notation as in (1.27)) $[\lambda, (t, x), (\sigma, \eta)] \mapsto [\mathbb{F}(x, t)]$, the class represented by any $\mathbb{F}(t, x)$ in the numerator of (1.23) s.t. $\mathbb{F}^k_\gamma = \sigma^k_\gamma$ for $\gamma = 1, \ldots, n$ and $\mathbb{F}^k_0 + \sum_{\gamma=1}^n \mathbb{F}^k_{\gamma\gamma} = \eta^k$ w.r.t. the chart φ_λ.

There is a way of writing Itô equations on a manifold so that all the coefficients, drift included, are vectors. It involves considering the additional structure of a linear connection ∇ on M, i.e. a covariant derivative

$$\nabla \colon TM \times \Gamma TM \to TM \tag{1.32}$$

which is a smooth function that maps $T_xM \times \Gamma TM$ to T_xM, is $\mathbb{R}$-bilinear, and satisfies the Leibniz rule $\nabla_{U_x}(fV) = f(x)\nabla_{U_x}V + (U_x f)V_x$. Equivalently, a connection is described through its Hessian

$$\nabla^2 \colon C^\infty M \to \Gamma(T^*M \otimes T^*M) \tag{1.33}$$

which is an $\mathbb{R}$-linear map satisfying $\nabla^2(fg) = f\nabla^2 g + g\nabla^2 f + \mathrm{d}f \otimes \mathrm{d}g + \mathrm{d}g \otimes \mathrm{d}f$ for all $f, g \in C^\infty M$. These two data are equivalent and related by

$$\langle \nabla^2_x f, V \otimes U\rangle = U_x(Vf) - (\nabla_{U_x} V)f \tag{1.34}$$

If Γ^{ij}_k are the Christoffel symbols of ∇ w.r.t. a chart φ (this means $\nabla_{\partial_x \varphi_i}\partial\varphi_j = \Gamma^k_{ij}(x)\partial_x\varphi_k$), the Hessian can be written as

$$\nabla^2_x f = (\partial^2_x \varphi_{ij} - \Gamma^k_{ij}(x)\partial_x\varphi_k)(f)\mathrm{d}_x\varphi^i \otimes \mathrm{d}_x\varphi^j \tag{1.35}$$

We will only be interested in connections modulo torsion, so it is not limiting for us to assume that a connection is symmetric or torsion-free, i.e. that its torsion tensor $\langle \tau_\nabla, U \otimes V\rangle = \nabla_U V - \nabla_V U - [U, V]$ vanishes, or equivalently that its Hessian is valued in $\Gamma(T^*M \odot T^*M)$. By far the most important example of such a connection is the Levi-Civita connection of a Riemannian metric $\mathscr{g}$; in this case the Hessian takes the form $\langle \nabla^2_x f, U_x \otimes V_x\rangle = \mathscr{g}(\nabla_{U_x}\mathrm{grad}^{\mathscr{g}} f, V_x)$. Torsion-free connections are relevant to our study of SDEs in that they correspond to the splittings of (1.12), i.e. a linear left inverse $\mathscr{q}$ to $\imath$ or a linear right inverse $\jmath$ to $\mathscr{p}$

$$0 \longrightarrow TM \underset{\mathscr{q}}{\overset{\imath}{\rightleftarrows}} \mathbb{T}M \underset{\mathscr{p}}{\overset{\jmath}{\rightleftarrows}} TM \odot TM \longrightarrow 0 \tag{1.36}$$

The existence of the bundle maps $\jmath$ and q are equivalent to one another and to the the isomorphism $(q, p)\colon \mathbb{T}M \to TM \oplus (TM \odot TM)$ (this is the well-known splitting lemma [Hat02, p.147], valid in the category of vector bundles). A torsion-free connection ∇ on M is equivalent to a splitting by setting

$$(q_x L_x) f \coloneqq L_x f - \langle \nabla_x^2 f, p_x L_x \rangle \tag{1.37}$$

We recall that, given $V \in \Gamma TM$, $W_x \in T_x M$, the "composition" $U_x(V) \in \mathbb{T}_x M$ is defined by $U_x(V) f \coloneqq U_x(y \mapsto V_y f)$, and we have

$$p_x(U_x(V)) = U_x \odot V_x, \quad q_x(U_x(V)) = \nabla_{U_x} V \tag{1.38}$$

Using that $\partial_x^2 \varphi_{ij} = \partial_x \varphi_i(\partial \varphi_j)$ and (1.35) we have

$$p_x \partial_x^2 \varphi_{ij} = \partial_x \varphi_i \odot \partial_x \varphi_j, \quad q_x \partial_x^2 \varphi_{ij} = \Gamma_{ij}^k(x) \partial_x \varphi_k \tag{1.39}$$

Another way to view this correspondence is by $\jmath^* \mathbb{d}_x f = \nabla_x^2 f$.

Now, given symmetric connections on N and M, a field of Schwartz morphisms $\mathbb{F} \in \Gamma\mathrm{Sch}(N, M)$ can be viewed as a field of block matrices

$$\begin{bmatrix} F & G \\ 0 & F \otimes F \end{bmatrix} (x, z)\colon T_z N \oplus (T_z N \odot T_z N) \to T_x M \oplus (T_x M \odot T_x M) \tag{1.40}$$

One can then require that $G \equiv 0$, so that $\mathbb{F}$ reduces to F, which defines the *Itô equation*

$$\mathrm{d}X_t = F(X_t, Z_t)\mathrm{d}Z_t \tag{1.41}$$

Such equations have been considered in [É90]. The data needed to define this equation is the same as that involved in the definition of the Stratonovich equation (1.5), namely an element of $\Gamma\mathrm{Hom}(TN, TM)$, but the meaning of the equation depends on the connections on N and M. In local coordinates, using (1.39) to specify $\mathbb{F}_{\alpha\beta}^k$ in (1.21) to the case $G \equiv 0$, this equation takes the form

$$\begin{aligned} \mathrm{d}X_t^k = {} & F_\gamma^k(X_t, Z_t)\mathrm{d}Z_t^\gamma \\ & + \tfrac{1}{2}\big({}^N\Gamma_{\alpha\beta}^\gamma(Z_t) F_\gamma^k(X_t, Z_t) - {}^M\Gamma_{ij}^k(X_t) F_\alpha^i F_\beta^j(X_t, Z_t)\big)\mathrm{d}[Z^\alpha, Z^\beta]_t \end{aligned} \tag{1.42}$$

Note that if the Christoffel symbols on both manifolds vanish the above equation reduces to its first line; however, unless a manifold is flat a chart cannot in general be chosen so that the Christoffel symbols vanish (except for at a single chosen point: these are called normal coordinates). Itô equations can be equivalently defined through the Itô integral

$$\int_0^\cdot \langle H_s, \mathrm{d}X_s \rangle \coloneqq \int_0^\cdot \langle q^* H_s, \mathbb{d}X_s \rangle \tag{1.43}$$

by proceeding as in (1.7).

Recall that an (M, ∇)-valued semimartingale is a *local martingale* if for all $f \in C^\infty M$

$$f(X) - \int_0^\cdot \langle p^* \nabla_{X_s}^2 f, \mathbb{d}X_s \rangle \tag{1.44}$$

is a real-valued local martingale (the integral is to be interpreted as half the quadratic variation of X along the bilinear form $\nabla^2 f$); this property coincides with the usual local martingale property when M is a vector space. In local coordinates an application of (1.35) and (1.17) shows that the local martingale property corresponds to the requirement that

$$\mathrm{d}X_t^k + \tfrac{1}{2}\Gamma_{ij}^k(X_t)\mathrm{d}[X^i, X^j]_t \tag{1.45}$$

be a real-valued local martingale for each k. The Itô integral (1.43) and Itô equations (1.41) on manifolds behave well w.r.t. local martingales: if the integrand or driver is a local martingale, so is the integral or solution; this is again seen in local coordinates (1.42).

In the following example we examine the case of diffusions, defined using Itô equations, in which the issue of the drift not being a vector is (partially) resolved:

Example 1.4 (Itô diffusion)**.** Example 1.3 specified to the above case (M has a symmetric connection, $G \equiv 0$ in (1.40)) becomes the equation

$$\mathrm{d}X_t = \sigma_\gamma(X_t,t)\mathrm{d}W_t^\gamma + \mu(X_t,t)\mathrm{d}t \tag{1.46}$$

where now $\mu(x,t) = \mathbb{F}(x,t) \in T_xM$ can legitimately be referred to as the "drift vector". Note however that in an arbitrary chart φ the drift will still carry a correction term:

$$\mathrm{d}X_t^k = \sigma_\gamma^k(X_t,t)\mathrm{d}W_t^\gamma + \left(\mu^k(X_t,t) - \frac{1}{2}\sum_{\gamma=1}^{n}\Gamma_{ij}^k(X_t)\sigma_\gamma^i\sigma_\gamma^j(X_t,t)\right)\mathrm{d}t \tag{1.47}$$

which reduces to the ordinary Itô lemma if $M = \mathbb{R}^m$ and the chart φ is a diffeomorphism of $\mathbb{R}^m$. The ${}^N\Gamma_{\alpha\beta}^\gamma$'s do not appear since the driver is already valued in a Euclidean space. The data needed to define such an equation coincides with that needed for (1.5), so we can define the bundle

$$\mathrm{Diff}_{\mathrm{Ito}}^n M := \mathrm{Diff}_{\mathrm{Strat}}^n M \to M \times \mathbb{R}_{\geq 0} \tag{1.48}$$

already defined in (1.9). Crucially, however, the Stratonovich and Itô calculi give different meanings to the equation defined by a section of this bundle; in particular, a torsion-free connection on M is required in the latter case. The "Itô" and "Strat" therefore do not represent differences in the bundles, which are identical, but only serve as a reminder of which calculus is being used to give the section the meaning of an SDE.

Itô equations on manifolds are the true generalisation of their Euclidean space-valued counterparts, but have the disadvantage of only being defined w.r.t. a specific connection. For instance, if $F \in \Gamma\mathrm{Diff}_{\mathrm{Ito}}^n$, M is Riemannian with M' a Riemannian submanifold s.t. for all z and $x \in M'$, $F(z,x)$ maps to T_xM', F does not in general define an Itô equation on M', since the Riemannian connection on M' is not in general the restriction of that of M. However, F, seen as a field of Schwartz morphisms, does define a Schwartz-Meyer equation on M' (with a G term that is in general non-zero w.r.t. to the Riemannian connection on M').

In the following table we summarise the advantages of these three ways of representing SDEs on manifolds:

	Stratonovich	Schwartz-Meyer/2-jet	Itô
Does not require ∇	✓	✓	
Uses Itô integration		✓	✓
Coefficients are vectors	✓		✓

It is natural to ask how these three types of equations are related to one another. In the case of diffusions, there exists a commutative diagram of bijections

$$
\begin{array}{ccc}
 & \Gamma\mathrm{Diff}^n_{\mathrm{Sch}} M & \\
\overset{\mathpzc{a}}{\nearrow} & & \overset{\mathpzc{b}}{\searrow} \\
\Gamma\mathrm{Diff}^n_{\mathrm{Strat}} M & \xrightarrow{\quad \mathpzc{c} \quad} & \Gamma\mathrm{Diff}^n_{\mathrm{Ito}} M
\end{array}
\tag{1.49}
$$

All three $\mathpzc{a}, \mathpzc{b}, \mathpzc{c}$ are the identity on the diffusion coefficients. The behaviour of $\mathpzc{a}, \mathpzc{b}, \mathpzc{c}$ on the Stratonovich, Schwartz-Meyer and Itô drifts is explained below

$$
\mathpzc{a} b \coloneqq b + \frac{1}{2}\sum_{\gamma=1}^{n} \sigma_\gamma(\sigma_\gamma), \quad \mathpzc{b}\eta \coloneqq \mathpzc{q}\eta, \quad \mathpzc{c} b \coloneqq b + \frac{1}{2}\sum_{\gamma=1}^{n} \nabla_{\sigma_\gamma}\sigma_\gamma \tag{1.50}
$$

Note that, while $\mathpzc{b}$ and $\mathpzc{c}$ depend on the connection, $\mathpzc{a}$ does not. If $\eta = \mathbb{F}_0 + \frac{1}{2}\sum_{\gamma=1}^n \mathbb{F}_{\gamma\gamma}$ is a Schwartz-Meyer drift, (1.15) and (1.38) force $\eta - \frac{1}{2}\sum_{\gamma=1}^n \sigma_\gamma(\sigma_\gamma)$ to lie in T_xM, which is thus $\mathpzc{a}^{-1}\eta$. Moreover, we have $\mathpzc{b}^{-1}\mu = \mathpzc{i}\mu + \frac{1}{2}\sum_{\gamma=1}^n \mathpzc{j}(\sigma_\gamma \odot \sigma_\gamma)$ and $\mathpzc{c}^{-1}\mu = b - \frac{1}{2}\sum_{\gamma=1}^n \nabla_{\sigma_\gamma}\sigma_\gamma$. $\mathpzc{a}, \mathpzc{b}, \mathpzc{c}$ define correspondences of SDEs in the sense that solutions are preserved (e.g. X is a solution of $F \in \mathrm{Diff}^n_{\mathrm{Strat}} M$ if and only if X is a solution of $\mathpzc{a}F$, and the same for $\mathpzc{b}, \mathpzc{c}$). This is immediate by the expression of such equations in charts, by (1.38) and the usual Itô-Stratonovich conversion formula.

Remark 1.5. What makes Itô-Stratonovich conversion formulae difficult to state in the case of a general manifold-valued semimartingale driver Z, is that the change of calculus involves the emergence of new drivers which are not naturally valued in the manifold where Z is valued (the quadratic covariation of Z). Nevertheless, the map $\mathpzc{a}$ can be defined in this general setting [É89, Lemma 7.22], though its inverse cannot canonically.

1.2 Manifolds embedded in $\mathbb{R}^d$

In this chapter we will mostly be concerned with manifolds embedded in $\mathbb{R}^d$: these can be studied using the extrinsic, canonical, $\mathbb{R}^d$-coordinates instead of non-canonical local ones. Let M be an m-dimensional smooth manifold embedded in $\mathbb{R}^d$. We assume M to be locally given by a non-degenerate Cartesian equation $F(x) = 0$: M can be described globally in this way if and only if it is closed and its embedding has trivial normal bundle; therefore, to preserve generality, we only assume F to be local. Throughout this chapter the letter x will denote a point in $\mathbb{R}^d$ and the letter y a point in M. Thus $F \colon \mathbb{R}^d \to \mathbb{R}^{d-m}$ is a submersion, which implies $JF(x)JF(x)^\intercal \in GL(\mathbb{R}, d-m)$ for all $x \in \mathbb{R}^d$ ($JF(x) \in \mathbb{R}^{(d-m)\times d}$ the Jacobian of F at x):

$$
JF(x)JF(x)^\intercal v^\intercal = 0 \;\Rightarrow\; (vJF(x))(vJF(x))^\intercal = vJF(x)JF(x)^\intercal v^\intercal = 0 \;\Rightarrow\; v = 0 \tag{1.51}
$$

Let π, defined on a tubular neighbourhood T of M in $\mathbb{R}^d$ be the Riemannian submersion

$$\pi(x) \coloneqq \arg\min\{|x-y| : y \in M\} \tag{1.52}$$

This map can be seen to exist by using the normal exponential map defined in [Pet06, p.132], and is constant on the affine $(d-m)$-dimensional slices of T which intersect M orthogonally: this is because the fibre $\pi^{-1}(y)$ coincides with the union of all geodesics in $\mathbb{R}^d$ (i.e. straight line segments) which start at y, with initial velocity orthogonal to M, each taken for t in some open interval containing 0. It is important also to remember that π is unique given the embedding of M (on a thin enough T such that it is well defined), whereas F is not canonically determined. In what follows we will be concerned with understanding which quantities are dependent on the chosen F and which instead only depend on the embedding of M. The only properties of π that we will need are that

$$F \circ \pi \equiv 0, \qquad \pi|_M = \mathbb{1}_M \Rightarrow \pi \circ \pi \equiv \pi \tag{1.53}$$

Differentiating these (the second up to order 2) we obtain

$$\begin{aligned} \frac{\partial F}{\partial x^h}(\pi(x))\frac{\partial \pi^h}{\partial x^k}(x) &= 0 \\ \frac{\partial \pi}{\partial x^h}(\pi(x))\frac{\partial \pi^h}{\partial x^k}(x) &= \frac{\partial \pi}{\partial x^k}(x) \\ \frac{\partial^2 \pi}{\partial x^a \partial x^b}(\pi(x))\frac{\partial \pi^a}{\partial x^i}\frac{\partial \pi^b}{\partial x^j}(x) + \frac{\partial \pi}{\partial x^h}(\pi(x))\frac{\partial^2 \pi^h}{\partial x^i \partial x^j}(x) &= \frac{\partial^2 \pi}{\partial x^i \partial x^j}(x) \end{aligned} \tag{1.54}$$

If $V_y \in T_M$ and X is a smooth curve s.t. $X_0 = y$ and $\dot{X}_0 = V(y)$, differentiating $\pi(X_t) = X_t$ results in $J\pi(y) = V_y$: this shows that $J\pi|_{TM} = \mathbb{1}_{TM}$. By a similar argument, the fact that $\pi^{-1}(y)$ is a straight line segment that intersects M orthogonally implies that $J\pi|_{T^\perp M} = \mathbb{1}_{T^\perp M}$ ($T_y^\perp M$ the normal bundle of M at y). These two statements mean that

$$P(y) = J\pi(y) \quad \text{for } y \in M \tag{1.55}$$

where $P(y)\colon T_y\mathbb{R}^d \to T_yM$ is the orthogonal projection onto the tangent bundle of M, which can be defined in terms of F as

$$\begin{aligned} &P(x) \coloneqq \mathbb{1} - Q(x) \qquad \text{where} \\ &Q(x) \coloneqq JF^\intercal(x)(JF(x)JF^\intercal(x))^{-1}JF(x) \in \mathbb{R}^{d\times d} \qquad \text{and we have} \\ &PQ(x) = 0 = QP(x), \quad QQ(x) = Q(x) = Q^\intercal(x), \quad PP(x) = P(x) = P^\intercal(x) \end{aligned} \tag{1.56}$$

The notation is borrowed from [CDL15]. Note that we can use F to define P, Q on a tubular neighbourhood of M, but these will only be independent of F on M. $Q(y)\colon T_y\mathbb{R}^d \to T_y^\perp M$ is the orthogonal projection onto the normal bundle. Another consequence of (1.54) (evaluated at $y \in M$) that will be useful is that, for

$V_y, W_y \in T_y\mathbb{R}^d$, and denoting $\overline{U}_y = P(y)U_y, \check{U}_y = Q(y)U_y$

$$\begin{aligned} &\frac{\partial^2\pi}{\partial x^i \partial x^j}(y)\overline{V}_y^i\overline{W}_y^j \in T_y^\perp M, \quad \frac{\partial^2\pi}{\partial x^i \partial x^j}(y)\overline{V}_y^i\check{W}_y^j \in T_y M, \frac{\partial^2\pi}{\partial x^i \partial x^j}(y)\check{V}_y^i\check{W}_y^j = 0 \\ \implies &\frac{\partial^2\pi}{\partial x^i \partial x^j}(y)V_y^i W_y^j = \underbrace{\frac{\partial^2\pi}{\partial x^i \partial x^j}(y)(\overline{V}_y^i\overline{W}_y^j)}_{\in T_y^\perp M} + \underbrace{\frac{\partial^2\pi}{\partial x^i \partial x^j}(y)(\overline{V}_y^i\check{W}_y^j + \check{V}_y^i\overline{W}_y^j)}_{\text{both terms } \in T_y M} \end{aligned} \tag{1.57}$$

Actually, to show that the third term statement in the first line, we need a separate argument:

Remark 1.6. Let $U \subseteq \mathbb{R}^d$, $f \colon U \to \mathbb{R}^e$, $y \in U$, $A_y, B_y \in T_y\mathbb{R}^d$. Then

$$\frac{\partial^2 f}{\partial x^i \partial x^j}(y)A_y^i B_y^j \tag{1.58}$$

only depends on f restricted to the affine plane (or line) centred in y and spanned by A_y, B_y. Indeed, intending with A the extension of A_y to a constant vector field on U, we can write

$$\frac{\partial^2 f}{\partial x^i \partial x^j}(y)A_y^i B_y^j = \frac{\partial}{\partial x^j}\bigg|_y \Big(\underbrace{\frac{\partial f}{\partial x^i}(x)A_x^i}_{=:g(x)}\Big)B_y^j \tag{1.59}$$

This is the directional derivative of g at y in the direction B_y, and therefore only depends on the restriction of g to the affine line $\operatorname{span}\{B_y\}$. But $g(x)$ is itself a directional derivative, and only depends on f restricted to the affine line $\operatorname{span}\{A_x\}$. Thus the whole expression only depends on f restricted to $\bigcup_{x \in \operatorname{span}\{B_y\}} \operatorname{span}\{A_x\} = \operatorname{span}\{A_y, B_y\}$.

This shows that the term in question only depends on π restricted to $\operatorname{span}\{\check{V}_y, \check{W}_y\}$, which is the constant y map, whose derivatives therefore vanish.

Remark 1.7. The other terms appearing in (1.57) have a description that should be more familiar to differential geometers:

$$\begin{aligned} \frac{\partial^2\pi}{\partial x^i \partial x^j}(y)\overline{V}_y^i\overline{W}_y^j &= {}^{\mathbb{R}^d}\nabla^\perp_{\overline{V}_y}\overline{W} := Q(y)^{\mathbb{R}^d}\nabla_{\overline{V}_y}\overline{W} = \mathrm{I\!I}(\overline{V}_y, \overline{W}_y) \\ -\frac{\partial^2\pi}{\partial x^i \partial x^j}(y)\overline{V}_y^i\check{W}_y^j &= {}^{\mathbb{R}^d}\nabla^\top_{\overline{V}_y}\check{W} := P(y)^{\mathbb{R}^d}\nabla_{\overline{V}_y}\check{W} \end{aligned} \tag{1.60}$$

where ${}^{\mathbb{R}^d}\nabla$ denotes covariant differentiation in $\mathbb{R}^d$ (i.e. just directional differentiation). Notice this is true independently of the chosen extension of $\overline{W}, \check{W}$ to local vector fields, a priori needed to give the RHSs a meaning. The first term is the second fundamental form of $\overline{V}_y, \overline{W}_y$ [Lee97, p.134], whereas the second term is the second fundamental tensor [Joso5, Def. 3.6.1]. If M is an open set of an affine subspace of M, π is a linear map and both terms vanish. We prove the first of the two equalities in (1.60), the second is proved similarly:

$$Q(y)^{\mathbb{R}^d}\nabla_{\overline{V}_y}\overline{W} = Q_j(y)\frac{\partial \overline{W}^j}{\partial x^i}(y)\overline{V}_y^i = -\frac{\partial Q_j}{\partial x^i}(y)\overline{W}_y^j\overline{V}_y^i = \frac{\partial^2\pi}{\partial x^i \partial x^j}(y)\overline{V}_y^i\overline{W}_y^j \tag{1.61}$$

where the second equality follows from the fact that $Q\overline{W} = 0$ (and that the derivative is taken in a tangential direction, i.e. $\overline{V}_y \in T_y M$), and the last equality is given by (1.64) below. Note that the terms of (1.60) are extrinsic, in the sense that they depend on the embedding of M, unlike

$$ {}^{M}\nabla_{\overline{V}_y}\overline{W}_y = P(y)^{\mathbb{R}^d}\nabla_{\overline{V}_y}\overline{W} \tag{1.62} $$

the Levi-Civita connection of the Riemannian metric on M, which is intrinsic to M.

Finally, it will be necessary to consider the relationship between the derivatives of P, Q and the second derivatives of π. We differentiate (1.55) at time 0 along a smooth curve Y_t in M with $Y_0 = 0$ and $\dot{Y}_0 = \overline{V}_y \in T_yM$ and obtain

$$ \frac{\partial P_k}{\partial x^h}(y)\overline{V}_y^h = \frac{\partial^2 \pi}{\partial x^i \partial x^j}(y)\overline{V}_y^i \tag{1.63} $$

from which we obtain, for $W \in T_yM$

$$ \begin{aligned} -\frac{\partial Q_k}{\partial x^h}(y)\overline{V}_y^h\overline{W}_y^k &= \frac{\partial P_k}{\partial x^h}(y)\overline{V}_y^h\overline{W}_y^k = \frac{\partial^2 \pi^k}{\partial x^i \partial x^j}(y)\overline{V}_y^i\overline{W}_y^j \in T_y^{\perp}M \\ -\frac{\partial Q_k}{\partial x^h}(y)\overline{V}_y^h\breve{W}_y^k &= \frac{\partial P_k}{\partial x^h}(y)\overline{V}_y^h\breve{W}_y^k = \frac{\partial^2 \pi}{\partial x^i \partial x^j}(y)\overline{V}_y^i\breve{W}_y^j \in T_yM \end{aligned} \tag{1.64} $$

where we have used (1.57).

We now consider a setup $\mathcal{S} = (\Omega, \mathcal{F}, P)$ satisfying the usual conditions, W an n-dimensional Brownian motion defined on $\mathcal{S}$. Consider the W-driven diffusion Stratonovich SDE

$$ \mathrm{d}X_t^k = \sigma_\gamma^k(X_t, t) \circ \mathrm{d}W_t^\gamma + b^k(X_t, t)\mathrm{d}t, \quad X_0 = y_0 \in M \tag{1.65} $$

As already discussed in Section 1.1, the natural condition on σ_γ, b which guarantees that X will stay on M for its lifetime is their tangency to M:

$$ Q(y)\sigma_\gamma(y,t) = 0 = Q(y)b(y,t) \quad \text{for all } y \in M, t \geq 0, \gamma = 1, \dots, n \tag{1.66} $$

Our focus, however, will be mostly on the Itô SDE

$$ \mathrm{d}X_t^k = \sigma_\gamma^k(X_t, t)\mathrm{d}W_t^\gamma + \mu^k(X_t, t)\mathrm{d}t, \quad X_0 = y_0 \in M \tag{1.67} $$

with smooth coefficients defined in $[0, +\infty) \times \mathbb{R}^d$; we do not assume them to be globally Lipschitz, so the solution might only exist up to a positive stopping time, not in general bounded from below by a positive deterministic constant. We are interested in deriving the "tangency condition" for the above SDE, i.e. a condition on the coefficients that will guarantee that the solution will not leave M. One way to impose this is to convert (1.67) to Stratonovich form

$$ \mathrm{d}X_t^k = \sigma_\gamma^k(X_t, t) \circ \mathrm{d}W_t^\gamma + \left(\mu^k - \frac{1}{2}\sum_{\gamma=1}^{n} \sigma_\gamma^h \frac{\partial \sigma_\gamma^k}{\partial x^h}\right)(X_t, t)\mathrm{d}t, \quad X_0 = y_0 \in M \tag{1.68} $$

and require (1.66):

$$\begin{cases} Q_k(y)\sigma^k_\gamma(y,t) = 0 \\ Q_k(y)\left(\mu^k - \frac{1}{2}\sum_{\gamma=1}^n \sigma^h_\gamma \frac{\partial \sigma^k_\gamma}{\partial x^h}\right)(y,t) = 0 \end{cases} \tag{1.69}$$

Now, given that $Q\sigma_\alpha$ vanishes on M, all its directional derivatives along the tangent directions σ_β will too, which gives, using (1.64)

$$0 = \frac{\partial(Q\sigma_\alpha)}{\partial x^h}\sigma^h_\beta = \frac{\partial Q_i}{\partial x^j}\sigma^i_\alpha\sigma^j_\beta + Q_k\frac{\partial \sigma^k_\alpha}{\partial x^h}\sigma^h_\beta \implies Q_k\frac{\partial \sigma^k_\alpha}{\partial x^h}\sigma^h_\beta = \frac{\partial^2 \pi}{\partial x^i \partial x^j}\sigma^i_\alpha\sigma^j_\beta \quad \text{on } M \tag{1.70}$$

We can thus reformulate the second equation in (1.69) to obtain

$$\begin{cases} Q_k(y)\sigma^k_\gamma(y,t) = 0 \\ Q_k(y)\mu^k(y,t) = \frac{1}{2}\sum_{\gamma=1}^n \frac{\partial^2 \pi}{\partial x^i \partial x^j}(y)\sigma^i_\gamma\sigma^j_\gamma(y,t) \end{cases} \tag{1.71}$$

This is useful because it removes the reliance of this constraint on the derivatives of σ, and can be interpreted as saying that the diffusion coefficients must be tangent to M and the Itô drift must instead lie on the space parallel to the tangent space of M, displaced by an amount which depends on the second fundamental form of M applied to the diffusion coefficients.

Remark 1.8 (Tangency of a second-order differential operator). (1.71) can also be derived by writing the second order tangency condition for $L^k_y\partial_y x_k + L^{ij}_y\partial^2_y x_{ij} = L_y \in \mathbb{T}_y\mathbb{R}^d$ to belong to $\mathbb{T}_y M$: this is done by writing $\mathbb{T}_y\pi L_y = L_y$ in $\mathbb{R}^d$-coordinates as

$$\begin{bmatrix} L^h_y \\ L^{ab}_y \end{bmatrix} = \begin{bmatrix} \frac{\partial \pi^h}{\partial x^k} & \frac{\partial^2 \pi^h}{\partial x^i \partial x^j} \\ 0 & \frac{\partial \pi^a}{\partial x^i}\frac{\partial \pi^b}{\partial x^j} \end{bmatrix}(y)\begin{bmatrix} L^k_y \\ L^{ij}_y \end{bmatrix} \tag{1.72}$$

and then applying it to $L_y = \sigma_\gamma(y,t), \eta(y,t)$, given in terms a field of Schwartz morphisms $\mathbb{F}$ as

$$\sigma^k_\gamma = \mathbb{F}^k_\gamma, \quad \eta^k = \mathbb{F}^k_0 + \frac{1}{2}\sum_{\gamma=1}^n \mathbb{F}^k_{\gamma\gamma} \tag{1.73}$$

Note that it would instead be incorrect to split $\mathbb{F}$ according to the Euclidean connection into a matrix with F and G terms as in (1.40), and then to require that F and G map to TM, since the splitting of $\mathbb{F}$ according to the connection on M will be different, i.e. the diagram

$$\begin{array}{ccc} \mathbb{T}\mathbb{R}^d & \xleftrightarrow{\cong} & T\mathbb{R}^d \oplus (T\mathbb{R}^d \odot T\mathbb{R}^d) \\ \uparrow & & \uparrow \\ \mathbb{T}M & \xleftrightarrow{\cong} & TM \oplus (TM \odot TM) \end{array} \tag{1.74}$$

does not commute.

We now compute the Hessian for embedded M: for $f \in C^\infty M$ we have

$$\langle {}^M\nabla^2_y f, \overline{V}_y \otimes \overline{U}_y \rangle = \langle {}^{\mathbb{R}^d}\nabla^2_y (f \circ \pi), \overline{V}_y \otimes \overline{U}_y \rangle \tag{1.75}$$

where we have used (1.34), (1.62) to reduce this to a computation of directional derivatives, and finally (1.64) (the argument is similar to (1.61)). ${}^{\mathbb{R}^d}\nabla^2$ of course is just the ordinary Hessian. We can now compute ${}^M\!q$, the splitting appearing in (1.36) w.r.t. the connection on M: if $\mathbb{T}_y M \ni L_y = L^k_y \partial_y x_k + L^{ij}_y \partial^2_y x_{ij}$, using (1.37) yields

$$\begin{aligned} ({}^M\!q_y L_y) f &= L_y(f) - \langle {}^M\nabla^2_y f, p_y L_y \rangle \\ &= L_y(f \circ \pi) - \langle {}^{\mathbb{R}^d}\nabla^2_y (f \circ \pi), L^{ij}_y \partial^2_y x_{ij} \rangle \\ &= \frac{\partial f}{\partial x^h}(y) \frac{\partial \pi^h}{\partial x^k}(y) L^k_y \end{aligned} \tag{1.76}$$

which means

$${}^M\!q_y = P(y) \circ {}^{\mathbb{R}^d}\!q_y \colon \mathbb{T}_y M \to T_y M \tag{1.77}$$

Therefore the condition on an arbitrary Schwartz morphism of being Itô w.r.t. to the Riemannian connection on M in the sense of Example 1.4 is ${}^M\!q \circ \mathbb{F} \circ {}^{\mathbb{R}^d}\!j = 0$, or ${}^M\!q \mathbb{F}_{\alpha\beta} = 0$, which in $\mathbb{R}^d$-coordinates is

$$P_k(y) \mathbb{F}^k_{\alpha\beta}(y,t) = 0 \tag{1.78}$$

Compare this with the stronger condition of $\mathbb{F}$ of being Itô w.r.t. to the connection on $\mathbb{R}^d$, which is $\mathbb{F}^k_{\alpha\beta}(y,t) = 0$. Thus, given an Itô equation $\mathbb{F}$ on M, defined as in (1.46) ($\sigma_\gamma = \mathbb{F}_\gamma$, $\overline{\mu} = \mathbb{F}_0$) we have that the drift in $\mathbb{R}^d$ of such equation is given by $\overline{\mu}^k + \frac{1}{2}\sum_{\gamma=1}^n \mathbb{F}^k_{\gamma\gamma}$, with the first term tangent to M and the second orthogonal to M, and equal to $\frac{1}{2}\sum_{\gamma=1}^n \frac{\partial \pi^h}{\partial x^i \partial x^j} \sigma^i \sigma^j_\gamma$, by Remark 1.8 and (1.78). Therefore an Itô equation on M with coefficients $\sigma_\gamma, \overline{\mu}$ is read in ambient coordinates as

$$\mathrm{d}X^k_t = \sigma^k_\gamma(Y_t,t)\mathrm{d}W^\gamma_t + \left(\overline{\mu}^k + \frac{1}{2}\sum_{\gamma=1}^n \frac{\partial^2 \pi^k}{\partial x^i \partial x^j}\sigma^i_\gamma \sigma^j_\gamma\right)(Y_t,t)\mathrm{d}t \tag{1.79}$$

Notice that the tangential part of the $\mathbb{R}^d$-drift, $\overline{\mu}$, is arbitrary, while its orthogonal part is determined by the diffusion coefficients, and the condition that the solution remain on M. The notion of M-valued local martingale also has a description in terms of ambient coordinates [É89, ¶4.10]: for an M-valued Itô process (such as the solution to (1.79)) the local martingale property is equivalent to requiring that the drift be orthogonal to M at each point (and thus determined by the diffusion coefficients; for (1.79) this means $\overline{\mu} = 0$). This condition is very reminiscent of the property of geodesics of having acceleration orthogonal to M [Lee97, Lemma 8.5].

Using all (1.50) and (1.77) it is easy to verify that converting between Stratonovich, Schwartz-Meyer and Itô equations on M is equivalent when treating the equations as being valued in M or in $\mathbb{R}^d$. By this we mean that, denoting with $\mathrm{Diff}^n_{\mathrm{Strat},M}\mathbb{R}^d$ the bundle of Stratonovich equations on $\mathbb{R}^d$ which restrict to equations on M (and analogously for the other two diffusion bundles) the maps a, b, c of (1.49) fit into the commutative

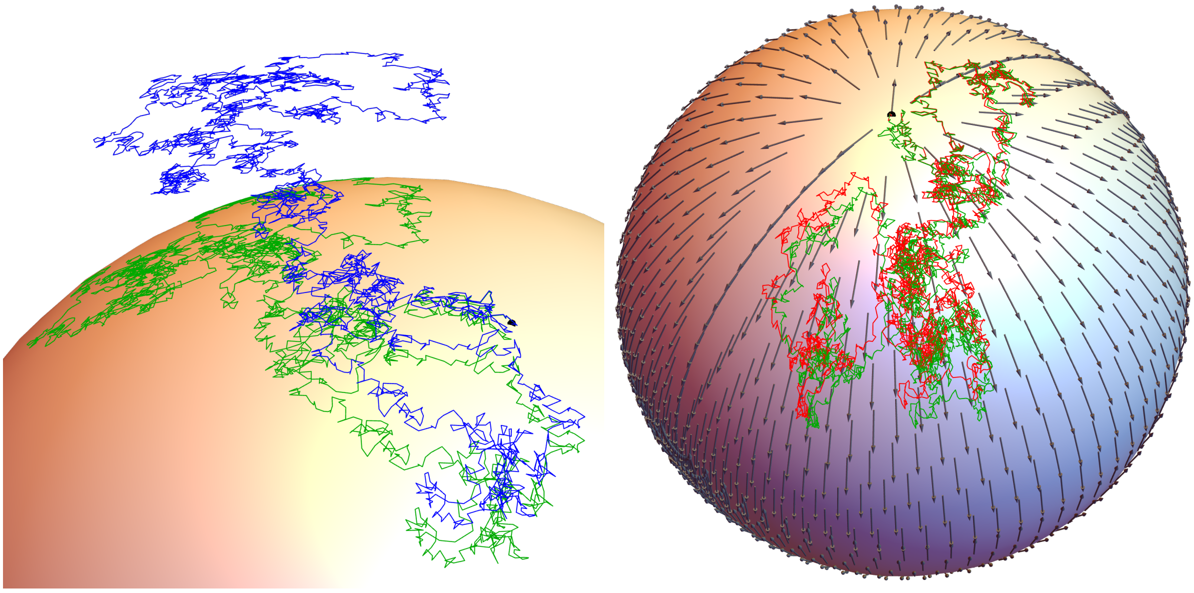


Figure 1.1: On the left a sample path of the solution to the Itô equation (blue) with the two diffusion coefficients $2(x^2+y^2+z^2)^{-1}(-y,x,0), 2(x^2+y^2+z^2)^{-1}(0,-z,y)$, which are tangent to $S^2 \hookrightarrow \mathbb{R}^3$, zero drift and initial condition $(0,1,0)$; in the same plot a sample path (using the same random seed) of the solution to the Stratonovich equation (green) defined by the same vector fields and initial condition. The solution to the Itô equation drifts radially outwards, while the solution to the Stratonovich equation remains on S^2. On the right we compare the same Stratonovich path with a sample path of the solution to the Itô equation (red) with the same diffusion coefficients and initial condition, but with the orthogonal drift term necessary to keep the solution on S^2 (1.71). The resulting solution is an S^2-valued local martingale, while the solution to the Stratonovich equation is not: this is illustrated by plotting the vector field on S^2 given by tangential component of the Itô drift possessed by the Stratonovich equation: this can be viewed as a manifold-valued drift component.

diagram

$$
\begin{array}{ccccc}
 & & \Gamma\mathrm{Diff}^n_{\mathrm{Sch},M}\mathbb{R}^d & & \\
 & \overset{\mathbb{R}^d_{\mathfrak{a}}}{\swarrow\!\!\nearrow} & & \overset{\mathbb{R}^d_{\mathfrak{b}}}{\nwarrow\!\!\searrow} & \\
\Gamma\mathrm{Diff}^n_{\mathrm{Strat},M}\mathbb{R}^d & & \xleftrightarrow{\ \mathbb{R}^d_{\mathfrak{c}}\ } & & \Gamma\mathrm{Diff}^n_{\mathrm{Ito},M}\mathbb{R}^d \\
\downarrow & & \downarrow & & \downarrow \\
 & & \Gamma\mathrm{Diff}^n_{\mathrm{Sch}}M & & \\
 & \overset{M_{\mathfrak{a}}}{\swarrow\!\!\nearrow} & & \overset{M_{\mathfrak{b}}}{\nwarrow\!\!\searrow} & \\
\Gamma\mathrm{Diff}^n_{\mathrm{Strat}}M & & \xleftrightarrow{\ M_{\mathfrak{c}}\ } & & \Gamma\mathrm{Diff}^n_{\mathrm{Ito}}M
\end{array}
\tag{1.80}
$$

where vertical arrows denote restriction. An embedding argument immediately allows us to extend this assertion to the case where $\mathbb{R}^d$ is substituted with a Riemannian manifold of which M is a Riemannian submanifold. This confirms there is no ambiguity in converting an M-valued SDE between its various forms.

Example 1.9 (Time dependent submanifold)**.** Observe that the tangency conditions (1.66) and (1.71) can be written respectively as

$$\begin{cases}(\mathbb{1}-J\widetilde{\pi})\sigma_\gamma=0\\(\mathbb{1}-J\widetilde{\pi})b=0\end{cases}\qquad\begin{cases}(\mathbb{1}-J\widetilde{\pi})\sigma_\gamma=0\\(\mathbb{1}-J\widetilde{\pi})\mu=\dfrac{1}{2}\displaystyle\sum_{\gamma=1}^{n}\frac{\partial^2\widetilde{\pi}}{\partial x^i\partial x^j}(y)\sigma_\gamma^i\sigma_\gamma^j\end{cases}\tag{1.81}$$

for *any* smooth map $\widetilde{\pi}$ defined on a tubular neighbourhood of M, with values in M, s.t. $\widetilde{\pi}|_M=\mathbb{1}_M$, by the same exact reasoning (for the Itô case we argue as in Remark 1.8). $J\widetilde{\pi}(y)$ is no longer the orthogonal projection $P(y)$, but still restricts to the identity on T_yM for $y\in M$, i.e. it has the property that $\ker(\mathbb{1}-J\widetilde{\pi})=TM$ on M. Allowing ourselves to consider all such tubular neighbourhood projections is useful in the following application. Given that we are considering time-dependent equations, it is very natural to also allow the submanifold M to be time-dependent. Making this precise entails considering a smooth $(m+1)$-dimensional manifold $\widetilde{M}$ embedded in $\mathbb{R}^{1+d}$, s.t. $M_t:=\widetilde{M}\cap\{x_0=t\}$ is a smooth m-dimensional manifold embedded in $\{x_0=t\}\times\mathbb{R}^d$. We are looking for conditions on σ, b (resp. μ) which are sufficient to guarantee the solution to (1.65) (resp. (1.67)) X_t to belong to M_t for all t for which it is defined. We then consider the $\mathbb{R}^{1+d}$-valued process (t,X_t), which satisfies the dynamics

$$\mathrm{d}\begin{bmatrix}t\\X_t\end{bmatrix}=\begin{bmatrix}0\\\sigma(X_t,t)\end{bmatrix}\circ\mathrm{d}W_t+\begin{bmatrix}1\\b(X_t,t)\end{bmatrix}\mathrm{d}t\quad\text{resp.}\quad=\begin{bmatrix}0\\\sigma(X_t,t)\end{bmatrix}\mathrm{d}W_t+\begin{bmatrix}1\\\mu(X_t,t)\end{bmatrix}\mathrm{d}t\tag{1.82}$$

Then, given a thin enough tubular neighbourhood of $\widetilde{M}$ in $\mathbb{R}^{1+d}$ consider the map

$$\widetilde{\pi}\colon\widetilde{T}\to\widetilde{M},\quad\widetilde{\pi}(t,x)=\pi_t(x)\tag{1.83}$$

where π_t is defined as in (1.52) for the manifold M_t. Notice that this does not in general coincide with the Riemannian projection of a tubular neighbourhood onto $\widetilde{M}$, which in general has no reason to preserve time, i.e. be expressible as a union of π_t's. The identity $J\widetilde{\pi}J\widetilde{\pi}=J\widetilde{\pi}$ can be written in block matrix form as

$$\left[\begin{array}{c|ccc}1&0&\cdots&0\\\hline J\pi_t\dot{\pi}_t+\dot{\pi}_t&&J\pi_tJ\pi_t&\end{array}\right]=\left[\begin{array}{c|ccc}1&0&\cdots&0\\\hline\dot{\pi}_t&&J\pi_t&\end{array}\right]\tag{1.84}$$

where we are denoting $\dot{\pi}_t(y)=\frac{\mathrm{d}}{\mathrm{d}t}\pi_t(y)$: this implies that at each point $y\in M_t$, $\dot{\pi}_t(y)\in T_y^\perp M_t$. This choice of the tubular neighbourhood projection will be further motivated later on, in Example 1.14, Example 1.20. In view of the above considerations, we can use it anyway to impose tangency of the SDE: this results in an unmodified condition on the diffusion coefficients, and the conditions on the orthogonal components of the Stratonovich and Itô drifts are given respectively by

$$\begin{aligned}(\mathbb{1}-J\pi_t)b(y,t)&=\dot{\pi}_t(y)\\(\mathbb{1}-J\pi_t)\mu(y,t)&=\frac{1}{2}\sum_{\gamma=1}^{n}\frac{\partial^2\pi_t}{\partial x^i\partial x^j}(y)\sigma_\gamma^i\sigma_\gamma^j+\dot{\pi}_t(y)\end{aligned}\tag{1.85}$$

which keep track of the evolution of M_t in time.

1.3 Projecting SDEs

In Section 1.1 we discussed three ways of representing SDEs on manifolds: Stratonovich, Schwartz-Meyer and Itô. In this section we will define, for each one of these representations, a natural projection of the SDE onto a submanifold. We will mostly take the ambient manifold to be $\mathbb{R}^d$, which will allow us to use the theory of the previous section to derive formulae for the projections in ambient coordinates.

Let M be a smooth submanifold of the smooth manifold D, let T be a tubular neighbourhood of M in D and

$$\pi\colon T \to M \text{ a smooth map which restricts to the identity on } M \tag{1.86}$$

If D is Riemannian π can be chosen as in (1.52), but this is not necessary. Let $F \in \Gamma\mathrm{Hom}(TN, TD)$ be a Stratonovich equation driven by an N-valued semimartingale Z, where N is another smooth manifold. We can then define the M-valued Stratonovich equation

$$M \times N \ni (y, z) \mapsto \widetilde{F}(y, z) \coloneqq T_y\pi \circ F(y, z) \in \mathrm{Hom}(T_zN, T_yM) \tag{1.87}$$

We call this Stratonovich SDE the *Stratonovich projection of* F.

Now consider the Z-driven, D-valued Schwartz-Meyer equation $\mathbb{F} \in \Gamma\mathrm{Sch}(N, M)$. We can project this SDE to an SDE on M too, by

$$M \times N \ni (y, z) \mapsto \widehat{\mathbb{F}}(y, z) \coloneqq \mathbb{T}_y\pi \circ \mathbb{F}(y, z) \in \mathrm{Sch}_{z,y}(N, M) \tag{1.88}$$

We call this Schwartz-Meyer SDE the *Itô-jet projection of* $\mathbb{F}$.

If N, D and M all carry torsion-free connections we can interpret a section $F \in \Gamma\mathrm{Hom}(TN, TD)$ as an Itô equation, and similarly for

$$M \times N \ni (y, z) \mapsto \overrightarrow{F}(y, z) \coloneqq T_y\pi \circ F(y, z) \in \mathrm{Hom}(T_zN, T_yM) \tag{1.89}$$

We call this Itô SDE the *Itô-vector projection* of F. Most often D will be Riemannian, so that Levi-Civita connections are defined on both D and M. Note that the Itô-vector projection is identical to the Stratonovich projection as a map, but the interpretations of the resulting sections as SDEs differ (and the Itô-vector projected SDE depends explicitly on the connections on all three manifolds). The names of these three projections are taken from [AB16], where they were first defined.

Remark 1.10 (Naturality of the SDE projections). Assume we have a commutative square

$$\begin{array}{ccc} D & \xrightarrow{\phi} & D' \\ \downarrow{\scriptstyle\pi} & & \downarrow{\scriptstyle\pi'} \\ M & \xrightarrow{\phi|_M} & M' \end{array} \tag{1.90}$$

where ϕ a diffeomorphism, D, M, π as above, and similarly for D', M', π'. Then functoriality of T and $\mathbb{T}$

imply that the Stratonovich and Itô-jet projections are natural in the sense that the squares

$$
\begin{array}{ccc}
\mathrm{Hom}(TN, TD) & \xrightarrow{T\phi} & \mathrm{Hom}(TN, TD') \\
\big\downarrow{\scriptstyle\sim} & & \big\downarrow{\scriptstyle\sim} \\
\mathrm{Hom}(TN, TM) & \xrightarrow{T\phi|_M} & \mathrm{Hom}(TN, TM')
\end{array}
\qquad
\begin{array}{ccc}
\mathrm{Sch}(N, D) & \xrightarrow{\mathbb{T}\phi} & \mathrm{Sch}(N, D') \\
\big\downarrow{\scriptstyle\frown} & & \big\downarrow{\scriptstyle\frown} \\
\mathrm{Sch}(N, M) & \xrightarrow{\mathbb{T}\phi|_M} & \mathrm{Sch}(N, M')
\end{array}
\tag{1.91}
$$

commute. The Itô-vector projection cannot be natural in the same way, since we are still free to modify the connections on all four manifolds. However, if D, D' are Riemannian and ϕ is a global isometry, the corresponding statement does hold for the Itô-vector projection as well: this is by naturality of the Levi-Civita connection [Lee97, Proposition 5.6].

Remark 1.11 (The Itô-vector projection preserves local martingales). Although the Itô-vector projection is natural w.r.t. a smaller class of maps, it has the advantage of preserving the local martingale property: by this we mean that if the driver is a local martingale, so must the solution to the Itô-vector-projected SDE be. This is shown simply by the good behaviour of Itô equations w.r.t. manifold-valued local martingales.

Remark 1.12. One might wonder whether it is possible to "push forward" SDEs according to an arbitrary smooth and surjective map $f \colon D \to D'$. If f is a surjective function admitting a smooth right inverse ι, then we may write the pushforward of, say, the Stratonovich SDE $\mathrm{d}X = F(X, Z) \circ \mathrm{d}Z$ as $\mathrm{d}Y = F(Z, \iota(Y)) \circ \mathrm{d}Y$. This condition on f essentially corresponds to the condition (1.86). For general smooth surjective f (such as the bundle projection of a non-trivial principal bundle), however, we do not see a way of defining a new closed form SDE on D'.

We will now restrict our attention to the projections of $\mathbb{R}^d$-valued diffusions onto the embedded manifold M. Focusing on diffusions has the advantage of allowing us to use the maps (1.49) to compare the projections. In other words we can ask if the vertical rectangles in the diagram

$$
\begin{array}{ccccc}
 & & \Gamma\mathrm{Diff}^n_{\mathrm{Sch}}\mathbb{R}^d & & \\
 & \overset{\mathbb{R}^d_{\mathfrak{a}}}{\swarrow\!\!\nearrow} & \big| & \overset{\mathbb{R}^d_{\mathfrak{b}}}{\nwarrow\!\!\searrow} & \\
\Gamma\mathrm{Diff}^n_{\mathrm{Strat}}\mathbb{R}^d & & \xleftrightarrow{\mathbb{R}^d_{\mathfrak{c}}} & & \Gamma\mathrm{Diff}^n_{\mathrm{Ito}}\mathbb{R}^d \\
\big\downarrow{\scriptstyle\sim} & & \big\downarrow{\scriptstyle\frown} & & \big\downarrow{\scriptstyle\to} \\
 & & \Gamma\mathrm{Diff}^n_{\mathrm{Sch}}M & & \\
 & \overset{M_{\mathfrak{a}}}{\swarrow\!\!\nearrow} & & \overset{M_{\mathfrak{b}}}{\nwarrow\!\!\searrow} & \\
\Gamma\mathrm{Diff}^n_{\mathrm{Strat}}M & & \xleftrightarrow{M_{\mathfrak{c}}} & & \Gamma\mathrm{Diff}^n_{\mathrm{Ito}}M
\end{array}
\tag{1.92}
$$

commute (compare with (1.80), in which the equations on top already restrict to equations on M). We will show that they do not, and that all combinations of possibilities regarding their non-commutativity are possible. Examples of these cases are to be found in Subsection 1.5.1 below. We recall the notation $\overline{V}_y := P(y)V_y$, $\check{V}_y := Q(y)V_y$ and begin by considering the $\mathbb{R}^d$-valued Stratonovich SDE (1.65). By (1.55) the coefficients of the Stratonovich projection of this SDE will just be the projected coefficients: $\widetilde{\sigma}_\gamma = \overline{\sigma}_\gamma, \widetilde{b} = \overline{b}$, so that the resulting Stratonovich equation is

$$\begin{aligned}\mathrm{d}Y_t &= \overline{\sigma}_\gamma(Y_t,t)\circ \mathrm{d}W_t^\gamma + \overline{b}(Y_t,t)\mathrm{d}t, \quad Y_0 = y_0\in M\\ &= \frac{\partial\pi}{\partial x^k}(Y_t)\sigma_\gamma^k(Y_t,t)\circ\mathrm{d}W_t^\gamma + \frac{\partial\pi}{\partial x^k}(Y_t)b^k(Y_t,t)\mathrm{d}t\end{aligned} \tag{1.93}$$

Throughout this chapter we will use X for the initial SDE and Y to denote the projected SDE. Now assume we start with (1.67), and want an Itô SDE on M. We can still use the Stratonovich projection by converting the SDE to Stratonovich form as in (1.68), projecting as above, and converting back to Itô form (by (1.80) this last conversion can be seen to occur interchangeably in M or in $\mathbb{R}^d$). We have

$$\begin{aligned}\mathrm{d}Y_t &= \overline{\sigma}_\gamma(Y_t,t)\circ\mathrm{d}W_t^\gamma + P_k(Y_t)\Big(\mu^k - \frac{1}{2}\sum_{\gamma=1}^n \sigma_\gamma^h\frac{\partial\sigma_\gamma^k}{\partial x^h}\Big)(Y_t,t)\mathrm{d}t\\ &= \overline{\sigma}_\gamma(Y_t,t)\mathrm{d}W_t^\gamma + \underbrace{\Bigg(\overline{\mu} + \frac{1}{2}\sum_{\gamma=1}^n\Big(\overline{\sigma}_\gamma^l\frac{\partial\overline{\sigma}_\gamma}{\partial x^l} - \sigma_\gamma^h P_k\frac{\partial\sigma_\gamma^k}{\partial x^h}\Big)\Bigg)}_{\widetilde{\mu}}(Y_t,t)\mathrm{d}t\end{aligned} \tag{1.94}$$

Using (1.64) we can split $\widetilde{\mu}$ in its orthogonal and tangential components: on M we have

$$\begin{aligned}\widetilde{\mu} &= \overline{\mu} + \frac{1}{2}\sum_{\gamma=1}^n\Bigg(\overline{\sigma}_\gamma^l\Big(\frac{\partial P_k}{\partial x^l}\sigma_\gamma^k + P_k\frac{\partial\sigma_\gamma^k}{\partial x^l}\Big) - \sigma_\gamma^h P_k\frac{\partial\sigma_\gamma^k}{\partial x^h}\Bigg)\\ &= \overline{\mu} + \frac{1}{2}\sum_{\gamma=1}^n\Big(\frac{\partial P_k}{\partial x^l}\overline{\sigma}_\gamma^l\breve{\sigma}_\gamma^k + \frac{\partial P_k}{\partial x^l}\overline{\sigma}_\gamma^l\overline{\sigma}_\gamma^k + \overline{\sigma}_\gamma^l P_k\frac{\partial\sigma_\gamma^k}{\partial x^l} - \sigma_\gamma^h P_k\frac{\partial\sigma_\gamma^k}{\partial x^h}\Big)\\ &= \underbrace{\overline{\mu} + \frac{1}{2}\sum_{\gamma=1}^n\Big(\frac{\partial^2\pi}{\partial x^i\partial x^j}\overline{\sigma}_\gamma^i\breve{\sigma}_\gamma^j - \breve{\sigma}_\gamma^h P_k\frac{\partial\sigma_\gamma^k}{\partial x^h}\Big)}_{\in TM} + \underbrace{\frac{1}{2}\sum_{\gamma=1}^n\frac{\partial^2\pi}{\partial x^i\partial x^j}\overline{\sigma}_\gamma^i\overline{\sigma}_\gamma^j}_{\in T^\perp M}\end{aligned} \tag{1.95}$$

with implied evaluation of all terms at (y,t).

We now move on to the Itô-jet projection. Let $\mathbb{F}\in\Gamma\mathrm{Diff}^n_{\mathrm{Sch}}\mathbb{R}^d$ as in (1.73), so that the Schwartz-Meyer equation it defines coincides with the Itô equation (1.67). We can then write (1.88) using matrix notation as

$$\begin{bmatrix}\mathrm{d}Y_t\\ \frac{1}{2}\mathrm{d}[Y]_t\end{bmatrix} = \begin{bmatrix}\frac{\partial\pi}{\partial x} & \frac{\partial^2\pi}{\partial x^2}\\ 0 & \frac{\partial\pi}{\partial x}\odot\frac{\partial\pi}{\partial x}\end{bmatrix}(Y_t)\begin{bmatrix}F & G\\ 0 & F\odot F\end{bmatrix}(Y_t,t)\begin{bmatrix}\mathrm{d}W_t\\ \frac{1}{2}\mathrm{d}[W]_t\end{bmatrix} \tag{1.96}$$

of which the first line reads

$$\begin{aligned}\mathrm{d}Y_t &= \frac{\partial\pi}{\partial x^k}(Y_t)\Big(F_\gamma^k(Y_t,t)\mathrm{d}W_t^\gamma + F_0^k(Y_t,t)\mathrm{d}t + \frac{1}{2}\sum_{\gamma=1}^n G_{\gamma\gamma}(Y_t,t)\mathrm{d}t\Big)\\ &\quad+\frac{1}{2}\sum_{\gamma=1}^n\frac{\partial^2\pi}{\partial x^i\partial x^j}(Y_t)F_\gamma^i F_\gamma^j(Y_t,t)\mathrm{d}t\\ &= \overline{\sigma}_\gamma(Y_t,t)\mathrm{d}W_t^\gamma + \underbrace{\Big(\overline{\mu} + \frac{1}{2}\sum_{\gamma=1}^n\frac{\partial^2\pi}{\partial x^i\partial x^j}\sigma_\gamma^i\sigma_\gamma^j\Big)}_{\widehat{\mu}}(Y_t,t)\mathrm{d}t\end{aligned} \tag{1.97}$$

Remark 1.13. We can write the Itô-jet-projected drift $\widehat{\mu}$ as the generator of the SDE, applied to the tubular neighbourhood projection π:

$$\widehat{\mu}(y,t) = \frac{\partial \pi}{\partial x^k}\mu^k(t,y) + \frac{1}{2}\sum_{\gamma=1}^{n}\frac{\partial^2 \pi}{\partial x^i \partial x^j}\sigma^i_\gamma \sigma^j_\gamma(t,y) = (\mathscr{L}_t \pi)(y) \tag{1.98}$$

In [AB16] the field of Schwartz morphisms $\mathbb{F}$ is taken to be induced by a (time-homogeneous) field of maps f as in (1.25). In this approach we can use functoriality of $\mathbb{T}$ to write

$$\widehat{\mathbb{F}}(y) = \mathbb{T}_y \pi \circ \mathbb{F}(y) = \mathbb{T}_y \pi \circ \mathbb{T}_0 f_y = \mathbb{T}_0(\pi \circ f_y) \tag{1.99}$$

thus obtaining an SDE defined by the field of (2-jets of) maps given by projecting the original field of maps onto M with the tubular neighbourhood projection π.

Finally, we consider the Itô-vector projection of (1.67). By (1.79), in coordinates this amounts to projecting (1.67) to the Itô SDE on M with diffusion coefficients given by $\overline{\sigma}_\gamma$ and drift

$$\overrightarrow{\mu} = \underbrace{\overline{\mu}}_{\in TM} + \underbrace{\frac{1}{2}\sum_{\gamma=1}^{n}\frac{\partial^2 \pi}{\partial x^i \partial x^j}\overline{\sigma}^i_\gamma \overline{\sigma}^j_\gamma}_{\in T^\perp M} \tag{1.100}$$

To summarise, all three projections of the Itô equation (1.67) agree on how to map the diffusion coefficients, and the orthogonal components of the drift terms will all be fixed by the constraint (1.71), while their tangential projections are given by (respectively Stratonovich, Itô-jet, Itô-vector)

(1.101)

$P\widetilde{\mu}$	$\overline{\mu} + \frac{1}{2}\sum_{\gamma=1}^{n}\left(\frac{\partial^2 \pi}{\partial x^i \partial x^j}\overline{\sigma}^i_\gamma \breve{\sigma}^j_\gamma - \breve{\sigma}^h_\gamma P_k \frac{\partial \sigma^k_\gamma}{\partial x^h}\right)$
$P\widehat{\mu}$	$\overline{\mu} + \sum_{\gamma=1}^{n}\frac{\partial^2 \pi}{\partial x^i \partial x^j}\overline{\sigma}^i_\gamma \breve{\sigma}^j_\gamma$
$P\overrightarrow{\mu}$	$\overline{\mu}$

By calculations similar to (1.95) we can compute the projections of (1.65) in Stratonovich form: again, all three projections will orthogonally project the diffusion coefficients, and behave as follows on the Stratonovich drifts.

(1.102)

$\widetilde{b}$	$\overline{b}$
$\widehat{b}$	$\overline{b} + \frac{1}{2}\sum_{\gamma=1}^{n}\left(\breve{\sigma}^h_\gamma P_k \frac{\partial \sigma^k_\gamma}{\partial x^h} + \frac{\partial^2 \pi}{\partial x^i \partial x^j}\overline{\sigma}^i_\gamma \breve{\sigma}^j_\gamma\right)$
$\overrightarrow{b}$	$\overline{b} + \frac{1}{2}\sum_{\gamma=1}^{n}\left(\breve{\sigma}^h_\gamma P_k \frac{\partial \sigma^k_\gamma}{\partial x^h} - \frac{\partial^2 \pi}{\partial x^i \partial x^j}\overline{\sigma}^i_\gamma \breve{\sigma}^j_\gamma\right)$

From now on we will consider (1.67) as being our starting point, unless otherwise mentioned, and thus refer to (1.101) when comparing the three projections.

We end this section with a brief comparison of the three projections, leaving a detailed analysis of their differences to Subsection 1.5.1. The three projections coincide if $\sigma_\gamma \in TM$ for $\gamma = 1, \ldots, n$ (which includes the ODE case $\sigma_\gamma = 0$), in which case the diffusion coefficients remain unaffected, and the tangent component of the projected drift is simply given by $\overline{\mu}$. If $\sigma_\gamma \in T^\perp M$ for $\gamma = 1, \ldots, n$ all three projections result in an ODE on M, and the Itô-jet and Itô-vector projections coincide. Another case in which the Itô-jet and Itô-vector projections coincide is when the second derivatives of π vanish: this occurs in particular if M is embedded affinely, i.e. it coincides with some open set of an affine space of $\mathbb{R}^d$. All three projections forget the orthogonal part of the (Itô or Stratonovich) drift. We observe from (1.101) that the Itô-jet and Itô-vector projections of (1.67) only depend on the values of the Itô-coefficients on M. The Stratonovich projection, instead, could additionally depend on the tangential components of the derivatives of the diffusion coefficients in the direction of their normal components. Naturally, the situation is reversed when projecting (1.65): here it is the Stratonovich projection that only depends on the values of the coefficients on M, while the Itô-jet and -vector projections might depend on the mentioned derivative term.

Example 1.14 (The projections in the case M time-dependent)**.** Recalling Example 1.9 (and the map $\widetilde{\pi}$ defined therein) we may ask whether there is a way to consider the three SDE projections in the case of M time-dependent. The most natural way to define this is to consider, as done in (1.82), the joint equation satisfied by (t, X_t), project its coefficients in the three ways onto $\widetilde{M}$, thus obtaining a solution of the form (t, Y_t): this uses that $\widetilde{\pi}^0(t, y) = t$ (with time the 0^{th} coordinate), which is instead not necessarily satisfied by the Riemannian tubular neighbourhood projection onto $\widetilde{M}$. It is easily checked that the formulae (1.101) for the tangential component of the drift of Y_t continue to hold with the substitution of π_t for π (so that also the projection onto the tangent space P is now time-dependent), whereas in all three cases the orthogonal component of the drift picks up the term $\dot{\pi}_t$, needed to keep the process on the evolving manifold M_t. In particular, in the Itô-jet case we have

$$\widehat{\mu}(y, t) = (\mathscr{L}_t \pi_t)(y) + \dot{\pi}_t(y) = \widetilde{\mathscr{L}} \widetilde{\pi}(t, y) \tag{1.103}$$

where $\mathscr{L}_t$ is the generator of X and $\widetilde{\mathscr{L}}$ is that of (t, X_t) (which can be considered as being a time-homogeneous Markov process). This identity extends the observation made in Remark 1.13. The same term $\dot{\pi}_t$ should be added to the Stratonovich drifts (1.102) for the extension to the case of M time-dependent.

1.4 The optimal projection

In the previous section we showed how to abstractly project manifold-valued SDEs onto submanifolds in three (possibly) different ways, and specialised these constructions to the case of $M \hookrightarrow \mathbb{R}^d$-valued diffusions. In this section we will seek the *optimal* projection of an SDE for X_t, which we write in Itô form as (1.67). Let

$$\mathrm{d}Y_t^k = \mathring{\sigma}_\gamma^k(Y_t, t)\mathrm{d}W_t^\gamma + \mathring{\mu}^k(Y_t, t)\mathrm{d}t, \quad X_0 = y_0 \in M \tag{1.104}$$

be the M-valued SDE to be defined, which we write in $\mathbb{R}^d$-coordinates. Its coefficients $\mathring{\sigma}_\gamma$ and $\mathring{\mu}$ are to be treated as unknowns, to be determined by the optimisation criteria that involve the minimisation of the quantities

$$E[|Y_t - X_t|^2], \qquad E[|Y_t - \pi(X_t)|^2], \qquad |E[Y_t - X_t]|^2 \tag{1.105}$$

asymptotically for small t. Before we define the optimality criteria precisely, it is important to note that such expectations are undefined if the solution to either SDE is explosive, or, in the second case, even if it exits the tubular neighbourhood of M on which π is defined. The problem must be slightly changed so as to ensure that we are minimising a well-defined quantity. One option is to take the above expectations on the event $\{t \leq \tau_r\}$, where

$$\tau_r := \min\{t \geq 0 : |(X_t, Y_t) - (y_0, y_0)|^2 \geq r^2\} \tag{1.106}$$

for some suitable $r > 0$. However, since for such optimality criteria the values of the vector fields of both SDEs outside the ball $B_{(y_0,y_0)}(r) \subseteq \mathbb{R}^{2d}$ are irrelevant, it is simpler to just assume that they vanish outside $B_{(y_0,y_0)}(2r)$. Since the optimisation criteria will only determine the value of $\mathring{\sigma}$, $\mathring{\mu}$ at the initial condition, this is really only an assumption on σ and μ. The following proposition reassures us that, at least in well-behaved cases, this does not alter the problem in a way that interferes with the optimisation (which, as will be seen shortly, only involves the Taylor expansions of order 2 of (1.105) in $t = 0$).

Lemma 1.15. *Let X, Y, y_0, τ_r be as above, U a neighbourhood of (y_0, y_0) in $\mathbb{R}^{2d}$ and assume that there exists deterministic $\varepsilon > 0$ s.t. $X_t, Y_t \in U$ for $t \in [0, \varepsilon]$. Let $f : U \times [0, \varepsilon] \to \mathbb{R}$ be continuous s.t. $f(y_0, y_0, 0) = 0$, and assume moreover that $E[\max_{0 \leq t \leq \varepsilon} |f(X_t, Y_t, t)|] < \infty$ (this holds, in particular, under the global Lipschitz assumptions that guarantee SDE exactness [RW00, Theorem 11.2]). Then for any $r > 0$ with $\overline{B}_r(y_0, y_0) \subseteq U$*

$$E[f(X_t, Y_t, t)] - E[f(X_t, Y_t, t); t < \tau_r] \tag{1.107}$$

belongs to $O(t^n)$ for all $n \in \mathbb{N}$ as $t \to 0$.

Proof. Fix r, and let $\tau := \tau_r$. The Itô formula yields the a decomposition $|(X_t, Y_t) - (y_0, y_0)|^2 = L_t + A_t$ with L_t sum of Brownian integrals and A_t time integral, all of which for $t \leq \tau \wedge \varepsilon$ have bounded integrand (by continuity of the SDE coefficients and compactness of $\overline{B}_r(y_0, y_0) \times [0, \varepsilon]$). $[L]_t$ can be expressed as a time integral with bounded integrand: let $R > 0$ bound the sum of the absolute values of all integrands mentioned for $t \in [0, \tau \wedge \varepsilon]$. Then, still for $t \leq \tau \wedge \varepsilon$ we have $|A_t|, [L]_t \leq Rt$, and for any $\xi > 0$ it holds that $|(X_t, Y_t) - (y_0, y_0)|^2 \leq L_t + R\xi$ for $0 \leq t \leq \varepsilon \wedge \xi$. Letting $\xi := r^2/(3R)$, on $[0, \varepsilon \wedge \xi]$ we have

$$\begin{aligned} P[t \geq \tau] &= P\big[\max_{0 \leq s \leq t} |(X_s, Y_s) - (y_0, y_0)|^2 \geq r^2\big] \\ &= P\big[\max_{0 \leq s \leq \tau \wedge t} |(X_s, Y_s) - (y_0, y_0)|^2 \geq r^2\big] \\ &\leq P\big[\max_{0 \leq s \leq \tau \wedge t} L_s > r^2/2\big] \\ &= P\big[\max_{0 \leq s \leq t} L_{\tau \wedge s} > r^2/2\big] \\ &\leq \exp\left(-\frac{r^4}{4Rt}\right) \end{aligned} \tag{1.108}$$

by the tail estimate [RW00, Theorem 37.8 p.77]. Now, for $t \in [0, \varepsilon \wedge \xi]$ by Cauchy-Schwarz

$$\begin{aligned}
&\big|E[f(X_t, Y_t, t)] - E[f(X_t, Y_t, t); t < \tau]\big| \\
&= \big|E[f(X_t, Y_t, t); t \geq \tau]\big| \\
&\leq E[f(X_t, Y_t, t)^2]^{1/2} P[t \geq \tau]^{1/2} \\
&\leq E\Big[\max_{[0,t]} f(X_s, Y_s, s)^2\Big]^{1/2} P[t \geq \tau]^{1/2} \\
&\lesssim \exp\Big(-\frac{r^4}{4Rt}\Big)
\end{aligned} \tag{1.109}$$

since the first factor also vanishes as $t \to 0$, by the hypotheses on f, X, Y and dominated convergence. ■

We proceed with the constrained optimisation problem, assuming all SDE coefficients to be compactly supported; this means all local martingales involved will be martingales, and that we may use Fubini to pass to the expectation inside integrals in $\mathrm{d}t$. If we can write the Taylor expansion of the strong error

$$E\big[|Y_t - X_t|^2\big] = a_1 t + a_2 t^2 + o(t^2) \tag{1.110}$$

a first goal could be to minimise the leading coefficient a_1 (of course there is no constant term because $Y_0 = y_0 = X_0$). Using Itô's formula, and intending with $\simeq$ equality of differentials up to differentials of martingales, we have

$$\begin{aligned}
&\mathrm{d}|Y_t - X_t|^2 \\
&= \mathrm{d}\sum_{k=1}^d (Y_t^k - X_t^k)^2 \\
&= 2\sum_{k=1}^d \big((Y_t^k - X_t^k)\mathrm{d}Y_t^k - (Y_t^k - X_t^k)\mathrm{d}X_t^k\big) + \sum_{k=1}^d \big(\mathrm{d}Y_t^k\mathrm{d}Y_t^k + \mathrm{d}X_t^k\mathrm{d}X_t^k - 2\mathrm{d}X_t^k\mathrm{d}Y_t^k\big) \\
&\simeq \sum_{k=1}^d \Bigg[2\Bigg(\sum_{\gamma=1}^n \int_0^t \big(\mathring{\sigma}_\gamma^k(Y_s, s) - \sigma_\gamma^k(X_s, s)\big)\mathrm{d}W_s^\gamma \\
&\qquad + \int_0^t \big(\mathring{\mu}^k(Y_s, s) - \mu^k(X_s, s)\big)\mathrm{d}s\Bigg)\big(\mathring{\mu}^k(Y_t, t) - \mu^k(X_t, t)\big) \\
&\qquad + \sum_{\gamma=1}^n \big(\mathring{\sigma}_\gamma^k(Y_t, t)^2 + \sigma_\gamma^k(X_t, t)^2 - 2\sigma_\gamma^k(X_t, t)\mathring{\sigma}_\gamma^k(Y_t, t)\big)\Bigg]\mathrm{d}t
\end{aligned}$$

We now compute the expectation:

$$\begin{aligned}
&E\big[|Y_t - X_t|^2\big] \\
&= \sum_{k=1}^d 2E\Bigg[\int_0^t \Bigg(\sum_{\gamma=1}^n \int_0^s \big(\mathring{\sigma}_\gamma^k(Y_u, u) - \sigma_\gamma^k(X_u, u)\big)\mathrm{d}W_u^\gamma\Bigg)\big(\mathring{\mu}^k(Y_s, s) - \mu^k(X_s, s)\big)\mathrm{d}s\Bigg] \\
&\qquad + 2E\Bigg[\int_0^t \Bigg(\int_0^s \big(\mathring{\mu}^k(Y_u, u) - \mu^k(X_u, u)\big)\mathrm{d}u\Bigg)\big(\mathring{\mu}^k(Y_s, s) - \mu^k(X_s, s)\big)\mathrm{d}s\Bigg]
\end{aligned}$$

$$+E\Bigg[\sum_{\gamma=1}^{n}\int_0^t\big(\mathring{\sigma}_\gamma^k\mathring{\sigma}_\gamma^k(Y_s,s)+\sigma_\gamma^k\sigma_\gamma^k(X_s,s)-2\sigma_\gamma^k(X_s,s)\mathring{\sigma}_\gamma^k(Y_s,s)\big)\mathrm{d}s\Bigg]$$

$$\begin{aligned}
&=\int_0^t E\Bigg[\sum_{k=1}^{d}2\bigg(\sum_{\gamma=1}^{n}\int_0^s\big(\mathring{\sigma}_\gamma^k(Y_u,u)-\sigma_\gamma^k(X_u,u)\big)\mathrm{d}W_u^\gamma\bigg)\big(\mathring{\mu}^k(Y_s,s)-\mu^k(X_s,s)\big)\\
&\qquad\qquad +2\bigg(\int_0^s\big(\mathring{\mu}^k(Y_u,u)-\mu^k(X_u,u)\big)\mathrm{d}u\bigg)\big(\mathring{\mu}^k(Y_s,s)-\mu^k(X_s,s)\big)\\
&\qquad\qquad +\sum_{\gamma=1}^{n}\big(\mathring{\sigma}_\gamma^k(Y_s,s)-\sigma_\gamma^k(X_s,s)\big)^2\Bigg]\mathrm{d}s\\
&=:\int_0^t E[Z_s]\mathrm{d}s
\end{aligned}$$

and differentiating, with reference to (1.110) we have

$$a_1=\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_0^+\int_0^t E[Z_s]\mathrm{d}s=\sum_{\gamma=1}^{n}|\mathring{\sigma}_\gamma(y_0,0)-\sigma_\gamma(y_0,0)|^2 \tag{1.111}$$

Since a_1 only depends on the diffusion coefficients, its minimisation is expressed by the constrained optimisation problem whose solution is simply given by projecting the σ_γ's onto TM:

$$\begin{cases}\text{minimise }\displaystyle\sum_{\gamma=1}^{n}|\mathring{\sigma}_\gamma-\sigma_\gamma|^2\\ \text{subject to }Q_h^k\mathring{\sigma}_\gamma^h=0\end{cases}\iff\mathring{\sigma}=\overline{\sigma}=P\sigma \tag{1.112}$$

Here we have omitted evaluation at the initial condition $(0,y_0)$. Since we have not obtained a condition on $\mathring{\mu}$ our SDE (1.104) is still underdetermined, and the condition would be satisfied by the Stratonovich projection of (1.67).

One idea to obtain a condition on $\mathring{\mu}$ would be to minimise a_2 in (1.110). This attempt, however, has the drawback that we are minimising the second Taylor coefficient of a function without its first vanishing (unless the σ_γ's are already tangent to start with: in this case the minimisation of a_2 can be seen to result in the three projections, which all coincide). Although this approach is discussed in [AB16], we will not do so here, as there are more sound optimisation criteria. Indeed, we can look at the Taylor expansion of the weak error

$$|E[Y_t-X_t]|^2=b_2t^2+o(t^2)\quad\text{as }t\to 0^+ \tag{1.113}$$

We compute the term on the left as

$$|E[Y_t-X_t]|^2=\bigg|\int_0^t E[\mathring{\mu}(Y_s,s)-\mu(X_s,s)]\mathrm{d}s\bigg|^2 \tag{1.114}$$

and

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left|\int_0^t E[\mathring{\mu}(Y_s,s)-\mu(X_s,s)]\mathrm{d}s\right|^2 \\
&= 2E[\mathring{\mu}(Y_t,t)-\mu(X_t,t)]\int_0^t E[\mathring{\mu}(Y_s,s)-\mu(X_s,s)]\mathrm{d}s \\
&\frac{\mathrm{d}^2}{\mathrm{d}t^2}\bigg|_0\left|\int_0^t E[\mathring{\mu}(Y_s,s)-\mu(X_s,s)]\mathrm{d}s\right|^2 = 2|\mathring{\mu}(y_0,0)-\mu(y_0,0)|^2
\end{aligned}
\tag{1.115}
$$

which confirms that (1.113) lacks a linear term, and we have

$$
b_2 = |\mathring{\mu}-\mu|^2 \tag{1.116}
$$

Requiring the minimisation of b_2 is thus independent of the minimisation of a_1 above, and results in the constrained optimisation problem

$$
\begin{cases}
\text{minimise } |\mathring{\mu}-\mu|^2 \\
\text{subject to } Q^k_h\mathring{\mu}^h_\gamma = \dfrac{1}{2}\displaystyle\sum_{\gamma=1}^n \frac{\partial^2\pi^k}{\partial x^i\partial x^j}\overline{\sigma}^i_\gamma\overline{\sigma}^j_\gamma
\end{cases}
\iff \mathring{\mu} = \overline{\mu} + \frac{1}{2}\sum_{\gamma=1}^n \frac{\partial^2\pi}{\partial x^i\partial x^j}\overline{\sigma}^i_\gamma\overline{\sigma}^j_\gamma
\tag{1.117}
$$

A quick glance at (1.101) shows we have proven the following

Theorem 1.16 (Optimality of the Itô-vector projection). *The coefficients $\mathring{\sigma}_\gamma, \mathring{\mu}$ of the M-valued SDE (1.104) that solve the constrained optimisation problem*

$$
\begin{cases}
\text{minimise } a_1 \text{ in (1.110) and } b_2 \text{ in (1.113)} \\
\text{subject to (1.71)}
\end{cases}
\tag{1.118}
$$

for all initial conditions $X_0 = Y_0 = y_0 \in M$ are given (uniquely for $t=0$) by the Itô-vector projection of the original SDE (1.67).

Remark 1.17. In defining the three projections in Section 1.3 we intended for the projected coefficients to still be time-dependent if the original ones were. The optimality requirement only fixes the coefficients at the initial condition, at time 0, i.e. $\mathring{\sigma}_\gamma(y_0,0), \mathring{\mu}(y_0,0)$. To retain the time-dependence we may consider the optimisation involving all time-translated initial conditions $Y_{t_0} = y_0$.

Remark 1.18. Note that the form (Itô or Stratonovich) the initial SDE is provided in is irrelevant: if we had begun with (1.65) instead of (1.67) the optimality criterion would still have led us to the Itô-vector projection, which for the Stratonovich drift would have taken the form $\overrightarrow{b}$ in (1.102). The only reason to start with an Itô SDE is that the calculations are simpler, and it is possible to express the optimal coefficients as functions of the values of the coefficients of the original SDE, without reference to their derivatives.

The optimisation of Theorem 1.16 has the disadvantage of coming from the two separate minimisations of a_1 and b_2, which are Taylor coefficients of different quantities. There is a different way of arriving at coefficients by successively minimising the Taylor coefficients of the same quantity, with the first minimisation resulting in a null term. The idea is to consider

$$E\big[|Y_t-\pi(X_t)|^2\big]=c_1t+c_2t^2+o(t^2) \tag{1.119}$$

where X,Y,τ are respectively as in (1.67), (1.104),(1.106), with the requirement on r that $B_r(y_0)$ be contained in the domain of π. The map π is the one defined in (1.52), although it can more generally satisfy (1.86). Letting $\mathring{\sigma}_\gamma, \mathring{\mu}$ resume their status as unknowns, we proceed with the calculations.

$$\begin{aligned}
&\mathrm{d}|Y_t-\pi(X_t)|^2\\
&=\mathrm{d}\sum_{k=1}^{d}(Y_t^k-\pi^k(X_t))^2\\
&=\sum_{k=1}^{d}\Bigg[2(Y_t^k-\pi(X_t^k))\mathrm{d}Y_t^k-2(Y_t^k-\pi(X_t^k))\frac{\partial\pi^k}{\partial x^h}(X_t)\mathrm{d}X_t^h+\mathrm{d}Y_t^k\mathrm{d}Y_t^k\\
&\qquad+\bigg(\frac{\partial\pi^k}{\partial x^i}\frac{\partial\pi^k}{\partial x^j}(X_t)-(Y_t^k-\pi^k(X_t))\frac{\partial^2\pi^k}{\partial x^i\partial x^j}(X_t)\bigg)\mathrm{d}X_t^i\mathrm{d}X_t^j-2\frac{\partial\pi^k}{\partial x^h}(X_t)\mathrm{d}X_t^h\mathrm{d}Y_t^k\Bigg]\\
&\simeq\sum_{k=1}^{d}\Bigg[2\big(Y_t^k-\pi^k(X_t)\big)\bigg(\mathring{\mu}^k(t,Y_t)-\frac{\partial\pi^k}{\partial x^h}\mu^h(t,X_t)-\frac{1}{2}\sum_{\gamma=1}^{n}\frac{\partial^2\pi^k}{\partial x^i\partial x^j}\sigma_\gamma^i\sigma_\gamma^j(X_t,t)\bigg)\\
&\qquad+\sum_{\gamma=1}^{n}\bigg(\mathring{\sigma}_\gamma^k\mathring{\sigma}_\gamma^k(t,Y_t)+\frac{\partial\pi^k}{\partial x^i}\frac{\partial\pi^k}{\partial x^j}\sigma_\gamma^i\sigma_\gamma^j(t,X_t)-2\frac{\partial\pi^k}{\partial x^h}\sigma_\gamma^h(t,X_t)\mathring{\sigma}_\gamma^k(t,Y_t)\bigg)\Bigg]\mathrm{d}t\\
&=:Z_t\mathrm{d}t
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}E\bigg[\int_0^t Z_s\mathrm{d}s\bigg]\\
&=E\Bigg[\sum_{k=1}^{d}2\big(Y_t^k-\pi^k(X_t)\big)\bigg(\mathring{\mu}^k(t,Y_t)-\frac{\partial\pi^k}{\partial x^h}\mu^h(t,X_t)-\frac{1}{2}\sum_{\gamma=1}^{n}\frac{\partial^2\pi}{\partial x^i\partial x^j}\sigma_\gamma^i\sigma_\gamma^j(X_t,t)\bigg)\\
&\qquad+\sum_{\gamma=1}^{n}\bigg(\mathring{\sigma}_\gamma^k(Y_t,t)-\frac{\partial\pi^k}{\partial x^h}\sigma_\gamma^h(X_t,t)\bigg)^2\Bigg]
\end{aligned} \tag{1.120}$$

and therefore

$$c_1=\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_0^+E\bigg[\int_0^t Z_s\mathrm{d}s\bigg]=\sum_{\gamma=1}^{n}|\mathring{\sigma}_\gamma-P\sigma_\gamma|^2 \tag{1.121}$$

(evaluation at $(y_0, 0)$ is implied). Thus c_1 vanishes if and only if $\mathring{\sigma} := P\sigma$. Continuing as before and we have

$$
\begin{aligned}
\mathrm{d}Z_t \simeq & \sum_{k=1}^{d} 2\big(Y_t^k - \pi^k(X_t)\big)\mathrm{d}(\ldots) \\
& +2\left(\mathring{\mu}^k(t, Y_t) - \frac{\partial \pi^k}{\partial x^h}\mu^h(t, X_t) - \frac{1}{2}\frac{\partial^2 \pi^k}{\partial x^i \partial x^j}\sigma^i_\gamma \sigma^j_\gamma(X_t, t)\right)^2 \mathrm{d}t \\
& +2\sum_{\gamma=1}^{n}\left(\mathring{\sigma}^k_\gamma(Y_t, t) - \frac{\partial \pi^k}{\partial x^h}\sigma^h_\gamma(X_t, t)\right)\mathrm{d}(\ldots) \\
& +2f(\sigma, J\sigma, H\sigma; \mathring{\sigma}, J\mathring{\sigma}, H\mathring{\sigma}; \mu, J\mu)|_{X_t, Y_t, t}\mathrm{d}t
\end{aligned}
\tag{1.122}
$$

for some smooth function f (J denotes Jacobian and H Hessian), which we denote f_t for short; the differentials $\mathrm{d}(\ldots)$ can be ignored, since their factors will vanish when evaluated below.

$$
\begin{aligned}
c_2 &= \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\bigg|_0^+ E[Z_t] \\
&= \sum_{\gamma,k}\left(\mathring{\mu}^k - \frac{\partial \pi}{\partial x^h}\mu^h - \frac{1}{2}\frac{\partial^2 \pi}{\partial x^i \partial x^j}\sigma^i_\gamma \sigma^j_\gamma\right)^2 + f_t
\end{aligned}
\tag{1.123}
$$

The constrained optimisation problem for the minimisation of c_2 conditional on the previous minimisation of c_1 is thus given by

$$
\begin{aligned}
&\begin{cases}
\text{minimise } \displaystyle\sum_{k=1}^{d}\left(\mathring{\mu}^k - \frac{\partial \pi^k}{\partial x^h}\mu^h - \frac{1}{2}\sum_{\gamma=1}^{n}\frac{\partial^2 \pi^k}{\partial x^i \partial x^j}\sigma^i_\gamma \sigma^j_\gamma\right)^2 \\
\text{subject to } Q^k_h \mathring{\mu}^h - \displaystyle\frac{1}{2}\sum_{\gamma=1}^{n}\frac{\partial^2 \pi^k}{\partial x^i \partial x^j}\mathring{\sigma}^i_\gamma \mathring{\sigma}^j_\gamma = 0
\end{cases} \\
&\begin{cases}
2\left(\mathring{\mu}^h - \displaystyle\frac{\partial \pi^h}{\partial x^l}\mu^l - \frac{1}{2}\sum_{\gamma=1}^{n}\frac{\partial^2 \pi^h}{\partial x^i \partial x^j}\sigma^i_\gamma \sigma^j_\gamma\right) - \displaystyle\sum_{k=1}^{d} Q^k_h \lambda^k = 0 \\
Q^k_h \mathring{\mu}^h - \displaystyle\frac{1}{2}\sum_{\gamma=1}^{n}\frac{\partial^2 \pi^k}{\partial x^i \partial x^j}\mathring{\sigma}^i_\gamma \mathring{\sigma}^j_\gamma = 0
\end{cases} \\
&\lambda \in T_y M, \quad \mu = P\mathring{\mu} + \frac{1}{2}\sum_{\gamma=1}^{n}\frac{\partial^2 \pi}{\partial x^i \partial x^j}\sigma^i_\gamma \sigma^j_\gamma
\end{aligned}
\tag{1.124}
$$

Comparing with (1.98) we see that we have proven the following

Theorem 1.19 (Optimality of the Itô-jet projection)**.** *The coefficients $\mathring{\sigma}_\gamma$, $\mathring{\mu}$ of the M-valued SDE (1.104) that solve the constrained optimisation problem*

$$
\begin{cases}
\text{minimise } c_1 \text{ and } c_2\text{, conditionally on the minimisation of } c_1\text{, in (1.119)} \\
\text{subject to (1.71)}
\end{cases}
\tag{1.125}
$$

for all initial conditions $X_0 = Y_0 = y_0 \in M$ are given (uniquely for $t = 0$) by the Itô-jet projection of the original SDE (1.67).

Remarks analogous to Remark 1.17 and Remark 1.18 hold for Theorem 1.19. The Itô-vector and Itô-jet projection therefore satisfy different optimality properties, while the Stratonovich projection is suboptimal in both senses. We end the section with the extension of the optimisations to the case of M time-dependent.

Example 1.20 (Optimality for M time-dependent)**.** Recall the case in which the submanifold M depends smoothly on time, for which we can define similar versions of all three projections Example 1.14. For Theorem 1.16 the optimisation criterion does not require reformulation, while the constraint is modified as described in Example 1.9: therefore the Itô-vector projection remains optimal in the case of M time-dependent. For Theorem 1.16 the natural generalisation is given by substituting π_t for π in (1.119). Since $|y - \pi_t(x)| = |(t, y) - \widetilde{\pi}(t, x)|$, by the definition of the Itô-jet projection in the case of M time-dependent (and since the calculations in this section never relied on π being the Riemannian tubular neighbourhood projection), we have that the time-dependent Itô-jet projection (1.103) is optimal in this case too.

1.5 Further considerations

In this final section we dig deeper into the details surrounding the Itô and Stratonovich projections of SDEs, and answer a few lingering questions.

1.5.1 Differences between the projections

In this subsection we will provide examples to justify our claim that the vertical rectangles of (1.92) do not commute.

We begin with an example in which the Itô-jet and -vector projections coincide, but are different from the Stratonovich projection. This example also shows how the dependence of the Stratonovich projection of (1.67) on the derivatives of the diffusion coefficients can be non-trivial.

Example 1.21. Take $M = \{(x, 0) : x \in \mathbb{R}\} \hookrightarrow \mathbb{R}^2, n = 1$ and the Itô SDEs

$$\mathrm{d}\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} Y_t \\ X_t \end{bmatrix} \mathrm{d}W_t, \quad \mathrm{d}\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} 0 \\ X_t \end{bmatrix} \mathrm{d}W_t \tag{1.126}$$

whose diffusion coefficients coincide, and are orthogonal to M, on M. Their Stratonovich projections onto the affine subspace $M = \mathbb{R}$ are respectively given by the ODEs

$$\dot{X}_t = -\frac{1}{2} X_t, \quad \dot{X}_t = 0 \tag{1.127}$$

The Itô-jet and -vector projections of the two SDEs above coincide (since their coefficients on M coincide) and are trivial. An example where Itô-jet = Itô-vector $\neq$ Stratonovich, and where the Itô projections are non-trivial can be obtained from this by increasing n to 2 and adding a tangent diffusion coefficient.

Next, we ask the question of when the Stratonovich and Itô-jet projections coincide. The following criterion is a rephrasing of [AB16, Theorem 4].

Remark 1.22 (Fibering property). In general the difference of the Stratonovich- and Itô-jet-projected drift can be written as

$$TM \ni \widehat{\mu} - \widetilde{\mu} = \frac{1}{2}\sum_{\gamma=1}^{n}\left(\frac{\partial^2\pi}{\partial x^i \partial x^j}\overline{\sigma}^i_\gamma \breve{\sigma}^j_\gamma + \breve{\sigma}^k_\gamma \frac{\partial \pi}{\partial x^h}\frac{\partial \sigma^h_\gamma}{\partial x^k}\right) = \frac{1}{2}\sum_{\gamma=1}^{n}\frac{\partial}{\partial x^k}\left(\frac{\partial \pi}{\partial x^h}\sigma^h_\gamma\right)\breve{\sigma}^k_\gamma \tag{1.128}$$

Therefore, if we assume that

$$\frac{\partial \pi}{\partial x^h}(x)\sigma^h_\gamma(x,t) \quad \text{is independent of } x \in \pi^{-1}(\pi(x)) \tag{1.129}$$

for x in a neighbourhood of M (again, if we are only interested in starting our equation at time zero, the above requirement only needs to be considered for $t = 0$), the derivative of the above quantity along any vector tangent to the fibre of π (which at points in M means orthogonal to M) vanishes: this means (1.128) vanishes and the Itô jet and Stratonovich projections are equal. Moreover, if, representing the original SDE in Stratonovich form as (1.65), we additionally have that

$$\frac{\partial \pi}{\partial x^k}(x)b^k(x,t) \quad \text{is independent of } x \in \pi^{-1}(\pi(x)) \tag{1.130}$$

then it is immediate to verify that $\pi(X_t)$ is a solution of the Stratonovich projection, and therefore that, letting Y be the solution to the Stratonovich=Itô-jet projection

$$Y_t = \pi(X_t) \tag{1.131}$$

up to the exit time of X_t from the tubular neighbourhood in which π is defined. Observe that in the absence of these conditions we cannot expect, in general, to obtain a closed form SDE for $\pi(X_t)$, as the coefficients will depend explicitly on X_t. This is even true if (1.129) holds but (1.130) does not, as can be shown simply by considering the ODE case $\sigma_\gamma = 0$.

Example 1.23. Let $M = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\} \hookrightarrow \mathbb{R}^2$. π is defined in $\mathbb{R}^2 \setminus \{0\}$ as $\pi(x,y) = (x^2+y^2)^{-1/2}(x,y)$. Consider the SDE, dependent on the real parameter a

$$\mathrm{d}\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = (X_t^2 + Y_t^2)^a \begin{bmatrix} Y_t \\ X_t \end{bmatrix} \mathrm{d}W_t \tag{1.132}$$

There is a single diffusion coefficient σ, decomposed as

$$\sigma(x,y) = \underbrace{(x^2+y^2)^{a-1}(x^2-y^2)\begin{bmatrix} -y \\ x \end{bmatrix}}_{\overline{\sigma}(x,y)\in T_{(x,y)}M} + \underbrace{(x^2+y^2)^{a-1}2xy\begin{bmatrix} x \\ y \end{bmatrix}}_{\breve{\sigma}(x,y)\in T^\perp_{(x,y)}M} \tag{1.133}$$

Moreover, for $(x,y) \in M$ we have

$$J\sigma(x,y) = \begin{bmatrix} 2axy & 2ay^2+1 \\ 2ax^2+1 & 2axy \end{bmatrix} \tag{1.134}$$

We have

$$
\begin{aligned}
J\pi(x,y) &= (x^2+y^2)^{-3/2}\begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} \\
H\pi^1(x,y) &= (x^2+y^2)^{-5/2}\begin{bmatrix} -3xy^2 & 2x^2y-y^3 \\ 2x^2y-y^3 & 2xy^2-x^3 \end{bmatrix} \\
H\pi^2(x,y) &= (x^2+y^2)^{-5/2}\begin{bmatrix} 2x^2y-y^3 & 2xy^2-x^3 \\ 2xy^2-x^3 & -3x^2y \end{bmatrix}
\end{aligned}
\tag{1.135}
$$

We compute, for $(x,y)\in M$

$$
\begin{aligned}
\frac{\partial^2\pi}{\partial x^i\partial x^j}\overline{\sigma}^i\breve{\sigma}^j &= -2xy(x^2-y^2)\begin{bmatrix} -y \\ x \end{bmatrix} \\
\breve{\sigma}^h P_k\frac{\partial\sigma^k}{\partial x^h} &= 2xy(2ax^4-2ay^4+x^2-y^2)\begin{bmatrix} -y \\ x \end{bmatrix}
\end{aligned}
\tag{1.136}
$$

We examine more closely the cases $a=0$, $a=-1$ and $a=1$. In the first case (already examined in [AB16, §4]), the two terms of (1.136) sum to zero, so that (1.128) vanishes and the Stratonovich and Itô-jet projections coincide. Indeed, the fibering property of (1.129) is verified, as it is easy to see $J\pi(\lambda x,\lambda y)\sigma(\lambda x,\lambda y)$ does not depend on $\lambda>0$. Moreover, since the Stratonovich drift of the equation is given by $-\frac{1}{2}(x,y)$ also (1.130) holds and the solution to the Stratonovich=Itô-jet-projected SDE equals the projection of the solution of the original SDE up to the (a.s. infinite) time it hits the origin. However the Itô-vector projection is distinct, which can be seen by observing that $P\widetilde{\mu}=P\widehat{\mu}$ (given by the first term in (1.136)) does not vanish, e.g. at the point $(\cos(\pi/6),\sin(\pi/6))$. If $a=-1$ the two terms in (1.136) coincide on M and therefore the Stratonovich projection is identical to the Itô-vector projection. The Itô-jet projection, however, is different, again by the nonvanishing of the first term in (1.136) at $(\cos(\pi/6),\sin(\pi/6))$. To generate a case where all three projections are distinct take $a=1$: all identities can be seen not to hold at the point $(\cos(\pi/6),\sin(\pi/6))$. This case shows that the only projection that preserves the local martingale property is the Itô-vector.

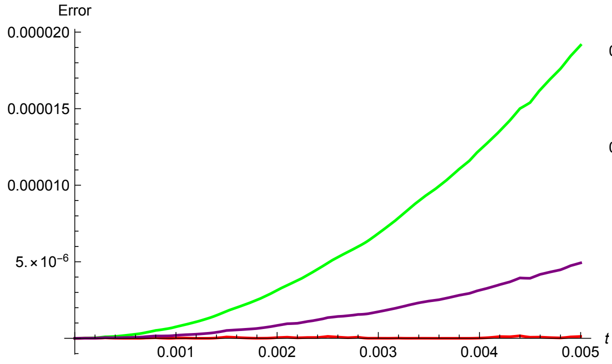
Example 1.24. Consider the case in which $\sigma_\gamma(x,t)=\sigma_\gamma(t)$ do not depend on the state of the solution. In this case, even if (1.67) and (1.65) are equivalent, the projections may still be all different. (1.101) however shows that

$$
\widehat{\mu}-\overrightarrow{\mu}=2(\widetilde{\mu}-\overrightarrow{\mu})
\tag{1.137}
$$

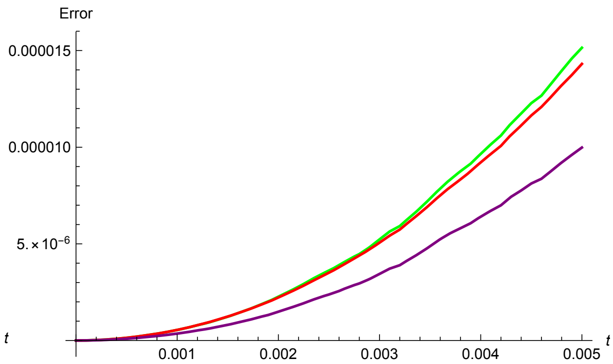
so that if any two projections coincide, they must all. An example where all projections are different is given by taking M, d as in Example 1.23 and the single, constant diffusion coefficient $\sigma=(1,1)$: all projections differ, for instance at the point $(1,0)$. An example where the projections all coincide is when $n=d$ and $\sigma^k_\gamma=\delta^k_\gamma$:

$$
\textstyle\sum_\gamma \overline{\sigma}^i_\gamma\breve{\sigma}^j_\gamma=\sum_\gamma P^i_\alpha\delta^\alpha_\gamma Q^j_\beta\delta^\beta_\gamma=\sum_\gamma P^i_\gamma Q^j_\gamma=P^i_\gamma Q^\gamma_j=0
\tag{1.138}
$$

If the original drift also vanishes, we are in the presence of the trivial SDE for Brownian motion, whose Itô and Stratonovich projections coincide with the process $\pi(W_t)$ up to the exit time of W from the domain of π, by the same reasoning of Remark 1.22.



Weak error (1.113)



Mean-squared error (1.119)

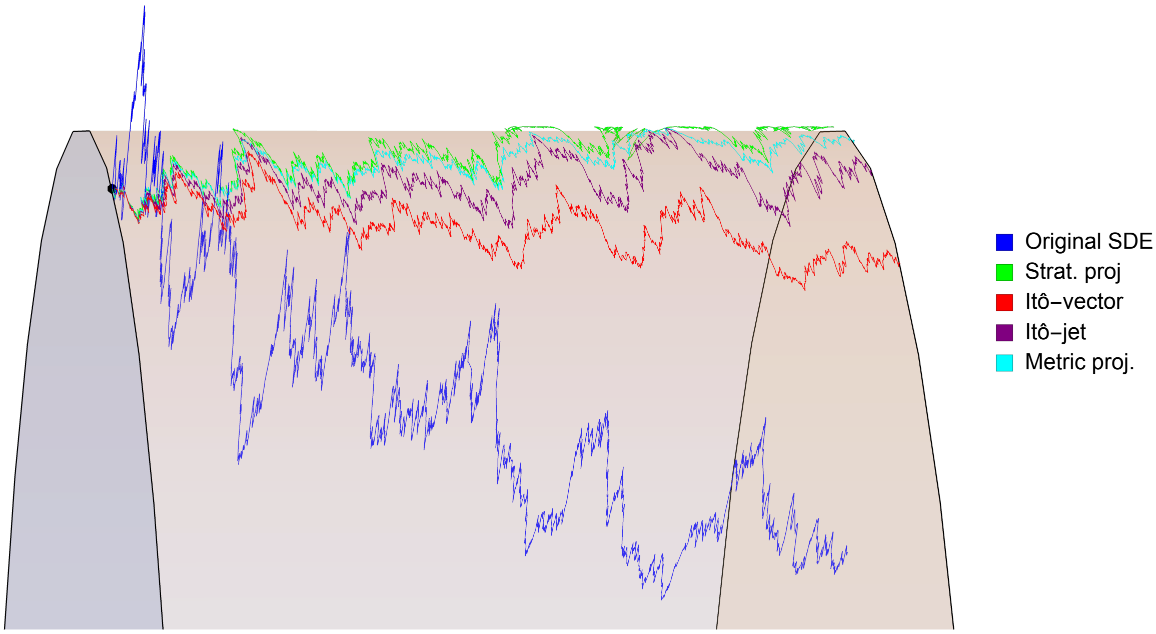


Figure 1.2: In these figures we focus on [Example 1.23, $a = 1$], with initial condition $(\cos(\pi/6), \sin(\pi/6))$, so that all projections are distinct. The two graphs above are respectively plots of the errors $|E[Y_t - X_t]|^2$ and $E[|Y_t - \pi(X_t)|^2]$ for Y_t the solution to the Stratonovich, Itô-vector and Itô-jet projections, with the expectation taken over 10^4 sample paths. We see confirmation of the fact that the Itô-vector projection performs better in the first error metric, that the Itô-jet projection does so in the second, and that the Stratonovich projection is markedly suboptimal in both senses (especially in the first, while in the second case it performs very similarly to the Itô-vector projection). The analogous plot for the error (1.110) is not included, as the results for the three projections are visually indistinguishable, in accordance with the fact that all three projections minimise a_1 (without it vanishing in this case). The figure below displays one sample path (t, Y_t) where Y_t is each of the following processes: the solution to the original SDE, to the three projected SDEs, and the metric projection π applied to the original solution. All sample paths are derived from the same random seed. Since the optimality criteria all involve taking expectation, we do not expect to be able to derive meaningful intuition from a single path, but it is nonetheless informative to have visual confirmation that all projections are distinct, but related.

In this section we have developed examples that cover all possible situations involving identities, and lack thereof, between the three projections. We summarise them in the table below:

$\widetilde{\mu} = \widehat{\mu} = \overrightarrow{\mu}$	$\sigma_\gamma \in TM$ and [Example 1.24, second case]
$\widetilde{\mu} \neq \widehat{\mu} = \overrightarrow{\mu}$	[Example 1.21, first SDE]
$\widetilde{\mu} = \widehat{\mu} \neq \overrightarrow{\mu}$	[Example 1.23, $a = 0$]
$\widetilde{\mu} = \overrightarrow{\mu} \neq \widehat{\mu}$	[Example 1.23, $a = -1$]
$\widetilde{\mu} \neq \widehat{\mu} \neq \overrightarrow{\mu} \neq \widetilde{\mu}$	[Example 1.23, $a = 1$] and [Example 1.24, first case]

1.5.2 Intrinsic optimality of the Itô projections

The fact that in (1.119) we are comparing two points, Y_t and $\pi(X_t)$, which lie in M opens up the possibility of substituting the Euclidean distance with the Riemannian distance of M, d_M, inside the expectation. One can then ask whether this leads to a different optimisation. Let U be a neighbourhood of the initial condition y_0 in $\mathbb{R}^d$, $V := U \cap M$, $\varphi \colon V \to \mathbb{R}^m$ a normal chart centred in y_0, $\overline{\varphi} := \varphi \circ \pi \colon U \to \mathbb{R}^m$. This means that if G_t is a geodesic in M starting at y_0, $\varphi(G_t) = vt$ where $\mathbb{R}^m \ni v = T_{y_0}\varphi(\dot{G}_0)$. As a consequence we have that, if $W_{y_0} \in T_{y_0}M$, picking the geodesic G with $G_0 = y_0$, $\dot{G}_{y_0} = W_{y_0}$, we have that

$$0 = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\bigg|_0 \varphi(G_t) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\bigg|_0 \overline{\varphi}(G_t) = \frac{\partial^2 \overline{\varphi}}{\partial x^i \partial x^j}(y_0)\dot{G}_0^i \dot{G}_0^j + \frac{\partial \overline{\varphi}}{\partial x^k}(y_0)\ddot{G}_0^k = \frac{\partial^2 \overline{\varphi}}{\partial x^i \partial x^j}(y_0) W_{y_0}^i W_{y_0}^j \quad (1.139)$$

since the acceleration of G is orthogonal to M. Now, the problem consists of choosing $\mathring{\sigma}_\gamma$ and $\mathring{\mu}$ in such a way that c_1' vanishes and c_2' is minimal in

$$E\big[{}^{\varphi}d_M\big(\varphi(Y_t), \varphi(\pi(X_t))\big)^2\big] = c_1' t + c_2' t^2 + o(t^2) \quad (1.140)$$

where ${}^{\varphi}d_M(a, b) := d_M(\varphi^{-1}(a), \varphi^{-1}(b))$. We have expressed d_M in normal coordinates in order to be able to use the estimates of [Nic12, Appendix A], which tell us that the derivatives of orders ≤ 3 of ${}^{\varphi}d_M$ agree with those of the squared distance function of $\mathbb{R}^m$ (in particular those of order 1 and 3 vanish). Since we are only interested in c_1' and c_2', this means we can substitute the LHS of (1.140) with

$$E\big[|\varphi(Y_t), \varphi(\pi(X_t))|^2\big] \quad (1.141)$$

Proceeding as in the computations of Section 1.4, we see that

$$c_1 = \sum_{\gamma=1}^{n} |J\varphi\mathring{\sigma}_\gamma - J\varphi P\sigma_\gamma|^2 \quad (1.142)$$

This quantity is made to vanish exactly as before, namely in the unique case $\mathring{\sigma}_\gamma = \overline{\sigma}_\gamma = \widehat{\sigma}_\gamma$. As for the drift, notice that since φ is a chart in M, minimising c_2 will only involve a condition on the tangential part of $\mathring{\mu}$, and is thus an unconstrained optimisation problem (the constraint (1.71) is then fulfilled by separately adding the required orthogonal term). Proceeding as in Section 1.4, we see that the quantity to be minimised is given by

$$\sum_{p=1}^{m} \left(\frac{\partial \overline{\varphi}^p}{\partial x^k}\mathring{\mu}^k - \frac{\partial \overline{\varphi}^p}{\partial x^k}\frac{\partial \pi^k}{\partial x^h}\mu^h - \frac{1}{2}\sum_{\gamma=1}^{n} \frac{\partial^2(\overline{\varphi}^p \circ \pi)}{\partial x^i \partial x^j}\sigma_\gamma^i \sigma_\gamma^j \right)^2 \quad (1.143)$$

which results in

$$\frac{\partial \overline{\varphi}}{\partial x^k}\mathring{\mu}^k = \frac{\partial \overline{\varphi}}{\partial x^k}\frac{\partial \pi^k}{\partial x^h}\mu^h + \frac{1}{2}\sum_{\gamma=1}^{n}\frac{\partial \overline{\varphi}}{\partial x^h}\frac{\partial^2 \pi^h}{\partial x^i \partial x^j}\sigma^i_\gamma \sigma^j_\gamma + \frac{\partial^2 \overline{\varphi}}{\partial x^i \partial x^j}\overline{\sigma}^i_\gamma \overline{\sigma}^j_\gamma \tag{1.144}$$

Since the last term vanishes by (1.139) we have that this formula for $\mathring{\mu}$ coincides with the Itô-jet projection $\widehat{\mu}$.

Remark 1.25 (Optimality for Riemannian ambient manifolds). This reformulation of the optimality criterion allows us to generalise the statement of Theorem 1.19 to the case where $\mathbb{R}^d$ is substituted with a general Riemannian manifold D, of which M is a Riemannian submanifold, (1.67) with a diffusion-type SDE on D (in any one of the three equivalent formulations), and the squared Euclidean norm in (1.119) is substituted with $d_D(Y_t, \pi(X_t))^2$. By considering a Nash embedding of D (and hence, transitively, of M) in $\mathbb{R}^r$ for large enough r, and extending the diffusion to a diffusion in $\mathbb{R}^r$, we have that the d_M-optimal projection and the $d_{\mathbb{R}^d}$-optimal projection both coincide with the Itô-jet projection. But since $d_{\mathbb{R}^d} \leq d_D \leq d_M$, this projection must also be d_D-optimal, as is immediate by comparing Taylor expansions.

We may also ask whether Theorem 1.16 admits a generalisation to the Riemannian case. This can be done by substituting the difference $Y_t - X_t$ with $\psi(Y_t) - \psi(X_t)$ in both (1.110) and (1.113), where ψ is any normal chart for the ambient Riemannian manifold D centred at the initial condition y_0, and the radius r appearing in (1.106) is chosen so that the ball of radius r centred in y_0 is contained in the domain of ψ. The proof of optimality is straightforward from Theorem 1.16 and the fact that $T_{y_0}\psi$ is a linear isometry, thus making the square

$$\begin{array}{ccc} T_{y_0}D & \xrightarrow{T_{y_0}\psi} & T_{y_0}\mathbb{R}^d \\ \big\downarrow {\scriptstyle P(y_0)} & & \big\downarrow {\scriptstyle P'(y_0)} \\ T_{y_0}M & \xrightarrow{T_{y_0}\psi|_M} & T_{y_0}M' \end{array} \tag{1.145}$$

(where $M' = \mathrm{Im}\psi$, and $P(y_0)$, $P'(y_0)$ are the metric projections) commute.

We have thus shown that both Theorem 1.16 and Theorem 1.19 can be reformulated so as to apply to the case of the ambient manifold being Riemannian.

1.5.3 Optimality criteria for the Stratonovich projection

It is surprising that the most naïve way to project the coefficients of an SDE is suboptimal according to the criteria introduced in this chapter. In this subsection we a (somewhat less compelling) way in which the Stratonovich projection can be considered optimal. This idea is already present in [AB16, §3.4].

As before, we start with the Stratonovich SDE (1.65). Define a second SDE

$$\mathrm{d}\Xi_t = -\sigma_\gamma(\Xi_t, t) \circ \mathrm{d}B^\gamma_t - b(\Xi_t, t)\mathrm{d}t, \quad \Xi_0 = y_0 \tag{1.146}$$

where B is another n-dimensional Brownian motion, with no specific relationship with W. Assume we are looking for coefficients $\mathring{\sigma}_\gamma$ and $\mathring{b}$ s.t., defining

$$\begin{aligned} \mathrm{d}Y_t &= \mathring{\sigma}_\gamma(Y_t, t) \circ \mathrm{d}W^\gamma_t + \mathring{b}(Y_t, t)\mathrm{d}t, \quad Y_0 = y_0 \\ \mathrm{d}\Upsilon_t &= -\mathring{\sigma}_\gamma(\Upsilon_t, t) \circ \mathrm{d}B^\gamma_t - \mathring{b}(\Upsilon_t, t)\mathrm{d}t, \quad \Upsilon_0 = y_0 \end{aligned} \tag{1.147}$$

the following quantity is minimised for small t (in the same sense as in Theorem 1.19):

$$\frac{1}{2}E\big[|Y_t - \pi(X_t)|^2\big] + \frac{1}{2}E\big[|\Upsilon_t - \pi(\Xi_t)|^2\big] \tag{1.148}$$

Note that the original input of the problem is the same as before, i.e. σ_γ and μ, but the quantity to be optimised is different. In [AB16] the SDEs with reflected Stratonovich coefficients are interpreted as representing a solution going backward in time: this fits in nicely with the interpretation of the Stratonovich integral of being time-symmetric (e.g. in the sense of the midpoint-evaluated Riemann sums that converge in L^2 to it). This interpretation is backed up by the fact that, if μ and $\mathring{\mu}$ denote the Itô drifts for X and Y, the SDE for Ξ, Υ can be equivalently written using the backwards Itô integral d^{b} (defined by endpoint evaluation)

$$\mathrm{d}^{\mathrm{b}}\Xi_t = -\sigma_\gamma(\Xi_t, t)\mathrm{d}^{\mathrm{b}}B^\gamma_t - \mu(\Xi_t, t)\mathrm{d}t, \qquad \mathrm{d}^{\mathrm{b}}\Upsilon_t = -\mathring{\sigma}_\gamma(\Upsilon_t, t)\mathrm{d}^{\mathrm{b}}B^\gamma_t - \mathring{\mu}(\Upsilon_t, t)\mathrm{d}t \tag{1.149}$$

so that (1.148) can be viewed as averaging an SDE going forward in time with one going backwards. This heuristic interpretation, however, is not necessary in the computations, and we can proceed by optimising (1.148) as is. Proceeding as above, this leads to the the diffusion coefficients being, as always, orthogonally projected ($\mathring{\sigma}_\gamma = \overline{\sigma}_\gamma$) and the constrained optimisation problem for the drift $\widetilde{\mu}$ given by

$$\begin{cases} \text{minimise} \quad \dfrac{1}{2}\displaystyle\sum_{k=1}^m \bigg(\widetilde{\mu}^k - \frac{\partial \pi^k}{\partial x^h}\mu^h - \frac{1}{2}\sum_{\gamma=1}^n \frac{\partial^2\pi^k}{\partial x^i \partial x^j}\sigma^i_\gamma\sigma^j_\gamma\bigg)^2 \\ +\dfrac{1}{2}\displaystyle\sum_{k=1}^m \bigg(-\widetilde{\mu}^k + \sum_{\gamma=1}^n \overline{\sigma}^l_\gamma \frac{\partial\overline{\sigma}^k_\gamma}{\partial x^l} - \frac{\partial \pi^k}{\partial x^h}\Big(-\mu^h + \sum_{\gamma=1}^n \sigma^l_\gamma\frac{\partial \sigma^h_\gamma}{\partial x^l}\Big) - \frac{1}{2}\sum_{\gamma=1}^n \frac{\partial^2\pi^k}{\partial x^i \partial x^j}\sigma^i_\gamma\sigma^j_\gamma\bigg)^2 \\ \text{subject to} \quad Q^k_h\widetilde{\mu}^h - \dfrac{1}{2}\displaystyle\sum_{\gamma=1}^n \frac{\partial^2\pi^k}{\partial x^i\partial x^j}\overline{\sigma}^i_\gamma\overline{\sigma}^j_\gamma = 0 \end{cases} \tag{1.150}$$

which is checked, by using Lagrange multipliers as above, to have solution the Stratonovich-projected drift (1.95). Therefore, the Stratonovich projection is optimal in this "time-symmetric" sense.

Conclusions and further directions

In this chapter we have shown, re-expressing and improving on the ideas of [AB16], how a concrete optimisation problem involving SDEs points towards the use of Itô calculus on manifolds, while the results given by adopting the more commonly used Stratonovich calculus are suboptimal. We believe that, apart from being of interest in their own right, these results offer some motivation for Chapter 2 and Chapter 4 below.

It would be interesting to extend this optimisation result to the case where the equation is driven by $(1/4, 1) \ni H$-fractional Brownian motion, in the sense of rough paths (which for $H > 1/2$ means in the sense of Young). Although this would amount to a generalisation of a Stratonovich equation, as seen in (1.102) the optimal coefficients can still be expressed as a function of the original ones and their derivatives, and similar formulae could be shown to hold in the case of fractional noise. The rough-Skorokhod conversion formula [CL19, CL20] could be of help here, although it would have to first be extended to cover the case in which the RDE has drift for the problem to be interesting. The considerations on Euler expansions discussed in Remark 5.5, once made precise, could also be of aid.

2

Non-geometric rough paths on manifolds

Project status. This chapter consists of a lightly edited version of the preprint [ABCRF20], co-authored with Damiano Brigo, Thomas Cass and John Armstrong. An abridged version of it has very recently been accepted for publication in the *Journal of the London Mathematical Society*.

Introduction

The theory of rough paths, first introduced in [Lyo98], has as its primary goal that of providing a rigorous mathematical framework for the study of differential equations driven by highly irregular inputs. The roughness of such signals renders the traditional definition of differentiation and integration inapplicable, and motivates the definition of *rough path*, a path X accompanied by functions, satisfying certain algebraic and analytic constraints, which postulate the values of its (otherwise undefined) iterated integrals. This concept leads to definitions of *rough integration* against the rough path $\boldsymbol{X}$ and of *rough differential equation* (*RDE*) driven by $\boldsymbol{X}$, which bear the important feature of being continuous in the signal $\boldsymbol{X}$, according to appropriately defined p-variation norms. Rough path theory applies to a wide variety of settings, including to the case in which X is given by the realisation of a stochastic process, for which it constitutes a pathwise approach to stochastic integration, extending the classical stochastic analysis of semimartingales.

An important feature that a rough path can satisfy is that of being *geometric*: this can be interpreted as the statement that it obeys the integration by parts and change of variable laws of first-order calculus, its irregularity notwithstanding. The theory of geometric rough paths has been the most studied [FV10b], and applies to semimartingales through the use of the Stratonovich integral. Other notions of stochastic integration, however, cannot be modelled by geometric rough paths, the Itô integral being the prime (but not the only [ER03]) example.

Since smooth manifolds are meant to provide a general setting for ordinary differential calculus to be carried out, it is natural to ask how "rougher" calculi can be defined in the curved setting. In the context of stochastic calculus, this question has led to a rich literature on Brownian motion on manifolds. More recently, it has been raised a number of times with regards to rough paths [CLL12, DS17, CDL15, BL15, Bai19]. In all cases, however, only the case of geometric rough paths has been discussed.

The main goal of this chapter is to construct a theory of manifold-valued rough paths of bounded p-variation, with $p < 3$, which are not assumed to be geometric. We will often refer to such rough paths as "non-geometric", as we have in the title of the chapter, even though what we actually mean is that they are not necessarily geometric. The regularity assumption ensures that we may draw on the familiar setting of [FH14] for vector space-valued rough paths; dropping this requirement would require the more complex algebraic tools of [HK15]. Our theory includes defining rough integration and differential equations, both from the intrinsic and extrinsic points of view, and showing how the classical notions of parallel transport and Cartan development can be extended to the case of non-geometric rough paths.

Although the definition of the Itô integral on manifolds has been known for decades, Stratonovich calculus has been preferred in the vast majority of the literature on stochastic differential geometry. Nevertheless, there are phenomena that are best captured by Itô calculus, particularly those which relate to the martingale property. In this spirit, Chapter 1 focuses on how a concrete problem involving the approximation of SDEs with ones defined on submanifolds necessitates the use of Itô notation, and that the result naturally provided by projecting the Stratonovich coefficients is suboptimal in general. The reason that Stratonovich integration and geometric rough paths are preferred in differential geometry is that they admit a simple coordinate-free description, as is also remarked on [Lyo98, p.219]. An important point, however, that we wish to make in this chapter is the following: an invariant theory of integration against non-geometric rough paths may also be given, albeit one that depends on the choice of a linear connection on the tangent bundle of the manifold. Although geometric rough path theory still retains the important property of being connection-invariant, all rough paths may be treated in a coordinate-free manner, since, while manifolds may not admit global coordinate systems, they always admit covariant derivatives. Overlooking this principle leads to the common misconception that Itô calculus/non-geometric rough integration cannot be carried out on manifolds, even in cases where a connection is already independently and canonically specified, e.g. when the manifold is Riemannian. In much of stochastic differential geometry the focus is not on the stochastic integral per se, which is viewed as a tool to investigate laws of processes defined on Riemannian Wiener space: in this context it is certainly justifiable to only work with the Stratonovich integral. Our emphasis here, however, is on pathwise integration itself: for this reason we believe it to be of value to build up the theory in a way that is faithful to the choice of the calculus, as specified through the rough path $\boldsymbol{X}$.

This chapter is organised as follows: in Section 2.1 we review the theory of vector space-valued rough paths of bounded $3 > p$-variation, controlled rough integrations and RDEs, relying (with a few modifications and additions) on [FH14].

In Section 2.2 we review the differential geometry which is necessary in the following sections.

In Section 2.3 we develop the theory at the heart of the chapter: this entails defining rough paths on manifolds and their controlled integrands in a coordinate-free manner by using pushforwards and pullbacks through charts, showing how the choice of a linear connection gives rise to a definition of rough integral, and defining RDEs in a similar spirit. We follow the "transfer principle" philosophy [É90] of replacing all instances of Euclidean spaces with smooth manifolds, which means that both the driving rough path and the solution are valued in (possibly different) manifolds. When we restrict our theory to semimartingales we recover the known framework for Itô integration and stochastic differential equations (SDEs) on manifolds [É89].

The next two sections contain the main contribution of this chapter. In Section 2.4 we switch from the local to the extrinsic framework, and show how our theory extends that of [CDL15] to non-geometric integrators and controlled integrands more general than one-forms. Our broader assumptions require us to make additional non-degeneracy requirements on the path X, which are not needed in the local setting. We also remark that in this section we are confining ourselves to the Riemannian case (with the metric being induced by an embedding), while in the rest of the chapter we allow for general connections.

Finally, in Section 2.5 we return to our local coordinate framework to carry out the constructions of parallel transport along rough paths and the resulting notion of Cartan development, or "rolling without slipping", a cornerstone of stochastic differential geometry which yields a convenient way of moving back and forth between the linear and curved setting. Since we are dealing with parallel transport as a TM-valued RDE driven by an M-valued rough path $\boldsymbol{X}$, the lack of geometricity leads us to require the choice of a connection not just on the tangent bundle of M but also of one on the tangent bundle of the manifold TM. The latter connection may not be chosen arbitrarily, and we identify criteria (formulated in terms of the former connection) that guarantee well-definedness, linearity, and, if M is Riemannian, isometricity of parallel transport. Different choices of such connection give rise to different definitions of parallel transport and Cartan development, which are only detectable at a second-order level, and all collapse to the same RDE when the rough path is geometric. Though we develop the theory in the most general way possible, three examples for how a connection on TM may be lifted to one on TTM are drawn from the literature; a case not analysed until now concerns the Levi-Civita connection of the Sasaki metric, which results in parallel transport coinciding with Stratonovich parallel transport. We end by seizing the opportunity to explore a few additional topics in stochastic analysis on manifolds, such as Cartan development in the presence of torsion, with a pathwise emphasis.

We hope that the framework laid out in this chapter may be used in the future to extend our understanding of manifold-valued rough paths, both deterministic and stochastic, and in Conclusions and further directions offer some ideas in this direction.

2.1 Background on rough paths

In this section we review the core theory of finite-dimensional vector space-valued (controlled) rough paths, and the corresponding notions of rough integrals and RDEs. We refer mainly to [FH14], with the caveat that we are in the setting of arbitrary control functions, as opposed to Hölder regularity. The former has has the advantage of being a parametrisation-invariant framework, and of allowing us to consider a larger class of paths (e.g. all semimartingales, and not just Brownian motion). Other authors have already been treating controlled rough paths in the setting of bounded p-variation [CL19, §2.4]. When a result in this first section is stated without proof, it is understood that the proof can be found in [FH14, Ch. 1-10]. Many of the more quantitative aspects of rough paths are left out, as they will not be relevant for the transposition of the theory to manifolds. Since our vector spaces are finite-dimensional, and since we will rely on arbitrary charts to make the manifold-valued theory coordinate-free, we will use fixed coordinates to express all of our formulae.

2.1.1 $\mathbb{R}^d$-valued rough paths

Throughout this chapter p will be a real number $\in [1,3)$; we will not exclude the case of $p \in [1,2)$ in which the theory reduces to Young integration, and remains valid with trivial adjustments. A *control* on $[0,T]$ is a continuous function ω defined on the subdiagonal $\Delta_T \coloneqq \{(s,t) \in [0,T]^2 \mid s \leq t\}$, s.t. $\omega(t,t) = 0$ for $0 \leq t \leq T$ and $\omega(s,u) + \omega(u,t) \leq \omega(s,t)$ for $0 \leq s \leq u \leq t \leq T$. ω will denote a control throughout this chapter, and should be thought of as being a fixed property of the (rough) path which relates to its parametrisation; the main example is the *Hölder control* $\omega(s,t) = t - s$. Given a path $X \colon [0,T] \to \mathbb{R}^d$ we will denote its increment $X_{st} \coloneqq X_t - X_s$. Let $\mathcal{C}^p_\omega([0,T],\mathbb{R}^d)$ denote the set of $\mathbb{R}^d$-valued continuous paths $X \colon [0,T] \to \mathbb{R}^d$ with

$$\sup_{0 \leq s < t \leq T} \frac{|X_{st}|}{\omega(s,t)^{1/p}} < \infty \tag{2.1}$$

For there to exist a control ω s.t. the above holds is equivalent to saying that X is a path of bounded p-variation [FV10b, Proposition 5.10]; if ω is the Hölder control we recover the definition of Hölder regularity. This kind of regularity is preserved by smooth maps [FV10b, §9.3].

Recall that if $X \in \mathcal{C}^p_\omega([0,T],\mathbb{R}^d)$, $H \in \mathcal{C}^q_\omega([0,T],\mathbb{R}^{e\times d})$ with $1/p + 1/q > 1$ (which happens, in particular, when $p = q \in [1,2)$) we may define the Young integral

$$\int_s^t H \mathrm{d}X \coloneqq \lim_{n\to\infty} \sum_{[u,v]\in\pi_n} H_u X_{uv} \tag{2.2}$$

where $(\pi_n)_n$ is a sequence of partitions on $[s,t]$ with vanishing step size; the resulting path $\int_0^\cdot H\mathrm{d}X$ belongs to $\mathcal{C}^p_\omega([0,T],\mathbb{R}^d)$. When the regularity requirement is no longer satisfied the Riemann sums no longer converge, and the definition of integral will require X and H to carry additional structure.

Definition 2.1 (Rough path)**.** A *p-rough path* controlled by ω on $[0,T]$, valued in $\mathbb{R}^d$ consists of a pair $\boldsymbol{X} = (X, \mathbb{X})$ with $X \in \mathcal{C}^p_\omega([0,T],\mathbb{R}^d)$ (the *trace*) and a continuous function $\mathbb{X} \colon \Delta_T \to (\mathbb{R}^d)^{\otimes 2} = \mathbb{R}^{d\times d}$ (the *second order part*) satisfying the regularity condition

$$\sup_{0 \leq s < t \leq T} \frac{|\mathbb{X}_{st}|}{\omega(s,t)^{2/p}} < \infty \tag{2.3}$$

with the property that the *Chen identity* holds: for all $0 \leq s \leq u \leq t \leq T$ and $\alpha, \beta = 1, \ldots, d$

$$\mathbb{X}^{\alpha\beta}_{st} = \mathbb{X}^{\alpha\beta}_{su} + X^\alpha_{su} X^\beta_{ut} + \mathbb{X}^{\alpha\beta}_{ut} \tag{2.4}$$

We denote the set of all such $\boldsymbol{X}$ as $\mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$ (note the difference in font with $\mathcal{C}$, used for simple paths). Its *bracket* path is given by

$$[\boldsymbol{X}]^{\alpha\beta}_{st} \coloneqq X^\alpha_{st} X^\beta_{st} - (\mathbb{X}^{\alpha\beta}_{st} + \mathbb{X}^{\beta\alpha}_{st}) \tag{2.5}$$

These are indeed the increments of an element of $\mathcal{C}^{p/2}_\omega([0,T], (\mathbb{R}^d)^{\odot 2})$-valued path, where $\odot$ denotes symmetric tensor product. We will say that $\boldsymbol{X}$ is *geometric* if $[\boldsymbol{X}] = 0$, and denote the set of these with $\mathscr{G}^p_\omega([0,T],\mathbb{R}^d)$.

The idea is that $\mathbb{X}_{st}$ represents the value of the (otherwise undefined) integral

$$\int_s^t \int_s^u \mathrm{d}X_r \otimes \mathrm{d}X_u = \int_s^t X_{su} \otimes \mathrm{d}X_u \tag{2.6}$$

In this interpretation it is easily checked that the Chen relation is simply the statement that the integral $\int X_u \otimes \mathrm{d}X_u$ is additive on consecutive time intervals, and the property of $\boldsymbol{X}$ of being geometric represents an integration by parts formula. Relaxing the Chen identity to

$$\mathbb{X}^{\alpha\beta}_{st} = \mathbb{X}^{\alpha\beta}_{su} + X^\alpha_{su} X^\beta_{ut} + \mathbb{X}^{\alpha\beta}_{ut} + \varepsilon^{\alpha\beta}_{st} \tag{2.7}$$

for some function of two parameters $\varepsilon_{st} \in o(\omega(s,t))$ as $t \searrow s$ for all s gives us the definition of *almost (geometric) rough path* and space of these denoted $\widetilde{\mathscr{C}}$ (and $\widetilde{\mathscr{G}}$ for almost geometric rough paths); this definition is motivated by the fact that the ε_{st}'s vanish in the limit of a sum over a sequence of partitions:

$$\sum_{[s,t]\in\pi} \varepsilon_{st} = \sum_{[s,t]\in\pi} \frac{\varepsilon_{st}}{\omega(s,t)} \omega(s,t) \leq \omega(0,T) \sup_{[s,t]\in\pi} \frac{\varepsilon_{st}}{\omega(s,t)} \xrightarrow{|\pi|\to 0} 0 \tag{2.8}$$

since $p < 3$ and $O(\omega(s,t)^{3/p}) \subseteq o(\omega(s,t))$. The same reasoning is also at the root of the following lemma [Lyo98, Theorem 3.3.1] [CDLL16, Proposition 3.5]. We write $\approx$ for equality up to an $\varepsilon_{st} \in o(\omega(s,t))$ as $t \searrow s$.

Lemma 2.2.

1. *If* $\boldsymbol{X}, \boldsymbol{Y} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$, $\boldsymbol{X} \approx \boldsymbol{Y} \Rightarrow \boldsymbol{X} = \boldsymbol{Y}$*;*
2. *Given* $\widetilde{\boldsymbol{X}} \in \widetilde{\mathscr{C}}^p_\omega([0,T],\mathbb{R}^d)$*, there exists a unique* $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$ *with* $\boldsymbol{X} \approx \widetilde{\boldsymbol{X}}$*, which is given by*

$$\boldsymbol{X}_{st} = \lim_{n\to\infty} \bigotimes_{[u,v]\in\pi_n} \widetilde{\boldsymbol{X}}_{uv} \tag{2.9}$$

where π_n *is any sequence of partitions of* $[s,t]$ *with vanishing step size. Moreover, if* $\widetilde{\boldsymbol{X}} \in \widetilde{\mathscr{G}}^p_\omega([0,T],\mathbb{R}^d)$*,* $\boldsymbol{X} \in \mathscr{G}^p_\omega([0,T],\mathbb{R}^d)$*.*

Both statements also hold when restricted to the level of paths $\in \mathcal{C}^p_\omega([0,T],\mathbb{R}^d)$*.*

If $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$ are related as in Lemma 2.2 we will say that latter is the rough path *associated* to the former.

Given $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$ we may associate a canonical element ${}_{\mathrm{g}}\boldsymbol{X} \in \mathscr{G}^p_\omega([0,T],\mathbb{R}^d)$, which we call its *geometrisation*, with trace equal to that of $\boldsymbol{X}$ and

$${}_{\mathrm{g}}\mathbb{X}^{\alpha\beta}_{st} := \tfrac{1}{2}(\mathbb{X}^{\alpha\beta}_{st} - \mathbb{X}^{\beta\alpha}_{st}) + \tfrac{1}{2} X^\alpha_{st} X^\beta_{st} \tag{2.10}$$

In other words, ${}_{\mathrm{g}}\mathbb{X}$ has the same antisymmetric part as $\mathbb{X}$ and symmetric part fixed by the trace and the geometricity condition, and it is easily checked that the Chen identity continues to hold.

2.1.2 Controlled paths and rough integration

We proceed to define the objects which are, in some sense, dual to rough paths, and are original to [Gub04]:

Definition 2.3 (Controlled path). Let $X \in \mathcal{C}^p_\omega([0,T],\mathbb{R}^d)$. An $\mathbb{R}^e$-valued, *X-controlled path*, or element of $\mathscr{D}^p_X([0,T],\mathbb{R}^e)$ is a pair $\boldsymbol{H} = (H, H')$, where $H \in \mathcal{C}^p_\omega([0,T],\mathbb{R}^e)$ (the *trace*), $H' \in \mathcal{C}^p_\omega([0,T],\mathbb{R}^{e\times d})$ (the *Gubinelli derivative* of H w.r.t. X), and

$$R^k_{st} \coloneqq H^k_{st} - H'^k_{\gamma;s} X^\gamma_{st}, \qquad \sup_{0\le s<t\le T} \frac{|R_{st}|}{\omega(s,t)^{2/p}} < \infty, \tag{2.11}$$

Here $\mathbb{R}^{e\times d}$ should be thought of as $\mathcal{L}(\mathbb{R}^d,\mathbb{R}^e)$ (where $\mathcal{L}$ means "linear maps"). We will identify $\mathbb{R}^n$-valued expressions with their coordinate expression throughout this chapter, e.g. we will write $X = (X^\gamma)$, $\boldsymbol{X} = (X^\gamma, \mathbb{X}^{\alpha\beta})$, $\boldsymbol{H} = (H^k, H'^k_\gamma)$. We will use $\approx_2$ as a shorthand for equality up to $O(\omega(s,t)^{2/p})$, i.e. (2.11) may be written as $H^k_{st} \approx_2 H'^k_\gamma X^\gamma_{st}$.

The following definition and theorem establishes that rough paths function as integrators, and that their controlled paths should be thought of as their admissible integrands.

Definition/Theorem 2.4 (Rough integral). Let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$ and $\boldsymbol{H} \in \mathscr{D}_X(\mathbb{R}^{e\times d})$. We then define, for $0 \le s \le t \le T$

$$\int_s^t \boldsymbol{H}\,\mathrm{d}\boldsymbol{X} \coloneqq \lim_{n\to\infty} \sum_{[u,v]\in\pi_n} H_{\gamma;u} X^\gamma_{uv} + H'_{\alpha\beta;u}\mathbb{X}^{\alpha\beta}_{uv} \tag{2.12}$$

where $(\pi_n)_n$ is a sequence of partitions on $[s,t]$ with vanishing step size. *This limit exists, is independent of such sequence and is obtained by applying Lemma 2.2 to*

$$H_{\gamma;s} X^\gamma_{st} + H'_{\alpha\beta;s}\mathbb{X}^{\alpha\beta}_{st} \tag{2.13}$$

Here H_t is an $\mathbb{R}^{e\times d}$-valued path and H'_t is a $\mathbb{R}^{e\times d\times d}$-valued path, with superscripts denoting $\mathbb{R}^e$-coordinates and subscripts denoting $\mathbb{R}^d$-coordinates; in $H'^k_{\alpha\beta}$ the coordinate of the Gubinelli derivative is α, i.e. the controlled path property now reads $H^k_{\beta;st} - H'^k_{\alpha\beta;s} X^\alpha_{st} \in O(\omega(s,t)^{2/p})$. We will often refer to controlled paths with trace valued in $\mathbb{R}^{e\times d}$ as *controlled integrands*. Clearly if $\widetilde{\boldsymbol{X}} \in \widetilde{\mathscr{C}}^p_\omega([0,T],\mathbb{R}^d)$ we may have substituted it for $\boldsymbol{X}$ in (2.12) and (2.13). We will often omit the integration extrema: in this case identities are to be intended to hold when the integral is taken on any interval. Also notice that it is obvious from the definition that the integral is linear in the integrand and additive on consecutive time intervals.

The condition of H admitting a Gubinelli derivative w.r.t. X is a strong condition, and one can only expect it to be satisfied when H bears a special relationship with X. One may also ask whether there are conditions on X under which any Gubinelli derivative H' is unique: this is not always true, since if X is too regular inside $\mathcal{C}^p_\omega([0,T],\mathbb{R}^d)$ the regularity requirement on H' becomes less stringent. A condition on X that rules this out, and guarantees uniqueness of the Gubinelli derivative is given by *true roughness* of X: this means that for all s in a dense set of $[0,T]$ and for all $\phi \in (\mathbb{R}^d)^*$

$$\limsup_{t\searrow s} \frac{|\langle\phi, X_{st}\rangle|}{\omega(s,t)^{2/p}} = \infty \tag{2.14}$$

It is satisfied, for instance, by a.a. sample paths of fractional Brownian motion with Hurst parameter $1/3 < H \le 1/2$, when considered as elements of $\mathcal{C}^p$, $1/H < p < 3$.

Theorem 2.5 (Uniqueness of the Gubinelli derivative)**.** *Let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$ with trace X truly rough, $(H, {}^1H'), (H, {}^2H') \in \mathscr{D}_X(\mathbb{R}^d)$. Then ${}^1H' = {}^2H'$.*

A corollary of this result is the uniqueness of the decomposition of the sum of a Young integral and a rough integral:

Theorem 2.6 (Doob-Meyer for rough paths)**.** *Let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$ with trace X truly rough, $Y \in \mathcal{C}^{p/2}([0,T],\mathbb{R}^d)$, ${}^1\boldsymbol{H}, {}^2\boldsymbol{H} \in \mathscr{D}_X(\mathbb{R}^{e\times d})$, ${}^1K, {}^2K \in \mathcal{C}^p([0,T],\mathbb{R}^{e\times d})$ then*

$$\int {}^1\boldsymbol{H}\mathrm{d}\boldsymbol{X} + \int {}^1K\mathrm{d}Y = \int {}^2\boldsymbol{H}\mathrm{d}\boldsymbol{X} + \int {}^2K\mathrm{d}Y \tag{2.15}$$

implies ${}^1\boldsymbol{H} = {}^2\boldsymbol{H}$ and ${}^1K = {}^2K$.

In most cases, as for Example 2.7 below, the Gubinelli derivative is defined in a canonical manner, and is intended to be computed accordingly, regardless of whether uniqueness holds or not.

Example 2.7 (Examples of canonically controlled paths)**.**

1. The simplest example of an X-controlled path is a smooth function $f \in C^\infty(\mathbb{R}^d,\mathbb{R}^e)$ applied to X: its Gubinelli derivative is given by $Df(X)$ (where $Df \in C^\infty(\mathbb{R}^d,\mathbb{R}^{e\times d})$ is the differential of f, with coordinates $\partial_\gamma f^k$) since

$$f^k(X)_{st} - \partial_\gamma f^k(X_s)X^\gamma_{st} \in O(|X_{st}|^2) \subseteq O(\omega(s,t)^{2/p}) \tag{2.16}$$

by Taylor's theorem. We call this X-controlled path $\boldsymbol{f}(X)$, but may omit the bold font if there is no ambiguity;

2. Let $\boldsymbol{X}, \boldsymbol{H}$ be as in Definition/Theorem 2.4, then the rough integral $\int_0^\cdot \boldsymbol{H}\mathrm{d}\boldsymbol{X}$ admits Gubinelli derivative H. We will sometimes denote the resulting element of $\mathscr{D}_X(\mathbb{R}^e)$ by $\boldsymbol{\int} \boldsymbol{H}\mathrm{d}\boldsymbol{X}$ (note the bold font used for the integral) if we want to emphasise that we are considering it a controlled path, but may omit the bold font if no ambiguity is possible;

3. Assume $\boldsymbol{H} \in \mathscr{D}_X(\mathbb{R}^e)$ and that $K \in \mathcal{C}^{p/2}_\omega([0,T],\mathbb{R}^e)$, then we may use H' as the Gubinelli derivative of $H + K$ and we have that $(H+K, H') \in \mathscr{D}_X(\mathbb{R}^e)$.

Example 2.8 (Difference of rough integrals against rough paths with common trace)**.**
Let ${}_1\boldsymbol{X} = (X, {}_1\mathbb{X}), {}_2\boldsymbol{X} = (X, {}_2\mathbb{X}) \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$, $\boldsymbol{H} \in \mathscr{D}^p_X(\mathbb{R}^{e\times d})$. Then it is easy to verify that there must exist a path $D \in \mathcal{C}^{p/2}_\omega(\mathbb{R}^{d\times d})$ s.t. ${}_2\mathbb{X}_{st} = {}_1\mathbb{X}_{st} + D_{st}$, and it is easily deduced from the (2.13)

$$\int \boldsymbol{H}\mathrm{d}_2\boldsymbol{X} = \int \boldsymbol{H}\mathrm{d}_1\boldsymbol{X} + \int H'\mathrm{d}D \tag{2.17}$$

where the second integral on the right is intended in the sense of Young. An important special case is when ${}_2\boldsymbol{X} = {}_{\mathrm{g}}\boldsymbol{X}$ for a rough path $\boldsymbol{X}$, in which case $D = \frac{1}{2}[\boldsymbol{X}]$. Note that this identity also holds at the level of controlled paths, since the Gubinelli derivatives (taken according to Example 2.7) both coincide with H. We will often use the notation

$$\circ\,\mathrm{d}\boldsymbol{X} \coloneqq \mathrm{d}_{\mathrm{g}}\boldsymbol{X} \tag{2.18}$$

which is motivated by Stratonovich calculus (see Remark 2.24 below).

A controlled path may be transformed into a rough path in a canonical fashion:

Definition 2.9 (Lift of a controlled path)**.** Let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$, $\boldsymbol{H} \in \mathscr{D}_X(\mathbb{R}^e)$. Define $\uparrow_{\mathbb{X}}\boldsymbol{H}$ to be the rough path associated to $\upharpoonleft_{\mathbb{X}}\boldsymbol{H}$, defined as

$$(\upharpoonleft_{\mathbb{X}}\boldsymbol{H})_{st} \coloneqq (H^k_{st}, H'^i_{\alpha;s}H'^j_{\beta;s}\mathbb{X}^{\alpha\beta}_{st}) \tag{2.19}$$

which is easily verified to belong to $\widetilde{\mathscr{C}}^p_\omega([0,T],\mathbb{R}^e)$ (and $\widetilde{\mathscr{G}}^p_\omega([0,T],\mathbb{R}^e)$ if $\boldsymbol{X}$ is geometric).

We would also like to show that the operation of lifting restricts to geometric rough paths: this is accomplished in the following

Lemma 2.10. *If* $\boldsymbol{X} \in \mathscr{G}^p([0,T],\mathbb{R}^d)$ *then* $\uparrow_{\mathbb{X}}\boldsymbol{H} \in \mathscr{G}^p([0,T],\mathbb{R}^e)$.

Proof. We cannot apply Lemma 2.2 directly to $\upharpoonleft_{\mathbb{X}}\boldsymbol{H}$, since it only satisfies $[\upharpoonleft_{\mathbb{X}}\boldsymbol{H}] \approx 0$:

$$\begin{aligned}
(\widetilde{\upharpoonleft}_{\mathbb{X}}\boldsymbol{H})^{ij}_{st} + (\widetilde{\upharpoonleft}_{\mathbb{X}}\boldsymbol{H})^{ij}_{st} &= H'^i_{\alpha;s}H'^j_{\beta;s}(\mathbb{X}^{\alpha\beta}_{st} + \mathbb{X}^{\beta\alpha}_{st}) \\
&= H'^i_{\alpha;s}X^\alpha_{st}H'^j_{\beta;s}X^\beta_{st} \\
&= (H^i_{st} - R^i_{st})(H^j_{st} - R^j_{st}) \\
&= H^i_{st}H^j_{st} - H^i_{st}R^j_{st} - H^j_{st}R^i_{st} + R^i_{st}R^j_{st}
\end{aligned}$$

where R is as in (2.11). We may therefore define $\widetilde{\upharpoonleft}_{\mathbb{X}}\boldsymbol{H}$ by $(\widetilde{\upharpoonleft}_{\mathbb{X}}\boldsymbol{H})^k \coloneqq (\upharpoonleft_{\mathbb{X}}\boldsymbol{H})^k$ and

$$(\widetilde{\upharpoonleft}_{\mathbb{X}}\boldsymbol{H})^{ij}_{st} \coloneqq (\upharpoonleft_{\mathbb{X}}\boldsymbol{H})^{ij}_{st} + H^i_{st}R^j_{st} - \tfrac{1}{2}R^i_{st}R^j_{st} \tag{2.20}$$

Then $(\widetilde{\upharpoonleft}_{\mathbb{X}}\boldsymbol{H})_{st} \approx (\upharpoonleft_{\mathbb{X}}\boldsymbol{H})_{st}$ (and is therefore still almost multiplicative), and $[\upharpoonleft_{\mathbb{X}}\boldsymbol{H}] = 0$ from which we conclude by the aforementioned lemma. ■

Example 2.11 (Lifts of controlled paths)**.**

1. Given a $f \in C^\infty(\mathbb{R}^d,\mathbb{R}^e)$ we define $f_*\boldsymbol{X} \coloneqq \uparrow_{\mathbb{X}}\boldsymbol{f}(X)$ the *pushforward* of $\boldsymbol{X}$ through f, and by Taylor's formula we have

$$(f_*\boldsymbol{X})_{st} \approx (\partial_\gamma f^k(X_s)X^\gamma_{st} + \tfrac{1}{2}\partial_{\alpha\beta}f^k(X_s)X^\alpha_{st}X^\beta_{st}, \partial_\alpha f^i\partial_\beta f^j(X_s)\mathbb{X}^{\alpha\beta}_{st}) \tag{2.21}$$

2. Rough integrals may be lifted to rough paths: if $\boldsymbol{H}$ is as in Definition/Theorem 2.4 we set $\int_s^t \boldsymbol{H}\mathrm{d}\boldsymbol{X} \coloneqq (\uparrow_{\mathbb{X}}\int \boldsymbol{H}\mathrm{d}\boldsymbol{X})_{st}$ and we have

$$\int_s^t \boldsymbol{H}\mathrm{d}\boldsymbol{X} \approx (H^k_{\gamma;s}X^\gamma_{st} + H'^k_{\alpha\beta;s}\mathbb{X}^{\alpha\beta}_{st}, H^i_{\alpha;s}H^j_{\beta;s}\mathbb{X}^{\alpha\beta}_{st}) \tag{2.22}$$

   Note, however, that (2.17) does not hold at the rough path level, since the lift on the LHS would be computed using ${}_2\mathbb{X}$, and the one on the RHS using ${}_1\mathbb{X}$, and this would affect the second order part of the rough integrals. For this reason we will mostly consider Itô-Stratonovich type corrections only at the trace level.

Whenever there is an ambiguity as to whether a function on Δ_T is a controlled or rough path we will rely on coordinate notation to distinguish these two possibilities, e.g. $(\int \boldsymbol{H} \mathrm{d}\boldsymbol{X})^k$ are the coordinates of the trace of the controlled/rough path, $(\int \boldsymbol{H} \mathrm{d}\boldsymbol{X})^k_\gamma = H^k_\gamma$ are the coordinates of the Gubinelli derivative of the controlled path (corresponding to the trace of H) and $(\int_s^t \boldsymbol{H} \mathrm{d}\boldsymbol{X})^{ij} \approx H^i_{\alpha;s} H^j_{\beta;s} \mathbb{X}^{\alpha\beta}_{st}$ those of the second order part of the rough path. We will often use coordinate notation inside the integral too, to track the action of the integrand on the integrator, e.g. $(\int \boldsymbol{H} \mathrm{d}\boldsymbol{X})^k =: \int \boldsymbol{H}^k_\gamma \mathrm{d}\boldsymbol{X}^\gamma$, with the understanding that we also need the second-order coordinates of $\boldsymbol{X}$ and $\boldsymbol{H}$ to compute this integral.

Proposition 2.12 (Operations on controlled paths)**.** *Let* $X \in \mathcal{C}^p_\omega([0,T],\mathbb{R}^d)$.

Change of controlling path. *Let* $\boldsymbol{H} \in \mathscr{D}_X(\mathbb{R}^e)$, $\boldsymbol{K} \in \mathscr{D}_H(\mathbb{R}^f)$, *then*

$$\boldsymbol{K} * H' \coloneqq (K^c, K'^c_k H'^k_\gamma) \in \mathscr{D}_X(\mathbb{R}^f) \tag{2.23}$$

In particular, if $\boldsymbol{K} = \boldsymbol{f}(H)$ *for* $f \in C^\infty(\mathbb{R}^e,\mathbb{R}^f)$ *we denote this* $f_*\boldsymbol{H}$ *and call it the pushforward of* $\boldsymbol{H}$ *through* f*;*

Leibniz rule. *Let* $\boldsymbol{H} \in \mathscr{D}_X(\mathbb{R}^{f\times e})$ *and* $\boldsymbol{K} \in \mathscr{D}_X(\mathbb{R}^{g\times f})$*, then*

$$\boldsymbol{K} \cdot \boldsymbol{H} \coloneqq (K^r_c H^c_k, K'^r_{\gamma c} H^c_k + K^r_c H'^c_{\gamma k}) \in \mathscr{D}_X(\mathbb{R}^{g\times e}) \tag{2.24}$$

Pullback. *Let* $g \in C^\infty(\mathbb{R}^d,\mathbb{R}^e)$, $\boldsymbol{H} \in \mathscr{D}_{g(X)}(\mathbb{R}^{f\times e})$*, then*

$$\begin{aligned} g^*\boldsymbol{H} &\coloneqq (\boldsymbol{H} * \underbrace{Dg(X)}_{=g(X)'}) \cdot \boldsymbol{Dg}(X) \in \mathscr{D}_X(\mathbb{R}^{f\times d}) \\ &= (H^c_k \partial_\gamma g^k(X), H'^c_{ij}\partial_\alpha g^i \partial_\beta g^j(X) + H^c_k \partial_{\alpha\beta} g^k(X)) \end{aligned} \tag{2.25}$$

Proof. Clearly all three paths belong to $\mathcal{C}^p_\omega$. We need to check that (2.11) holds in all three cases. In the case of the change of controlling path we have

$$K^c_{st} - K'^c_{k;s} H'^k_{\gamma;s} X^\gamma_{st} \approx_2 K^c_{st} - K'^c_{k;s} H^k_{st} \approx_2 0 \tag{2.26}$$

As for the Leibniz rule, consider the matrix multiplication function

$$m \colon \mathbb{R}^{g\times f} \times \mathbb{R}^{f\times e} \to \mathbb{R}^{g\times e}, \quad (z^r_c, y^c_k) \mapsto (z^r_c y^c_k) \tag{2.27}$$

It is easily verified that $\boldsymbol{K} \cdot \boldsymbol{H} = m_*(\boldsymbol{H},\boldsymbol{K})$, the pushforward of controlled paths being defined in the step above.

The case of the pullback readily follows from its expression as a combination of the two above constructions. ■

Proposition 2.13 (Compatibility)**.** *Let* $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$ *and* $\boldsymbol{H} \in \mathscr{D}_X(\mathbb{R}^e)$*:*

1. *Lifting is compatible with change of controlling path in the sense that, for* $\boldsymbol{K} \in \mathscr{D}_H(\mathbb{R}^f)$ *we have*

$$\uparrow_{\mathbb{H}} \boldsymbol{K} = \uparrow_{\mathbb{X}} (\boldsymbol{K} * H') \tag{2.28}$$

where $\mathbb{H}$ denotes the second order part of the rough path $\uparrow_{\mathbb{X}} \boldsymbol{H}$. In particular, for $f \in C^\infty(\mathbb{R}^e, \mathbb{R}^f)$ pushforward of rough and controlled paths are related through lift by $f_* \uparrow_{\mathbb{X}} \boldsymbol{H} = \uparrow_{\mathbb{X}} f_* \boldsymbol{H}$. Moreover, $f_*(g_* \boldsymbol{X}) = (f \circ g)_* \boldsymbol{X}$ for appropriately valued smooth maps f, g;

2. *Lifting is compatible with geometrisation in the sense that*

$$_{\mathrm{g}}(\uparrow_{\mathbb{X}} \boldsymbol{H}) = \uparrow_{{}_{\mathrm{g}}\mathbb{X}} \boldsymbol{H} \tag{2.29}$$

   In particular, pushforward of rough paths and rough integration preserve geometricity;

3. *For appropriately value smooth maps f, g and controlled integrands $\boldsymbol{K}$ we have*

$$(f \circ g)^* \boldsymbol{K} = g^*(f^* \boldsymbol{K}) \tag{2.30}$$

Proof. As for the first claim, the two rough paths agree on the trace K and second order part

$$(\uparrow_{\mathbb{H}} \boldsymbol{K})^{ab}_{st} \approx K'^a_{i;s} K'^b_{j;s} \mathbb{H}^{ij}_{st} \approx K'^a_{i;s} K'^b_{j;s} H'^i_{\alpha;s} H'^j_{\beta;s} \mathbb{X}^{\alpha\beta}_{st} = (\uparrow_{\mathbb{X}} (\boldsymbol{K} * H'))^{ab} \tag{2.31}$$

Identity of the two rough paths therefore holds by Lemma 2.2. Now, taking $\boldsymbol{K} = \boldsymbol{f}(H)$ and the definitions of pushforward this yields

$$f_* \uparrow_{\mathbb{X}} \boldsymbol{H} = \uparrow_{\mathbb{H}} (\boldsymbol{f}(H)) = \uparrow_{\mathbb{X}} f_* \boldsymbol{H} \tag{2.32}$$

Taking, furthermore, $\boldsymbol{H} = \boldsymbol{g}(X)$ we obtain

$$f_*(g_* \boldsymbol{X}) = f_* \uparrow_{\mathbb{X}} \boldsymbol{g}(X) = \uparrow_{\mathbb{X}} (f_* \boldsymbol{g}(X)) = \uparrow_{\mathbb{X}} (\boldsymbol{f} \circ \boldsymbol{g}(X)) = (f \circ g)_* \boldsymbol{X} \tag{2.33}$$

As for the second claim, the two rough paths have the same trace, and therefore the same symmetric part of the second order part, and antisymmetric part equal to half of

$$\begin{aligned} {}_{\mathrm{g}}(\uparrow_{\mathbb{X}} \boldsymbol{H})^{ij} - {}_{\mathrm{g}}(\uparrow_{\mathbb{X}} \boldsymbol{H})^{ji} &= (\uparrow_{\mathbb{X}} \boldsymbol{H})^{ij} - (\uparrow_{\mathbb{X}} \boldsymbol{H})^{ji} \\ &\approx H'^i_{\alpha;s} H'^j_{\beta;s} (\mathbb{X}^{\alpha\beta}_{st} - \mathbb{X}^{\beta\alpha}_{st}) \\ &\approx (\uparrow_{{}_{\mathrm{g}}\mathbb{X}} \boldsymbol{H})^{ij} - (\uparrow_{{}_{\mathrm{g}}\mathbb{X}} \boldsymbol{H})^{ji} \end{aligned} \tag{2.34}$$

Therefore ${}_{\mathrm{g}}(\uparrow_{\mathbb{X}} \boldsymbol{H}) \approx \uparrow_{{}_{\mathrm{g}}\mathbb{X}} \boldsymbol{H}$ and we conclude again by Lemma 2.2.

The final statement is verified using a similar comparison of the expressions in coordinates. ■

The following is a rough path version of the Itô lemma. Note how the formula simplifies to a first order chain rule in the case of $\boldsymbol{X}$ geometric. It is followed by the rough path-version of the Kunita-Watanabe identity, where the bracket path plays the role of quadratic covariation matrix.

Theorem 2.14 (Itô lemma for rough paths)**.** *Let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T], \mathbb{R}^d)$ and $f \in C^\infty(\mathbb{R}^d, \mathbb{R}^e)$. Then*

$$f(X) = f(X_0) + \int_0^\cdot Df(X) \mathrm{d}\boldsymbol{X} + \frac{1}{2} \int_0^\cdot D^2 f(X) \mathrm{d}[\boldsymbol{X}] \tag{2.35}$$

Moreover, the Gubinelli derivatives of the LHS and RHS, computed canonically according to Example 2.7 agree, *thus giving rise to an identity in $\mathscr{D}_X(\mathbb{R}^e)$, and after applying $\uparrow_{\mathbb{X}}$, to one in $\mathscr{C}^p_\omega([0,T],\mathbb{R}^e)$ (with the term $f(X_0)$ only influencing the trace).*

Proof. The path-level statement is proved in [FH14, Proposition 5.6]. The Gubinelli derivative of the LHS according to [Example 2.7, 1.] is $Df(X)$, which coincides with the Gubinelli derivative of the LHS according to [Example 2.7, 2.,3.] (since the bracket path, and thus the Young integral has higher regularity). ■

Note how the integral of the exact one-form $Df(X)$ does not require the whole of $\boldsymbol{X}$: this is because its Gubinelli derivative, $D^2f(X)$, is symmetric. Only the symmetric part of $\mathbb{X}$ is needed: the pair $(X, \odot\mathbb{X})$ (with $\odot$ denoting the symmetrisation operator) is called a *reduced rough path.*

Proposition 2.15 (Kunita-Watanabe identity for rough paths)**.** *Let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$, $\boldsymbol{H} \in \mathscr{D}_X(\mathbb{R}^e)$. Then*

$$[\uparrow_{\mathbb{X}}\boldsymbol{H}]^{ij}_{st} \approx H^i_{st}H^j_{st} - H'^i_{\alpha;s}H'^j_{\beta;s}(\mathbb{X}^{\alpha\beta}_{st} + \mathbb{X}^{\beta\alpha}_{st}) \tag{2.36}$$

so in particular (if $e = f \times d$ in the second case below)

$$[f_*\boldsymbol{X}]^{ij}_{st} = \int_s^t \partial_\alpha f^i(X)\partial_\beta f^j(X)\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}, \quad \left[\int \boldsymbol{H}\mathrm{d}\boldsymbol{X}\right]^{ij}_{st} = \int_s^t H^i_\alpha H^j_\beta \mathrm{d}[\boldsymbol{X}]^{\alpha\beta} \tag{2.37}$$

Proof. The first claim is immediate from (2.5), Example 2.11. The bracket of a pushforward is computed as

$$\begin{aligned}
[f_*\boldsymbol{X}]^{ij}_{st} &\approx f^i(X)_{st}f^j(X)_{st} - \partial_\alpha f^i(X_s)\partial_\beta f^j(X_s)(\mathbb{X}^{\alpha\beta}_{st} + \mathbb{X}^{\beta\alpha}_{st}) \\
&\approx \partial_\alpha f^i(X_s)\partial_\beta f^j(X_s)(X^\alpha_{st}X^\beta_{st} - (\mathbb{X}^{\alpha\beta}_{st} + \mathbb{X}^{\beta\alpha}_{st})) \\
&= \partial_\alpha f^i(X_s)\partial_\beta f^j(X_s)[\boldsymbol{X}]^{\alpha\beta}_{st} \\
&\approx \int \partial_i f(X)\partial_j f(X)\mathrm{d}[\boldsymbol{X}]^{ij}
\end{aligned} \tag{2.38}$$

and since the integral is additive on consecutive intervals we conclude that we have equality by uniqueness in Lemma 2.2. As for the rough integral

$$\begin{aligned}
\left[\int \boldsymbol{H}\mathrm{d}\boldsymbol{X}\right]^{ij}_{st} &\approx (H^i_{\gamma;s}X^\gamma_{st} + H'^i_{\alpha\beta;s}\mathbb{X}^{\alpha\beta}_{st})(H^j_{\gamma;s}X^\gamma_{st} + H'^j_{\alpha\beta;s}\mathbb{X}^{\alpha\beta}_{st}) - H^i_{\alpha;s}H^j_{\beta;s}\mathbb{X}^{\alpha\beta}_{st} \\
&\approx H^i_{\alpha;s}H^j_{\beta;s}(X^\alpha_{st}X^\beta_{st} - (\mathbb{X}^{\alpha\beta}_{st} + \mathbb{X}^{\beta\alpha}_{st})) \\
&= H^i_{\alpha;s}H^j_{\beta;s}[\boldsymbol{X}]^{\alpha\beta}_{st} \\
&\approx \int_s^t H^i_\alpha H^j_\beta \mathrm{d}[\boldsymbol{X}]^{\alpha\beta}
\end{aligned} \tag{2.39}$$

and conclude as before that equality holds. ■

The fact that the rough integral can be canonically considered a rough path in its own right naturally leads to the question of associativity, which is answered in the affirmative:

Theorem 2.16 (Associativity of the rough integral)**.** *Let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$, $\boldsymbol{H} \in \mathscr{D}_X(\mathbb{R}^{e\times d})$,*

$\boldsymbol{I} \coloneqq \int \boldsymbol{H} \mathrm{d}\boldsymbol{X} \in \mathscr{D}_X(\mathbb{R}^e)$, $\boldsymbol{Y} \coloneqq \uparrow_{\mathbb{X}} \boldsymbol{I}$, $\boldsymbol{K} \in \mathscr{D}_I(\mathbb{R}^{f \times e})$. *Then*

$$\left(\int \boldsymbol{K} \mathrm{d}\boldsymbol{Y} \right) * I' = \int (\boldsymbol{K} * I') \cdot \boldsymbol{H} \mathrm{d}\boldsymbol{X} \in \mathscr{D}_X(\mathbb{R}^f) \tag{2.40}$$

As a result, the identity $\int \boldsymbol{K} \mathrm{d}\boldsymbol{Y} = \int (\boldsymbol{K} * H) \cdot \boldsymbol{H} \mathrm{d}\boldsymbol{X}$ *also holds in* $\mathscr{C}^p_\omega([0,T], \mathbb{R}^f)$.

Proof. At the level of the trace we have

$$\begin{aligned} \int_s^t \boldsymbol{K}^c_k \mathrm{d}\boldsymbol{Y}^k &\approx K^c_{k;s} Y^k_{st} + K'^c_{ij;s} \mathbb{Y}^{ij}_{st} \\ &\approx K^c_{k;s}(H^k_{\gamma;s} X^\gamma_{st} + H'^k_{\alpha\beta;s} \mathbb{X}^{\alpha\beta}_{st}) + K'^c_{ij;s} H^i_{\alpha;s} H^j_{\beta;s} \mathbb{X}^{\alpha\beta}_{st} \\ &\approx \int_s^t (\boldsymbol{K} * H) \cdot \boldsymbol{H} \mathrm{d}\boldsymbol{X} \end{aligned} \tag{2.41}$$

which proves the identity of the traces, since $I' = H$. Their Gubinelli derivatives w.r.t. X, as computed according to [Example 2.7, 2.] and Proposition 2.12 both coincide with $(K^c_k H^k_\gamma)$. Passing to the lift on this identity we have

$$\Big\uparrow_{\mathbb{X}} \int (\boldsymbol{K} * H) \cdot \boldsymbol{H} \mathrm{d}\boldsymbol{X} = \Big\uparrow_{\mathbb{X}} \int \boldsymbol{K} \mathrm{d}\boldsymbol{Y} * H = \Big\uparrow_{\mathbb{Y}} \int \boldsymbol{K} \mathrm{d}\boldsymbol{Y} \tag{2.42}$$

where we have used [Proposition 2.13, 1.] in the second identity. This is the identity required in the second statement. ■

The next proposition expresses the degree to which pushforward of rough paths and pullback of controlled paths fail to be adjoint operators under the rough integral pairing; in particular the adjunction does hold when the integrator is geometric or when g is an affine map.

Corollary 2.17. *Let* $\boldsymbol{X}, \boldsymbol{H}, g$ *be as in [Proposition 2.12, Pullback]. Then*

$$\left(\int \boldsymbol{H} \mathrm{d}(g_* \boldsymbol{X}) \right) * Dg(X) = \int g^* \boldsymbol{H} \mathrm{d}\boldsymbol{X} + \frac{1}{2} \int H \cdot D^2 g(X) \mathrm{d}[\boldsymbol{X}] \tag{2.43}$$

where, as usual, the identity holds in $\mathscr{D}_X(\mathbb{R}^e)$ *according to Example 2.7 and thus in* $\mathscr{C}^p_\omega([0,T], \mathbb{R}^e)$.

Proof. Plugging in the the expression for $g_* \boldsymbol{X}$ given by Theorem 2.14 and applying Theorem 2.16 we have

$$\begin{aligned} \left(\int \boldsymbol{H} \mathrm{d}(g_* \boldsymbol{X}) \right) * Dg(X) &= \left(\int \boldsymbol{H} \mathrm{d}\left(\int \boldsymbol{Dg}(X) \mathrm{d}\boldsymbol{X} + \tfrac{1}{2} \int D^2 g(X) \mathrm{d}[\boldsymbol{X}]\right) \right) * Dg(X) \\ &= \int (\boldsymbol{H} * Dg(X)) \cdot \boldsymbol{Dg}(X) \mathrm{d}\boldsymbol{X} + \frac{1}{2} \int H \cdot D^2 g(X) \mathrm{d}[\boldsymbol{X}] \\ &= \int g^* \boldsymbol{H} \mathrm{d}\boldsymbol{X} + \frac{1}{2} \int H \cdot D^2 g(X) \mathrm{d}[\boldsymbol{X}] \end{aligned} \tag{2.44}$$

As usual, the more regular Young integral only contributes to the trace of the X-controlled/rough paths in question. ■

2.1.3 Rough differential equations

We proceed to discuss a central theme of rough path theory: that of rough differential equations, or RDEs.

Definition 2.18. Let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$, $F \in C^\infty(\mathbb{R}^{e+d},\mathbb{R}^{e\times d})$. A *controlled solution* to the RDE

$$\mathrm{d}\boldsymbol{Y} = F(Y,X)\mathrm{d}\boldsymbol{X}, \quad Y_0 = y_0 \tag{2.45}$$

(which we will write in coordinates as $\mathrm{d}\boldsymbol{Y}^k = F^k_\gamma(Y,X)\mathrm{d}\boldsymbol{X}^\gamma$ when we wish to emphasise the action of the field of linear maps F on the driver $\boldsymbol{X}$) is an element $\boldsymbol{Y} \in \mathscr{D}_X(\mathbb{R}^e)$ s.t.

$$\boldsymbol{Y} = y_0 + \int F_*(\boldsymbol{Y},\boldsymbol{X})\mathrm{d}\boldsymbol{X} \in \mathscr{D}_X(\mathbb{R}^e) \tag{2.46}$$

where $F_*(\boldsymbol{Y},\boldsymbol{X})$ is the pushforward of the $\mathbb{R}^{e+d}$-valued X-controlled path with trace (Y,X) and Gubinelli derivative $(Y',\mathbb{1})$. We will call $\uparrow_{\mathbb{X}}\boldsymbol{Y}$ (which we will denote again $\boldsymbol{Y}$) a *rough path solution* to (2.45).

We will sometimes write $\mathrm{d}Y$ (without the bold font for Y) on the LHS of (2.45) when only referring to the trace level of the solution. Note that the definition of controlled solution implies the requirement $Y' = F(Y,X)$ and

$$F_*(\boldsymbol{Y},\boldsymbol{X}) = (F^k_\gamma(Y,X), \partial_\alpha F^k_\beta(Y,X) + F^h_\alpha \partial_h F^k_\beta(Y,X)) \tag{2.47}$$

Since a solution of either type is entirely determined by its trace and F, $\boldsymbol{X}$ we will often just use the term *solution* without specifying which type we intend.

Remark 2.19. Usually only RDEs of the form $\mathrm{d}\boldsymbol{Y} = F(Y)\mathrm{d}\boldsymbol{X}$ are considered. (2.45) can be considered as a special case of this by simply "doubling the variables", i.e. considering the joint RDE

$$\mathrm{d}\begin{pmatrix}\boldsymbol{X}\\ \boldsymbol{Y}\end{pmatrix} = \begin{pmatrix}\mathbb{1}\\ F(Y,X)\end{pmatrix}\mathrm{d}\boldsymbol{X} \tag{2.48}$$

We have chosen to consider RDEs that also depend on X since this will become a compulsory requirement when $\boldsymbol{X}$ is manifold-valued; this framework is taken from [É89], where it is used in the context of manifolds-valued SDEs. A side benefit of introducing this dependence is that rough integrals may now be seen as a particular case of RDEs, namely when F is independent of Y. In particular, thanks to Theorem 2.14, controlled paths given by one-forms may be viewed as particular cases of solutions to RDEs driven by the rough path $(\boldsymbol{X},[\boldsymbol{X}])$.

Example 2.20 (RDEs driven by rough paths with common trace)**.** Let ${}_1\boldsymbol{X}, {}_2\boldsymbol{X}, D$ be as in Example 2.8. Then, recalling the convention used in the previous chapter whereby Greek letters index the driver and Latin ones the solution, we have the following identity of controlled solutions

$$\mathrm{d}Y^k = F^k_\gamma(Y,X)\mathrm{d}_2\boldsymbol{X}^\gamma \iff \mathrm{d}Y^k = F^k_\gamma(Y,X)\mathrm{d}_1\boldsymbol{X}^k + (\partial_\alpha F^k_\beta + F^h_\alpha\partial_h F^k_\beta)(Y,X)\mathrm{d}D^{\alpha\beta} \tag{2.49}$$

Note this identity does not hold for rough path solutions, for the reason provided in Example 2.11. The second expression is an RDE driven by the rough path with trace (X,D) and second order part $\mathbb{X}$. This is particularly important when ${}_2\boldsymbol{X} = {}_\mathrm{g}\boldsymbol{X}$, $D = \frac{1}{2}[\boldsymbol{X}]$ for a rough path $\boldsymbol{X}$, as it informs us that every RDE may be rewritten as an RDE driven by the geometric rough path $({}_\mathrm{g}\boldsymbol{X},[\boldsymbol{X}])$.

The following theorem is proved in [CDL15, Corollary 2.17, Theorem 4.2], and its proof carries over to the case of $\boldsymbol{X}$ non-geometric (thanks to Example 2.20) and with F depending on X (thanks to Remark 2.19).

We will say that $\boldsymbol{Y}$ is a controlled/rough path solution *up to time* $S \leq T$ if it is a solution to (2.45) where the driving rough path is substituted with $\boldsymbol{X}|_{[0,R]} \in \mathscr{C}^p_\omega([0,R],\mathbb{R}^d)$, for all $R < S$. Note that, according to this terminology, a solution up to time T is a not necessarily a solution on the whole of $[0,T]$ (the former may explode precisely at time T, while for the latter we have $Y_T = y_0 + \int_0^T F_*(\boldsymbol{Y},\boldsymbol{X})\mathrm{d}\boldsymbol{X}$): to distinguish the two we will call the latter a *global solution*.

Theorem 2.21 (Local existence and uniqueness). *Precisely one of the following two possibility holds w.r.t.* (2.45)

1. *A global solution exists;*
2. *There exists an $S \leq T$ and a solution up to time S, with $Y_{[0,S)}$ not contained in any compact set of $\mathbb{R}^e$.*

Moreover, in either case, the solution is unique on the interval on which it is defined.

The following lemma further specifies that the exit time from an open neighbourhood is bounded from below, uniformly in the initial time and initial condition (ranging in a precompact neighbourhood) of an RDE with fixed driver $\boldsymbol{X}$. It can be found in [CDL15, Corollary 2.17], and its proof carries over to the setting considered here once again by Example 2.20 and Remark 2.19 (and using the obvious fact that $X_{[0,T]}$ is compact).

Lemma 2.22. *Let $U, V \subseteq \mathbb{R}^e$ be open with $V \supseteq \overline{U}$ compact. Then there exists a $\delta > 0$ s.t. for all $t_0 \in [0,T]$ and $y_0 \in U$ the unique solution to*

$$\mathrm{d}Y = F(Y,X)\mathrm{d}\boldsymbol{X}, \quad Y_{t_0} = y_0 \tag{2.50}$$

is defined and satisfies $Y \in V$ on $[t_0, (t_0+\delta)\wedge T]$.

Although this thesis is not about global existence, we will need the following lemma that guarantees it in an important special case.

Lemma 2.23. *Let F be as in Definition 2.18 with*

$$F^k_\gamma(y,x) = A^k_{\gamma h}(x)y^h + b^k_\gamma(x) \tag{2.51}$$

for some $A \in C^\infty(\mathbb{R}^d, \mathbb{R}^{e\times e\times d})$, $b \in C^\infty(\mathbb{R}^d,\mathbb{R}^{e\times d})$. Then (2.45) *admits a global solution.*

Proof. First of all, observe that

$$(\partial_\alpha F^k_\beta + F^h_\alpha \partial_h F^k_\beta)(y,x) = (\partial_\alpha A^k_{\beta h} + A^l_{\alpha h}A^k_{\beta l})(x)y^h + (\partial_\alpha b^k_\beta + b^h_\alpha A^k_{h\beta})(x) \tag{2.52}$$

has the same form as F: by Example 2.20 we may therefore assume $\boldsymbol{X}$ is geometric. Now, by Theorem 2.16, we may view the equation as linear, driven by the rough path $\int (A(X), b(X))\mathrm{d}\boldsymbol{X}$, and may therefore directly apply the result for global existence of linear RDEs [FV10b, Theorem 10.53]. ∎

2.1.4 Stochastic rough paths

Finally, we address the topic of stochastic processes lifted to rough paths. We denote $\mathcal{S}(\Omega,[0,T],\mathbb{R}^d)$ the set of $\mathbb{R}^d$-valued continuous adapted semimartingales defined up to time T on some stochastic setup

$(\Omega, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ satisfying the usual conditions. We may define $\boldsymbol{X}, \widehat{\boldsymbol{X}} \in \mathscr{C}^p([0,T],\mathbb{R}^d)$ a.s. by Stratonovich and Itô integration respectively

$$\mathbb{X}^{\alpha\beta}_{st} := \int_s^t X^\alpha_{su} \circ \mathrm{d}X^\beta_u, \quad \widehat{\mathbb{X}}^{\alpha\beta}_{st} := \int_s^t X^\alpha_{su} \mathrm{d}X^\beta_u, \quad \mathbb{X}^{\alpha\beta}_{st} = \widehat{\mathbb{X}}^{\alpha\beta}_{st} + \tfrac{1}{2}[X]^{\alpha\beta}_{st} \tag{2.53}$$

where $[X]$ denotes the quadratic covariation tensor of X.

Remark 2.24. We have $[\widehat{\boldsymbol{X}}] = [X]$ and $[\boldsymbol{X}] = 0$ a.s., so ${}_{\mathrm{g}}\widehat{\boldsymbol{X}} = \boldsymbol{X}$. In general, rough path theory applied to semimartingales extends the usual stochastic calculus, i.e. Itô/Stratonovich stochastic integrals agree a.s. with the path-by-path computed rough integrals w.r.t. the Itô/Stratonovich lifts [FH14, Proposition 5.1, Corollary 5.2], and the strong solution to an Itô/Stratonovich SDE coincides a.s. with the path-by-path computed solution to the RDE driven by the Itô/Stratonovich-enhanced rough path [FH14, Theorem 9.1] (these results are only shown for Brownian integrators, but may be extended to general continuous semimartingales, e.g. by reducing to the Brownian case by splitting the integrator into its bounded variation and local martingale parts and applying the Dubins-Schwarz theorem to the latter).

The following statement made in the same spirit, which we could not find in the literature, will be important later on.

Proposition 2.25. *Let $X \in \mathcal{S}(\Omega, [0,T])$ and $f \in C^\infty(\mathbb{R}^d, \mathbb{R}^e)$. Then $f_*\boldsymbol{X}$ and $f_*\widehat{\boldsymbol{X}}$ coincide a.s. with the lifts of the semimartingale $f(X)$ computed respectively through Stratonovich and Itô integration.*

Proof. We begin with the Itô case. By the classical Itô formula and Remark 2.24 we have that, a.s.

$$\begin{aligned}
\int_s^t f^i(X)\mathrm{d}f^j(X) &= \int_s^t f^i(X)\partial_\gamma f^j(X)\mathrm{d}X + \frac{1}{2}\int_s^t f^i(X)\partial_{\alpha\beta}f(X)\mathrm{d}[X]^{\alpha\beta} \\
&= \int_s^t f^i\partial_\gamma f^j(X)\mathrm{d}\widehat{\boldsymbol{X}}^\gamma + \frac{1}{2}\int_s^t f^i\partial_{\alpha\beta}f(X)\mathrm{d}[\widehat{\boldsymbol{X}}]^{\alpha\beta} \\
&\approx f^i\partial_\gamma f^j(X_s)X^\gamma_{st} + (\partial_\alpha f^i\partial_\beta f^j + f^i\partial_{\alpha\beta}f^j)(X_s)\widehat{\mathbb{X}}^{\alpha\beta}_{st} \\
&\quad + \tfrac{1}{2}f^i\partial_{\alpha\beta}f^j(X_s)[\widehat{\boldsymbol{X}}]^{\alpha\beta}_{st} \\
&= f^i\partial_\gamma f^j(X_s)X^\gamma_{st} + \partial_\alpha f^i\partial_\beta f^j\widehat{\mathbb{X}}^{\alpha\beta}_{st} + \tfrac{1}{2}f^i\partial_{\alpha\beta}f^j(X_s)X^\alpha_{st}X^\beta_{st} \\
&\approx f^i(X_s)f^j(X)_{st} + \partial_\alpha f^i\partial_\beta f^j\widehat{\mathbb{X}}^{\alpha\beta}_{st}
\end{aligned} \tag{2.54}$$

Therefore a.s.

$$\int_s^t f^i(X)_{su}\mathrm{d}f^j(X_u) = \int_s^t f^i(X)\mathrm{d}f^j(X) - f^i(X_s)f^j(X)_{st} \approx \partial_\alpha f^i\partial_\beta f^j\widehat{\mathbb{X}}^{\alpha\beta}_{st} \tag{2.55}$$

and we conclude by Lemma 2.2. The Stratonovich case is handled analogously, with the only difference that the first order change of variable formula holds, and that brackets vanish. ■

For other examples of stochastic rough paths, which include lifts of Gaussian and Markov processes, we refer to [FV10b, Ch. III]. Though these rough paths are mostly geometric, examples of non-geometric, non-semimartingale stochastic rough paths also exist in the literature [QX18].

2.2 Background on differential geometry

In this section we review the differential geometry needed in the rest of this chapter. We begin by recalling various equivalent notions of connections on manifolds, and proceed to specialise this study to the case where M is a Riemannian submanifold of $\mathbb{R}^d$. We follow [Lee97] and [KN96] for classical differential geometry (with an occasional glance at [Nak03] for expressions in local coordinates), and [Lee97, Ch. 8], [CDL15], Section 1.2 and [Dri04] for the extrinsic theory.

2.2.1 Linear connections

Let M be a smooth m-dimensional manifold, and $\tau M\colon TM \to M$ its tangent bundle; throughout this chapter we will identify fibre bundles with their projection. We will denote the tangent map of a smooth map of manifolds $f\colon M \to N$ by $\tau f\colon \tau M \to \tau N$ (a morphism in the category of vector bundles), the map of total spaces by $Tf\colon TM \to TN$ (a smooth map), and by $T_x f$ its restriction to the tangent space $T_x M$ (a linear map). In this subsection we review equivalent notions of a connection on a manifold. Given a smooth fibre bundle $\pi\colon E \to M$ we denote $\Gamma\pi$ its $C^\infty M$-module of sections and $E_A := \pi^{-1}(A)$ for $A \subseteq M$, $E_x := E_{\{x\}}$ for $x \in M$.

Definition 2.26 (Covariant derivative)**.** A *linear connection*, or *covariant derivative* on a smooth vector bundle $\pi\colon E \to M$ is a map

$$\nabla\colon \Gamma\tau M \times \Gamma\pi \to \Gamma\pi, \quad (U, e) \mapsto \nabla_U e, \quad (\nabla_U e)(x) =: \nabla_{U(x)} e \tag{2.56}$$

which is linear in both arguments and which satisfies the Leibniz rule $\nabla_{U(x)}(fe) = f(x)\nabla_{U(x)}e + (U(x)f)e(x)$ for $f \in C^\infty M$.

The notation $\nabla_{U(x)}e$ is justified by the fact that the value of the section $\nabla_U e$ only depends on the value of $U(x)$ (and on the value of e on any curve whose tangent vector at x is $U(x)$); in general we will denote vectors based at $x \in M$ as $U(x), V(x), \ldots$, reserving $U, V, \ldots$ for vector fields (or just vectors based at an unspecified point). Covariant derivatives will mainly be considered on the tangent bundle τM: in this case it is automatically extended to the whole of $\bigoplus_{k,l\in\mathbb{N}} \tau M^{\otimes k} \otimes \tau^* M^{\otimes l}$ given a few compatibility conditions [Lee97, Lemma 4.6]. A linear connection on τM is equivalently defined by a *Hessian*, as described in (1.34). Given a chart φ, recall that we denote $\partial_k\varphi(x)$ the basis elements of the tangent space $T_x M$ defined by φ, and we abbreviate $\partial_k := \partial_k\varphi$ if there is no risk of ambiguity. Moreover, we denote Γ^k_{ij} the Christoffel symbols of ∇ w.r.t. φ: this means $\nabla_{\partial_i}\partial_j = \Gamma^k_{ij}\partial_k$, and therefore

$$\nabla_U V = (U^h \partial_h V^k + U^i V^j \Gamma^k_{ij})\partial_k \tag{2.57}$$

and if $\omega \in \Gamma\tau^* M$

$$\nabla_U \omega = (U^i \partial_i \omega_j - U^i \omega_k \Gamma^k_{ij})\mathrm{d}^j \tag{2.58}$$

where $\mathrm{d}^k := \mathrm{d}\varphi^k$ are the elements of the dual basis of $\{\partial_k\varphi\}_k$. Given two charts $\varphi, \overline{\varphi}$ defined on overlapping domains, the (non-tensorial) transformation rule of the Christoffel symbols is

$$\Gamma^{\overline{k}}_{\overline{i}\overline{j}} = \partial^{\overline{k}}_k \partial^i_{\overline{i}} \partial^j_{\overline{j}} \Gamma^k_{ij} + \partial^{\overline{k}}_h \partial^h_{\overline{i}\overline{j}} \tag{2.59}$$

where overlined indices refer to $\overline{\varphi}$ and simple indices to φ, and the ∂'s refer to the derivatives of the change of chart, e.g. $\partial^{h}_{\overline{ij}}(x) := \partial_{\overline{ij}}(\varphi^h \circ \overline{\varphi}^{-1})(x)$. The Hessian can be written in coordinates as

$$\nabla^2 f = (\partial_{ij} - \Gamma^k_{ij}\partial_k) f \,\mathrm{d}\varphi^i \otimes \mathrm{d}\varphi^j \tag{2.60}$$

Those connections whose Hessians are valued in $\Gamma(\tau^* M^{\odot 2})$, or equivalently with vanishing *torsion* tensor $\langle \mathcal{T}, U \otimes V\rangle := \nabla_U V - \nabla_V U - [U,V]$ are called *torsion-free*. In local coordinates $\mathcal{T}^k_{ij} = \Gamma^k_{ij} - \Gamma^k_{ji}$. We can associate to any connection ∇ a torsion-free one by ${}^{\odot}\nabla_U V := \nabla_U V - \frac{1}{2}\langle \mathcal{T}, U \otimes V\rangle$ or equivalently by projecting its Hessian onto $T^*M \odot T^*M$: the symmetrised connection will then define the same set of (parametrised) geodesics.

Let g be a Riemannian metric, i.e. an element of $\Gamma(\tau^* M^{\odot 2})$ which is nowhere vanishing and positive-definite at all points (many, but not all, of the considerations made in this chapter about Riemannian metrics can be extended to pseudo-Riemannian ones). A connection ∇ is *metric* w.r.t. g if $\nabla g = 0$, or in local coordinates

$$g_{ij,k} - g_{hj}\Gamma^h_{ki} - g_{ih}\Gamma^h_{kj} = 0 \tag{2.61}$$

where indices after the comma denote coordinate partial differentiation, i.e. $g_{ij,k} := \partial_k g_{ij}$, and g_{ij} the components of the metric in the same chart (g^{ij} will denote the inverse of g_{ij}, i.e. $g^{ik} g_{kj} = \delta^i_j$). We will also use indices after a semicolon to denote covariant differentiation, e.g. $g_{ij;k} := (\nabla g)_{ijk}$. There is precisely one such connection which is also torsion-free, called the *Levi-Civita* connection of g, which we denote ${}^g\nabla$, and its Christoffel symbols are given by

$${}^g\Gamma^k_{ij} = \tfrac{1}{2} g^{kh}(g_{hj,i} + g_{ih,j} - g_{ij,h}) \tag{2.62}$$

When on a Riemannian manifold we will sometimes use the *musical isomorphisms* $^\flat : \tau M \to \tau^* M$ with inverse $^\sharp$. In coordinates these are given by performing "index gymnastics" w.r.t. g, i.e. $V_i := V^\flat_i := g_{ij}V^j$, $\omega^i := (\omega^\sharp)^i = \omega_j g^{ij}$. Similar raising and lowering of indices will be performed with arbitrary tensors.

Remark 2.27. If ∇ is g-metric, it is not true in general that ${}^{\odot}\nabla$ is metric. Denoting $\mathcal{T}^{\,k}_{\;ij}$ the components of the torsion tensor, we have that the difference between ∇ and ${}^g\nabla$ is quantified by the *contorsion* tensor

$$\mathcal{K}^k_{ij} := \tfrac{1}{2}(\mathcal{T}^{\,k}_{\;ij} + \mathcal{T}_{i\;j}^{\;k} + \mathcal{T}_{j\;i}^{\;k}) = \Gamma^k_{ij} - {}^g\Gamma^k_{ij} \tag{2.63}$$

which has symmetric part $\frac{1}{2}(\mathcal{T}_{i\;j}^{\;k} + \mathcal{T}_{j\;i}^{\;k})$.

The *curvature tensor* associated to a connection ∇ is

$$\mathcal{R}(U,V)W := \nabla_U\nabla_V W - \nabla_V\nabla_U W - \nabla_{[U,V]}W \tag{2.64}$$

where $[U,V]$ denotes the Lie bracket of vector fields, which vanishes if the vectors are given by the local basis sections ∂_k defined by a chart. We denote the coefficients

$$\mathcal{R}_{ijk}{}^h := \langle R(\partial_i,\partial_j)\partial_k, \mathrm{d}^h\rangle = \Gamma^h_{jk,i} - \Gamma^h_{ik,j} + \Gamma^h_{il}\Gamma^l_{jk} - \Gamma^h_{jl}\Gamma^l_{ik} \tag{2.65}$$

and warn the reader that the ordering of the indices is not standard in the literature (this convention is, for instance, the one followed by [Lee97, YI73]). The curvature tensor satisfies the symmetry

$$\mathscr{R}_{ijk}{}^{h} = -\mathscr{R}_{jik}{}^{h} \tag{2.66}$$

Moreover, if ∇ is torsion-free

$$\mathscr{R}_{ijk}{}^{h} + \mathscr{R}_{jki}{}^{h} + \mathscr{R}_{kij}{}^{h} = 0 \tag{2.67}$$

Moreover, if ∇ is $\mathscr{g}$-metric (but not necessarily torsion-free)

$$\mathscr{R}_{ijkh} = -\mathscr{R}_{ijhk} \tag{2.68}$$

and finally if $\nabla = {}^{\mathscr{g}}\nabla$

$$\mathscr{R}_{ijkh} = \mathscr{R}_{khij} \tag{2.69}$$

These symmetries are often stated directly in the Levi-Civita case, but hold under the more general hypotheses stated above, as can be seen from a careful reading of their proof [Lee97, Proposition 7.4]. We also recall the definition of *Ricci tensor* a symmetric tensor field defined as a contraction of the curvature tensor, and whose components we still denote (without ambiguity, thanks to the different number of indices) by the symbol $\mathscr{R}$:

$$\mathscr{R}_{ij} = -\mathscr{R}_{ki}{}^{k}{}_{j} = -\mathscr{R}_{hikj}\,\mathscr{g}^{hk} \tag{2.70}$$

Given a smooth fibre bundle $\pi\colon E \to M$ with typical fibre the smooth n-dimensional manifold R, its *vertical bundle* $V\pi$ is the subbundle of τE with total space $VE := \ker(T\pi\colon TE \to TM)$, and we have $V_{e(x)}E = T_{e(x)}E_x$, i.e. elements of the total space of $V\pi$ are vectors tangent to the fibres of π. Recall that for a smooth map of manifolds $f \in C^\infty(P, Q)$ and a fibre bundle $\rho\colon D \to Q$, we define the *pullback* bundle

$$f^*\rho\colon \{(p,d) \in P \times D \mid f(p) = \rho(d)\} := f^*Q \to P, \quad (p,d) \mapsto p \tag{2.71}$$

and there is a bundle map $f^*\rho \to \rho$

$$\begin{array}{ccc} f^*Q & \xrightarrow{\ \mathrm{pr}_2\ } & D \\ {\scriptstyle f^*\rho}\big\downarrow & & \big\downarrow{\scriptstyle \rho} \\ P & \xrightarrow{\ f\ } & Q \end{array} \tag{2.72}$$

The *vertical lift* of π is defined as the fibre bundle isomorphism

$$\begin{aligned} &\pi^*\pi \to V\pi, \quad E_x \times E_x \ni (e(x), U(x)) \mapsto \mathscr{v}(e(x))U(x) \\ &\mathscr{v}(e(x))U(x)(f \in C^\infty E) := \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_0 f(e(x) + tU(x)) \end{aligned} \tag{2.73}$$

An *Ehresmann connection* is a vector bundle $\eta\colon H \to E$ which is complementary to $V\pi$, i.e. $H \oplus VE = TE$. When π is a vector bundle, Ehresmann connections and a covariant derivatives are equivalent by further requiring of the former that, denoting the sum and scalar multiplication map by

$$\Sigma\colon E \oplus E \to E, \qquad \Lambda_a\colon E \to E,\ a \in \mathbb{R} \tag{2.74}$$

$T\Sigma$ map the subbundle $\{(\alpha(e), \beta(e)) \in H_e \oplus H_e \mid e \in E\} \leq T(E \oplus E)$ to H and that $T\Lambda_a$ map H to itself for all $a \in \mathbb{R}$. This in particular implies that $H_{0_x} = T_x M$ where we are identifying M with the zero section of TM. In order to describe the correspondence we first define the *horizontal lift* (relative to an Ehresmann connection $\eta\colon H \to M$ on the fibre bundle π) as the fibre bundle isomorphism

$$\hbar\colon \pi^*\tau M \to \eta, \quad E_x \times T_x M \ni (e(x), U(x)) \mapsto \hbar(e(x))U(x) \coloneqq T_{e(x)}\pi\big|_{H_{e(x)}}^{-1}(U(x)) \tag{2.75}$$

i.e. $\hbar$ is a splitting of the short exact sequence of vector bundles:

$$0 \longrightarrow V\pi \longrightarrow \tau E \underset{T\pi}{\overset{\hbar}{\rightleftarrows}} \pi^*\tau M \longrightarrow 0 \tag{2.76}$$

The Ehresmann connection associated to a covariant derivative (where π now is a vector bundle) is given in terms of its horizontal lift as

$$\hbar(e(x))U(x) \coloneqq T_x e(U(x)) - \mathcal{v}(e(x))\nabla_{U(x)} e \tag{2.77}$$

for any section $e \in \Gamma\pi$ whose value at x is $e(x)$ (the independence on the section e is checked by using the usual characterisation of tensoriality [Lee97, Lemma 2.4], i.e. by showing that $\hbar(fe(x))U(x) = f(x)\hbar(e(x))U(x)$: this is easily done in local coordinates).

If we have a chart $\varphi\colon A \to \mathbb{R}^m$ for $A \subseteq M$, a chart $\phi\colon B \to \mathbb{R}^n$ for $B \subseteq R$ (the typical fibre of π, an arbitrary n-dimensional manifold) and a trivialisation $\Phi\colon E_A \to A \times R$, the triple (φ, ϕ, Φ) defines a chart

$$(\varphi \times \phi) \circ \Phi\colon \Phi^{-1}(A \times B) \to \mathbb{R}^m \times \mathbb{R}^n \tag{2.78}$$

We will call the resulting coordinates *product coordinates*. If π is a vector bundle, R can (and always will) be taken equal to $\mathbb{R}^n$ and ϕ to the identity, and if $\pi = \tau M$ or $\tau^* M$, Φ can be defined canonically in terms of φ as $T\varphi$ or $T^*\varphi^{-1}$. In these cases we will speak of *induced coordinates*.

Convention 2.28. In what follows we will be working on the manifolds TM (or T^*M) and E. It will therefore be helpful to establish conventions regarding indexing of the product and induced coordinates. In the absence of other manifolds, ambiguities as to the chart, etc. we will denote with Greek indices $\alpha, \beta, \gamma, \ldots = 1, \ldots, m$ the coordinates on M, with Latin indices $i, j, k, \ldots = m+1, \ldots, m+n$ the coordinates on E in excess of the aforementioned coordinates of the base space M and with $\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\gamma}, \ldots = m+1, \ldots, 2m$ the induced coordinates on TM in excess of those on M. More specifically, $\widetilde{\gamma} \coloneqq m + \gamma$, and we will take this into account when using the Einstein convention, e.g. $a_{\alpha\beta} b^{\widetilde{\beta}\gamma} = \sum_{\beta=1}^m a_{\alpha\beta} b^{(m+\beta)\gamma}$. Moreover, we will use capital letters $I, J, K, \ldots = 1, \ldots, m+n$ to denote indices that run through all coordinates on E, and capital letters $A, B, C, \ldots = 1, \ldots, 2m$ to denote indices that run through all the coordinates on TM. The following diagrams should help explain this arrangement:

$$
\begin{aligned}
E: &\quad \underbrace{(\overbrace{x^1,\ldots,x^m}^{\alpha,\beta,\gamma,\ldots},\overbrace{y^1,\ldots,y^n}^{i,j,k,\ldots})}_{I,J,K,\ldots} \\
TM: &\quad \underbrace{(\overbrace{x^1,\ldots,x^m}^{\alpha,\beta,\gamma,\ldots},\overbrace{\widetilde{x}^{\widetilde{1}},\ldots,\widetilde{x}^{\widetilde{m}}}^{\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma},\ldots})}_{A,B,C,\ldots}
\end{aligned}
\tag{2.79}
$$

A similar convention is followed when T^*M is replaced with TM.

It is important to point out the following potential source of confusion. If $V(x) \in T_xM$ it can be either viewed as a vector in the vector space T_xM, with coordinates $V^\gamma(x)$, or as a point in the manifold TM, with coordinates

$$(V(x)^\gamma, V(x)^{\widetilde{\gamma}}) = (x^\gamma, V^\gamma(x)) \tag{2.80}$$

Note the different meaning of $V(x)^\gamma$ and $V^\gamma(x)$; in any case, this ambiguity will be avoided by always considering elements as vectors whenever otherwise mentioned. The use of the twiddled indices is seen when considering vectors in TTM and TT^*M.

Finally, we mention that the use of Greek/Latin indices will also be used in the separate case in which we are dealing with two different manifolds M and N, to distinguish between coordinates on the two manifolds.

In the case of π a vector bundle the change of product coordinates from φ, Φ to $\overline{\varphi}$, $\overline{\Phi}$ can be written as

$$\partial_K^{\overline{K}}(y) = \begin{pmatrix} \partial_\gamma^{\overline{\gamma}}(x) & 0 \\ \partial_\gamma \lambda_k^{\overline{k}}(x)y^k & \lambda_k^{\overline{k}}(x) \end{pmatrix}, \quad (\overline{\Phi} \circ \Phi^{-1})(x,y) = (x, \lambda(x)y) \tag{2.81}$$

for $x = \pi(y)$ and $\lambda \in C^\infty(\varphi(A), \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n))$. It is worthwhile to specify this to the cases of $\pi = \tau M$ (where $\Phi = T\varphi$) and τ^*M ($\Phi = T^*\varphi^{-1}$), so $\lambda_{\widetilde{\gamma}}^{\widetilde{\overline{\gamma}}} = \partial_\gamma^{\overline{\gamma}}$ and $\lambda_{\widetilde{\gamma}}^{\widetilde{\overline{\gamma}}} = \partial_{\overline{\gamma}}^{\gamma}$ respectively, and

$$\pi = \tau M: \quad \partial_C^{\overline{C}}(y) = \begin{pmatrix} \partial_\gamma^{\overline{\gamma}} & \partial_{\widetilde{\gamma}}^{\overline{\gamma}} \\ \partial_\gamma^{\widetilde{\overline{\gamma}}} & \partial_{\widetilde{\gamma}}^{\widetilde{\overline{\gamma}}} \end{pmatrix}(y) = \begin{pmatrix} \partial_\gamma^{\overline{\gamma}}(x) & 0 \\ \partial_{\gamma\alpha}^{\overline{\gamma}}(x)y^\alpha & \partial_\gamma^{\overline{\gamma}}(x) \end{pmatrix} \tag{2.82}$$

$$\pi = \tau^*M: \quad \partial_C^{\overline{C}}(y) = \begin{pmatrix} \partial_\gamma^{\overline{\gamma}} & \partial_{\widetilde{\gamma}}^{\overline{\gamma}} \\ \partial_\gamma^{\widetilde{\overline{\gamma}}} & \partial_{\widetilde{\gamma}}^{\widetilde{\overline{\gamma}}} \end{pmatrix}(y) = \begin{pmatrix} \partial_{\overline{\gamma}}^{\gamma}(x) & 0 \\ \partial_{\overline{\beta}\overline{\gamma}}^{\alpha}\partial_\gamma^{\overline{\beta}}(x)y_\alpha & \partial_{\overline{\gamma}}^{\gamma}(x) \end{pmatrix} \tag{2.83}$$

The expression of the horizontal lift in induced coordinates in the case of $\pi = \tau M$ is well known and given by (see for example [É89, p.115])

$$(\hbar(V)U)^\gamma = U^\gamma, \quad (\hbar(V)U)^{\widetilde{\gamma}} = -\Gamma_{\alpha\beta}^\gamma V^\beta U^\alpha \tag{2.84}$$

Note that the coordinates of a horizontal lift are not only linear in the vector being lifted, but in the point in TM (or T^*M) at which the lift is based: this is a consequence of the linearity of the connection, and will be important to guarantee linearity of parallel transport in Section 2.5.

It will be helpful to define the *frame bundle* $\phi M: FM \to M$, the subbundle of $\tau M^{\oplus m}$ whose fibre at $x \in M$ is given by all m-frames (i.e. ordered bases) of T_xM. Since FM is an open subspace of $TM^{\oplus m}$ it makes sense to use the product coordinates of the latter for the former: these are canonically defined in terms of a chart on M by pairs $(\lambda, \gamma) \in \{1, \ldots, m\}^2$ with the first referring to the copy of TM, i.e. if $y \in F_x M$ then $y_\lambda \coloneqq \mathrm{pr}_\lambda(y) \in T_x M$ has coordinates $y^\gamma_\lambda = y^{(\lambda,\gamma)}$. If M is Riemannian we may additionally consider the *orthonormal frame bundle* $oM \colon OM \to M$, i.e. the subbundle of ϕM with total space consisting of orthonormal frames.

We define the *fundamental horizontal vector fields* $\mathscr{H}_\lambda \in \Gamma\tau FM$, $\lambda = 1, \ldots, m$ [Hsu02, p.39] by the property $T_y \mathrm{pr}_\gamma(\mathscr{H}_\lambda(y)) = \hbar(y_\gamma) y_\lambda$, or in coordinates

$$\mathscr{H}^\gamma_\lambda(y) = y^\gamma, \quad \mathscr{H}^{(\mu,\gamma)}_\nu(y) = -\Gamma^\gamma_{\alpha\beta}(x) y^\beta_\mu y^\alpha_\nu \tag{2.85}$$

with $y \in F_x M$. If M is Riemannian and ∇ is metric these vector fields restrict to elements of $\Gamma\tau OM$.

To end this subsection, we briefly describe what it means for a smooth map of manifolds to preserve connections. Here we are following [É89].

Definition 2.29 (Affine map)**.** Let ${}^M\nabla$ (${}^N\nabla$) be a linear connection on the tangent bundle of the smooth manifold M (N). We will say that $f \in C^\infty(M, N)$ is *affine* if

$$\forall U, V \in \Gamma\tau M \quad T_x f({}^M\nabla_{U(x)} V) = {}^N\nabla_{T_x f(U(x))} Tf(V) \tag{2.86}$$

Note that the RHS is well defined, as $Tf(V)$ need only be defined on a curve tangent to $U(x)$ at x. The name is justified by the fact that the terminology coincides with the usual notion of affinity for smooth maps of Euclidean spaces. Other examples of affine maps are isometries of Riemannian manifolds (Riemannian isomorphisms that is — local isometries are not affine in general). In terms of the Hessians affinity of f reads

$$T^*_x f^{\otimes 2}({}^N\nabla^2 g)(f(x)) = {}^M\nabla(g \circ f)(x), \quad g \in C^\infty N \tag{2.87}$$

Symmetrising this identity yields the notion of *symmetric affinity*: this is equivalent to the requirement that f preserve parametrised geodesics, with full affinity holding if f additionally preserves torsion. The most useful characterisation of affinity, however, is the local one

$$({}^{M,N}\nabla^2 f)^k_{\alpha\beta}(x) \coloneqq \partial_{\alpha\beta} f^k(x) + {}^N\Gamma^k_{ij}(f(x)) \partial_\alpha f^i \partial_\beta f^j(x) - {}^M\Gamma^\gamma_{\alpha\beta} \partial_\gamma f^k(x) = 0 \tag{2.88}$$

which symmetrised yields the condition for symmetric affinity:

$$\partial_{\alpha\beta} f^k(x) = \tfrac{1}{2}({}^M\Gamma^\gamma_{\alpha\beta} + {}^M\Gamma^\gamma_{\beta\alpha}) \partial_\gamma f^k(x) - \tfrac{1}{2}({}^N\Gamma^k_{ij} + {}^N\Gamma^k_{ji})(f(x)) \partial_\alpha f^i \partial_\beta f^j(x) \tag{2.89}$$

Of course, there is no difference between the two if both connections are torsion-free. Symmetrised expressions will be of interest to us because of the symmetry of the bracket of a rough path; to lighten the notation we will add $(\alpha\beta)$ in an expression to mean that we are symmetrising it w.r.t. to the indices α, β. For instance, symmetric affinity can be written as $({}^{M,N}\nabla^2 f)^k_{\alpha\beta}(x) \overset{(\alpha\beta)}{=} 0$ or more succinctly still as $({}^{M,N}\nabla^2 f)^k_{(\alpha\beta)}(x) = 0$.

Example 2.30 (Affinity and fibre bundles)**.** It will be important to consider whether the projection map of a fibre bundle $\pi \colon E \to M$ is an affine map w.r.t. to chosen linear connections $\widetilde{\nabla}$ on E and ∇ on M. By (2.88), the condition of π of being affine reads in coordinates

$$\partial_{IJ}\pi^\gamma = \partial_K\pi^\gamma\widetilde{\Gamma}^K_{IJ} - \Gamma^\gamma_{\alpha\beta}\partial_I\tau M^\alpha\partial_J\tau M^\beta \tag{2.90}$$

where the $\widetilde{\Gamma}$'s denote the Christoffel symbols of $\widetilde{\nabla}$. Keeping in mind that $(\varphi \circ \pi \circ T\varphi^{-1})$ is the map $(x^1, \ldots, x^m, y^1, \ldots, y^n) \mapsto (x^1, \ldots, x^m)$ we compute

$$\partial_{IJ}\pi^\gamma = 0, \quad \partial_\beta\pi^\alpha = \delta^\alpha_\beta, \quad \partial_k\pi^\gamma = 0 \tag{2.91}$$

and (2.90) becomes

$$\widetilde{\Gamma}^\gamma_{\alpha\beta} = \Gamma^\gamma_{\alpha\beta}, \quad \widetilde{\Gamma}^\gamma_{\alpha\widetilde{\beta}} = 0, \quad \widetilde{\Gamma}^\gamma_{\widetilde{\alpha}\beta} = 0, \quad \widetilde{\Gamma}^\gamma_{\widetilde{\alpha}\widetilde{\beta}} = 0 \tag{2.92}$$

It is similarly checked that if π is a vector bundle the condition of the inclusions of the fibres (as flat spaces) of being affine reads

$$\widetilde{\Gamma}^K_{\widetilde{\alpha}\widetilde{\beta}} = 0 \tag{2.93}$$

and the condition of the inclusion of M (as the zero section) of being affine reads

$$\widetilde{\Gamma}^\gamma_{\alpha\beta}(0_x) = \Gamma^\gamma_{\alpha\beta}(x) \tag{2.94}$$

Replacing symmetric affinity with affinity results in the above coordinate expressions being symmetrised in the bottom two indices of each Christoffel symbol, e.g. the symmetrisation of $\widetilde{\Gamma}^\gamma_{\alpha\widetilde{\beta}} = 0$ is $\widetilde{\Gamma}^\gamma_{\alpha\widetilde{\beta}} + \widetilde{\Gamma}^\gamma_{\widetilde{\beta}\alpha} = 0$ (not $\widetilde{\Gamma}^\gamma_{\alpha\widetilde{\beta}} + \widetilde{\Gamma}^\gamma_{\widetilde{\alpha}\beta} = 0$).

2.2.2 Embedded manifolds

Although we have chosen to write this chapter mainly in the framework of intrinsic manifolds and local coordinates, we will relate our work to [CDL15], in which manifolds are embedded in Euclidean space. In this subsection we revisit some of the notions of the previous subsection, assuming that M is Riemannian and isometrically embedded in $\mathbb{R}^d$, $\imath\colon M \hookrightarrow \mathbb{R}^d$ (this is always possible by the Nash embedding theorem, for high enough d). This means that the connection on M will always be the Levi-Civita connection of the induced metric: this setting is less general than the one considered in the previous subsection, where non-metric connections with torsion were considered. In order to precisely distinguish between extrinsic and intrinsic formulae, we will always distinguish between objects on M (which will be treated using local coordinates, indexed by Greek letters $\alpha, \beta, \gamma, \ldots$) and their counterparts on $\mathcal{M} \coloneqq \imath(M)$ (treated using ambient coordinates, indexed by Latin letters $a, b, c, \ldots$). For instance $T_y\mathcal{M}$ and $T_y^\perp\mathcal{M}$ (the normal space) are subspaces of $T_y\mathbb{R}^d$, $T_y\mathbb{R}^d = T_y\mathcal{M} \oplus T_y^\perp\mathcal{M}$, and $T_x\imath\colon T_xM \to T_{\imath(x)}\mathcal{M}$ is an isomorphism.

The projection maps P and Q are defined ambiently as in Section 1.2. As for the Riemannian tubular neighbourhood projection, also defined therein, we make the following distinction: $\pi\colon A \to M$ takes values in the intrinsic manifold, and $\Pi \coloneqq \imath \circ \pi\colon A \to \mathcal{M}$ for some tubular neighbourhood A of $\mathcal{M}$. The important features of π and Π are

$$\pi \circ \imath = \mathbb{1}_M, \quad \Pi \circ \Pi = \Pi, \quad \imath = \Pi \circ \imath \tag{2.95}$$

We may express the Levi-Civita covariant derivative ∇ of M in ambient coordinates as follows:

$$T_x\imath\nabla_{U(x)}V = P(\imath(x))^d\nabla_{T_x\imath U(x)}(T\imath V), \quad \nabla_{U(x)}\omega = {}^d\nabla_{T_x\imath U(x)}(\omega \circ T\pi) \tag{2.96}$$

for $U(x) \in T_xM$, $V \in \Gamma\tau M$, $\omega \in \Gamma\tau^*M$, where we have extended $T\imath(V)$ to a vector field on A in an arbitrary smooth way, and ${}^d\nabla$ is the canonical covariant derivative on $\mathbb{R}^d$ given by taking directional derivatives. We emphasise once again that computations are carried out in ambient coordinates, i.e. ∂_c is differentiation in the c-th variable of $\mathbb{R}^d$, and sums go from 1 to d. The Levi-Civita Hessian is the given by [É89, (4.9)]

$$\nabla^2 f = T^*\imath^{\otimes 2}\, {}^d\nabla^2(f \circ \pi), \quad f \in C^\infty M \tag{2.97}$$

where ${}^d\nabla$ is the usual Hessian in $\mathbb{R}^d$.

We now express the Christoffel symbols according to some local chart in terms of $\imath, \pi$. (1.63) can be restated as

$$\partial_a P_b P^a(y) = \partial_{ab}\Pi P^a(y), \quad y \in \mathcal{M} \tag{2.98}$$

(and in particular the LHS is independent of the extension of P to a tubular neighbourhood). Another useful fact about the second derivatives of Π is the following identity, obtained by differentiating the second identity in (2.95) twice at $y \in M$:

$$\partial_{ce}\Pi P^c_a P^e_b(y) + P_c\partial_{ab}\Pi^c(y) = \partial_{ab}\Pi(y) \tag{2.99}$$

We will now show that, according to any chart on M

$$\Gamma^\gamma_{\alpha\beta}(x) = \partial_c\pi^\gamma(\imath(x))\partial_{\alpha\beta}\imath^c(x) \tag{2.100}$$

To prove this identity, let $\widetilde{\partial}_\alpha, \widetilde{\partial}_\beta$ be extensions to A of $T\imath\partial_\alpha = \partial_\alpha\imath, T\imath\partial_\beta = \partial_\beta\imath$ respectively. (2.96) implies

$$\Gamma^\gamma_{\alpha\beta}(x) = \partial_c\pi^\gamma({}^d\nabla_{\widetilde{\partial}_\alpha}\widetilde{\partial}_\beta)^c(\imath(x))$$

and

$$\begin{aligned}({}^d\nabla_{\widetilde{\partial}_\alpha}\widetilde{\partial}_\beta)(\imath(x)) &= \partial_e(\widetilde{\partial}_\beta\imath)(\imath(x))\partial_\alpha\imath^e(x) \\ &= \partial_e(\partial_\beta\imath \circ \pi)(\imath(x))\partial_\alpha\imath^e(x) \\ &= \partial_{\gamma\beta}\imath(x)\partial_e\pi^\gamma(\imath(x))\partial_\alpha\imath^e(x) \\ &= \partial_{\gamma\beta}\imath(x)\delta^\gamma_\alpha \\ &= \partial_{\alpha\beta}\imath(x)\end{aligned}$$

which concludes the argument.

2.3 Rough paths, rough integration and RDEs on manifolds

In this section M and N will denote smooth m- and n-dimensional manifolds respectively. Given a control ω on $[0,T]$ we say that a continuous path $X \colon [0,T] \to M$ lies in $\mathcal{C}^p_\omega([0,T],M)$ if for all $f \in C^\infty M$, $f(X) \in \mathcal{C}^p([0,T],\mathbb{R})$; on vector spaces, on vector spaces, this agrees with the ordinary definition by invariance of bounded p-variation under smooth maps. Equivalently, $X \in \mathcal{C}^p_\omega([0,T],M)$ if for all charts φ, $\varphi(X) \in \mathcal{C}^p([a,b],\mathbb{R}^m)$ whenever $X|_{[a,b]}$ is contained in the domain of φ.

Example 2.31 (Path in a fibre bundle)**.** Let $\pi\colon E \to M$ be a smooth fibre bundle with typical fibre R. A path $H \in \mathcal{C}^p_\omega([0,T],E)$ is characterised as follows: for every local trivialisation $\Phi\colon E_A \to A \times R$ and for every $0 \leq a \leq b \leq T$ s.t. $H|_{[a,b]} \subseteq E_A$, we have $\mathrm{pr}_1 \circ \Phi(H|_{[a,b]}) \in \mathcal{C}^p_\omega([a,b],A)$ and $\mathrm{pr}_2 \circ \Phi(H|_{[a,b]}) \in \mathcal{C}^p_\omega([a,b],R)$. Examples of such paths are given by smooth sections $\sigma \in \Gamma\pi$ evaluated at $X \in \mathcal{C}^p_\omega([a,b],M)$.

Young integration on a manifolds is simple to perform. Let $p \in [1,2)$, $X \in \mathcal{C}^p_\omega([0,T],M)$, $H \in \mathcal{C}^p_\omega([0,T],\mathcal{L}(\tau M,\mathbb{R}^e))$ in the fibre of X. We can then define the *Young integral*

$$\int_0^T H^k \mathrm{d}X := \lim_{|\pi|\to 0} \sum_{[s,t]\in\pi} H^k_{\gamma;s} X^\gamma_{st} \tag{2.101}$$

where $X^\gamma := \varphi^\gamma(X)$ and $H^k_\gamma := (H \circ (T_X\varphi)^{-1})^k_\gamma$ for any chart $\varphi\colon A \to \mathbb{R}^m$ with $X_{[s,t]} \subseteq A$. It is simple to check with a change of charts that Riemann summands do not depend on the chart up to $O(\omega(s,t)^{2/p}) \subseteq o(\omega(s,t))$, and the limit is well defined and converges, since it converges in every chart. Similarly, given a field of linear homomorphisms $V \in \Gamma\mathcal{L}(\tau M,\tau N)$ (here $\mathcal{L}(\tau M,\tau N)$ is the bundle $\mathcal{L}(TM,TN) \to N \times M$ with fibres $\mathcal{L}(TM,TN)_{y,x} := \mathcal{L}(T_xM,T_yN)$) we can define the Young differential equation by

$$\mathrm{d}Y = V(Y,X)\mathrm{d}X \quad \iff \quad \mathrm{d}Y^k = V^k_\gamma(Y,X)\mathrm{d}X^\gamma \tag{2.102}$$

where the coordinates are taken to be w.r.t. arbitrary charts on M and N. We will give definitions of rough paths, their controlled paths, rough integrals and RDEs in the same spirit, relying on the theory of Section 2.1.

Definition 2.32 (Rough path on a manifold)**.** Given an atlas $(\varphi\colon A_\varphi \to \mathbb{R}^m)_\varphi$ of M, an M-valued $[2,3) \ni$ *p-rough path* controlled by ω on $[0,T]$, $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],M)$, consists of a collection of rough paths ${}^\varphi\boldsymbol{X} = ({}^\varphi X, {}^\varphi\mathbb{X}) \in \mathscr{C}^\alpha_\omega([a_\varphi,b_\varphi],\mathbb{R}^m)$, where the intervals $[a_\varphi,b_\varphi]$ are chosen so that their union is $[0,T]$ and no two overlap in a single point, and with the property that for all charts $\varphi, \overline{\varphi}$ in the atlas s.t. $[a_\varphi,b_\varphi] \cap [a_{\overline{\varphi}},b_{\overline{\varphi}}] \neq \varnothing$

$$(\overline{\varphi} \circ \varphi^{-1})_* {}^\varphi\boldsymbol{X} = {}^{\overline{\varphi}}\boldsymbol{X} \in \mathscr{C}^p_\omega([a_\varphi,b_\varphi] \cap [a_{\overline{\varphi}},b_{\overline{\varphi}}],\mathbb{R}^m) \tag{2.103}$$

The *trace* of $\boldsymbol{X}$ is the path $t \mapsto X_t := \varphi^{-1}({}^\varphi X_t) \in M$ whenever $t \in [a_\varphi,b_\varphi]$ (independently of φ), $X \in \mathcal{C}^p([0,T],M)$. $\boldsymbol{X}$ is *geometric*, $\boldsymbol{X} \in \mathscr{G}^p_\omega([0,T],M)$, if ${}^\varphi\boldsymbol{X}$ is geometric for all φ.

It makes sense to allow the mappings $\varphi \mapsto {}^\varphi\boldsymbol{X}$ and $\varphi \mapsto [a_\varphi,b_\varphi]$ to be multi-valued, so that the same chart can be used multiple times (e.g. if the trace X goes back and forth between charts). To define a rough path on M with trace X we only need as many charts as it takes to cover $X_{[0,T]}$: once the compatibility condition (2.103) is satisfied for one such cover, for any further chart ψ, ${}^\psi\boldsymbol{X}$ is unambiguously defined thanks to [Proposition 2.13, 1.]: indeed, for charts $\varphi, \overline{\varphi}$ in the original covering

$$(\psi \circ \overline{\varphi}^{-1})_* {}^{\overline{\varphi}}\boldsymbol{X} = (\psi \circ \varphi^{-1} \circ \varphi \circ \overline{\varphi}^{-1})_* {}^{\overline{\varphi}}\boldsymbol{X} = (\psi \circ \varphi^{-1})_* (\varphi \circ \overline{\varphi}^{-1})_* {}^{\overline{\varphi}}\boldsymbol{X} = (\psi \circ \overline{\varphi}^{-1})_* {}^\varphi\boldsymbol{X} \tag{2.104}$$

The definitions of $\mathscr{G}^p_\omega([0,T],M)$ and of the geometrisation map $\mathscr{C}^p_\omega([0,T],M) \to \mathscr{G}^p_\omega([0,T],M)$ are well defined in charts, thanks to the fact that pushforward commutes with geometrisation [Proposition 2.13, 2.]. The bracket of a manifold-valued rough path is defined in charts, i.e.

$$ {}^{\varphi}[\boldsymbol{X}]_{st} \coloneqq [{}^{\varphi}\boldsymbol{X}]_{st}, \quad {}^{\overline{\varphi}}[\boldsymbol{X}]^{\overline{\alpha}\overline{\beta}}_{st} \approx \partial^{\overline{\alpha}}_{\alpha}\partial^{\overline{\beta}}_{\beta}({}^{\varphi}X_s){}^{\varphi}[\boldsymbol{X}]^{\alpha\beta}_{st} \tag{2.105} $$

for charts $\varphi, \overline{\varphi}$, thanks to Proposition 2.15. It should be noted that $[\boldsymbol{X}]$ is not an M-valued path; rather it can be viewed as a path valued in $TM \otimes TM$ with $[\boldsymbol{X}]_t$ in the fibre of X_t.

A very similar definition given at the end of [BL15], where the authors allow for more general than smooth (e.g. only Lipschitz) transition maps, focusing on geometric rough paths. Here we will not be concerned with finding the minimal working framework for defining rough paths on manifolds, rather we develop this theory in the familiar context of smooth manifolds (in certain cases endowed with extra structure), keeping in mind that many results can be generalised to the C^2 or Lipschitz setting.

Rough paths on manifolds can be pushed forward by smooth maps: if $f \in C^\infty(M,N)$ and $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],M)$, $f_*\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],N)$ is defined by, for a chart ψ on N

$$ {}^{\psi}(f_*\boldsymbol{X}) \coloneqq (\psi \circ f \circ \varphi^{-1})_* {}^{\varphi}\boldsymbol{X} \tag{2.106} $$

independently of the chart φ on M. Following [É89] we define an M-valued *semimartingale* to be a stochastic process defined on some setup $(\Omega, \mathcal{F}, P)$ satisfying the usual conditions with the property that $f(X)$ is a real-valued semimartingale for all $f \in C^\infty M$, and denote the set of those defined on the interval $[0,T]$ as $\mathcal{S}(\Omega,[0,T];M)$. If M is a finite-dimensional $\mathbb{R}$-vector space the two notions of $\mathcal{S}(\Omega,[0,T];V)$ coincide thanks to Itô's formula.

Example 2.33 (Itô and Stratonovich rough paths on M)**.** Let $X \in \mathcal{S}(\Omega,[0,T];M)$. We can define its Stratonovich and Itô lifts respectively by lifting ${}^{\varphi}X$ to ${}^{\varphi}\mathbb{X}$ and ${}^{\varphi}\widehat{\mathbb{X}}$ defined in (2.53) on all stochastic intervals $[a,b]$ s.t. $X_{[a,b]} \subseteq A_\varphi$ (the domain of φ) for $t \in [a,b]$. Crucially, these a.s. define M-valued stochastic rough paths thanks to Proposition 2.25, and just as in the linear case we have $\widehat{{}_{g}\boldsymbol{X}} = \boldsymbol{X}$. These definitions, together with Remark 2.24 allow us to restrict all the rough path theory that follows to the semimartingale context and recover the theory of stochastic calculus on manifolds (Stratonovich and Itô integrals, SDEs, etc.) as presented in [É89].

We proceed with the definition of controlled paths, specifically in the case of integrands. While for the definition of rough path we used pushforward to force compatibility, for controlled paths we require it through pullbacks.

Definition 2.34 (Controlled integrand)**.** Let $X \in \mathcal{C}^p_\omega([0,T],M)$. We define an $\mathbb{R}^e$-valued *X-controlled integrand* $\boldsymbol{H} = ({}^{\varphi}H, {}^{\varphi}H') \in \mathscr{D}_X(\mathcal{L}(\tau M, \mathbb{R}^e))$ to be a collection ${}^{\varphi}\boldsymbol{H} \in \mathscr{D}_{{}^{\varphi}X|_{[a_\varphi,b_\varphi]}}(\mathbb{R}^{m\times e})$ with $\varphi, a_\varphi, b_\varphi$ as in Definition 2.32 and

$$ (\varphi \circ \overline{\varphi}^{-1})^* {}^{\varphi}\boldsymbol{H} = {}^{\overline{\varphi}}\boldsymbol{H}, \quad \text{i.e. } H_{\overline{\gamma}} = H_\gamma \partial^\gamma_{\overline{\gamma}},\ H'_{\overline{\alpha}\overline{\beta}} = H'_{\alpha\beta}\partial^\alpha_{\overline{\alpha}}\partial^\beta_{\overline{\beta}} + H_\gamma \partial^\gamma_{\overline{\alpha}\overline{\beta}} \tag{2.107} $$

The *trace* of $\boldsymbol{H}$ is the path $H \coloneqq {}^{\varphi}H \circ T_X\varphi$, which is valued in the fibre of X of the bundle $\mathcal{L}(\tau M, \mathbb{R}^e) = (\tau^* M)^e$.

As for $\mathbb{R}^d$-valued controlled paths, the most immediate example is given by the evaluation of a one-form $\sigma \in \Gamma\mathcal{L}(\tau M, \mathbb{R}^e)$: in coordinates this amounts to $\boldsymbol{\sigma}(X) = (\sigma^k_\gamma(X), \partial_\alpha \sigma^k_\beta(X))$.

A smooth map of manifolds $f \in C^\infty(M, N)$ defines the *pullback* of controlled integrands: if $X \in \mathcal{C}^p_\omega([0,T], M)$ and $\boldsymbol{H} \in \mathscr{D}_{f(X)}(\mathcal{L}(\tau N, \mathbb{R}^e))$

$$^\varphi(f^*\boldsymbol{H}) \coloneqq (\psi \circ f \circ \varphi^{-1})^{*\psi}\boldsymbol{H} \tag{2.108}$$

is defined independently of the chart φ on M by Proposition 2.12 and is checked to be an element of $\mathscr{D}_X(\mathcal{L}(\tau M, \mathbb{R}^e))$.

Remark 2.35 (General controlled paths). More in general, let $\pi \colon E \to M$ be a smooth fibre bundle with typical fibre the n-dimensional manifold R. We may define π-valued X*-controlled path* as a pair $\boldsymbol{H} = (H, H')$ with $H \in \mathcal{C}^p_\omega([0,T], E)$ in the fibre of X, $H' \in \mathcal{C}^p_\omega([0,T], \mathcal{L}(TM, TE))$ in the fibre of (X, H), with H'_t a section of $T_{H_t}\pi$ for all $t \in [0,T]$, i.e. $T_{H_t}\pi \circ H'_t = \mathbb{1}_{TM}$. Moreover, we require that for all charts $\varphi \colon A \to \mathbb{R}^m$, $\phi \colon B \to \mathbb{R}^n$ with $A \subseteq M$, $B \subseteq R$ open and local trivialisations $\Phi \colon E_A \to A \times R$, calling, for all $[a,b] \subseteq [0,T]$ s.t. $X_{[a,b]} \subseteq A$, $\Phi(H_{[a,b]}) \subseteq A \times B$, and $s \in [a,b]$

$$\begin{aligned} &\mathscr{F} \coloneqq (\varphi, \phi, \Phi), \quad {}^\varphi X_s \coloneqq \varphi(X_s) \in \mathbb{R}^m, \quad ({}^\varphi X_s, {}^\mathscr{F} H_s) \coloneqq (\varphi \times \phi) \circ \Phi(H_s) \in \mathbb{R}^m \times \mathbb{R}^n, \\ &(\mathbb{1}_{\mathbb{R}^m}, {}^\mathscr{F} H'_s) \coloneqq T_{\Phi(H_s)}(\varphi \times \phi) \circ T_{H_s}\Phi \circ H'_s \circ (T_{X_s}\varphi)^{-1} \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^m \oplus \mathbb{R}^n) \end{aligned} \tag{2.109}$$

(the assumptions on $\boldsymbol{H}$ imply that the above coordinate expressions have this form) we have

$$^\mathscr{F}\boldsymbol{H} \in \mathscr{D}^p_{^\varphi X}([a,b], \mathbb{R}^n), \quad \text{i.e. } {}^\mathscr{F}H_{st} - {}^\mathscr{F}H'_s {}^\varphi X_{st} \in O(\omega(s,t)^{2/p}) \tag{2.110}$$

This requirement can be checked to not depend on the choice of $\mathscr{F}$. Moreover, this definition can be seen to restrict to Definition 2.34 for the choice $\pi = \mathcal{L}(\tau^* M, \mathbb{R}^e)$, for which the charts ϕ and the trivialisation Φ can be chosen canonically in terms of φ. In practice, however, there are no applications of this more general notion of controlled path within our scope, and we will rely solely on the definition of controlled integrand.

We now give the definition of rough integral on manifolds. Note that this definition already exists for the Stratonovich and Itô integral of semimartingale [É89, p.93, p.109] (already covered in this thesis in Section 1.1), and thanks to Remark 2.24 the notion below extends these when applied to Example 2.33. We will avoid deriving the integral using Schwartz-Meyer theory, which is cumbersome to formulate for rough paths, and define the rough integral directly in coordinates, without appeals to second-order forms. It is easily checked that the naïve definition of the rough integral in charts $\int \boldsymbol{H} \mathrm{d}\boldsymbol{X} \coloneqq \int \boldsymbol{H}_\gamma \mathrm{d}\boldsymbol{X}^\gamma$ fails to be coordinate-invariant due to the bracket correction in the change of variable formula; to come up with an intrinsic notion we must rely on a connection on τM.

Definition 2.36. Assume τM is endowed with a linear connection ∇ and let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T], M)$, $\boldsymbol{H} \in \mathscr{D}_X(\mathcal{L}(\tau M, \mathbb{R}^e))$. We define the *rough integral*

$$\int_0^\cdot \boldsymbol{H} \mathrm{d}_\nabla \boldsymbol{X} \coloneqq \sum_{[s_\varphi, t_\varphi]} \int_{s_\varphi}^{t_\varphi} \boldsymbol{H}_\gamma \mathrm{d}\boldsymbol{X}^\gamma + \frac{1}{2}\int_{s_\varphi}^{t_\varphi} H_\gamma \Gamma^\gamma_{\alpha\beta}(X) \mathrm{d}[\boldsymbol{X}]^{\alpha\beta} \in \mathcal{C}^p_\omega([0,T], \mathbb{R}^e) \tag{2.111}$$

where we are summing over a finite partition of $[0, \cdot]$ whose intervals $[s_\varphi, t_\varphi]$ are indexed by charts φ with the property that each $[s_\varphi, t_\varphi]$ is contained in the domain of φ, and the coordinates in the integrals are taken w.r.t.

to these charts. Moreover, this path can be augmented with the unique second order part $\approx H^i_{\alpha;s}H^j_{\beta;s}\mathbb{X}^{\alpha\beta}_{st}$ where the coordinates α, β are taken w.r.t. to any chart that contains $X_{[s,t]}$, thus defining an element of $\mathscr{C}^p_\omega([0,T],\mathbb{R}^e)$.

We will often write d_M for d_∇, especially when more than one manifold is involved. Note how we have defined the rough integral directly as a rough path, without passing through the notion of controlled path. To do so would have required to define what it means for an $\mathbb{R}^e$-valued path to be controlled by an M-valued one: this is possible by applying the generalised definition of Remark 2.35 to the trivial vector bundle over M with fibre $\mathbb{R}^e$. However, it is much simpler to bypass this step, and we shall do so for solutions of RDEs as well.

Theorem 2.37. *Definition 2.36 is sound: it depends neither on the partition or on the charts chosen for each interval.*

Proof. We begin by dealing with the trace. The first assertion immediately follows from the second by comparing two integrals taken w.r.t. two different partitions with that taken w.r.t. their common refinement (identity holds by additivity of the rough integral on consecutive time intervals). We then consider two charts $\varphi, \overline{\varphi}$, the latter of whose indices we denote using overlines. Then by Corollary 2.17 we have

$$\int \boldsymbol{H}_{\overline{\gamma}}\mathrm{d}\boldsymbol{X}^{\overline{\gamma}} = \int \boldsymbol{H}_\gamma \mathrm{d}\boldsymbol{X}^\gamma + \frac{1}{2}\int H_{\overline{\gamma}}\partial^{\overline{\gamma}}_{\alpha\beta}(X)\mathrm{d}[\boldsymbol{X}]^{\alpha\beta} \tag{2.112}$$

Moreover, using Proposition 2.15 and (2.59) we have

$$\begin{aligned}\int H_{\overline{\gamma}}\Gamma^{\overline{\gamma}}_{\overline{\alpha}\overline{\beta}}(X)\mathrm{d}[\boldsymbol{X}]^{\overline{\alpha}\overline{\beta}} &= \int (H_\gamma \partial^\gamma_{\overline{\gamma}}(X)) \cdot (\partial^{\overline{\gamma}}_\lambda \partial^\mu_{\overline{\alpha}}\partial^\nu_{\overline{\beta}}\Gamma^\lambda_{\mu\nu} + \partial^\lambda_{\overline{\alpha}\overline{\beta}}\partial^{\overline{\gamma}}_\lambda)(X) \cdot (\partial^{\overline{\alpha}}_\alpha \partial^{\overline{\beta}}_\beta(X)\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}) \\ &= \int (H_\gamma\Gamma^\gamma_{\alpha\beta}(X) + H_\gamma \partial^\gamma_{\overline{\alpha}\overline{\beta}}\partial^{\overline{\alpha}}_\alpha\partial^{\overline{\beta}}_\beta(X))\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}\end{aligned} \tag{2.113}$$

Putting these two identities together, we have

$$\begin{aligned}\int \boldsymbol{H}_\gamma \mathrm{d}\boldsymbol{X}^\gamma + \int H_\gamma \Gamma^\gamma_{\alpha\beta}(X)\mathrm{d}[\boldsymbol{X}]^{\alpha\beta} &= \int \boldsymbol{H}_\gamma \mathrm{d}\boldsymbol{X}^\gamma + \int H_\gamma\Gamma^\gamma_{\alpha\beta}(X)\mathrm{d}[\boldsymbol{X}]^{\alpha\beta} \\ &\quad + \int H_\gamma(\partial^\gamma_{\overline{\gamma}}\partial^{\overline{\gamma}}_{\alpha\beta} + \partial^\gamma_{\overline{\alpha}\overline{\beta}}\partial^{\overline{\alpha}}_\alpha\partial^{\overline{\beta}}_\beta)(X)\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}\end{aligned} \tag{2.114}$$

But

$$\partial^\gamma_{\overline{\gamma}}\partial^{\overline{\gamma}}_{\alpha\beta} + \partial^\gamma_{\overline{\gamma}}\partial^\delta_{\overline{\alpha}\overline{\beta}}\partial^{\overline{\gamma}}_\delta\partial^{\overline{\alpha}}_\alpha\partial^{\overline{\beta}}_\beta = \partial_{\alpha\beta}((\overline{\varphi}^\gamma \circ \varphi^{-1}) \circ (\varphi \circ \overline{\varphi}^{-1})) = \partial_{\alpha\beta}\varphi^\gamma = 0 \tag{2.115}$$

which yields the desired identity. As for the second order part, we have

$$H^i_{\overline{\alpha};s}H^j_{\overline{\beta};s}\mathbb{X}^{\overline{\alpha}\overline{\beta}}_{st} \approx (H^i_{\gamma;s}\partial^\gamma_{\overline{\alpha}}(X_s)H^j_{\delta;s}\partial^\delta_{\overline{\beta}}(X_s)) \cdot (\partial^{\overline{\alpha}}_\alpha\partial^{\overline{\beta}}_\beta(X_s)\mathbb{X}^{\alpha\beta}_{st}) = H^i_{\alpha;s}H^j_{\beta;s}\mathbb{X}^{\alpha\beta}_{st} \tag{2.116}$$

This concludes the proof. ■

We proceed by proving a few properties of the rough integral on manifolds. The first of these (cf. [É89, p.109] in the case of the Itô integral) tells us that the definition of the rough integral is indeed the one that yields the correct change of variable formula, i.e. in which the second derivative is replaced with the Hessian. Note that Proposition 2.15 allows us to define the integral of an element of $K \in \mathcal{C}^p([0,T],\mathcal{L}(\tau M^{\otimes 2},\mathbb{R}^e))$

above X, against $[\boldsymbol{X}]$ in coordinates as

$$\int K\mathrm{d}[\boldsymbol{X}] \coloneqq \int K_{\alpha\beta}\mathrm{d}[\boldsymbol{X}]^{\alpha\beta} \tag{2.117}$$

This is the analogue of [É89, Definition 3.9] in the rough path context.

Proposition 2.38 (Properties of the rough integral on manifolds)**.**

Exact integrands. *For* $f \in C^\infty(M, \mathbb{R}^e)$ $\quad f(X) - f(X_0) = \int_0^\cdot \boldsymbol{df}(X)\mathrm{d}_\nabla \boldsymbol{X} + \frac{1}{2}\int_0^\cdot \nabla^2 f(X)\mathrm{d}[\boldsymbol{X}]$*;*

Geometric integrators. $\int \boldsymbol{H}\mathrm{d}_\nabla \boldsymbol{X}$ *does not depend on the torsion of* ∇*, and if* $\boldsymbol{X}$ *is geometric it is altogether independent of* ∇*;*

Pushforward-pullback behaviour. *For* $\boldsymbol{X} \in \mathscr{C}^p(M)$*,* $f \in C^\infty(M, N)$*,* $\boldsymbol{H} \in \mathscr{D}^p_{f(X)}(\mathcal{L}(\tau N, \mathbb{R}^e))$*,* ${}^M\nabla$*,* ${}^N\nabla$ *connections on* N *and* M *respectively*

$$\int \boldsymbol{H}\mathrm{d}_N f_*\boldsymbol{X} - \int f^*\boldsymbol{H}\mathrm{d}_M\boldsymbol{X} = \frac{1}{2}\int H_k({}^{M,N}\nabla^2 f)^k_{\alpha\beta}(X)\mathrm{d}[\boldsymbol{X}]^{\alpha\beta} \tag{2.118}$$

where ${}^{M,N}\nabla^2 f$ *is defined in* (2.88)*. In particular the RHS above vanishes whenever* $\boldsymbol{X}$ *is geometric or* f *is symmetrically affine.*

Proof. It suffices to show all three statements in a single chart. The first follows immediately from (2.60) and Theorem 2.14. The second is evident from the fact that the bracket of a geometric rough path vanishes, and that even when it does not it is a symmetric tensor. The third is handled by using Corollary 2.17. ■

Example 2.39 (Tensorial expansion of the rough integral)**.** The Taylor-type approximation

$$\int_s^t \boldsymbol{H}\mathrm{d}_\nabla \boldsymbol{X} \approx H_{\gamma;s}X^\gamma_{st} + H'_{\alpha\beta;s}\mathbb{X}^{\alpha\beta}_{st} + \tfrac{1}{2}H_{\gamma;s}\Gamma^\gamma_{\alpha\beta}(X_s)[\boldsymbol{X}]^{\alpha\beta}_{st} \tag{2.119}$$

is coordinate-invariant up to $o(\omega(s,t))$, but the single terms in it are not. We may rewrite it as

$$\int_s^t \boldsymbol{H}\mathrm{d}_\nabla \boldsymbol{X} \approx H_{\gamma;s}(X^\gamma_{st} + \tfrac{1}{2}\Gamma^\gamma_{\alpha\beta}(X_s)X^\alpha_{st}X^\beta_{st}) + ({}^\odot\nabla \boldsymbol{H})_{\alpha\beta;s}\mathbb{X}^{\alpha\beta}_{st} \tag{2.120}$$

where for a connection ∇

$$(\nabla \boldsymbol{H})^k_{\alpha\beta} \coloneqq H'^k_{\alpha\beta} - H^k_\gamma\Gamma^\gamma_{\alpha\beta}(X) \tag{2.121}$$

(and therefore $({}^\odot\nabla \boldsymbol{H})_{\alpha\beta} = H'^k_{\alpha\beta} - H^k_\gamma\Gamma^\gamma_{(\alpha\beta)}(X)$, where we are symmetrising the bottom two indices). $\nabla\boldsymbol{H}$ is defined by analogy with (2.58), i.e. if $\boldsymbol{H} = \boldsymbol{\omega}(X)$ for a one-form $\omega \in \Gamma\mathcal{L}(\tau M, \mathbb{R}^e)$, then $\nabla\boldsymbol{H} = \nabla\omega(X)$. Now all four individual terms $H_{\gamma;s}$, $({}^\odot\nabla \boldsymbol{H})_{\alpha\beta;s}$ (and even $(\nabla\boldsymbol{H})_{\alpha\beta;s}$), $X_{st} + \frac{1}{2}\Gamma^\gamma_{\alpha\beta}(X_s)X^\alpha X^\beta$ and $\mathbb{X}^{\alpha\beta}_{st}$ transform as tensors, in the latter two cases up to an $o(\omega(s,t))$. Note that omitting the symmetrisation in ${}^\odot\nabla\boldsymbol{H}$ will result in an incorrect expansion, since the accordingly modified expansion (2.120) will not be almost additive, due to the extra term involving the evaluation of the torsion against $X_{su} \wedge X_{ut}$:

$$\Gamma^\gamma_{\alpha\beta}(X_s)(X^\alpha_{su}X^\beta_{ut} - X^\beta_{su}X^\alpha_{ut}) \tag{2.122}$$

If $\boldsymbol{X}$ is geometric, the symmetrisation can be omitted by writing the expansion as $H_{\gamma;s}(X^\gamma_{st} + \frac{1}{2}\Gamma^\gamma_{\alpha\beta}(X_s)\mathbb{X}^{\alpha\beta}_{st}) + (\nabla \boldsymbol{H})_{\alpha\beta;s}\mathbb{X}^{\alpha\beta}_{st}$ (in this case, of course, the connection is purely auxiliary). All of this seems to suggest that manifolds endowed with non-torsionfree connections are not the correct environment for non-geometric rough integration.

Example 2.40 (Itô-Stratonovich correction on manifolds)**.** We compare the integral of $\boldsymbol{H} \in \mathscr{D}_X(\mathcal{L}(\tau M, \mathbb{R}^e))$ against $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T], M)$ and its geometrisation: by Example 2.8 we have, at the path level

$$\int \boldsymbol{H} \circ \mathrm{d}\boldsymbol{X} - \int \boldsymbol{H}\mathrm{d}_\nabla \boldsymbol{X} = \frac{1}{2}\int \nabla \boldsymbol{H}\mathrm{d}[\boldsymbol{X}] \tag{2.123}$$

This identity is the analogue of [Dri04, Theorem 5.17] in the context of rough paths. Note how our ability of writing the above Itô-Stratonovich correction formula in terms of an integral against $\mathrm{d}[\boldsymbol{X}]$ is due to the fact that we are integrating controlled paths. In our context of rough integration this is a necessity, but in stochastic calculus on manifolds one can integrate a much larger class of $\mathcal{L}(TM, \mathbb{R}^e)$-valued processes above X, and for these the correction formula will involve the quadratic covariations of H and X.

Example 2.41 (Riemannian rough integral)**.** We may define a τM-valued X-controlled path $\boldsymbol{P} \in \mathscr{D}_X(\tau M)$ by a collection of ${}^\varphi\boldsymbol{P} \in \mathscr{D}_{\varphi X}(\mathbb{R}^m)$ satisfying the change of coordinates

$$P^{\overline{\gamma}} = \partial^{\overline{\gamma}}_\gamma P^\gamma,\ P'^{\overline{\beta}}_{\overline{\alpha}} = \partial^{\overline{\beta}}_\beta P'^{\beta}_\alpha \partial^{\alpha}_{\overline{\alpha}} + \partial^{\overline{\beta}}_{\gamma\alpha} P^\gamma \partial^\alpha_{\overline{\alpha}} \tag{2.124}$$

where we are writing the first index as a superscript since we view $P \in T_X M$ (this definition would fall under the more general Remark 2.35). Vector fields evaluated at X (along with their coordinate partial derivatives) are obvious examples. Now, if g is a Riemannian metric on M, it is natural to define, for $\boldsymbol{P} \in \mathscr{D}_X(\tau M)$ the controlled integrand

$$\boldsymbol{P}^\flat = (P^\flat_\gamma, P^{\flat\prime}_{\alpha\beta}) \coloneqq (g_{\gamma\delta}(X)P^\delta, g_{\beta\delta,\alpha}(X)P^\delta + g_{\beta\delta}(X)P'^{\delta}_\alpha) \in \mathscr{D}_X(\tau^* M) \tag{2.125}$$

and if ∇ is a connection on M (which need not be metric) we may integrate $\boldsymbol{P}$ thanks to the Riemannian metric:

$$\int g(\boldsymbol{P}, \mathrm{d}_\nabla \boldsymbol{X}) \coloneqq \int \boldsymbol{P}^\flat \mathrm{d}_\nabla \boldsymbol{X} \tag{2.126}$$

These definitions can be extended to the multivariate case, i.e. when we replace τM with $\tau M^{\oplus e}$.

We will now define RDEs driven by manifold-valued rough paths and with solutions valued in a second manifold. The semimartingale-analogue of the definition below can be found in [É90, p.428]. A heuristic derivation of the coordinate expression can be derived by writing the "intrinsic differential" on a manifold P endowed with a connection as $\mathrm{d}_P\boldsymbol{Z}^k \coloneqq \mathrm{d}\boldsymbol{Z}^c + \frac{1}{2}\Gamma^c_{ab}(Z)\mathrm{d}[\boldsymbol{Z}]^{ab}$ and writing the identity $\mathrm{d}_N\boldsymbol{Y}^k = F^k_\gamma(Y,X)\mathrm{d}_M\boldsymbol{X}^\gamma$:

$$\mathrm{d}\boldsymbol{Y}^k + {}^N\Gamma^k_{ij}(Y)\mathrm{d}[\boldsymbol{Y}]^{ij} = F^k_\gamma(Y,X)(\mathrm{d}\boldsymbol{X}^\gamma + {}^M\Gamma^\gamma_{\alpha\beta}(X)\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}) \tag{2.127}$$

Swapping in $\mathrm{d}[\boldsymbol{Y}]^{ij} = F^i_\alpha F^j_\beta(Y,X)\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}$ (which can be done since the terms of regularity $p/2$ do not contribute to the bracket) then yields (2.129) below.

Definition 2.42 (RDEs on manifolds)**.** Let $F \in \Gamma\mathcal{L}(\tau M, \tau N)$, $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T], M)$ and $y_0 \in N$. We define a *solution* to the RDE

$$\mathrm{d}_N \boldsymbol{Y} = F(Y,X)\mathrm{d}_M \boldsymbol{X}, \quad Y_0 = y_0 \tag{2.128}$$

to mean a N-valued rough path $\boldsymbol{Y}$ with $Y_0 = y_0$ s.t. for any two charts on M and N, and on any interval restricted to which X and Y are contained in the respective domains, the following RDE (in the sense of Definition 2.18)

$$\mathrm{d}\boldsymbol{Y}^k = F^k_\gamma(Y,X)\mathrm{d}\boldsymbol{X}^\gamma + \tfrac{1}{2}(F^k_\gamma(Y,X)^M\Gamma^\gamma_{\alpha\beta}(X) - {}^N\Gamma^k_{ij}(Y)F^i_\alpha F^j_\beta(Y,X))\mathrm{d}[\boldsymbol{X}]^{\alpha\beta} \tag{2.129}$$

where coordinates are taken (invariantly) w.r.t. the two charts, holds. Note that this implies $\mathbb{Y}^{ij}_{st} \approx F^i_\alpha F^j_\beta(Y_s, X_s)\mathbb{X}^{\alpha\beta}_{st}$.

The coordinate-independence check is analogous to that performed in Theorem 2.37 and is therefore omitted. Analogously to the vector space-valued case, notions of global and local solutions can be defined and distinguished, and the smoothness of F ensures local existence and uniqueness of the solution. These results can, as usual, be proved via "patching" and applying Theorem 2.21. Also note that, just as for the rough integral, only the connection modulo its torsion is relevant, and is not relevant at all when $\boldsymbol{X}$ is geometric, in which case the usual coordinate expression $\mathrm{d}\boldsymbol{Y}^k = F^k_\gamma(Y,X)\mathrm{d}\boldsymbol{X}^\gamma$ holds: for this reason we shall omit the M and N subscripts to the differentials in this case.

Remark 2.43 (Local existence and uniqueness). The local existence and uniqueness theorem Theorem 2.21 extends verbatim to the case of RDEs on manifolds Definition 2.42 (where compacts are determined by the manifold topology), by an embedding argument (the only think to keep in mind is that the embedding must be proper, so that explosion in M is synonymous with explosion in the ambient $\mathbb{R}^d$). A similar (though stronger) theorem for semimartingales can be found in [É90, Theorem 4].

The next two examples only deals with manifold-valued semimartingales SDEs, but can be viewed in context of RDEs thanks to Example 2.33.

Example 2.44 (Local martingales)**.** Recall that if M is endowed with a connection ∇, an M-valued local martingale X is an M-valued semimartingale s.t. for all $f \in C^\infty M$, $f(X) - \frac{1}{2}\int \nabla^2 f(X)\mathrm{d}[X]$ is a local martingale in $\mathbb{R}^d$, or in local coordinates

$$\mathrm{d}_\nabla X = \mathrm{d}X^\gamma + \tfrac{1}{2}\Gamma^\gamma_{\alpha\beta}(X)\mathrm{d}[X]^{\alpha\beta} \tag{2.130}$$

is the differential of a local martingale in $\mathbb{R}^d$. As observed in [É90], it is easy to see that the martingale-preserving property of Itô SDEs carries over to the manifold setting: if X is an M-valued local martingale and Y is the solution to (2.129) (where $\boldsymbol{X}$ is given by the Itô lift of X) is an N-valued local martingale.

Example 2.45 (Itô diffusions)**.** Let $M = \mathbb{R}^{1+m}$, $X_t = (t, B_t)$ where B is an m-dimensional Brownian motion. Then F can be viewed as a collection of $1+m$ vector fields $F_0, F_1, \ldots, F_m \in \Gamma\tau M$, which we take to not depend on X. It is well known that the solution to the Stratonovich SDE $\mathrm{d}Y = F_0(Y)\mathrm{d}t + F_\gamma(Y)\circ\mathrm{d}B^\gamma_t$ is a diffusion with generator $\mathscr{L} := F_0 + \frac{1}{2}\sum_{\gamma=1}^m F^2_\gamma$ (where F^2_γ denotes the differential operator

$F_\gamma^2 f(x) \coloneqq F_\gamma(y \mapsto F_\gamma f(y)))$: this means that for all $f \in C^\infty M$

$$f(Y) - \int_0^\cdot \mathscr{L} f(Y)\mathrm{d}t \tag{2.131}$$

is a local martingale. Itô diffusions on manifolds can also be considered: the solution to $\mathrm{d}_\nabla Y = F_0(Y)\mathrm{d}t + F(Y)\mathrm{d}B$ (intended in the same intrinsic sense as Definition 2.42) is a diffusion with generator $\mathscr{L} \coloneqq F_0 + \frac{1}{2}\sum_{\gamma=1}^m \langle \nabla^2 \cdot, F_\gamma \otimes F_\gamma \rangle$. We may verify this claim in local coordinates:

$$\begin{aligned} \mathrm{d}f(Y) &= \partial_k f(Y)\mathrm{d}Y^k + \partial^2_{ij} f(Y)\mathrm{d}[Y]^{ij} \\ &= \partial_k f(Y)\big(F^k_\gamma(Y)\mathrm{d}B^\gamma_t + (F^k_0 - \tfrac{1}{2}\Gamma^k_{ij}\textstyle\sum_\gamma F^i_\gamma F^j_\gamma)(Y)\mathrm{d}t\big) + \tfrac{1}{2}\textstyle\sum_\gamma \partial_{ij} f F^i_\gamma F^j_\gamma(Y)\mathrm{d}t \\ &= \partial_k f(Y) F^k_\gamma(Y)\mathrm{d}B^\gamma_t + (F^k_0 + \tfrac{1}{2}\textstyle\sum_\gamma (\partial_{ij} f - \partial_k f \Gamma^k_{ij}) F^i_\gamma F^j_\gamma)(Y)\mathrm{d}t \end{aligned} \tag{2.132}$$

from which the conclusion follows using (2.60). This example carries over to the case in which F depends on t, in which case $\mathscr{L}$ will also be time-dependent.

2.4 The extrinsic viewpoint

In this chapter we have mostly chosen to adopt a local perspective on differential geometry. This choice is motivated by the fact that the most natural definition of rough and controlled paths involve charts, and that therefore the resulting theory would most easily be handled using local coordinates. While we shall continue with this approach in the next section, one of our objectives is to compare our results with those of the other main paper on this topic, [CDL15], in which manifolds are handled using an extrinsic approach. To do this, we will revisit the main definitions of the previous section assuming that all manifolds are smoothly embedded in Euclidean space, and using ambient Euclidean coordinates to express our formulae. We will show that our results do indeed extend those of [CDL15], in which only geometric rough paths and one-form integrands are considered. One of the most interesting aspect of this section, however, is that for things to generalise in the correct manner to the case of general controlled integrands, additional non-degeneracy hypotheses will have to be placed on the class of integrands; these are always satisfied if X is truly rough.

We will use the notation introduced and referenced in Subsection 2.2.2 for embedded manifolds endowed with the Levi-Civita connection of the induced Riemannian metric. We begin by stating when an $\mathbb{R}^d$-valued rough path may be considered to lie on $\mathcal{M}$: this will entail not only the obvious requirement on the trace, but also a condition on the second order part.

Definition 2.46 (Constrained rough path)**.** Let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$. We will say $\boldsymbol{X}$ is *constrained to* $\mathcal{M}$ if $\Pi_*\boldsymbol{X} = \boldsymbol{X}$, and denote the set of $\mathcal{M}$-constrained p-rough paths controlled by ω with $\mathscr{C}^p_\omega([0,T],\mathcal{M})$ and its subset of geometric ones with $\mathscr{G}^p_\omega([0,T],\mathcal{M})$.

$\imath_*$ defines bijections $\mathscr{C}^p_\omega([0,T],M) \to \mathscr{C}^p_\omega([0,T],\mathcal{M})$ with inverse π_*, but we still choose to distinguish the two notions, since local coordinates are used in the former case, while $\mathscr{C}^p_\omega([0,T],\mathcal{M}) \subseteq \mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$. An equivalent way of stating Definition 2.46 for an $\mathbb{R}^d$-valued rough path $\boldsymbol{X}$ is as follows: $X \in \mathcal{C}^p_\omega([0,T],\mathcal{M})$, or equivalently by Taylor's formula

$$X^c_{st} = \Pi^c(X)_{st} \approx P^c_d(X_s)X^d_{st} + \tfrac{1}{2}\partial_{ab}\Pi^c(X_s)X^a_{st}X^b_{st} \tag{2.133}$$

and

$$\mathbb{X}^{cd}_{st} \approx P^c_a P^d_b(X_s)\mathbb{X}^{ab}_{st} \Leftrightarrow Q^c_a(X_s)\mathbb{X}^{ab}_{st} \approx 0 \Leftrightarrow Q^d_b(X_s)\mathbb{X}^{ab}_{st} \approx 0 \tag{2.134}$$

Moreover, these imply

$$[\boldsymbol{X}]^{cd}_{st} \approx P^c_a P^d_b(X_s)[\boldsymbol{X}]^{ab}_{st} \tag{2.135}$$

We note straight away that this definition extends the characterisation [CDL15, Corollary 3.32 (2)] to the non-geometric setting; the characterisation [CDL15, Corlollary 3.32 (1)] (which states that $Q_a I_b(X_s)(\mathbb{X}^{ab}_{st} - \mathbb{X}^{ba}_{st}) \approx 0$) does not hold, however, for non-geometric rough paths, as the symmetric part of their second order part is not determined by their trace (a counterexample is easily found by taking $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathcal{M})$ and then adding to $\mathbb{X}$ any path $Z \in \mathcal{C}^{p/2}_\omega([0,T],(\mathbb{R}^d)^{\odot 2})$ s.t. $Q^d_b(X_s)Z^{ab}_{st} \not\approx 0$).

Instead of defining a notion of "constrained controlled path" we directly define a notion of rough integral "on $\mathcal{M}$" which is valid for any path in $\mathbb{R}^{e\times d}$ that is controlled by the trace of the integrator. We will then show that, under an additional hypothesis on the integrand, this integral only depends on the restriction of the integrand (and indeed just of its trace) to $T_X M$.

Definition 2.47 (Constrained rough integral)**.** Let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathcal{M})$, $\boldsymbol{H} \in \mathscr{D}_X(\mathbb{R}^{e\times d})$. We define the *$\mathcal{M}$-constrained rough integral* of $\boldsymbol{H}$ against $\boldsymbol{X}$ (both as an element of $\mathscr{D}_X(\mathbb{R}^e)$ and as one of $\mathscr{C}^p_\omega([0,T],\mathbb{R}^e)$) as

$$\int \boldsymbol{H}\mathrm{d}_{\mathcal{M}}\boldsymbol{X} \coloneqq \int \Pi^*\boldsymbol{H}\mathrm{d}\boldsymbol{X} = \int (\boldsymbol{H}\cdot\boldsymbol{P}(X))\mathrm{d}\boldsymbol{X} \tag{2.136}$$

The identity above is shown by the following simple calculation (we will reuse the letters e, d as indices without the risk of ambiguity)

$$\begin{aligned}\int_s^t \Pi^*\boldsymbol{H}\mathrm{d}\boldsymbol{X} &\approx H_{d;s}P^d_c(X_s)X^c_{st} + \big(H'_{ef;s}P^e_a P^f_b(X_s) + H_{d;s}\partial_{ab}\Pi^d(X_s)\big)\mathbb{X}^{ab}_{st} \\ &\approx H_{d;s}P^d_c(X_s)X^c_{st} + \big(H'_{eh;s}P^h_f(X_s) + H_{d;s}\partial_{eh}\Pi^d P^h_f(X_s)\big)P^e_a P^f_b(X_s)\mathbb{X}^{ab}_{st} \\ &\approx H_{d;s}P^d_c(X_s)X^c_{st} + \big(H'_{eh;s}P^h_f(X_s) + H_{d;s}\partial_e P^d_h P^h_f(X_s)\big)P^e_a P^f_b(X_s)\mathbb{X}^{ab}_{st} \\ &\approx H_{d;s}P^d_c(X_s)X^c_{st} + \big(H'_{ah;s}P^h_b(X_s) + H_{d;s}\partial_a P^d_b(X_s)\big)\mathbb{X}^{ab}_{st} \\ &\approx \int_s^t (\boldsymbol{H}\cdot\boldsymbol{P}(X))\mathrm{d}\boldsymbol{X}\end{aligned} \tag{2.137}$$

where we have used that $\boldsymbol{X}$ is constrained to M, the properties of P and (2.98); at the level of Gubinelli derivatives/second order parts the identity is obvious.

Using Corollary 2.17 we compute the correction formula for the traces of the ordinary and constrained rough integrals

$$\int \boldsymbol{H}\mathrm{d}\boldsymbol{X} - \int \boldsymbol{H}\mathrm{d}_{\mathcal{M}}\boldsymbol{X} = \int \boldsymbol{H}\mathrm{d}\Pi_*\boldsymbol{X} - \int \Pi^*\boldsymbol{H}\mathrm{d}\boldsymbol{X} = \frac{1}{2}\int H_c\partial_{ab}\Pi^c(X)\mathrm{d}[\boldsymbol{X}] \tag{2.138}$$

while their second-order parts both agree with

$$\mathbb{Y}^{ij}_{st} \approx H^i_{c;s}H^j_{d;s}P^c_a P^d_b(X_s)\mathbb{X}^{ab}_{st} \approx H^i_{a;s}H^j_{b;s}\mathbb{X}^{ab}_{st} \tag{2.139}$$

In particular, if $\boldsymbol{X}$ is geometric

$$\int \boldsymbol{H}\mathrm{d}_{\mathcal{M}}\boldsymbol{X} = \int \boldsymbol{H}\mathrm{d}\boldsymbol{X} \tag{2.140}$$

and hence agrees with [CDL15, Definition 3.24] when restricted to the case of one-forms (see Example 2.49 below).

Also note that if $\boldsymbol{X}$ is the Itô or Stratonovich stochastic rough path associated to a semimartingale, the above definition coincides, thanks to Remark 2.24, with the usual Itô and Stratonovich integrals, given in extrinsic form in [Dri04, Definition 5.13].

Now, it is clear that if $\Pi^*\boldsymbol{H}$ (or equivalently $\pi^*\boldsymbol{H}$, since $\imath_*$ is injective) vanishes, $\int \boldsymbol{H}\mathrm{d}_{\mathcal{M}}\boldsymbol{X}$ also vanishes, and we may conclude that the integral depends only on the restriction of $\boldsymbol{H}$ to $\mathcal{M}$ in the sense that $\Pi^*\boldsymbol{H} = \Pi^*\boldsymbol{K} \Rightarrow \int \boldsymbol{K}\mathrm{d}_{\mathcal{M}}\boldsymbol{X} = \int \boldsymbol{H}\mathrm{d}_{\mathcal{M}}\boldsymbol{X}$. This, however, falls short of our goal of generalising [CDL15, Corollary 3.35] (or rather one implication — we will address the second one in Remark 2.52 below), which states, in our notation, that if $\boldsymbol{X} \in \mathscr{G}^p_\omega([0,T],\mathcal{M})$ then $\int f(X)\mathrm{d}\boldsymbol{X} = 0$ for all $f \in \Gamma\mathcal{L}(\mathbb{R}^d,\mathbb{R}^e)$ s.t. $\imath^* f = 0$. The point is that the requirement is only placed on the trace $f(X)$ of the integrand, not on the whole controlled path. Unfortunately, without further assumptions, the obvious generalisation to the setting of general controlled integrands of this statement fails. The example below exhibits two ways in which this can occur.

Example 2.48. Take $\mathcal{M}$ to be the unit circle S^1 in $\mathbb{R}^2$, so Π is given by

$$\Pi\colon \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}^2, \quad (x,y) \mapsto \frac{(x,y)}{\sqrt{x^2+y^2}} \tag{2.141}$$

Let $\boldsymbol{Z} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^2)$ given by

$$Z_t \coloneqq (1,0), \quad \mathbb{Z}_{st} \coloneqq \begin{pmatrix} t-s & 0 \\ 0 & t-s \end{pmatrix} \tag{2.142}$$

which satisfies the Chen identity thanks to the constancy of the trace. Define $\boldsymbol{X} \coloneqq \Pi_*\boldsymbol{Z} \in \mathscr{C}^p_\omega([0,T],\mathcal{M})$: it is checked that

$$X_t = (1,0), \quad \mathbb{X}_{st} \approx \begin{pmatrix} 0 & 0 \\ 0 & t-s \end{pmatrix} \tag{2.143}$$

Now let

$$H_t = (H_{1;t}, H_{2;t}) \coloneqq (1,0), \quad H' \coloneqq 0_{2\times 2} \tag{2.144}$$

Trivially, $(H,H') \eqqcolon \boldsymbol{H} \in \mathscr{D}_X(\mathbb{R}^{1\times d})$, and we compute

$$\begin{aligned} \int_s^t \boldsymbol{H}\mathrm{d}\boldsymbol{X} &\approx H_{d;s}P^d_c(X_s)X^c_{st} + (H'_{cd;s}P^c_aP^d_b(X_s) + H_{c;s}\partial_{ab}\Pi^c(X_s))\mathbb{X}^{ab}_{st} \\ &= H_{1;s}\partial_{22}\Pi^1(X_s)\mathbb{X}^{22}_{st} \\ &= s-t \end{aligned} \tag{2.145}$$

despite the fact that $H|_{T_XM} = 0$ (and even $H'|_{T_XM^{\otimes 2}} = 0$).

Another example is given as follows: let $\mathcal{M} = \mathbb{R}^d$ with $d = 2$ (or embed in $\mathbb{R}^3$ if we want non-zero codimension) and let $\boldsymbol{X}$ be the geometric rough path

$$X_t \coloneqq (0,0), \quad \mathbb{X}_{st} \coloneqq \begin{pmatrix} 0 & t-s \\ t-s & 0 \end{pmatrix} \tag{2.146}$$

and $\boldsymbol{H}$ be given by

$$H \coloneqq 0, \quad H'_{st} \coloneqq \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{2.147}$$

Again membership to $\mathscr{D}_X(\mathbb{R}^{2\times 1})$ is trivially satisfied and proceeding as above we compute

$$\int_s^t \boldsymbol{H} \mathrm{d}\boldsymbol{X} = H'_{cd;s} P^c_a P^d_b(X_s) \mathbb{X}^{ab}_{st} = H'_{12} \mathbb{X}^{12}_{st} = t - s \tag{2.148}$$

To summarise, in the first example we were able to have $H|_{T_X M} = 0$, $H'|_{T_X M^{\otimes 2}} = 0$, but the manifold had to be non-flat ($D^2\Pi \neq 0$) and the rough path had to be chosen to be non-geometric (if not $\mathbb{X}^{22} = (X^2_{st})^2 \approx 0$, assuming X is chosen to be in $\mathcal{C}^{p/2}_\omega([0,T],\mathbb{R}^d)$, which is necessary to produce a counterexample). In the second example we were able to choose a geometric rough path, and even a flat manifold, but it was not the case that $H'|_{T_X M^{\otimes 2}} = 0$ (although it still held that $H|_{T_X M} = 0$). Finding a similar example in which $\boldsymbol{X}$ is geometric and $H|_{T_X M} = 0$, $H'|_{T_X M^{\otimes 2}} = 0$ would be more difficult, as is clear from the fact that in this case $\int_s^t \boldsymbol{H} \mathrm{d}_{\mathcal{M}} \boldsymbol{X} \approx \frac{1}{2}\partial_{ab}\Pi(X_s) X^a_{st} X^b_{st}$ (a consequence of (2.140), (2.133) and (2.134)): for the counterexample to work X cannot be truly rough, but at the same time not too regular either.

As will be shown later in Corollary 2.54, this type of behaviour can be ruled out whenever X is truly rough when viewed as being $\mathcal{M}$-valued — it is therefore not accidental that in the examples above X was chosen to be constant (and in particular an element of $\mathcal{C}^{p/2}_\omega([0,T],\mathbb{R}^d)$). A case which is instead always well-behaved is that of 1-form integrands:

Example 2.49 (1-form integrands)**.** Let $f \in \Gamma\mathcal{L}(\tau\mathbb{R}^d, \mathbb{R}^e)$ be a 1-form defined on $\mathbb{R}^d$, and assume for the moment that $f(X)|_{T_X M} = 0$. Then, by differentiating $f = f_d Q^d$ we obtain

$$\boldsymbol{f}(X) = (f_d Q^d_c(X), \partial_a f_d Q^d_b(X) - f_d \partial_{ab}\Pi^d(X))$$

which implies

$$\begin{aligned} \Pi^* \boldsymbol{f}(X) &= (f_d Q^d_e P^e_c(X), \partial_e f_d Q^d_f P^e_a P^f_b(X) - f_d \partial_{ef}\Pi^d P^e_a P^f_b(X) + f_d \partial_{ab}\Pi^d(X)) \\ &= (0, f_d \partial_{ef}\Pi^d (P^e_a Q^f_b + Q^e_a P^f_b + Q^e_a Q^f_b)(X)) \end{aligned}$$

so that

$$\int_s^t \boldsymbol{f}(X) \mathrm{d}_{\mathcal{M}} \boldsymbol{X} \approx f_d \partial_{ef}\Pi^d (P^e_a Q^f_b + Q^e_a P^f_b + Q^e_a Q^f_b)(X_s) \mathbb{X}^{ab}_{st} \approx 0$$

by (2.134). By linearity, this implies that for a general 1-form f, $\int \boldsymbol{f}(X) \mathrm{d}_{\mathcal{M}} \boldsymbol{X}$ only depends on $f(X)|_{T_X M}$.

The same conclusion follows if we realise that the formula [CDL15, (3.17)]

$$\int_s^t f(X) \mathrm{d}\boldsymbol{X} \approx f_d P^d_c(X_s) X^c_{st} + (\nabla f)_{ef}(X_s) P^e_a P^f_b(X_s) \mathbb{X}^{ab}_{st} \tag{2.149}$$

extends to the case of non-geometric rough paths and that $\nabla f(x)$ (defined in (2.96)) only depends on $f(x)|_{T_x M}$ for $x \in M$. This is the extrinsic version of (2.120) applied to one-form integrands, and the same expansion would hold for arbitrary controlled paths, by defining

$$(\nabla \boldsymbol{H})_{ab} \coloneqq H'_{ab} + H_c \partial_{ab}\Pi^c(X) \tag{2.150}$$

This shows that the true roughness assumption is not necessary to integrate ambient one-forms against constrained rough paths in a manner which is only dependent upon their restriction to τM. We have also shown that Definition 2.47 extends [CDL15, Definition 3.24].

Example 2.50 (Affine subspaces)**.** If M is an affine subspace of $\mathbb{R}^d$ then P is constant, and

$$\int_s^t \boldsymbol{H} \mathrm{d}_{\mathcal{M}} \boldsymbol{X} \approx H_{d;s} P^d_c X^c_{st} + H'_{ef;s} P^e_a P^f_b \mathbb{X}^{ab}_{st} \tag{2.151}$$

and in particular only depends on $H|_{T_X M}, H'|_{T_X M^{\otimes 2}}$. The true roughness hypothesis is still necessary if we want dependence only on $H|_{T_X M}$, as demonstrated by Example 2.48.

Example 2.51 (Itô-Stratonovich corrections on embedded manifolds)**.** By Proposition 2.13 the geometrisation of an $\mathcal{M}$-constrained rough path is still constrained, and we may use Example 2.8 to compute the trace-level difference of the constrained integrals against $\boldsymbol{X}$ and its geometrisation as

$$\int \boldsymbol{H} \circ \mathrm{d}\boldsymbol{X} - \int \boldsymbol{H} \mathrm{d}_{\mathcal{M}} \boldsymbol{X} = \frac{1}{2} \int (\nabla \boldsymbol{H}) \mathrm{d}[\boldsymbol{X}] \tag{2.152}$$

This is the extrinsic version of Example 2.40.

Remark 2.52. In [CDL15, Corollary 3.20] it is shown that, for $\boldsymbol{X} \in \mathscr{G}^p_\omega([0,T], \mathbb{R}^d)$ with X valued in $\mathcal{M}$, the condition

$$\int f(X) \mathrm{d}\boldsymbol{X} = 0 \quad \forall f \in \Gamma\mathcal{L}(\tau\mathbb{R}^d, \mathbb{R}^e) \text{ s.t. } \Pi^* f = 0 \tag{2.153}$$

implies (2.134) and thus $\boldsymbol{X} \in \mathscr{G}^p_\omega([0,T], \mathcal{M})$. In order to attempt to generalise this statement to the non-geometric case one would have to pick which of the integrals in (2.136) to use; in both cases, however, the statement becomes trivial since, and even replacing the quantifier over one-forms with one over all controlled integrands $\boldsymbol{H}$, we are dealing with the integral against $\boldsymbol{X}$ of a controlled path with trace $H_c P^c(X) = 0$: if X is truly rough this implies that the whole integrand, and thus the integral, vanishes, regardless of the behaviour of $\mathbb{X}$. Note that using the ordinary $\mathbb{R}^d$-integral in place of the constrained integral (the two coincide for geometric rough paths by (2.140)) is not meaningful either: indeed, if (2.153) implied $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T], \mathcal{M})$ for non-geometric $\boldsymbol{X}$, by Example 2.8

$$\int f(X) \circ \mathrm{d}\boldsymbol{X} = \int f(X) \mathrm{d}\boldsymbol{X} + \frac{1}{2} \int Df(X) \mathrm{d}[\boldsymbol{X}] = \frac{1}{2} \int \partial_a f_b(X) \mathrm{d}[\boldsymbol{X}]^{ab} \tag{2.154}$$

which would have to be zero by Example 2.49 and the fact that $\mathscr{C}^p_\omega([0,T], \mathcal{M})$ is closed under geometrisation (or by Theorem 2.6). But this is not the case if we pick $\mathcal{M}$, $\boldsymbol{X}$ as in [Example 2.48, first example], $e = 1$ and $f(x^1, x^2) = (x^1, x^2)$ (which restricts to 0 on $T\mathcal{M}$) we have $[\boldsymbol{X}]^{ab}_{st} = 2\delta^{a2}\delta^{b2}(s-t)$ and therefore $\frac{1}{2}\int \partial_a f_b(X) \mathrm{d}[\boldsymbol{X}]^{ab} = s - t \neq 0$, a contradiction.

The only way (that we can think of) to characterise non-geometric rough integrals in terms of ambient ones would be to endow $\mathbb{R}^d$ with a connection s.t. $\imath$ is symmetrically affine (which can always be done [É90, Lemma 15]) and replacing the integral in (2.153) with the rough integral in $\mathbb{R}^d$ taken w.r.t. this connection, in the intrinsic sense of Definition 2.36. This, however, falls short of the goal of characterising constrained rough paths in terms of notions that do not involve manifolds.

We still have not related the constrained rough integral with its intrinsic counterpart, defined in Definition 2.36. This is done as follows:

Theorem 2.53. *Let* $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathcal{M})$*,* $\boldsymbol{H} \in \mathscr{D}_X(\mathbb{R}^{e\times d})$*. Then*

$$\int \boldsymbol{H}\mathrm{d}_{\mathcal{M}}\boldsymbol{X} = \int \imath^*\boldsymbol{H}\mathrm{d}_M\pi_*\boldsymbol{X} \tag{2.155}$$

Proof. Applying Proposition 2.38 to $\imath$ we obtain

$$\begin{aligned}\int \boldsymbol{H}\mathrm{d}_{\mathcal{M}}\boldsymbol{X} &= \int \Pi^*\boldsymbol{H}\mathrm{d}(\imath\circ\pi)_*\boldsymbol{X} \\ &= \int \imath^*\Pi^*\boldsymbol{H}\mathrm{d}\pi_*\boldsymbol{X} + \frac{1}{2}\int H_d P^d_c(X)({}^{M,\mathbb{R}^d}\nabla^2\imath)^c_{\alpha\beta}(\pi(X))\mathrm{d}[\pi_*\boldsymbol{X}]^{\alpha\beta}\end{aligned} \tag{2.156}$$

Now,

$$\int \imath^*\Pi^*\boldsymbol{H}\mathrm{d}\pi_*\boldsymbol{X} = \int (\Pi\circ\imath)^*\boldsymbol{H}\mathrm{d}\pi_*\boldsymbol{X} = \int \imath^*\boldsymbol{H}\mathrm{d}\pi_*\boldsymbol{X} \tag{2.157}$$

and applying (2.88) and (2.100), for $x \in M$, $y := \imath(x)$

$$\begin{aligned}({}^{M,\mathbb{R}^d}\nabla^2\imath)^c_{\alpha\beta}(x) &= \partial_{\alpha\beta}\imath^c(x) - \Gamma^\gamma_{\alpha\beta}\partial_\gamma(y)\imath^c(x) \\ &= \partial_{\alpha\beta}\imath^c(x) - \partial_e\pi^\gamma(y)\partial_{\alpha\beta}\imath^e\partial_\gamma\imath^c(x) \\ &= \partial_{\alpha\beta}\imath^c(x) - \partial_e(\imath\circ\pi)^c(y)\partial_{\alpha\beta}\imath^e(x) \\ &= \partial_{\alpha\beta}\imath^c(x) - P^c_e(x)\partial_{\alpha\beta}\imath^e(y) \\ &= Q^c_e(y)\partial_{\alpha\beta}\imath^e(x)\end{aligned} \tag{2.158}$$

which implies

$$H_d P^d_c(X)({}^{M,\mathbb{R}^d}\nabla^2\imath)^c_{\alpha\beta}(\pi(X)) = H_d P^d_c Q^c_e(X)\partial_{\alpha\beta}\imath^e(\pi(X)) = 0 \tag{2.159}$$

concluding the proof. ∎

The following corollary makes sense in light of the fact that true roughness (2.14) is invariant under diffeomorphisms and can thus be defined for manifold-valued paths, in charts.

Corollary 2.54. *If* X *is s.t.* $\pi(X)$ *is truly rough,* $\int \boldsymbol{H}\mathrm{d}_{\mathcal{M}}\boldsymbol{X}$ *only depends on* $H|_{T_X M}$*.*

Proof. Let $\boldsymbol{H}, \boldsymbol{K}$ be s.t. $H|_{T_X M} = K|_{T_X M}$. Then by Theorem 2.53

$$\int \boldsymbol{K}\mathrm{d}_{\mathcal{M}}\boldsymbol{X} - \int \boldsymbol{H}\mathrm{d}_{\mathcal{M}}\boldsymbol{X} = \int \imath^*(\boldsymbol{K}-\boldsymbol{H})\mathrm{d}_M\pi_*\boldsymbol{X} = 0$$

since $(K-H)\circ T\imath = 0$ and true roughness of X imply $\imath^*(\boldsymbol{K}-\boldsymbol{H}) = 0$. ∎

We now turn to the extrinsic treatment of RDEs. Let ${}^{\mathcal{N}}\imath := N \hookrightarrow \mathbb{R}^e$ be a Nash embedding of another Riemannian manifold N, ${}^{\mathcal{N}}\imath(N) =: \mathcal{N}$, and ${}^{\mathcal{N}}\pi, {}^{\mathcal{N}}\Pi$ its Riemannian tubular neighbourhood projections (we will also left superscripts to denote the inclusion/projections relative to $\mathcal{M}$ accordingly). Let $F \in \Gamma\mathcal{L}(\tau\mathbb{R}^d, \tau\mathbb{R}^e)$ restrict to an element of $\Gamma\mathcal{L}(\tau\mathcal{M},\tau\mathcal{N})$ (this means $F(y,x)$ maps $T_x\mathcal{M}$ to $T_y\mathcal{N}$ for $x \in M$, $y \in N$): just as in the intrinsic setting the expression $\mathrm{d}\boldsymbol{Y}^k = F^k_c(Y,X)\mathrm{d}\boldsymbol{X}^c$ is ill-defined, in the extrinsic setting Y will, for $\boldsymbol{X}$ non-geometric, exit $\mathcal{M}$ when the equation is started on $\mathcal{M}$. Proceeding heuristically to derive the extrinsic counterpart to the local RDE formula (2.129), with the idea that $\mathrm{d}_{\mathcal{L}}\boldsymbol{Z} = P_c(Z)\mathrm{d}\boldsymbol{Z}^c$ for an embedded manifold $\mathcal{L}$ and $\boldsymbol{Z} \in \mathscr{C}^p_\omega([0,T],\mathcal{L})$, we interpret

$$\mathrm{d}_{\mathcal{N}}\boldsymbol{Y}^k = F^k_c(Y,X)\mathrm{d}_{\mathcal{M}}\boldsymbol{X}^c \tag{2.160}$$

as ${}^{\mathcal{N}}P^k_h(Y)\mathrm{d}\boldsymbol{Y}^h = F^k_d(Y,X){}^{\mathcal{M}}P^d_c(X)\mathrm{d}\boldsymbol{X}^c$ or in Davie form (using $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathcal{M})$, imposing $\boldsymbol{Y} \in \mathscr{C}^p_\omega([0,T],\mathcal{N})$ and using (2.98)) as

$$\begin{aligned} {}^{\mathcal{N}}P^k_h(Y_s)Y^h_{st} + \partial_{ij}{}^{\mathcal{N}}\Pi^k(Y_s)\mathbb{Y}^{ij}_{st} &\approx F^k_d(Y,X){}^{\mathcal{M}}P^d_c(X)\boldsymbol{X}^c_{st} \\ \mathbb{Y}^{ij}_{st} &\approx F^i_dF^j_b(Y_s,X_s)\mathbb{X}^{ab}_{st} \end{aligned} \tag{2.161}$$

Note that we have chosen not to expand the RHS of the first line into first and second-order parts. Now, by (2.133) applied to Y, we may rewrite this as

$$Y^k_{st} - \tfrac{1}{2}\partial_{ij}{}^{\mathcal{N}}\Pi^k(Y_s)Y^i_{st}Y^j_{st} + \partial_{ij}{}^{\mathcal{N}}\Pi^k(Y_s)\mathbb{Y}^{ij}_{st} \approx F^k_d(Y,X){}^{\mathcal{M}}P^d_c(X)\boldsymbol{X}^c_{st} \tag{2.162}$$

or as (2.164) in the definition below.

Definition 2.55 (Constrained RDE)**.** Given $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathcal{M})$, $y_0 \in \mathcal{N}$ and $F \in \Gamma\mathcal{L}(\tau\mathbb{R}^d,\tau\mathbb{R}^e)$ which restricts to an element of $\Gamma\mathcal{L}(\tau\mathcal{M},\tau\mathcal{N})$ we will write

$$\mathrm{d}_{\mathcal{N}}\boldsymbol{Y}^k = F^k_c(Y,X)\mathrm{d}_{\mathcal{M}}\boldsymbol{X}^c, \quad Y_0 = y_0 \tag{2.163}$$

to mean

$$\mathrm{d}\boldsymbol{Y}^k = F^k_d(Y,X){}^{\mathcal{M}}P^d_c(X)\mathrm{d}\boldsymbol{X}^c + \tfrac{1}{2}\partial_{ij}{}^{\mathcal{N}}\Pi^k(Y)F^i_aF^j_b(Y,X)\mathrm{d}[\boldsymbol{X}]^{ab}, \quad Y_0 = y_0 \tag{2.164}$$

and say that $\boldsymbol{Y}$ solves the *$\mathcal{N}$-constrained RDE* driven by the $\mathcal{M}$-constrained rough path $\boldsymbol{X}$.

The next proposition legitimises this formula.

Theorem 2.56. *Let $\boldsymbol{X}, y_0, F$ be as in Definition 2.55.*

1. *The solution to (2.164) only depends on $(F(y,x)|_{T_x\mathcal{M}})_{x\in\mathcal{M},y\in\mathcal{N}}$ and belongs to $\mathscr{C}^p_\omega([0,T],\mathcal{N})$;*
2. *If $\boldsymbol{X}$ is geometric, so is $\boldsymbol{Y}$ and the equation can be rewritten as $\mathrm{d}\boldsymbol{Y}^k = F^k_c(Y,X)\mathrm{d}\boldsymbol{X}^c$;*
3. *$\boldsymbol{Y} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^e)$ satisfies (2.164) if and only if ${}^{\mathcal{N}}\pi_*\boldsymbol{Y}$ solves the RDE driven by $\boldsymbol{W} := {}^{\mathcal{M}}\pi_*\boldsymbol{X}$*

$$\mathrm{d}\boldsymbol{Z} = (T^{\mathcal{N}}\pi \circ F({}^{\mathcal{N}}\imath(Z), {}^{\mathcal{M}}\imath(W)) \circ T^{\mathcal{M}}\imath)\mathrm{d}\boldsymbol{W} \tag{2.165}$$

   in the sense of Definition 2.42.

Proof. In this proof we will draw on the entirety of the theory of Subsection 2.2.2 and the present section, and will therefore omit the precise equations which motivate our computations. The first part of 1. follows from

3.; we therefore proceed to show that $\boldsymbol{Y}$ is $\mathcal{N}$-constrained. We have, omitting all evaluations at Y_s and X_s and relying on indices to distinguish maps referring to $\mathcal{M}$ and $\mathcal{N}$ (e.g. $P^c_d := {}^{\mathcal{M}}P^c_d(X_s)$, $\partial_{ij}\Pi^k := \partial_{ij}{}^{\mathcal{N}}\Pi^k(Y_s)$)

$$\begin{aligned} &P^k_h Y^h_{st} + \tfrac{1}{2}\partial_{ij}\Pi^k Y^i_{st} Y^j_{st} \\ \approx\ &P^k_h \big[F^h_d P^d_c X^c_{st} + (\partial_a F^h_c P^c_b + F^l_a \partial_l F^h_c P^c_b + F^h_c \partial_a P^c_b)\mathbb{X}^{ab}_{st} + \tfrac{1}{2}\partial_{ij}\Pi^k F^i_a F^j_b [\boldsymbol{X}]^{ab}\big] \\ &+\tfrac{1}{2}\partial_{ij}\Pi^k F^i_a F^j_b X^a_{st} X^b_{st} \end{aligned} \tag{2.166}$$

We calculate

$$\begin{aligned} P^k_h F^h_d P^d_c &= F^k_d P^d_c \\ P^k_h(\partial_a F^h_c P^c_b + F^h_c \partial_a P^c_b) &= \partial_a(F^k_c P^c_b) - Q^k_h \partial_a (F^h_c P^c_b) \\ &= \partial_a(F^k_c P^c_b) - \partial_a(Q^k_h F^h_c P^c_b) \\ &= \partial_a(F^k_c P^c_b) \\ &= \partial_a F^k_c P^c_b + F^k_c \partial_a P^c_b \\ P^k_h F^l_a \partial_l F^h_c P^c_b &= F^l_a \partial_l (P^k_h F^h_c P^c_b) - F^l_a \partial_l P^k_h F^h_c P^c_b \\ &= F^l_a \partial_l (F^k_c P^c_b) - \partial_{ij}\Pi^k F^i_a F^j_b \\ &= F^l_a \partial_l F^k_c P^c_b - \partial_{ij}\Pi^k F^i_a F^j_b \\ P^k_h \partial_{ij}\Pi^h F^i_a F^j_b [\boldsymbol{X}]^{ab}_{st} &\approx P^k_h \partial_{ij}\Pi^h F^i_c F^j_d P^c_a P^d_b [\boldsymbol{X}]^{ab}_{st} \\ &\approx P^k_h \partial_{ij}\Pi^h P^i_l P^j_p F^l_c F^p_d P^c_a P^d_b [\boldsymbol{X}]^{ab}_{st} \\ &\approx 0 \end{aligned} \tag{2.167}$$

Substituting these in (2.166)

$$\begin{aligned} &P^k_h Y^h_{st} + \tfrac{1}{2}\partial_{ij}\Pi^k Y^i_{st} Y^j_{st} \\ \approx\ &F^k_d P^d_c X^c_{st} + (\partial_a F^k_c P^c_b + F^k_c \partial_a P^c_b + F^l_a \partial_l F^h_c P^c_b)\mathbb{X}^{ab}_{st} + \tfrac{1}{2}\partial_{ij}\Pi^k F^i_a F^j_b (X^a_{st} X^b_{st} - 2\mathbb{X}^{ab}_{st}) \\ \approx\ &Y^k_{st} \end{aligned} \tag{2.168}$$

To prove 2. we proceed in a similar fashion: if $\boldsymbol{X}$ is geometric, we have

$$\begin{aligned} Y^k_{st} &\approx F^k_d P^d_c X^c_{st} + (\partial_a F^k_c P^c_b + F^k_c \partial_a P^c_b + F^l_a \partial_l F^h_c P^c_b)\mathbb{X}^{ab}_{st} \\ &\approx F^k_d (X^d_{st} - \tfrac{1}{2}\partial_{ab}\Pi^d X^a_{st} X^b_{st}) + (\partial_a F^k_b + F^l_a \partial_l F^h_b)\mathbb{X}^{ab}_{st} + F^k_d \partial_{ab}\Pi^d (\tfrac{1}{2} X^a_{st} X^b_{st}) \\ &\approx F^k_d X^d_{st} + (\partial_a F^k_b + F^l_a \partial_l F^h_b)\mathbb{X}^{ab}_{st} \end{aligned} \tag{2.169}$$

and $\boldsymbol{Y}$ is geometric because it is the solution to an RDE driven by an $\mathbb{R}^d$-valued geometric rough path.

The proof of 3. is analogous to that of Theorem 2.53. ■

RDEs can be used to generate elements of $\mathscr{C}^p_\omega([0,T],\mathcal{M})$ starting from any unconstrained rough path (cf. [CDL15, Example 4.12, Proposition 4.13] for the geometric case):

Example 2.57 (Projection construction of constrained rough paths)**.** Let $\boldsymbol{Z} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$. Then the solution $\boldsymbol{X}$ to

$$\mathrm{d}_{\mathcal{M}}\boldsymbol{X}^k = P^k_c(X)\mathrm{d}_{\mathbb{R}^d}\boldsymbol{Z}^c, \quad X_0 = x_0 \in M \tag{2.170}$$

i.e.

$$\mathrm{d}\boldsymbol{X}^k = P^k_c(X)\mathrm{d}\boldsymbol{Z}^c + \tfrac{1}{2}\partial_{ij}\Pi^k P^i_a P^j_b(X)\mathrm{d}[\boldsymbol{Z}]^{ab} \tag{2.171}$$

belongs to $\mathscr{C}^p_\omega([0,T],\mathcal{M})$ by Theorem 2.56. Here P and Π refer to the embedded manifold $\mathcal{M}$. Moreover it is checked, using (2.133) and (2.134) that if $\boldsymbol{Z} \in \mathscr{C}^p_\omega([0,T],\mathcal{M})$ with $Z_0 = x_0$, then $\boldsymbol{X} = \boldsymbol{Z}$, i.e. this defines a projection $\mathscr{C}^p_\omega([0,T],\mathbb{R}^d) \twoheadrightarrow \mathscr{C}^p_\omega([0,T],\mathcal{M})$.

2.5 Parallel transport and Cartan development

In this section we will discuss parallel transport and Cartan development (or "rolling without slipping") along/of non-geometric rough paths on manifolds. The topic has already been addressed in the geometric case (in the extrinsic setting) in [CDL15]; not assuming geometricity however introduces several complications. The literature on Itô calculus of semimartingales on manifolds also features similar topics [É90, p.435-440]; however, because of the adjustments that need to be made for the rough path setting, and because of the greater generality with which the theory is approached (even when restricted to semimartingales), the material presented in this chapter will only use the material introduced in its first three sections. We will rely on local coordinates for our computations, and will not explore parallel transport and development in the extrinsic context.

We will tackle parallel transport along non-geometric rough paths by first studying the more general case of RDEs with solutions valued in fibre bundles above the manifold in which the driver is valued; we will progressively restrict our attention to more tractable and interesting cases until we reach the case of the horizontal lift, i.e. in which the equation is the natural generalisation of the parallel transport equation; this will then be used to define Cartan (anti)development, a technique first used to give an intrinsic definition of Brownian motion on Riemannian manifolds [EE71]. We will see that treating non-geometric rough paths entails adding Itô-type corrections to the classical formulae, and that the terms appearing in the resulting equations will have to satisfy second-order conditions for properties that are usually taken for granted (well-definedness, linearity, metricity) to hold.

More precisely, we will consider an m-dimensional smooth manifold M whose tangent bundle is endowed with a linear connection ∇ which we will think of as fixed throughout this section; given a fibre bundle $\pi\colon E \to M$ and a linear connection $\widetilde{\nabla}$ on τE (note we do not require a connection on the bundle π), we are interested in equations of the form

$$\mathrm{d}_{\widetilde{\nabla}}\boldsymbol{Y} = F(Y)\mathrm{d}_\nabla \boldsymbol{X}, \quad Y_0 = y_0 \in E_o \tag{2.172}$$

where F is a section of the bundle $\mathcal{L}_E(\tau M, \tau E)$ (the E subscript denotes the base space: this means we are dealing with a bundle over E, not $E \times M$, i.e. the fibre at $y \in E$ is given by $\mathcal{L}(T_{\pi(y)}M, T_yE)$) and where $X_0 = o \in M$ is a basepoint on the manifold which will be fixed throughout this section. The first thing to notice is that such equations are not of the form Definition 2.42, since F is not defined for all pairs $(y,x) \in E \times M$; we proceed to introduce the tools that are needed to give this type of equation a meaning. Throughout this section we will use the notation in Subsection 2.2.1, and in particular Convention 2.28 for indexing vectors based in the total space of fibre bundles.

Since we want the solution Y to stay in the fibre of X, the very first thing to require of F is that it be a right inverse to $T\pi$:

$$T_y\pi \circ F(y) = \mathbb{1}_{T_xM} \iff \delta^\alpha_\beta = (T_y\pi \circ F(y))^\alpha_\beta = (T_y\pi)^\alpha_K F^K_\beta(y) = \delta^\alpha_K F^K_\beta = F^\alpha_\beta(y) \tag{2.173}$$

We will assume this condition to hold throughout this section unless otherwise stated. For $W \in \Gamma\tau M$ we define

$$FW \in \Gamma\tau E, \quad (FW)(y) \coloneqq F(y)W(\pi(y)) \tag{2.174}$$

In this section we will understand all expressions as being evaluated at (x,y) with $y \in E_x$ unless otherwise specified. The following definition will be of importance in the study of non-geometric RDEs on fibre bundles:

Definition 2.58. We define $\widetilde{F} \coloneqq \widetilde{F}(\widetilde{\nabla}, F)$ by

$$\langle \widetilde{F}, U \otimes V\rangle \coloneqq F\nabla_U V - \widetilde{\nabla}_{FU} FV \in TE, \quad \text{for } U, V \in \Gamma\tau M \tag{2.175}$$

Lemma 2.59. *For $U, V \in \Gamma\tau M$ we have*

$$\begin{aligned}
(\widetilde{\nabla}_{FU}FV)^\gamma &= U^\alpha\partial_\alpha V^\gamma + U^\alpha V^\beta \widetilde{\Gamma}^\gamma_{\alpha\beta} + F^i_\alpha U^\alpha V^\beta \widetilde{\Gamma}^\gamma_{i\beta} + U^\alpha F^j_\beta V^\beta \widetilde{\Gamma}^\gamma_{\alpha j} + F^i_\alpha F^j_\beta U^\alpha V^\beta \widetilde{\Gamma}^\gamma_{ij} \\
(\widetilde{\nabla}_{FU}FV)^k &= U^\alpha(\partial_\alpha F^k_\gamma V^\gamma + F^k_\gamma \partial_\alpha V^\gamma) + F^i_\alpha U^\alpha \partial_i F^k_\gamma V^\gamma \\
&\quad + U^\alpha V^\beta \widetilde{\Gamma}^k_{\alpha\beta} + F^i_\alpha U^\alpha V^\beta \widetilde{\Gamma}^k_{i\beta} + U^\alpha F^j_\beta V^\beta \widetilde{\Gamma}^k_{\alpha j} + F^i_\alpha F^j_\beta U^\alpha V^\beta \widetilde{\Gamma}^k_{ij}
\end{aligned} \tag{2.176}$$

so we have $\widetilde{F} \in \Gamma\mathcal{L}_E(\tau M^{\otimes 2}, \tau E)$, and

$$\widetilde{F}_{\alpha\beta}{}^\gamma = \Gamma^\gamma_{\alpha\beta} - (\widetilde{\Gamma}^\gamma_{\alpha\beta} + F^i_\alpha \widetilde{\Gamma}^\gamma_{i\beta} + F^j_\beta \widetilde{\Gamma}^\gamma_{\alpha j} + F^i_\alpha F^j_\beta \widetilde{\Gamma}^\gamma_{ij}) \tag{2.177}$$

$$\widetilde{F}_{\alpha\beta}{}^k = F^k_\gamma \Gamma^\gamma_{\alpha\beta} - (\partial_\alpha F^k_\beta + F^h_\alpha \partial_h F^k_\beta + \widetilde{\Gamma}^k_{\alpha\beta} + F^i_\alpha \widetilde{\Gamma}^k_{i\beta} + F^j_\beta \widetilde{\Gamma}^k_{\alpha j} + \widetilde{\Gamma}^k_{ij} F^i_\alpha F^j_\beta) \tag{2.178}$$

Proof. We compute

$$\begin{aligned}
\partial_I(FV)^K &= \partial_I(F^K_\gamma(V^\gamma \circ \pi)) \\
&= \partial_I F^K_\gamma V^\gamma + F^K_\gamma \partial_\beta V^\gamma \partial_I \pi^\beta \\
&= \begin{cases} \partial_\alpha V^\gamma & K = \gamma \le m,\ I = \alpha \le m \\ 0 & K = \gamma \le m,\ I = i > m \\ \partial_\alpha F^k_\gamma V^\gamma + F^k_\gamma \partial_\alpha V^\gamma & K = k > m,\ I = \alpha \le m \\ \partial_i F^k_\gamma V^\gamma & K = k > m,\ I = i > m \end{cases}
\end{aligned} \tag{2.179}$$

Substituting these terms in

$$(\widetilde{\nabla}_{FU}FV)^K = (FU)^I \partial_I (FV)^K + (FU)^I (FV)^J \widetilde{\Gamma}^K_{IJ} \tag{2.180}$$

yields the desired expressions.

We must now show that $\widetilde{F}$ is bilinear, thus legitimising our use of the notation $\langle \widetilde{F}, U \otimes V\rangle$: this is easily done by computing the RHS of (2.175) thanks to the previously computed expression, and seeing that the derivatives of V cancel out, leaving us with the desired expressions for $\widetilde{F}_{\alpha\beta}{}^\gamma$ and $\widetilde{F}_{\alpha\beta}{}^k$. ■

The task is now to extend F to all pairs (y, x) where y does not necessarily lie in E_x (the existence of such extensions is not hard to show: see [É89, Lemma 8.16, Proof of Proposition 8.15]), and to investigate when the resulting (2.172), which can now be understood as in Definition 2.42, is independent of the extension. To do so we introduce the following condition on $\widetilde{\nabla}$ and F (with ∇ thought of as fixed).

Condition 2.60. Assuming (2.173) holds, for all $U \in \Gamma\tau M$ we have

$$T\pi\widetilde{\nabla}_{FU}(FU) = \nabla_U U \tag{2.181}$$

i.e. $T\pi\langle\widetilde{F}, U \otimes U\rangle = 0$.

We have purposefully stated the condition with two copies of the same vector field U, instead of $T\pi\widetilde{\nabla}_{FU}(FV) = \nabla_U V$: this is motivated by the fact that we only need the symmetrisation of the latter identity (obtained through polarisation), since these terms will turn out to be the coefficients of the bracket. Requiring the unsymmetrised version of the condition would perhaps have been more natural, but is not as sharp; similar comments hold for Condition 2.67 and Condition 2.79 below. Recall the round bracket notation for symmetrisation.

Lemma 2.61. *Condition 2.60 is equivalent to $\widetilde{F}$ restricting to an element of $\Gamma\mathcal{L}_E(\tau M^{\odot 2}, V\pi)$, or in product coordinates to $\widetilde{F}_{(\alpha\beta)}{}^{\gamma} = 0$, i.e.*

$$\widetilde{\Gamma}^{\gamma}_{\alpha\beta} + \widetilde{\Gamma}^{\gamma}_{i\beta}F^i_\alpha + \widetilde{\Gamma}^{\gamma}_{\alpha j}F^j_\beta + \widetilde{\Gamma}^{\gamma}_{ij}F^i_\alpha F^j_\beta \overset{(\alpha\beta)}{=} \Gamma^{\gamma}_{\alpha\beta} \tag{2.182}$$

Moreover, the condition being satisfied for all choices of F is equivalent to τM being symmetrically affine w.r.t. $\widetilde{\nabla}, \nabla$ (with the "only if" statement only valid for $m \geq 2$).

Proof. The first characterisation of Condition 2.60 is obvious, and the expression in local coordinates is a direct consequence of Lemma 2.59 and polarisation.

As for the second statement, we must check that the conditions on the symmetrised Christoffel symbols stated in Example 2.30 hold. The "if" part is immediate. For the converse, first of all reading the identity with $F = 0$ yields $\widetilde{\Gamma}^{\gamma}_{\alpha\beta} \overset{(\alpha\beta)}{=} \Gamma^{\gamma}_{\alpha\beta}$, and the identity may be rewritten as

$$(\widetilde{\Gamma}^{\gamma}_{\alpha j} + \widetilde{\Gamma}^{\gamma}_{j\alpha})F^j_\beta + (\widetilde{\Gamma}^{\gamma}_{i\beta} + \widetilde{\Gamma}^{\gamma}_{\beta i})F^i_\alpha + (\widetilde{\Gamma}^{\gamma}_{ij} + \widetilde{\Gamma}^{\gamma}_{ji})F^i_\alpha F^j_\beta = 0 \tag{2.183}$$

Now read the identity for arbitrary but fixed $\alpha \neq \beta$ (which is possible since $m \geq 2$) and j, and with $F^k_\gamma := \delta^{kj}\delta_{\gamma\beta}$ (this is possible since the coefficients F^k_γ are completely arbitrary: we do not even have to argue coordinate-independence, as everything is local and we may take F to be supported in the domain of the chart): this yields $\widetilde{\Gamma}^{\gamma}_{\alpha j} \overset{(\alpha j)}{=} 0$, reducing our identity to $(\widetilde{\Gamma}^{\gamma}_{ij} + \widetilde{\Gamma}^{\gamma}_{ji})F^i_\alpha F^j_\beta = 0$. We may then fix arbitrary i, j and pick F exactly as above to conclude $\Gamma^{\gamma}_{ij} \overset{(ij)}{=} 0$. ■

The next result establishes the link between the condition and well-definedness of RDEs in fibre bundles.

Theorem 2.62. *If Condition 2.60 holds (2.172) is well defined, i.e. it is independent of the extension of F to an element of $\Gamma\mathcal{L}_{E\times M}(\tau M, \tau E)$. In this case $\pi(Y) = X$ and the coordinate expression of the RDE reduces to its* *vertical component and is given by*

$$\begin{aligned}\mathrm{d}\boldsymbol{Y}^k &= F^k_\gamma(Y)\mathrm{d}\boldsymbol{X}^\gamma \\ &+\tfrac{1}{2}\big(F^k_\gamma(Y)\Gamma^\gamma_{\alpha\beta}(X) - (\widetilde{\Gamma}^k_{\alpha\beta} + \widetilde{\Gamma}^k_{\alpha j}F^j_\beta + \widetilde{\Gamma}^k_{i\beta}F^i_\alpha + \widetilde{\Gamma}^k_{ij}F^i_\alpha F^j_\beta)(Y)\big)\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}\end{aligned} \tag{2.184}$$

which may be written at the trace level as

$$\mathrm{d}Y^k = F^k_\gamma(Y) \circ \mathrm{d}\boldsymbol{X}^\gamma + \tfrac{1}{2}\widetilde{F}_{\alpha\beta}{}^k(Y)\mathrm{d}[\boldsymbol{X}]^{\alpha\beta} \tag{2.185}$$

Moreover, if $\boldsymbol{X}$ is geometric the equation is always well defined and independent of the connections ∇ and $\widetilde{\nabla}$.

Proof. Denoting still with $F = F(y,x)$ an arbitrary extension of F as a section of the bundle $\mathcal{L}_{E\times M}(\tau M, \tau E)$ (i.e. $F(y,\pi(y)) = F(y)$), we have that the first m coordinates of the local form of (2.172) is given by

$$\mathrm{d}\boldsymbol{Y}^\gamma = F^\gamma_\alpha(Y,X)\mathrm{d}\boldsymbol{X}^\alpha + \tfrac{1}{2}\big(\Gamma^\gamma_{(\alpha\beta)}(X) - \widetilde{\Gamma}^\gamma_{(IJ)}(Y)F^I_\alpha F^J_\beta(Y,X)\big)\mathrm{d}[\boldsymbol{X}]^{\alpha\beta} \tag{2.186}$$

where we have symmetrised the second order part thanks to the symmetry of the tensor $[\boldsymbol{X}]$. Notice that by (2.173), by hypothesis and Lemma 2.61 we have that on pairs (Y,X) s.t. $Y \in T_X M$ the coefficient of $\mathrm{d}\boldsymbol{X}^\alpha$ equals δ^α_γ and that of $\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}$ vanishes. Now consider the RDE defined only locally in the domain of the chart (i.e. without the claim that the following is a coordinate-invariant expression)

$$\mathrm{d}\begin{pmatrix}\boldsymbol{Y}^\gamma \\ \boldsymbol{Y}^k\end{pmatrix} = \begin{pmatrix}\mathrm{d}\boldsymbol{X}^\gamma \\ F^k_\gamma(Y,X)\mathrm{d}\boldsymbol{X}^\gamma + \tfrac{1}{2}(F^k_\gamma(Y,X)\Gamma^\gamma_{\alpha\beta}(X) - \widetilde{\Gamma}^k_{IJ}(Y)F^I_\alpha F^J_\beta(Y,X))\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}\end{pmatrix} \tag{2.187}$$

The solution to this RDE stays in the fibre of the trace of the driver X. But the solution to this RDE must also solve (2.186), since it takes its values in the locus in which the coefficients of the two coincide (here we are using the obvious principle that the solution of an RDE does not change if the coefficients are modified away from the solution).

If $\boldsymbol{X}$ is geometric $[\boldsymbol{X}]$ vanishes altogether and we may show well-definedness in the same manner, and is independent of the connections on the source and target manifolds since the driver is geometric. ■

For the remainder of this section we assume Condition 2.60 is satisfied unless otherwise stated. An even stronger requirement, which we will instead **not** assume to hold by default, is:

Condition 2.63. $F\nabla_U U = \widetilde{\nabla}_{FU} FU$ for all $U \in \Gamma\tau M$, i.e. $\widetilde{F} = 0$, or in local coordinates (assuming Condition 2.60 already holds)

$$F^k_\gamma \Gamma^\gamma_{\alpha\beta} \overset{(\alpha\beta)}{=} \partial_\alpha F^k_\beta + F^h_\alpha \partial_h F^k_\beta + \widetilde{\Gamma}^k_{\alpha\beta} + F^i_\alpha \widetilde{\Gamma}^k_{i\beta} + F^j_\beta \widetilde{\Gamma}^k_{\alpha j} + \widetilde{\Gamma}^k_{ij} F^i_\alpha F^j_\beta \tag{2.188}$$

The following is immediately inferred through Theorem 2.62.

Corollary 2.64. *If Condition 2.63 holds then the trace-level of (2.172) is well defined and equivalent to*

$$\mathrm{d}Y = F(Y) \circ \mathrm{d}\boldsymbol{X} \tag{2.189}$$

We will mostly use the geometrised form of our fibre bundle-valued RDEs, regardless of whether Condition 2.63 holds or not, keeping in mind that the second-order part of the solution is given in terms of the original rough path $\boldsymbol{X}$ as $\mathbb{Y}^{ij}_{st} \approx F^i_a F^j_b(Y_s)\mathbb{X}^{ab}_{st}$.

In the following example we show how it is possible to generate integrands by solving cotangent bundle-valued RDEs.

Example 2.65 (Solutions to RDEs as controlled integrands)**.** If $\pi = \tau^* M$ we may make the solution to (2.172) into an element of $\mathscr{D}_X(\tau^* M) = \mathscr{D}_X(\mathcal{L}(\tau M, \mathbb{R}))$ (on any interval $[0, S]$ on which the solution is defined) as follows. Working in induced coordinates, it is natural to define $Y'_{\alpha\beta} := F^{\widetilde{\beta}}_{\alpha}(Y)$: since the resulting $\boldsymbol{Y}$ is a controlled path in each coordinate chart, we need only check that it satisfies the transformation rule required in Definition 2.34 to conclude it is a controlled integrand. At the path level this is already guaranteed by the soundness of Definition 2.42, but for Gubinelli derivatives it must be verified, as we have not defined solutions to RDEs on manifolds as controlled paths: by (2.83) we have

$$\begin{aligned} Y'_{\overline{\alpha}\overline{\beta}} &= F^{\widetilde{\overline{\beta}}}_{\overline{\alpha}}(Y) \\ &= \partial^{\widetilde{\overline{\beta}}}_{B} F^{B}_{\alpha}(Y)\partial^{\alpha}_{\overline{\alpha}} \\ &= \partial^{\widetilde{\overline{\beta}}}_{\beta} \delta^{\beta}_{\alpha} \partial^{\alpha}_{\overline{\alpha}} + \partial^{\widetilde{\overline{\beta}}}_{\widetilde{\beta}} F^{\widetilde{\beta}}_{\alpha}(Y)\partial^{\alpha}_{\overline{\alpha}} \\ &= \partial^{\gamma}_{\overline{\alpha}\overline{\beta}} Y_{\gamma} + \partial^{\alpha}_{\overline{\alpha}} \partial^{\beta}_{\overline{\beta}} Y'_{\alpha\beta} \end{aligned} \tag{2.190}$$

which is precisely the transformation rule required to be satisfied by Gubinelli derivatives under [Proposition 2.12, Pullback]. We have thus given a meaning to the scalar rough integral $\int \boldsymbol{Y} \mathrm{d}\boldsymbol{X}$. Examples of connections on the cotangent bundle, analogous to the complete, horizontal and Sasaki lifts, defined below for the tangent bundle, are given in [YI73, p.269, p.286] and [SA11]; it would be interesting to verify that Condition 2.60 holds for these connections. If we start with the bundle $\mathcal{L}(\tau M, \mathbb{R}^e) = (\tau^* M)^{\oplus e}$, we may similarly define $\mathbb{R}^e$-valued rough integrals (here we need a connection on $\tau\mathcal{L}(TM, \mathbb{R}^e)$). It is similarly checked that taking $\pi = \tau M$ (as done below) results in the solution of (2.172) being a τM-valued controlled path in the sense of Example 2.41 (with Gubinelli derivatives $Y'^{\beta}_{\alpha} = F^{\widetilde{\beta}}_{\alpha}$), which can then be integrated against $\boldsymbol{X}$ once a Riemannian metric on τM is given.

We proceed with the main theory. **For the remainder of this section we let π be a vector bundle** unless otherwise stated. We will say that $\widetilde{U} \in \Gamma\tau E$ is *linear* if

$$\widetilde{U}^{\gamma}(y) = U^{\gamma}(x), \quad \widetilde{U}^{k}(y) = \widetilde{U}^{k}_{h}(x) y^{h} \tag{2.191}$$

with $x = \pi(y)$ and for locally defined functions $U^{\gamma}, \widetilde{U}^{k}_{h}$. We will say that F (which we are assuming satisfies (2.173) and Condition 2.60) is *linear* if FU is a linear vector field for all $U \in \Gamma\tau M$. We will continue to assume $x := \pi(y)$ below.

Lemma 2.66. *The condition of $U \in \Gamma\tau E$ of being linear is intrinsic. Therefore that of F of being such is too, and in coordinates it amounts to*

$$F^{k}_{\gamma}(y) = F^{k}_{\gamma h}(x) y^{h} \tag{2.192}$$

for locally defined functions $F^{k}_{\gamma h}$.

Proof. With the notations of (2.81) we have

$$\begin{aligned}\widetilde{U}^{\overline{\gamma}}(y) &= \partial_K^{\overline{\gamma}}(x)\widetilde{U}^K(y) = \partial_\gamma^{\overline{\gamma}}U^\gamma(x) \\ \widetilde{U}^{\overline{k}}(y) &= \partial_K^{\overline{k}}(x)\widetilde{U}^K(y) \\ &= \partial_\gamma\lambda_k^{\overline{k}}(x)y^kU^\gamma(x) + \lambda_k^{\overline{k}}(x)\widetilde{U}_h^k(y)y^h \\ &= (\partial_\gamma\lambda_k^{\overline{k}}(\lambda^{-1})_{\overline{h}}^kU^\gamma + \lambda_k^{\overline{k}}\widetilde{U}_h^k(\lambda^{-1})_{\overline{h}}^h)(x)y^{\overline{h}}\end{aligned} \tag{2.193}$$

and we may therefore set $U^{\overline{\gamma}} = \partial_\gamma^{\overline{\gamma}}U^\gamma$, $\widetilde{U}_{\overline{h}}^{\overline{k}} = \partial_\gamma\lambda_k^{\overline{k}}(\lambda^{-1})_{\overline{h}}^kU^\gamma + \lambda_k^{\overline{k}}\widetilde{U}_h^k(\lambda^{-1})_{\overline{h}}^h$. The linearity condition on F is not stated with reference to a particular coordinate system, and is therefore invariant under change of coordinates because linearity of vector fields is. Finally picking an arbitrary $U \in \Gamma\tau M$, we have $(FU)^\gamma(y) = U^\gamma(x)$ by (2.173) and for $(FU)^k(y) = F_\gamma^k(y)U^\gamma(x)$ to be of the form $\widetilde{U}_h^k(x)y^h$ for all U we need $F_\gamma^k(y)$ to be of the form $F_{\gamma h}^k(x)y^h$ (for the "only if" implication simply pick $U^\gamma = \delta_\beta^\gamma$ with $\beta = 1, \ldots, m$). ∎

Note that the k index in $F_{\gamma h}^k(x)y^h$ represents a coordinate in T_yE, whereas h represents a coordinate in E_x; following Convention 2.28 we will not place a twiddle on the upper index, as we view $F_{\gamma h}^k$ as the coordinates of a linear map between vector spaces. **For the remainder of this section we will assume F to be linear**, and we will be concerned with the question of whether this implies that the resulting (2.172) is also linear, i.e. that its coordinate expression is linear in the conventional sense. To this end, we introduce the following condition on F and $\widetilde{\nabla}$ (which does not involve the connection ∇ at all, to the extent that $\widetilde{\nabla}$ is not defined in terms of it).

Condition 2.67. Assume F is linear. $\widetilde{\nabla}_{FU}(FU)$, or equivalently $\langle\widetilde{F}, U \otimes U\rangle$, is a linear vector field for all $U \in \Gamma\tau M$.

Lemma 2.68. *Let F be linear and Condition 2.60 hold. Then Condition 2.67 is equivalent to $\widetilde{F}|_{TM^{\odot 2}}$ lying in the image of the map*

$$\begin{aligned}\Gamma(\tau^*M^{\odot 2} \otimes \pi^* \otimes \pi) = \Gamma\mathcal{L}_M(\tau M^{\odot 2} \otimes \pi, \pi) &\to \Gamma\mathcal{L}_E(\tau M^{\odot 2}, V\pi) \\ G &\mapsto \big(e \mapsto (U \odot V \mapsto \mathfrak{v}(e)\langle G, U \odot V \otimes e\rangle)\big)\end{aligned} \tag{2.194}$$

where $\mathfrak{v}(e)\colon E \to V_e\pi$ denotes the vertical lift isomorphism based at e. In other words, we may write its coordinates (symmetrising in the first two indices) as $\widetilde{F}_{(\alpha\beta)h}{}^\gamma = 0$ and

$$\begin{aligned}\widetilde{F}_{\alpha\beta h}{}^k y^h \overset{(\alpha\beta)}{=} F_{\gamma h}^k\Gamma_{\alpha\beta}^\gamma y^h - (\partial_\alpha F_{\beta h}^k y^h + F_{\alpha h}^l F_{\beta l}^k y^h \\ + \widetilde{\Gamma}_{ij}^k F_{\alpha h}^i F_{\beta l}^j y^h y^l + \widetilde{\Gamma}_{\alpha\beta}^k + F_{\alpha h}^i \widetilde{\Gamma}_{i\beta}^k y^h + F_{\beta h}^j \widetilde{\Gamma}_{\alpha j}^k y^h)\end{aligned} \tag{2.195}$$

and it follows that an equivalent formulation of the condition is that the expression

$$\widetilde{\Gamma}_{\alpha\beta}^k + \widetilde{\Gamma}_{\alpha j}^k F_{\beta h}^j y^h + \widetilde{\Gamma}_{i\beta}^k F_{\alpha h}^i y^h + \widetilde{\Gamma}_{ij}^k F_{\alpha h}^i F_{\beta l}^j y^h y^l, \quad (\alpha\beta) \tag{2.196}$$

is linear in the y coordinates.

Moreover, the condition being satisfied for all choices of F as above (without assuming Condition 2.60 is) is equivalent to the stronger requirement that $\widetilde{\nabla}_{\widetilde{U}}\widetilde{U}$ be linear for all linear $\widetilde{U} \in \Gamma\tau E$ (with the "only if" statement *only valid for $m \geq 2$), which in coordinates reads*

$$\begin{aligned} &\widetilde{\Gamma}^\gamma_{(\alpha\beta)} \text{ constant in } y, \quad \widetilde{\Gamma}^\gamma_{(i\beta)} = \widetilde{\Gamma}^\gamma_{(\alpha j)} = \widetilde{\Gamma}^\gamma_{(ij)} = 0 \\ &\widetilde{\Gamma}^k_{(\alpha\beta)} \text{ linear in } y, \quad \widetilde{\Gamma}^k_{(\alpha j)}, \widetilde{\Gamma}^k_{(i\beta)} \text{ constant in } y, \quad \widetilde{\Gamma}^k_{(ij)} = 0 \end{aligned} \tag{2.197}$$

Note that in (2.195) we are not able to provide an expression for $\widetilde{F}_{\alpha\beta h}{}^k$, since some of the terms on the RHS are non-linear (recall that the $\Gamma^\gamma_{\alpha\beta}$'s and $F^k_{\gamma h}$ are evaluated at x, but the $\widetilde{\Gamma}^K_{IJ}$ are non-linearly evaluated at y, and moreover there are quadratic terms).

Proof of Lemma 2.68. The first characterisation of Condition 2.67 is just a reformulation of the second, which is evident by (2.178), Lemma 2.61 and the definition of linear vector field. The third follows from the second by subtracting terms that are already linear in y.

As for the second statement, we first observe that linearity of $\widetilde{\nabla}_{FU} FU$ without requiring Condition 2.60 entails the additional requirement that (by (2.177), rewritten to account for the linearity of F) the expression

$$\Gamma^\gamma_{\alpha\beta} - (\widetilde{\Gamma}^\gamma_{\alpha\beta} + F^i_{\alpha h}\widetilde{\Gamma}^\gamma_{i\beta} y^h + F^j_{\beta h}\widetilde{\Gamma}^\gamma_{\alpha j} y^h + F^i_{\alpha h} F^j_{\beta l}\widetilde{\Gamma}^\gamma_{ij} y^h y^l), \quad (\alpha\beta) \tag{2.198}$$

be constant in y. Then by arguing as in the proof of Lemma 2.61 by progressively disregarding constant (resp. linear) terms in (2.198) (resp. (2.196)) we may conclude that linearity of $\widetilde{\nabla}_{FU} FU$ for all F and U as above is equivalent to (2.197).

Now, writing $(\nabla_{\widetilde{U}} \widetilde{V})^K = \widetilde{U}^I \partial_I \widetilde{V}^K + \widetilde{U}^I \widetilde{V}^J \widetilde{\Gamma}^K_{IJ}$ for $\widetilde{U}, \widetilde{V}$ linear (with notation as in (2.191)) we obtain

$$\begin{aligned} (\nabla_{\widetilde{U}} \widetilde{V})^\gamma &= U^\alpha \partial_\alpha V^\gamma + U^\alpha V^\beta \widetilde{\Gamma}^\gamma_{\alpha\beta} + U^\alpha \widetilde{V}^j_h \widetilde{\Gamma}^\gamma_{\alpha j} y^h + \widetilde{U}^i_h V^\beta \widetilde{\Gamma}^\gamma_{i\beta} y^h + \widetilde{U}^i_h \widetilde{V}^j_l \widetilde{\Gamma}^\gamma_{ij} y^h y^l \\ (\nabla_{\widetilde{U}} \widetilde{V})^k &= U^\alpha \partial_\alpha \widetilde{V}^k_h y^h + \widetilde{U}^i_h \widetilde{V}^k_i y^h + U^\alpha V^\beta \widetilde{\Gamma}^k_{\alpha\beta} + U^\alpha \widetilde{V}^j_h \widetilde{\Gamma}^k_{\alpha j} y^h + \widetilde{U}^i_h V^\beta \widetilde{\Gamma}^k_{i\beta} y^h \\ &\quad + \widetilde{U}^i_h \widetilde{V}^j_l \widetilde{\Gamma}^k_{ij} y^h y^l \end{aligned} \tag{2.199}$$

As usual, we rely on the symbols involved to infer whether a function is evaluated at $y \in E$ or at $x = \pi(y)$. We then see, by arbitrarity of $U^\gamma, V^\gamma, \widetilde{U}^k_h, \widetilde{V}^k_h \in C^\infty M$, polarisation, and the usual elimination procedure, that linearity of $\nabla_{\widetilde{U}} \widetilde{U}$ is equivalent to (2.197). ■

Assuming Condition 2.60 is satisfied we may consider the *flow map* associated to F and $\boldsymbol{X}$ at times $0 \leq s \leq t \leq T$

$$\begin{aligned} &\Phi_{ts} = \Phi(F, \boldsymbol{X})_{ts} \colon E_{X_s} \to E_{X_t}, \quad y \mapsto Y_t \\ &\text{where} \quad \mathrm{d}Y = F(Y)\mathrm{d}\boldsymbol{X},\ Y_s = y \end{aligned} \tag{2.200}$$

which is defined as long as Y_t is defined, and by uniqueness we have

$$\Phi_{tu} \circ \Phi_{us} = \Phi_{ts} \tag{2.201}$$

for $0 \leq s \leq u \leq t \leq T$ whenever one of the two sides is defined. The following theorem justifies our interest in the linearity condition.

Theorem 2.69. *Let F be linear and satisfy Condition 2.60 and Condition 2.67. Then (2.172) can be written in coordinates as*

$$\mathrm{d}Y^k = F^k_{\gamma h}(X)Y^h \circ \mathrm{d}\boldsymbol{X}^\gamma + \tfrac{1}{2}\widetilde{F}_{(\alpha\beta)h}{}^k(X)Y^h\mathrm{d}[\boldsymbol{X}]^{\alpha\beta} \tag{2.202}$$

and admits a global solution. Moreover, Φ_{ts} defines linear isomorphisms $E_{X_s} \cong E_{X_t}$ for all $0 \leq s \leq t \leq T$. These statements also hold, independently of $\widetilde{F}$, if $\boldsymbol{X}$ is geometric.

Proof. The first statement is a restatement of (2.184) to the case in which Condition 2.67 is satisfied. We may argue global existence by Theorem 2.21 and Remark 2.43: indeed, assume that there exists $S \leq T$ such that $Y_{[0,S)}$ is not contained in any compact set of M. Since $\pi(Y) = X$ on $[0, S)$, we must have that $\lim_{t\to S^-} \pi(Y_t) = X_S$, i.e. Y must "explode vertically". This, however, is not possible either, since if we may pick a system of product coordinates which contains X_S, this would mean that the coordinate solution to (2.184) must only be defined for $t < S$, which is ruled out by Lemma 2.23.

Standard uniqueness arguments apply in charts to show that Φ_{ts} is a linear monomorphism (and thus an isomorphism, by dimensionality) when X_s, X_t are contained in a single chart, and these can be combined to yield the global statement by "patching" $X_{[0,T]}$ with finitely many charts and applying (2.201). ■

We will denote $\Phi_{st} := \Phi_{ts}^{-1}$ for $0 \leq s \leq t \leq T$. We proceed to study the local dynamics satisfied by $t \mapsto \Phi_{t0}$ and $t \mapsto \Phi_{0t}$. Fix coordinates for the vector space $E_o = E_{X_0}$, which we denote with the symbols $i^\circ, j^\circ, k^\circ, \ldots$; we continue to denote with $\alpha, \beta, \gamma \ldots$ and $i, j, k \ldots$ the local coordinates in and above a neighbourhood containing X_t; we do not intend for the former indices to bear any relationship with the latter (e.g. k° and k appearing in a common expression have nothing to do with each other).

Proposition 2.70. *The coordinate expressions $\Phi^k_{k^\circ;t0}$ and $\Phi^{k^\circ}_{k;0t}$ respectively solve the RDEs (at the trace level) driven by $({}_{\mathrm{g}}\boldsymbol{X}, [\boldsymbol{X}])$*

$$\begin{aligned}
\mathrm{d}\Phi^k_{k^\circ;t0} &= F^k_{\gamma h}(X_t)\Phi^h_{k^\circ;t0} \circ \mathrm{d}\boldsymbol{X}^\gamma_t + \tfrac{1}{2}\widetilde{F}_{(\alpha\beta)h}{}^k(X_t)\Phi^h_{k^\circ;t0}\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}_t \\
\mathrm{d}\Phi^{k^\circ}_{k;0t} &= -\Phi^{k^\circ}_{h;0t}F^h_{\gamma k}(X_t) \circ \mathrm{d}\boldsymbol{X}^\gamma_t - \tfrac{1}{2}\Phi^{k^\circ}_{h;0t}\widetilde{F}_{(\alpha\beta)k}{}^h(X_t)\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}_t
\end{aligned} \tag{2.203}$$

Proof. The statement is local, and we may confine ourselves to the domain of a single set of product coordinates containing X_t. By Theorem 2.69 we have

$$\begin{aligned}
(\mathrm{d}\Phi^k_{k^\circ;t0})y &= \mathrm{d}(\Phi^k_{k^\circ;t0}y) \\
&= \mathrm{d}Y_t \\
&= F^k_{\gamma h}(X_t)Y^h_t \circ \mathrm{d}\boldsymbol{X}^\gamma_t + \tfrac{1}{2}\widetilde{F}_{(\alpha\beta)h}{}^k(X_t)Y^h_t\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}_t \\
&= F^k_{\gamma h}(X_t)\Phi^k_{h^\circ;t0}y \circ \mathrm{d}\boldsymbol{X}^\gamma_t + \tfrac{1}{2}\widetilde{F}_{(\alpha\beta)h}{}^k(X_t)\Phi^h_{k^\circ;t0}y\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}_t \\
&= (F^k_{\gamma h}(X_t)\Phi^k_{h^\circ;t0} \circ \mathrm{d}\boldsymbol{X}^\gamma_t + \tfrac{1}{2}\widetilde{F}_{(\alpha\beta)h}{}^k(X_t)\Phi^h_{k^\circ;t0}\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}_t)y
\end{aligned} \tag{2.204}$$

We may therefore conclude, by arbitrarity of $y \in E_o$, that the first of the two RDEs holds. As for the second,

we have

$$
\begin{aligned}
0 &= \mathrm{d}\delta^{k^\circ}_{h^\circ} \\
&= \mathrm{d}(\Phi^{k^\circ}_{k;0t}\Phi^{k}_{h^\circ;t0}) \\
&= (\mathrm{d}\Phi^{k^\circ}_{k;0t})\Phi^{k}_{h^\circ;t0} + \Phi^{k^\circ}_{k;0t}(F^k_{\gamma h}(X_t)\Phi^h_{h^\circ;t0} \circ \mathrm{d}\boldsymbol{X}^\gamma_t + \tfrac{1}{2}\widetilde{F}_{(\alpha\beta)h}{}^k(X_t)\Phi^h_{h^\circ;t0}\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}_t)
\end{aligned}
\tag{2.205}
$$

which we rewrite as

$$
\begin{aligned}
\mathrm{d}\Phi^{k^\circ}_{l;0t} &= -\Phi^{k^\circ}_{k;0t}(F^k_{\gamma h}(X_t)\Phi^h_{h^\circ;t0} \circ \mathrm{d}\boldsymbol{X}^\gamma_t + \tfrac{1}{2}\widetilde{F}_{(\alpha\beta)h}{}^k(X_t)\Phi^h_{h^\circ;t0}\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}_t)\Phi^{h^\circ}_{l;0t} \\
&= -\Phi^{k^\circ}_{k;0t}F^k_{\gamma l}(X_t) \circ \mathrm{d}\boldsymbol{X}^\gamma_t - \tfrac{1}{2}\Phi^{k^\circ}_{k;0t}\widetilde{F}_{(\alpha\beta)l}{}^k(X_t)\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}_t
\end{aligned}
\tag{2.206}
$$

thus concluding the proof. ■

For the remainder of this section we will let $\pi = \tau M$ unless otherwise stated. An important feature of the equation in this case is that we can integrate the inverse of the flow map to obtain a $T_o M$-valued rough path. Note that, although we have not explicitly defined controlled integrands with values in an arbitrary finite-dimensional vector space V, this is done simply by choosing a basis of V and setting $\mathscr{D}_X(\mathcal{L}(\tau M, V)) := \mathscr{D}_X(\mathcal{L}(\tau M, \mathbb{R}^m))$ under the corresponding isomorphism $V \cong \mathbb{R}^m$ (all the needed constructions are easily seen not to depend on the choice of the basis). The next lemma states the change of coordinate formula satisfied by $F^\gamma_{\alpha\beta}$:

Lemma 2.71.

$$
F^{\widetilde{\overline{\gamma}}}_{\overline{\alpha}\overline{\beta}} = \partial^{\overline{\gamma}}_{\gamma}\partial^{\alpha}_{\overline{\alpha}}\partial^{\beta}_{\overline{\beta}}F^\gamma_{\alpha\beta} - \partial^{\overline{\gamma}}_{\gamma}\partial^{\gamma}_{\overline{\alpha}\overline{\beta}}
\tag{2.207}
$$

Proof. By (2.82) we have

$$
\begin{aligned}
F^{\widetilde{\gamma}}_{\alpha\beta}\partial^{\beta}_{\overline{\beta}}y^{\overline{\beta}} &= F^{\widetilde{\gamma}}_{\alpha\beta}y^\beta \\
&= F^{\widetilde{\gamma}}_{\alpha} \\
&= \partial^{\widetilde{\gamma}}_{\overline{C}}F^{\overline{C}}_{\overline{\alpha}}\partial^{\overline{\alpha}}_{\alpha} \\
&= \partial^{\widetilde{\gamma}}_{\overline{\gamma}}\delta^{\overline{\gamma}}_{\overline{\alpha}}\partial^{\overline{\alpha}}_{\alpha} + \partial^{\widetilde{\gamma}}_{\widetilde{\overline{\gamma}}}F^{\widetilde{\overline{\gamma}}}_{\overline{\alpha}}\partial^{\overline{\alpha}}_{\alpha} \\
&= \partial^{\gamma}_{\overline{\gamma}\overline{\beta}}y^{\overline{\beta}}\delta^{\overline{\gamma}}_{\overline{\alpha}}\partial^{\overline{\alpha}}_{\alpha} + \partial^{\gamma}_{\overline{\gamma}}F^{\widetilde{\overline{\gamma}}}_{\overline{\alpha}\overline{\beta}}y^{\overline{\beta}}\partial^{\overline{\alpha}}_{\alpha} \\
&= (\partial^{\gamma}_{\overline{\alpha}\overline{\beta}}\partial^{\overline{\alpha}}_{\alpha} + \partial^{\gamma}_{\overline{\gamma}}F^{\widetilde{\overline{\gamma}}}_{\overline{\alpha}\overline{\beta}}\partial^{\overline{\alpha}}_{\alpha})y^{\overline{\beta}}
\end{aligned}
\tag{2.208}
$$

from which

$$
F^{\widetilde{\gamma}}_{\alpha\beta}\partial^{\beta}_{\overline{\beta}} = \partial^{\gamma}_{\overline{\alpha}\overline{\beta}}\partial^{\overline{\alpha}}_{\alpha} + \partial^{\gamma}_{\overline{\gamma}}F^{\widetilde{\overline{\gamma}}}_{\overline{\alpha}\overline{\beta}}\partial^{\overline{\alpha}}_{\alpha}
\tag{2.209}
$$

thanks to the arbitrarity of y, and we may conclude. ■

As there will be we no risk of ambiguity, we shall reassign $F^\gamma_{\alpha\beta} := F^{\widetilde{\gamma}}_{\alpha\beta}$, and since now, in view of Lemma 2.68, $\widetilde{F}$ may be viewed as restricting to an element of $\Gamma(\tau^* M^{\odot 2} \otimes \tau^* M \otimes \tau M)$ it also makes sense to set $\widetilde{F}_{(\alpha\beta)\gamma}{}^\delta := \widetilde{F}_{(\alpha\beta)\gamma}{}^{\widetilde{\delta}}$. The tensor field $\widetilde{F}$ may now be given the following interpretation: its evaluation against $(U \odot V) \otimes W$ consists of taking the (symmetrisation of the) defect in commutativity between covariant derivatives and horizontal lift, $F\nabla_U V - \widetilde{\nabla}_{FU}FV$ and mapping its vertical part at $W \in TM$ down isomorphically onto TM.

Proposition 2.72. $\boldsymbol{\Phi}_{0\cdot} \in \mathscr{D}_X(\mathcal{L}(\tau M, T_o M))$*, where* $\Phi'^{\gamma^\circ}_{\alpha\beta;0t} \coloneqq -\Phi^{\gamma^\circ}_{\gamma;0t} F^{\gamma}_{\alpha\beta}(X_t)$*.*

Proof. The local condition is satisfied in each coordinate chart thanks to Proposition 2.70. We must check that the compatibility condition of Definition 2.34 is met: again, this is obvious at the trace level, and for Gubinelli derivatives we have, by Lemma 2.71

$$\begin{aligned} \Phi'^{\gamma^\circ}_{\overline{\alpha}\overline{\beta}} &= -\Phi^{\gamma^\circ}_{\overline{\gamma}} F^{\overline{\gamma}}_{\overline{\alpha}\overline{\beta}} \\ &= -\Phi^{\gamma^\circ}_{\overline{\gamma}} F^{\overline{\gamma}}_{\overline{\alpha}\overline{\beta}} \\ &= -\Phi^{\gamma^\circ}_{\gamma} \partial^{\gamma}_{\overline{\gamma}} (\partial^{\overline{\gamma}}_{\delta} \partial^{\alpha}_{\overline{\alpha}} \partial^{\beta}_{\overline{\beta}} F^{\delta}_{\alpha\beta} - \partial^{\overline{\gamma}}_{\delta} \partial^{\delta}_{\overline{\alpha}\overline{\beta}}) \\ &= -\Phi^{\gamma^\circ}_{\gamma} (\partial^{\alpha}_{\overline{\alpha}} \partial^{\beta}_{\overline{\beta}} F^{\gamma}_{\alpha\beta} - \partial^{\gamma}_{\overline{\alpha}\overline{\beta}}) \\ &= \Phi'^{\gamma^\circ}_{\alpha\beta} \partial^{\alpha}_{\overline{\alpha}} \partial^{\beta}_{\overline{\beta}} + \Phi^{\gamma^\circ}_{\gamma} \partial^{\gamma}_{\overline{\alpha}\overline{\beta}} \end{aligned} \tag{2.210}$$

Thus concluding the proof. ■

We now restrict our attention for the last time: **from now on we will consider the case in which F is given by the horizontal lift $\hbar$** unless otherwise stated. This means we are interested in $\widetilde{\nabla}$-differentiating horizontal vector fields w.r.t. horizontal directions, with Condition 2.60 fixing the horizontal part of such covariant derivatives, while Condition 2.67 and the optional Condition 2.63 impose limitations on their vertical part. Here we should remark on the similarity with horizontal connections in sub-Riemannian geometry [CC09, Definition 7.4.1], although our setting is more specific (i.e. not all sub-Riemannian manifolds arise as the total space of a vector or even fibre bundle), and the requirements on the connection is somewhat different (on the one hand we are only interested in $\widetilde{\nabla}_U$ with U horizontal, and on the other also consider the vertical components of such covariant derivatives). In coordinates

$$\begin{aligned} F^{\gamma}_{\alpha\beta} &= -\Gamma^{\gamma}_{\alpha\beta} \\ \widetilde{F}_{\alpha\beta\delta}{}^{\gamma} y^{\delta} &\overset{(\alpha\beta)}{=} -\Gamma^{\gamma}_{\varepsilon\delta}\Gamma^{\varepsilon}_{\alpha\beta} y^{\delta} - (-\Gamma^{\gamma}_{\beta\delta,\alpha} y^{\delta} + \Gamma^{\varepsilon}_{\alpha\delta}\Gamma^{\gamma}_{\beta\varepsilon} y^{\delta} \\ &\quad + \widetilde{\Gamma}^{\widetilde{\gamma}}_{\widetilde{\xi}\widetilde{\eta}} \Gamma^{\xi}_{\alpha\delta} \Gamma^{\eta}_{\beta\varepsilon} y^{\delta} y^{\varepsilon} + \widetilde{\Gamma}^{\widetilde{\gamma}}_{\alpha\beta} - \Gamma^{\xi}_{\alpha\delta} \widetilde{\Gamma}^{\widetilde{\gamma}}_{\widetilde{\xi}\beta} y^{\delta} - \Gamma^{\eta}_{\beta\delta} \widetilde{\Gamma}^{\widetilde{\gamma}}_{\alpha\widetilde{\eta}} y^{\delta}) \end{aligned} \tag{2.211}$$

Note how Lemma 2.71 agrees with (2.59).

We are now in a position to be able to provide the natural generalisation of parallel transport of vectors and Cartan (anti)development to the setting of non-geometric rough paths, with τTM endowed with a linear connection. Since the development of a path is not guaranteed to remain in the manifold for all time, it will be helpful to define the following variations of the rough path spaces (note the use of the double closing parenthesis):

$$\begin{aligned} \mathscr{C}^p_\omega([0,T)], M) &\coloneqq \mathscr{C}^p_\omega([0,T], M) \cup \{\boldsymbol{X} \in \mathscr{C}^p_\omega([0,S), M) \text{ for some } S \leq T \\ &\qquad \text{and } \nexists \text{ compact } K \subseteq M \text{ s.t. } X_{[0,S)} \subseteq K\} \\ \mathscr{C}^p_\omega([0,\leq T)], M) &\coloneqq \mathscr{C}^p_\omega([0,T], M) \cup \bigcup_{0 \leq S \leq T} \mathscr{C}^p_\omega([0,S), M) \end{aligned} \tag{2.212}$$

Note that $\mathscr{C}^p_\omega([0,T)], M) \subseteq \mathscr{C}^p_\omega([0,\leq T)], M)$, and we also will use these notations when M is a vector space. Moreover, we will add a modifier in the rough path sets to denote those rough paths which are started at a specific point. The following notions are defined whenever Condition 2.60 and Condition 2.67 are met.

Definition 2.73. Let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T)], M, o)$. We will denote

$$/\!/(\boldsymbol{X})_{ts} \coloneqq \Phi(\hbar, \boldsymbol{X})_{ts} \colon T_{X_s}M \xrightarrow{\cong} T_{X_t}M \tag{2.213}$$

which by (2.202) is well defined for all s, t at which X is defined, and call it *parallel transport* of vectors along the rough path $\boldsymbol{X}$. We will denote $/\!/_{ts} \coloneqq /\!/(\boldsymbol{X})_{ts}$, $/\!/_t \coloneqq /\!/(\boldsymbol{X})_t \coloneqq /\!/(\boldsymbol{X})_{t0} \colon T_oM \to T_{X_t}M$ and $\backslash\!\backslash_t \coloneqq /\!/(\boldsymbol{X})_{0t} = /\!/_t^{-1}$ when there is no ambiguity as to the rough path.

Remark 2.74 (There is no alternate notion of "backward parallel transport"). A rough path $\boldsymbol{X}$ canonically defines a rough path $\overleftarrow{\boldsymbol{X}} = (\overleftarrow{X}, \overleftarrow{\mathbb{X}})$ above the inverted path $\overleftarrow{X}_t \coloneqq X_{T-t}$. This is done by imposing the Chen identity to hold for all $0 \le s, u, t \le T$ (not just $s \le u \le t$), and results in $\overleftarrow{\mathbb{X}}_{st} = -\mathbb{X}_{T-t,T-s} + X^{\otimes 2}_{T-t,T-s}$ for $0 \le s \le t \le T$. It is shown that if $\boldsymbol{H} \in \mathscr{D}_X$ then $\overleftarrow{\boldsymbol{H}} \in \mathscr{D}_{\overleftarrow{X}}$, where $\overleftarrow{\boldsymbol{H}}_t \coloneqq \boldsymbol{H}_{T-t}$, and that

$$\int_0^T \overleftarrow{\boldsymbol{H}} \mathrm{d}\overleftarrow{\boldsymbol{X}} = -\int_0^T \boldsymbol{H} \mathrm{d}\boldsymbol{X} \tag{2.214}$$

at the trace level. It can then be concluded (by a uniqueness argument) that

$$\begin{cases} \mathrm{d}Y = y_0 + \int F(Y)\mathrm{d}\boldsymbol{X} \\ \mathrm{d}\overleftarrow{Y} = Y_T + \int F(\overleftarrow{Y})\mathrm{d}\overleftarrow{\boldsymbol{X}} \end{cases} \implies \overleftarrow{Y}_t = Y_{T-t} \tag{2.215}$$

which implies that, denoting with Φ the flow map of the RDE defined by $F, \boldsymbol{X}$ and with $\overleftarrow{\Phi}$ the one defined by $F, \overleftarrow{\boldsymbol{X}}$, $\overleftarrow{\Phi} = \Phi^{-1}$. Therefore, once a rough path is fixed, the definition of $\backslash\!\backslash$ given above and the one obtained by defining the parallel transport RDE w.r.t. $\overleftarrow{\boldsymbol{X}}$ coincide.

Definition 2.75. Let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T)], M, o)$. Using Proposition 2.72 we will denote

$$\overline{\bigcirc}(\boldsymbol{X})_\cdot \coloneqq \int_0^\cdot \backslash\!\backslash(\boldsymbol{X})_s \mathrm{d}_\nabla \boldsymbol{X}_s \in \mathscr{C}^p_\omega([0, \le T)], T_oM, 0_o) \tag{2.216}$$

which we call the *antidevelopment* of $\boldsymbol{X}$. If $\boldsymbol{Z} = \overline{\bigcirc}(\boldsymbol{X})$ (up to the time at which $\boldsymbol{X}$ is defined) we will denote $\boldsymbol{X} = \underline{\bigcirc}(\boldsymbol{Z})$ and call $\boldsymbol{X}$ the *development* of $\boldsymbol{Z}$.

In coordinates (2.216) amounts to

$$\mathrm{d}\overline{\bigcirc}^{\gamma^\circ}(\boldsymbol{X}) = \backslash\!\backslash_\gamma^{\gamma^\circ} \mathrm{d}\boldsymbol{X}_t^\gamma + \tfrac{1}{2}\backslash\!\backslash_\gamma^{\gamma^\circ} \Gamma^\gamma_{\alpha\beta} \mathrm{d}[\boldsymbol{X}]^{\alpha\beta} \tag{2.217}$$

For the moment we have only defined development of a rough path which already is the antidevelopment of an M-valued one. If we start from an arbitrary $\boldsymbol{Z} \in \mathscr{C}^p_\omega([0,T], T_oM, 0_o)$ with $Z_0 = 0_o$ we would like to invert Definition 2.75 and define its development as the solution to the path-dependent RDE

$$\mathrm{d}_\nabla \underline{\bigcirc}(\boldsymbol{Z}) = /\!/(\underline{\bigcirc}(\boldsymbol{Z}))\mathrm{d}\boldsymbol{Z}, \quad \underline{\bigcirc}(\boldsymbol{Z})_0 = o \tag{2.218}$$

Heuristically, this means that in an infinitesimal time interval $[t_0, t_0 + \mathrm{d}t]$ we are translating the differential

$\mathrm{d}\boldsymbol{Z}_{t_0} \in T_{Z_{t_0}}T_oM$ so that it is based at the origin 0_o, parallel-transporting it along the already-developed portion of the rough path $\boldsymbol{X}_{[0,t_0]} \coloneqq \underline{\bigcirc}(\boldsymbol{Z})_{[0,t_0]}$ so that it is now based at X_{t_0}, and then using it to "roll T_oM on M along $\boldsymbol{Z}$ without slipping" for time $\mathrm{d}t$. The problem, of course, is that we have not defined such (adaptedly) path-dependent RDEs. Moreover, it should be noted that even once this equation is given a meaning, contrary to the case of parallel transport there is no reason why the solution should not explode (see [Dri18, Corollary 1.36] for general criteria that rule this out for $\boldsymbol{X}$ geometric).

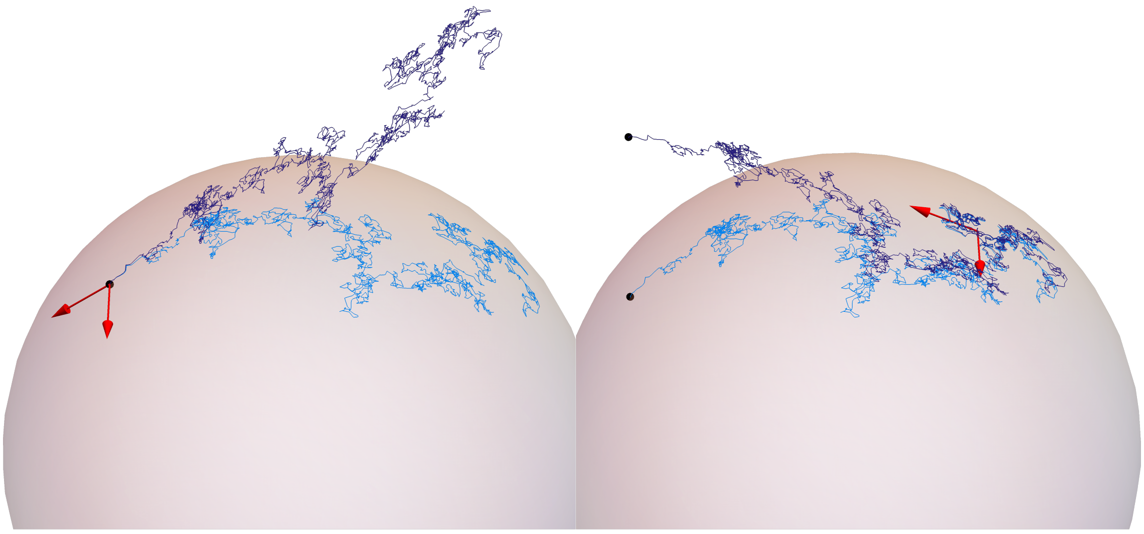


Figure 2.1: A 2-dimensional Brownian path in $T_{(1,0,0)}S^2$, plotted in dark blue, and its Stratonovich development onto S^2, plotted in light blue. The "rolling without slipping" motion is shown at the initial time (on the left) and at a later time (on the right), with the parallel frame based at the point of contact between the tangent space and the manifold.

The trick (already well-known for geometric rough paths) to give (2.218) a meaning is to consider it jointly with a parallel frame: this transforms the path-dependent RDE into a state-dependent one.

Theorem 2.76. *Let $\boldsymbol{Z} \in \mathscr{C}^p_\omega([0,T], T_oM)$. Then $\boldsymbol{X} = \underline{\bigcirc}(\boldsymbol{Z}) \in \mathscr{C}^p_\omega([0,T], M)$ (possibly up to its exit time from M) if and only if $\boldsymbol{X}$ is the unique solution to*

$$\begin{cases} \mathrm{d}/\!/^{\gamma}_{\gamma^\circ} = \hbar^{\gamma}_{\alpha\beta}/\!/^{\beta}_{\gamma^\circ}/\!/^{\alpha}_{\delta^\circ} \circ \mathrm{d}\boldsymbol{Z}^{\delta^\circ} + \frac{1}{2}\widetilde{F}(\widetilde{\nabla},\hbar)_{\alpha\beta\delta}{}^{\gamma}/\!/^{\alpha}_{\alpha^\circ}/\!/^{\beta}_{\beta^\circ}/\!/^{\delta}_{\gamma^\circ}\mathrm{d}[\boldsymbol{Z}]^{\alpha^\circ\beta^\circ} \\ \qquad = -\Gamma^{\gamma}_{\alpha\beta}/\!/^{\beta}_{\gamma^\circ}/\!/^{\alpha}_{\delta^\circ} \circ \mathrm{d}\boldsymbol{Z}^{\delta^\circ} \\ \qquad\quad +\frac{1}{2}\big[-\Gamma^{\gamma}_{\varepsilon\delta}\Gamma^{\varepsilon}_{\alpha\beta}/\!/^{\delta}_{\gamma^\circ} - (-\Gamma^{\gamma}_{\beta\delta,\alpha}/\!/^{\delta}_{\gamma^\circ} + \Gamma^{\varepsilon}_{\alpha\delta}\Gamma^{\gamma}_{\beta\varepsilon}/\!/^{\delta}_{\gamma^\circ} \\ \qquad\quad +\widetilde{\Gamma}^{\widetilde{\gamma}}_{\widetilde{\xi}\widetilde{\eta}}\Gamma^{\xi}_{\alpha\delta}\Gamma^{\eta}_{\beta\varepsilon}/\!/^{\delta}_{\gamma^\circ}/\!/^{\varepsilon}_{\gamma^\circ} + \widetilde{\Gamma}^{\widetilde{\gamma}}_{\alpha\beta} - \Gamma^{\xi}_{\alpha\delta}\widetilde{\Gamma}^{\widetilde{\gamma}}_{\widetilde{\xi}\beta}/\!/^{\delta}_{\gamma^\circ} - \Gamma^{\eta}_{\beta\delta}\widetilde{\Gamma}^{\widetilde{\gamma}}_{\alpha\widetilde{\eta}}/\!/^{\delta}_{\gamma^\circ})\big]/\!/^{\alpha}_{\alpha^\circ}/\!/^{\beta}_{\beta^\circ}\mathrm{d}[\boldsymbol{Z}]^{\alpha^\circ\beta^\circ} \\ \mathrm{d}\boldsymbol{X}^{\gamma} = /\!/^{\gamma}_{\gamma^\circ}\mathrm{d}\boldsymbol{Z}^{\gamma^\circ} - \frac{1}{2}\Gamma^{\gamma}_{\alpha\beta}/\!/^{\alpha}_{\alpha^\circ}/\!/^{\beta}_{\beta^\circ}\mathrm{d}[\boldsymbol{Z}]^{\alpha^\circ\beta^\circ} \end{cases} \tag{2.219}$$

$$X_0 = o, \quad \{/\!/_{\gamma^\circ;0}\}_{\gamma^\circ=1,\ldots,m} \text{ any basis of } T_oM$$

with the Γ's evaluated at X_t and the $\widetilde{\Gamma}$'s evaluated at $/\!/_{\gamma^\circ;t} = /\!/_{\gamma^\circ}(\boldsymbol{X})_t$.

$\underline{\bigcirc}$ therefore defines a surjective map $\mathscr{C}^p_\omega([0,T], T_oM, 0_o) \twoheadrightarrow \mathscr{C}^p_\omega([0,T)], M, o)$ with right inverse $\overline{\bigcirc}\colon \mathscr{C}^p_\omega([0,T)], M, o) \hookrightarrow \mathscr{C}^p_\omega([0, \leq T)], T_oM, 0_o)$ (composed with any map prolonging an element

of $\mathscr{C}^p_\omega([0,S),T_oM,0_o)$ up to time T, e.g. trivially). In particular, if M is compact $\bigcirc\!\!\!\!\_$ takes values in $\mathscr{C}^p_\omega([0,T],M,o)$, i.e. development exists for all time.

If $\boldsymbol{Z}$ is geometric this equation may be stated more elegantly as taking values in the frame bundle $\phi M\colon FM \to M$, and defined by the fundamental horizontal vector fields (2.85), i.e.

$$\mathrm{d}\boldsymbol{Y} = \mathscr{H}_{\lambda^\circ}(Y)\mathrm{d}\boldsymbol{Z}^{\lambda^\circ},\ Y_0 \in F_oM \quad \Longrightarrow \quad \underline{\bigcirc}(\boldsymbol{Z}) = \phi M_* \boldsymbol{Y},\ Y = /\!/(\boldsymbol{X}) \tag{2.220}$$

In this context, compare (2.219) with [Hsu02, (3.3.9) p.86], which is stated in the case of X a Brownian motion, although the formula generalises to more general processes/geometric rough paths. We have decided not to consider frame bundle-valued RDEs in the non-geometric case, since this would require defining a connection on FM, which is a delicate matter (some comments to this effect are provided in [É90, p.439] in the case of the complete lift, though these do not contain an exhaustive description of the connection on τFM). We have preferred to define development in a coordinate-free manner by simply declaring $\boldsymbol{X}$ to be the development of $\boldsymbol{Z}$ if $\boldsymbol{Z}$ is the antidevelopment of $\boldsymbol{X}$ (as done in Definition 2.75), and only relying on the local description involving the parallel frame (seen as m vectors which are parallel-transported individually) as an alternative characterisation, useful for explicit computations; in this approach only parallel transport of vectors is needed.

Proof of Theorem 2.76. By (2.217) $\boldsymbol{X} = \underline{\bigcirc}(\boldsymbol{Z})$ means

$$\begin{aligned} &\mathrm{d}\boldsymbol{Z}^{\gamma^\circ}_t = \backslash\!\backslash^{\gamma^\circ}_\gamma \mathrm{d}\boldsymbol{X}^\gamma + \tfrac{1}{2}\backslash\!\backslash^{\gamma^\circ}_\gamma \Gamma^\gamma_{\alpha\beta}\mathrm{d}[\boldsymbol{X}]^{\alpha\beta}, \quad Z_0 = 0_o \\ \Rightarrow\quad &\mathrm{d}[\boldsymbol{Z}]^{\alpha^\circ\beta^\circ} = \backslash\!\backslash^{\alpha^\circ}_\alpha \backslash\!\backslash^{\beta^\circ}_\beta \mathrm{d}[\boldsymbol{X}]^{\alpha\beta} \end{aligned} \tag{2.221}$$

and we have

$$\begin{aligned} &/\!/^\gamma_{\gamma^\circ}\mathrm{d}\boldsymbol{Z}^{\gamma^\circ} - \tfrac{1}{2}\Gamma^\gamma_{\alpha\beta}/\!/^\alpha_{\alpha^\circ}/\!/^\beta_{\beta^\circ}\mathrm{d}[\boldsymbol{Z}]^{\alpha^\circ\beta^\circ}_t \\ &= /\!/^\gamma_{\gamma^\circ}\backslash\!\backslash^{\gamma^\circ}_\delta \mathrm{d}\boldsymbol{X}^\delta + \tfrac{1}{2}/\!/^\gamma_{\gamma^\circ}\backslash\!\backslash^{\gamma^\circ}_\delta \Gamma^\delta_{\alpha\beta}\mathrm{d}[\boldsymbol{X}]^{\alpha\beta} - \tfrac{1}{2}\Gamma^\gamma_{\alpha\beta}/\!/^\alpha_{\alpha^\circ}/\!/^\beta_{\beta^\circ}\backslash\!\backslash^{\alpha^\circ}_\mu\backslash\!\backslash^{\beta^\circ}_\nu \mathrm{d}[\boldsymbol{X}]^{\mu\nu} \\ &= \mathrm{d}\boldsymbol{X}^\gamma_t \end{aligned} \tag{2.222}$$

By Proposition 2.70 and (2.211) we have

$$\begin{aligned} \mathrm{d}/\!/^\gamma_{\gamma^\circ} &= F^\gamma_{\varepsilon\delta}/\!/^\delta_{\gamma^\circ} \circ \mathrm{d}\boldsymbol{X}^\varepsilon + \tfrac{1}{2}\widetilde{F}_{(\alpha\beta)\delta}{}^\gamma /\!/^\delta_{\gamma^\circ}\mathrm{d}[\boldsymbol{X}]^{\alpha\beta} \\ &= F^\gamma_{\varepsilon\delta}/\!/^\delta_{\gamma^\circ}\mathrm{d}\boldsymbol{X}^\varepsilon + \tfrac{1}{2}\big[\widetilde{F}_{(\alpha\beta)\delta}{}^\gamma + \partial_\alpha F^\gamma_{\beta\delta} + F^\varepsilon_{\alpha\delta}F^\gamma_{\beta\varepsilon}\big]/\!/^\delta_{\gamma^\circ}\mathrm{d}[\boldsymbol{X}]^{\alpha\beta} \\ &= -\Gamma^\gamma_{\varepsilon\delta}/\!/^\delta_{\gamma^\circ}/\!/^\varepsilon_{\varepsilon^\circ}\mathrm{d}\boldsymbol{Z}^{\varepsilon^\circ} \\ &\quad + \tfrac{1}{2}\big[\widetilde{F}_{(\alpha\beta)\delta}{}^\gamma - \Gamma^\gamma_{\beta\delta,\alpha} + \Gamma^\varepsilon_{\alpha\delta}\Gamma^\gamma_{\beta\varepsilon} + \Gamma^\gamma_{\varepsilon\delta}\Gamma^\varepsilon_{\alpha\beta}\big]/\!/^\delta_{\gamma^\circ}/\!/^\alpha_{\alpha^\circ}/\!/^\beta_{\beta^\circ}\mathrm{d}[\boldsymbol{X}]^{\alpha\beta} \\ &= -\Gamma^\gamma_{\varepsilon\delta}/\!/^\delta_{\gamma^\circ}/\!/^\varepsilon_{\varepsilon^\circ} \circ \mathrm{d}\boldsymbol{Z}^{\varepsilon^\circ} + \tfrac{1}{2}\widetilde{F}_{(\alpha\beta)\delta}{}^\gamma /\!/^\delta_{\gamma^\circ}/\!/^\alpha_{\alpha^\circ}/\!/^\beta_{\beta^\circ}\mathrm{d}[\boldsymbol{X}]^{\alpha\beta} \end{aligned} \tag{2.223}$$

where in the last step the Gubinelli derivative of $-\Gamma^\gamma_{\beta\delta}(X)/\!/^\delta_{\gamma^\circ}/\!/^\varepsilon_{\varepsilon^\circ}$ w.r.t. Z^α is computed thanks to the previous step and (2.222). Retracing these steps proves the converse. Note that we do not need to show the coordinate invariance of (2.219), as we have shown it is equivalent to $\boldsymbol{Z} = \overline{\bigcirc}(\boldsymbol{X})$, which is defined in Definition 2.75 without reference to a coordinate system.

The map $\underline{\bigcirc}$ is then well defined by uniqueness of RDE solutions applied to the $(m+m^2)$-dimensional system in each coordinate patch, and its right inverse is $\overline{\bigcirc}$ by definition. It only remains to show that $\underline{\bigcirc}(\boldsymbol{Z})$ is either defined up to time T or that it is defined up to and excluding some $S \leq T$ with the image of its trace not contained in any compact of M. Assume $(X, /\!/)$ is defined up to time S with $X_{[0,S)}$ contained in a compact K of M. Therefore there exists $t_n \searrow S$ s.t. $\lim X_{t_n} = \overline{x} \in K$. We now show that for any neighbourhood V of $\overline{x}$ there exists s_0 s.t. $X_{[s_0,S)} \subseteq V$. Consider the image of (2.219) (defined in V) through a change of coordinates Φ that maps the $/\!/$ components to a compact, and extend the resulting coefficients smoothly. Now picking a second neighbourhood U of $\overline{x}$ s.t. $\mathrm{Im}(U) \subseteq U, \overline{U} \subseteq V$, an application of Lemma 2.22 proves the claim by picking s_0 s.t. $s_0 < S < s_0 + \delta$. (The change of coordinates was necessary because we need to be able to start the equation for $(X, /\!/)$ at an arbitrary point in TU^m.) We may then reason as in the proof of Theorem 2.69 to conclude that $(X, /\!/)_{[0,S)}$ must also lie in a compact of TM^m, and a second application of Lemma 2.22 (arguing as in [CDL15, Theorem 4.2]) then shows that the solution may be prolonged past S (or with its limit if $S = T$). This concludes the proof. ∎

The following result is proven in [É89, Theorem 8.22] in the case of Stratonovich parallel transport, and, interestingly, it carries over to the non-geometric case.

Corollary 2.77. *At the trace level we have*

$$\overline{\bigcirc}(\boldsymbol{X})_{\cdot} \coloneqq \int_0^{\cdot} \backslash\!\backslash_s \circ \mathrm{d}\boldsymbol{X}_s \tag{2.224}$$

and we may replace

$$\mathrm{d}X^\gamma = /\!/^\gamma_{\gamma^\circ} \circ \mathrm{d}\boldsymbol{Z}^{\gamma^\circ} \tag{2.225}$$

for the second equation of (2.219).

Remark 2.78. We emphasise that this does not mean that the (anti)development of a rough path coincides with that of its geometrisation (including at the trace level): in (2.224) parallel transport is still carried out with reference to the original non-geometric $\boldsymbol{X}$ (and thus depends on the choice of $\widetilde{\nabla}$), and in the case of development, the first equation of (2.219) still has the $\mathrm{d}[\boldsymbol{Z}]$ terms, which are not present when developing ${}_{\mathrm{g}}\boldsymbol{Z}$. Moreover, at the second order level $\mathbb{X}^{\alpha\beta}_{st} \approx /\!/^{\alpha}_{\alpha^\circ;s} /\!/^{\beta}_{\beta^\circ;s} \mathbb{Z}^{\alpha^\circ\beta^\circ}_{st}$ locally in terms of the original rough path $\boldsymbol{Z}$.

Proof of Corollary 2.77. By Proposition 2.70 and (2.217) we have, at the trace level

$$\backslash\!\backslash^{\gamma^\circ}_{\gamma} \circ \mathrm{d}\boldsymbol{X} = \backslash\!\backslash^{\gamma^\circ}_{\gamma} \mathrm{d}\boldsymbol{X}^\gamma_t + \tfrac{1}{2} \backslash\!\backslash^{\gamma^\circ}_{\gamma} \Gamma^\gamma_{\alpha\beta} \mathrm{d}[\boldsymbol{X}]^{\alpha\beta} = \backslash\!\backslash^{\gamma^\circ}_{\gamma} \mathrm{d}_\nabla \boldsymbol{X}^\gamma \tag{2.226}$$

and the second claim is proved analogously by using (2.219). ∎

If M is Riemannian and ∇ is metric we may further ask under what hypotheses the $/\!/_{ts}$'s are linear isometries $T_{X_s}M \cong T_{X_t}M$. The following condition does not actually require F to be given by horizontal lift, although we will only apply it in that case.

Condition 2.79. Let g be a Riemannian metric on M and ∇ be g-metric: $\langle \widetilde{F}, U \otimes U \otimes V \rangle \in \Gamma\tau M$ is g-orthogonal to V for all $U, V \in \Gamma\tau M$.

Note that we are not requiring $\widetilde{\nabla}$ to be metric w.r.t. a Riemannian metric on the manifold TM. The statement of this condition in coordinates is given in the following lemma, whose proof is immediate by polarisation.

Lemma 2.80. *In coordinates Condition 2.79 corresponds to*

$$\widetilde{F}_{(\alpha\beta)(\gamma\delta)} = 0 \tag{2.227}$$

Theorem 2.81. *If Condition 2.79 holds, or if $\boldsymbol{X}$ is geometric, $/\!/(\boldsymbol{X})_{ts}$ is a linear isometry for all $0 \le s, t \le T$.*

The following pattern has emerged: for each property (well-definedness, linearity, and metricity, each required at the level of generality considered) we have a first-order condition (respectively (2.173), F linear, and ∇ g-metric — as shall be seen in the proof below) and a second order condition (respectively Condition 2.60, Condition 2.67, Condition 2.79). The first-order conditions are necessary when considering the geometric (or even smooth) case, whereas the second-order conditions become relevant once the driving rough path is no longer geometric. All three conditions are automatically satisfied when Condition 2.63 holds.

Proof of Theorem 2.81. We may assume $s = 0$; then for $y, z \in T_o M$ by Proposition 2.70 we have

$$\begin{aligned}
&\mathrm{d}\langle g(X), /\!/_{\alpha^\circ} \otimes /\!/_{\beta^\circ}\rangle \\
&= \mathrm{d}(g_{\alpha\beta} /\!/^{\alpha}_{\alpha^\circ} /\!/^{\beta}_{\beta^\circ}) \\
&= g_{\alpha\beta,\gamma} /\!/^{\alpha}_{\alpha^\circ} /\!/^{\beta}_{\beta^\circ} \circ \mathrm{d}\boldsymbol{X}^\gamma \\
&\quad + g_{\alpha\beta} /\!/^{\delta}_{\alpha^\circ} /\!/^{\beta}_{\beta^\circ} (-\Gamma^{\alpha}_{\gamma\delta} \circ \mathrm{d}\boldsymbol{X}^\gamma + \tfrac{1}{2} \widetilde{F}_{(\xi\eta)\delta}{}^{\alpha} \mathrm{d}[\boldsymbol{X}]^{\xi\eta}) \\
&\quad + g_{\alpha\beta} /\!/^{\alpha}_{\alpha^\circ} /\!/^{\delta}_{\beta^\circ} (-\Gamma^{\beta}_{\gamma\delta} \circ \mathrm{d}\boldsymbol{X}^\gamma + \tfrac{1}{2} \widetilde{F}_{(\xi\eta)\delta}{}^{\beta} \mathrm{d}[\boldsymbol{X}]^{\xi\eta}) \\
&= /\!/^{\alpha}_{\alpha^\circ} /\!/^{\beta}_{\beta^\circ} \big[(g_{\alpha\beta,\gamma} - g_{\alpha\delta}\Gamma^{\delta}_{\gamma\beta} - g_{\delta\beta}\Gamma^{\delta}_{\gamma\alpha}) \circ \mathrm{d}\boldsymbol{X}^\gamma + \widetilde{F}_{(\xi\eta)(\alpha\beta)} \mathrm{d}[\boldsymbol{X}]^{\xi\eta}\big]
\end{aligned} \tag{2.228}$$

which vanishes by metricity of ∇, (2.61) and Condition 2.79 or by vanishing of the bracket in the case of $\boldsymbol{X}$ geometric (note how by Theorem 2.6 the hypotheses are sharp in the case of X truly rough). ∎

We will now provide three examples of connections $\widetilde{\nabla}$ on τM which it makes sense to consider. The first two, for which we refer to [YI73], can be viewed as "lifts" of the connection ∇ (which is not assumed to be metric or torsion-free), while the third consists of assuming M is Riemannian, defining a Riemannian metric on the manifold TM, and taking its Levi-Civita connection.

Example 2.82 (The complete lift of ∇)**.** The *complete lift* $\widetilde{\nabla}$ of ∇ is the linear connection on τTM whose Christoffel symbols in induced coordinates are given as functions of the Christoffel symbols Γ^k_{ij} of ∇ w.r.t. to φ as follows:

$$\begin{aligned}
&\widetilde{\Gamma}^{\gamma}_{\alpha\beta}(x,y) = \Gamma^{\gamma}_{\alpha\beta}(x), \quad \widetilde{\Gamma}^{\gamma}_{\alpha\widetilde{\beta}}(x,y) = \widetilde{\Gamma}^{\gamma}_{\widetilde{\alpha}\beta}(x,y) = \widetilde{\Gamma}^{\gamma}_{\widetilde{\alpha}\widetilde{\beta}}(x,y) = 0 \\
&\widetilde{\Gamma}^{\widetilde{\gamma}}_{\alpha\beta}(x,y) = \partial_\lambda \Gamma^{\gamma}_{\alpha\beta}(x) y^\lambda, \quad \widetilde{\Gamma}^{\widetilde{\gamma}}_{\alpha\widetilde{\beta}}(x,y) = \Gamma^{\gamma}_{\alpha\beta}(x), \quad \widetilde{\Gamma}^{\widetilde{\gamma}}_{\widetilde{\alpha}\beta}(x,y) = \Gamma^{\gamma}_{\alpha\beta}(x), \quad \widetilde{\Gamma}^{\widetilde{\gamma}}_{\widetilde{\alpha}\widetilde{\beta}}(x,y) = 0
\end{aligned} \tag{2.229}$$

From these and Example 2.30 it follows that τM is an affine map w.r.t. $\widetilde{\nabla}$, ∇. This connection admits the following simple description: having defined the complete lift of $V \in \Gamma\tau M$ as $\widetilde{V} \in \Gamma\tau TM$ given in induced coordinates by

$$\widetilde{V}^{\gamma}(x,y) \coloneqq V^{\gamma}(x), \quad \widetilde{V}^{\widetilde{\gamma}}(x,y) \coloneqq y^\lambda \partial_\lambda V^\gamma(x) \tag{2.230}$$

(this is checked to be a sound definition; note that no further connection is needed to perform this lift) $\widetilde{\nabla}$ is characterised by the condition

$$\widetilde{\nabla}_{\widetilde{U}}\widetilde{V} = \widetilde{\nabla_U V}, \quad U, V \in \Gamma\tau M \tag{2.231}$$

We will only need the local description of $\widetilde{\nabla}$. However, we remark that the complete lift can be extended to tensor fields, and in particular to Riemannian metrics g, thus yielding a pseudo-Riemannian metric $\widetilde{g}$ on TM (with metric signature (m, m)) whose components are given by

$$\begin{pmatrix} \widetilde{g}_{\alpha\beta} & g_{\alpha\widetilde{\beta}} \\ g_{\widetilde{\alpha}\beta} & g_{\widetilde{\alpha}\widetilde{\beta}} \end{pmatrix}(x, y) = \begin{pmatrix} \partial_\lambda g_{\alpha\beta}(x) y^\lambda & g_{\alpha\beta}(x) \\ g_{\alpha\beta}(x) & 0 \end{pmatrix} \tag{2.232}$$

If ∇ is g-metric, then $\widetilde{\nabla}$ is $\widetilde{g}$-metric, and if ∇ is torsion-free then so is $\widetilde{\nabla}$; therefore $\widetilde{{}^{g}\nabla} = {}^{\widetilde{g}}\nabla$. In general, $\widetilde{\nabla}$ has the property that its geodesics are given by the Jacobi fields of ∇.

It is easily checked using the theory in this section that Condition 2.60 and Condition 2.67 are satisfied for all F in the case of the complete lift, and in the case of parallel transport with ∇ torsion-free we have

$$\widetilde{F}_{\alpha\beta\delta}{}^{\gamma} = \mathscr{R}_{\alpha\delta\beta}{}^{\gamma} \tag{2.233}$$

Condition 2.79, however, is not satisfied even when ∇ is Levi-Civita, since

$$\begin{aligned} \widetilde{F}_{(\alpha\beta)(\delta\gamma)} &= \tfrac{1}{4}(\mathscr{R}_{\alpha\delta\beta\gamma} + \mathscr{R}_{\beta\delta\alpha\gamma} + \mathscr{R}_{\alpha\gamma\beta\delta} + \mathscr{R}_{\beta\gamma\alpha\delta}) \\ &= \tfrac{1}{2}(\mathscr{R}_{\alpha\delta\beta\gamma} + \mathscr{R}_{\alpha\gamma\beta\delta}) \end{aligned} \tag{2.234}$$

which does not vanish in general. The resulting parallel transport equation was first studied, for semimartingales, in [DG78] and subsequently in [Mey82, (27)] (we caution the reader that the convention regarding the indices of the curvature tensor differ from the ones used in (2.65)), and it was realised in [É90, p.437] that this type of parallel transport fits into the framework of SDEs of the type defined in Definition 2.42.

Example 2.83 (The horizontal lift of ∇)**.** The second lift of a connection which we examine is the *horizontal lift* of ∇, which we also denote $\widetilde{\nabla}$ (ambiguity will easily be avoided, since we will always use each connection separately). Its Christoffel symbols in induced coordinates are similar to those of the complete lift, with one important difference:

$$\begin{aligned} &\widetilde{\Gamma}^{\gamma}_{\alpha\beta}(x,y) = \Gamma^{\gamma}_{\alpha\beta}(x), \quad \widetilde{\Gamma}^{\gamma}_{\alpha\widetilde{\beta}}(x,y) = \widetilde{\Gamma}^{\gamma}_{\widetilde{\alpha}\beta}(x,y) = \widetilde{\Gamma}^{\gamma}_{\widetilde{\alpha}\widetilde{\beta}}(x,y) = \widetilde{\Gamma}^{\widetilde{\gamma}}_{\widetilde{\alpha}\widetilde{\beta}}(x,y) = 0 \\ &\widetilde{\Gamma}^{\widetilde{\gamma}}_{\alpha\beta}(x,y) = (\partial_\lambda \Gamma^{\gamma}_{\alpha\beta} - \mathscr{R}_{\lambda\alpha\beta}{}^{\gamma})(x) y^\lambda, \; \widetilde{\Gamma}^{\widetilde{\gamma}}_{\alpha\widetilde{\beta}}(x,y) = \Gamma^{\gamma}_{\alpha\beta}(x), \; \widetilde{\Gamma}^{\widetilde{\gamma}}_{\widetilde{\alpha}\beta}(x,y) = \Gamma^{\gamma}_{\alpha\beta}(x) \end{aligned} \tag{2.235}$$

As for the complete lift, τM is an affine map w.r.t. $\widetilde{\nabla}, \nabla$; the extra term appearing in $\widetilde{\Gamma}^{\widetilde{\gamma}}_{\alpha\beta}$, however, causes $\widetilde{\nabla}$ to have non-vanishing torsion in general even if ∇ is torsion-free. Just as for the complete lift, the horizontal lift of a connection is motivated by a broader construction which involves lifting other objects defined on M, such as vector fields. However, unlike the case of the complete lift, these lifts require a connection on τM to begin with, and are performed in a way which is related to (2.77); we do not provide more details here. If ∇ is g-metric, then $\widetilde{\nabla}$ is $\widetilde{g}$-metric, where $\widetilde{g}$ is the pseudo-Riemannian metric (2.232) (although, unlike the case of the complete lift, $\widetilde{{}^{g}\nabla} \neq {}^{\widetilde{g}}\nabla$ because the former has torsion in general). The characterisation of geodesics of the horizontal lift of a connection is more complicated than that of its complete lift, but it still holds that τM maps $\widetilde{\nabla}$-geodesics to ∇-geodesics. Moreover, it holds that horizontal lifts of geodesics (namely curves in TM

above geodesics whose tangent vectors are horizontal, i.e. parallel transports above geodesics) define geodesics w.r.t. the horizontal lift: this is seen from [YI73, (9.4) p.115].

Like the complete lift, the horizontal lift results in Condition 2.60 and Condition 2.67 being satisfied for all F, but in the case of F given by horizontal lift it additionally satisfies Condition 2.63. Therefore the resulting parallel transport is, at the trace level, the same as geometric/Stratonovich parallel transport, a conclusion which is also noted in [Mey82, É90].

Example 2.84 (The Sasaki metric)**.** Let g be a Riemannian metric on M. We can lift g to a Riemannian metric $\widetilde{g}$ on TM, called the *Sasaki metric* in the following way: recalling the notations introduced in Subsection 2.2.1 for vertical and horizontal bundles, we declare for all $U(x) \in T_x M$

$$V_{U(x)}\tau M \perp H_{U(x)}, \quad \widetilde{g}|_{V_{U(x)}\tau M} \coloneqq (v(U(x))^{-1})^* g, \quad \widetilde{g}|_{H_{U(x)}} \coloneqq (\hbar(U(x))^{-1})^* g \tag{2.236}$$

In induced coordinates, this is amounts to

$$\begin{pmatrix} \widetilde{g}_{\alpha\beta} & \widetilde{g}_{\alpha\widetilde{\beta}} \\ \widetilde{g}_{\widetilde{\alpha}\beta} & \widetilde{g}_{\widetilde{\alpha}\widetilde{\beta}} \end{pmatrix}(x,y) = \begin{pmatrix} g_{\alpha\beta}(x) + g_{\delta\varepsilon}\Gamma^{\delta}_{\mu\alpha}\Gamma^{\varepsilon}_{\nu\beta}(x)y^\mu y^\nu & \Gamma^{\gamma}_{\alpha\lambda} g_{\gamma\beta}(x)y^\lambda \\ \Gamma^{\gamma}_{\lambda\beta} g_{\alpha\gamma}(x)y^\lambda & g_{\alpha\beta}(x) \end{pmatrix} \tag{2.237}$$

and

$$\begin{pmatrix} \widetilde{g}^{\alpha\beta} & \widetilde{g}^{\alpha\widetilde{\beta}} \\ \widetilde{g}^{\widetilde{\alpha}\beta} & \widetilde{g}^{\widetilde{\alpha}\widetilde{\beta}} \end{pmatrix}(x,y) = \begin{pmatrix} g^{\alpha\beta}(x) & -\Gamma^{\beta}_{\lambda\gamma} g^{\alpha\gamma}(x)y^\lambda \\ -\Gamma^{\alpha}_{\gamma\lambda} g^{\gamma\beta}(x)y^\lambda & g^{\alpha\beta}(x) + g^{\delta\varepsilon}\Gamma^{\alpha}_{\delta\mu}\Gamma^{\beta}_{\varepsilon\nu}(x)y^\mu y^\nu \end{pmatrix} \tag{2.238}$$

where the Γ's are the Christoffel symbols of ${}^g\nabla$. The horizontal lift of ${}^g\nabla$ is $\widetilde{g}$-metric, but does not coincide with ${}^{\widetilde{g}}\nabla$ due to torsion. We will call ${}^{\widetilde{g}}\nabla$ the *Sasaki lift* of ${}^g\nabla$ (even though, strictly speaking, it is the metric that we are lifting). The Christoffel symbols of ${}^{\widetilde{g}}\nabla$ in induced coordinates have more complex expressions than the ones for the other two lifts of connections, and are given as functions of the Christoffel symbols of ${}^g\nabla$ and of the components of its curvature tensor by

$$\begin{aligned}
\widetilde{\Gamma}^{\gamma}_{\alpha\beta}(x,y) &= \Gamma^{\gamma}_{\alpha\beta}(x) + \tfrac{1}{2}(\mathscr{R}_{\mu\delta\alpha}{}^{\gamma}\Gamma^{\delta}_{\lambda\beta} + \mathscr{R}_{\mu\delta\beta}{}^{\gamma}\Gamma^{\delta}_{\alpha\lambda})(x)y^\lambda y^\mu \\
\widetilde{\Gamma}^{\gamma}_{\widetilde{\alpha}\beta}(x,y) &= \tfrac{1}{2}\mathscr{R}_{\lambda\alpha\beta}{}^{\gamma}(x)y^\lambda, \quad \widetilde{\Gamma}^{\gamma}_{\alpha\widetilde{\beta}}(x,y) = \tfrac{1}{2}\mathscr{R}_{\lambda\beta\alpha}{}^{\gamma}(x)y^\lambda, \quad \widetilde{\Gamma}^{\gamma}_{\widetilde{\alpha}\widetilde{\beta}}(x,y) = \Gamma^{\widetilde{\gamma}}_{\widetilde{\alpha}\widetilde{\beta}}(x,y) = 0 \\
\widetilde{\Gamma}^{\widetilde{\gamma}}_{\alpha\beta}(x,y) &= \tfrac{1}{2}(\mathscr{R}_{\alpha\lambda\beta}{}^{\gamma} + \mathscr{R}_{\beta\lambda\alpha}{}^{\gamma} + 2\partial_\lambda\Gamma^{\gamma}_{\alpha\beta})(x)y^\lambda \\
&\quad + \tfrac{1}{2}\Gamma^{\gamma}_{\nu\delta}(\mathscr{R}_{\varepsilon\mu\alpha}{}^{\delta}\Gamma^{\varepsilon}_{\lambda\beta} + \mathscr{R}_{\varepsilon\mu\beta}{}^{\delta}\Gamma^{\varepsilon}_{\alpha\lambda})(x)y^\lambda y^\mu y^\nu \\
\widetilde{\Gamma}^{\widetilde{\gamma}}_{\widetilde{\alpha}\beta}(x,y) &= \Gamma^{\gamma}_{\alpha\beta}(x) - \tfrac{1}{2}\Gamma^{\gamma}_{\mu\delta}\mathscr{R}_{\lambda\alpha\beta}{}^{\delta}(x)y^\lambda y^\mu, \\
\widetilde{\Gamma}^{\widetilde{\gamma}}_{\alpha\widetilde{\beta}}(x,y) &= \Gamma^{\gamma}_{\alpha\beta}(x) - \tfrac{1}{2}\Gamma^{\gamma}_{\mu\delta}\mathscr{R}_{\lambda\beta\alpha}{}^{\delta}(x)y^\lambda y^\mu
\end{aligned} \tag{2.239}$$

These symbols are taken from [Sas58] with one important caveat: the $\mathscr{R}_{\alpha\beta\gamma}{}^{\delta}$'s therein have been transcribed into $\mathscr{R}_{\gamma\beta\alpha}{}^{\delta}$'s here. This is because the author follows a different ordering in the coordinate expression of the curvature tensor. This convention is not stated in the paper, but it can be deduced by computing any one of the Christoffel symbols involving a curvature term. This check can be performed by using the fact that the horizontal lift of ∇ is $\widehat{g}$-metric and (2.63). Let $\widetilde{\mathscr{T}}$ denote the torsion tensor of the horizontal lift of ∇: its only non-zero component is given by

$$\widetilde{\mathscr{T}}{}^{\widetilde{\gamma}}_{\alpha\beta}(x,y) = (\mathscr{R}_{\lambda\beta\alpha}{}^{\gamma} - \mathscr{R}_{\lambda\alpha\beta}{}^{\gamma})(x)y^\lambda \tag{2.240}$$

Thus $\widetilde{\mathcal{T}}_{\alpha}{}^{\gamma}{}_{\widetilde{\beta}}(x, y) = 0$ and, performing index gymnastics w.r.t. $\widetilde{g}$ and using (2.66), (2.67), (2.68) and (2.69) we obtain

$$\begin{aligned} \widetilde{\mathcal{T}}_{\widetilde{\alpha}}{}^{\gamma}{}_{\beta}(x, y) &= \widehat{g}_{\widetilde{\alpha}\widetilde{\varepsilon}}\widehat{g}^{\gamma\delta}(x, y)(\mathcal{R}_{\lambda\beta\delta}{}^{\varepsilon} - \mathcal{R}_{\lambda\delta\beta}{}^{\varepsilon})(x)y^{\lambda} \\ &= g_{\alpha\varepsilon}g^{\gamma\delta}(\mathcal{R}_{\lambda\beta\delta}{}^{\varepsilon} - \mathcal{R}_{\lambda\delta\beta}{}^{\varepsilon})(x)y^{\lambda} \\ &= g_{\alpha\varepsilon}g^{\gamma\delta}(\mathcal{R}_{\lambda\beta\delta}{}^{\varepsilon} + \mathcal{R}_{\delta\lambda\beta}{}^{\varepsilon})(x)y^{\lambda} \\ &= -g_{\alpha\varepsilon}g^{\gamma\delta}\mathcal{R}_{\beta\delta\lambda}{}^{\varepsilon}(x)y^{\lambda} \\ &= -g^{\gamma\delta}\mathcal{R}_{\beta\delta\lambda\alpha}(x)y^{\lambda} \\ &= -g^{\gamma\delta}\mathcal{R}_{\lambda\alpha\beta\delta}(x)y^{\lambda} \\ &= -\mathcal{R}_{\lambda\alpha\beta}{}^{\gamma}(x)y^{\lambda} \end{aligned} \tag{2.241}$$

Then, since $\widetilde{\mathcal{K}}^{\gamma}{}_{\widetilde{\alpha}\beta} = \frac{1}{2}\widetilde{\mathcal{T}}_{\widetilde{\alpha}}{}^{\gamma}{}_{\beta}$ (where $\mathcal{K}$ is the contorsion tensor defined in (2.63)), (2.235) yield the value of $\widetilde{\Gamma}^{\gamma}_{\widetilde{\alpha}\beta}$ in (2.239). Similarly to the case of the horizontal lift of a connection, the horizontal lift of a Riemannian geodesic is a geodesic w.r.t. Sasaki metric.

"Sasaki parallel transport" has not, to our knowledge, been considered in the literature. Like for the horizontal lift, Condition 2.63 is satisfied w.r.t. to the Sasaki lift of ∇ when F is given by horizontal lift, and the resulting definition of parallel transport is therefore equivalent to the geometrised one. However, unlike the complete and horizontal lifts, Condition 2.60 cannot be expected to hold for general F, since the Sasaki lift does not make τM symmetrically affine: this can be seen, again with reference to Example 2.30, by noting that, for instance $\widetilde{\Gamma}^{\gamma}_{\widetilde{\alpha}\beta}(x, y) = \frac{1}{2}\mathcal{R}_{\lambda\alpha\beta}{}^{\gamma}(x)y^{\lambda}$ is not, in general, antisymmetric in α, β. This means we may not, in general, define arbitrary equations (2.172) when $E = TM$ is given the Sasaki lift of ∇.

Example 2.85 (Local martingales and Brownian motion)**.** It is well known that Stratonovich (anti)development preserves local martingales and if M is Riemannian it preserves Brownian motions. In our setting the first statement always holds in all cases (assuming Condition 2.60 and Condition 2.67 hold, so that (anti)development is defined), as can be easily seen from the local characterisation of manifold-valued martingales (2.130), and the local expressions (2.217), (2.219). This is also observed (by a different argument) in [É90, p.440].

As for the preservation of Brownian motion, we first recall that the Levy criterion on manifolds [É89, Proposition 5.18] immediately implies the following local characterisation of Brownian motion of a Riemannian manifold (M, g): X is a Brownian motion on M if and only if it is a local martingale and

$$\mathrm{d}[X]^{\alpha\beta} = g^{\alpha\beta}(X)\mathrm{d}t \tag{2.242}$$

If Z is a T_oM-valued Brownian motion, then if Condition 2.79 holds, we have for $X = \mathrm{d}\underset{\sim}{\bigcirc}(Z)$

$$\mathrm{d}[X]^{\alpha\beta} = /\!/^{\alpha}_{\alpha^\circ}/\!/^{\beta}_{\beta^\circ}\mathrm{d}[Z]^{\alpha^\circ\beta^\circ} = /\!/^{\alpha}_{\alpha^\circ}/\!/^{\beta}_{\beta^\circ}\delta^{\alpha^\circ\beta^\circ}\mathrm{d}t = g^{\alpha\beta}(X)\mathrm{d}t \tag{2.243}$$

where the last identity holds thanks to the fact that $g^{\alpha^\circ\beta^\circ}(o) = \delta^{\alpha^\circ\beta^\circ}$ and Theorem 2.81. That antidevelopment maps Brownian motions to Brownian motions under the same hypotheses is checked analogously.

We may therefore conclude that (anti)development defined w.r.t. the complete, horizontal and Sasaki lifts to preserve local martingales, but only that defined w.r.t. the horizontal and Sasaki lifts to preserve Brownian motion.

We also note that we should expect (anti)development taken w.r.t. two different $\widetilde{\nabla}$'s to be different pathwise, even when both satisfy the linearity and metricity conditions. For Brownian motion this might mean that the law of the (anti)developments coincide (i.e. they are both Brownian motions), despite the paths defined by the same state $\omega \in \Omega$ being different. Another way of generating pathwise-distinct Brownian motions through (anti)development of the same Brownian motion is by adding a contorsion term (see Remark 2.27) to the Levi-Civita connection ∇ and taking Stratonovich development. In general, by the Itô isometry the cross-covariance matrix of the Itô antidevelopments ${}^1\overline{\bigcirc}(X)$ and ${}^2\overline{\bigcirc}(X)$ of the same M-valued semimartingale X taken w.r.t. ${}^1\nabla, {}^1\widetilde{\nabla}$ on the one hand and ${}^2\nabla, {}^2\widetilde{\nabla}$ on the other is given by $E[{}^1\overline{\bigcirc}^{\alpha^\circ}(X){}^2\overline{\bigcirc}^{\beta^\circ}(X)] = E[\int {}^1\backslash\!\backslash^{\alpha^\circ}_\alpha {}^2\backslash\!\backslash^{\beta^\circ}_\beta \mathrm{d}[X]^{\alpha\beta}]$, with ${}^k\backslash\!\backslash, k = 1, 2$ denoting the respective parallel transports above X.

Example 2.86 ($/\!/$ along Brownian motion on Einstein manifolds w.r.t. the complete lift)**.** We assume (M, g) is an Einstein manifold, i.e. a Riemannian manifold whose Ricci tensor is proportional to the metric tensor, $\mathscr{R}_{\alpha\beta} = \lambda g_{\alpha\beta}$ with $\lambda \in \mathbb{R}$ (the best known such example is the sphere, in all dimensions). Let Z be a Brownian motion on T_oM and X its Stratonovich development, an M-valued Brownian motion, and we compare the behaviour of Stratonovich parallel transport $/\!/(X)$ with parallel transport defined w.r.t. to the complete lift $\widetilde{\nabla}$ of the Levi-Civita connection ∇, which we denote $\widetilde{/\!/}(X)$. By proceeding as in the proof ofTheorem 2.81 and Example 2.85 we compute

$$\begin{aligned} \mathrm{d}g(/\!/_{\alpha^\circ}, \widetilde{/\!/}_{\beta^\circ}) &= \tfrac{1}{2} g_{\alpha\beta} /\!/^\alpha_{\alpha^\circ} \mathscr{R}_{\xi\gamma\eta}{}^\beta \widetilde{/\!/}^\gamma_{\beta^\circ} g^{\xi\eta} \mathrm{d}t \\ &= -\tfrac{1}{2} \mathscr{R}_{\alpha\beta} /\!/^\alpha_{\alpha^\circ} \widetilde{/\!/}^\beta_{\beta^\circ} \mathrm{d}t \\ &= -\tfrac{\lambda}{2} g(/\!/_{\alpha^\circ}, \widetilde{/\!/}_{\beta^\circ}) \mathrm{d}t \end{aligned} \tag{2.244}$$

which implies $g(/\!/_{\alpha^\circ}, \widetilde{/\!/}_{\beta^\circ}) = \exp(-\lambda t/2)\delta^{\alpha^\circ\beta^\circ}$, and similarly $g(\widetilde{/\!/}_{\alpha^\circ}, \widetilde{/\!/}_{\beta^\circ}) = \exp(-\lambda t)\delta^{\alpha^\circ\beta^\circ}$. In other words $\widetilde{/\!/}(X)$ preserves orthogonality, but not orthonormality, since it consists of a scaling by the above exponential factor. Note that this behaviour of $\widetilde{/\!/}$ can only be expected to hold along the Brownian motion X, and not along $\widetilde{X} := \widetilde{\bigcirc}(Z)$, the development of Z taken according to the complete lift of $\widetilde{\nabla}$, which is not in general a Brownian motion (even given the Einstein assumption): this can be seen by writing $\mathrm{d}\widetilde{X} = \sum_{\gamma^\circ} \widetilde{/\!/}^\alpha_{\gamma^\circ}(\widetilde{X})\widetilde{/\!/}^\beta_{\gamma^\circ}(\widetilde{X})\mathrm{d}t$ and by showing that the SDE satisfied by $\sum_{\gamma^\circ} \widetilde{/\!/}^\alpha_{\gamma^\circ}(\widetilde{X})\widetilde{/\!/}^\beta_{\gamma^\circ}(\widetilde{X})$ has an extra drift term when compared to that satisfied by $g^{\alpha\beta}(\widetilde{X})$.

Example 2.87 (Linearising rough integrals and rewriting Driver's integration by parts formula)**.** Antidevelopment can be used to write rough integrals against M-valued rough paths as ones against T_oM-valued ones. Let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0, T], M, o)$ and $\boldsymbol{H} \in \mathscr{D}_X(\mathcal{L}(\tau M, \mathbb{R}^e))$. Then it is checked that

$$\begin{aligned} /\!/^* \boldsymbol{H} &= ((/\!/^* H)^c_{\gamma^\circ}, (/\!/^* H)'^c_{\alpha^\circ\beta^\circ}) := (H^c_\gamma /\!/^\gamma_{\gamma^\circ}, H'^c_{\alpha\beta} /\!/^\alpha_{\alpha^\circ} /\!/^\beta_{\beta^\circ} - H^c_\gamma \Gamma^\gamma_{\alpha\beta} /\!/^\alpha_{\alpha^\circ} /\!/^\beta_{\beta^\circ}) \\ &\quad \in \mathscr{D}_Z(\mathcal{L}(T_oM, \mathbb{R}^e)) \end{aligned} \tag{2.245}$$

with $\boldsymbol{Z} := \overline{\bigcirc}(\boldsymbol{X})$ (independently of the chart used for the coordinates α, β, γ) and that

$$\int \boldsymbol{H} \mathrm{d}_\nabla \boldsymbol{X} = \int /\!/^* \boldsymbol{H} \mathrm{d}\overline{\bigcirc}(\boldsymbol{X}) \tag{2.246}$$

Note how, in particular, this is independent of the connection $\widetilde{\nabla}$ on τTM used to define $/\!/$ and $\overline{\bigcirc}$. Now, assume that M is Riemannian, ∇ is metric and Condition 2.79 holds. Then for $\boldsymbol{P} \in \mathscr{D}_X(\tau M^{\oplus e})$, and referring to Example 2.41 for the notation

$$
\begin{aligned}
\int g(\boldsymbol{P}, \mathrm{d}_\nabla \boldsymbol{X}) &= \int \boldsymbol{P}^\flat \mathrm{d}_\nabla \boldsymbol{X} \\
&= \int /\!/^* \boldsymbol{P}^\flat \mathrm{d}\overline{\bigcirc}(\boldsymbol{X}) \\
&= \int (\backslash\backslash \boldsymbol{P})^\flat \mathrm{d}\overline{\bigcirc}(\boldsymbol{X}) \\
&= \int \backslash\backslash \boldsymbol{P} \cdot \mathrm{d}\overline{\bigcirc}(\boldsymbol{X})
\end{aligned}
\tag{2.247}
$$

where the dot product denotes the metric at o and

$$
\backslash\backslash \boldsymbol{P} = \big((\backslash\backslash P)^{\gamma^\circ}, (\backslash\backslash P)'^{\beta^\circ}_{\alpha^\circ}\big) \coloneqq \big(\backslash\backslash^{\gamma^\circ}_{\gamma} P^\gamma, \textstyle\sum_{\beta,\eta}(\Gamma^\beta_{\xi\eta} \backslash\backslash^{\beta^\circ}_{\eta} /\!/^{\xi}_{\alpha^\circ} P^\beta + \backslash\backslash^{\beta^\circ}_{\beta} P^\beta_\alpha /\!/^{\alpha}_{\alpha^\circ})\big)
\tag{2.248}
$$

The converses of these statements, i.e.

$$
\begin{aligned}
&\int \boldsymbol{K} \mathrm{d}\boldsymbol{Z} = \int \backslash\backslash^* \boldsymbol{K} \mathrm{d}\underline{\bigcirc}(\boldsymbol{X}), \quad \int \boldsymbol{Q} \cdot \mathrm{d}\boldsymbol{Z} = \int g(/\!/\boldsymbol{Q}, \mathrm{d}\underline{\bigcirc}(\boldsymbol{Z})) \\
&\text{with } \backslash\backslash^* \boldsymbol{K} = ((\backslash\backslash^* K)_\gamma, (\backslash\backslash^* K)_{\alpha\beta}) \coloneqq (K_{\gamma^\circ} \backslash\backslash^{\gamma^\circ}_{\gamma}, K_{\alpha^\circ\beta^\circ} \backslash\backslash^{\alpha^\circ}_{\alpha} \backslash\backslash^{\beta^\circ}_{\beta} + K_{\beta^\circ} \textstyle\sum_\eta \Gamma^\beta_{\alpha\eta} /\!/^{\beta^\circ}_{\eta}) \\
&\text{and } /\!/\boldsymbol{Q} = \big((/\!/Q)^\gamma, (/\!/Q)^\beta_\alpha\big) \coloneqq (/\!/^{\gamma}_{\gamma^\circ} Q^{\gamma^\circ}, \Gamma^\beta_{\alpha\eta} /\!/^{\eta}_{\beta^\circ} Q^{\beta^\circ} + /\!/^{\beta}_{\beta^\circ} Q^{\beta^\circ}_{\alpha^\circ} \backslash\backslash^{\alpha^\circ}_{\alpha}) \\
&\text{for } \boldsymbol{K} \in \mathscr{D}_Z(\mathcal{L}(T_oM, \mathbb{R}^e)),\ \boldsymbol{Q} \in \mathscr{D}_Z(T_oM^{\oplus e})
\end{aligned}
\tag{2.249}
$$

are similarly shown to hold.

As an application of the latter, we show how the integration by parts formula [Dri04, Theorem 7.32] can be rewritten as an Itô integral on M. Let Z be a T_oM-valued Brownian motion, X its Stratonovich development, H a Cameron-Martin process above X, $h \coloneqq \backslash\backslash H$ with $h = \int u \mathrm{d}t$, $U \coloneqq /\!/u$ (for the precise terminology pertaining to curved Wiener space see the above reference). Then we may write the integration by parts formula, i.e. a formula for the adjoint of the gradient operator

$$
\begin{aligned}
\mathscr{D}^* H &= \int \Big(u^{\cdot} + \frac{1}{2} \backslash\backslash^{\cdot}_{\gamma} \mathscr{R}^{\gamma}{}_{\beta}(X) /\!/^{\beta}_{\beta^\circ} h^{\beta^\circ}\Big) \cdot \mathrm{d}Z \\
&= \int g\Big(U + \frac{1}{2} \mathscr{R}^{\cdot}{}_{\gamma}(X) H^\gamma, \mathrm{d}_\nabla X\Big)
\end{aligned}
\tag{2.250}
$$

as an Itô integral on M. Moreover, if u admits a Gubinelli derivative w.r.t. Z, so does U w.r.t. X, and Example 2.40, Example 2.41 may be combined to yield the expression of this as a Stratonovich integral on M, plus a correction term involving the covariant derivative of Ricci tensor.

Example 2.88 (Torsion)**.** In general, RDEs of the form Definition 2.42 are independent of the torsion of the connections on the source and target manifolds. For parallel transport, however, torsion of ∇ directly affects the field $F = \hbar$ that defines the RDE, and to that extent it influences the definition of $/\!\!/$ and therefore that of $\underline{\bigcirc}$ and $\overline{\bigcirc}$ (both for the trace and second order levels of the rough paths considered). The torsion of $\widetilde{\nabla}$, instead, plays no role. To exhibit the relevance of torsion for parallel transport and development we need only focus on smooth paths. Take $M = \mathbb{R}^3$ with its canonical coordinates, and ∇ with constant Christoffel symbols $\Gamma^1_{23} = 1 = -\Gamma^1_{32}$, $\Gamma^2_{31} = 1 = -\Gamma^2_{13}$, $\Gamma^3_{12} = 1 = -\Gamma^3_{21}$ and $\Gamma^k_{ij} = 0$ otherwise. This connection has the same geodesics as the Euclidean connection (straight lines), but, as described in [ua], parallel transport along geodesics looks like a spinning rugby ball, as illustrated in Figure 2.2 for an orthonormal frame. While the Euclidean connection and ∇ agree on geodesics, they define different notions of developments: identifying $T_0\mathbb{R}^3 = \mathbb{R}^3$ we have $\underline{\bigcirc} = \mathbb{1}$ according to the former, while this is not the case for the latter, as shown in Figure 2.3.

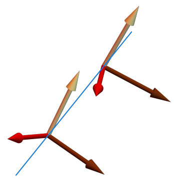


Figure 2.2

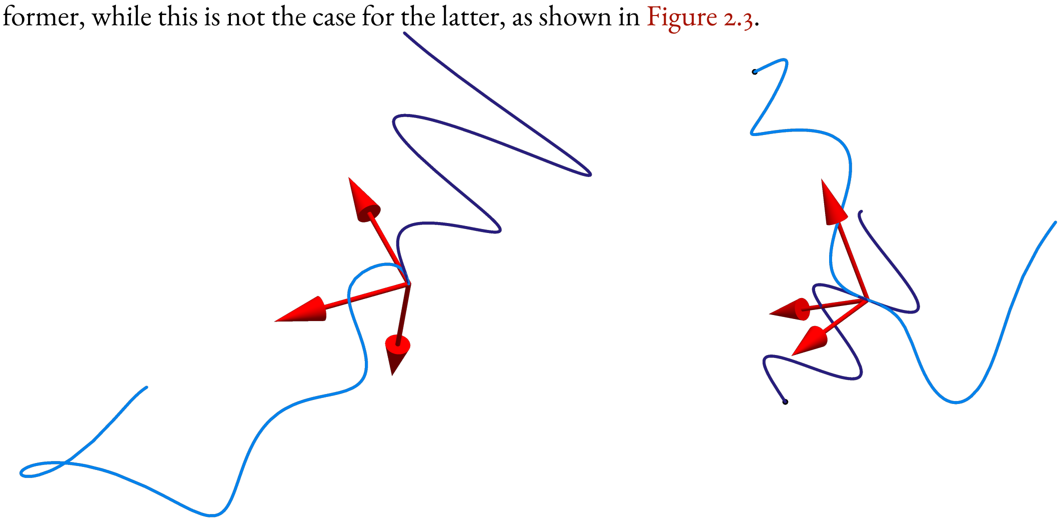


Figure 2.3: This figure relates to Example 2.88. The two plots are analogous to those of Figure 2.1, with the manifold in question being $\mathbb{R}^3$, endowed with the connection defined above, and the path being developed (in dark blue) is the parametrised smooth curve $X_t := (2t\cos(t), 10\sin(t), 3t)$. The two copies M and T_0M of $\mathbb{R}^3$ are superimposed, with coinciding axes in the first plot. We observe how the two curves are not identical, which would be the case if the connection on $\mathbb{R}^3$ were the Euclidean one. Also note how X and $\underline{\bigcirc}(X)$ are tangent curves at their point of contact.

Example 2.89. In this example we confine ourselves to geometric/Stratonovich development, and we consider the question of whether, given a sub-vector space $P \subseteq T_oM$, there exists a submanifold $N \subseteq M$ of the same dimension as P and tangent to it at o, at least defined in a neighbourhood of o, with the property that for all P-valued (rough) paths $\boldsymbol{Z}$ taking values in P and starting at 0_o, $\underline{\bigcirc}(\boldsymbol{Z})$ is valued in N. Since this must hold when Z is a straight line, and since straight lines develop to geodesics, if such N exists it must (at least locally) be given by $\exp(P)$. More generally, segments of affine line segments also develop to geodesic segments, and considering the case of Z a piecewise linear path leads to the conclusion that $\exp(P)$ must be a totally-geodesic submanifold of M, since it must contain every piecewise geodesic path started at o.

At the other extreme, we may be interested in showing that, when Z is a P-valued Brownian motion, $\underline{\bigcirc}(Z)$ admits a density w.r.t. to a (hence any) Lebesgue measure on M. If $P = T_oM$ we should expect this to hold in view of the fact that the vectors $T_y\phi M(\mathscr{H}_{\lambda^\circ}(y))$ (with ϕM the projection map of frame bundle) span T_xM for any $y \in F_xM$. Now let P be of dimension $k < m$. It is possible to show that the first two orders of the iterated Lie brackets of the fundamental horizontal vector fields, projected down onto TM, are given respectively by torsion and curvature:

$$\begin{aligned} T_y\phi M^\gamma[\mathscr{H}_{\mu^\circ}, \mathscr{H}_{\nu^\circ}] &= \mathscr{T}^\gamma_{\beta\alpha}(x) y^{(\alpha,\mu^\circ)} y^{(\beta,\nu^\circ)} \\ T_y\phi M^\gamma[\mathscr{H}_{\lambda^\circ}, [\mathscr{H}_{\mu^\circ}, \mathscr{H}_{\nu^\circ}]] &= \mathscr{R}_{\alpha\beta\delta}{}^{\gamma}(x) y^{(\alpha,\mu^\circ)} y^{(\beta,\nu^\circ)} y^{(\delta,\lambda^\circ)} \end{aligned} \tag{2.251}$$

What is needed, in concrete cases, to conclude that $\underline{\bigcirc}(Z)$ admits a smooth density is a version of the Hörmander condition which, stated in coordinates, only applies to the first e_1 components of an $(e_1 + e_2)$-dimensional SDE $\mathrm{d}Y = F_\gamma(Y)\mathrm{d}Z^\gamma$, and correspondingly only requires that $\mathrm{Lie}(F_\gamma \colon \gamma = 1, \ldots, d)(y)$ span $\mathbb{R}^{e_1}$. This is because we are only interested in the existence of the density of $\underline{\bigcirc}(Z)$, and not of the parallel frame above it, with which the development SDE is jointly written. The full solution is not going to admit a density in the frame bundle in general, regardless of the dimension of the Brownian motion: for example, if ∇ is Riemannian the parallel frame will be constrained to the orthonormal frame bundle. Even restricting to orthonormal frames, it is not clear to us that a density exists for arbitrary stochastic processes (a rather degenerate counterexample goes as follows: take $M = \mathbb{R}^2$ and take the stochastic process given by a random straight line out of the origin: the orthonormal basis will then be constant). See Figure 2.4 for an example of what this condition holding at different orders or not holding looks like. Note that there are possible intermediate answers to the questions above, i.e. there may be a submanifold N of dimension anywhere between that of P and that of M which contains all developments of P-valued rough paths. In view of [CF10] we can expect all these considerations, once made rigorous, to carry over to the case of Gaussian RDEs.

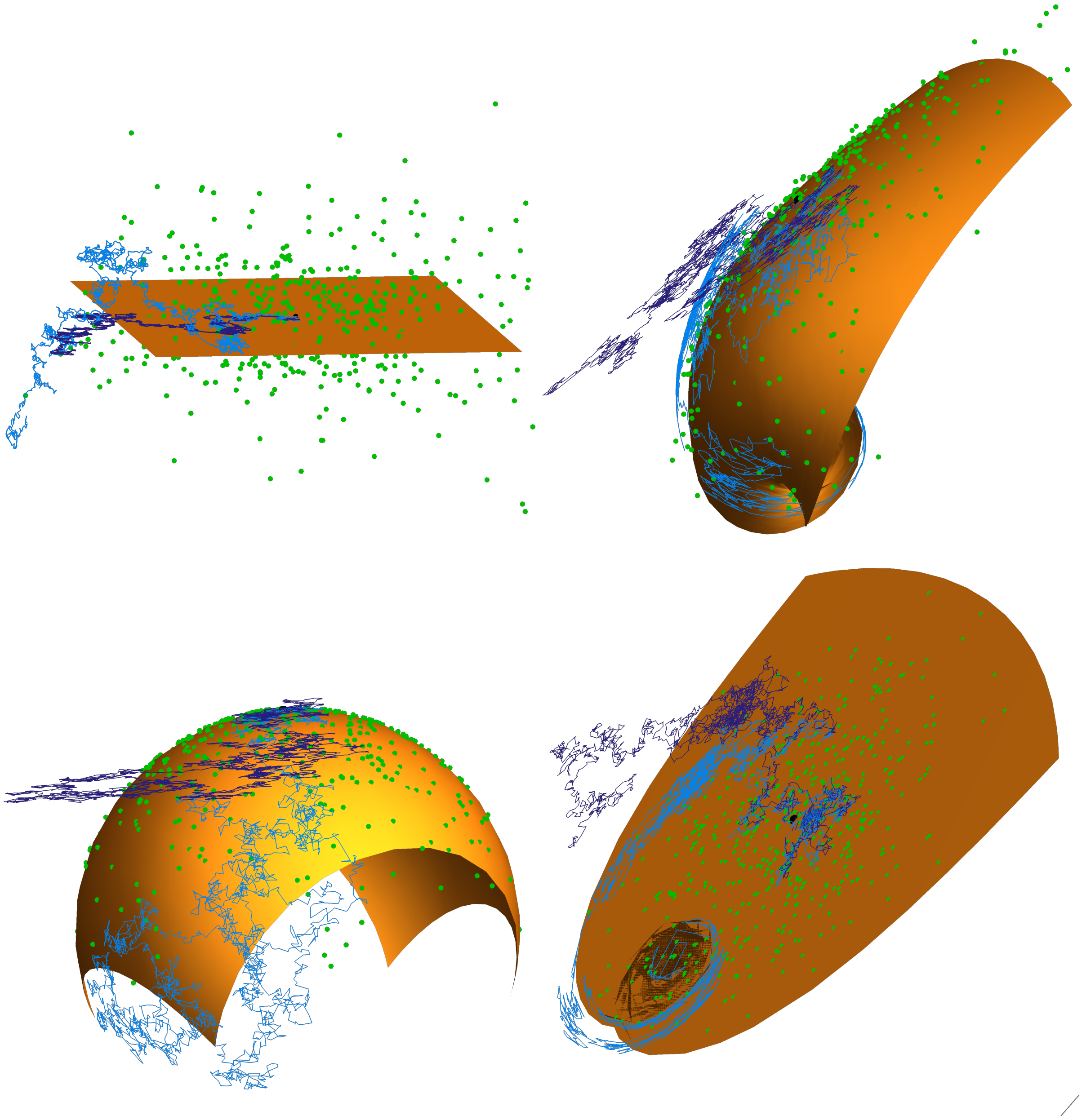


Figure 2.4: The purpose of these plots is the study of different behaviours of the development of a 2-dimensional Brownian motion within a 3-dimensional manifold M, specifically with regards to the question of whether it remains constrained to a 2 dimensional submanifold, or whether it admits a density. In the upper left we are considering $M = \mathbb{R}^3$ with the antisymmetric connection of Example 2.88, while in the other three cases $M = \mathbb{R}^3 \setminus \{0\}$ with the connection whose Christoffel symbols are given by taking those for the Euclidean metric, written in spherical coordinates, and setting the Γ^r's to zero (the geodesics in this connection are therefore circles centred at the origin and rays out of the origin, and M admits the foliation into the affine submanifolds given by concentric 2-spheres centred at the origin). In each case we have plotted one 2-dimensional Brownian path valued in some subplane P of T_oM (in dark blue), its development onto M (in light blue), a cloud of ~ 500 points consisting of the developments of the Brownian motion at a terminal time (in green; for improved visibility this time is less than the terminal time of the blue paths), and the locally nondegenerate surface parametrised by the exponential map applied to P. In the first case o is the origin and $P = \exp(P)$ is just the xy-plane, and we see how the point cloud (as well as the developed path) takes up three-dimensional space well: this is consistent with the Hörmander condition being satisfied at order 1, thanks to torsion (2.251). In the other three plots $o = (1, 0, 0)$ and P is a plane intersecting the xy-plane in the line $(0, t, 0)$ and with different inclinations w.r.t. the z-axis: $\pi/4$ in the upper right, 0 in the bottom left and $\pi/2$ in the bottom right (this means in latter two cases P is respectively tangent to the unit sphere, and coincides with the xy-plane). Note that we have rotated the plots for improved visibility of all the components. In the plot on the upper right we see how the point cloud and the developed path do not quite adhere to $\exp(P)$, consistent with the Hörmander condition being satisfied at order 2, thanks to curvature, but not at order 1, since the connection is torsion-free. In the other two cases $\exp(P)$ is an affine submanifold of M (in the first case a sphere with the Levi-Civita connection — a leaf in the aforementioned foliation — and in the second case the punctured xy-plane with a non-Euclidean connection), and therefore development remains constrained to the surface plotted in orange, and does not admit a density w.r.t. 3-dimensional Lebesgue measure.

Conclusions and further directions

In this chapter we have developed the basic theory of non-geometric rough paths on manifolds, both in the intrinsic and extrinsic frameworks, and shown how the classical notions of parallel transport and Cartan development carry over in a natural manner, but with non-trivial modification, to our setting.

We believe a couple of topics of the last section deserve closer attention. It would be interesting to explore additional examples of connections on τTM which may result in definitions of parallel transport different from the geometric/Stratonovich one: the Levi-Civita connection of the Cheeger-Gromoll metric [MT88] and the connections defined in [AT03, §5] could be worthwhile to test. Secondly, we would like to study the laws of developments of positive codimension Brownian motion, as discussed in Example 2.89, and to formulate the version of Hörmander's theorem necessary to show (in certain cases) the existence of the density.

3

A combinatorial approach to geometric rough paths and their controlled paths

Project status. This a chapter consists of a paper, written jointly with Thomas Cass, Christian Litterer and Bruce Driver, an almost identical version of which has very recently been accepted for publication in the *Journal of the London Mathematical Society*.

Introduction

In this chapter we use algebraic and combinatorial methods to explore the basic structure theory of weakly geometric rough paths of arbitrary roughness. Our approach allows us to work directly with the rough paths and leads to a clean separation of analysis and algebra, yielding explicit combinatorial descriptions of the resulting objects. The theory of geometric rough paths is well known and is usually deduced from the corresponding properties of the (lifts of) smooth paths, by taking closures in a suitable rough path metric. In finite dimensions, this means that identities for geometric rough paths readily extend to the weakly geometric setting. However, this extension is predicated on the close relation of weakly geometric and geometric rough paths established by Friz and Victoir [FV10b] for paths with values in finite-dimensional spaces. It is presently not clear if a similar relation holds in infinite dimensions. Similarly, the smooth approximation arguments are not available when studying more general branched rough paths.

Another goal of this chapter is to unify Lyons's original approach and Gubinelli's "linearised" version [Gub04] which deals with controlled rough paths, or *controlled paths* as we call them here, to avoid ambiguity. This is widely considered to be the most general and modern approach to rough path theory. Many fundamental results, such as the definition and convergence of controlled-rough integrals, are not present in the literature, stated in this setting. The fact that we work with controlled paths provides further motivation to avoid smooth approximation: controlled paths are only indirectly defined in terms of their reference rough path, and a smooth approximation of the rough path does not automatically yield one of the controlled path.

The structure theory for $[2,3) \ni p$-rough paths is well known [FH14] and has been developed with the manifold-valued theory in mind [CDL15]. These identities, however, do not extend in a straightforward manner to more irregular geometric rough paths. The most important example is the formula for the rough path lift of an controlled path. The idea of lifting controlled paths goes back to Gubinelli, who obtained an inductive formula valid for branched rough paths [Gub10, Remark 8.7]. The closed-form formula involving ordered shuffles first appeared in [LCL07], but is not stated in the controlled setting and is not maximally general, as it is only stated for integrals of symmetric $\mathrm{Lip}(\gamma - 1)$ functions. Our definition and proof works for an arbitrary controlled path, without assumptions on the symmetries of the derivatives, which are not present, for instance, in general RDE solutions. Many other identities involving rough and controlled paths follow from this, and lay the groundwork for the basic theory of arbitrarily irregular geometric rough paths on manifolds. Although the fact that we are dealing with geometric rough paths means that many of the identities involving integration will resemble their counterparts in ordinary calculus, the underlying rough path theory that makes it possible to prove them requires some combinatorics which is, at times, quite complex. For this reason, we have made it a priority to state all results in a clear, coordinate-free manner, without resorting to the coordinate notation that is used only in some of the proofs. This way, our formulae can more easily be referenced, and extended to new settings, such as infinite-dimensional vector spaces.

The chapter is structured as follows: in Section 3.1 we introduce algebraic preliminaries and notations, in particular the shuffle and ordered shuffle, and prove two technical lemmas that will be needed in the next section. Section 3.2 is the main section, in which we prove the fundamental structural identities of geometric rough path theory, such as associativity of rough integration. In Section 3.3 we give a very brief account of how the results of the previous can be used to transfer the theory of rough paths to manifolds, both in the intrinsic and extrinsic frameworks, and in Conclusions and further directions we mention a couple of extensions/applications that could follow from the work of this chapter.

3.1 Tensor bialgebras

We begin with a concise review of bialgebras defined on tensor algebras, for which we refer to [Man06, Chapter I] [Wei18, Chapter 2]. Given $n_1, \ldots, n_m \in \mathbb{N}$ (which may be 0) we define $\mathrm{Sh}(n_1, \ldots, n_m)$ to be the subset of the permutation group $\mathfrak{S}_{n_1+\ldots+n_m}$ of $(n_1, \ldots, n_m)$-*shuffles*, i.e. permutations σ with the property that

$$\sigma(n_1 + \ldots + n_{i-1} + 1) < \sigma(n_1 + \ldots + n_{i-1} + 2) < \ldots < \sigma(n_1 + \ldots + n_i) \tag{3.1}$$

for $i = 1, \ldots, m$ (with $n_0 := 0$). We will additionally call σ an $(n_1, \ldots, n_m)$-*ordered shuffle* if

$$\sigma(n_1) \le \sigma(n_1 + n_2) \le \ldots \le \sigma(n_1 + \ldots + n_m) \tag{3.2}$$

and we denote the set of these with $\overline{\mathrm{Sh}}(n_1, \ldots, n_m)$. If $n_i = 0$ for some i we have $\mathrm{Sh}(n_1, \ldots, n_m) = \mathrm{Sh}(n_1, \ldots, \widehat{n}_i, \ldots, n_m)$, $\overline{\mathrm{Sh}}(n_1, \ldots, n_m) = \overline{\mathrm{Sh}}(n_1, \ldots, \widehat{n}_i, \ldots, n_m)$ (with $\widehat{\ }$ denoting omission).

In this chapter the letters $U, V, W, \ldots$ will always be finite-dimensional $\mathbb{R}$-vector spaces. Given such a vector space V, a permutation $\sigma \in \mathfrak{S}_n$ induces a linear isomorphism

$$\sigma_*\colon V^{\otimes n} \to V^{\otimes n}, \quad v_1 \otimes \cdots \otimes v_n \mapsto v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(n)} \tag{3.3}$$

If $\{e_i\}_{i \in I}$ is a basis of V we may write $a \in V^{\otimes n}$ uniquely as $a^{(i_1,\dots,i_n)} e_{i_1} \otimes \cdots \otimes e_{i_n} =: a^{\boldsymbol{i}} e_{\boldsymbol{i}}$ where there is a sum on the ordered n-tuple of basis indices $\boldsymbol{i} \coloneqq (i_1, \dots, i_n)$ (following the Einstein convention), and

$$\sigma_*(a) = a^{\boldsymbol{i}} e_{\sigma_* \boldsymbol{i}} = a^{\sigma_*^{-1} \boldsymbol{j}} e_{\boldsymbol{j}}$$

where

$$\rho_* \boldsymbol{i} \coloneqq (i_{\rho(1)}, \dots, i_{\rho(n)}) \text{ for } \rho \in \mathfrak{S}_n \tag{3.4}$$

We let $I^\bullet \coloneqq \bigcup_{n \in \mathbb{N}} I^n$ be the set of I-valued tuples; this includes the empty tuple $()$, and we use $I^\bullet_*$ to denote the set of all such non-empty tuples. We will sometimes identify a tuple $(k_1, \dots, k_n)$ with the corresponding tensor $e_{k_1} \otimes \cdots \otimes e_{k_n}$ according to the chosen basis. Note that $\sigma_*^{-1} \boldsymbol{i}$ is the tuple obtained by "permuting $\boldsymbol{i}$ according to σ", e.g. if

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 5 & 2 & 4 \end{pmatrix} \in \mathrm{Sh}(3,2), \quad \sigma^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 2 & 5 & 3 \end{pmatrix} \in \mathfrak{S}_5$$

then

$$\sigma_*^{-1}(i_1, i_2, i_3, i_4, i_5) = (i_1, i_4, i_2, i_5, i_3)$$

Note that the composition rule (both for tensors and tuples) is

$$(\sigma \circ \rho)_* = \rho_* \sigma_* \tag{3.5}$$

Indeed, denoting $w_k \coloneqq v_{\sigma(k)}$ we have

$$\begin{aligned}
(\sigma \circ \rho)_*(v_1 \otimes \cdots \otimes v_n) &= v_{\sigma(\rho(1))} \otimes \cdots \otimes v_{\sigma(\rho(n))} \\
&= w_{\rho(1)} \otimes \cdots \otimes w_{\rho(n)} \\
&= \rho_*(w_1 \otimes \cdots \otimes w_n) \\
&= \rho_* \sigma_*(v_1 \otimes \cdots \otimes v_n)
\end{aligned}$$

For a tuple $\boldsymbol{i} = (i_1, \dots, i_n)$ we will denote $|\boldsymbol{i}| \coloneqq n$ its length, and given two tuples $\boldsymbol{i}, \boldsymbol{j}$ we write $\boldsymbol{ij}$ for their concatenation. We will denote $T(V) \coloneqq \bigoplus_{n=0}^{\infty} V^{\otimes n}$ and for $a \in T(V)$, with a^n its projection onto $V^{\otimes n}$; this has a distinct meaning to the notation $a^{\boldsymbol{i}}$ for tuples $\boldsymbol{i} \in I^\bullet$, explained above. When we are considering the tensor products of the dual V^* of a vector space V, or more generally the space of linear maps $\mathcal{L}(V, W)$ from V to another vector space W, we will replace superscripts with subscripts and vice-versa.

We denote by $(T(V), \otimes, \Delta_\sqcup)$ the tensor bialgebra of V, i.e. the product is given by the ordinary tensor product, which in coordinates reads

$$(a \otimes b)^{\boldsymbol{k}} = \sum_{\boldsymbol{ij} = \boldsymbol{k}} a^{\boldsymbol{i}} b^{\boldsymbol{j}} \tag{3.6}$$

and the *shuffle coproduct* is defined on elementary tensors (and extended linearly) as

$$\begin{aligned}\Delta_{ш}\colon T(V) &\to T(V)\boxtimes T(V)\\ v_1\otimes\cdots\otimes v_n &\mapsto \sum_{\substack{k=0,\ldots,n\\ \sigma\in\mathrm{Sh}(k,n-k)}} (v_{\sigma(1)}\otimes\cdots\otimes v_{\sigma(k)})\boxtimes(v_{\sigma(k+1)}\otimes\cdots\otimes v_{\sigma(n)})\end{aligned}\tag{3.7}$$

Here use the symbol $\boxtimes$ to denote external tensor product, reserving $\otimes$ for the algebra product. In coordinates this reads, for $a\in T(V)$

$$(\Delta_{ш}a)^{\boldsymbol{i},\boldsymbol{j}} = \sum_{\substack{\sigma\in\mathrm{Sh}(|\boldsymbol{i}|,|\boldsymbol{j}|)\\ \boldsymbol{k}=\sigma_*^{-1}(\boldsymbol{ij})}} a^{\boldsymbol{k}} =: \sum_{\boldsymbol{k}\in\mathrm{Sh}(\boldsymbol{i},\boldsymbol{j})} a^{\boldsymbol{k}} \tag{3.8}$$

where $\boldsymbol{i}$ denotes the index of the first $\boxtimes$-factor and $\boldsymbol{j}$ that of the second. In order to give a precise meaning to $\mathrm{Sh}(\boldsymbol{i},\boldsymbol{j})$ we must introduce multiset notation. Recall that a multiset is like a set (in that the order of its elements is not taken into account), but with the difference that the same element may appear more than once; we will denote multisets with double braces, e.g. $\{\{1,2,2,2,3,3\}\}$. If A and B are multisets we write $A\subseteq B$ if each element of A belongs to B counted with its multiplicity, e.g. $\{\{2,2,3,3\}\}\subseteq\{\{1,2,2,2,3,3\}\}$ but $\{\{1,1,2,2,3,3\}\}\not\subseteq\{\{1,2,2,2,3,3\}\}$, and $A=B$ is defined to mean $A\subseteq B$ and $B\subseteq A$. With this in mind, we are defining

$$\mathrm{Sh}(\boldsymbol{i},\boldsymbol{j}) \coloneqq \{\{\sigma_*^{-1}(\boldsymbol{ij}) \mid \sigma\in\mathrm{Sh}(|\boldsymbol{i}|,|\boldsymbol{j}|)\}\} \tag{3.9}$$

This means that the tuple $\boldsymbol{k}$ appears as many times as there are σ's with the property that $\boldsymbol{k}=\sigma_*^{-1}(\boldsymbol{ij})$. Similar multisets will be defined without explicit mention from now on.

The bialgebra that is graded dual to $(T(V),\otimes,\Delta_{ш})$ is given by $(T(V^*),ш,\Delta_\otimes)$. Note that we are using the notion of graded duality for bialgebras (see [Foi13, §1.5]), which is different to ordinary duality: in a nutshell, this just means that we are taking the dual of each (finite-dimensional) direct summand and that the product (coproduct) in one bialgebra is the dual to the product (coproduct) in the other; "dual" here makes sense because products and coproducts respect the grading. This allows us to avoid considering formal series of tensors, which are unnecessary when considering rough paths without their full signatures, and retain most of the usual properties of duality. $ш=\Delta_{ш}^*$ is the *shuffle product* given by

$$\begin{aligned}ш\colon T(V^*)\boxtimes T(V^*) &\to T(V^*),\\ (v^1\otimes\cdots\otimes v^n)\boxtimes(v^{n+1}\otimes\cdots\otimes v^{n+m}) &\mapsto \sum_{\sigma\in\mathrm{Sh}(n,m)} v^{\sigma^{-1}(1)}\otimes\cdots\otimes v^{\sigma^{-1}(n+m)}\end{aligned}\tag{3.10}$$

which in coordinates (using subscripts, since we are working in V^*) reads

$$(a\,ш\,b)_{\boldsymbol{k}} = \sum_{\substack{|\boldsymbol{i}|=0,\ldots,|\boldsymbol{k}|\\ \sigma\in\mathrm{Sh}(|\boldsymbol{i}|,|\boldsymbol{k}|-|\boldsymbol{i}|)\\ \boldsymbol{ij}=\sigma_*\boldsymbol{k}}} a_{\boldsymbol{i}}b_{\boldsymbol{j}} =: \sum_{(\boldsymbol{i},\boldsymbol{j})\in\mathrm{Sh}^{-1}(\boldsymbol{k})} a_{\boldsymbol{i}}b_{\boldsymbol{j}} \tag{3.11}$$

Note that with this notation $(\boldsymbol{i},\boldsymbol{j})\in\mathrm{Sh}^{-1}(\boldsymbol{k}) \Leftrightarrow \boldsymbol{k}\in\mathrm{Sh}(\boldsymbol{ij})$; in particular $\boldsymbol{i}$ or $\boldsymbol{j}$ may be empty. The coproduct $\Delta_\otimes = \otimes^*$ is the *deconcatenation coproduct* given by

$$\begin{aligned}\Delta_\otimes \colon T(V^*) &\to T(V^*) \boxtimes T(V^*) \\ v^1 \otimes \cdots \otimes v^n &\mapsto \sum_{k=0}^{n} (v^1 \otimes \cdots \otimes v^k) \boxtimes (v^{k+1} \otimes \cdots \otimes v^n)\end{aligned} \tag{3.12}$$

or in coordinates

$$(\Delta_\otimes a)_{\boldsymbol{i},\boldsymbol{j}} = a_{\boldsymbol{ij}} \tag{3.13}$$

Recall that in every bialgebra with coproduct Δ we may define its reduced coproduct $\widetilde{\Delta} a := \Delta a - a \boxtimes 1 - 1 \boxtimes a$, which is also coassociative. Also recall that in a coalgebra (C, Δ) (for us C will always be a tensor algebra) the (reduced) coproduct can be iterated, as coassociativity guarantees that $\Delta^m \colon C \to C^{\boxtimes m}$ has a unique meaning (note that under this convention $\Delta^2 = \Delta$, $\Delta^1 := \mathbb{1}_C$, $\Delta^0 := 1_{\mathbb{R}}$). Since the above bialgebras are graded and connected the reduced (iterated) coproduct factors as

$$\widetilde{\Delta}^m = \pi_{\geq 1}^{\boxtimes m} \circ \Delta^m \tag{3.14}$$

where $\pi_{\geq 1} \colon T(V) \twoheadrightarrow \bigoplus_{n \geq 1} V^{\otimes n}$ is the projection onto tensor products of positive order.

We may define the *ordered shuffle coproduct* $\Delta_{\overline{ш}}$ and the *ordered shuffle product* $\overline{ш}$ by requiring shuffles in (3.7) and (3.10) to be ordered. This does not, in fact, define a real (co)product, because $\overline{ш}$ fails to be associative: indeed, it satisfies the alternative relation

$$a \overline{ш} (b \overline{ш} c) = (a \overline{ш} b + b \overline{ш} a) \overline{ш} c$$

This property makes $(T(V), \overline{ш})$ a Zinbiel algebra [EFP15]. Whenever we iterate $\overline{ш}$ or $\Delta_{\overline{ш}}$ we will be carrying out composition from left to right, i.e. inductively

$$\begin{aligned} a_1 \overline{ш} \cdots \overline{ш} a_n &:= (a_1 \overline{ш} a_2) \overline{ш} a_3 \overline{ш} \cdots \overline{ш} a_n \\ \Delta_{\overline{ш}}^m a &:= \sum_{(a)_{\overline{ш}}^m} a_{(1)} \boxtimes \cdots \boxtimes a_{(m)} \\ &:= \sum_{(a)_{\overline{ш}}^{m-1}} \Big(\Delta_{\overline{ш}} a_{(1)} \Big) \boxtimes a_{(2)} \boxtimes \cdots \boxtimes a_{(m-1)} \end{aligned} \tag{3.15}$$

This guarantees that the coordinate expression provided for the unordered shuffle carries over to the ordered case, with $\overline{\mathrm{Sh}}$ instead of Sh, e.g.

$$\begin{aligned} (\widetilde{\Delta}_{\overline{ш}}^m a)^{\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m} &= [|\boldsymbol{k}^1|, \ldots, |\boldsymbol{k}^m| \geq 1] \sum_{\substack{\sigma \in \overline{\mathrm{Sh}}(|\boldsymbol{k}^1|, \ldots, |\boldsymbol{k}^m|) \\ \boldsymbol{k} = \sigma_*^{-1}(\boldsymbol{k}^1 \ldots \boldsymbol{k}^m)}} a^{\boldsymbol{k}} \\ &=: [|\boldsymbol{k}^1|, \ldots, |\boldsymbol{k}^m| \geq 1] \sum_{\boldsymbol{k} \in \overline{\mathrm{Sh}}(\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m)} a^{\boldsymbol{k}} \end{aligned} \tag{3.16}$$

Here the square bracket has binary value depending on the truth value of the proposition it contains, and is present because the coproduct is reduced; the set of permutations over which the sum is taken is given by ordered shuffles, reflecting the fact that we are dealing with the ordered shuffle coproduct.

Before proceeding, we take a moment to motivate our interest in shuffles and ordered shuffles, although this will become much clearer in Section 3.2. While it is well known that the former are used to express products of iterated integrals of a (smooth) path X

$$\begin{aligned}&\Big(\int_{s<u_1<\ldots<u_n<t} \mathrm{d}X^{i_1}_{u_1}\cdots \mathrm{d}X^{i_m}_{u_m}\Big)\Big(\int_{s<v_1<\ldots<v_n<t} \mathrm{d}X^{j_1}_{v_1}\cdots \mathrm{d}X^{j_n}_{v_n}\Big)\\ &= \sum_{\boldsymbol{k}\in \mathrm{Sh}(\boldsymbol{i},\boldsymbol{j})} \int_{s<r_1<\ldots<r_{n+m}<t} \mathrm{d}X^{k_1}_{r_1}\cdots \mathrm{d}X^{k_{m+n}}_{r_{m+n}}\end{aligned}$$

the role of ordered shuffles in the study of iterated integrals of paths is less appreciated. One way to motivate their significance is as follows: let Y be the solution to the ODE

$$\mathrm{d}Y = V(Y)\mathrm{d}X = V(Y)\dot{X}\mathrm{d}t$$

with X V-valued and Y W-valued. We fix bases on both vector spaces and use Greek indices for V and Latin ones for W — this will be the convention later on as well. Substituting formal Euler expansions, and defining $V^k_{\boldsymbol{\gamma}}(y) := V_{\gamma_1}\cdots V^k_{\gamma_n}(y)$ for a tuple $\boldsymbol{\gamma} = (\gamma_1,\cdots,\gamma_n)$, with the product denoting iterated composition of vector fields (i.e. $V_\gamma f(y) := \partial_k f(y) V^k_\gamma(y)$ for a function f of y)

$$\begin{aligned}&\int_{s<u_1<\ldots<u_m<t} \mathrm{d}Y^{k_1}_{u_1}\cdots \mathrm{d}Y^{k_m}_{u_m}\\ &= \int_{s<u_1<\ldots<u_m<t} \mathrm{d}\Big(V^{k_1}_{\boldsymbol{\gamma}^1}(Y_s)\boldsymbol{X}^{\boldsymbol{\gamma}^1}_{su_1}\Big)\cdots \mathrm{d}\Big(V^{k_m}_{\boldsymbol{\gamma}^m}(Y_s)\boldsymbol{X}^{\boldsymbol{\gamma}^m}_{su_m}\Big)\\ &= V^{k_1}_{\boldsymbol{\gamma}^1}(Y_s)\cdots V^{k_m}_{\boldsymbol{\gamma}^m}(Y_s)\int_{s<u_1<\ldots<u_m<t} \mathrm{d}\big(\textstyle\int_{s<r^1_1<\ldots<r^1_{n_1}<u_n} \mathrm{d}X^{\gamma^1_1}_{r^1_1}\cdots \mathrm{d}X^{\gamma^1_{n_1}}_{r^1_{n_1}}\big)\cdots\\ &\quad\cdots \mathrm{d}\big(\textstyle\int_{s<r^m_1<\ldots<r^m_{n_m}<u_m} \mathrm{d}X^{\gamma^m_1}_{r^m_1}\cdots \mathrm{d}X^{\gamma^m_{n_m}}_{r^m_{n_m}}\big)\\ &= V^{k_1}_{\boldsymbol{\gamma}^1}(Y_s)\cdots V^{k_m}_{\boldsymbol{\gamma}^m}(Y_s)\\ &\quad\cdot\int_{\substack{s<r^1_1<\ldots<r^1_{n_1-1}<r^1_{n_1}\\ \vdots\\ s<r^m_1<\ldots<r^m_{n_m-1}<r^m_{n_m}\\ s<r^1_{n_1}<\ldots<r^m_{n_m}<t}} \mathrm{d}X^{\gamma^1_1}_{r^1_1}\cdots \mathrm{d}X^{\gamma^1_{n_1}}_{r^1_{n_1}}\cdots\cdots \mathrm{d}X^{\gamma^m_{n_1}}_{r^m_1}\cdots \mathrm{d}X^{\gamma^m_{n_m}}_{r^m_{n_m}}\\ &= \sum_{\boldsymbol{\gamma}^1,\ldots,\boldsymbol{\gamma}^m} V^{k}_{\boldsymbol{\gamma}^1}(Y_s)\cdots V^{k}_{\boldsymbol{\gamma}^m}(Y_s)\sum_{\boldsymbol{\gamma}\in\overline{\mathrm{Sh}}(\boldsymbol{\gamma}^1,\ldots,\boldsymbol{\gamma}^m)}\int_{s<v_1<\ldots<v_{n:=\sum_i n_i}} \mathrm{d}X^{\gamma_1}_{v_1}\cdots \mathrm{d}X^{\gamma_n}_{v_n}\end{aligned} \tag{3.17}$$

In other words, ordered shuffles index the sum involved in the expression of the iterated integrals of Y in terms of those of X. The relevance of ordered shuffles in the similar cases of linear RDEs and $\mathrm{Lip}(\gamma-1)$ functions was observed in [LCL07, p.72-75].

We will now prove a couple of technical results that will be used in Section 3.2; these are of our own formulation and cannot be found in the literature referenced at the beginning. They will be stated in terms of tuples (although they are essentially statements about shuffles), so they can be readily deployed when dealing with rough paths. Unshuffling and concatenating satisfy the following commutativity relation:

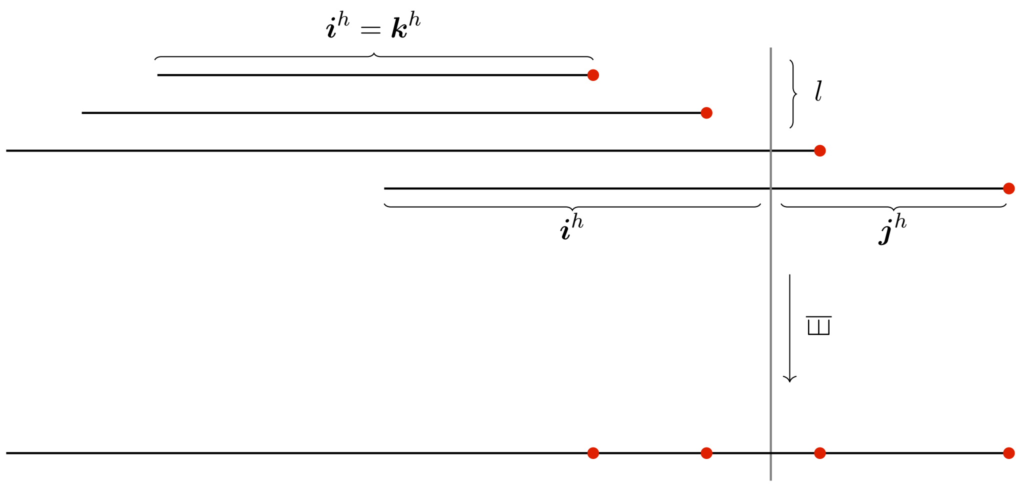
$$\begin{array}{ccc} T(V) & \xrightarrow{\Delta_{\sqcup\!\sqcup}^m} & T(V)^{\boxtimes m} \\ \uparrow{\scriptstyle\otimes} & & \uparrow{\scriptstyle\otimes_{1,m+1}\boxtimes\cdots\boxtimes\otimes_{m,2m}} \\ T(V)^{\boxtimes 2} & \xrightarrow{\Delta_{\sqcup\!\sqcup}^m\boxtimes\Delta_{\sqcup\!\sqcup}^m} & T(V)^{\boxtimes 2m} \end{array}$$

The following lemma can be viewed as the counterpart to this statement in the context of ordered shuffles. Since it is essentially combinatorial in nature, we will state it in terms of tuples. Recall that $I_*^\bullet$ denotes the set of all I-valued tuples of any positive order.

Lemma 3.1. *Let $\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m \in I_*^\bullet$. The following identity of multisets holds:*

$$\begin{aligned} &\{\{(\boldsymbol{i},\boldsymbol{j}) \in I^\bullet \times I^\bullet \mid \boldsymbol{ij} \in \overline{\mathrm{Sh}}(\boldsymbol{k}^1,\ldots,\boldsymbol{k}^m)\}\} \\ = &\{\{(\boldsymbol{i},\boldsymbol{j}) \in I^\bullet \times I^\bullet \mid \boldsymbol{j} \in \overline{\mathrm{Sh}}(\boldsymbol{j}^{l+1},\ldots,\boldsymbol{j}^m); \\ &\boldsymbol{i} \in \mathrm{Sh}(\overline{\mathrm{Sh}}(\boldsymbol{i}^1,\ldots,\boldsymbol{i}^l), \mathrm{Sh}(\boldsymbol{i}^{l+1},\ldots,\boldsymbol{i}^m)); \\ &\text{where } \boldsymbol{i}^h = \boldsymbol{k}^h \text{ for } h \le l; \\ &\boldsymbol{i}^h\boldsymbol{j}^h = \boldsymbol{k}^h \text{ and } |\boldsymbol{j}^h| \ge 1, \text{ for } h \ge l+1; \\ &\text{with } l = 0,\ldots,m\}\} \end{aligned} \tag{3.18}$$

The following picture is meant to illustrate the idea of the statement: the horizontal lines represent tuples, and the red bullet points represent their terminal elements.



Note that we are taking into account multiplicities in the multiset $\boldsymbol{k} \in \overline{\mathrm{Sh}}(\boldsymbol{k}^1,\ldots,\boldsymbol{k}^m)$, i.e. if the same tuple $\boldsymbol{k}$ belongs twice to $\overline{\mathrm{Sh}}(\boldsymbol{k}^1,\ldots,\boldsymbol{k}^m)$, then any pair $(\boldsymbol{i},\boldsymbol{j})$ such that $\boldsymbol{ij} = \boldsymbol{k}$ appears twice in the multiset; an analogous remark holds for the RHS and for all similarly defined multisets. $\mathrm{Sh}(\overline{\mathrm{Sh}}(\boldsymbol{i}^1,\ldots,\boldsymbol{i}^l), \mathrm{Sh}(\boldsymbol{i}^{l+1},\ldots,\boldsymbol{i}^m))$ here stands for

$$\{\{\boldsymbol{i} \in \mathrm{Sh}(\boldsymbol{a},\boldsymbol{b}) \mid \boldsymbol{a} \in \overline{\mathrm{Sh}}(\boldsymbol{i}^1,\ldots,\boldsymbol{i}^l),\ \boldsymbol{b} \in \mathrm{Sh}(\boldsymbol{i}^{l+1},\ldots,\boldsymbol{i}^m)\}\}$$

the multiset of tuples obtained by shuffling $\boldsymbol{i}^1,\ldots,\boldsymbol{i}^m$, the first l with order. When $l = 0$ or m this reduces respectively to $\mathrm{Sh}(\mathrm{Sh}(\boldsymbol{i}^1,\ldots,\boldsymbol{i}^m)) = \mathrm{Sh}(\boldsymbol{i}^1,\ldots,\boldsymbol{i}^m)$, $\mathrm{Sh}(\overline{\mathrm{Sh}}(\boldsymbol{i}^1,\ldots,\boldsymbol{i}^m)) = \overline{\mathrm{Sh}}(\boldsymbol{i}^1,\ldots,\boldsymbol{i}^m)$, since the only possible way of shuffling a single is to leave it unchanged. The proof of this lemma is most easily understood when going through its steps with reference to the example that immediately follows it.

Proof of Lemma 3.1. Let $A(\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m)$ denote the first multiset defined above and $B(\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m)$ the second. For tuples $\boldsymbol{\ell}^1, \ldots, \boldsymbol{\ell}^n$ and $1 \leq a \leq b \leq n$ define $\boldsymbol{\ell}^{a:b} := \boldsymbol{\ell}^a \ldots \boldsymbol{\ell}^b \in I^\bullet$ (juxtaposition) and for a tuple $\boldsymbol{\ell}$ and $1 < c \leq d \leq |\boldsymbol{\ell}|$ $\boldsymbol{\ell}_{c:d} := (\ell_c, \ldots, \ell_d)$, and let $(\boldsymbol{i}, \boldsymbol{j}) \in A(\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m)$. This means there exists $\sigma \in \overline{\mathrm{Sh}}(|\boldsymbol{k}^1|, \ldots, |\boldsymbol{k}^m|)$ with $\boldsymbol{ij} = \sigma_*^{-1} \boldsymbol{k}^{1:m}$. Let

$$l := \begin{cases} 0 & \text{if } |\boldsymbol{i}| < \sigma(|\boldsymbol{k}^1|) \\ m & \text{if } \boldsymbol{j} = () \\ \text{s.t. } \sigma(|\boldsymbol{k}^{1:l}|) \leq |\boldsymbol{i}| < \sigma(|\boldsymbol{k}^{1:l+1}|) & \text{otherwise} \end{cases}$$

which is exists and is unique since σ is an ordered shuffle. We then let $\boldsymbol{i}^h := \boldsymbol{k}^h$ for $h \leq l$ and for $h \geq l+1$

$$\boldsymbol{i}^h \boldsymbol{j}^h := \boldsymbol{k}^h, \quad \sigma(|\boldsymbol{k}^{1:h-1}| + |\boldsymbol{i}^h|) \leq |\boldsymbol{i}| < \sigma(|\boldsymbol{k}^{1:h-1}| + |\boldsymbol{i}^h| + 1) \tag{3.19}$$

where $|\boldsymbol{i}^h|$ (and hence $\boldsymbol{i}^h$) is unique since σ is a shuffle. Now, it cannot be the case that for $h \geq l+1$ we have $|\boldsymbol{i}^h| = |\boldsymbol{k}^h|$, for this would violate the definition of l: this implies $|\boldsymbol{j}^h| \geq 1$. Moreover, we have $\boldsymbol{i} = \sigma_*^{-1} \boldsymbol{i}^{1:m}$ and $\boldsymbol{j} = \sigma_*^{-1} \boldsymbol{j}^{l+1:m}$, where we are defining the right hand sides using the same expression as before (3.4), but by considering the numberings on $\boldsymbol{i}^{1:m}$, $\boldsymbol{j}^{l+1:m}$ to be those inherited as subtuples of $\boldsymbol{k}^{1:m}$: this is because $\boldsymbol{i}^h$ occupies the segment of $\boldsymbol{k}^{1:m}$ numbered with $[|\boldsymbol{k}^{1:h-1}| + 1, |\boldsymbol{k}^{1:h-1}| + |\boldsymbol{i}^h|]$, all of which σ maps into $[1, |\boldsymbol{i}|]$, by (3.19) and again by the shuffle property of σ; similarly, $\boldsymbol{j}^h$ occupies the segment numbered $[|\boldsymbol{k}^{1:h-1}| + |\boldsymbol{i}^h| + 1, |\boldsymbol{k}^{1:h}|]$ which gets mapped above $|\boldsymbol{i}|$. By construction σ (once domains are renumbered) shuffles $\boldsymbol{i}^{1:l}$ and $\boldsymbol{j}^{l+1:m}$ with order, since these are the tuples that contain the $k^h_{|\boldsymbol{k}^h|}$'s, and $\boldsymbol{i}^{l+1,m}$ without order. If $\rho \in \mathrm{Sh}(n_1, \ldots, n_m)$ and $S \subseteq \{1, \ldots, n_1 + \ldots + n_m\}$, $\rho|_S$ is still a shuffle, with the additional order constraints on those $(n_1 + \ldots + n_q)$'s that belong to S: therefore, we have that $\boldsymbol{j} \in \overline{\mathrm{Sh}}(\boldsymbol{j}^{l+1}, \ldots, \boldsymbol{j}^m)$ and $\boldsymbol{i} \in \mathrm{Sh}(\overline{\mathrm{Sh}}(\boldsymbol{i}^1, \ldots, \boldsymbol{i}^l), \mathrm{Sh}(\boldsymbol{i}^{l+1}, \ldots, \boldsymbol{i}^m))$. This shows $A(\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m) \subseteq B(\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m)$.

Conversely, let $(\boldsymbol{i}, \boldsymbol{j}) \in B(\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m)$, with $l, \boldsymbol{i}^h, \boldsymbol{j}^h$ as in (3.18). $\boldsymbol{ij}$ is obtained by an ordered shuffle of $\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m$: that the order of each $\boldsymbol{k}^h$, $h \leq l$ is preserved is immediate since $\boldsymbol{i}^h = \boldsymbol{k}^h$; that the order of each $\boldsymbol{k}^h$, $h > l$ is preserved is a consequence of the fact that the order of $\boldsymbol{i}^h$ is preserved, that the order of $\boldsymbol{j}^h$ is preserved, and that $\boldsymbol{i}$ comes before $\boldsymbol{j}$ in the juxtaposition $\boldsymbol{ij}$; that the shuffle of the $\boldsymbol{k}^h$'s is ordered is a consequence of the fact that the shuffles of $\boldsymbol{i}^1, \ldots, \boldsymbol{i}^l$ and $\boldsymbol{j}^{l+1}, \ldots, \boldsymbol{j}^{l+1}$ are ordered. This shows $B(\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m) \subseteq A(\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m)$; also note that in both inclusions multiplicities are indeed counted, since the correspondence between the underlying permutation in $A(\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m)$ and the pair of underlying permutations in $B(\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m)$ is bijective. ■

Example 3.2. We illustrate the idea behind this lemma with an example. Let

$$m = 4; \quad |\boldsymbol{k}^1| = 2,\ |\boldsymbol{k}^2| = 3,\ |\boldsymbol{k}^3| = 4,\ |\boldsymbol{k}^4| = 4$$

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 \\ 4 & 6 & 3 & 5 & 8 & 7 & 9 & 11 & 12 & 1 & 2 & 10 & 13 \end{pmatrix} \in \overline{\mathrm{Sh}}(2,3,4,4)$$

$$\sigma_*^{-1}(\boldsymbol{k}^1\boldsymbol{k}^2\boldsymbol{k}^3\boldsymbol{k}^4) = (\underbrace{k^4_1, k^4_2, k^2_1, k^1_1, k^2_2, {\color{red}k^1_2}, k^3_1, {\color{red}k^2_3}}_{=:\boldsymbol{i}}, \underbrace{k^3_2, k^4_3, k^3_3, {\color{red}k^3_4}, {\color{red}k^4_4}}_{=:\boldsymbol{j}})$$

$$(\boldsymbol{i},\boldsymbol{j}) \in A(\boldsymbol{k}^1,\boldsymbol{k}^2,\boldsymbol{k}^3,\boldsymbol{k}^4)$$

where we have coloured in red the terminal elements of the $\boldsymbol{k}^h$'s (and recall that it is necessary to renumber $\boldsymbol{k}^1\boldsymbol{k}^2\boldsymbol{k}^3\boldsymbol{k}^4$ from 1 to 13 before applying (3.4), and then change the numbering back) and we have

$$\begin{aligned}
&(\boldsymbol{i},\boldsymbol{j}) \in B(\boldsymbol{k}^1,\boldsymbol{k}^2,\boldsymbol{k}^3,\boldsymbol{k}^4) \quad \text{with } l=2;\\
&\boldsymbol{i}^1 = (k^1_1,k^1_2),\ \boldsymbol{i}^2 = (k^2_1,k^2_2,k^2_3),\ \boldsymbol{i}^3 = (k^3_1),\ \boldsymbol{i}^4 = (k^4_1,k^4_2);\\
&\boldsymbol{j}^1 = (k^3_2,k^3_3,k^3_4),\ \boldsymbol{j}^2 = (k^4_3,k^4_4)
\end{aligned}$$

since $\boldsymbol{k}^h = \boldsymbol{i}^h$ for $h=1,2$, $\boldsymbol{k}^h = \boldsymbol{i}^h\boldsymbol{j}^h$ for $h=3,4$, and

$$\begin{aligned}
&\boldsymbol{j} \in \overline{\mathrm{Sh}}(\boldsymbol{j}^1,\boldsymbol{j}^2); \quad \boldsymbol{i} \in \mathrm{Sh}(\boldsymbol{a},\boldsymbol{b})\\
&\text{with } \boldsymbol{a} := (k^2_1,k^1_1,k^2_2,k^1_2,k^2_3) \in \overline{\mathrm{Sh}}(\boldsymbol{i}^1,\boldsymbol{i}^2)\\
&\qquad \boldsymbol{b} := (k^4_1,k^4_2,k^3_1) \in \mathrm{Sh}(\boldsymbol{i}^3,\boldsymbol{i}^4)
\end{aligned}$$

Note that neither of the two Sh's above can be replaced with $\overline{\mathrm{Sh}}$.

The reduced version of this Lemma 3.1 would involve restricting it to non-empty tuples $\boldsymbol{k}^1,\ldots,\boldsymbol{k}^m$; we will need the dual of this statement. We will use the notation $((\boldsymbol{i}^1,\ldots,\boldsymbol{i}^l),(\boldsymbol{i}^{l+1},\ldots,\boldsymbol{i}^m)) \in (\overline{\mathrm{Sh}}^{-1},\mathrm{Sh}^{-1})(\mathrm{Sh}^{-1}(\boldsymbol{i}))$ as a shorthand for $(\boldsymbol{i}^1,\ldots,\boldsymbol{i}^l) \in \overline{\mathrm{Sh}}^{-1}(\boldsymbol{a})$, $(\boldsymbol{i}^{l+1},\ldots,\boldsymbol{i}^m) \in \mathrm{Sh}^{-1}(\boldsymbol{b}) : (\boldsymbol{a},\boldsymbol{b}) \in \mathrm{Sh}^{-1}(\boldsymbol{i})$, and m is fixed.

Corollary 3.3. *For I-valued tuples $\boldsymbol{i},\boldsymbol{j}$ the following identity of multisets holds:*

$$\begin{aligned}
&\{\{(\boldsymbol{k}^1,\ldots,\boldsymbol{k}^m) \in \overline{\mathrm{Sh}}^{-1}(\boldsymbol{i}\boldsymbol{j}) \mid |\boldsymbol{k}^1|,\ldots,|\boldsymbol{k}^m| \geq 1\}\}\\
=&\{\{(\boldsymbol{k}^1,\ldots,\boldsymbol{k}^m) \in (I^\bullet_*)^m \mid (\boldsymbol{j}^{l+1},\ldots,\boldsymbol{j}^m) \in \overline{\mathrm{Sh}}^{-1}(\boldsymbol{j});\\
&((\boldsymbol{i}^1,\ldots,\boldsymbol{i}^l),(\boldsymbol{i}^{l+1},\ldots,\boldsymbol{i}^m)) \in (\overline{\mathrm{Sh}}^{-1},\mathrm{Sh}^{-1})(\mathrm{Sh}^{-1}(\boldsymbol{i}));\\
&\boldsymbol{k}^h = \boldsymbol{i}^h,\ h \leq l;\ \boldsymbol{k}^h = \boldsymbol{i}^h\boldsymbol{j}^h,\ |\boldsymbol{j}^h| \geq 1,\ h \geq l+1;\\
&l = 0,\ldots,m\}\}
\end{aligned} \tag{3.20}$$

Proof of Corollary 3.3. Let $C(\boldsymbol{i},\boldsymbol{j})$ denote the first multiset above and $D(\boldsymbol{i},\boldsymbol{j})$ the second, and recall the names $A(\boldsymbol{k}^1,\ldots,\boldsymbol{k}^m)$, $B(\boldsymbol{k}^1,\ldots,\boldsymbol{k}^m)$ for the sets of Lemma 3.1. We then have (taking into account multiplicities)

$$\begin{aligned}
(\boldsymbol{k}^1,\ldots,\boldsymbol{k}^m) \in C(\boldsymbol{i},\boldsymbol{j}) &\iff (\boldsymbol{i},\boldsymbol{j}) \in A(\boldsymbol{k}^1,\ldots,\boldsymbol{k}^m)\\
&\iff (\boldsymbol{i},\boldsymbol{j}) \in B(\boldsymbol{k}^1,\ldots,\boldsymbol{k}^m)\\
&\iff (\boldsymbol{k}^1,\ldots,\boldsymbol{k}^m) \in D(\boldsymbol{i},\boldsymbol{j})
\end{aligned}$$

thus concluding the proof. ■

Next we discuss another combinatorial relation involving ordered and unordered shuffles; similarly to the earlier lemma, an example is provided after the proof that helps to .

Lemma 3.4. *Let* $n := n_1 + \ldots + n_m$, $n^l := n_1 + \ldots + n_l$ *for* $l = 1, \ldots, m$, *and* $\boldsymbol{k}^1, \ldots, \boldsymbol{k}^n \in I_*^\bullet$. *We have*

$$\begin{aligned} &\mathrm{Sh}(\overline{\mathrm{Sh}}(\boldsymbol{k}^1, \ldots, \boldsymbol{k}^{n_1}), \ldots, \overline{\mathrm{Sh}}(\boldsymbol{k}^{n^{m-1}+1}, \ldots, \boldsymbol{k}^n)) \\ =& \bigsqcup_{\pi \in \mathrm{Sh}(n_1,\ldots,n_m)} \overline{\mathrm{Sh}}(\boldsymbol{k}^{\pi^{-1}(1)}, \ldots, \boldsymbol{k}^{\pi^{-1}(n)}) \end{aligned} \tag{3.21}$$

Here $\bigsqcup$ denotes disjoint union of multisets, e.g. if the same tuple appears in sets corresponding to two different σ's it should be counted twice.

Proof of Lemma 3.4. Let $N_l := |\boldsymbol{k}^{n^{l-1}+1}| + \ldots + |\boldsymbol{k}^{n^l}|$ for $l = 1, \ldots, m$, and $N^l := N_1 + \ldots + N_l$ for $l = 1, \ldots, m$, and $N := N^m$. We have

$$\begin{aligned} &\mathrm{Sh}(\overline{\mathrm{Sh}}(\boldsymbol{k}^1, \ldots, \boldsymbol{k}^{n_1}), \ldots, \overline{\mathrm{Sh}}(\boldsymbol{k}^{n^{m-1}+1}, \ldots, \boldsymbol{k}^n)) \\ =&\{\{\sigma_*^{-1}(\boldsymbol{h}^1 \ldots \boldsymbol{h}^m) \mid \sigma \in \mathrm{Sh}(N_1, \ldots, N_m); \\ &\boldsymbol{h}^l \in \overline{\mathrm{Sh}}(\boldsymbol{k}^{n^{l-1}+1}, \ldots, \boldsymbol{k}^{n^l}),\ l = 1, \ldots, m\}\} \\ =&\{\{\sigma_*^{-1}(\boldsymbol{h}^1 \ldots \boldsymbol{h}^m) \mid \sigma \in \mathrm{Sh}(N_1, \ldots, N_m); \\ &\boldsymbol{h}^l = \rho_{l*}^{-1}(\boldsymbol{k}^{n^{l-1}+1} \ldots \boldsymbol{k}^{n^l}); \\ &\rho_l \in \overline{\mathrm{Sh}}(|\boldsymbol{k}^{n^{l-1}+1}|, \ldots, |\boldsymbol{k}^{n^l}|),\ l = 1, \ldots, m\}\} \end{aligned}$$

Now, denoting $(\rho_1, \ldots, \rho_m)$ the element of $\mathfrak{S}_N$ which acts on $\{N^{l-1} + 1, \ldots, N^l\}$ with ρ_l, we continue the chain of identities

$$\begin{aligned} =&\{\{\sigma_*^{-1}(\rho_1, \ldots, \rho_m)_*^{-1}(\boldsymbol{k}^1 \ldots \boldsymbol{k}^{n_m}) \mid \sigma \in \mathrm{Sh}(N_1, \ldots, N_m); \\ &\rho_l \in \overline{\mathrm{Sh}}(|\boldsymbol{k}^{n^{l-1}+1}|, \ldots, |\boldsymbol{k}^{n^l}|),\ l = 1, \ldots, m\}\} \\ =&\{\{(\sigma \circ (\rho_1, \ldots, \rho_m))_*^{-1}(\boldsymbol{k}^1 \ldots \boldsymbol{k}^{n_m}) \mid \sigma \in \mathrm{Sh}(N_1, \ldots, N_m); \\ &\rho_l \in \overline{\mathrm{Sh}}(|\boldsymbol{k}^{n^{l-1}+1}|, \ldots, |\boldsymbol{k}^{n^l}|),\ l = 1, \ldots, m\}\} \end{aligned}$$

since $(\rho_1^{-1}, \ldots, \rho_m^{-1}) = (\rho_1, \ldots, \rho_m)^{-1}$ and thanks to the composition rule (3.5). Let π denote the restriction of $\sigma \circ (\rho_1, \ldots, \rho_m)$ to the set

$$T := \{t_1, \ldots, t_n\}, \quad t_l := |\boldsymbol{k}^1| + \ldots + |\boldsymbol{k}^l|$$

Since the $\boldsymbol{k}^h$'s are all non-empty, T is a subset of $\{1, \ldots, N\}$ of cardinality n, so after renumbering it we can consider π as an element of $\mathfrak{S}_n$. Now, since ρ_l is an ordered shuffle, it preserves the ordering of $\{|\boldsymbol{k}^{n^{l-1}+1}|, \ldots, |\boldsymbol{k}^{n^{l-1}+1}| + \ldots + |\boldsymbol{k}^{n^l}|\}$, and since σ is a shuffle it preserves the ordering $\{\rho_l(|\boldsymbol{k}^{n^{l-1}+1}|), \ldots, \rho_l(|\boldsymbol{k}^{n^l}|)\} \subseteq \{N^{l-1} + 1, \ldots, N^l\}$. These two facts imply $\pi \in \mathrm{Sh}(n_1, \ldots, n_m)$, and we have

$$\begin{aligned} &\{\sigma \circ (\rho_1, \ldots, \rho_m) \mid \rho_l \in \overline{\mathrm{Sh}}(|\boldsymbol{k}^{n^{l-1}+1}|, \ldots, |\boldsymbol{k}^{n^l}|),\ l = 1, \ldots, m\} \\ =&\ \{\tau \in \mathrm{Sh}(|\boldsymbol{k}^1|, \ldots, |\boldsymbol{k}^n|) \mid \tau(t_{\pi^{-1}(1)}) < \ldots < \tau(t_{\pi^{-1}(n)})\} \\ =:&\ \overline{\mathrm{Sh}}^\pi(|\boldsymbol{k}^1|, \ldots, |\boldsymbol{k}^n|) \end{aligned}$$

since any $\tau \in \mathrm{Sh}(|\boldsymbol{k}^1|,\ldots,|\boldsymbol{k}^n|)$ with $\tau(t_{\pi^{-1}(1)}) < \ldots < \tau(t_{\pi^{-1}(n)})$ for some $\pi \in \mathfrak{S}_n$ factors uniquely as $\sigma \circ (\rho_1,\ldots,\rho_m)$ with σ acting on T with π: this is evident from the fact that each ρ_l acts on the segment $[N^{l-1}+1, N^l]$ and σ acts on the whole segment $[1,N]$ but without altering the order in each $[N^{l-1}+1, N^l]$. This implies

$$\begin{aligned}&\{\sigma \circ (\rho_1,\ldots,\rho_m) \mid \sigma \in \mathrm{Sh}(N_1,\ldots,N_m);\\ &\rho_l \in \overline{\mathrm{Sh}}(|\boldsymbol{k}^{n^{l-1}+1}|,\ldots,|\boldsymbol{k}^{n^l}|),\ l=1,\ldots,m\}\\ = &\bigsqcup_{\pi \in \mathrm{Sh}(n_1,\ldots,n_m)} \overline{\mathrm{Sh}}^\pi(|\boldsymbol{k}^1|,\ldots,|\boldsymbol{k}^n|)\end{aligned}$$

because as σ ranges over $\mathrm{Sh}(N_1,\ldots,N_m)$ all $\mathrm{Sh}(n_1,\ldots,n_m) \ni \pi$'s are obtained, and the $\overline{\mathrm{Sh}}^\pi$'s are mutually disjoint since the ordered shuffle relations imposed by different π's are mutually exclusive. Since

$$\overline{\mathrm{Sh}}(\boldsymbol{k}^{\pi^{-1}(1)},\ldots,\boldsymbol{k}^{\pi^{-1}(n)}) = \{\{\tau_*^{-1}(\boldsymbol{k}^1\ldots\boldsymbol{k}^n) \mid \tau \in \overline{\mathrm{Sh}}^\pi(|\boldsymbol{k}^1|,\ldots,|\boldsymbol{k}^n|)\}\}$$

the proof is concluded. ■

Example 3.5. We illustrate the idea behind this lemma with an example. Let

$$n_1 = 2,\ n_2 = 1; \quad |\boldsymbol{k}^1| = 3,\ |\boldsymbol{k}^2| = 2,\ |\boldsymbol{k}^3| = 3 \tag{3.22}$$

and with the notations of the proof let

$$\begin{aligned}\rho_1 &= \begin{pmatrix}1&2&3&4&5\\1&3&4&2&5\end{pmatrix} \in \overline{\mathrm{Sh}}(3,2),\ \rho_2 = \begin{pmatrix}1&2&3\\1&2&3\end{pmatrix} \in \overline{\mathrm{Sh}}(3),\\ \Rightarrow\ (\rho_1,\rho_2) &= \begin{pmatrix}1&2&3&4&5&6&7&8\\1&3&4&2&5&6&7&8\end{pmatrix}\\ \sigma &= \begin{pmatrix}1&2&3&4&5&6&7&8\\2&3&6&7&8&1&4&5\end{pmatrix} \in \mathrm{Sh}(5,3)\\ \Rightarrow\ \tau \coloneqq \sigma\circ(\rho_1,\rho_2) &= \begin{pmatrix}1&2&3&4&5&6&7&8\\2&6&7&3&8&1&4&5\end{pmatrix} \in \overline{\mathrm{Sh}}^\pi(3,2,3)\\ \text{with } \pi &= \begin{pmatrix}1&2&3\\2&3&1\end{pmatrix} \in \mathrm{Sh}(2,1)\end{aligned}$$

because the restriction of $\sigma \circ (\rho_1,\rho_2)$ to $\{|\boldsymbol{k}^1|, |\boldsymbol{k}^1+\boldsymbol{k}^2|, |\boldsymbol{k}^1+\boldsymbol{k}^2+\boldsymbol{k}^3|\}$ is $\left(\begin{smallmatrix}3&5&8\\7&8&5\end{smallmatrix}\right)$, which coincides with π after renumbering domain and codomain. Also note how, given τ and the numbers (3.22) one can recover σ, ρ_1, ρ_2: first obtain ρ_1, ρ_2 considering how τ orders the segments $[1,5]$, $[6,8]$ (and renumbering) and $\sigma = \tau \circ (\rho_1,\ldots,\rho_m)$. We therefore have, writing terminal elements in red

$$\begin{aligned}(\sigma\circ(\rho_1,\rho_2))_*^{-1}(\boldsymbol{k}_1\boldsymbol{k}_2\boldsymbol{k}_3) &= (k_1^3, k_1^1, k_1^2, k_2^3, k_3^3, k_2^1, k_3^1, k_2^2)\\ &\in \mathrm{Sh}(\overline{\mathrm{Sh}}(\boldsymbol{k}^1,\boldsymbol{k}^2), \overline{\mathrm{Sh}}(\boldsymbol{k}^3))\\ &\in \overline{\mathrm{Sh}}(\boldsymbol{k}^3,\boldsymbol{k}^1,\boldsymbol{k}^2) = \overline{\mathrm{Sh}}(\boldsymbol{k}^{\pi^{-1}(1)},\boldsymbol{k}^{\pi^{-1}(2)},\boldsymbol{k}^{\pi^{-1}(3)})\end{aligned}$$

which can also be written as

$$(k_1^3, k_1^1, k_1^2, k_2^3, {\color{red}k_3^3}, k_2^1, {\color{red}k_3^1}, {\color{red}k_2^2}) = \eta_*^{-1}(\boldsymbol{k}^3, \boldsymbol{k}^1, \boldsymbol{k}^2) \in \overline{\mathrm{Sh}}(\boldsymbol{k}^3, \boldsymbol{k}^1, \boldsymbol{k}^2)$$
$$\text{with } \eta = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 4 & 5 & 2 & 6 & 7 & 3 & 8 \end{pmatrix} \in \overline{\mathrm{Sh}}(3, 3, 2)$$

We primarily use Lemma 3.4 in the two cases $n_1 = \ldots = n_m = 1$ with m arbitrary, and $m = 2$ with n_1, n_2 arbitrary; the former admits the following concise reformulation. Given a vector space W, let $\odot$ denote symmetric tensor product, and

$$\odot_m := \frac{1}{m!} \sum_{\pi \in \mathfrak{S}_m} \pi_* \colon W^{\otimes m} \twoheadrightarrow W^{\odot m}, \quad w_1 \otimes \cdots \otimes w_m \mapsto w_1 \odot \cdots \odot w_m \tag{3.23}$$

denote the symmetrisation map. When referring to the external tensor product we will replace the symbol $\odot$ with $\boxdot$.

Corollary 3.6. *The diagram*

$$\begin{array}{ccc} & & T(V)^{\boxtimes m} \\ & \nearrow^{\widetilde{\Delta}^m_{\shuffle}} & \\ T(V) & & \Big\downarrow {\scriptstyle m!\boxdot_m} \\ & \searrow_{\widetilde{\Delta}^m_{\shuffle}} & \\ & & T(V)^{\boxtimes m} \end{array} \tag{3.24}$$

commutes.

Proof. The statement in coordinates reads

$$\sum_{\substack{\pi \in \mathfrak{S}_m \\ \boldsymbol{h} \in \overline{\mathrm{Sh}}(\boldsymbol{k}^{\pi(1)}, \ldots, \boldsymbol{k}^{\pi(m)})}} a^{\boldsymbol{h}} = \sum_{\boldsymbol{k} \in \mathrm{Sh}(\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m)} a^{\boldsymbol{k}}$$

for non-empty tuples $\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m$. Indeed, we have

$$\begin{aligned} (m! \boxdot \widetilde{\Delta}_{\shuffle} a)^{\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m} &= \Big(\sum_{\substack{\pi \in \mathfrak{S}_m \\ (\widetilde{a})^m_{\shuffle}}} \pi_*(a_{(1)} \boxtimes \cdots \boxtimes a_{(m)}) \Big)^{\boldsymbol{k}^1, \ldots, \boldsymbol{k}^m} \\ &= \sum_{\pi \in \mathfrak{S}_m} \sum_{(\widetilde{a})^m_{\shuffle}} a_{(\pi(1))}^{\boldsymbol{k}^1} \cdots a_{(\pi(m))}^{\boldsymbol{k}^m} \\ &= \sum_{\pi \in \mathfrak{S}_m} \sum_{(\widetilde{a})^m_{\shuffle}} a_{(1)}^{\boldsymbol{k}^{\pi(1)}} \cdots a_{(m)}^{\boldsymbol{k}^{\pi(m)}} \\ &= \sum_{\pi \in \mathfrak{S}_m} (\widetilde{\Delta}^m_{\shuffle} a)^{\boldsymbol{k}^{\pi(1)}, \ldots, \boldsymbol{k}^{\pi(m)}} \\ &= \sum_{\substack{\pi \in \mathfrak{S}_m \\ \boldsymbol{h} \in \overline{\mathrm{Sh}}(\boldsymbol{k}^{\pi(1)}, \ldots, \boldsymbol{k}^{\pi(m)})}} a^{\boldsymbol{h}} \end{aligned}$$

To prove the claim in coordinates we must show the identity of sets

$$\bigsqcup_{\pi \in \mathfrak{S}_m} \overline{\mathrm{Sh}}(\boldsymbol{k}^{\pi(1)}, \dots, \boldsymbol{k}^{\pi(m)}) = \mathrm{Sh}(\boldsymbol{k}^1, \dots, \boldsymbol{k}^m)$$

for tuples $\boldsymbol{k}^1, \dots, \boldsymbol{k}^m$ of positive order. This is precisely Lemma 3.4 with $n_1 = \dots = n_m = 1$, since $\mathrm{Sh}(n_1, \dots, n_m) = \mathfrak{S}_m$ and we may replace π^{-1} with π. ■

3.2 Geometric rough paths

We denote $T^N(V)$ the vector subspace of $T(V)$ given by all tensors of degree $\leq N$, and super/subscripts of $\leq N, \geq M$ denote truncations of the algebra to tensors of the degrees expressed (e.g. for $a \in T(V)$ to belong to $T^N(V)$ it means that $a^n = 0$ for $n > N$, or equivalently $a = a^{\leq N}$). We will similarly use $[M, N]$ as a super/subscript to denote tensors of degrees n with $M \leq n \leq N$. Control functions (as defined in Subsection 2.1.1) will be denoted ω.

Definition 3.7 (Weakly geometric rough path)**.** Let $T > 0, p \geq 1$ and ω be a control on $[0, T]$. A *p-weakly geometric rough path* $\boldsymbol{X}$ controlled by ω, defined on $[0, T]$ and with values in V may be defined as a continuous map

$$\boldsymbol{X} \colon \Delta_T \to T^{\lfloor p \rfloor}(V) \tag{3.25}$$

with $\boldsymbol{X}^0 = 1$ and satisfying the following properties, which we first present in coordinate-free form and subsequently in coordinates w.r.t. a basis of V:

Regularity. $\displaystyle\sup_{0 \leq s < t \leq T} \frac{|\boldsymbol{X}^n_{st}|}{\omega(s,t)^{n/p}} < \infty$, or $\displaystyle\sup_{0 \leq s < t \leq T} \frac{|\boldsymbol{X}^{\boldsymbol{k}}_{st}|}{\omega(s,t)^{|\boldsymbol{k}|/p}} < \infty$ for $n = |\boldsymbol{k}| = 1, \dots \lfloor p \rfloor$;

Multiplicativity. $(\boldsymbol{X}_{su} \otimes \boldsymbol{X}_{ut})^{\leq \lfloor p \rfloor} = \boldsymbol{X}_{st}$, i.e. $\displaystyle\boldsymbol{X}^{\boldsymbol{k}}_{st} = \sum_{(\boldsymbol{i},\boldsymbol{j}) = \boldsymbol{k}} \boldsymbol{X}^{\boldsymbol{i}}_{su} \boldsymbol{X}^{\boldsymbol{j}}_{ut}$ for $|\boldsymbol{k}| \leq \lfloor p \rfloor$ and $0 \leq s \leq u \leq t \leq T$;

Integration by parts. $\boldsymbol{X}_{st} \boxtimes \boldsymbol{X}_{st} = \Delta_{\sqcup\!\sqcup} \boldsymbol{X}_{st}$, or $\displaystyle\boldsymbol{X}^{\boldsymbol{i}}_{st} \boldsymbol{X}^{\boldsymbol{j}}_{st} = \sum_{\boldsymbol{k} \in \mathrm{Sh}(\boldsymbol{i},\boldsymbol{j})} \boldsymbol{X}^{\boldsymbol{k}}_{st}$ for all $0 \leq s \leq t \leq T$.

Let $\mathscr{C}^p_\omega([0, T], V)$ denote the set of all such maps.

In the following we will refer to such objects as "rough paths", dropping the "weakly geometric", since these are the only rough paths that we will be considering in this chapter. We will sometimes refer to the third property above as geometricity, since it distinguishes weakly geometric rough paths among the more general branched rough paths, which are treated in Chapter 4 since they require a different algebra. We will always denote rough paths in bold. The last condition is usually stated by saying that $\boldsymbol{X}$ takes values in the group $G^{\lfloor p \rfloor}(V)$, defined in [CDLL16, Definition 2.9]. We will denote $X \coloneqq \boldsymbol{X}^1$ the *trace* of X: when equipped with an initial value X_0 (which will often be provided) this is an element of $\mathcal{C}^p_\omega([0, T], V)$ defined as the set of continuous paths $Y \colon [0, T] \to V$ s.t., denoting $Y_{st} \coloneqq Y_t - Y_s$ (a notation that will be used for paths throughout)

$$\sup_{0 \leq s < t \leq T} \frac{|Y_{st}|}{\omega(s,t)^{1/p}} < \infty \tag{3.26}$$

It is sufficient to define $\boldsymbol{X}$ to take values in $T^{\lfloor p \rfloor}(V)$, as [Lyo98, Theorem 2.2.1] shows that there exists a unique extension of $\widehat{\boldsymbol{X}}$ of $\boldsymbol{X}$ to $T(V)$ which satisfies the above three properties, and for $m > \lfloor p \rfloor$ is given by

$$\widehat{\boldsymbol{X}}^m_{st} = \lim_{n \to \infty} \Big(\bigotimes_{[u,v] \in \pi_n} \boldsymbol{X}_{uv} \Big)^m \tag{3.27}$$

where $(\pi_n)_n$ is any sequence of partitions on $[s,t]$ with vanishing step size as $n \to \infty$.

The following proposition states that the symmetric part of a weakly geometric rough path is entirely determined by its trace. Given $\ell \in \mathcal{L}(T^N(V), U)$ ($\mathcal{L}$ denotes the space of linear maps) and $a \in T(V)$ we will denote $\langle \ell, a \rangle = \ell(a)$ the evaluation of ℓ on a. We will always identify $\mathcal{L}(\mathbb{R}, U) = U$ by setting $\ell \mapsto \ell(1)$.

Proposition 3.8. *For $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T], V)$ we have $n! \odot_n \boldsymbol{X}^n_{st} = X^{\otimes n}_{st}$.*

Proof. We proceed by induction on n. For $n = 0, 1$ there is nothing to prove. For the inductive step we will need the following fact: each $\pi \in \mathfrak{S}_{n+1}$ can be expressed uniquely as $\sigma \circ \rho$ with ρ in the stabiliser of $n+1$ (a subgroup of $\mathfrak{S}_{n+1}$ isomorphic to $\mathfrak{S}_n$) and $\sigma \in \mathrm{Sh}(n,1)$: indeed, if $m \coloneqq \pi(n+1)$ we may set

$$\sigma(k) \coloneqq \begin{cases} k & 1 \le k \le m-1 \\ k+1 & m \le k \le n \\ m & k = n+1 \end{cases} \qquad \rho(k) \coloneqq \begin{cases} \pi(k) & 1 \le \pi(k) \le m-1 \\ \pi(k) - 1 & m \le \pi(k) \le n \\ n+1 & k = n+1 \end{cases}$$

Uniqueness follows from a counting argument, since there are $n!$ choices for ρ and $n+1$ for σ. We then compute

$$\begin{aligned} (n+1)! \odot_{n+1} \boldsymbol{X}^{n+1}_{st} &= \sum_{\pi \in \mathfrak{S}_{n+1}} \pi_* \boldsymbol{X}^{n+1}_{st} \\ &= \sum_{\rho \in \mathfrak{S}_n} \rho_* \sum_{\sigma \in \mathrm{Sh}(n,1)} \sigma_* \boldsymbol{X}^{n+1}_{st} \\ &= \sum_{\rho \in \mathfrak{S}_n} \rho_* (\Delta_\sqcup\!\sqcup \boldsymbol{X}_{st})^{n,1} \\ &= \sum_{\rho \in \mathfrak{S}_n} \rho_* \boldsymbol{X}^n_{st} \otimes X_{st} \\ &= X^{\otimes(n+1)}_{st} \end{aligned}$$

where we have used the geometricity axiom and the inductive hypothesis. ∎

We proceed to define the objects that can be regarded as dual to rough paths, original to [Gub04]. In what follows we will write $\approx_m$ between two real-valued quantities dependent on $0 \le s \le t \le T$ to mean that their difference lies in $O(\omega(s,t)^{m/p})$ as $t \searrow s$, and simply $\approx$ to mean $\approx_{\lfloor p \rfloor + 1}$. We will frequently use the following properties, which are trivial to check:

$$\begin{aligned} a_{st} \approx_m b_{st} \approx_n c_{st} &\Rightarrow a_{st} \approx_{n \wedge m} c_{st} \\ a_{st} \approx_m b_{st},\ c_{st} \approx_n 0 &\Rightarrow a_{st} c_{st} \approx_{m+n} b_{st} c_{st} \end{aligned} \tag{3.28}$$

from which we deduce more generally

$$\begin{aligned} &a_{st} \approx_{m_1} b_{st},\ a_{st}, b_{st} \approx_{n_1} 0, \quad c_{st} \approx_{m_2} d_{st},\ c_{st}, d_{st} \approx_{n_2} 0 \\ \Rightarrow\ &a_{st}c_{st} \approx_{m_1+n_2} b_{st}c_{st} \approx_{m_2+n_1} b_{st}d_{st} \\ \Rightarrow\ &a_{st}c_{st} \approx_{(m_1+n_2)\wedge(m_2+n_1)} b_{st}d_{st} \end{aligned} \tag{3.29}$$

If a continuous map $\widetilde{\boldsymbol{X}}\colon \Delta_T \to T^{\lfloor p \rfloor}(V)$ satisfies the regularity and integration by parts conditions, and satisfies the multiplicativity condition with a "$\approx$" replacing the "$=$" (*almost multiplicative*), it defines a rough path by [Lyo98, Theorem 3.3.1], by taking the limit (3.27) (w.r.t. $\widetilde{\boldsymbol{X}}$), and this rough path $\boldsymbol{X}$ is unique with the property that $\widetilde{\boldsymbol{X}}_{st} \approx \boldsymbol{X}_{st}$. The following lemma tells us that this is also true if the integration by parts condition only holds with an $\approx$ (*almost geometric*). In light of this, we will break with the literature in defining an *almost rough path* as an $\widetilde{\boldsymbol{X}}$ that satisfies the regularity condition in Definition 3.7 and is almost multiplicative and almost geometric.

Proposition 3.9 (Almost rough paths)**.** *Let $\widetilde{\boldsymbol{X}}$ be a V-valued almost p-rough path. Then there exists a unique p-rough path $\boldsymbol{X}$ with the property that $\boldsymbol{X}_{st} \approx \widetilde{\boldsymbol{X}}_{st}$.*

Proof. We use that the shuffle algebra is free abelian on the Lyndon words (e.g. over the rationals) [Reu93, Theorem 6.1] to define an intermediate $\overline{\boldsymbol{X}}$: set $\overline{\boldsymbol{X}}^{\boldsymbol{h}} \coloneqq \widetilde{\boldsymbol{X}}^{\boldsymbol{h}}$ if $\boldsymbol{h}$ is a Lyndon word with $|\boldsymbol{h}| \leq \lfloor p \rfloor$, and for a tuple $\boldsymbol{k}$ with $|\boldsymbol{k}| < \lfloor p \rfloor$ expressed (uniquely up to order of factors) as $\sum_\lambda c_\lambda \boldsymbol{k}^1_\lambda \sqcup\!\sqcup \cdots \sqcup\!\sqcup \boldsymbol{k}^{n_\lambda}_\lambda$ with $c_\lambda \in \mathbb{R}$ and the $\boldsymbol{k}^j_\lambda$'s (not necessarily distinct) Lyndon words, set

$$\overline{\boldsymbol{X}}^{\boldsymbol{k}} \coloneqq \sum_\lambda c_\lambda \overline{\boldsymbol{X}}^{\boldsymbol{k}^1_\lambda} \cdots \overline{\boldsymbol{X}}^{\boldsymbol{k}^{n_\lambda}_\lambda} = \sum_\lambda c_\lambda \widetilde{\boldsymbol{X}}^{\boldsymbol{k}^1_\lambda} \cdots \widetilde{\boldsymbol{X}}^{\boldsymbol{k}^{n_\lambda}_\lambda} \approx \sum_\lambda c_\lambda \langle \boldsymbol{k}^1_\lambda \sqcup\!\sqcup \cdots \sqcup\!\sqcup \boldsymbol{k}^{n_\lambda}_\lambda, \widetilde{\boldsymbol{X}} \rangle = \widetilde{\boldsymbol{X}}^{\boldsymbol{k}}$$

since $\widetilde{\boldsymbol{X}}$ is almost geometric. $\overline{\boldsymbol{X}}$ is then $\approx \widetilde{\boldsymbol{X}}$, it satisfies integration by parts (exactly) by construction, and is still almost multiplicative since

$$\overline{\boldsymbol{X}}^{\boldsymbol{k}} \approx \boldsymbol{X}^{\boldsymbol{k}}_{st} = \sum_{(\boldsymbol{i},\boldsymbol{j})=\boldsymbol{k}} \boldsymbol{X}^{\boldsymbol{i}}_{su} \boldsymbol{X}^{\boldsymbol{j}}_{ut} \approx \sum_{(\boldsymbol{i},\boldsymbol{j})=\boldsymbol{k}} \overline{\boldsymbol{X}}^{\boldsymbol{i}}_{su} \overline{\boldsymbol{X}}^{\boldsymbol{j}}_{ut}$$

Existence of $\boldsymbol{X}$ then follows immediately by applying the above-referenced result to $\overline{\boldsymbol{X}}$ and uniqueness follows from the fact that if $\boldsymbol{X}'$ is a second p-rough path satisfying the statement of this proposition, then we have $\boldsymbol{X}' \approx \widetilde{\boldsymbol{X}} \approx \overline{\boldsymbol{X}} \Rightarrow \boldsymbol{X}' = \boldsymbol{X}$ again by the same result. ■

Definition 3.10. Let $\boldsymbol{X}$ be as above and U another vector space. An U-valued $\boldsymbol{X}$*-controlled path* $\boldsymbol{H}$ is an element of $\mathcal{C}^p_\omega([0,T], \mathcal{L}(T^{\lfloor p \rfloor - 1}(V), U))$ (where ω is the control for $\boldsymbol{X}$) s.t. for $n = 0, \ldots, \lfloor p \rfloor - 2$ and each $a \in V^{\otimes n}$

$$\langle \boldsymbol{H}_{n;t}, a \rangle \approx_{\lfloor p \rfloor - n} \langle \boldsymbol{H}_{[n,\lfloor p \rfloor - 1];s}, \boldsymbol{X}^{\leq \lfloor p \rfloor - 1 - n}_{st} \otimes a \rangle \tag{3.30}$$

Here $\boldsymbol{H}_n$ denotes the n-th level of $\boldsymbol{H}$. Denote $\mathscr{D}_{\boldsymbol{X}}(U)$ the vector space of all U-valued $\boldsymbol{X}$-controlled paths.

The maps $\boldsymbol{H}_n \coloneqq \boldsymbol{H}|_{V^{\otimes n}}$ are known as the *Gubinelli derivatives* of H and $H \coloneqq \boldsymbol{H}_0 \in U$ is called the *trace* of $\boldsymbol{H}$ (note the discrepancy with rough paths: for these the trace is the order-1 component). Note that the defining condition only involves $\boldsymbol{X}^{\leq \lfloor p \rfloor - 1}$, and that it holds automatically at level $\lfloor p \rfloor - 1$ by regularity of

$\boldsymbol{H}$. In coordinates it reads

$$\boldsymbol{H}^k_{\boldsymbol{\beta};t} \approx_{\lfloor p \rfloor - |\boldsymbol{\beta}|} \sum_{|\boldsymbol{\alpha}|=0}^{\lfloor p \rfloor - 1 - |\boldsymbol{\beta}|} \boldsymbol{H}^k_{(\boldsymbol{\alpha},\boldsymbol{\beta});s} \boldsymbol{X}^{\boldsymbol{\alpha}}_{st}, \quad 0 \leq |\boldsymbol{\beta}| \leq \lfloor p \rfloor - 2 \tag{3.31}$$

Here the superscript k refers to the value of $\boldsymbol{H}$ in U (and will often be omitted when unnecessary), and the sum is not only on the length $|\boldsymbol{\alpha}|$ of the tuple $\boldsymbol{\alpha}$, but on the tuple itself. For the branched version of this definition, see [HK15].

An important case is when $U = \mathcal{L}(V, W)$ for another vector space W: by the tensor-hom adjunction we then have

$$\mathcal{L}(T^{\lfloor p \rfloor - 1}(V), \mathcal{L}(V, W)) = \mathcal{L}(T^{\lfloor p \rfloor - 1}(V) \otimes V, W) = \mathcal{L}\Big(\bigoplus_{n=1}^{\lfloor p \rfloor} V^{\otimes n}, W\Big) \tag{3.32}$$

We will use angle brackets and coordinate notation for linear maps accordingly, i.e. the last slot in a bracket or in a tuple will refer to the copy of V in the target space of the original linear map. We will call controlled paths valued in $\mathcal{L}(V, W)$ *W-valued controlled integrands*, and we may rewrite (3.30) as

$$\langle \boldsymbol{H}_t, b \rangle \approx_{\lfloor p \rfloor - n + 1} \langle \boldsymbol{H}_{[n, \lfloor p \rfloor];s}, \boldsymbol{X}^{[0, \lfloor p \rfloor - n]}_{st} \otimes b \rangle \in W, \quad b \in V^{\otimes n}, \quad n = 1, \ldots, \lfloor p \rfloor - 1 \tag{3.33}$$

or in coordinates

$$\boldsymbol{H}^k_{\boldsymbol{\beta};t} \approx_{\lfloor p \rfloor - |\boldsymbol{\beta}| + 1} \sum_{|\boldsymbol{\alpha}|=0}^{\lfloor p \rfloor - |\boldsymbol{\beta}|} \boldsymbol{H}^k_{(\boldsymbol{\alpha},\boldsymbol{\beta});s} \boldsymbol{X}^{\boldsymbol{\alpha}}_{st}, \quad 1 \leq |\boldsymbol{\beta}| \leq \lfloor p \rfloor - 1 \tag{3.34}$$

The next example contains a very important example of controlled path.

Example 3.11 (Smooth functions of X)**.** Let $F \in C^\infty(V, U)$, then

$$t \mapsto (F(X_t), DF(X_t), \ldots, D^{\lfloor p \rfloor - 1} F(X_t)) \in \mathcal{L}(T^{\lfloor p \rfloor - 1}(V), U) \tag{3.35}$$

is an $\boldsymbol{X}$-controlled path, which we denote simply $F(X)$, or $\boldsymbol{F}(X)$ if we want to emphasise that we are considering the full controlled path, not just its trace. Indeed, denoting by $\partial_{\boldsymbol{\gamma}} F$ the order-$|\boldsymbol{\gamma}|$ partial derivative of F in the directions of the chosen basis determined by the tuple $\boldsymbol{\gamma}$, we have, for $0 \leq |\boldsymbol{\beta}| \leq \lfloor p \rfloor - 2$

$$\begin{aligned}
&F(X_t)_{\boldsymbol{\beta}} - \sum_{|\boldsymbol{\alpha}|=0}^{\lfloor p \rfloor - 1 - |\boldsymbol{\beta}|} F(X_s)_{(\boldsymbol{\alpha},\boldsymbol{\beta})} \boldsymbol{X}^{\boldsymbol{\alpha}}_{st} \\
&= \partial_{\boldsymbol{\beta}} F(X_t) - \sum_{|\boldsymbol{\alpha}|=0}^{\lfloor p \rfloor - 1 - |\boldsymbol{\beta}|} \partial_{\boldsymbol{\alpha},\boldsymbol{\beta}} F(X_s) \boldsymbol{X}^{\boldsymbol{\alpha}}_{st} \\
&= \partial_{\boldsymbol{\beta}} F(X_t) - \sum_{n=0}^{\lfloor p \rfloor - 1 - |\boldsymbol{\beta}|} \frac{1}{n!} \partial_{\boldsymbol{\alpha},\boldsymbol{\beta}} F(X_s) X^{\alpha_1}_{st} \cdots X^{\alpha_n}_{st} \\
&\approx_{\lfloor p \rfloor - |\boldsymbol{\beta}|} 0
\end{aligned}$$

where we have used Proposition 3.8 together with the symmetry of higher differentials and Taylor's approximation. Note that the symmetry of Gubinelli derivatives is a special feature of this kind of controlled path, and cannot be expected to hold in general. When $U = \mathcal{L}(V, W)$ we shall call F an W*-valued one-form*, and we adopt the convention

$$\langle F(X), v_1 \otimes \cdots \otimes v_{n+1} \rangle = D^n F(X)(v_1, \ldots, v_n)(v_{n+1}) \in W$$

or in coordinates $F(X)_{\boldsymbol{\alpha},\beta} = \partial_{\boldsymbol{\alpha}} F_\beta(X)$.

The next lemma is necessary for the definition of rough integral.

Lemma 3.12. *Let* $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T], V)$ *and* $\boldsymbol{H} \in \mathscr{D}_{\boldsymbol{X}}(\mathcal{L}(V, W))$. *Define, for* $0 \le s \le t \le T$

$$\Xi_{st} \coloneqq \langle \boldsymbol{H}_s, \boldsymbol{X}^{\ge 1}_{st} \rangle \in W \tag{3.36}$$

where the evaluation is taken under the identification (3.32). *Then* Ξ *is almost additive: for all* $0 \le s \le u \le t \le T$

$$\Xi_{st} - \Xi_{su} - \Xi_{ut} \approx 0 \tag{3.37}$$

Proof. Using the multiplicativity axiom, the regularity of $\boldsymbol{H}$ and (3.34) (together with (3.28)) we may write

$$\begin{aligned}
&\Xi_{st} - \Xi_{su} - \Xi_{ut} \\
&= \sum_{|\boldsymbol{\gamma}|=1}^{\lfloor p \rfloor} (\boldsymbol{H}_{\boldsymbol{\gamma};s} \boldsymbol{X}^{\boldsymbol{\gamma}}_{st} - \boldsymbol{H}_{\boldsymbol{\gamma};s} \boldsymbol{X}^{\boldsymbol{\gamma}}_{su} - \boldsymbol{H}_{\boldsymbol{\gamma};u} \boldsymbol{X}^{\boldsymbol{\gamma}}_{ut}) \\
&= \sum_{|\boldsymbol{\gamma}|=1}^{\lfloor p \rfloor} (\boldsymbol{H}_{\boldsymbol{\gamma};s} (\boldsymbol{X}^{\boldsymbol{\gamma}}_{st} - \boldsymbol{X}^{\boldsymbol{\gamma}}_{su} - \boldsymbol{X}^{\boldsymbol{\gamma}}_{ut}) - \boldsymbol{H}_{\boldsymbol{\gamma};su} \boldsymbol{X}^{\boldsymbol{\gamma}}_{ut}) \\
&= \sum_{|\boldsymbol{\gamma}|=1}^{\lfloor p \rfloor} \Bigg(\boldsymbol{H}_{\boldsymbol{\gamma};s} \sum_{\substack{(\boldsymbol{\alpha},\boldsymbol{\beta})=\boldsymbol{\gamma} \\ |\boldsymbol{\alpha}|,|\boldsymbol{\beta}|\ge 1}} \boldsymbol{X}^{\boldsymbol{\alpha}}_{su} \boldsymbol{X}^{\boldsymbol{\beta}}_{ut} - \boldsymbol{H}_{\boldsymbol{\gamma};su} \boldsymbol{X}^{\boldsymbol{\gamma}}_{ut} \Bigg) \\
&\approx \sum_{\substack{|\boldsymbol{\alpha}|,|\boldsymbol{\beta}|\ge 1 \\ |\boldsymbol{\alpha}|+|\boldsymbol{\beta}| \le \lfloor p \rfloor}} \boldsymbol{H}_{(\boldsymbol{\alpha},\boldsymbol{\beta});s} \boldsymbol{X}^{\boldsymbol{\alpha}}_{su} \boldsymbol{X}^{\boldsymbol{\beta}}_{ut} - \sum_{|\boldsymbol{\varepsilon}|=1}^{\lfloor p \rfloor - 1} \boldsymbol{H}_{\boldsymbol{\varepsilon};su} \boldsymbol{X}^{\boldsymbol{\varepsilon}}_{ut} \\
&\approx \sum_{\substack{|\boldsymbol{\alpha}|,|\boldsymbol{\beta}|\ge 1 \\ |\boldsymbol{\alpha}|+|\boldsymbol{\beta}| \le \lfloor p \rfloor}} \boldsymbol{H}_{(\boldsymbol{\alpha},\boldsymbol{\beta});s} \boldsymbol{X}^{\boldsymbol{\alpha}}_{su} \boldsymbol{X}^{\boldsymbol{\beta}}_{ut} - \sum_{|\boldsymbol{\varepsilon}|=1}^{\lfloor p \rfloor - 1} \sum_{|\boldsymbol{\delta}|=1}^{\lfloor p \rfloor - |\boldsymbol{\varepsilon}|} \boldsymbol{H}_{(\boldsymbol{\delta},\boldsymbol{\varepsilon});s} \boldsymbol{X}^{\boldsymbol{\delta}}_{su} \boldsymbol{X}^{\boldsymbol{\varepsilon}}_{ut} \\
&= 0
\end{aligned}$$

■

Definition 3.13 (Rough integral)**.** Let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T], V)$, $\boldsymbol{H} \in \mathscr{D}_{\boldsymbol{X}}(\mathcal{L}(V, W))$ be as above. We define

$$\int_0^\cdot \boldsymbol{H} \, \mathrm{d}\boldsymbol{X} \colon [0,T] \to W \tag{3.38}$$

to be the unique path $I \in \mathcal{C}_\omega([0,T], W)$ with the property that $I_{st} \approx \Xi_{st}$, which exists by [Lyo98, Theorem

3.3.1], and is given by

$$I_{st} = \lim_{n\to\infty} \sum_{[u,v]\in\pi_n} \Xi_{uv}$$

for a sequence of partitions $(\pi_n)_n$ with vanishing step size. We can make $\boldsymbol{\int} \boldsymbol{H} \mathrm{d}\boldsymbol{X}$ into an $\boldsymbol{X}$-controlled path by defining, for $1 \le n \le \lfloor p \rfloor - 1$

$$\left(\boldsymbol{\int} \boldsymbol{H} \mathrm{d}\boldsymbol{X}\right)_n \coloneqq \boldsymbol{H}_{n-1} \in \mathcal{L}(V^{\otimes n-1}, \mathcal{L}(V,W)) = \mathcal{L}(V^{\otimes n}, W)$$

Note the use of bold font for the integral sign, which emphasises membership to $\mathscr{D}_{\boldsymbol{X}}(W)$.

An $\boldsymbol{X}$-controlled path can be made into a rough path in its own right. We use (3.17) as a blueprint for the following definition, where we truncate at the correct order to avoid infinite sums.

Definition 3.14 (Lift of a controlled path)**.** Let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],V)$, $\boldsymbol{H} \in \mathscr{D}_{\boldsymbol{X}}(U)$. Define $\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H} \colon \Delta_T \to T^{\lfloor p \rfloor}(U)$ (notice the partial arrow notation to indicate almost multiplicativity & geometricity) by

$$(\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^0_{st} \coloneqq 1, \quad (\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^1_{st} \coloneqq H_{st} \tag{3.39}$$

and for $2 \le m \le \lfloor p \rfloor$

$$\begin{aligned}(\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^m_{st} &\coloneqq \langle \boldsymbol{H}^{\boxtimes m}_s, \widetilde{\Delta}^m_{\sqcup\!\sqcup} \boldsymbol{X}_{st} \rangle \\ &= \sum_{\substack{n_1,\ldots,n_m \ge 1 \\ n \coloneqq n_1+\ldots+n_m \le \lfloor p \rfloor}} \langle \boldsymbol{H}_{n_1;s} \boxtimes \cdots \boxtimes \boldsymbol{H}_{n_m;s}, (\widetilde{\Delta}_{\sqcup\!\sqcup}\boldsymbol{X})^{n_1,\ldots,n_m}_{st} \rangle\end{aligned} \tag{3.40}$$

As it is shown in Theorem 3.15 below, [Lyo98, Theorem 3.3.1] applies to this functional, and given any sequence of partitions $(\pi_n)_n$ with vanishing step size

$$(\uparrow_{\boldsymbol{X}}\boldsymbol{H})_{st} \coloneqq \lim_{n\to\infty} \Big(\bigotimes_{[u,v]\in\pi_n} (\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})_{uv} \Big) \tag{3.41}$$

defines an element of $\mathscr{C}^p_\omega([0,T],U)$, which we call the *lift* of $\boldsymbol{H}$ to rough path w.r.t. $\boldsymbol{X}$.

Note how it was necessary to distinguish the case $m = 1$ above: this is due to the fact that we do not have the $\lfloor p \rfloor^{\text{th}}$ Gubinelli derivative, and therefore (3.40) would only be accurate at order $\lfloor p \rfloor$ (though in all explicit cases presented here these are known, and (3.40) is applicable for $m = 1$ too; an example where the case distinction is essential would be Example 3.11 with F only $(\lfloor p \rfloor - 1)$-times differentiable). (3.40) can be written dually as

$$(\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^m_{st} = \langle \boldsymbol{H}^{\sqcup\!\sqcup m}_{\ge 1;s}, \boldsymbol{X}_{st} \rangle = \sum_{\substack{n_1,\ldots,n_m \ge 1 \\ n \coloneqq n_1+\ldots+n_m \le \lfloor p \rfloor}} \langle \boldsymbol{H}^{n_1}_s \sqcup\!\sqcup \cdots \sqcup\!\sqcup \boldsymbol{H}^{n_m}_s, \boldsymbol{X}^n_{st} \rangle \tag{3.42}$$

and in coordinates as

$$
\begin{aligned}
(\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^{()} = 1, \quad (\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^{k} = H^k \\
(\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^{(k_1,\ldots,k_m)}_{st} = \sum_{\substack{|\boldsymbol{\gamma}^1|,\ldots,|\boldsymbol{\gamma}^m|\geq 1\\ |\boldsymbol{\gamma}^1|+\ldots+|\boldsymbol{\gamma}^m|\leq\lfloor p\rfloor\\ \boldsymbol{\gamma}\in\overline{\mathrm{Sh}}(\boldsymbol{\gamma}^1,\ldots,\boldsymbol{\gamma}^m)}} \boldsymbol{H}^{k_1}_{\boldsymbol{\gamma}^1;s}\cdots\boldsymbol{H}^{k_m}_{\boldsymbol{\gamma}^m;s}\boldsymbol{X}^{\boldsymbol{\gamma}}_{st}
\end{aligned}
\tag{3.43}
$$

In explicit calculations we will use $\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H}$, for which we have a combinatorial expression, as a proxy for the true lift $\uparrow_{\boldsymbol{X}}\boldsymbol{H}$.

The following is one of the main theorems in this chapter. It can be compared with [LCL07, Theorem 4.6], which applies to the special case of integrals of $\mathrm{Lip}(\gamma)$ forms, covered in Example 3.17 below. Their proof makes use of the symmetry of $\mathrm{Lip}(\gamma)$ forms, while the lemma below does not require it.

Theorem 3.15. *$\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H}$ is an almost rough path. Therefore the limit taken in (3.41) exists and defines a U-valued p-weakly geometric rough path, controlled by ω on $[0,T]$, with trace H.*

Proof. We begin by showing almost multiplicativity, i.e. that for $|\boldsymbol{k}| = 0,\ldots,\lfloor p\rfloor$ and $0\leq s\leq u\leq t\leq T$

$$
\sum_{(\boldsymbol{i},\boldsymbol{j})=\boldsymbol{k}} (\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^{\boldsymbol{i}}_{su}(\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^{\boldsymbol{j}}_{ut} \approx (\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^{\boldsymbol{k}}_{st}
$$

For $|\boldsymbol{k}| = 0$ this is trivial and for $|\boldsymbol{k}| = 1$ it coincides with the statement that H is a path. For $|\boldsymbol{k}| = 2$ (which presupposes $\lfloor p\rfloor\geq 2$) we have

$$
\begin{aligned}
&\sum_{(\boldsymbol{i},\boldsymbol{j})=(k_1,k_2)} (\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^{\boldsymbol{i}}_{su}(\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^{\boldsymbol{j}}_{ut} \\
&= (\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^{(k_1,k_2)}_{su} + (\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^{k_1}_{su}(\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^{k_2}_{ut} + (\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^{(k_1,k_2)}_{ut} \\
&= \sum_{\substack{|\boldsymbol{\alpha}^1|,|\boldsymbol{\alpha}^2|\geq 1\\ |\boldsymbol{\alpha}^1|+|\boldsymbol{\alpha}^2|\leq\lfloor p\rfloor\\ \boldsymbol{\alpha}\in\overline{\mathrm{Sh}}(\boldsymbol{\alpha}^1,\boldsymbol{\alpha}^2)}} (\boldsymbol{H}^{k_1}_{\boldsymbol{\alpha}^1;s}\boldsymbol{H}^{k_2}_{\boldsymbol{\alpha}^2;s}\boldsymbol{X}^{\boldsymbol{\alpha}}_{su}) + H^{k_1}_{su}H^{k_2}_{ut} + \sum_{\substack{|\boldsymbol{\beta}^1|,|\boldsymbol{\beta}^2|\geq 1\\ |\boldsymbol{\beta}^1|+|\boldsymbol{\beta}^2|\leq\lfloor p\rfloor\\ \boldsymbol{\beta}\in\overline{\mathrm{Sh}}(\boldsymbol{\beta}^1,\boldsymbol{\beta}^2)}} (\boldsymbol{H}^{k_1}_{\boldsymbol{\beta}^1;u}\boldsymbol{H}^{k_2}_{\boldsymbol{\beta}^2;u}\boldsymbol{X}^{\boldsymbol{\beta}}_{ut})
\end{aligned}
\tag{3.44}
$$

We continue the calculation by re-expanding all the $\boldsymbol{H}$ terms at s and using (3.28):

$$
H^{k_1}_{su} \approx_{\lfloor p\rfloor} \sum_{|\boldsymbol{\alpha}^1|=1}^{\lfloor p\rfloor-1} \boldsymbol{H}^{k_1}_{\boldsymbol{\alpha}^1;s}\boldsymbol{X}^{\boldsymbol{\alpha}^1}_{su}
$$

$$
H^{k_2}_{ut} \approx_{\lfloor p\rfloor} \sum_{|\boldsymbol{\beta}|=1}^{\lfloor p\rfloor-1} \boldsymbol{H}_{\boldsymbol{\beta};u}\boldsymbol{X}^{\boldsymbol{\beta}}_{ut} \approx_{\lfloor p\rfloor} \sum_{\substack{|\boldsymbol{\beta}|=1,\ldots,\lfloor p\rfloor-1\\ |\boldsymbol{\alpha}^2|=0,\ldots,\lfloor p\rfloor-1-|\boldsymbol{\beta}|}} \boldsymbol{H}^{k_2}_{(\boldsymbol{\alpha}^2,\boldsymbol{\beta});s}\boldsymbol{X}^{\boldsymbol{\alpha}^2}_{su}\boldsymbol{X}^{\boldsymbol{\beta}}_{ut}
$$

These two identities, the fact that $H_{su}, H_{ut} \approx_1 0$ and (3.29) imply

$$H^{k_1}_{su} H^{k_2}_{ut} \approx \sum_{\substack{|\boldsymbol{\beta}|=1,\dots,\lfloor p\rfloor-1 \\ |\boldsymbol{\alpha}^1|=1,\dots,\lfloor p\rfloor-1 \\ |\boldsymbol{\alpha}^2|=0,\dots,\lfloor p\rfloor-1-|\boldsymbol{\beta}| \\ \boldsymbol{\alpha}\in\mathrm{Sh}(\boldsymbol{\alpha}^1,\boldsymbol{\alpha}^2)}} \boldsymbol{H}^{k_1}_{\boldsymbol{\alpha}^1;s} \boldsymbol{H}^{k_2}_{(\boldsymbol{\alpha}^2,\boldsymbol{\beta});s} \boldsymbol{X}^{\boldsymbol{\alpha}}_{su} \boldsymbol{X}^{\boldsymbol{\beta}}_{ut}$$

Similarly

$$\boldsymbol{H}^{k_1}_{\boldsymbol{\beta}^1;u} \boldsymbol{H}^{k_2}_{\boldsymbol{\beta}^2;u} \boldsymbol{X}^{\boldsymbol{\beta}}_{ut} \approx \sum_{\substack{|\boldsymbol{\alpha}^1|=0,\dots,\lfloor p\rfloor-1-|\boldsymbol{\beta}^1| \\ |\boldsymbol{\alpha}^2|=0,\dots,\lfloor p\rfloor-1-|\boldsymbol{\beta}^2| \\ \boldsymbol{\alpha}\in\mathrm{Sh}(\boldsymbol{\alpha}^1,\boldsymbol{\alpha}^2)}} \boldsymbol{H}^{k_1}_{(\boldsymbol{\alpha}^1,\boldsymbol{\beta}^1);s} \boldsymbol{H}^{k_2}_{(\boldsymbol{\alpha}^2,\boldsymbol{\beta}^2);s} \boldsymbol{X}^{\boldsymbol{\alpha}}_{su} \boldsymbol{X}^{\boldsymbol{\beta}}_{ut}$$

Incorporating these computations in (3.44) we obtain

$$\begin{aligned}
&\sum_{(\boldsymbol{i},\boldsymbol{j})=(k_1,k_2)} (\mathbb{1}_{\boldsymbol{X}}\boldsymbol{H})^{\boldsymbol{i}}_{su} (\mathbb{1}_{\boldsymbol{X}}\boldsymbol{H})^{\boldsymbol{j}}_{ut} \\
&\approx \sum_{\substack{|\boldsymbol{\alpha}^1|,|\boldsymbol{\alpha}^2|\geq 1 \\ |\boldsymbol{\alpha}^1|+|\boldsymbol{\alpha}^2|\leq\lfloor p\rfloor \\ \boldsymbol{\alpha}\in\overline{\mathrm{Sh}}(\boldsymbol{\alpha}^1,\boldsymbol{\alpha}^2)}} (\boldsymbol{H}^{k_1}_{\boldsymbol{\alpha}^1;s} \boldsymbol{H}^{k_2}_{\boldsymbol{\alpha}^2;s} \boldsymbol{X}^{\boldsymbol{\alpha}}_{su}) + \sum_{\substack{|\boldsymbol{\beta}|=1,\dots,\lfloor p\rfloor-1 \\ |\boldsymbol{\alpha}^1|=1,\dots,\lfloor p\rfloor-1 \\ |\boldsymbol{\alpha}^2|=1,\dots,\lfloor p\rfloor-1-|\boldsymbol{\beta}| \\ \boldsymbol{\alpha}\in\mathrm{Sh}(\boldsymbol{\alpha}^1,\boldsymbol{\alpha}^2)}} (\boldsymbol{H}^{k_1}_{\boldsymbol{\alpha}^1;s} \boldsymbol{H}^{k_2}_{(\boldsymbol{\alpha}^2,\boldsymbol{\beta});s} \boldsymbol{X}^{\boldsymbol{\alpha}}_{su} \boldsymbol{X}^{\boldsymbol{\beta}}_{ut}) \\
&\quad + \sum_{\substack{|\boldsymbol{\beta}^1|,|\boldsymbol{\beta}^2|\geq 1 \\ |\boldsymbol{\beta}^1|+|\boldsymbol{\beta}^2|\leq\lfloor p\rfloor \\ |\boldsymbol{\alpha}^1|=0,\dots,\lfloor p\rfloor-1-|\boldsymbol{\beta}^1| \\ |\boldsymbol{\alpha}^2|=0,\dots,\lfloor p\rfloor-1-|\boldsymbol{\beta}^2| \\ \boldsymbol{\alpha}\in\mathrm{Sh}(\boldsymbol{\alpha}^1,\boldsymbol{\alpha}^2) \\ \boldsymbol{\beta}\in\overline{\mathrm{Sh}}(\boldsymbol{\beta}^1,\boldsymbol{\beta}^2)}} (\boldsymbol{H}^{k_1}_{(\boldsymbol{\alpha}^1,\boldsymbol{\beta}^1);s} \boldsymbol{H}^{k_2}_{(\boldsymbol{\alpha}^2,\boldsymbol{\beta}^2);s} \boldsymbol{X}^{\boldsymbol{\alpha}}_{su} \boldsymbol{X}^{\boldsymbol{\beta}}_{ut}) \\
&= \sum_{\substack{|\boldsymbol{\gamma}^1|,|\boldsymbol{\gamma}^2|\geq 1 \\ |\boldsymbol{\gamma}^1|+|\boldsymbol{\gamma}^2|\leq\lfloor p\rfloor}} \boldsymbol{H}^{k_1}_{\boldsymbol{\gamma}^1;s} \boldsymbol{H}^{k_2}_{\boldsymbol{\gamma}^2;s} \sum_{\substack{l=0,1,2 \\ \boldsymbol{\alpha}^h=\boldsymbol{\gamma}^h,\ h\leq l \\ (\boldsymbol{\alpha}^h,\boldsymbol{\beta}^h)=\boldsymbol{\gamma}^h,\ |\boldsymbol{\beta}^h|\geq 1,\ h\geq l+1 \\ \boldsymbol{\alpha}\in\mathrm{Sh}(\overline{\mathrm{Sh}}(\boldsymbol{\alpha}^1,\dots,\boldsymbol{\alpha}^l),\mathrm{Sh}(\boldsymbol{\alpha}^{l+1},\dots,\boldsymbol{\alpha}^2)) \\ \boldsymbol{\beta}\in\overline{\mathrm{Sh}}(\boldsymbol{\beta}^{l+1},\dots,\boldsymbol{\beta}^2)}} \boldsymbol{X}^{\boldsymbol{\alpha}}_{su} \boldsymbol{X}^{\boldsymbol{\beta}}_{ut} \\
&= \sum_{\substack{|\boldsymbol{\gamma}^1|,|\boldsymbol{\gamma}^2|\geq 1 \\ |\boldsymbol{\gamma}^1|+|\boldsymbol{\gamma}^2|\leq\lfloor p\rfloor}} \boldsymbol{H}^{k_1}_{\boldsymbol{\gamma}^1;s} \boldsymbol{H}^{k_2}_{\boldsymbol{\gamma}^2;s} \sum_{\substack{\boldsymbol{\gamma}\in\overline{\mathrm{Sh}}(\boldsymbol{\gamma}^1,\boldsymbol{\gamma}^2) \\ (\boldsymbol{\alpha},\boldsymbol{\beta})=\boldsymbol{\gamma}}} \boldsymbol{X}^{\boldsymbol{\alpha}}_{su} \boldsymbol{X}^{\boldsymbol{\beta}}_{ut} \\
&= (\mathbb{1}_{\boldsymbol{X}}\boldsymbol{H})^{(k_1,k_2)}_{st}
\end{aligned}$$

where we have used Lemma 3.1 in the second-last identity and multiplicativity of $\boldsymbol{X}$ in the last. The case of $m := |\boldsymbol{k}| \geq 3$ (which presupposes $\lfloor p\rfloor \geq 3$) is handled similarly, but has to be distinguished from the previous case since the middle terms are not the same.

$$\begin{aligned}
&\sum_{(\boldsymbol{i},\boldsymbol{j})=\boldsymbol{k}} (\mathbb{1}_{\boldsymbol{X}}\boldsymbol{H})^{\boldsymbol{i}}_{su} (\mathbb{1}_{\boldsymbol{X}}\boldsymbol{H})^{\boldsymbol{j}}_{ut} \\
&= \sum_{\substack{|\boldsymbol{\alpha}^1|,\dots,|\boldsymbol{\alpha}^m|\geq 1 \\ |\boldsymbol{\alpha}^1|+\dots+|\boldsymbol{\alpha}^m|\leq\lfloor p\rfloor \\ \boldsymbol{\alpha}\in\overline{\mathrm{Sh}}(\boldsymbol{\alpha}^1,\dots,\boldsymbol{\alpha}^m)}} \boldsymbol{H}^{k_1}_{\boldsymbol{\alpha}^1;s} \cdots \boldsymbol{H}^{k_m}_{\boldsymbol{\alpha}^m;s} \boldsymbol{X}^{\boldsymbol{\alpha}}_{su} + \Bigg(\sum_{\substack{|\boldsymbol{\alpha}^1|,\dots,|\boldsymbol{\alpha}^{m-1}|\geq 1 \\ |\boldsymbol{\alpha}^1|+\dots+|\boldsymbol{\alpha}^{m-1}|\leq\lfloor p\rfloor \\ \boldsymbol{\alpha}\in\overline{\mathrm{Sh}}(\boldsymbol{\alpha}^1,\dots,\boldsymbol{\alpha}^{m-1})}} \boldsymbol{H}^{k_1}_{\boldsymbol{\alpha}^1;s} \cdots \boldsymbol{H}^{k_{m-1}}_{\boldsymbol{\alpha}^{m-1};s} \boldsymbol{X}^{\boldsymbol{\alpha}}_{su} \Bigg) H^{k_m}_{ut}
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{l=2}^{m-2}\Bigg(\sum_{\substack{|\boldsymbol{\alpha}^1|,\dots,|\boldsymbol{\alpha}^l|\geq 1\\ |\boldsymbol{\alpha}^1|+\dots+|\boldsymbol{\alpha}^l|\leq\lfloor p\rfloor\\ \boldsymbol{\alpha}\in\overline{\mathrm{Sh}}(\boldsymbol{\alpha}^1,\dots,\boldsymbol{\alpha}^l)}}\boldsymbol{H}^{k_1}_{\boldsymbol{\alpha}^1;s}\cdots\boldsymbol{H}^{k_l}_{\boldsymbol{\alpha}^l;s}\boldsymbol{X}^{\boldsymbol{\alpha}}_{su}\Bigg)\Bigg(\sum_{\substack{|\boldsymbol{\beta}^{l+1}|,\dots,|\boldsymbol{\beta}^m|\geq 1\\ |\boldsymbol{\beta}^{l+1}|+\dots+|\boldsymbol{\beta}^m|\leq\lfloor p\rfloor\\ \boldsymbol{\beta}\in\overline{\mathrm{Sh}}(\boldsymbol{\beta}^{l+1},\dots,\boldsymbol{\beta}^m)}}\boldsymbol{H}^{k_{l+1}}_{\boldsymbol{\beta}^{l+1};u}\cdots\boldsymbol{H}^{k_m}_{\boldsymbol{\beta}^m;u}\boldsymbol{X}^{\boldsymbol{\beta}}_{ut}\Bigg)\\
&+H^{k_1}_{su}\Bigg(\sum_{\substack{|\boldsymbol{\beta}^2|,\dots,|\boldsymbol{\beta}^m|\geq 1\\ |\boldsymbol{\beta}^2|+\dots+|\boldsymbol{\beta}^m|\leq\lfloor p\rfloor\\ \boldsymbol{\beta}\in\overline{\mathrm{Sh}}(\boldsymbol{\beta}^2,\dots,\boldsymbol{\beta}^m)}}\boldsymbol{H}^{k_2}_{\boldsymbol{\beta}^2;u}\cdots\boldsymbol{H}^{k_m}_{\boldsymbol{\beta}^m;u}\boldsymbol{X}^{\boldsymbol{\beta}}_{ut}\Bigg)+\sum_{\substack{|\boldsymbol{\beta}^1|,\dots,|\boldsymbol{\beta}^m|\geq 1\\ |\boldsymbol{\beta}^1|+\dots+|\boldsymbol{\beta}^m|\leq\lfloor p\rfloor\\ \boldsymbol{\beta}\in\overline{\mathrm{Sh}}(\boldsymbol{\beta}^1,\dots,\boldsymbol{\beta}^m)}}\boldsymbol{H}^{k_1}_{\boldsymbol{\beta}^1;u}\cdots\boldsymbol{H}^{k_m}_{\boldsymbol{\beta}^m;u}\boldsymbol{X}^{\boldsymbol{\beta}}_{ut}\\
\approx&\sum_{\substack{|\boldsymbol{\alpha}^1|,\dots,|\boldsymbol{\alpha}^m|\geq 1\\ |\boldsymbol{\alpha}^1|+\dots+|\boldsymbol{\alpha}^m|\leq\lfloor p\rfloor\\ \boldsymbol{\alpha}\in\overline{\mathrm{Sh}}(\boldsymbol{\alpha}^1,\dots,\boldsymbol{\alpha}^m)}}\boldsymbol{H}^{k_1}_{\boldsymbol{\alpha}^1;s}\cdots\boldsymbol{H}^{k_m}_{\boldsymbol{\alpha}^m;s}\boldsymbol{X}^{\boldsymbol{\alpha}}_{su}\\
&+\Bigg(\sum_{\substack{|\boldsymbol{\alpha}^1|,\dots,|\boldsymbol{\alpha}^{m-1}|\geq 1\\ |\boldsymbol{\alpha}^1|+\dots+|\boldsymbol{\alpha}^{m-1}|\leq\lfloor p\rfloor\\ \boldsymbol{\alpha}\in\overline{\mathrm{Sh}}(\boldsymbol{\alpha}^1,\dots,\boldsymbol{\alpha}^{m-1})}}\boldsymbol{H}^{k_1}_{\boldsymbol{\alpha}^1;s}\cdots\boldsymbol{H}^{k_{m-1}}_{\boldsymbol{\alpha}^{m-1};s}\boldsymbol{X}^{\boldsymbol{\alpha}}_{su}\Bigg)\Bigg(\sum_{\substack{|\boldsymbol{\beta}^m|=1,\dots,\lfloor p\rfloor-1\\ |\boldsymbol{\alpha}^m|=0,\dots,\lfloor p\rfloor-1-|\boldsymbol{\beta}|}}\boldsymbol{H}^{k_m}_{(\boldsymbol{\alpha}^m,\boldsymbol{\beta}^m);s}\boldsymbol{X}^{\boldsymbol{\alpha}^m}_{su}\boldsymbol{X}^{\boldsymbol{\beta}^m}_{ut}\Bigg)\\
&+\sum_{l=2}^{m-2}\Bigg[\Bigg(\sum_{\substack{|\boldsymbol{\alpha}^1|,\dots,|\boldsymbol{\alpha}^l|\geq 1\\ |\boldsymbol{\alpha}^1|+\dots+|\boldsymbol{\alpha}^l|\leq\lfloor p\rfloor\\ \boldsymbol{\alpha}\in\overline{\mathrm{Sh}}(\boldsymbol{\alpha}^1,\dots,\boldsymbol{\alpha}^l)}}\boldsymbol{H}^{k_1}_{\boldsymbol{\alpha}^1;s}\cdots\boldsymbol{H}^{k_l}_{\boldsymbol{\alpha}^l;s}\boldsymbol{X}^{\boldsymbol{\alpha}}_{su}\Bigg)\\
&\cdot\Bigg(\sum_{\substack{|\boldsymbol{\beta}^{l+1}|,\dots,|\boldsymbol{\beta}^m|\geq 1\\ |\boldsymbol{\beta}^{l+1}|+\dots+|\boldsymbol{\beta}^m|\leq\lfloor p\rfloor\\ |\boldsymbol{\alpha}^h|=0,\dots,\lfloor p\rfloor-1-|\boldsymbol{\beta}^h|,\ h\geq l+1\\ \boldsymbol{\alpha}\in\mathrm{Sh}(\boldsymbol{\alpha}^{l+1},\dots,\boldsymbol{\alpha}^m)\\ \boldsymbol{\beta}\in\overline{\mathrm{Sh}}(\boldsymbol{\beta}^{l+1},\dots,\boldsymbol{\beta}^m)}}\boldsymbol{H}^{k_{l+1}}_{(\boldsymbol{\alpha}^{l+1},\boldsymbol{\beta}^{l+1});s}\cdots\boldsymbol{H}^{k_m}_{(\boldsymbol{\alpha}^m,\boldsymbol{\beta}^m);s}\boldsymbol{X}^{\boldsymbol{\alpha}}_{su}\boldsymbol{X}^{\boldsymbol{\beta}}_{ut}\Bigg)\Bigg]\\
&+\Bigg(\sum_{|\boldsymbol{\alpha}^1|=1}^{\lfloor p\rfloor-1}\boldsymbol{H}^{k_1}_{\boldsymbol{\alpha}^1;s}\boldsymbol{X}^{\boldsymbol{\alpha}^1}_{su}\Bigg)\Bigg(\sum_{\substack{|\boldsymbol{\beta}^2|,\dots,|\boldsymbol{\beta}^m|\geq 1\\ |\boldsymbol{\beta}^2|+\dots+|\boldsymbol{\beta}^m|\leq\lfloor p\rfloor\\ |\boldsymbol{\alpha}^h|=0,\dots,\lfloor p\rfloor-1-|\boldsymbol{\beta}^h|,\ h\geq 2\\ \boldsymbol{\alpha}\in\mathrm{Sh}(\boldsymbol{\alpha}^2,\dots,\boldsymbol{\alpha}^m)\\ \boldsymbol{\beta}\in\overline{\mathrm{Sh}}(\boldsymbol{\beta}^2,\dots,\boldsymbol{\beta}^m)}}\boldsymbol{H}^{k_2}_{(\boldsymbol{\alpha}^2,\boldsymbol{\beta}^2);s}\cdots\boldsymbol{H}^{k_m}_{(\boldsymbol{\alpha}^m,\boldsymbol{\beta}^m);s}\boldsymbol{X}^{\boldsymbol{\alpha}}_{su}\boldsymbol{X}^{\boldsymbol{\beta}}_{ut}\Bigg)+\\
+&\sum_{\substack{|\boldsymbol{\beta}^1|,\dots,|\boldsymbol{\beta}^m|\geq 1\\ |\boldsymbol{\beta}^1|+\dots+|\boldsymbol{\beta}^m|\leq\lfloor p\rfloor\\ |\boldsymbol{\alpha}^h|=0,\dots,\lfloor p\rfloor-1-|\boldsymbol{\beta}^h|\\ \boldsymbol{\alpha}\in\mathrm{Sh}(\boldsymbol{\alpha}^1,\dots,\boldsymbol{\alpha}^m)\\ \boldsymbol{\beta}\in\overline{\mathrm{Sh}}(\boldsymbol{\beta}^1,\dots,\boldsymbol{\beta}^m)}}\boldsymbol{H}^{k_1}_{(\boldsymbol{\alpha}^1,\boldsymbol{\beta}^1);s}\cdots\boldsymbol{H}^{k_m}_{(\boldsymbol{\alpha}^m,\boldsymbol{\beta}^m);s}\boldsymbol{X}^{\boldsymbol{\alpha}}_{su}\boldsymbol{X}^{\boldsymbol{\beta}}_{ut}\\
=&\sum_{\substack{|\boldsymbol{\gamma}^1|,\dots,|\boldsymbol{\gamma}^m|\geq 1\\ |\boldsymbol{\gamma}^1|+\dots+|\boldsymbol{\gamma}^m|\leq\lfloor p\rfloor}}\boldsymbol{H}^{k_1}_{\boldsymbol{\gamma}^1;s}\cdots\boldsymbol{H}^{k_m}_{\boldsymbol{\gamma}^m;s}\sum_{\substack{l=0,\dots,m\\ \boldsymbol{\alpha}^h=\boldsymbol{\gamma}^h,\ h\leq l\\ (\boldsymbol{\alpha}^h,\boldsymbol{\beta}^h)=\boldsymbol{\gamma}^h,\ |\boldsymbol{\beta}^h|\geq 1,\ h\geq l+1\\ \boldsymbol{\alpha}\in\mathrm{Sh}(\overline{\mathrm{Sh}}(\boldsymbol{\alpha}^1,\dots,\boldsymbol{\alpha}^l),\mathrm{Sh}(\boldsymbol{\alpha}^{l+1},\dots,\boldsymbol{\alpha}^2))\\ \boldsymbol{\beta}\in\overline{\mathrm{Sh}}(\boldsymbol{\beta}^{l+1},\dots,\boldsymbol{\beta}^2)}}\boldsymbol{X}^{\boldsymbol{\alpha}}_{su}\boldsymbol{X}^{\boldsymbol{\beta}}_{ut}\\
=&\sum_{\substack{|\boldsymbol{\gamma}^1|,\dots,|\boldsymbol{\gamma}^m|\geq 1\\ |\boldsymbol{\gamma}^1|+\dots+|\boldsymbol{\gamma}^m|\leq\lfloor p\rfloor}}\boldsymbol{H}^{k_1}_{\boldsymbol{\gamma}^1;s}\cdots\boldsymbol{H}^{k_m}_{\boldsymbol{\gamma}^m;s}\sum_{\substack{\boldsymbol{\gamma}\in\overline{\mathrm{Sh}}(\boldsymbol{\gamma}^1,\dots,\boldsymbol{\gamma}^m)\\ (\boldsymbol{\alpha},\boldsymbol{\beta})=\boldsymbol{\gamma}}}\boldsymbol{X}^{\boldsymbol{\alpha}}_{su}\boldsymbol{X}^{\boldsymbol{\beta}}_{ut}\\
=&\,(\mathbb{1}_{\boldsymbol{X}}\boldsymbol{H})^{\boldsymbol{k}}_{st}
\end{aligned}
$$

We proceed with the proof of geometricity. Again using (3.29) we have

$$
\begin{aligned}
&(\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^{\boldsymbol{i}}_{st}(\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^{\boldsymbol{j}}_{st} \\
\approx& \sum_{\substack{|\boldsymbol{\alpha}^1|,\ldots,|\boldsymbol{\alpha}^m|,|\boldsymbol{\beta}^1|,\ldots,|\boldsymbol{\beta}^n|\geq 1 \\ |\boldsymbol{\alpha}^1|+\ldots+|\boldsymbol{\alpha}^m|+|\boldsymbol{\beta}^1|+\ldots+|\boldsymbol{\beta}^n|\leq\lfloor p\rfloor \\ \boldsymbol{\alpha}\in\overline{\mathrm{Sh}}(\boldsymbol{\alpha}^1,\ldots,\boldsymbol{\alpha}^m) \\ \boldsymbol{\beta}\in\overline{\mathrm{Sh}}(\boldsymbol{\beta}^1,\ldots,\boldsymbol{\beta}^n)}} \boldsymbol{H}^{i_1}_{\boldsymbol{\alpha}^1;s}\cdots\boldsymbol{H}^{i_m}_{\boldsymbol{\alpha}^m;s}\boldsymbol{H}^{j_1}_{\boldsymbol{\beta}^1;s}\cdots\boldsymbol{H}^{j_n}_{\boldsymbol{\beta}^n;s}\boldsymbol{X}^{\boldsymbol{\alpha}}_{st}\boldsymbol{X}^{\boldsymbol{\beta}}_{st} \\
=& \sum_{\substack{|\boldsymbol{\alpha}^1|,\ldots,|\boldsymbol{\alpha}^m|,|\boldsymbol{\beta}^1|,\ldots,|\boldsymbol{\beta}^n|\geq 1 \\ |\boldsymbol{\alpha}^1|+\ldots+|\boldsymbol{\alpha}^m|+|\boldsymbol{\beta}^1|+\ldots+|\boldsymbol{\beta}^n|\leq\lfloor p\rfloor}} \boldsymbol{H}^{i_1}_{\boldsymbol{\alpha}^1;s}\cdots\boldsymbol{H}^{i_m}_{\boldsymbol{\alpha}^m;s}\boldsymbol{H}^{j_1}_{\boldsymbol{\beta}^1;s}\cdots\boldsymbol{H}^{j_n}_{\boldsymbol{\beta}^n;s} \\
&\cdot \sum_{\boldsymbol{\gamma}\in\mathrm{Sh}(\overline{\mathrm{Sh}}(\boldsymbol{\alpha}^1,\ldots,\boldsymbol{\alpha}^m),\overline{\mathrm{Sh}}(\boldsymbol{\beta}^1,\ldots,\boldsymbol{\beta}^n))} \boldsymbol{X}^{\boldsymbol{\gamma}}_{st} \\
=& \sum_{\boldsymbol{k}\in\mathrm{Sh}(\boldsymbol{i},\boldsymbol{j})} \sum_{\substack{|\boldsymbol{\gamma}^1|,\ldots,|\boldsymbol{\gamma}^{m+n}|\geq 1 \\ |\boldsymbol{\gamma}^1|+\ldots+|\boldsymbol{\gamma}^{m+n}|\leq\lfloor p\rfloor}} \boldsymbol{H}^{k_1}_{\boldsymbol{\gamma}^1;s}\cdots\boldsymbol{H}^{k_{m+n}}_{\boldsymbol{\gamma}^{m+n};s} \sum_{\boldsymbol{\gamma}\in\overline{\mathrm{Sh}}(\boldsymbol{\gamma}^1,\ldots,\boldsymbol{\gamma}^{m+n})} \boldsymbol{X}^{\boldsymbol{\gamma}}_{st} \\
=& \sum_{\boldsymbol{k}\in\mathrm{Sh}(\boldsymbol{i},\boldsymbol{j})} (\upharpoonleft_{\boldsymbol{X}}\boldsymbol{H})^{\boldsymbol{k}}_{st}
\end{aligned}
$$

where we have used Lemma 3.4 (with $m = 2$) in the second last identity. We may therefore apply Proposition 3.9 to conclude the proof. ■

This construction immediately yields a couple of important examples of rough path:

Example 3.16 (Pushforward of rough paths)**.** Let $\boldsymbol{X}$, F be as in Example 3.11. We denote

$$F_*\boldsymbol{X} := \upharpoonleft_{\boldsymbol{X}} F(X) \tag{3.45}$$

and call it the *pushforward* of $\boldsymbol{X}$ through F. This is a rough path with trace $F(X)$.

Example 3.17 (Rough integrals as rough paths)**.** Let $\boldsymbol{X}$, $\boldsymbol{H}$ be as in Definition 3.13. We continue to denote $\int \boldsymbol{H}\mathrm{d}\boldsymbol{X} := \upharpoonleft_{\boldsymbol{X}} \int \boldsymbol{H}\mathrm{d}\boldsymbol{X}$ (relying on context to distinguish between whether we intend the integral as a controlled or rough path).

We can use these notions to reinterpret the following well-known fact about weakly geometric rough paths. Notice how, in particular, this implies that the rough integral of an exact one-form is entirely determined by its trace. (Incidentally, arbitrary one-forms do not require the whole rough path for the integral to be defined either, just the terms $(\odot_{n-1} \otimes \mathbb{1})\boldsymbol{X}^n$ for $n = 1, \ldots, \lfloor p \rfloor$.)

Proposition 3.18 (Change of variable formula)**.** *Let $\boldsymbol{X}$ be as above, $F \in C^\infty(V, W)$, then the following identity*

$$\boldsymbol{F}(X) = F(X_0) + \int DF(X)\mathrm{d}\boldsymbol{X} \tag{3.46}$$

holds in $\mathscr{D}_{\boldsymbol{X}}(W)$. Therefore, the corresponding identity in $\mathscr{C}^p_\omega([0,T], W)$ holds as well:

$$F_*\boldsymbol{X} = F(X_0) + \int DF(X)\mathrm{d}\boldsymbol{X} \tag{3.47}$$

where in both cases the constant $F(X_0)$ is only added to the trace of the integral.

Proof. For the trace we have, by Taylor's formula and Proposition 3.8

$$
\begin{aligned}
F(X)_{st} &\approx \sum_{n=1}^{\lfloor p \rfloor} \frac{1}{n!} \langle D^n F(X_s), X_{st}^{\otimes n} \rangle \\
&= \sum_{n=1}^{\lfloor p \rfloor} \langle D^n F(X_s), \boldsymbol{X}_{st}^n \rangle \\
&\approx \int_s^t DF(X) \mathrm{d}\boldsymbol{X}
\end{aligned}
$$

and therefore [Lyo98, Theorem 3.3.1] implies $F(X)_{0t} = \int_0^t DF(X)\mathrm{d}\boldsymbol{X}$. The other claims follow trivially. ■

Remark 3.19. A similar formula would hold for more general controlled paths, i.e.

$$
\boldsymbol{H} = H_0 + \int \boldsymbol{H}' \mathrm{d}\boldsymbol{X} \tag{3.48}
$$

(together with its rough path counterpart, given by passing to the $\uparrow_{\boldsymbol{X}}$ on both sides), provided that we have a $\lfloor p \rfloor$-th Gubinelli derivative, needed to define the controlled integrand $\boldsymbol{H}'$.

We would now like to show that a path controlled by the lift of a controlled path is controlled by the original rough path in a canonical fashion.

Definition 3.20 (Change of controlling rough path)**.** Let $\boldsymbol{X} \in \mathscr{C}_\omega^p([0,T],V)$, $\boldsymbol{H} \in \mathscr{D}_{\boldsymbol{X}}(U)$ (also used to denote its rough path lift), S another vector space and $\boldsymbol{K} \in \mathscr{D}_{\boldsymbol{H}}(S)$. We then define

$$
(\boldsymbol{K} * \boldsymbol{H})_n := \sum_{m=1}^{n} \boldsymbol{K}_m \circ \boldsymbol{H}^{\boxtimes m} \circ \widetilde{\Delta}_{\sqcup\!\sqcup}^m |_{V^{\otimes n}} \tag{3.49}
$$

which for $n = 0$ reduces to $(\boldsymbol{K} * \boldsymbol{H})_0 := K$. In coordinates this means

$$
(\boldsymbol{K} * \boldsymbol{H})_\gamma^c := \sum_{\substack{m=1,\ldots,|\gamma| \\ (\gamma^1,\ldots,\gamma^m) \in \overline{\mathrm{Sh}}^{-1}(\gamma) \\ |\gamma^1|,\ldots,|\gamma^m| \geq 1}} \boldsymbol{K}_{\boldsymbol{k}}^c \boldsymbol{H}_{\gamma^1}^{k_1} \cdots \boldsymbol{H}_{\gamma^m}^{k_m}, \quad |\gamma| \geq 1 \tag{3.50}
$$

and $(\boldsymbol{K} * \boldsymbol{H})_{()}^c = K^c$.

In Proposition 3.22 below we show that this defines a controlled path. The next example features a case in which the reduced ordered shuffle coproduct can be replaced with its unordered counterpart; this however is not the general case.

Example 3.21. Let $\boldsymbol{X}, \boldsymbol{H}, \boldsymbol{K}$ be as above, with $\boldsymbol{K} := F(H)$ for $F \in C^\infty(U,S)$. Then, since $\boldsymbol{K}_m$ is symmetric we may rewrite (3.49) by using the unordered shuffle coproduct: by Corollary 3.6

$$
(F(H) * \boldsymbol{H})_n = \sum_{m \geq 1} \frac{1}{m!} D^m F(H) \circ \boldsymbol{H}^{\boxtimes m} \circ \widetilde{\Delta}_{\sqcup\!\sqcup}^m |_{V^{\otimes n}}, \quad n \geq 1 \tag{3.51}
$$

or in coordinates

$$(F(H)*\boldsymbol{H})^c_\gamma := \sum_{\substack{m=1,\ldots,|\gamma| \\ (\gamma^1,\ldots,\gamma^m)\in \mathrm{Sh}^{-1}(\gamma) \\ |\gamma^1|,\ldots,|\gamma^m|\geq 1}} \frac{1}{m!}\partial_{\boldsymbol{k}} F^c(H)\boldsymbol{H}^{k_1}_{\gamma^1}\cdots \boldsymbol{H}^{k_m}_{\gamma^m}, \quad |\gamma|\geq 1 \tag{3.52}$$

When $\boldsymbol{H}$ is also given by a smooth function this is known as the Faà di Bruno formula for the higher derivatives of a composition of functions

$$\partial_\gamma (F\circ G)^c(X) := \sum_{\substack{m=1,\ldots,|\gamma| \\ (\gamma^1,\ldots,\gamma^m)\in \mathrm{Sh}^{-1}(\gamma) \\ |\gamma^1|,\ldots,|\gamma^m|\geq 1}} \frac{1}{m!}\partial_{\boldsymbol{k}} F^c(H)\partial_{\gamma^1}G^{k_1}\cdots \partial_{\gamma^m}G^{k_m}(X) \tag{3.53}$$

We will denote the $\boldsymbol{X}$-controlled path $F(H)*\boldsymbol{H} := F_*\boldsymbol{H} \in \mathscr{D}_{\boldsymbol{X}}(S)$ and call it the *pushforward* of $\boldsymbol{H}$ through F. Note how this is distinct from $\boldsymbol{F}(H)\in\mathscr{D}_{\boldsymbol{H}}(S)$, and note how $F_*\overline{X} = \boldsymbol{F}(X)$ where $\overline{X}$ denotes the controlled path X with zero Gubinelli derivatives. An easy application of Proposition 3.18 shows the following change of variable formula for controlled paths:

$$F_*\boldsymbol{H} = \Big(F(H_0) + \int DF(H)\mathrm{d}(\uparrow_{\boldsymbol{X}}\boldsymbol{H})\Big) * \boldsymbol{H} \tag{3.54}$$

Proposition 3.22. *The map $\boldsymbol{K}*\boldsymbol{H}\in\mathcal{L}(T^{\lfloor p\rfloor-1}(V),S)$ of Definition 3.20 is an element of $\mathscr{D}_{\boldsymbol{X}}(S)$.*

Proof. Clearly $\boldsymbol{K}*\boldsymbol{H}\in\mathcal{C}^p_\omega([0,T],\mathcal{L}(T^{\lfloor p\rfloor-1}(V),U))$, so it remains to show (3.31). We preliminarily write

$$\begin{aligned}\boldsymbol{K}^c_{\boldsymbol{j};st} &\approx_{\lfloor p\rfloor-|\boldsymbol{j}|} \sum_{|\boldsymbol{i}|=1}^{\lfloor p\rfloor-|\boldsymbol{j}|-1} \boldsymbol{K}^c_{(\boldsymbol{i},\boldsymbol{j});s}\boldsymbol{H}^{\boldsymbol{i}}_{st} \\ &\approx \sum_{\substack{m=1,\ldots,\lfloor p\rfloor-|\boldsymbol{j}|-1 \\ |\boldsymbol{\alpha}^1|,\ldots,|\boldsymbol{\alpha}^m|\geq 1 \\ |\boldsymbol{\alpha}^1|+\ldots+|\boldsymbol{\alpha}^m|\leq\lfloor p\rfloor \\ \boldsymbol{\alpha}\in\overline{\mathrm{Sh}}(\boldsymbol{\alpha}^1,\ldots,\boldsymbol{\alpha}^m)}} \boldsymbol{K}^c_{(\boldsymbol{i},\boldsymbol{j});s}\boldsymbol{H}^{i_1}_{\boldsymbol{\alpha}^1;s}\cdots\boldsymbol{H}^{i_m}_{\boldsymbol{\alpha}^m;s}\boldsymbol{X}^{\boldsymbol{\alpha}}_{st}\end{aligned} \tag{3.55}$$

and

$$\boldsymbol{H}^j_{\boldsymbol{\delta};st} \approx_{\lfloor p\rfloor-|\boldsymbol{\delta}|} \sum_{|\boldsymbol{\alpha}|=1}^{\lfloor p\rfloor-|\boldsymbol{\delta}|-1} \boldsymbol{H}^j_{(\boldsymbol{\alpha},\boldsymbol{\delta});s}\boldsymbol{X}^{\boldsymbol{\alpha}}_{st} \tag{3.56}$$

Now, let $1\leq|\boldsymbol{\beta}|\leq\lfloor p\rfloor-2$:

$$\begin{aligned}&(\boldsymbol{K}*\boldsymbol{H})^c_{\boldsymbol{\beta};t} \\ =& \sum_{\substack{n=1,\ldots,|\boldsymbol{\beta}| \\ |\boldsymbol{\beta}^1|,\ldots,|\boldsymbol{\beta}^n|\geq 1 \\ (\boldsymbol{\beta}^1,\ldots,\boldsymbol{\beta}^n)\in\overline{\mathrm{Sh}}^{-1}(\boldsymbol{\beta})}} \boldsymbol{K}^c_{\boldsymbol{j};t}\boldsymbol{H}^{j_1}_{\boldsymbol{\beta}^1;t}\cdots\boldsymbol{H}^{j_n}_{\boldsymbol{\beta}^n;t}\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\substack{n=1,\dots,|\boldsymbol{\beta}| \\ |\boldsymbol{\beta}^1|,\dots,|\boldsymbol{\beta}^n|\geq 1 \\ (\boldsymbol{\beta}^1,\dots,\boldsymbol{\beta}^n)\in\overline{\mathrm{Sh}}^{-1}(\boldsymbol{\beta})}} (\boldsymbol{K}^c_{\boldsymbol{j};s} + \boldsymbol{K}^c_{\boldsymbol{j};st})(\boldsymbol{H}^{j_1}_{\boldsymbol{\beta}^1;s} + \boldsymbol{H}^{j_1}_{\boldsymbol{\beta}^1;st}) \cdots (\boldsymbol{H}^{j_n}_{\boldsymbol{\beta}^n;s} + \boldsymbol{H}^{j_n}_{\boldsymbol{\beta}^n;st}) \\
&\approx_{\lfloor p\rfloor - |\boldsymbol{\beta}|} \sum_{\substack{n=1,\dots,|\boldsymbol{\beta}| \\ |\boldsymbol{\beta}^1|,\dots,|\boldsymbol{\beta}^n|\geq 1 \\ (\boldsymbol{\beta}^1,\dots,\boldsymbol{\beta}^n)\in\overline{\mathrm{Sh}}^{-1}(\boldsymbol{\beta})}} \Big(\boldsymbol{K}^c_{\boldsymbol{j};s}\boldsymbol{H}^{j_1}_{\boldsymbol{\beta}^1;s} \cdots \boldsymbol{H}^{j_n}_{\boldsymbol{\beta}^n;s} + \sum_{\substack{\epsilon_0,\dots,\epsilon_n = 0,1 \\ (\epsilon_0,\dots,\epsilon_n)\neq(0,\dots,0)}} \xi_{\epsilon_0}\eta^1_{\epsilon_1}\cdots\eta^n_{\epsilon_n} \Big)
\end{aligned}
$$

where $\xi_0 := \boldsymbol{K}^c_{\boldsymbol{j};s}$, $\eta^l_0 := \boldsymbol{H}^{j_l}_{\boldsymbol{\beta}^l;s}$, and ξ_1, η^l_1 are given by the RHSs of (3.55) and (3.56) (with $\boldsymbol{\delta} := \boldsymbol{\beta}^l$) respectively; the $\approx_{\lfloor p\rfloor - |\boldsymbol{\beta}|}$ in the last line above holds since $|\boldsymbol{j}|, |\boldsymbol{\beta}^l| \leq |\boldsymbol{\beta}|$ and the ϵ_l's are not all zero. We can rewrite this as

$$
\begin{aligned}
&(\boldsymbol{K} * \boldsymbol{H})^c_{\boldsymbol{\beta};t} \\
&\approx_{\lfloor p\rfloor - |\boldsymbol{\beta}|} \sum_{\substack{n=1,\dots,|\boldsymbol{\beta}| \\ |\boldsymbol{\beta}^1|,\dots,|\boldsymbol{\beta}^n|\geq 1 \\ (\boldsymbol{\beta}^1,\dots,\boldsymbol{\beta}^n)\in\overline{\mathrm{Sh}}^{-1}(\boldsymbol{\beta}) \\ m=0,\dots,\lfloor p\rfloor - n-1 \\ |\boldsymbol{\alpha}^1|,\dots,|\boldsymbol{\alpha}^m|\geq 1 \\ |\boldsymbol{\alpha}^1|+\dots+|\boldsymbol{\alpha}^m|\leq\lfloor p\rfloor \\ \boldsymbol{\delta}\in\overline{\mathrm{Sh}}(\boldsymbol{\alpha}^1,\dots,\boldsymbol{\alpha}^m) \\ \boldsymbol{\alpha}^{m+l}=0,\dots,\lfloor p\rfloor - |\boldsymbol{\beta}^l|-1}} (\boldsymbol{K}^c_{(\boldsymbol{i},\boldsymbol{j});s}\boldsymbol{H}^{i_1}_{\boldsymbol{\alpha}^1;s}\cdots\boldsymbol{H}^{i_m}_{\boldsymbol{\alpha}^m;s}\boldsymbol{X}^{\boldsymbol{\delta}}_{st}) \cdot \\
&\qquad\qquad \cdot (\boldsymbol{H}^{j_1}_{(\boldsymbol{\alpha}^{m+1},\boldsymbol{\beta}^1);s}\boldsymbol{X}^{\boldsymbol{\alpha}^{m+1}}_{st}) \cdots (\boldsymbol{H}^{j_n}_{(\boldsymbol{\alpha}^{m+n},\boldsymbol{\beta}^n);s}\boldsymbol{X}^{\boldsymbol{\alpha}^{m+n}}_{st}) \\
&\approx_{\lfloor p\rfloor - |\boldsymbol{\beta}|} \sum_{\substack{q=1,\dots,\lfloor p\rfloor - 1 \\ n=1,\dots,|\boldsymbol{\beta}| \\ |\boldsymbol{\alpha}^1|,\dots,|\boldsymbol{\alpha}^{q-n}|\geq 1 \\ |\boldsymbol{\alpha}^1|+\dots+|\boldsymbol{\alpha}^q|\leq\lfloor p\rfloor - |\boldsymbol{\beta}|-1 \\ |\boldsymbol{\beta}^1|,\dots,|\boldsymbol{\beta}^n|\geq 1 \\ \boldsymbol{\alpha}\in\mathrm{Sh}(\overline{\mathrm{Sh}}(\boldsymbol{\alpha}^1,\dots,\boldsymbol{\alpha}^{q-n}),\mathrm{Sh}(\boldsymbol{\alpha}^{q-n+1},\dots,\boldsymbol{\alpha}^q)) \\ (\boldsymbol{\beta}^1,\dots,\boldsymbol{\beta}^n)\in\overline{\mathrm{Sh}}^{-1}(\boldsymbol{\beta}) \\ \boldsymbol{\gamma}^l=\boldsymbol{\alpha}^l,\ l\leq q-n \\ \boldsymbol{\gamma}^l=(\boldsymbol{\alpha}^l,\boldsymbol{\beta}^l),\ l\geq q-n+1}} \boldsymbol{K}^c_{\boldsymbol{k};s}\boldsymbol{H}^{k_1}_{\boldsymbol{\gamma}^1;s}\cdots\boldsymbol{H}^{k_q}_{\boldsymbol{\gamma}^q;s}\boldsymbol{X}^{\boldsymbol{\alpha}}_{st} \\
&= \sum_{|\boldsymbol{\alpha}|=0}^{\lfloor p\rfloor - |\boldsymbol{\beta}|-1} \boldsymbol{K}^c_{\boldsymbol{k};s}\Big(\sum_{\substack{q=1,\dots,|\boldsymbol{\alpha}|+|\boldsymbol{\beta}| \\ l=0,\dots,q \\ ((\boldsymbol{\alpha}^1,\dots,\boldsymbol{\alpha}^l),(\boldsymbol{\alpha}^{l+1},\dots,\boldsymbol{\alpha}^q))\in(\overline{\mathrm{Sh}}^{-1},\mathrm{Sh}^{-1})(\mathrm{Sh}^{-1}(\boldsymbol{\alpha})) \\ (\boldsymbol{\beta}^{l+1},\dots,\boldsymbol{\beta}^q)\in\overline{\mathrm{Sh}}^{-1}(\boldsymbol{\beta}) \\ \boldsymbol{\gamma}^h=\boldsymbol{\alpha}^h,\ |\boldsymbol{\alpha}^h|\geq 1,\ h\leq l \\ \boldsymbol{\gamma}^h=(\boldsymbol{\alpha}^h,\boldsymbol{\beta}^h),\ |\boldsymbol{\beta}^h|\geq 1,\ h\geq l+1}} \boldsymbol{H}^{k_1}_{\boldsymbol{\gamma}^1;s}\cdots\boldsymbol{H}^{k_q}_{\boldsymbol{\gamma}^q;s}\Big)\boldsymbol{X}^{\boldsymbol{\alpha}}_{st} \\
&= \sum_{|\boldsymbol{\alpha}|=0}^{\lfloor p\rfloor - |\boldsymbol{\beta}|-1} \Big(\sum_{\substack{(\boldsymbol{\gamma}^1,\dots,\boldsymbol{\gamma}^q)\in\overline{\mathrm{Sh}}^{-1}(\boldsymbol{\alpha},\boldsymbol{\beta}) \\ |\boldsymbol{\gamma}^1|,\dots,|\boldsymbol{\gamma}^q|\geq 1}} \boldsymbol{K}^c_{\boldsymbol{k};s}\boldsymbol{H}^{k_1}_{\boldsymbol{\gamma}^1;s}\cdots\boldsymbol{H}^{k_q}_{\boldsymbol{\gamma}^q;s}\Big)\boldsymbol{X}^{\boldsymbol{\alpha}}_{st} \\
&= \sum_{|\boldsymbol{\alpha}|=0}^{\lfloor p\rfloor - |\boldsymbol{\beta}|-1} (\boldsymbol{K} * \boldsymbol{H})^c_{(\boldsymbol{\alpha},\boldsymbol{\beta});s}\boldsymbol{X}^{\boldsymbol{\alpha}}_{st}
\end{aligned}
$$

as needed. The second-last identity above is given by Corollary 3.3. The proof of (3.31) for $(\boldsymbol{K} * \boldsymbol{H})_0 = K$ is a much simplified version of the proof above. ■

Another application of the change of controlling path construction is a Leibniz rule for controlled paths.

Definition 3.23 (Leibniz rule for controlled paths)**.** Let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],V)$, U_i be vector spaces for $i = 1,2,3$, $\boldsymbol{H} \in \mathscr{D}_{\boldsymbol{X}}(\mathcal{L}(U_1,U_2))$, $\boldsymbol{K} \in \mathscr{D}_{\boldsymbol{X}}(\mathcal{L}(U_2,U_3))$. Define $(\boldsymbol{K}\cdot\boldsymbol{H})_0 := K \circ H$ and for $n = 1,\ldots,\lfloor p \rfloor - 1$

$$(\boldsymbol{K}\cdot\boldsymbol{H})_n := \times \circ (\boldsymbol{K} \boxtimes \boldsymbol{H}) \circ \Delta_{\sqcup\!\sqcup} \tag{3.57}$$

where $\times \colon \mathcal{L}(U_2,U_3) \boxtimes \mathcal{L}(U_1,U_2) \to \mathcal{L}(U_1,U_3)$ is ordinary composition of linear maps.

In coordinates

$$(\boldsymbol{K}\cdot\boldsymbol{H})^{\binom{i}{j}}_{\boldsymbol{\gamma}} = \sum_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in \mathrm{Sh}^{-1}(\boldsymbol{\gamma})} \boldsymbol{K}^{\binom{i}{k}}_{\boldsymbol{\alpha}} \boldsymbol{H}^{\binom{k}{j}}_{\boldsymbol{\beta}} \tag{3.58}$$

Here we are using the notation $\binom{h}{l}$ to denote indices in spaces of linear maps, the upper index referring to the codomain and the lower to the domain; this allows us to use the Einstein convention on such superscripts. When the domain coincides with V (i.e. it is a controlled integrand) we will often place it after the bottom tuple, e.g. $\boldsymbol{H}^{\binom{k}{\delta}}_{\gamma} = \boldsymbol{H}^k_{(\gamma,\delta)}$, as done previously. The presence of the unreduced $\Delta_{\sqcup\!\sqcup}$ in the previous definition might seem strange at first, but it is easily justified as follows:

Proposition 3.24. *$\boldsymbol{K}\cdot\boldsymbol{H}$ defines an element of $\mathscr{D}_{\boldsymbol{X}}(\mathcal{L}(U_1,U_3))$ which coincides with $\times(K,H) * (\boldsymbol{K},\boldsymbol{H})$. In particular, if $\boldsymbol{H} = A(X)$, $\boldsymbol{K} = B(X)$ for smooth functions A, B then*

$$\boldsymbol{H}\cdot\boldsymbol{K} = A(\cdot)B(\cdot)(X) \tag{3.59}$$

with $A(\cdot)B(\cdot)$ denoting the function $x \mapsto A(x)B(x)$.

Proof. We apply Example 3.21 with the function $\times$ and the controlled path $(\boldsymbol{K},\boldsymbol{H}) \in \mathscr{D}_{\boldsymbol{X}}(\mathcal{L}(U_2,U_3) \oplus \mathcal{L}(U_1,U_2))$. Denoting by $\binom{h}{\cdot}$ coordinates for the second direct summand and by $\binom{\cdot}{k}$ those for the first, we have

$$\times^{\binom{i}{j}}(\kappa,\eta) = \eta^{\binom{i}{l}}\kappa^{\binom{l}{j}}$$
$$\partial_{\binom{h}{p}} \times^{\binom{i}{j}} (\kappa,\eta) = \delta^{ih}\kappa^{\binom{p}{j}}, \quad \partial_{\binom{q}{k}} \times^{\binom{i}{j}} (\kappa,\eta) = \delta^{jk}\eta^{\binom{i}{q}}$$
$$\partial_{\binom{h}{p},\binom{p}{k}} \times^{\binom{i}{j}} (\kappa,\eta) = \delta^{jk}\delta^{hi} = \partial_{\binom{p}{k},\binom{h}{p}} \times^{\binom{i}{j}} (\kappa,\eta)$$

and all other derivatives vanish. Therefore

$$(\times(K,H) * (\boldsymbol{K},\boldsymbol{H}))^{\binom{i}{j}}_{\boldsymbol{\gamma}} = \boldsymbol{K}^{\binom{i}{l}}_{\boldsymbol{\gamma}} H^{\binom{l}{j}} + K^{\binom{i}{l}}\boldsymbol{H}^{\binom{l}{j}}_{\boldsymbol{\gamma}} + \sum_{\substack{(\boldsymbol{\alpha},\boldsymbol{\beta})\in \mathrm{Sh}^{-1}(\boldsymbol{\gamma}) \\ |\boldsymbol{\alpha}|,|\boldsymbol{\beta}|\geq 1}} \boldsymbol{K}^{\binom{i}{l}}_{\boldsymbol{\alpha}} \boldsymbol{H}^{\binom{l}{k}}_{\boldsymbol{\beta}}$$

The factor $1/2!$ is not present in the sum, since each non-vanishing second derivative is counted twice, as emphasised above. This expression coincides with (3.58). The last statement holds since $A(\cdot)B(\cdot) = \times(A,B)$. ■

Next we define a notion of pullback for controlled integrands.

Definition 3.25. Let $\boldsymbol{X}$ be as above, $F \in C^\infty(V, W)$, $\boldsymbol{H} \in \mathscr{D}_{F_*\boldsymbol{X}}(\mathcal{L}(W, U))$. Let

$$F^*\boldsymbol{H} := (\boldsymbol{H} * F(X)) \cdot DF(X) \in \mathscr{D}_{\boldsymbol{X}}(\mathcal{L}(V, U)) \tag{3.60}$$

We will not need the coordinate expression of the pullback of a controlled path, although it can still be derived as done in other cases. The next proposition reassures us of the compatibility and associativity of some of the operations defined up to now.

Proposition 3.26.

1. *Let $\boldsymbol{X}, \boldsymbol{H}, \boldsymbol{H}, \boldsymbol{K}$ be as in Example 3.21, then*

$$\uparrow_{\boldsymbol{X}}(\boldsymbol{K} * \boldsymbol{H}) = \uparrow_{\boldsymbol{H}}\boldsymbol{K} \in \mathscr{C}^p_\omega([0,T], S) \tag{3.61}$$

In particular lifting commutes with pushforwards

$$F_*(\uparrow_{\boldsymbol{X}}\boldsymbol{H}) = \uparrow_{\boldsymbol{X}}(F_*\boldsymbol{H}) \tag{3.62}$$

and furthermore $F_*(G_*\boldsymbol{X}) = (F \circ G)_*\boldsymbol{X}$ for appropriately valued smooth functions F, G;

2. $(\boldsymbol{J} * \boldsymbol{K}) * \boldsymbol{H} =: \boldsymbol{J} * \boldsymbol{K} * \boldsymbol{H} := \boldsymbol{J} * (\boldsymbol{K} * \boldsymbol{H})$ *for appropriately valued controlled paths $\boldsymbol{H}, \boldsymbol{K}, \boldsymbol{J}$, and in particular $F_*G_*\boldsymbol{H} = (F \circ G)_*\boldsymbol{H}$ for appropriately valued smooth functions F, G;*

3. $(\boldsymbol{J} \cdot \boldsymbol{K}) * \boldsymbol{H} = (\boldsymbol{J} * \boldsymbol{H}) \cdot (\boldsymbol{K} * \boldsymbol{H})$ *for appropriately valued controlled paths $\boldsymbol{H}, \boldsymbol{K}, \boldsymbol{J}$; in particular, taking $\boldsymbol{J} = A(H)$, $\boldsymbol{K} = B(H)$ we have $(A(\cdot)B(\cdot))_*\boldsymbol{H} = A_*\boldsymbol{H} \cdot B_*\boldsymbol{H}$;*

4. $(\boldsymbol{J} \cdot \boldsymbol{K}) \cdot \boldsymbol{H} =: \boldsymbol{J} \cdot \boldsymbol{K} \cdot \boldsymbol{H} := \boldsymbol{J} \cdot (\boldsymbol{K} \cdot \boldsymbol{H})$ *for appropriately valued controlled paths $\boldsymbol{H}, \boldsymbol{K}, \boldsymbol{J}$;*

5. $F^*(G^*\boldsymbol{H}) = (G \circ F)^*\boldsymbol{H}$ *for appropriately valued smooth maps F, G.*

Proof. We begin with 1.:

$$\begin{aligned}
&\upharpoonleft_{\boldsymbol{X}}(\boldsymbol{K} * \boldsymbol{H})^{(c_1,\dots,c_m)}_{st} \\
&= \sum_{\substack{|\boldsymbol{\gamma}^1|,\dots,|\boldsymbol{\gamma}^m| \geq 1 \\ |\boldsymbol{\gamma}^1|+\dots+|\boldsymbol{\gamma}^m| \leq \lfloor p \rfloor \\ \boldsymbol{\gamma} \in \overline{\mathrm{Sh}}(\boldsymbol{\gamma}^1,\dots,\boldsymbol{\gamma}^m)}} (\boldsymbol{K} * \boldsymbol{H})^{c_1}_{\boldsymbol{\gamma}^1;s} \cdots (\boldsymbol{K} * \boldsymbol{H})^{c_m}_{\boldsymbol{\gamma}^m;s} \boldsymbol{X}^{\boldsymbol{\gamma}}_{st} \\
&= \sum_{\substack{|\boldsymbol{\gamma}^1|,\dots,|\boldsymbol{\gamma}^m| \geq 1 \\ |\boldsymbol{\gamma}^1|+\dots+|\boldsymbol{\gamma}^m| \leq \lfloor p \rfloor \\ \boldsymbol{\gamma} \in \overline{\mathrm{Sh}}(\boldsymbol{\gamma}^1,\dots,\boldsymbol{\gamma}^m) \\ n_l = 1,\dots,|\boldsymbol{\gamma}^l| \\ (\boldsymbol{\gamma}^{l1},\dots,\boldsymbol{\gamma}^{ln_l}) \in \overline{\mathrm{Sh}}^{-1}(\boldsymbol{\gamma}^l)}} (\boldsymbol{K}^{c_1}_{\boldsymbol{k}^1;s} \boldsymbol{H}^{k^1_1}_{\boldsymbol{\gamma}^{11};s} \cdots \boldsymbol{H}^{k^1_{n_1}}_{\boldsymbol{\gamma}^{1n_1};s}) \cdots (\boldsymbol{K}^{c_m}_{\boldsymbol{k}^m;s} \boldsymbol{H}^{k^m_1}_{\boldsymbol{\gamma}^{m1};s} \cdots \boldsymbol{H}^{k^m_{n_m}}_{\boldsymbol{\gamma}^{mn_m};s}) \boldsymbol{X}^{\boldsymbol{\gamma}}_{st} \\
&= \sum_{\substack{|\boldsymbol{k}^1|,\dots,|\boldsymbol{k}^m| \geq 1 \\ |\boldsymbol{k}^1|+\dots+|\boldsymbol{k}^m| \leq \lfloor p \rfloor \\ \boldsymbol{h} \in \overline{\mathrm{Sh}}(\boldsymbol{k}^1,\dots,\boldsymbol{k}^m)}} \boldsymbol{K}^{c_1}_{\boldsymbol{k}^1;s} \cdots \boldsymbol{K}^{c_m}_{\boldsymbol{k}^m;s} \sum_{\substack{|\boldsymbol{\delta}^1|,\dots,|\boldsymbol{\delta}^q| \geq 1 \\ |\boldsymbol{\delta}^1|+\dots+|\boldsymbol{\delta}^q| \leq \lfloor p \rfloor \\ \boldsymbol{\delta} \in \overline{\mathrm{Sh}}(\boldsymbol{\delta}^1,\dots,\boldsymbol{\delta}^q)}} \boldsymbol{H}^{h_1}_{\boldsymbol{\delta}^1;s} \cdots \boldsymbol{H}^{h_q}_{\boldsymbol{\delta}^q;s} \boldsymbol{X}^{\boldsymbol{\delta}}_{st} \\
&= (\upharpoonleft_{\boldsymbol{H}}\boldsymbol{K})^{c_1,\dots,c_m}
\end{aligned}$$

and the statement follows by [Lyo98, Theorem 3.3.1]. As for the second statement

$$F_*(\uparrow_{\boldsymbol{X}} \boldsymbol{H}) = \uparrow_{\boldsymbol{H}} F(H) = \uparrow_{\boldsymbol{X}} (F(H) * \boldsymbol{H}) = \uparrow_{\boldsymbol{X}} (F_* \boldsymbol{H})$$

and

$$F_*(G_* \boldsymbol{X}) = F_*(\uparrow_{\boldsymbol{X}} \boldsymbol{G}(X)) = \uparrow_{\boldsymbol{X}} (F_* \boldsymbol{G}(X)) = \uparrow_{\boldsymbol{X}} (F_* G_* \overline{X}) = \uparrow_{\boldsymbol{X}} ((F \circ G)_* \overline{X})$$

(where $\overline{X}$ is the $\boldsymbol{X}$-controlled path with trace X and zero Gubinelli derivatives) and the conclusion is implied by 2. below.

The proof of 2. is straightforward (with the second claim deduced from the Faà di Bruno formula (3.53)).

As for 3., we have

$$\begin{aligned}
((\boldsymbol{J} \cdot \boldsymbol{K}) * \boldsymbol{H}))^{\binom{a}{b}}_{\boldsymbol{\gamma}} &= \sum_{\substack{m=1,\ldots,|\boldsymbol{\gamma}| \\ (\boldsymbol{\gamma}^1,\ldots,\boldsymbol{\gamma}^m) \in \overline{\mathrm{Sh}}^{-1}(\boldsymbol{\gamma}) \\ |\boldsymbol{\gamma}^1|,\ldots,|\boldsymbol{\gamma}^m| \geq 1}} (\boldsymbol{J} \cdot \boldsymbol{K})^{\binom{a}{b}}_{\boldsymbol{k}} \boldsymbol{H}^{k_1}_{\boldsymbol{\gamma}^1} \cdots \boldsymbol{H}^{k_m}_{\boldsymbol{\gamma}^m} \\
&= \sum_{\substack{m=1,\ldots,|\boldsymbol{\gamma}| \\ (\boldsymbol{\gamma}^1,\ldots,\boldsymbol{\gamma}^m) \in \overline{\mathrm{Sh}}^{-1}(\boldsymbol{\gamma}) \\ |\boldsymbol{\gamma}^1|,\ldots,|\boldsymbol{\gamma}^m| \geq 1 \\ (\boldsymbol{i},\boldsymbol{j}) \in \mathrm{Sh}^{-1}(\boldsymbol{k})}} \boldsymbol{J}^{\binom{a}{c}}_{\boldsymbol{i}} \boldsymbol{K}^{\binom{c}{b}}_{\boldsymbol{j}} \boldsymbol{H}^{k_1}_{\boldsymbol{\gamma}^1} \cdots \boldsymbol{H}^{k_m}_{\boldsymbol{\gamma}^m} \\
&= \sum_{\substack{l+q=1,\ldots,|\boldsymbol{\gamma}| \\ (\boldsymbol{\alpha},\boldsymbol{\beta}) \in \mathrm{Sh}^{-1}(\boldsymbol{\gamma}) \\ (\boldsymbol{\alpha}^1,\ldots,\boldsymbol{\alpha}^l) \in \overline{\mathrm{Sh}}^{-1}(\boldsymbol{\alpha}) \\ (\boldsymbol{\beta}^1,\ldots,\boldsymbol{\beta}^q) \in \overline{\mathrm{Sh}}^{-1}(\boldsymbol{\beta}) \\ |\boldsymbol{\alpha}^1|,\ldots,|\boldsymbol{\alpha}^l|,|\boldsymbol{\beta}^1|,\ldots,|\boldsymbol{\beta}^q| \geq 1}} \boldsymbol{J}^{\binom{a}{c}}_{\boldsymbol{i}} \boldsymbol{K}^{\binom{c}{b}}_{\boldsymbol{j}} \boldsymbol{H}^{i_1}_{\boldsymbol{\alpha}^1} \cdots \boldsymbol{H}^{i_l}_{\boldsymbol{\alpha}^l} \boldsymbol{H}^{j_1}_{\boldsymbol{\beta}^1} \cdots \boldsymbol{H}^{j_q}_{\boldsymbol{\beta}^q} \\
&= (\boldsymbol{J} * \boldsymbol{H}) \cdot (\boldsymbol{K} * \boldsymbol{H})
\end{aligned}$$

where the second last identity follows from Lemma 3.4 (with $m = 2$).

As for 4., it is easy to show, using associativity of composition and of $\Delta_\sqcup$, that both sides coincide with $\times^3 \circ (\boldsymbol{J} \boxtimes \boldsymbol{K} \boxtimes \boldsymbol{H}) \circ \Delta^3_\sqcup$ where $\times^3$ denotes composition of three linear maps.

Finally, 5. is shown as follows:

$$\begin{aligned}
F^* G^* \boldsymbol{H} &= (G^* \boldsymbol{H} * \boldsymbol{F}(X)) \cdot \boldsymbol{DF}(X) \\
&= (((\boldsymbol{H} * \boldsymbol{G}(F(X))) \cdot \boldsymbol{DG}(F(X))) * \boldsymbol{F}(X)) \cdot \boldsymbol{DF}(X) \\
&= (\boldsymbol{H} * \boldsymbol{G}(F(X)) * \boldsymbol{F}(X)) \cdot (\boldsymbol{DG}(F(X)) * \boldsymbol{F}(X)) \cdot \boldsymbol{DF}(X) \\
&= ((\boldsymbol{H} * \boldsymbol{G} \circ \boldsymbol{F}(X)) \cdot \boldsymbol{D}(\boldsymbol{G} \circ \boldsymbol{F})(X) \\
&= (G \circ F)^* \boldsymbol{H}
\end{aligned}$$

Here, we have used the previous points 2., 3. and 4. in the proposition, as well as the fact that

$$\boldsymbol{G}(F(X)) * \boldsymbol{F}(X) = G_* \boldsymbol{F}(X) = G_* F_* \overline{X} = (G \circ F)_* \overline{X} = \boldsymbol{G} \circ \boldsymbol{F}(X)$$

and similarly that

$$
\begin{aligned}
(\boldsymbol{DG}(F(X)) * \boldsymbol{F}(X)) \cdot \boldsymbol{DF}(X) &= \boldsymbol{DG} \circ \boldsymbol{F}(X) \cdot \boldsymbol{DF}(X) \\
&= \boldsymbol{DG} \circ \boldsymbol{F}(\cdot)\boldsymbol{DF}(\cdot)(X) \\
&= \boldsymbol{D}(\boldsymbol{G} \circ \boldsymbol{F})(X)
\end{aligned}
$$

where we have used Proposition 3.24. ■

In the next theorem we prove the property, well-known in both ordinary and stochastic calculus, which allows to "substitute the differential". This will be especially convenient when manipulating RDEs. We have only introduced the theory necessary to handle weakly geometric rough paths, and the theorem is therefore stated in this context, but one can expect this type of result to also hold true in other settings, such as Itô calculus and branched rough paths.

Theorem 3.27 (Associativity of the rough integral). *Let* $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],V)$, $\boldsymbol{H} \in \mathscr{D}_{\boldsymbol{X}}(\mathcal{L}(V,W))$, $I := \int \boldsymbol{H}\mathrm{d}\boldsymbol{X}$, $\boldsymbol{I}$ *the canonical controlled/rough path above* I, $\boldsymbol{K} \in \mathscr{D}_{\boldsymbol{I}}(\mathcal{L}(W,U))$. *Then*

$$
\left(\int \boldsymbol{K}\mathrm{d}\boldsymbol{I}\right) * \boldsymbol{I} = \int (\boldsymbol{K} * \boldsymbol{I}) \cdot \boldsymbol{H}\mathrm{d}\boldsymbol{X} \quad \in \mathscr{D}_{\boldsymbol{X}}(U) \tag{3.63}
$$

and therefore

$$
\int \boldsymbol{K}\mathrm{d}\boldsymbol{I} = \int (\boldsymbol{K} * \boldsymbol{I}) \cdot \boldsymbol{H}\mathrm{d}\boldsymbol{X} \quad \in \mathscr{C}^p_\omega([0,T],U) \tag{3.64}
$$

Proof. In this proof we will denote, for a tuple $\boldsymbol{\gamma}$, $\gamma^{\cdot}$ its last entry and $\boldsymbol{\gamma}^-$ the tuple obtained by removing $\gamma^{\cdot}$, so $\boldsymbol{\gamma} = (\boldsymbol{\gamma}^-, \gamma^{\cdot})$. Moreover, we will interchangeably use the two indexing notations for controlled integrands, e.g. $\boldsymbol{H}^k_{\boldsymbol{\gamma}} = \boldsymbol{H}^{\binom{k}{\gamma^{\cdot}}}_{\boldsymbol{\gamma}^-}$. For $|\boldsymbol{\gamma}| \geq 1$ we then have

$$
\begin{aligned}
\left(\int (\boldsymbol{K} * \boldsymbol{I}) \cdot \boldsymbol{H}\mathrm{d}\boldsymbol{X}\right)^c_{\boldsymbol{\gamma}} &= ((\boldsymbol{K} * \boldsymbol{I}) \cdot \boldsymbol{H})^c_{\boldsymbol{\gamma}} \\
&= \sum_{(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \mathrm{Sh}^{-1}(\boldsymbol{\gamma}^-)} (\boldsymbol{K} * \boldsymbol{I})^{\binom{c}{h}}_{\boldsymbol{\alpha}} \boldsymbol{H}^{\binom{h}{\gamma^{\cdot}}}_{\boldsymbol{\beta}} \\
&= \sum_{\substack{(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \mathrm{Sh}^{-1}(\boldsymbol{\gamma}^-) \\ (\boldsymbol{\alpha}^1,\ldots,\boldsymbol{\alpha}^n) \in \overline{\mathrm{Sh}}^{-1}(\boldsymbol{\alpha}) \\ |\boldsymbol{\alpha}^1|,\ldots,|\boldsymbol{\alpha}^n| \geq 1}} \boldsymbol{K}^{\binom{c}{h}}_{\boldsymbol{k}} \boldsymbol{H}^{k_1}_{\boldsymbol{\alpha}^1} \cdots \boldsymbol{H}^{k_n}_{\boldsymbol{\alpha}^n} \boldsymbol{H}^{\binom{h}{\gamma^{\cdot}}}_{\boldsymbol{\beta}} \\
&= \sum_{\substack{(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \mathrm{Sh}^{-1}(\boldsymbol{\gamma}^-) \\ (\boldsymbol{\alpha}^1,\ldots,\boldsymbol{\alpha}^n) \in \overline{\mathrm{Sh}}^{-1}(\boldsymbol{\alpha}) \\ |\boldsymbol{\alpha}^1|,\ldots,|\boldsymbol{\alpha}^n| \geq 1}} \boldsymbol{K}^c_{(\boldsymbol{k},h)} \boldsymbol{H}^{k_1}_{\boldsymbol{\alpha}^1} \cdots \boldsymbol{H}^{k_n}_{\boldsymbol{\alpha}^n} \boldsymbol{H}^h_{(\boldsymbol{\beta},\gamma^{\cdot})} \\
&= \sum_{\substack{m=1,\ldots,|\boldsymbol{\gamma}| \\ (\boldsymbol{\gamma}^1,\ldots,\boldsymbol{\gamma}^m) \in \overline{\mathrm{Sh}}^{-1}(\boldsymbol{\gamma}) \\ |\boldsymbol{\gamma}^1|,\ldots,|\boldsymbol{\gamma}^m| \geq 1}} \boldsymbol{K}^c_{\boldsymbol{k}} \boldsymbol{H}^{k_1}_{\boldsymbol{\gamma}^1} \cdots \boldsymbol{H}^{k_m}_{\boldsymbol{\gamma}^m}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{\substack{m=1,\ldots,|\gamma| \\ (\gamma^1,\ldots,\gamma^m)\in\overline{\mathrm{Sh}}^{-1}(\gamma) \\ |\gamma^1|,\ldots,|\gamma^m|\geq 1}} \left(\int \boldsymbol{K}\mathrm{d}\boldsymbol{I}\right)^c_{\boldsymbol{k}} \boldsymbol{I}^{k_1}_{\gamma^1}\cdots\boldsymbol{I}^{k_m}_{\gamma^m} \\
&= \left(\left(\int \boldsymbol{K}\mathrm{d}\boldsymbol{I}\right) * \boldsymbol{I}\right)^c_\gamma
\end{aligned}
$$

At the trace level, we have, through a similar argument

$$
\begin{aligned}
\left(\int (\boldsymbol{K} * \boldsymbol{I}) \cdot \boldsymbol{H}\mathrm{d}\boldsymbol{X}\right)^c_{\gamma;st} &\approx \sum_{|\gamma|=1}^{\lfloor p\rfloor} ((\boldsymbol{K} * \boldsymbol{I}) \cdot \boldsymbol{H})^c_{\gamma;s}\boldsymbol{X}^\gamma_{st} \\
&= \sum_{\substack{|\gamma|=1,\ldots\lfloor p\rfloor \\ m=1,\ldots,|\gamma| \\ (\gamma^1,\ldots,\gamma^m)\in\overline{\mathrm{Sh}}^{-1}(\gamma) \\ |\gamma^1|,\ldots,|\gamma^m|\geq 1}} \boldsymbol{K}^c_{\boldsymbol{k};s}\boldsymbol{H}^{k_1}_{\gamma^1;s}\cdots\boldsymbol{H}^{k_m}_{\gamma^m;s}\boldsymbol{X}^\gamma_{st} \\
&\approx \int \boldsymbol{K}^c\mathrm{d}\boldsymbol{I}
\end{aligned}
$$

As for the statement at the level of rough paths, we have

$$
\begin{aligned}
\int \boldsymbol{K}\mathrm{d}\boldsymbol{I} &= \Big\uparrow_{\boldsymbol{I}}\int \boldsymbol{K}\mathrm{d}\boldsymbol{I} \\
&= \Big\uparrow_{\boldsymbol{X}}\left[\left(\int \boldsymbol{K}\mathrm{d}\boldsymbol{I}\right) * \boldsymbol{I}\right] \\
&= \Big\uparrow_{\boldsymbol{X}}\left[\int (\boldsymbol{K} * \boldsymbol{I}) \cdot \boldsymbol{H}\mathrm{d}\boldsymbol{X}\right] \\
&= \int (\boldsymbol{K} * \boldsymbol{I}) \cdot \boldsymbol{H}\mathrm{d}\boldsymbol{X}
\end{aligned}
$$

where we have used 1. in Proposition 3.26 and the previous statement. ■

The next result, for which geometricity is essential, tells us that F_* and F^* behave as adjoint operators under the rough integral pairing. Its proof is an immediate consequence of Proposition 3.18 and Theorem 3.27.

Theorem 3.28 (Pushforward-pullback adjunction)**.** *Let $\boldsymbol{X}$, $\boldsymbol{H}$, F be as above, then the identity of controlled paths*

$$
\left(\int \boldsymbol{H}\mathrm{d}F_*\boldsymbol{X}\right) * F(X) = \int F^*\boldsymbol{H}\mathrm{d}\boldsymbol{X} \tag{3.65}
$$

holds, and therefore so does the corresponding one of rough paths

$$
\int \boldsymbol{H}\mathrm{d}F_*\boldsymbol{X} = \int F^*\boldsymbol{H}\mathrm{d}\boldsymbol{X} \tag{3.66}
$$

Next we move on to the topic of rough differential equations (RDEs). We will introduce two equivalent notions of solution to an RDE. Given a field of linear maps $F \in C^\infty(W, \mathcal{L}(V, W))$ and a smooth map

$g \in C^\infty(W, U)$ we define, for $y \in W$

$$Fg(y) := Dg(y) \circ F(y) \in \mathcal{L}(V, U) \tag{3.67}$$

and inductively

$$F^n g(y) := F(\eta \mapsto F^{n-1} g(\eta))|_{\eta = y} \in \mathcal{L}(V, \mathcal{L}(V^{\otimes n-1}, U)) = \mathcal{L}(V^{\otimes n}, U) \tag{3.68}$$

In coordinates we denote

$$F_\gamma g^c(y) := Fg(y)^c_\gamma = \partial_k g(y) F^k_\gamma(y) \;\Rightarrow\; F^n g(y)^c_{(\gamma_1, \ldots, \gamma_n)} = F_{\gamma_1} \cdots F_{\gamma_n} g^c(y)$$

We will also use the compact notation $F_{\boldsymbol{\gamma}} g^c(y)$ for the latter.

Remark 3.29. Note that $F_{\gamma_1} \cdots F_{\gamma_n} g^c(y)$ can be read right to left as well as left to right, i.e. it is equal to $F_{\gamma_1} \cdots F_{\gamma_{n-1}}(F_{\gamma_n} g^c)(y)$ for $n \geq 2$. This can be seen by induction on n (with the quantifier $\forall g$ inside the inductive hypothesis). For $n = 2$ the statement is tautological. For the inductive step we have

$$\begin{aligned} F_{\gamma_1} \cdots F_{\gamma_{n+1}} g^c(y) &= F_{\gamma_1}(F_{\gamma_2} \cdots F_{\gamma_{n+1}} g^c)(y) \\ &= F_{\gamma_1}(F_{\gamma_2} \cdots F_{\gamma_n}(F_{\gamma_{n+1}} g^c))(y) \\ &= F_{\gamma_1} \cdots F_{\gamma_n}(F_{\gamma_{n+1}} g^c)(y) \end{aligned}$$

where in the second identity we have used the inductive hypothesis.

Definition 3.30 (Davie solution to an RDE)**.** Let $F \in C^\infty(W, \mathcal{L}(V, W))$. A *solution* to the RDE

$$\mathrm{d}Y = F(Y)\mathrm{d}\boldsymbol{X}, \quad Y_0 = y_0 \tag{3.69}$$

is a path $Y \in C([0, T], W)$ starting at y_0 with the property that for all $g \in C^\infty(W)$.

$$g(Y)_{st} \approx \sum_{n=1}^{\lfloor p \rfloor} \langle F^n g(Y_s), \boldsymbol{X}^n_{st} \rangle \tag{3.70}$$

Proposition 3.31 (Gubinelli solution to an RDE)**.** *Let Y be a solution to* (5.26)*. Then*

$$\boldsymbol{Y} := (Y, F(Y) = F\mathbb{1}(Y), \ldots, F^{\lfloor p \rfloor - 1}\mathbb{1}(Y)) \in \mathscr{D}_{\boldsymbol{X}}(W) \tag{3.71}$$

and moreover

$$\boldsymbol{Y} = y_0 + \int F_* \boldsymbol{Y} \mathrm{d}\boldsymbol{X} \quad \in \mathscr{D}_{\boldsymbol{X}}(W) \tag{3.72}$$

Conversely, if an $\boldsymbol{X}$-controlled controlled path satisfying the above identity, its trace satisfies Definition 3.30.

In order to prove this proposition we will make use of the following

Lemma 3.32. *The $\boldsymbol{X}$-controlled path $\boldsymbol{Y}$ has the form* (3.71) *if and only if $\boldsymbol{Y}_n = (F_* \boldsymbol{Y})_{n-1}$ for $n \geq 1$. Moreover, in this case*

$$((Fg)_* \boldsymbol{Y})_{n-1} = F^n g(Y), \quad n \geq 1$$

with g as above.

Proof. If we prove

$$(Fg)_*(Y, F(Y), \dots, F^{\lfloor p \rfloor -1}\mathbb{1}(Y))_{n-1} = F^n g(Y)$$

we will have shown both the second statement and the "only if" part of the first (just choose $g = \mathbb{1}$). In order to show this, we first establish the identity

$$F_{\boldsymbol{\gamma}} g^c(y) = \sum_{\substack{(\boldsymbol{\gamma}^1,\dots,\boldsymbol{\gamma}^m)\in \overline{\mathrm{Sh}}^{-1}(\boldsymbol{\gamma}) \\ |\boldsymbol{\gamma}^1|,\dots,|\boldsymbol{\gamma}^m|\geq 1}} \partial_{\boldsymbol{k}} g^c F_{\boldsymbol{\gamma}^1}\mathbb{1}^{k_1}\cdots F_{\boldsymbol{\gamma}^m}\mathbb{1}^{k_m}(y) \tag{3.73}$$

Note that this is not a closed form formula for $F^n g$, since iterated compositions of the vector fields $F_{\boldsymbol{\gamma}}$ also appear on the RHS, but it will be useful for us nonetheless. We proceed by induction on $|\boldsymbol{\gamma}|$. For $|\boldsymbol{\gamma}| = 1$ there is nothing to show. For the inductive step, using Remark 3.29 we have

$$\begin{aligned}
&F_{(\gamma_1,\dots,\gamma_{n+1})} g^c(y) \\
&= F_{\gamma_1}(F_{\gamma_2,\dots,\gamma_{n+1}} g^c)(y) \\
&= F_{\gamma_1} \sum_{\substack{(\boldsymbol{\gamma}^1,\dots,\boldsymbol{\gamma}^m)\in \overline{\mathrm{Sh}}^{-1}(\gamma_2,\dots,\gamma_{n+1}) \\ |\boldsymbol{\gamma}^1|,\dots,|\boldsymbol{\gamma}^m|\geq 1}} \partial_{\boldsymbol{k}} g^c F_{\boldsymbol{\gamma}^1}\mathbb{1}^{k_1}\cdots F_{\boldsymbol{\gamma}^m}\mathbb{1}^{k_m}(y) \\
&= \sum_{\substack{(\boldsymbol{\gamma}^1,\dots,\boldsymbol{\gamma}^m)\in \overline{\mathrm{Sh}}^{-1}(\gamma_2,\dots,\gamma_{n+1}) \\ |\boldsymbol{\gamma}^1|,\dots,|\boldsymbol{\gamma}^m|\geq 1}} \Big(\partial_{h,\boldsymbol{k}} g^c F_{\boldsymbol{\gamma}^1}\mathbb{1}^{k_1}\cdots F_{\boldsymbol{\gamma}^m}\mathbb{1}^{k_m} \\
&\qquad\qquad + \sum_{l=1}^m \partial_{\boldsymbol{k}} g^c F_{\boldsymbol{\gamma}^1}\mathbb{1}^{k_1}\cdots \partial_h(F_{\boldsymbol{\gamma}^l}\mathbb{1}^{k_l})\cdots F_{\boldsymbol{\gamma}^m}\mathbb{1}^{k_m}\Big) F^h_{\gamma_1}(y) \\
&= \sum_{\substack{(\boldsymbol{\gamma}^1,\dots,\boldsymbol{\gamma}^m)\in \overline{\mathrm{Sh}}^{-1}(\gamma_2,\dots,\gamma_{n+1}) \\ |\boldsymbol{\gamma}^1|,\dots,|\boldsymbol{\gamma}^m|\geq 1}} \Big(\partial_{h,\boldsymbol{k}} g^c F^h_{\gamma_1} F_{\boldsymbol{\gamma}^1}\mathbb{1}^{k_1}\cdots F_{\boldsymbol{\gamma}^m}\mathbb{1}^{k_m} \\
&\qquad\qquad + \sum_{l=1}^m \partial_{\boldsymbol{k}} g^c F_{\boldsymbol{\gamma}^1}\mathbb{1}^{k_1}\cdots F_{\gamma_1,\boldsymbol{\gamma}^l}\mathbb{1}^{k_l}\cdots F_{\boldsymbol{\gamma}^m}\mathbb{1}^{k_m}\Big)(y) \\
&= \sum_{\substack{(\boldsymbol{\gamma}^1,\dots,\boldsymbol{\gamma}^m)\in \overline{\mathrm{Sh}}^{-1}(\gamma_1,\dots,\gamma_{n+1}) \\ |\boldsymbol{\gamma}^1|,\dots,|\boldsymbol{\gamma}^m|\geq 1}} \partial_{\boldsymbol{k}} g^c F_{\boldsymbol{\gamma}^1}\mathbb{1}^{k_1}\cdots F_{\boldsymbol{\gamma}^m}\mathbb{1}^{k_m}(y)
\end{aligned}$$

Now, for $n = |\boldsymbol{\gamma}| \geq 1$ we have

$$\begin{aligned}
((Fg)_*\boldsymbol{Y})^c_{\boldsymbol{\gamma}} &= \sum_{(\boldsymbol{\gamma}^1,\dots,\boldsymbol{\gamma}^m)\in \overline{\mathrm{Sh}}^{-1}(\gamma_1,\dots,\gamma_{n-1})} \partial_{\boldsymbol{k}}(F_{\gamma_n} g^c) F_{\boldsymbol{\gamma}^1}\mathbb{1}^{k_1}\cdots F_{\boldsymbol{\gamma}^m}\mathbb{1}^{k_m}(Y) \\
&= F_{(\gamma_1,\dots,\gamma_{n-1})}(F_{\gamma_n} g^c)(y) \\
&= F_{\boldsymbol{\gamma}} g^c(y)
\end{aligned}$$

where we have used Remark 3.29 and (3.73).

We now show the "if" implication of the first statement. Namely, we need to show that if $\boldsymbol{Y} \in \mathscr{D}_{\boldsymbol{X}}(W)$ has the property that $\boldsymbol{Y}_n = (F_*\boldsymbol{Y})_{n-1}$ for $n = 1, \ldots, \lfloor p \rfloor$ then $\boldsymbol{Y}_n = F^n\mathbb{1}(Y)$. We show this by induction on n. For $n = 1$ the assertion is obvious. For the inductive step we have

$$\begin{aligned}
\boldsymbol{Y}^h_\gamma &= \sum_{\substack{(\gamma^1,\ldots,\gamma^m)\in\overline{\mathrm{Sh}}^{-1}(\gamma)\\ |\gamma^1|,\ldots,|\gamma^m|\geq 1}} \partial_{\boldsymbol{k}} F^h(Y)\boldsymbol{Y}^{k_1}_{\gamma^1}\cdots\boldsymbol{Y}^{k_1}_{\gamma^1} \\
&= \sum_{\substack{(\gamma^1,\ldots,\gamma^m)\in\overline{\mathrm{Sh}}^{-1}(\gamma)\\ |\gamma^1|,\ldots,|\gamma^m|\geq 1}} \partial_{\boldsymbol{k}} F^h F_{\gamma^1}\mathbb{1}^{k_1}\cdots F_{\gamma^1}\mathbb{1}^{k_1}(Y) \\
&= F_\gamma \mathbb{1}^h(Y)
\end{aligned}$$

where we have used (3.73) and the inductive hypothesis. ■

Proof of Proposition 3.31. Let Y be a Davie solution to the RDE. Taking g in (4.30) to be $\mathbb{1}, F, \ldots, F^{\lfloor p \rfloor - 2}\mathbb{1}$ proves that $\boldsymbol{Y}$ defined in (3.71) is indeed an element of $\mathscr{D}_{\boldsymbol{X}}(W)$. By Lemma 3.32 we then have $F_*\boldsymbol{Y} = (F(Y), \ldots, F^{\lfloor p \rfloor}\mathbb{1}(Y))$ and by Definition 3.13

$$\int_s^t F_*\boldsymbol{Y}\mathrm{d}\boldsymbol{X} \approx \langle (F_*\boldsymbol{Y})_s, \boldsymbol{X}^{\geq 1}_{st}\rangle \approx Y_{st}$$

again by the Davie definition. Since both the left and RHSs are increments of paths, we conclude by [Lyo98, Theorem 3.3.1] that identity must hold. Therefore, since $Y_0 = y_0$, (3.72) holds at the trace level, and for $n \geq 1$

$$\left(y_0 + \int F_*\boldsymbol{Y}\mathrm{d}\boldsymbol{X}\right)_n = (F_*\boldsymbol{Y})_{n-1} = \boldsymbol{Y}_n$$

Conversely, assume that there exists some $\boldsymbol{Y} \in \mathscr{D}_{\boldsymbol{X}}(W)$ s.t. (3.72) holds: this implies that for $m \geq 0$

$$(F_*\boldsymbol{Y})_m = \left(y_0 + \int F_*\boldsymbol{Y}\mathrm{d}\boldsymbol{X}\right)_{m+1} = \boldsymbol{Y}_{m+1}$$

and therefore by Lemma 3.32 $\boldsymbol{Y}$ must have the form (3.71). Finally, $Y_0 = y_0$ and for $g \in C^\infty(W)$ and $\boldsymbol{Y} := \uparrow_{\boldsymbol{X}} \boldsymbol{Y}$

$$\begin{aligned}
g(Y)_{st} &= \int_s^t Dg(Y)\mathrm{d}\boldsymbol{Y} \\
&= \int_s^t (Dg(Y) * \boldsymbol{Y}) \cdot (F(Y) * \boldsymbol{Y})\mathrm{d}\boldsymbol{X} \\
&= \int_s^t (Dg(Y) \cdot F(Y)) * \boldsymbol{Y}\mathrm{d}\boldsymbol{X} \\
&= \int_s^t (Dg(\cdot)F(\cdot))(Y) * \boldsymbol{Y}\mathrm{d}\boldsymbol{X} \\
&= \int_s^t (Fg)_*\boldsymbol{Y}\mathrm{d}\boldsymbol{X}
\end{aligned}$$

$$\approx \sum_{n=1}^{\lfloor p \rfloor} \langle F^n g(Y_s), \boldsymbol{X}^n_{st} \rangle$$

where we have used Theorem 3.27, Proposition 3.26 and Lemma 3.32. This concludes the proof. ■

The above proposition tells us that once we have the solution in the sense of Definition 3.30 we can obtain an $\boldsymbol{X}$-controlled path, and thus by Definition 3.14 a rough path. If we want to emphasise the existence of one of these superstructures we will write

$$\mathrm{d}\boldsymbol{Y} = F(Y)\mathrm{d}\boldsymbol{X} \quad \text{and,} \quad Y_0 = y_0 \tag{3.74}$$

i.e. $\boldsymbol{Y} :=\uparrow_{\boldsymbol{X}} \boldsymbol{Y}$, again abusing notation by denoting $\boldsymbol{Y}$ both the controlled and the rough path. Notice that the initial condition only involves the trace.

The next result will be instrumental in defining RDEs on manifolds in a coordinate-invariant manner.

Theorem 3.33 (Change of variable formula for RDE solutions)**.** *Let $\boldsymbol{X}, F, \boldsymbol{Y}$ be as above, $g \in C^\infty(W, U)$. Then $(\boldsymbol{Y}, g_*\boldsymbol{Y})$ jointly solve the RDE*

$$\mathrm{d}\begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{Z} \end{pmatrix} = \begin{pmatrix} F(Y) \\ Dg(Y)F(Y) \end{pmatrix} \mathrm{d}\boldsymbol{X} \tag{3.75}$$

In particular, if g is invertible, Defining $C^\infty(U, \mathcal{L}(V, U)) \ni F_g(z) := Dg(g^{-1}(Z))F(g^{-1}(Z))$, $g_*\boldsymbol{Y}$ coincides with the rough path solution to

$$\mathrm{d}\boldsymbol{Z} = F_g(Z)\mathrm{d}\boldsymbol{X} \tag{3.76}$$

Proof. Using Proposition 3.18, Theorem 3.27 and Proposition 3.26 we have

$$\begin{aligned} \mathrm{d}(g_*\boldsymbol{Y}) &= y_0 + \int Dg(Y)\mathrm{d}\boldsymbol{Y} \\ &= y_0 + \int Dg_*\boldsymbol{Y} \cdot F_*\boldsymbol{Y}\mathrm{d}\boldsymbol{X} \\ &= y_0 + \int (Dg(\cdot)F(\cdot))_*\boldsymbol{Y}\mathrm{d}\boldsymbol{X} \end{aligned}$$

This proves the first claim; as for the second, we continue

$$\begin{aligned} \mathrm{d}(g_*\boldsymbol{Y}) &= \int (Dg(\cdot)F(\cdot))_*\boldsymbol{Y}\mathrm{d}\boldsymbol{X} \\ &= \int (Dg(g^{-1}(\cdot))F(g^{-1}(\cdot)))_* g_*\boldsymbol{Y}\mathrm{d}\boldsymbol{X} \end{aligned}$$

where we have again used Proposition 3.26. This concludes the proof. ■

3.3 Geometric rough paths on manifolds

In this short section we comment very briefly on how the theory of the previous can be used to define geometric $[1, \infty) \ni p$-rough paths on manifolds, and the corresponding notions of rough integral and RDE. We omit all details except for those that pertain to the extrinsic formulation, since the local theory proceeds in complete analogy with that for geometric $[1,3) \ni p$-rough paths, a special case treated in Chapter 2.

Geometric $[1,\infty) \ni p$-rough paths on manifolds are defined in charts as in Definition 2.32, using pushforwards to enforce compatibility, with associativity of pushforwards Proposition 3.26 guaranteeing independence of the atlas. Controlled integrands are also defined in charts, using pullbacks, as in Definition 2.34. The rough integral is defined as in Definition 2.36, but without the extra term involving the connection and bracket term. The independence of the integral of the coordinate system can be shown using the pushforward-pullback adjunction Theorem 3.28. RDEs driven by an M-valued rough path with solution in a second manifold N are also defined as in (2.129), again, without a dependence on covariant derivatives on M and N, with the solution well-defined as a rough path thanks to Theorem 3.33. As usual, we have local existence and uniqueness since the coefficients (and atlas) are C^∞. The formulae for parallel transport and Cartan development are classical, and the latter can be defined by the fundamental horizontal vector fields (2.220). The well-known equations in local coordinates read

$$\mathrm{d}A^\gamma = -\Gamma^\gamma_{\alpha\beta}(X)A^\beta \mathrm{d}\boldsymbol{X}^\alpha \tag{3.77}$$

for parallel transport of the vector A above X and

$$\begin{cases} \mathrm{d}\boldsymbol{Y}^k = A^k_\gamma \mathrm{d}\boldsymbol{Z}^\gamma \\ \mathrm{d}A^k_\gamma = -\Gamma^k_{ij}(Y)A^i_\alpha A^j_\gamma \mathrm{d}\boldsymbol{Z}^\alpha \end{cases} \tag{3.78}$$

for development of the T_oM-valued rough path $\boldsymbol{Z}$.

In [CDL15] the topic of manifold-valued theory of rough paths, rough integration (specifically of one-forms) and RDEs was treated from the extrinsic point of view. Here $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$ is defined in [CDL15, Definition 3.17] to be *constrained* to a smoothly embedded manifold M if its trace is M-valued and for all one-forms $F \in \Gamma\mathcal{L}(T\mathbb{R}^d, W)$ (Γ denoting the space of sections)

$$\forall x \in M \; F(x)|_{T_xM} = 0 \;\Rightarrow\; \int F(X)\mathrm{d}\boldsymbol{X} = 0 \in \mathscr{C}^p_\omega([0,T],W) \tag{3.79}$$

In [CDL15, Corollary 3.32, Proposition 3.35] this is shown to be equivalent to the trace X being M-valued and $(I \otimes Q(X_s))\boldsymbol{X}^2_{st} \approx 0$, or equivalently to $(P(X_s) \otimes P(X_s))\boldsymbol{X}^2_{st} \approx \boldsymbol{X}^2_{st}$, where for $x \in M$ $P(x)$ is the orthogonal projection $T_x\mathbb{R}^d \twoheadrightarrow T_xM$ and $Q := \mathbb{1} - P$. Moreover, one may replace $\boldsymbol{X}^2_{st}$ with its antisymmetric part $(\wedge \boldsymbol{X}^2)_{st}$ in these identities (because $\odot\boldsymbol{X}^2$ is already fixed by the trace).

This approach carries over to the case of higher p considered here. If $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^d)$ we may say that it is *constrained* to the smoothly embedded manifold M if $\pi_*\boldsymbol{X} = \boldsymbol{X}$, where π is the Riemannian projection of a tubular neighbourhood U of M onto M (i.e. it maps a point in U to the unique point on M closest to it — this is well defined and smooth on a thin enough tubular neighbourhood). This extends the definition of [CDL15] since $D_x\pi = P(x)$. In order to generalise the equivalent condition $(I \otimes Q(X_s))(\wedge\boldsymbol{X})_{st}$ we can take the log of our original condition, i.e. $\log \pi_*\boldsymbol{X} = \log \boldsymbol{X}$: this has the advantage of eliminating all the redundancies of the former (as explained in [LS06, p.767]), and its precise coordinate expression can be derived by using [FV10b, Definition 7.20], but at higher orders cannot be described in terms of antisymmetric tensors. The Chen-Strichartz formula [Bau04, Theorem 1.1] expresses $\log \boldsymbol{X}$ as a Lie polynomial; the task of expressing it in a basis of the Lie algebra is more complex still [Rei17].

Care must be taken when defining the rough integral of a controlled integrand, since it is no longer the case that for an $\boldsymbol{X}$-controlled integrand $\boldsymbol{H}$, $\int \boldsymbol{H} \mathrm{d}\boldsymbol{X}$ (defined in the ordinary sense, where $\boldsymbol{X}$ is considered an element of $\mathscr{C}^p_\omega([0,T], \mathbb{R}^d)$) does not always only depend on the trace H of $\boldsymbol{X}$ restricted to TM, although this is indeed the case when $\boldsymbol{H}$ is given by a one-form: this is because if $F \in \Gamma\mathcal{L}(T\mathbb{R}^d, W)$ (Γ denoting the space of sections) vanishes on TM (i.e. $F(x)P(x) = 0$ for $x \in M$) by Theorem 3.28 we have

$$\begin{aligned} &\int F(X)\mathrm{d}\boldsymbol{X} = \int F(X)\mathrm{d}\pi_*\boldsymbol{X} = \int \pi^*\boldsymbol{F}(X)\mathrm{d}\boldsymbol{X} = 0 \\ \text{since} \quad &\pi^*\boldsymbol{F}(X) = (\boldsymbol{F}(X) * \boldsymbol{\pi}(X)) \cdot \boldsymbol{D\pi}(X) = \boldsymbol{F}(X) \cdot \boldsymbol{P}(X) = \boldsymbol{F}(\cdot)\boldsymbol{P}(\cdot)(X) = 0 \end{aligned} \tag{3.80}$$

where we have used Proposition 3.24. This then implies that if $F, G \in \Gamma\mathcal{L}(T\mathbb{R}^d, W)$ restrict to the same element of $\Gamma\mathcal{L}(TM, W)$ then $\int F(X)\mathrm{d}\boldsymbol{X} = \int G(X)\mathrm{d}\boldsymbol{X}$. Similarly, we have that if $\boldsymbol{H}, \boldsymbol{K} \in \mathscr{D}_{\boldsymbol{X}}(\mathcal{L}(TM, W))$ are such that $\pi^*\boldsymbol{H} = \pi^*\boldsymbol{K}$ then $\int \boldsymbol{H}\mathrm{d}\boldsymbol{X} = \int \boldsymbol{K}\mathrm{d}\boldsymbol{X}$, but this involves conditions on all levels of $\boldsymbol{H}$, not just the trace.

Finally, the original definition of constrained rough path given by integration (3.79) also carries over to higher p. The fact that this is implied by $\pi_*\boldsymbol{X} = \boldsymbol{X}$ was shown in (3.80). For the converse, we rewrite the identity as $(\mathbb{1} - \pi)_*\boldsymbol{X} = 0$: the trace level is implied by the fact that X is M-valued, and at orders ≥ 1 the identity $\int Q(X)\mathrm{d}\boldsymbol{X} = (\mathbb{1} - \pi)_*\boldsymbol{X}$ is straightforward to check.

Conclusions and further directions

In this chapter we have provided a self-contained treatment of all the basic structural aspects of finite-dimensional geometric rough paths and their controlled paths. Our combinatorial approach combines Lyons' original theory with Gubinelli's subsequent approach, and is applied to show how rough path theory naturally extends to intrinsic and extrinsic manifolds.

Although we believe that spelling out all the algebraic relations between rough and controlled paths is already of interesting in the finite-dimensional case, it is really in infinite dimensions that this becomes necessary, due to the impracticability of the approach via smooth approximations. For this reason, further value could be added to the material of this chapter if it were to be extended to the case in which the vector spaces are Banach. While we expect all the fundamental identities (expressed without reference to a basis) to carry over, their proofs would require non-trivial modification, since they make use of coordinates, duality, etc. Some of the delicate aspects of rough paths in infinite dimensions are considered in [Wei18].

It would also be interesting to leverage the identities of this chapter to see whether Markovian rough paths [FV10b, Ch.16] can be expressed in a more explicit manner. One could then apply such a representation to study the Cartan development of these rough paths on manifolds, in particular with regards to the question of whether Markovianity is preserved (as it does in the case of Brownian motion) and, if so, how the expression of the generator of the developed rough path relates to that of that of the rough path being developed.

4

A transfer principle for branched rough paths

Project status. This chapter may be considered a finished project, though I have not yet uploaded it to arXiv or submitted it to a journal at the time of writing.

Introduction

As we have already mentioned a number of times in this thesis, Malliavin's original transfer principle consists of the formal equivalence between ordinary calculus of manifold-valued smooth curves and Stratonvich calculus of manifold-valued semimartingales. Building on results of Meyer [Mey81, Mey82], in [É90] Emery discovered the Itô transfer principle, the rule that makes it possible to define Itô integration and RDEs on a manifold with connection: condensed into a single formula, this consists of defining the Itô differential $\mathrm{d}_\nabla X^\gamma := \mathrm{d}X^\gamma + \frac{1}{2}\Gamma^\gamma_{\alpha\beta}(X)\mathrm{d}[X]^{\alpha\beta}$. In Chapter 2 we readapted this idea to non-geometric rough paths of bounded $[2,3) \ni p$-variation, additionally exploring its extrinsic formulation and the local theory of parallel transport and development. In Chapter 3 we instead studied the combinatorial structure of rough paths and their controlled paths, and used this to derive the trivial transfer principle for geometric rough paths. The goal of this chapter is to combine the challenges stemming from non-geometricity with those due to low path regularity, in order to formulate a transfer principle that is valid for general branched rough paths.

Formulating the transfer principle in this context is significantly more complex than in the $2 \leq p < 3$ case. This is mainly due to the way in which branched rough paths and the solutions to their RDEs transform under the action of a smooth function. Such change of variable formulae were discovered by Kelly in his PhD thesis [Kel12], and are the main technical tool needed here. An interesting, but complicating feature of these is that RDE solutions transform in a strictly more complex manner than the trace of the branched rough path does, something which only becomes visible when $p \geq 3$. As a result, one should expect a transfer principle for RDEs on manifolds to be more complex than the one that suffices for rough integration. For this reason, we have only focused on the latter, leaving the former for future investigation.

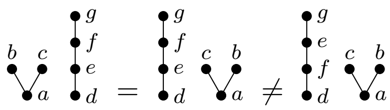
In Section 4.1 we begin by reviewing the background on branched rough paths, introduced by Gubinelli in [Gub10]. This includes a discussion of Kelly's bracket extension, needed for the change of variable formula. Next, in Section 4.2 we define the lift of a controlled path: the proof that this defines a rough path is more involved than its counterpart in the geometric case, and requires some delicate manipulations of rooted forests. When pushing forward a branched rough path, it is necessary (for the final goal of this chapter) to also push forward the simple bracket extension. This can be viewed as a generalisation of the rule for obtaining the quadratic covariation of Itô integrals, and requires an additional consistency relation on the simple bracket. Section 4.3 is a digression into a topic that has received relatively little attention in the literature, but which has potentially interesting consequences within our scope. Quasi-geometric rough paths [Bel20] constitute a class of rough paths that lie somewhere between geometric ones and the most general branched ones. They are compact enough to be defined on the tensor algebra, but do not obey the ordinary integration by parts rule formulated in terms of shuffles: rather, they are defined on Hoffman's quasi-shuffle algebra [Hof00]. The main result of this section is a characterisation of them which, in a nutshell, says that they are precisely those branched rough paths for which RDE solutions do not transform in a more complicated way than the trace of the original rough path does. We take the opportunity to give a brief survey on some facts about geometric, quasi-geometric and branched rough paths that are likely known by experts, but not easy to find in the literature. In Section 4.4 we define the transfer principle necessary to define rough integrals of one-forms against branched rough paths valued in a manifold carrying a covariant derivative. This entails defining certain higher-order Christoffel symbols (which are nevertheless completely determined by the connection), and involves a few subtleties pertaining to their symmetry (or lack thereof). While this chapter may be considered finished work, the section Conclusions and further directions is unusually long. In it, we mention several ways to build on the material covered, some within fairly easy reach and others which would require significant further effort.

4.1 Background on $\mathbb{R}^d$-valued branched rough paths

4.1.1 The Connes-Kreimer and Grossman-Larson Hopf algebras

In this subsection we will go over the algebraic prerequisites to the rest of the chapter. We will follow mainly [Foi13] (see also [Hof03]) for Hopf algebras of forests and [MM65, Man06] for the more general theory of Hopf algebras; when details and proofs are omitted it is intended that they are to be found therein. Most of the choices in setup and notation will follow [HK15] (for instance in the decision to define the Grossman-Larson Hopf algebra using forests rather than trees with unlabelled root, as done in most of the literature), but will deviate from in some aspects that will be motivated later on (for instance in the use of inhomogeneous gradings, see for instance [TZ20]).

In this chapter we will be interested in *A-decorated non-planar rooted forests*, where A is a finite alphabet. These are finite acyclic graphs with a finite number of vertices, which are labelled with elements of A, and each connected component of which - a *tree* - has a preferred vertex, its *root*. In graphical representations the root of a tree will be identified as its single lowermost vertex. The term "non-planar" refers to the fact that the trees in a forest, and the *children* of each vertex (the vertices attached to it that are further away from the root) are not given an order. An example is



in general for $a, \dots, g \in A$. The first and second terms are equal by non-planarity, and the third is different (unless $e = f$). Trees such as this term - ones in which each vertex has at most one child - are called *ladders*.

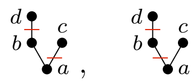
Call the set of such forests $\mathcal{F}^A$ and its subset of trees $\mathcal{T}^A$; note that we are considering an empty forest $\varnothing \in \mathcal{F}^A$ (which is not considered a tree, as it has no root). We will say that a non-empty forest is *proper* if it is not a tree. We will denote forests using letters $f, g, h, \dots$ and trees with $r, s, t, \dots$ We will write $\nu \in f$ to mean that ν is a vertex of f and we denote $\ell(\nu)$ its label. We will assume that A comes with a *weighting*, i.e. each element $a \in A$ has a weight $|a| \in \mathbb{N}^*$. This induces a grading on $\mathcal{F}^A$ by setting $|f| \coloneqq \sum_{\nu \in f} |\ell(\nu)|$, the *degree* of f. We will instead denote $\# f$ the number of vertices of f; more in general, throughout this chapter, we will use $\#$ to denote cardinality, reserving $|\cdot|$ for weightings and gradings. It will sometimes be helpful to write $f = t_1 \cdots t_n$ when the forest f is composed of the individual trees $t_1, \dots, t_n$, which are called its *factors*; note that this product, which will also be defined between forests and denoted simply $\cdot$, is the free abelian one. It will also be helpful to use the notation $t = [f]_a$ when the tree t is given by joining each root in the forest f to a new root labelled $a \in A$, and note that $[\varnothing]_a \coloneqq \bullet^a$.

An important case for the alphabet is

$$[d] \coloneqq \{1, \dots, d\} \tag{4.1}$$

for some $d \in \mathbb{N}^*$, with the homogeneous grading $|\gamma| \equiv 1$ for $\gamma = 1, \dots, d$. In this case we will denote $\mathcal{F}^A \eqqcolon \mathcal{F}^d$, and similarly replace all A superscripts with d's. We will usually use Greek letters for elements of $[d]$, reserving $a, b, c, \dots$ for more general labels.

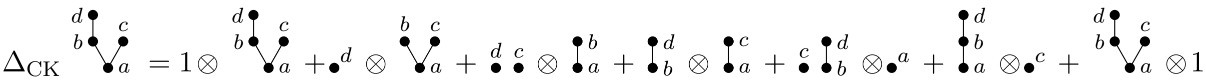
We will now introduce algebraic operations on $\mathbb{R}\langle \mathcal{F}^A \rangle$, the graded $\mathbb{R}$-vector space generated by $\mathcal{F}^A$, of which we identify the subspace generated by the empty forest with $\mathbb{R}$, by $\varnothing = 1$. A non-total *cut* C of $t \in \mathcal{T}^A$ is a subset of its edges. It is called *admissible* if it has the property that every increasing path in t contains at most one element of C. For example



both define cuts of the underlying tree, but only the first is admissible. The admissible cut $\varnothing$ is called the *trivial cut*. Deleting the edges in a non-total cut C transforms t into a forest t_C; if C is admissible we call $\overline{t}_C$ the tree containing its root (think of the portion of t below the cut) and $\underline{t}_C$ the forest comprised of all other factors of f (think of the portion of t above the cut). The trivial cut can be thought of as a cut above the leaves, since $\overline{t}_\varnothing = t$, $\underline{t}_\varnothing = \varnothing$. We also consider the *total cut* $\forall$, which is declared admissible and for which we set $\overline{t}_\forall = \varnothing$, $\underline{t}_\forall = t$; this cut, which does not correspond to any set of edges, should be thought of as a cut below the root. The set of cuts of $t \in \mathcal{T}^A$ (including $\forall$) is denoted $\mathrm{Cut}(t)$ and its subset of admissible ones $\mathrm{Cut}^*(t)$. We will also speak of cuts of a forest $t_1 \cdots t_n$: this is just a collection of cuts, one for each t_k. The *Connes-Kreimer* *coproduct* is given by

$$\Delta_{\mathrm{CK}} \colon \mathbb{R}\langle \mathcal{F}^A \rangle \to \mathbb{R}\langle \mathcal{F}^A \rangle \otimes \mathbb{R}\langle \mathcal{F}^A \rangle, \quad \Delta_{\mathrm{CK}} t \coloneqq \sum_{C \in \mathrm{Cut}^*(t)} \underline{t}_C \otimes \overline{t}_C \ \text{ for } t \in \mathcal{T}^A \tag{4.2}$$

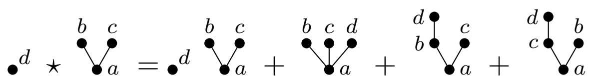
and required to be an algebra morphism according to the free abelian product of trees, i.e. $\Delta_{\mathrm{CK}}(t_1 \cdots t_n) = \Delta_{\mathrm{CK}} t_1 \cdots \Delta_{\mathrm{CK}} t_n$ with the product on the right given factor-wise, and extending linearly. For example



We now define the operation that is dual to Δ_{CK}, in a sense that will be made precise below. For $t_1 \cdots t_n = \mathscr{f}, \mathscr{g} \in \mathcal{F}^A$ we will say that $\mathscr{h}$ is obtained by *grafting* $\mathscr{f}$ onto $\mathscr{g}$, denoted $\mathscr{h} \in \mathscr{f} \frown \mathscr{g}$, if $\mathscr{h}$ is obtained by taking each factor t_k and either joining its root to a vertex of $\mathscr{g}$ (by adding an extra edge) or multiplying it with $\mathscr{g}$ (i.e. making it one of the factors of $\mathscr{h}$). Note that when we sum over $\mathscr{h} \in \mathscr{f} \frown \mathscr{g}$ we are not doing so over all distinct forests that are given by grafting $\mathscr{f}$ onto $\mathscr{g}$, but over distinct ways of grafting: the point is that there may be two distinct vertices in $\mathscr{g}$ s.t. grafting $\mathscr{f}$ onto them results in two identical labelled forests; for this reason $\mathscr{f} \frown \mathscr{g}$ is best thought as a multiset, not a set. Also note that $\varnothing \frown \mathscr{g}$ and $\mathscr{g} \frown \varnothing$ both consist of the singleton $\{\mathscr{g}\}$. We then define the *Grossman-Larson product*

$$\star \colon \mathbb{R}\langle \mathcal{F}^A \rangle \otimes \mathbb{R}\langle \mathcal{F}^A \rangle \to \langle \mathcal{F}^A \rangle, \quad \mathscr{f} \star \mathscr{g} \coloneqq \sum_{\mathscr{h} \in \mathscr{f} \frown \mathscr{g}} \mathscr{h} \tag{4.3}$$

and extending linearly. An example is



Note how the last two summands are the same if $b = c$. There is one operation that is left to define: the coproduct dual to the free abelian product of forests. We define the *Grossman-Larson coproduct*

$$\Delta_{\mathrm{GL}} \colon \mathbb{R}\langle \mathcal{F}^A \rangle \to \mathbb{R}\langle \mathcal{F}^A \rangle \otimes \mathbb{R}\langle \mathcal{F}^A \rangle, \quad \Delta_{\mathrm{GL}}(t_1 \cdots t_n) \coloneqq \sum_{I \sqcup J = \{1,\ldots,n\}} t_I \otimes t_J \tag{4.4}$$

where we are summing over all subsets I of the set with n elements, with J its complement, and for $K \subseteq \{1, \ldots, n\}$ we are defining $t_K \coloneqq \prod_{k \in K} t_k$. We now define the algebraic structure into which we would like these operations to fit.

Definition 4.1 (Connected graded bialgebra)**.** A *connected graded Hopf algebra* is a triple $(H, \times, \Delta)$ where $H = \bigoplus_{n \in \mathbb{N}} H^n$ is a graded real vector space, $\times \colon H \otimes H \to H$ (the *product*) and $\Delta \colon H \to H \otimes H$ (the *coproduct*) are linear functions, and the following axioms are satisfied for all $x, y, z \in H$:

Associativity. $(x \times y) \times z = x \times (y \times z)$;

Unit. There exists a *unit*, i.e. a linear map $\iota \colon \mathbb{R} \to H$ s.t. $x \times \iota(1) = x = \iota(1) \times x$ and $\Delta \circ \iota = \iota \otimes \iota$;

Coassociativity. $(\mathbb{1}_H \otimes \Delta) \circ \Delta = (\Delta \otimes \mathbb{1}_H) \circ \Delta$;

Counit. There exists a *counit*, i.e. a linear map $\epsilon \colon H \to \mathbb{R}$ s.t. $(\epsilon \otimes \mathbb{1}_H) \circ \Delta = \mathbb{1}_H = (\mathbb{1}_H \otimes \epsilon) \circ \Delta$ and $\epsilon(x \times y) = \epsilon x \times \epsilon y$;

Compatibility. $\Delta(x \times y) = \Delta x \times \Delta y$ and $\epsilon \circ \iota = \mathbb{1}_\mathbb{R}$;

Grading. $H^i \times H^j \subseteq H^{i+j}$ and $\Delta H^n \subseteq \bigoplus_{i+j=n} H^i \otimes H^j$;

Connectedness. $H^0 \cong \mathbb{R}$.

From these axioms, which define a connected graded bialgebra, it is possible to show the following further property, that turns H into a Hopf algebra:

Antipode. There exists an *antipode*, i.e. a linear map $\mathcal{S} \colon H \to H$ s.t. $\times \circ (\mathcal{S} \otimes \mathbb{1}_H) \circ \Delta = \iota \circ \epsilon = \times \circ (\mathbb{1}_H \otimes \mathcal{S}) \circ \Delta$.

We also consider the following two optional properties, denoting $\tau \colon H \otimes H \to H \otimes H$ the switch of factors:

Commutativity $\times \circ \tau = \times$;

Cocommutativity $\tau \circ \Delta = \Delta$.

If the unit and co-unit exist they are unique, and moreover we have $\operatorname{Im} \iota = H^0$, $\operatorname{Ker} \epsilon = \bigoplus_{n \geq 1} H^n$, and we will use ι to identify $H^0 = \mathbb{R}$. Less trivially, uniqueness also holds for the antipode: this fact, the proof of which uses the grading in an essential way, means that the antipode does not constitute additional structure; in a similar spirit, it can be shown that a bialgebra morphism, defined in the obvious way, automatically preserves the antipode. The *reduced coproduct* is defined by

$$\widetilde{\Delta} x \coloneqq \Delta x - x \otimes 1 - 1 \otimes x \in \bigoplus_{i,j \geq 1} H^i \otimes H^j \tag{4.5}$$

We may also iterate the coproduct by defining

$$\begin{aligned} &\Delta^m \colon H \to H^{\otimes m}, \quad \Delta^0 \coloneqq 1_\mathbb{R},\ \Delta^1 \coloneqq \mathbb{1},\ \Delta^2 \coloneqq \Delta, \\ &\Delta^m \coloneqq (\mathbb{1}_H \otimes \Delta^{m-1}) \circ \Delta = (\Delta^{m-1} \otimes \mathbb{1}_H) \circ \Delta \quad \text{for } m \geq 3 \end{aligned} \tag{4.6}$$

with the last identity holding by coassociativity. We can similarly iterate the reduced coproduct, $\widetilde{\Delta}^m$. We will use Sweedler notation

$$\Delta^m x \eqqcolon \sum_{(x)^m} x_{(1)} \otimes \cdots \otimes x_{(m)} \tag{4.7}$$

and we can modify the subscript $(x)^m$ to reflect whether we are reducing the coproduct, i.e. $(\widetilde{x})^m$, and/or to specify the specific coproduct used, e.g. $(x)^m_{\mathrm{CK}}$, $(x)^m_{\mathrm{GL}}$, and the superscript ^{m} will be omitted when it is 2.

An element $x \in H$ is *primitive* if $\Delta x = 1 \otimes x + x \otimes 1$ and *grouplike* if $\Delta x = x \otimes x$. The set of primitive elements will be denoted $\mathcal{P}(H)$ and forms a Lie algebra with bracket $[x, y] \coloneqq x \times y - y \times x$, in which $\mathcal{S}x = -x$. The set of grouplike elements will be denoted $\mathcal{G}(H)$ and forms a group in which $\mathcal{S}x = x^{-1}$. Such statements are not difficult to prove, e.g. for $x \in \mathcal{G}(H)$

$$\mathcal{S}x \times x = \times \circ (\mathcal{S} \otimes \mathbb{1}_H)(x \otimes x) = (\times \circ (\mathcal{S} \otimes \mathbb{1}_H) \circ \Delta)x = 1$$

We will adopt contractions in notation with obvious meaning such as $\mathcal{G}^A_{\mathrm{GL}} \coloneqq \mathcal{G}(\mathcal{H}^A_{\mathrm{GL}})$ (with $\mathcal{H}^A_{\mathrm{GL}}$ defined below).

Definition 4.2 (The Connes-Kreimer and Grossman-Larson bialgebras)**.** We call the triple $\mathcal{H}^A_{\mathrm{CK}} \coloneqq (\mathbb{R}\langle\mathcal{F}^A\rangle, \cdot, \Delta_{\mathrm{CK}})$ (where $\cdot$ denotes the free abelian product of forests) the *Connes-Kreimer Hopf algebra*, and the triple $\mathcal{H}^A_{\mathrm{GL}} \coloneqq (\mathbb{R}\langle\mathcal{F}^A\rangle, \star, \Delta_{\mathrm{GL}})$ the *Grossman-Larson Hopf algebra*.

That $\mathcal{H}^A_{\mathrm{CK}}$ and $\mathcal{H}^A_{\mathrm{GL}}$ are bialgebras is non-trivial. Before stating this result, we define the pairing that establishes $\mathcal{H}^A_{\mathrm{CK}}$ and $\mathcal{H}^A_{\mathrm{GL}}$ as dual to one another. For $\ell \in \mathcal{F}^A$ define $\mathcal{N}(\ell)$ to be the number of label-preserving order automorphisms of ℓ: this is recursively given by

$$\mathcal{N}(\varnothing) = 1, \qquad \mathcal{N}([\ell]_a) = \mathcal{N}(\ell), \qquad \mathcal{N}(\mathfrak{s}_1^{k_1}\cdots\mathfrak{s}_m^{k_m}) = \prod_{i=1}^m k_i!\mathcal{N}(\mathfrak{s}_i)^{k_i} \tag{4.8}$$

where $\mathfrak{s}_1 \cdots \mathfrak{s}_m$ are pairwise distinct trees when taking the labelling into account. We define the pairing

$$\langle\cdot,\cdot\rangle \colon \mathcal{H}^A_{\mathrm{CK}} \otimes \mathcal{H}^A_{\mathrm{GL}} \to \mathbb{R}, \quad \langle \ell, g\rangle \coloneqq \mathcal{N}(\ell)\delta_{\ell g} \tag{4.9}$$

where δ denotes Kronecker delta, and we are extending with bilinearity. This induces a corresponding pairing on $\mathbb{R}\langle\mathcal{F}^A\rangle^{\otimes n}$ by applying the above pairing factorwise on elementary tensors, and multiplying. The purpose of this pairing is for $\star$ to be dual to Δ_{CK} and $\cdot$ to Δ_{GL} with respect to it. The reason for the $\mathcal{N}(\ell)$ factor is explained by the following example: calling $\delta(\cdot,\cdot)$ the pairing induced by the Kronecker delta on basis elements, we have

$$\delta\big(\Delta_{\mathrm{CK}} [\bullet_a \bullet_a]_b \,, \bullet_a \otimes [\bullet_a]_b\big) = 2 \neq 1 = \delta\big([\bullet_a \bullet_a]_b \,, \bullet_a \star [\bullet_a]_b\big)$$

since there are two distinct cuts that result in a non-zero evaluation on the left, but only one way of grafting that does so on the right. The pairing which takes into account the order of the automorphism group instead works:

$$\big\langle \Delta_{\mathrm{CK}} [\bullet_a \bullet_a]_b \,, \bullet_a \otimes [\bullet_a]_b \big\rangle = 2 = \big\langle [\bullet_a \bullet_a]_b \,, \bullet_a \star [\bullet_a]_b \big\rangle$$

It can be observed that something similar thing occurs with the operations $\cdot$ and Δ_{GL}. We now state the result that summarises the content of this subsection:

Theorem 4.3 (The pair $(\mathcal{H}^A_{\mathrm{CK}}, \mathcal{H}^A_{\mathrm{GL}})$)**.** *$\mathcal{H}^A_{\mathrm{CK}}$ and $\mathcal{H}^A_{\mathrm{GL}}$ are connected graded Hopf algebras, the former commutative and the latter cocommutative, and the map* (4.9) *defines a graded bialgebra pairing, i.e.*

$$\langle \Delta_{\mathrm{CK}} z, x \otimes y\rangle = \langle z, x \star y\rangle, \qquad \langle x \otimes y, \Delta_{\mathrm{GL}} z\rangle = \langle xy, z\rangle \tag{4.10}$$

and

$$i \neq j \Rightarrow \langle (\mathcal{H}^A_{\mathrm{CK}})_i, (\mathcal{H}^A_{\mathrm{GL}})^j\rangle = 0, \quad \langle x, (\mathcal{H}^A_{\mathrm{GL}})^{|x|}\rangle = 0 \Rightarrow x = 0, \quad \langle (\mathcal{H}^A_{\mathrm{CK}})_{|y|}, y\rangle = 0 \Rightarrow y = 0 \tag{4.11}$$

Note how the above notion of graded duality is different from ordinary duality: the former is equivalent to an isomorphism $\bigoplus_{n\in\mathbb{N}} H^n \cong \bigoplus_{n\in\mathbb{N}} H_n$, where H_n is the dual H^n, defined unambiguously if H_n is finite-dimensional, as is the case here. Graded duality has the advantage of not introducing direct products (which would be needed in the study of full signatures, but are not when only dealing with rough paths), while still maintaining a lot of the necessary functoriality. For example, if $f\colon \bigoplus_{n\in\mathbb{N}} H^n \to \bigoplus_{n\in\mathbb{N}} K^n$ is a linear for which there exists m s.t. $f(H_n) \subseteq K_{n+m}$ for all n, it induces a unique map $f^*\colon \bigoplus_{n\in\mathbb{N}} K_n \to \bigoplus_{n\in\mathbb{N}} H_n$ s.t. $\langle f^*(y), x\rangle = \langle y, f(x)\rangle$.

The (uniquely determined) Connes-Kreimer antipode is given by

$$\mathcal{S}_{\mathrm{CK}}(t) = \sum_{\forall \neq C \in \mathrm{Cut}(t)} (-1)^{|C|+1} t_C, \quad t \in \mathcal{T}^A \tag{4.12}$$

(note that we are summing over all non-total cuts, not just the admissible ones) and extended as an algebra morphism. In a graded dual pair of Hopf algebras, the antipodes are graded dual to one another, so we can obtain the Grossman-Larson antipode as $\mathcal{S}_{\mathrm{GL}} = \mathcal{S}^*_{\mathrm{CK}}$.

Note that the fact that the pairing of $\mathcal{H}^A_{\mathrm{CK}}$ and $\mathcal{H}^A_{\mathrm{GL}}$ is not the Kronecker one implies that covariant and contravariant components no longer coincide: we reserve sub/super-scripting for the former, i.e. we denote

$$x^{f} \coloneqq \langle f, x\rangle, \quad y_{f} \coloneqq \langle y, f\rangle \qquad \text{for } x \in \mathcal{H}^A_{\mathrm{GL}},\ y \in \mathcal{H}^A_{\mathrm{CK}},\ f \in \mathcal{F}^A \tag{4.13}$$

The contravariant component of x w.r.t. f, on the other hand, is $\delta(f, x)$, meaning that we can express x as the finite sum $\sum_{f \in \mathcal{F}^A} \delta(f, x) f$ (and similarly for $y \in \mathcal{H}^A_{\mathrm{GL}}$). As a consequence $\langle y, x\rangle \neq y_f x^f$, rather

$$\begin{aligned}
\langle y, x\rangle &= \Big\langle \sum_{g \in \mathcal{F}^A} \delta(y, g) g, \sum_{f \in \mathcal{F}^A} \delta(x, f) f \Big\rangle \\
&= \Big\langle \sum_{g \in \mathcal{F}^A} \mathcal{N}(g)^{-1} \langle g, y\rangle g, \sum_{f \in \mathcal{F}^A} \mathcal{N}(f)^{-1} \langle f, x\rangle f \Big\rangle \\
&= \sum_{f, g \in \mathcal{F}^A} \mathcal{N}(g)^{-1} \mathcal{N}(f)^{-1} y_g x^f \langle g, f\rangle \\
&= \sum_{f \in \mathcal{F}^A} \mathcal{N}(f)^{-1} y_f x^f
\end{aligned} \tag{4.14}$$

We conclude this subsection with a couple of remarks.

Remark 4.4 (Different pairing used in [Kel12, HK15]). In these works $\mathcal{H}^A_{\mathrm{CK}}$ and $\mathcal{H}^A_{\mathrm{GL}}$ are paired using $\delta(\cdot,\cdot)$, not the pairing (4.9) that takes into account order automorphisms. For this reason, some of their identities will require minor modification for them to fit into the framework used here.

Remark 4.5 (Forest bialgebras over abstract vector spaces). $\mathcal{H}^A_{\mathrm{CK}}$ and $\mathcal{H}^A_{\mathrm{GL}}$ may be considered "bialgebras over $\mathbb{R}^A$", in the sense that $(\mathcal{H}^A_{\mathrm{CK}})_1 = \mathbb{R}^A = (\mathcal{H}^A_{\mathrm{GL}})^1$ canonically. The theory needed to replace $\mathbb{R}^A$ with a possibly infinite-dimensional (locally-convex) abstract vector space is developed in [Wei18]. Since we confine ourselves to the finite-dimensional case, for our purposes it makes more sense to begin by fixing coordinates, and later make the manifold-valued theory coordinate-free by considering suitably compatible families of rough paths defined w.r.t. arbitrary charts.

4.1.2 Rough paths, their controlled paths, rough integration and RDEs

In this subsection we introduce the topic of branched rough paths, original to [Gub10]. We will follow [HK15], and omit proofs and details that can be found therein; in light of Remark 4.4 we will make small adjustments to some formulae, without modifying their meaning in any significant way. We preface the main definition with a couple of preliminary ones. For $T \geq 0$ let $\Delta_T := \{(s,t) \in [0,T]^2 \mid s \leq t\}$. A *control* on $[0,T]$ is a continuous function $\omega\colon \Delta_T \to [0,+\infty)$ s.t. $\omega(t,t) = 0$ for $0 \leq t \leq T$ and is superadditive, i.e. $\omega(s,u) + \omega(u,t) \leq \omega(s,t)$ for $0 \leq s \leq u \leq t \leq T$. Throughout this chapter p will denote a real number $\in [1,+\infty)$. We will denote $(\mathcal{H}^A_{\mathrm{GL}})^n$ the sub-vector space of $\mathcal{H}^A_{\mathrm{GL}}$ generated by forest of weight n and $(\mathcal{H}^A_{\mathrm{GL}})^{\leq n}$ that generated by forest of degree $k \leq n$. We will also use similar sub/superscripts for projection on such subspaces, e.g. $x^{\leq n}$ is the projection of $x \in \mathcal{H}^A_{\mathrm{GL}}$ onto $(\mathcal{H}^A_{\mathrm{GL}})^{\leq n}$. When referring to $\mathcal{H}_{\mathrm{CK}}^A$ we will use subscripts instead of superscripts, to emphasize duality.

Definition 4.6 (Branched rough path)**.** An $\mathbb{R}^A$-valued *p-branched rough path* (of *inhomogeneous regularity* given by the weighting on A) on $[0,T]$ controlled by ω is a continuous map

$$\boldsymbol{X}\colon \Delta_T \to (\mathcal{H}^A_{\mathrm{GL}})^{\leq \lfloor p \rfloor}, \quad (s,t) \mapsto \boldsymbol{X}_{st} \tag{4.15}$$

s.t. $\boldsymbol{X}^{\varnothing} \equiv 1$ and satisfying the following three axioms:

Regularity. $\displaystyle\sup_{0\leq s<t\leq T} \frac{|\boldsymbol{X}^{f}_{st}|}{\omega(s,t)^{|f|/p}} < \infty$ for $f \in (\mathcal{F}^A)^{|\leq p|}$;

Multiplicativity. $\boldsymbol{X}_{st} = (\boldsymbol{X}_{su} \star \boldsymbol{X}_{ut})^{\leq\lfloor p \rfloor}$, or in coordinates $\boldsymbol{X}^{f}_{st} = \displaystyle\sum_{(f)_{\mathrm{CK}}} \boldsymbol{X}^{f_{(1)}}_{su} \boldsymbol{X}^{f_{(2)}}_{ut}$ for $f \in (\mathcal{F}^d)^{\leq \lfloor p \rfloor}$, and $0 \leq s \leq u \leq t \leq T$;

Grouplikeness. $\Delta_{\mathrm{GL}} \boldsymbol{X}_{st} = \boldsymbol{X}_{st} \otimes \boldsymbol{X}_{st}$, or in coordinates $\boldsymbol{X}^{fg}_{st} = \boldsymbol{X}^{f}_{st}\boldsymbol{X}^{g}_{st}$ for $f, g \in \mathcal{F}^A$ with $|f| + |g| \leq \lfloor p \rfloor$, and $0 \leq s \leq t \leq T$.

We denote the set of these $\mathscr{C}^p_\omega([0,T], \mathbb{R}^A)$.

The intuitive meaning of a branched rough path is given by the following recursive set of identities:

$$\boldsymbol{X}^{\varnothing}_{st} = 1, \qquad \boldsymbol{X}^{t_1 \cdots t_n}_{st} = \boldsymbol{X}^{t_1}_{st} \cdots \boldsymbol{X}^{t_n}_{st}, \qquad \boldsymbol{X}^{[f]_a}_{st} = \int_s^t \boldsymbol{X}^{f}_{su} \mathrm{d}X^a_u \tag{4.16}$$

for $t_k \in \mathcal{T}^A, f \in \mathcal{F}^A$. While the second identity is implied by the definition (and the first is actually required), the second is only to be taken heuristically, as the integral is not well defined in general. Of course, when $|f| \geq \lfloor p \rfloor$ the term $\boldsymbol{X}^{[f]_a}_{st}$ could be defined by taking the above identity literally in the sense of Young: this, together with the identity for products, would automatically define $\boldsymbol{X}^{g}_{st}$ for any $g \in \mathcal{F}^A$ and is called the *Lyons extension* of $\boldsymbol{X}$; in this chapter, however, we will always consider $\boldsymbol{X}$ to be truncated at order $\lfloor p \rfloor$: this will ensure that all of our sums are finite, yet precise at the necessary order. When equipped with an initial value X_0, the components of $\boldsymbol{X}$ indexed by single labelled vertices are the increments of components X^a of a continuous function $X\colon [0,T] \to \mathbb{R}^A$ called the *trace*; X is a member of $\mathcal{C}^p_\omega([0,T], \mathbb{R}^A)$, the set of functions $Y\colon [0,T] \to \mathbb{R}^A$ with the property that for $a \in A$

$$\sup_{0\leq s<t\leq T} \frac{|Y^a_{st}|}{\omega(s,t)^{|a|/p}} \tag{4.17}$$

where $Y_{st} := Y_t - Y_s$ is the increment. Note that the definitions of $\mathscr{C}^p_\omega([0,T], \mathbb{R}^A)$ and $\mathcal{C}$ depend on the weights assigned to elements of A. Also note that our setup accommodates rough paths of inhomogeneous

ω-Hölder regularity, but only ones that are integer multiples of a given p^{-1}: this choice is justified by the fact that the bracket terms considered in Subsection 4.1.3 (the only reason that has prompted us to consider inhomogeneous regularities) satisfies this property.

In what follows we will write $\approx_m$ between two real-valued quantities dependent on $0 \leq s \leq t \leq T$ to mean that their difference lies in $O(\omega(s,t)^{m/p})$ as $t \searrow s$, and simply $\approx$ (*almost* equal) to mean $\approx_{\lfloor p \rfloor+1}$. We will need the following simple readaptation of a well-known principle: its proof is identical to that of Proposition 3.9 (the basis of trees of $\mathcal{H}^A_{\mathrm{CK}}$ plays the role of the Lyndon basis).

Proposition 4.7 (Almost rough paths)**.** *Let $\widetilde{\boldsymbol{X}}$ be as in Definition 4.6, with the difference that the $=$'s signs in the multiplicativity and grouplikeness axioms are replaced with $\approx$'s. Then there exists a unique p-rough path $\boldsymbol{X}$ with the property that $\boldsymbol{X}_{st} \approx \widetilde{\boldsymbol{X}}_{st}$.*

We now give the definition of path controlled by a branched rough path.

Definition 4.8. Let $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^A)$. An $\mathbb{R}^e$-valued $\boldsymbol{X}$-controlled path $\boldsymbol{H}$ is an element of $\mathcal{C}^p([0,T],(\mathcal{H}^{A}_{\mathrm{CK}}{}^{\leq \lfloor p \rfloor -1})^{\times e})$ with homogeneous grading $\equiv 1$ on the target, and s.t. for $n = 0, \ldots, \lfloor p \rfloor - 2$

$$\mathbb{R}^e \ni \langle \boldsymbol{H}_{n;t}, y \rangle \approx_{\lfloor p \rfloor - n} \langle \boldsymbol{H}_s, \boldsymbol{X}_{st} \star y \rangle, \quad y \in (\mathcal{H}^A_{\mathrm{GL}})^n \tag{4.18}$$

Call the set of these $\mathscr{D}_{\boldsymbol{X}}(\mathbb{R}^e)$.

Note that $\boldsymbol{H}$ is a vector space-valued path; the homogeneous grading on the e-fold cartesian product of $(\mathcal{H}^A_{\mathrm{CK}})^{\leq \lfloor p \rfloor -1}$ just reflects the fact that its components, indexed by $[e] \times (\mathcal{F}^A)^{\leq \lfloor p \rfloor -1}$ (recall that we are denoting $[e] \coloneqq \{1, \ldots, e\}$), all have regularity p: this is because in all cases of interest $\boldsymbol{H}$ will be defined explicitly in terms of the whole of $\boldsymbol{X}$ and will therefore be, in general, as regular as the least regular component of X. We will denote the components of $\boldsymbol{H}$ as $\boldsymbol{H}^k_{\mathcal{f}}$ where $\mathcal{f} \in (\mathcal{F}^A)^{\leq \lfloor p \rfloor -1}$ and $k = 1, \ldots, e$; the terms $H \coloneqq \boldsymbol{H}_\varnothing \in \mathcal{C}^p([0,T],\mathbb{R}^e)$ will be called the *trace*, and the rest its *Gubinelli derivatives*. Note that in (4.18) the pairing is intended as componentwise on the upper index, which we will often omit. Using (4.14) we can express (4.18) as

$$\begin{aligned}
\boldsymbol{H}_{\mathcal{f};t} &\approx_{\lfloor p \rfloor - |\mathcal{f}|} \langle \boldsymbol{H}_s, \boldsymbol{X}_{st} \star \mathcal{f} \rangle \\
&= \langle \Delta_{\mathrm{CK}} \boldsymbol{H}_s, \boldsymbol{X}_{st} \otimes \mathcal{f} \rangle \\
&= \sum_{\mathcal{g},\mathcal{h} \in \mathcal{F}^A} \mathcal{N}(\mathcal{f})^{-1} \mathcal{N}(\mathcal{g})^{-1} \langle \Delta_{\mathrm{CK}} \boldsymbol{H}_s, \mathcal{g} \otimes \mathcal{h} \rangle \langle \mathcal{g} \otimes \mathcal{h}, \boldsymbol{X}_{st} \otimes \mathcal{f} \rangle \\
&= \sum_{\mathcal{g},\mathcal{h} \in \mathcal{F}^A} \mathcal{N}(\mathcal{f})^{-1} \mathcal{N}(\mathcal{g})^{-1} \langle \boldsymbol{H}_s, \mathcal{g} \star \mathcal{h} \rangle \boldsymbol{X}^{\mathcal{g}}_{st} \langle \mathcal{h}, \mathcal{f} \rangle \\
&\approx_{\lfloor p \rfloor - |\mathcal{f}|} \sum_{\substack{\mathcal{g} \in \mathcal{F}^A \\ |\mathcal{g}| \leq \lfloor p \rfloor - |\mathcal{f}|}} \mathcal{N}(\mathcal{g})^{-1} \langle \boldsymbol{H}_s, \mathcal{g} \star \mathcal{f} \rangle \boldsymbol{X}^{\mathcal{g}}_{st}
\end{aligned}$$

or in other words

$$\boldsymbol{H}_{\mathcal{f};t} \approx_{\lfloor p \rfloor - |\mathcal{f}|} \sum_{\substack{\mathcal{g} \in \mathcal{F}^A \\ |\mathcal{g}| \leq \lfloor p \rfloor - |\mathcal{f}| \\ \mathcal{h} \in \mathcal{g} \frown \mathcal{f}}} \mathcal{N}(\mathcal{g})^{-1} \boldsymbol{H}_{\mathcal{h};s} \boldsymbol{X}^{\mathcal{g}}_{st} \tag{4.19}$$

which at the trace level reads

$$H_t \approx_{\lfloor p \rfloor} \langle \boldsymbol{H}_s, \boldsymbol{X}_{st} \rangle = \sum_{\mathscr{g} \in \mathcal{F}^A,\ |\mathscr{g}| \le \lfloor p \rfloor} \mathcal{N}(\mathscr{f})^{-1} \boldsymbol{H}_{\mathscr{g};s} \boldsymbol{X}^{\mathscr{g}}_{st} \tag{4.20}$$

Example 4.9. To convince oneself that this is indeed the form of the expansion that is self-consistent at higher orders, consider the ODE

$$\mathrm{d}Y = V(Y)\mathrm{d}X, \quad V \in C^\infty(\mathbb{R}^e, \mathbb{R}^{e \times d}) \tag{4.21}$$

The order-4 expansion of Y in terms of the branched iterated integrals of X (obtained by performing iterated substitutions $Y_{st} \leftarrow \int_s^t V(Y)\mathrm{d}X$ and Taylor expansions of $V(Y)$ of the necessary order) is given by

$$\begin{aligned}
Y^k_t \approx\ & Y^k_s \boldsymbol{X}^1_{st} + V^k_\gamma(Y_s)\boldsymbol{X}^{\bullet_\gamma}_{st} + \partial_h V^k_\beta V^h_\alpha(Y_s)\boldsymbol{X}^{[\alpha]_\beta}_{st} + \partial_h V^k_\gamma \partial_l V^h_\beta V^l_\alpha(Y_s)\boldsymbol{X}^{[[\alpha]_\beta]_\gamma}_{st} + \tfrac{1}{2}\partial_{ij}V^k_\gamma V^i_\alpha V^j_\beta(Y_s)\boldsymbol{X}^{[\alpha\beta]_\gamma}_{st} \\
& + \tfrac{1}{6}\partial_{ijh}V^k_\delta V^i_\alpha V^j_\beta V^h_\gamma(Y_s)\boldsymbol{X}^{[\alpha\beta\gamma]_\delta}_{st} + \partial_{ij}V^k_\delta \partial_h V^i_\beta V^j_\gamma V^h_\alpha(Y_s)\boldsymbol{X}^{[[\alpha]_\beta\gamma]_\delta}_{st} \\
& + \partial_h V^k_\delta \partial_l V^h_\gamma \partial_p V^l_\beta V^p_\alpha(Y_s)\boldsymbol{X}^{[[[\alpha]_\beta]_\gamma]_\delta}_{st} + \tfrac{1}{2}\partial_h V^k_\delta \partial_{ij} V^h_\gamma V^j_\alpha V^j_\beta(Y_s)\boldsymbol{X}^{[[\alpha\beta]_\gamma]_\delta}_{st}
\end{aligned}$$

with the Einstein convention implying a sum on the single indices (not on distinct labelled trees - this is what the fractions are for). We therefore have

$$\text{coefficient of } \boldsymbol{X}^{[\alpha\beta]_\gamma}_{st} = \begin{cases} \partial_{ij}V^k_\gamma V^i_\alpha V^j_\beta(Y_s) & \alpha \ne \beta \\ \tfrac{1}{2}\partial_{ij}V^k_\gamma V^i_\alpha V^j_\beta(Y_s) & \alpha = \beta \end{cases}$$

$$\text{coefficient of } \boldsymbol{X}^{[\alpha\beta\gamma]_\delta}_{st} = \begin{cases} \tfrac{1}{6}\partial_{ijh}V^k_\delta V^i_\alpha V^j_\beta V^h_\gamma(Y_s) & \alpha = \beta = \gamma \\ \tfrac{1}{2}\partial_{ijh}V^k_\delta V^i_\alpha V^j_\beta V^h_\gamma(Y_s) & \alpha \ne \beta = \gamma \ \vee\ \alpha = \beta \ne \gamma \ \vee\ \alpha = \gamma \ne \beta \\ \partial_{ijh}V^k_\delta V^i_\alpha V^j_\beta V^h_\gamma(Y_s) & \alpha \ne \beta \ne \gamma \ne \alpha \end{cases}$$

and a statement similar to the first for the last term in the expansion. Now, setting this expression equal to (4.19) with $\mathscr{f} = \varnothing$ already fixes all the $\boldsymbol{Y}_{\mathscr{g}}$'s to be equal to the terms above involving derivatives and products of the V^h_c's without the fractions, e.g.

$$\boldsymbol{Y}^k_{[\alpha\alpha]_\gamma} = \partial_{ij}V^k_\gamma V^i_\alpha V^j_\alpha(Y)$$

Re-expanding this term, we have

$$\begin{aligned}
\boldsymbol{Y}^k_{[\alpha\alpha]_\gamma;t} &\approx \partial_{ij}V^k_\gamma V^i_\alpha V^j_\alpha(Y_s) + (\partial_{ijh}V^k_\gamma V^i_\alpha V^j_\alpha V^h_\beta(Y_s) + 2\partial_{ij}V^k_\gamma \partial_h V^i_\alpha V^j_\alpha V^h_\beta(Y_s))X^\beta_{st} \\
&= \boldsymbol{Y}^k_{[\alpha\alpha]_\gamma;s}\boldsymbol{X}^1_{st} + (\boldsymbol{Y}^k_{[\alpha\alpha\beta]_\gamma;s} + 2\boldsymbol{Y}^k_{[[\beta]_\alpha\alpha]_\gamma;s})\boldsymbol{X}^{\bullet_\beta}_{st}
\end{aligned}$$

which is precisely the expression predicted by (4.19).

Remark 4.10. As exemplified by the calculations above, sums of the sort $\sum_{\mathscr{f}} \mathcal{N}(\mathscr{f})^{-1}\varphi(\mathscr{f})$ can be replaced with ones $\sum_{\widetilde{\mathscr{f}}}\sum_\ell \mathcal{N}(\widetilde{\mathscr{f}})^{-1}\varphi(\widetilde{\mathscr{f}}_\ell)$, where $\widetilde{\mathscr{f}}$ are unlabelled forests (ranging in the set corresponding to that of the $\mathscr{f}$'s), we are additionally summing over all possible labellings ℓ on each $\widetilde{\mathscr{f}}$ (i.e. maps from the set of vertices of each $\widetilde{\mathscr{f}}$

to A), $\mathcal{N}(\widetilde{\mathcal{f}})$ is the number of unlabelled order automorphisms of $\widetilde{\mathcal{f}}$, and $\widetilde{\mathcal{f}}_\ell$ is the forest $\widetilde{\mathcal{f}}$ labelled with ℓ. This is because in the latter type of sum each term $\mathcal{N}(\widetilde{\mathcal{f}})^{-1}\varphi(\mathcal{f})$ appears $|\mathrm{Aut}(\widetilde{\mathcal{f}})/\mathrm{Aut}(\mathcal{f})| = \mathcal{N}(\widetilde{\mathcal{f}})/\mathcal{N}(\mathcal{f})$ times.

Example 4.11 (Smooth functions of controlled paths)**.** If $\boldsymbol{H} \in \mathscr{D}_{\boldsymbol{X}}(\mathbb{R}^e)$ and $f \in C^\infty(\mathbb{R}^e, \mathbb{R}^c)$ we can define an $\boldsymbol{X}$-controlled path with trace $f(H)$ by

$$(f_*\boldsymbol{H})_{\mathcal{f}} \coloneqq \sum_{\mathcal{f}_1\cdots\mathcal{f}_n = \mathcal{f}} \partial_{k_1,\ldots,k_n} f(H) \boldsymbol{H}^{k_1}_{\mathcal{f}_1} \cdots \boldsymbol{H}^{k_n}_{\mathcal{f}_n} \tag{4.22}$$

where we are summing on all distinct ways of expressing the forest $\mathcal{f}$ as a product of forests $\mathcal{f}_1 \cdots \mathcal{f}_n$. Again, here "distinct" is intended in the sense of multisets: if $\mathcal{f}$ is the product of trees $t_1 \cdots t_n$, each term in the sum corresponds to a partition of the multiset $\{\{t_1, \ldots, t_n\}\}$; for instance if $\mathcal{f} = \mathcal{s}\mathcal{s}t$ with $\mathcal{s}, t \in \mathcal{T}^A$, the term corresponding to $\mathcal{f} = \mathcal{s} \cdot \mathcal{s}t$ appears not once, but twice. Note that the absence of the $1/n!$ factor, present in [HK15], is due to the different dual pairing, as explained in Remark 4.4 (and exemplified in Example 4.9). In particular, if $e = d$ and we take $\boldsymbol{H}$ to be the $\boldsymbol{X}$-controlled path defined by X itself, i.e. $\boldsymbol{X}^a_{\mathcal{f}} = \delta^{\bullet^a}_{\mathcal{f}}$, we have the expression for a controlled path given by a function of the trace of X:

$$\langle f(X), \bullet^{a_1} \cdots \bullet^{a_n} \rangle \coloneqq \partial_{a_1,\ldots,a_n} f(X) \tag{4.23}$$

and zero on all other forests; in this case (4.19) reduces to the usual Taylor expansion. In this case we will often just write $f(X)$ to denote the above controlled path, since it only depends on f and the trace X.

We continue by defining rough integration. We will call elements of $\mathscr{D}_{\boldsymbol{X}}(\mathbb{R}^{e\times A})$ $\boldsymbol{X}$*-controlled integrands*, and we will use subscripts for the A index, i.e. $\boldsymbol{H}^k_{\mathcal{f},a}$ for $k \in [e],\ a \in A,\ \mathcal{f} \in \mathcal{F}^A$. Setting, for $a \in A$

$$\langle \mathcal{f}, \boldsymbol{X}^a_{st} \rangle \coloneqq \langle [\mathcal{f}]_a, \boldsymbol{X}_{st} \rangle, \quad \mathcal{f} \in \mathcal{F}^A$$

It is shown that

$$\langle \boldsymbol{H}_{a;s}, \boldsymbol{X}^a_{st} \rangle - \langle \boldsymbol{H}_{a;s}, \boldsymbol{X}^a_{su} \rangle - \langle \boldsymbol{H}_{a;u}, \boldsymbol{X}^a_{ut} \rangle \approx 0$$

for $s \le u \le t$, enabling the following

Definition 4.12 (Rough integral)**.** We define the *rough integral* as the unique path with increments $\approx \langle \boldsymbol{H}_{a;s}, \boldsymbol{X}^a_{st} \rangle$, i.e.

$$\int_s^t \boldsymbol{H} \mathrm{d}\boldsymbol{X} \coloneqq \lim_{n\to\infty} \sum_{[u,v]\in\pi_n} \langle \boldsymbol{H}_{a;u}, \boldsymbol{X}^a_{uv} \rangle \tag{4.24}$$

where $(\pi_n)_n$ is any sequence of partitions on $[s, t]$ with vanishing step size as $n \to \infty$.

This limit is shown to be well defined, independently of $(\pi_n)_n$, and taking $s = 0,\ t \in [0, T]$ above yields an element of $\mathcal{C}^p([0, T], \mathbb{R}^e)$. In coordinates we have, again by (4.6)

$$\int_s^t \boldsymbol{H} \mathrm{d}\boldsymbol{X} \approx \sum_{\substack{\mathcal{f}\in\mathcal{F}^A,\ a\in A \\ |\mathcal{f}|+|a|\le\lfloor p \rfloor}} \mathcal{N}(\mathcal{f})^{-1} \boldsymbol{H}_{\mathcal{f},a;s} \boldsymbol{X}^{[\mathcal{f}]_a}_{st} \tag{4.25}$$

This becomes an element $\int_0^\cdot \boldsymbol{H}\mathrm{d}\boldsymbol{X} \in \mathscr{D}_{\boldsymbol{X}}(\mathbb{R}^e)$ by setting $(\int_0^\cdot \boldsymbol{H}\mathrm{d}\boldsymbol{X})_\varnothing$ to the above path, and

$$\left(\int_0^\cdot \boldsymbol{H}\mathrm{d}\boldsymbol{X}\right)_{[f]_a} := \boldsymbol{H}_{f,a}, \quad \left(\int_0^\cdot \boldsymbol{H}\mathrm{d}\boldsymbol{X}\right)_{g} := 0, \quad \text{for } g \in \mathcal{F}^A \setminus (\mathcal{T}^A \cup \{\varnothing\}) \tag{4.26}$$

We move on to the topic of rough differential equations, or RDEs. Let $F \in C^\infty(\mathbb{R}^e, \mathbb{R}^{e\times d})$. We wish to give meaning to the expression

$$\mathrm{d}\boldsymbol{Y} = F(Y)\mathrm{d}\boldsymbol{X}, \quad Y_0 = y_0 \in \mathbb{R}^e \tag{4.27}$$

Definition 4.13 (RDE)**.** We will say $\boldsymbol{Y} \in \mathscr{D}_{\boldsymbol{X}}(\mathbb{R}^e)$ is a *controlled solution* to (4.27) if

$$\boldsymbol{Y}_t = y_0 + \int_0^t F_* \boldsymbol{Y} \mathrm{d}\boldsymbol{X} \tag{4.28}$$

where the $\boldsymbol{X}$-controlled paths Y and F_*Y are defined by the rules (4.26) and (4.22).

Note that if such a $\boldsymbol{Y}$ exists all its Gubinelli derivatives are automatically fixed by the trace Y and the smooth function F. Their expression can be computed more explicitly in terms of recursively-defined smooth functions of Y as

$$\begin{aligned} &\boldsymbol{Y}_{[f]_a;t} = F_{[f]_a}(Y_t), \qquad \boldsymbol{Y}_{g} = 0 \quad \text{for } g \in \mathcal{F}^A \setminus (\mathcal{T}^A \cup \{\varnothing\}) \\ &\text{where} \quad F_\varnothing := \mathbb{1}_{\mathbb{R}^e}, \quad F_{[t_1\cdots t_n]_a} := \partial_{k_1\cdots k_n} F_a F^{k_1}_{t_1} \cdots F^{k_n}_{t_n} \end{aligned} \tag{4.29}$$

We can use this and (4.25) to express the trace level of (4.28) as

$$Y_{st} \approx \sum_{t \in \mathcal{T}^A,\, |t| \le \lfloor p \rfloor} \mathcal{N}(t)^{-1} F_t(Y_s) \boldsymbol{X}^t_{st}, \quad Y_0 = y_0 \tag{4.30}$$

This is known as the *Davie solution*, and it is equivalent to the notion of controlled solution in the sense that (4.30) holds if and only if there exists a controlled solution, which is necessarily given by (4.29). Another interesting feature of the coefficients F_t is how they behave when evaluated against Grossman-Larson products: it can be shown that

$$F_{(t_1\cdots t_n)\star s} = \partial_{k_1\cdots k_n} F_s F^{k_1}_{t_1} \cdots F^{k_1}_{t_n}, \qquad t_1, \ldots, t_n, s \in \mathcal{T}^A \tag{4.31}$$

Taking $s = \bullet^c$ (and setting F to be zero on proper non-empty forests) reduces this identity to (4.29).

4.1.3 Kelly's bracket extension

The lack of constraints on the product structure of branched rough paths results in rough integration against $\boldsymbol{X}$ not being sufficiently rich to express increments of functions of $\boldsymbol{X}$-driven RDEs, not even of X itself. In this section we will review the material of [Kel12, Ch. 5], which remedies this lack of a change of variable formula by means of an ingenious procedure that consists of enlarging $\boldsymbol{X}$ by recursively adding new trace components and coherently lifting to a rough path; details and proofs not included here are intended to be found therein.

Given the weighted alphabet A we consider the enlarged alphabet consisting of adding to A all non-trivial proper forests in $\mathcal{F}^A$

$$\widehat{A} := A \sqcup (\mathcal{F}^A \setminus (\mathcal{T}^A \sqcup \{\varnothing\})) \tag{4.32}$$

and its subalphabet consisting of commutative sequences, i.e. multisets, of letters in A

$$\widetilde{A} \coloneqq \{(a_1 \cdots a_n) \mid n \in \mathbb{N},\ a_1, \ldots, a_n \in A\} \tag{4.33}$$

We are considering $A \subseteq \widetilde{A} \subseteq \widehat{A}$, with the second inclusion given by identifying $a_1 \cdots a_n = \bullet^{a_1} \cdots \bullet^{a_n}$. We denote the sets and Hopf algebras constructed w.r.t. $\widetilde{A}$, $\widehat{A}$ correspondingly, e.g. $\widetilde{\mathcal{H}}^A_{\mathrm{CK}}$ and $\widehat{\mathcal{H}}^d_{\mathrm{CK}}$. Elements of $\widetilde{\mathcal{F}}^A$ are forests labelled with commutative products of elements of A, and elements of $\widehat{\mathcal{F}}^d$ are forests labelled with A and A-labelled non-trivial proper forests or single vertices. We will use round brackets to denote these new types of labels. The weighting on $\widehat{\mathcal{F}}^d$ (and accordingly that on its subset $\widetilde{\mathcal{F}}^A$) is just given by summing the weights of the labels as elements of $\mathcal{F}^A$, i.e. counting up the total number of elements in A.

Define the bilinear "root labelling" map

$$\mathcal{J} \colon \mathbb{R}\langle \mathcal{F}^A \rangle \otimes \mathbb{R}\langle \mathcal{F}^A \rangle \to \mathbb{R}\langle \widehat{\mathcal{F}}^d \rangle, \qquad f \otimes g \mapsto \begin{cases} 0 & g = \varnothing \vee (g \in \mathcal{T}^A \wedge \# g > 1) \\ [f]_{(g)} & \text{otherwise} \end{cases} \tag{4.34}$$

and the *bracket polynomial* maps

$$\ll\cdot\gg \coloneqq \mathbb{1} - \mathcal{J} \circ \widetilde{\Delta}_{\mathrm{CK}} \colon \mathbb{R}\langle \mathcal{F}^A \rangle \to \mathbb{R}\langle \mathcal{F}^A \rangle \tag{4.35}$$

where recall that $\widetilde{\Delta}_{\mathrm{CK}}$ denotes the reduced Connes-Kreimer coproduct. For products of single vertices (which we identify with their labels) this reduces to

$$\begin{gathered} \ll c_1 \cdots c_n \gg \coloneqq c_1 \cdots c_n - \sum_{I \sqcup J = \{1,\ldots,n\}} [\bullet^{a_1} \cdots \bullet^{a_r}]_{(b_1 \cdots b_q)} \\ I = \{i_1, \ldots, i_r\},\ J = \{j_1, \ldots, j_q\}, \quad a_k \coloneqq c_{i_k},\ b_k \coloneqq c_{j_k} \end{gathered} \tag{4.36}$$

Definition 4.14 (Bracket extension)**.** A (*full*) *bracket extension* $\widehat{\boldsymbol{X}}$ of $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T], \mathbb{R}^A)$ is a p-rough path over the alphabet $\widehat{A}$ extending the existing one over A, and with the property that

$$\widehat{\boldsymbol{X}}^{(f)} = \langle \ll f \gg, \widehat{\boldsymbol{X}} \rangle, \qquad f \in \mathcal{F}^A \tag{4.37}$$

A *simple bracket extension* of $\boldsymbol{X}$ is a p-rough path $\boldsymbol{X}$ over the alphabet $\widetilde{A}$ with the property that

$$\widetilde{\boldsymbol{X}}^{(a_1 \cdots a_n)} = \langle \ll a_1 \cdots a_n \gg, \widetilde{\boldsymbol{X}} \rangle, \qquad a_1, \ldots, a_n \in A \tag{4.38}$$

All of this is equivalent to expressing the evaluation of $\boldsymbol{X}$ against a forest in terms of evaluations of $\widehat{\boldsymbol{X}}$ against trees as

$$\boldsymbol{X}^f = \langle \mathcal{J} \circ \Delta_{\mathrm{CK}}(f), \widehat{\boldsymbol{X}} \rangle \tag{4.39}$$

(with the simple case obtained by picking $f = \bullet^{a_1} \cdots \bullet^{a_n}$) where we are using the unreduced coproduct, with $[\varnothing]_{(g)} = \bullet^{(g)}$; note that this reduces to the trivial identity $\boldsymbol{X}^t = \boldsymbol{X}^t$ when $t \in \mathcal{T}^A$, since the only term considered on the RHS is the one corresponding to the cut that separates the root from everything else.

A full bracket extension automatically defines a simple bracket extension by taking the trees in the forest f of (4.37) to be given by a product of single vertices. The way bracket extensions are shown to exist is by noticing that the defining condition fixes the trace term $X^{(t_1\cdots t_n)}$ as being canonically determined by $\boldsymbol{X}^{t_1\cdots t_n} = \boldsymbol{X}^{t_1}\cdots\boldsymbol{X}^{t_n}$ (part of the original rough path) and lower degree rough path terms in the extension. In [Kel12, Proposition 5.3.9] it is proved that every branched rough path admits a full bracket extension. The proof works by recursively performing the following two steps: (i) show that the $\boldsymbol{X}^{(t_1\cdots t_n)}_{st}$'s of some fixed degree are actually path increments and (ii) lift them jointly and consistently with the rough path defined in previous iterations. Note that the second step is highly non-canonical: there are many different choices of a bracket extension. The point of view taken in this chapter is slightly different: the bracket extension, simple or full, will be fixed and part of the original data, since subsequent constructions will explicitly depend on it. In practical (stochastic) cases it is realistic to hope that the bracket terms can be defined canonically through the same mechanism which is used to define the original branched rough path. The main purpose of the bracket extension is the following result, which we state directly for signal-dependent RDEs:

Theorem 4.15 (Change of variable formula for RDE solutions [Kel12, Theorem 5.3.11])**.**
Let $\widehat{\boldsymbol{X}}$ be a bracket extension of $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T],\mathbb{R}^A)$, and $\boldsymbol{Y}$ be a solution to (4.27) (driven by the original rough path $\boldsymbol{X}$). For $g \in C^\infty\mathbb{R}^e$ we have (at the trace level)

$$
\begin{aligned}
g(Y)_{st} = &\int_s^t \partial_k g(Y) F^k_a(Y)\mathrm{d}\boldsymbol{X}^a \\
&+ \sum_{n=2}^{\lfloor p \rfloor} \frac{1}{n!}\int_s^t \partial_{k_1,\ldots,k_n} g(Y) F^{k_1}_{t_1}\cdots F^{k_n}_{t_n}(Y)\mathrm{d}\widehat{\boldsymbol{X}}^{(t_1\cdots t_n)}
\end{aligned}
\tag{4.40}
$$

where the t_i's range in $\mathcal{T}^A$, the k_i's in $[e]$ and the controlled integrands are defined using (4.29). In particular, for a simple bracket extension $\widetilde{\boldsymbol{X}}$ of $\boldsymbol{X}$ and $g \in C^\infty\mathbb{R}^d$

$$
g(X)_{st} = \sum_{n=1}^{\lfloor p \rfloor} \frac{1}{n!}\int_s^t \partial_{a_1,\ldots,a_n} g(X)\mathrm{d}\widetilde{\boldsymbol{X}}^{(a_1\cdots a_n)} \tag{4.41}
$$

where the a_i's range in A.

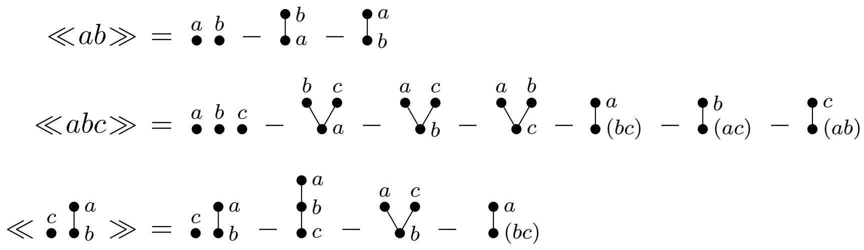
Note how in the above change of variable formulae the only terms of $\widehat{\boldsymbol{X}}$ (and $\widetilde{\boldsymbol{X}}$) that are needed to define the rough integrals are those $\widehat{\boldsymbol{X}}^{\widehat{f}}$ with $\widehat{f} \in \widehat{\mathcal{F}}^d$ in which the only new possible vertex labelled with an element of $\widehat{A} \setminus A$ is the root: this is because the trace of $\widehat{X}$ is only defined in terms of forests with such a labelling, and because integrands are $\boldsymbol{X}$-controlled (as opposed to, more generally, $\widehat{\boldsymbol{X}}$-controlled). However, it still makes more sense to consider the whole bracket extension - which enables us to consider $\widehat{\boldsymbol{X}}^{f}$ for any $f \in \widehat{\mathcal{F}}^d$ — since the terms in which the new labels appear in higher vertices of f will become relevant in the next section when defining the lift of a controlled path.

Remark 4.16. Note how, in light of (4.26), (4.40) and (4.41) respectively define $\widehat{\boldsymbol{X}}$- and $\widetilde{\boldsymbol{X}}$-controlled paths that are distinct to the $\boldsymbol{X}$-controlled ones given by the formulae (4.22) and (4.23). Although the latter do not require the bracket extension, the former have the advantage of vanishing on proper forests, precisely the property that is required to represent them as integrals. In the second case $g(X)$ can be $\widetilde{\boldsymbol{X}}$-controlled as

$$
\langle g(X), [\bullet^{a_1}\cdots\bullet^{a_i}]_{(b_1\cdots b_j)}\rangle = \mathcal{N}(b_1\cdots b_j)^{-1}\partial_{a_1,\ldots,a_i,b_1,\ldots,b_j} f(X) \tag{4.42}
$$

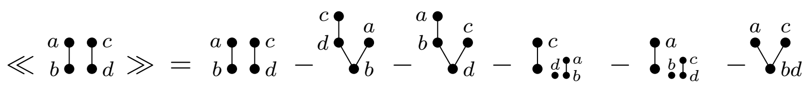
and zero on other types of trees. Here $\mathcal{N}(b_1\cdots b_j)$ is the order of the automorphism group of the corresponding forest with j single nodes. In expansions, we will often be summing not on the label $(b_1 \cdots b_j)$ (i.e. the multiset), but on the tuple $(b_1, \ldots, b_j)$, which means that the $\mathcal{N}(b_1 \cdots b_j)^{-1}$ will be replaced with $j!^{-1}$. This is a special case of Remark 4.10.

Example 4.17. We provide the expression of all bracket polynomials of up to order 3, which are sufficient for the above change of variable formula when $p < 4$.



(4.43)

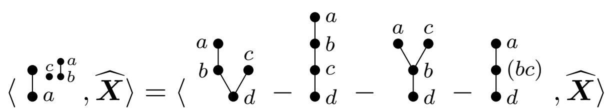
In order to see trees with labels in $\widehat{A} \setminus \widetilde{A}$ (forests that are not products of single vertices, that is) one must go one level higher: here is an example in which the forest has 4 vertices.



4.2 The extended lift of a controlled path

In this section we will show how, given $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T], \mathbb{R}^A)$ endowed with bracket extension $\widehat{\boldsymbol{X}}$, one can lift $\boldsymbol{H} \in \mathscr{D}_{\boldsymbol{X}}(\mathbb{R}^e)$ to a rough path $\uparrow_{\widehat{\boldsymbol{X}}} \boldsymbol{H} \in \mathscr{C}^p_\omega([0,T], \mathbb{R}^e)$ in a canonical fashion. Since this lift depends not only on $\boldsymbol{X}$ but also on $\widehat{\boldsymbol{X}}$, we will assume the bracket extension to be fixed and part of the initial data; we will denote $\widetilde{\boldsymbol{X}}$ the simple bracket extension determined by $\widehat{\boldsymbol{X}}$, which will be sufficient for certain constructions. Special attention will be given to the case of pushforwards, i.e. in which $\boldsymbol{H}$ is given by a smooth function of the trace X, for which a canonical simple bracket extension can be defined which only depends on $\widetilde{\boldsymbol{X}}$.

We must begin by imposing a new condition on our bracket extension: the relations that define it need to be required not only when the new label is on the root, but also higher up. For instance, one might expect it to be the case that



Given labelled forests $\mathscr{f}, \mathscr{g}$ we denote $\mathscr{f} \frown_\nu \mathscr{g}$ the forest obtained by grafting each root of $\mathscr{f}$ onto the vertex ν of $\mathscr{g}$; when we write $\nu \in \mathscr{g}$ we allow for the additional case $\nu = -$ which means we are simply multiplying $\mathscr{f}\mathscr{g}$. Note that by taking $\mathscr{g} = \varnothing$, $\nu = -$ below we recover the definition of ordinary bracket (4.39).

Bracket consistency. $\langle \mathscr{f} \frown_\nu \mathscr{g}, \widehat{\boldsymbol{X}} \rangle = \sum_{(\mathscr{f})_{\mathrm{CK}}} \langle [\mathscr{f}_{(1)}]_{(\mathscr{f}_{(2)})} \frown_\nu \mathscr{g}, \widehat{\boldsymbol{X}} \rangle$ for $\mathscr{f}, \mathscr{g} \in \mathcal{F}^A$, and $\nu \in \mathscr{g}$.

The condition can be rewritten as

$$
\begin{aligned}
\langle \bullet^{(\ell)} \frown_\nu g, \widehat{\boldsymbol{X}} \rangle &= \Big\langle \ell \frown_\nu g - \sum_{(\widetilde{\ell})_{\mathrm{CK}}} [\ell_{(1)}]_{(\ell_{(2)})} \frown_\nu g, \widehat{\boldsymbol{X}} \Big\rangle \\
&= \langle \ll \ell \gg \frown_\nu g, \widehat{\boldsymbol{X}} \rangle
\end{aligned}
\tag{4.44}
$$

where the last expression is defined by extending the $\frown_\nu$ operator linearly. We will refer to bracket extensions that satisfy this property as *consistent*, and denote respectively

$$
\widehat{\mathscr{C}}^p_\omega([0,T], \mathbb{R}^A), \qquad \widetilde{\mathscr{C}}^p_\omega([0,T], \mathbb{R}^A) \tag{4.45}
$$

the set of consistent bracket extensions and the set of consistent simple bracket extensions (i.e. in which the defining relation is only required with ℓ a product of single vertices) of their restriction to elements of $\mathscr{C}^p_\omega([0,T], \mathbb{R}^A)$. These sets are respectively contained in $\mathscr{C}^p_\omega([0,T], \mathbb{R}^{\widehat{A}})$ and $\mathscr{C}^p_\omega([0,T], \mathbb{R}^{\widetilde{A}})$. It is not difficult to construct bracket extensions that violate consistency: indeed, given any consistent bracket extension of a (homogeneously graded) $[3,4) \ni p$-rough path, it is possible to generate an inconsistent one simply by adding non-trivial paths of bounded $p/3$-variation to the terms indexed with trees $[\bullet_{(ab)}]_c$. The condition must therefore be required. We will henceforth assume $\widehat{\boldsymbol{X}}$ ($\widetilde{\boldsymbol{X}}$) to be a consistent (simple) bracket extension of $\boldsymbol{X}$; while this will not be needed in Theorem 4.24, it will in the other main theorem of this section, Theorem 4.25, and in general is a desirable property.

Example 4.18 (Consistency for $3 \le p < 4$)**.** When $A = [d]$ and $3 \le p < 4$ the only requirement needed for a bracket extension to be consistent is

$$
\langle [\bullet_{(\alpha\beta)}]_\gamma, \widetilde{\boldsymbol{X}} \rangle = \langle [\bullet_\alpha \bullet_\beta]_\gamma - [[\bullet_\alpha]_\beta]_\gamma - [[\bullet_\beta]_\alpha]_\gamma, \boldsymbol{X} \rangle \tag{4.46}
$$

If we begin with an arbitrary bracket extension, it is possible to simply replace the RHS above with the left. Indeed, it is easily checked that the Chen identity holds for $\langle [\bullet_{(\alpha\beta)}]_\gamma, \boldsymbol{X} \rangle$, and this is the only term that needs checking because there are no higher-order ones that could be affected by the substitution. For general p this is no longer true, and an existence theorem similar to [Kel12, Proposition 5.3.9] should be proved. Such a result would be, however, of little practical importance, since in explicit examples one would expect consistency to follow automatically. This will be the case for the rough path defined in Chapter 6.

Let $\boldsymbol{H} \in \mathscr{D}_{\boldsymbol{X}}(\mathbb{R}^e)$ and assume we want to postulate the rough path term $\langle [\bullet_i \bullet_j]_k, \boldsymbol{H}_{st} \rangle$ for some $i, j, k \in \{1, \ldots, e\}$. The idea is to define it by expanding H using (4.20). Proceeding formally, we set

$$
\begin{aligned}
\boldsymbol{H}^{[\bullet_i \bullet_j]_k}_{st} &:= \int_{s<u,v<w<t} \mathrm{d}H^i_u \mathrm{d}H^j_v \mathrm{d}H^k_w \\
&= \sum_{\substack{\ell, g, h \in \mathcal{F}^A \\ |\ell| + |g| + |h| \le \lfloor p \rfloor}} \frac{\boldsymbol{H}^i_{\ell;s} \boldsymbol{H}^j_{g;s} \boldsymbol{H}^k_{h;s}}{\mathcal{N}(\ell)\mathcal{N}(g)\mathcal{N}(h)} \int_{s<u,v<w<t} \mathrm{d}\boldsymbol{X}^\ell_{su} \mathrm{d}\boldsymbol{X}^g_{sv} \mathrm{d}\boldsymbol{X}^h_{sw}
\end{aligned}
$$

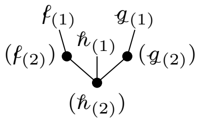
We now need to simplify the terms such as $\mathrm{d}\boldsymbol{X}^\ell_u$: if $\ell = [h]_a$ is a tree, (4.16) suggests substituting $\mathrm{d}\boldsymbol{X}^\ell_{su} = \boldsymbol{X}^h_{su} \mathrm{d}X^a_u$. If ℓ is not a tree we can use the bracket extension to express $\boldsymbol{X}^\ell$ as a sum of terms of the form $\widehat{\boldsymbol{X}}^{t}$ with $t \in \widehat{\mathcal{T}}^{A}$, using (4.39). We can therefore perform the substitution

$$\mathrm{d}\boldsymbol{X}^{f}_{su} = \mathrm{d}\sum_{(f)_{\mathrm{CK}}} \langle [f_{(1)}]_{(f_{(2)})}, \widehat{\boldsymbol{X}}_{su}\rangle = \sum_{(f)_{\mathrm{CK}}} \boldsymbol{X}^{f_{(1)}}_{su}\,\mathrm{d}\widehat{X}^{(f_{(2)})}_{u}$$

where, as usual, $\boldsymbol{X}^{(t)} = 0$ if $t \in \mathcal{T}^{A}$ is not a single vertex. Doing the same for g, h we are can conclude the calculation above as

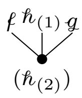
$$\boldsymbol{H}_{st} = \sum_{\substack{f,g,h\in\mathcal{F}^{A}\\ |f|+|g|+|h|\le\lfloor p\rfloor\\ (f)_{\mathrm{CK}},(g)_{\mathrm{CK}},(h)_{\mathrm{CK}}}} \frac{\boldsymbol{H}^{i}_{f;s}\boldsymbol{H}^{j}_{g;s}\boldsymbol{H}^{k}_{h;s}}{\mathcal{N}(f)\mathcal{N}(g)\mathcal{N}(h)} \int_{s<u,v<w<t} \boldsymbol{X}^{f_{(1)}}_{su}\boldsymbol{X}^{g_{(1)}}_{sv}\boldsymbol{X}^{h_{(1)}}_{sw}\,\mathrm{d}\widehat{X}^{(f_{(2)})}_{u}\,\mathrm{d}\widehat{X}^{(g_{(2)})}_{v}\,\mathrm{d}\widehat{X}^{(h_{(2)})}_{w}$$

$$= \sum_{\substack{f,g,h\in\mathcal{F}^{A}\\ |f|+|g|+|h|\le\lfloor p\rfloor\\ (f)_{\mathrm{CK}},(g)_{\mathrm{CK}},(h)_{\mathrm{CK}}}} \frac{\boldsymbol{H}^{i}_{f;s}\boldsymbol{H}^{j}_{g;s}\boldsymbol{H}^{k}_{h;s}}{\mathcal{N}(f)\mathcal{N}(g)\mathcal{N}(h)} \langle \quad , \widehat{\boldsymbol{X}}_{st}\rangle$$



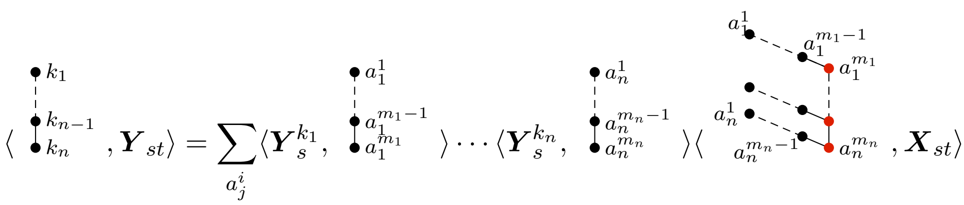
Where the tree in the last expression is constructed by joining the roots of each of the forests $f_{(1)}, g_{(1)}, h_{(1)}$ to the vertices below each, and we are only summing over terms in which all of $f_{(2)}, g_{(2)}, h_{(2)}$ are proper forests or single labels in A, i.e. in which they are labels in the alphabet $\widehat{A}$. Note how, by bracket consistency, we can replace this expression with

$$\sum_{\substack{f,g,h\in\mathcal{F}^{A}\\ |f|+|g|+|h|\le\lfloor p\rfloor\\ (h)_{\mathrm{CK}}}} \frac{\boldsymbol{H}^{i}_{f;s}\boldsymbol{H}^{j}_{g;s}\boldsymbol{H}^{k}_{h;s}}{\mathcal{N}(f)\mathcal{N}(g)\mathcal{N}(h)} \langle \quad , \widehat{\boldsymbol{X}}_{st}\rangle \tag{4.47}$$



The use of the bracket labels $(h_{(2)})$ cannot avoided: this is because the vertex is not a leaf.

Remark 4.19 (Ordered shuffle). It is instructive to see how this construction specifies to the ordered shuffle when $\boldsymbol{X}$ is geometric rough path: in this case we only need to sum over ladder trees, and the above formula becomes



and the ordered shuffle emerges by applying integration by parts, i.e. summing over all possible ways of collapsing the last tree onto linear trees in ways that maintain the ordering (this is the map ϕ_g defined in [HK15, (4.9)], named ϕ in (4.73) below): in this case this means respecting the ordering of each diagonal segment and the vertical segment of red vertices, which corresponds to the ordered indices in the ordered shuffle.

Remark 4.20. The procedure sketched above (and more precisely Definition 4.21 below) continues to work without modification when $\boldsymbol{H}$ is an $\widehat{\boldsymbol{X}}$-controlled path s.t. $\boldsymbol{H}_{f} = 0$ for $f \in \widehat{\mathcal{F}}^{A} \setminus (\widehat{\mathcal{T}}^{A} \cup \mathcal{F}^{A})$. This means, for example, that $\widehat{\boldsymbol{X}}$-driven RDE solutions can be lifted (note that $\varnothing \in \mathcal{F}^A$, so we are allowing $\boldsymbol{H}_\varnothing \neq 0$). Another example of this idea will be used when constructing bracket extensions of the lift.

However, when $\boldsymbol{H}_{f} \neq 0$ for some proper forest with at least one label in $\widehat{A} \setminus A$, or even $\widetilde{A} \setminus A$, we do not see a way in which the lift can be performed with $\widehat{\boldsymbol{X}}$ alone. An example would be when $\boldsymbol{H}$ is given by a smooth function of $\widehat{X}$, or even of $\widetilde{X}$. What would be needed in this case is a bracket extension of a bracket extension of $\boldsymbol{X}$, and there is no reason why the data of such a rough path, whose terms are indexed by "forests labelled with forests labelled with A-labelled forests" should be contained in $\widehat{\boldsymbol{X}}$. For instance, the following identity

$$\widetilde{\boldsymbol{X}}^{\bullet_\gamma \bullet_{(\alpha\beta)}}_{st} = \Big\langle\, [\gamma]_{(\alpha\beta)} + [(\alpha\beta)]_{\gamma} + \bullet^{(\alpha\beta\gamma)} - \big(\bullet^{(\gamma\,[\beta]_\alpha)} + \bullet^{(\gamma\,[\alpha]_\beta)} \big), \widehat{\boldsymbol{X}} \Big\rangle \tag{4.48}$$

which is easy to show directly using (4.43), shows how "bracketing" cannot be considered an associative operation (though it is for quasi-geometric rough paths, discussed in Section 4.3 below). Similar examples with more indices suggest that even $\widehat{\boldsymbol{X}}$ is not sufficient to express evaluations of $\widetilde{\boldsymbol{X}}$ against forests in terms of ones against trees.

In order to obtain a rough path that is rich enough to define lifts of all of its controlled paths, one would have to iterate the bracket extension only a finite number of times, after which further bracket extensions would be negligible (this is because the minimum regularity of the new trace terms in each iteration always increases by one). This, however, is not needed in the applications we have in mind.

We are ready to define the lift construction precisely. This will take some work, and it is convenient to establish some notation. Denote the labelling of a forest g by $\nu \mapsto \ell_g(\nu)$. Given a forest g, forests h^ν and labels a^ν for each vertex $\nu \in g$ we denote $\ast\{g; (a^\nu, h^\nu)^\nu\}$ the forest constructed by performing the following for each vertex $\nu \in g$: (re)label it a^ν and then graft the forest h^ν onto it (i.e. connect the root of each tree in h^ν to ν by adding an edge). To make calculations more readable, we will omit the upper bound on collections of forests, with the understanding that our sums are finite since they only contain terms $\widehat{\boldsymbol{X}}^{f}_{st}$ with $|f| \leq \lfloor p \rfloor$. The following result should be compared with the recursive formula [Gub10, Remark 8.7], which however does not use the bracket extension, and cannot therefore apply to the case in which the controlled path being lifted does not vanish on proper forests.

Definition 4.21 (Branched lift of a controlled path)**.** Let $\widehat{\boldsymbol{X}} \in \widehat{\mathscr{C}}^p_\omega([0,T], \mathbb{R}^A)$ restricting to $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T], \mathbb{R}^A)$ and $\boldsymbol{H} \in \mathscr{D}_{\boldsymbol{X}}(\mathbb{R}^e)$ we define $\boldsymbol{H}^{\bullet^k} = H^k$ and for $t \in \mathcal{T}^e$ with $\#t \geq 2$

$$(\upharpoonleft_{\widehat{\boldsymbol{X}}} \boldsymbol{H})^t_{st} \coloneqq \sum_{\substack{\{f^\nu\}^{\nu \in t} \subseteq \mathcal{F}^A \setminus \{\varnothing\} \\ (f^\nu)_{\mathrm{CK}}}} \Big(\prod_{\nu \in t} \mathcal{N}(f^\nu)^{-1} \boldsymbol{H}^{\ell_t(\nu)}_{f^\nu;s} \Big) \langle \ast\{t; ((f^\nu_{(2)}), f^\nu_{(1)})^{\nu \in t}\}, \widehat{\boldsymbol{X}}_{st} \rangle \tag{4.49}$$

and extending to $\mathcal{F}^e$ with products. We define $\uparrow_{\widehat{\boldsymbol{X}}} \boldsymbol{H}$ to be the unique rough path close to $\upharpoonleft_{\widehat{\boldsymbol{X}}} \boldsymbol{H}$ (this requires Theorem 4.24 below).

The sum in (4.49) is taken over all A-labelled forests f^ν such that $\sum_{\nu \in t} |f^\nu| \leq \lfloor p \rfloor$, with ν a vertex of t, and additionally using Sweedler notation for each f^ν, i.e. summing over all admissible cuts of each. Also note that $(f^\nu_{(2)})$ is a label in $\widehat{A}$, while $f^\nu_{(1)}$ is an A-labelled forest, and we are only summing over terms for which $f^\nu_{(2)}$ is a single vertex or a non-trivial proper forest. Before proving the main result of this section we focus on a couple of special cases in which the full bracket extension is actually not needed; for an example in which it generally is, one can take the controlled path to be given by a smooth function of an RDE solution.

Example 4.22 (Lifts of rough integrals and RDEs)**.** When the controlled path is given by a rough integral against $\boldsymbol{X}$, and in particular an $\boldsymbol{X}$-driven RDE solution, its rough path lift only requires $\boldsymbol{X}$ and not its bracket extension; this is because the controlled path vanishes on proper forests. Therefore, the only f^ν's considered in the sum of (4.49) are trees, and the only cuts considered in the $(f^\nu)_{\mathrm{CK}}$'s are those that disconnect the root from the rest of the tree f^ν. The $\ast$ terms are then just given by growing $[e]$-labelled trees out of all vertices of t:

$$\Big(\int_s^t \boldsymbol{H} \mathrm{d}\boldsymbol{X}\Big)^t = \sum_{\substack{\{f^\nu\}^{\nu\in t} \subseteq \mathcal{F}^A \setminus \{\varnothing\} \\ a^\nu \in A}} \Big(\prod_{\nu \in t} \mathcal{N}(f^\nu)^{-1} \boldsymbol{H}^{\ell_t(\nu)}_{[f^\nu]_{a^\nu};s}\Big) \langle \ast\{t; (a^\nu, f^\nu)^{\nu\in t}\}, \widehat{\boldsymbol{X}}_{st} \rangle \tag{4.50}$$

The expression for the $\boldsymbol{X}$-controlled path of an RDE $\boldsymbol{Y}$ in terms of its trace (4.29) can be substituted in this formula. RDE lifts have been studied in a more quantitative manner for geometric rough paths [FV10b, §10.4] under the name *full RDE solutions*, by defining them via smooth approximation of the trace. While such technique permits one to sidestep the algebra in the geometric (finite-dimensional) case, this is not possible for non-geometric branched rough paths, whose terms cannot be realised as limits of Stieltjes integrals of regularisations of the underlying path.

Example 4.23 (Pushforwards)**.** When the controlled path is $f(X)$ (4.23) we call its rough path lift the *pushforward* $f_*\boldsymbol{X}$. It only depends on the simple bracket extension $\widetilde{\boldsymbol{X}}$: this is because the only forests f over which we need to sum in (4.49) are products of single vertices, and as a consequence all the trees $\ast$ are already indexed by letters in $\widetilde{A}$. Using the alternative controlled path of (4.16) results in the same definition of $f_*\boldsymbol{X}$: this identity actually holds at the level of almost rough paths, which in both cases is given by

$$\upharpoonleft_{\widetilde{\boldsymbol{X}}} f(X)^t_{st} = \sum_{\substack{\boldsymbol{\alpha}^\nu, \boldsymbol{\beta}^\nu \\ |\boldsymbol{\beta}^\nu|>0}} \Big(\prod_{\nu\in t} \frac{1}{|\boldsymbol{\alpha}^\nu|!|\boldsymbol{\beta}^\nu|!} \partial_{\boldsymbol{\alpha}^\nu\boldsymbol{\beta}^\nu} f^{\ell_t(\nu)}(X_s)\Big) \langle \widetilde{\boldsymbol{X}}_{st}, \ast\{t; ((\boldsymbol{\beta}^\nu), \bullet^{\boldsymbol{\alpha}^\nu})^{\nu\in t}\}\rangle \tag{4.51}$$

where we are summing over tuples $\boldsymbol{\alpha}^\nu$ and $\boldsymbol{\beta}^\nu$, one each for each $\nu \in t$, $\boldsymbol{\alpha}^\nu\boldsymbol{\beta}^\nu$ denotes their concatenation, and $\bullet^{\boldsymbol{\alpha}^\nu}$ denotes the forest $\bullet^{\alpha_1^\nu} \cdots \bullet^{\alpha_n^\nu}$ where $\boldsymbol{\alpha}^\nu = (\alpha_1^\nu, \ldots, \alpha_n^\nu)$. The presence of the factorials is due to the fact that we are summing over tuples, not forests, and that over $\boldsymbol{\alpha}^\nu$, $\boldsymbol{\beta}^\nu$ individually: this change of variable and factor uses an argument involving binomial coefficients explained in the proof of Theorem 4.25.

Theorem 4.24. *$\upharpoonleft_{\widehat{\boldsymbol{X}}}\boldsymbol{H}$ is almost multiplicative, and $\uparrow_{\widehat{\boldsymbol{X}}}\boldsymbol{H}$ therefore defines a branched rough path.*

Proof. This does not require bracket consistency. We write out the string of identities that prove the first claim, and then carefully comment on each one (this includes explaining the notation used). When t is a single vertex, the statement reduces to that of H being a path. For $t \in \mathcal{T}^A$ with $\#t \geq 2$ we have

$$\begin{aligned} &\sum_{(t)_{\mathrm{CK}}} (\upharpoonleft_{\widehat{\boldsymbol{X}}}\boldsymbol{H})^{t_{(1)}}_{su} (\upharpoonleft_{\widehat{\boldsymbol{X}}}\boldsymbol{H})^{t_{(2)}}_{ut} \\ \approx &\sum_{(t)_{\mathrm{CK}}, (g^\mu)_{\mathrm{CK}}, (h^\nu)_{\mathrm{CK}}} \Big(\prod_{\mu\in t_{(1)}} \mathcal{N}(g^\mu)^{-1} \boldsymbol{H}^{\ell_t(\mu)}_{g^\mu;s}\Big) \langle \ast\{t_{(1)}; ((g^\mu_{(2)}), g^\mu_{(1)})^\mu\}, \widehat{\boldsymbol{X}}_{su}\rangle \\ &\qquad \Big(\prod_{\nu\in t_{(2)}} \mathcal{N}(h^\nu)^{-1} \boldsymbol{H}^{\ell_t(\nu)}_{h^\nu;u}\Big) \langle \ast\{t_{(2)}; ((h^\nu_{(2)}), h^\nu_{(1)})^\nu\}, \widehat{\boldsymbol{X}}_{ut}\rangle \end{aligned} \tag{4.52}$$

$$\approx \sum_{(t)_{\mathrm{CK}},(g^\mu)_{\mathrm{CK}},(h^\nu)_{\mathrm{CK}}} \Big(\prod_{\mu \in t_{(1)}} \mathcal{N}(g^\mu)^{-1} \boldsymbol{H}^{\ell_t(\mu)}_{g^\mu;s} \Big) \langle \ast\{t_{(1)}; ((g^\mu_{(2)}), g^\mu_{(1)})^\mu\}, \widehat{\boldsymbol{X}}_{su} \rangle \tag{4.53}$$
$$\Big(\prod_{\nu \in t_{(2)}} \mathcal{N}(h^\nu)^{-1} \sum_{j^\nu;\, k^\nu \in j^\nu \frown h^\nu} \mathcal{N}(j^\nu)^{-1} \boldsymbol{H}^{\ell_t(\nu)}_{k^\nu;s} \boldsymbol{X}^{j^\nu}_{su} \Big)$$
$$\langle \ast\{t_{(2)}; ((h^\nu_{(2)}), h^\nu_{(1)})^\nu\}, \widehat{\boldsymbol{X}}_{ut} \rangle$$

$$= \sum_{(t)_{\mathrm{CK}},(g^\mu)_{\mathrm{CK}}(k^\nu)^3_{\mathrm{CK}}} \Big(\prod_{\mu \in t_{(1)}} \mathcal{N}(g^\mu)^{-1} \boldsymbol{H}^{\ell_t(\mu)}_{g^\mu;s} \Big) \langle \ast\{t_{(1)}; ((g^\mu_{(2)}), g^\mu_{(1)})^\mu\}, \widehat{\boldsymbol{X}}_{su} \rangle \tag{4.54}$$
$$\Big(\prod_{\nu \in t_{(2)}} \mathcal{N}(k^\nu)^{-1} \boldsymbol{H}^{\ell_t(\nu)}_{k^\nu;s} \boldsymbol{X}^{k^\nu_{(1)}}_{su} \Big) \langle \ast\{t_{(2)}; ((k^\nu_{(3)}), k^\nu_{(2)})^\nu\}, \widehat{\boldsymbol{X}}_{ut} \rangle$$

$$= \sum_{\substack{(f^\lambda)_{\mathrm{CK}} \\ C \in \mathrm{Cut}^*(t) \\ D^\lambda \in \mathrm{Cut}^*_C(f^\lambda_{(1)})}} \Big(\prod_{\lambda \in t} \mathcal{N}(f^\lambda)^{-1} \boldsymbol{H}^{\ell_t(\lambda)}_{f^\lambda;s} \Big) \langle \ast\{\underline{t}_C; ((f^\mu_{(2)}), f^\mu_{(1)})^\mu\} \prod_{\nu \in \overline{t}_C} \underline{f^\nu_{(1)}}_{D^\nu}, \widehat{\boldsymbol{X}}_{su} \rangle \tag{4.55}$$
$$\langle \ast\{\overline{t}_C; ((f^\nu_{(2)}), \overline{f^\nu_{(1)}}_{D^\nu})^\nu\}, \widehat{\boldsymbol{X}}_{ut} \rangle$$

$$= \sum_{(f^\lambda)_{\mathrm{CK}}} \Big(\prod_{\lambda \in t} \mathcal{N}(f^\lambda)^{-1} \boldsymbol{H}^{\ell_t(\lambda)}_{f^\lambda;s} \Big) \langle \Delta_{\mathrm{CK}} \ast\{t; ((f^\lambda_{(2)}), f^\lambda_{(1)})^\lambda\}, \widehat{\boldsymbol{X}}_{su} \otimes \widehat{\boldsymbol{X}}_{ut} \rangle \tag{4.56}$$

$$= \sum_{(f^\lambda)_{\mathrm{CK}}} \Big(\prod_{\lambda \in t} \mathcal{N}(f^\lambda)^{-1} \boldsymbol{H}^{\ell_t(\lambda)}_{f^\lambda;s} \Big) \langle \ast\{t; ((f^\lambda_{(2)}), f^\lambda_{(1)})^\lambda\}, \widehat{\boldsymbol{X}}_{st} \rangle \tag{4.57}$$

$$= \big(\textstyle\int_{\widehat{\boldsymbol{X}}} \boldsymbol{H} \big)^t_{st} \tag{4.58}$$

(4.52) Here we are summing not just over cuts but over the non-empty forests g^μ and h^ν, with μ ranging over the vertices of $t_{(1)}$ and ν over those of $t_{(2)}$. A similar comment holds for subsequent identities. This step consists of substituting the definition of $\big(\int_{\widehat{\boldsymbol{X}}} \boldsymbol{H}\big)$, with the caveat that if $t_{(1)}$ or $t_{(2)}$ have a single vertex we are instead expanding the trace of the controlled path H according to (4.19), and using the definition of bracket extension (4.39) to express evaluations of $\widehat{\boldsymbol{X}}$ against a forest. The identity does indeed hold approximately, since we are assuming t has at least two vertices: if in one summand, for one of the factors it only holds that $\approx_{\lfloor p \rfloor}$, the presence of a second factor (which has at least one order of regularity, since the $\ast$'s are defined by summing over non-empty forests) means that $\approx$ holds for the summand as a whole.

(4.53) In this step we are re-expanding each $\boldsymbol{H}^{\ell_t(\nu)}_{h^\nu;u}$ at s, again using (4.19). Once again, the $\approx$ holds thanks to the presence of the other factor.

(4.54) uses the following combinatorial fact: defining multisets

$$A := \{\{(k, j, h) \mid h, j \in \mathcal{F}^A,\ k \in j \frown h\}\}, \qquad B := \{\{(k, C) \mid k \in \mathcal{F}^A,\ C \in \mathrm{Cut}^*(k)\}\}$$

and the map

$$f \colon B \to A, \quad (k, C) \mapsto (k, \underline{k}_C, \overline{k}_C)$$

(recall the notation for cuts used in (4.2), and that $\#$ denotes cardinality) it holds that for $(k, j, h) \in A$

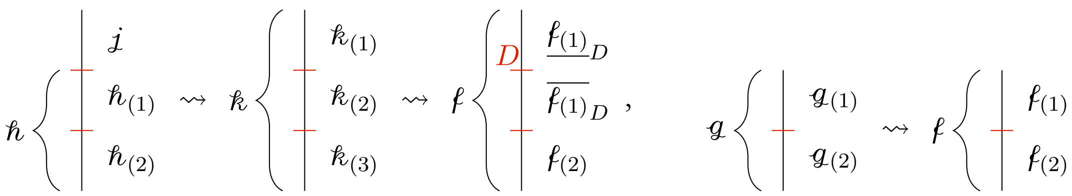
$$\begin{aligned} \# f^{-1}(k, j, h) &= \delta(\Delta_{\mathrm{CK}} k, j \otimes h) \\ &= \mathcal{N}(j)^{-1} \mathcal{N}(h)^{-1} \langle \Delta_{\mathrm{CK}} k, j \otimes h \rangle \end{aligned}$$

$$= \mathcal{N}(j)^{-1}\mathcal{N}(h)^{-1}\langle k, j \star h\rangle$$

$$= \frac{\mathcal{N}(k)}{\mathcal{N}(j)\mathcal{N}(h)}\delta(k, j \star h)$$

and $\delta(k, j\star h)$ is the number of times that $(k, j, h) \in A$. This means that when we go from summing over $\varnothing \neq h^\nu, j^\nu, k^\nu \in j^\nu \frown h^\nu$ to $k^\nu, (k^\nu)_{\mathrm{CK}}$ we must replace the factor $\mathcal{N}(j^\nu)\mathcal{N}(h^\nu)$ with $\mathcal{N}(k^\nu)$. The sum over each $\varnothing \neq h^\nu$, j^ν and $(h^\nu)_{\mathrm{CK}}$ becomes a sum over $\varnothing \neq k^\nu$, $(k^\nu)_{\mathrm{CK}}$ and $(k^\nu_{(2)})_{\mathrm{CK}}$, which by coassociativity we may more simply as one over $(k)^3_{\mathrm{CK}}$ (notation as in (4.7)). Since h was non-empty, we should disallow the total cut in the first of these coproducts; however, since this would result in $k_{(3)} = \varnothing$ (and thus $(k_{(3)})$ not being a valid label), the corresponding term would be null, so it is not incorrect to sum over $(k)^3_{\mathrm{CK}}$.

(4.55) Here λ ranges over all the vertices of t, and we write $f^\lambda = g^\lambda$ if $\lambda \in t_{(1)}$, $f^\lambda = h^\lambda$ if $\lambda \in t_{(2)}$. We have written out the coproduct on t more explicitly by summing over admissible cuts C. The thing to keep in mind with this substitution is that there was a double coproduct on k^ν but only an ordinary coproduct on g^μ. To reflect this when summing over the f^λ's, we additionally sum over $D^\lambda \in \mathrm{Cut}^*_C(f^\lambda_{(1)})$: this set denotes the set of admissible cuts of the forest $f^\lambda_{(1)}$ if C contains no edges below λ — i.e. $\lambda \in \overline{t}_C$ — and $\varnothing$ otherwise (which includes the case $C = \forall$) — i.e. $\lambda \in \underline{t}_C$. This means that D^λ is the trivial cut if the vertex λ has a cut below it — whereas it ranges over all admissible cuts of $f^\lambda_{(1)}$ when λ has no cut below it. The following diagram illustrates the changes of variable that have occurred in the last two steps; the vertical bars represents a forest cut in two or three places, and the μ, ν superscripts are omitted for brevity.

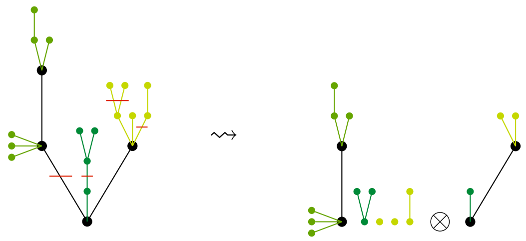


Note that we have also moved the term that previously was $\boldsymbol{X}_{su}^{k^\nu_{(1)}}$ into the first factor by including the product $\prod_{\nu\in\overline{t}_C} \underline{f^\nu_{(1)}}_{D^\nu}$ in the angle bracket.

(4.56) is based on the following fact regarding admissible cuts of a $\ast$: for labels a^λ and forests b^λ, $\mathrm{Cut}^*(\ast\{t; (a^\lambda, b^\lambda)^\lambda\})$ is in one-to-one correspondence with $\bigcup_{C\in\mathrm{Cut}^*(t)}(C \cup \bigsqcup_{\lambda\in t}\mathrm{Cut}^*_C(b^\lambda))$. Therefore

$$\begin{aligned}
&\Delta_{\mathrm{CK}} \ast\{t; (a^\lambda, b^\lambda)^\lambda\} \\
&= \sum_{\substack{C\in\mathrm{Cut}^*(t)\\ D^\lambda\in\mathrm{Cut}^*_C(b^\lambda)}} \underline{\ast}\{t; (a^\lambda, b^\lambda)^\lambda\}_{C\cup\bigcup_\lambda D^\lambda} \otimes \overline{\ast}\{t; (a^\lambda, b^\lambda)^\lambda\}_{C\cup\bigcup_\lambda D^\lambda} \\
&= \sum_{\substack{C\in\mathrm{Cut}^*(t)\\ D^\lambda\in\mathrm{Cut}^*_C(b^\lambda)}} \Big(\ast\{\underline{t}_C; (a^\mu, b^\mu)^\mu\}\prod_{\nu\in\overline{t}_C} \underline{b}^\nu_{D^\nu}\Big) \otimes \ast\{\overline{t}_C; (a^\nu, \overline{b}^\nu_{D^\nu})^\nu\}
\end{aligned}$$

Here is an example of a term in this calculation, in which t is black, the cut is red, and the f^λ are different shades of green to better distinguish them:



(4.57) is multiplicativity of $\widehat{\boldsymbol{X}}$ and finally,

(4.58) is the definition, keeping in mind once again that $\#t \geq 2$.

As for the second claim, regularity and the fact that $\upharpoonleft_{\widehat{\boldsymbol{X}}} \boldsymbol{H}$ is defined on $\mathcal{F}^e$ by extending with products, making it grouplike, allows us to apply Proposition 4.7 to conclude the proof. ■

It would be nice to define a canonical bracket extension for $\upharpoonleft_{\widehat{\boldsymbol{X}}} \boldsymbol{H}$ using $\widehat{\boldsymbol{X}}$, but the description of such a rough path appears to be quite complicated in general. However, in the special case of pushforwards — the only case that will be needed later on — the simple bracket extension admits a concise integral representation, defined in terms of $\widetilde{\boldsymbol{X}}$ alone. From now until the end of this chapter we will take $A = [d]$. Note how bracket consistency is crucial for the following result.

Theorem 4.25 (Simple bracket extension of pushforwards)**.** *Let $f \in C^\infty(\mathbb{R}^d, \mathbb{R}^e)$, then we can define a simple bracket extension of $f_* \boldsymbol{X}$ by*

$$f_* \widetilde{\boldsymbol{X}}^{(k_1 \cdots k_m)} = \sum_{|\gamma^1|, \ldots, |\gamma^m| > 0} \frac{1}{|\gamma^1|! \cdots |\gamma^m|!} \int \partial_{\gamma^m} f^{k_1} \cdots \partial_{\gamma^m} f^{k_m}(X) \mathrm{d} \widetilde{\boldsymbol{X}}^{(\gamma^1 \ldots \gamma^m)} \tag{4.59}$$

where we are summing over tuples (with $|\gamma^1| + \cdots + |\gamma^m| \leq \lfloor p \rfloor$), $\gamma^1 \ldots \gamma^m$ denotes their concatenation, and the lift of the integral is performed according to Example 4.22.

Proof. That $f_* \widetilde{\boldsymbol{X}}$ agrees with $f_* \boldsymbol{X}$ on $\mathcal{F}^d$ has already been checked in Example 4.23. Also note that, even though the integral in (4.59) is $\widetilde{\boldsymbol{X}}$-controlled, it can be lifted using $\widetilde{\boldsymbol{X}}$ alone, since it is an integral — no bracket terms in excess of those already contained in $\widetilde{\boldsymbol{X}}$ are needed. What remains to be shown is that

$$\langle \ll k_1 \cdots k_m \gg \frown_\nu t, f_* \widetilde{\boldsymbol{X}} \rangle = \langle \bullet^{(k_1 \cdots k_m)} \frown_\nu t, f_* \widetilde{\boldsymbol{X}} \rangle$$

Since the same proof works independently of the position of the vertex ν, we will prove it only at the ground level, i.e.

$$\langle \ll k_1 \cdots k_m \gg, f_* \widetilde{\boldsymbol{X}} \rangle = f_* \widetilde{\boldsymbol{X}}^{(k_1 \cdots k_m)}$$

First of all, we write the controlled path components: the only non-zero components of (4.59) (recall the notation $\bullet^{\boldsymbol{\alpha}} := \bullet^{\alpha_1} \cdots \bullet^{\alpha_n}$ for $\boldsymbol{\alpha} = (\alpha_1 \cdots \alpha_n)$, and note that this is distinct from $\bullet^{(\boldsymbol{\alpha})} := \bullet^{(\alpha_1 \cdots \alpha_n)}$)

$$\begin{aligned}
&\Bigg(\sum_{|\gamma^1|,\ldots,|\gamma^m|>0} \frac{1}{|\gamma^1|!\cdots|\gamma^m|!} \int \partial_{\gamma^m} f^{k_1} \cdots \partial_{\gamma^m} f^{k_m}(X) \mathrm{d}\widetilde{\boldsymbol{X}}^{(\gamma^1 \ldots \gamma^m)} \Bigg)_{[\,\overset{\boldsymbol{\alpha}}{\bullet}\,]_{(\boldsymbol{\beta})}} \\
&= \sum_{\substack{(\boldsymbol{\beta}^1 \ldots \boldsymbol{\beta}^m) = (\boldsymbol{\beta}) \\ |\boldsymbol{\beta}^1|,\ldots,|\boldsymbol{\beta}^m|>0}} \frac{1}{|\boldsymbol{\beta}^1|!\cdots|\boldsymbol{\beta}^m|!} \partial_{\boldsymbol{\beta}^1} f^{k_1} \cdots \partial_{\boldsymbol{\beta}^m} f^{k_m}(X)_{\overset{\boldsymbol{\alpha}}{\bullet}} \\
&= \sum_{\substack{(\boldsymbol{\beta}^1 \ldots \boldsymbol{\beta}^m) = (\boldsymbol{\beta}) \\ |\boldsymbol{\beta}^1|,\ldots,|\boldsymbol{\beta}^m|>0 \\ (\boldsymbol{\alpha}^1,\ldots,\boldsymbol{\alpha}^m) \in \mathrm{Sh}^{-1}(\boldsymbol{\alpha})}} \frac{1}{|\boldsymbol{\beta}^1|!\cdots|\boldsymbol{\beta}^m|!} \partial_{\boldsymbol{\alpha}^1\boldsymbol{\beta}^1} f^{k_1} \cdots \partial_{\boldsymbol{\alpha}^m\boldsymbol{\beta}^m} f^{k_m}(X)
\end{aligned} \tag{4.60}$$

where recall that there is a sum on non-empty tuples γ^l in the integral on the left. In the last sum, $\boldsymbol{\beta}^1, \ldots, \boldsymbol{\beta}^m$ range over all tuples s.t. the multiset given by their concatenation coincides with the multiset defined by $\boldsymbol{\beta}$, and the notation for unshuffles is explained in Chapter 3 (and note that the Sh^{-1}'s should be considered multisets, e.g. $(1,1)$ appears twice in $\mathrm{Sh}^{-1}(1,1)$). In the following calculation, on which we comment below, we omit evaluations of functions at (X_s) and the subscripts $_{st}$ to the rough paths

$$\begin{aligned}
&\langle \ll k_1 \cdots k_m \gg, f_* \widetilde{\boldsymbol{X}} \rangle \\
&= \langle \bullet^{\boldsymbol{k}} - \sum_{\substack{(\boldsymbol{i},\boldsymbol{j}) \in \mathrm{Sh}^{-1}\boldsymbol{k} \\ |\boldsymbol{i}|,|\boldsymbol{j}|>0}} [\bullet^{\boldsymbol{i}}\,]_{(\boldsymbol{j})}, f_* \widetilde{\boldsymbol{X}} \rangle && (4.61) \\
&= \sum_{|\gamma^1|,\ldots,|\gamma^m|>0} \frac{1}{|\gamma^1|!\cdots|\gamma^m|!} \langle \bullet^{\gamma^1 \ldots \gamma^m}, \boldsymbol{X} \rangle \\
&\quad - \sum_{\substack{(\boldsymbol{i},\boldsymbol{j}) \in \mathrm{Sh}^{-1}\boldsymbol{k} \\ |\boldsymbol{i}|,|\boldsymbol{j}|>0 \\ |\gamma^1|,\ldots,|\gamma^h|>0 \\ \boldsymbol{\delta};|\boldsymbol{\varepsilon}|>0 \\ (\boldsymbol{\delta}^1 \ldots \boldsymbol{\delta}^{m-h}) = (\boldsymbol{\delta}) \\ |\boldsymbol{\delta}^1|,\ldots,|\boldsymbol{\delta}^{m-h}|>0 \\ (\boldsymbol{\varepsilon}^1,\ldots,\boldsymbol{\varepsilon}^{m-h}) \in \mathrm{Sh}^{-1}(\boldsymbol{\varepsilon})}} \frac{1}{|\gamma^1|!\cdots|\gamma^h|!|\boldsymbol{\varepsilon}|!} \partial_{\gamma^1} f^{i_1} \cdots \partial_{\gamma^h} f^{i_h} \\
&\qquad \cdot \frac{1}{|\boldsymbol{\delta}^1|!\cdots|\boldsymbol{\delta}^{m-h}|!} \partial_{\boldsymbol{\delta}^1\boldsymbol{\varepsilon}^1} f^{j_1} \cdots \partial_{\boldsymbol{\delta}^{m-h}\boldsymbol{\varepsilon}^{m-h}} f^{j_{m-h}} \langle [\gamma^1 \ldots \gamma^h \boldsymbol{\varepsilon}]_{(\boldsymbol{\delta})}, \widetilde{\boldsymbol{X}} \rangle && (4.62) \\
&= \sum_{|\gamma^1|,\ldots,|\gamma^m|>0} \frac{1}{|\gamma^1|!\cdots|\gamma^m|!} \langle \bullet^{\gamma^1 \ldots \gamma^m}, \boldsymbol{X} \rangle \\
&\quad - \sum_{\substack{\boldsymbol{\alpha}^1,\ldots,\boldsymbol{\alpha}^m \\ \boldsymbol{\beta}^1,\ldots,\boldsymbol{\beta}^m \\ |\boldsymbol{\alpha}|,|\boldsymbol{\beta}|>0 \\ \exists l\ |\boldsymbol{\beta}^l|=0}} \frac{1}{|\boldsymbol{\alpha}^1|!|\boldsymbol{\beta}^1|!\cdots|\boldsymbol{\alpha}^m|!|\boldsymbol{\beta}^m|!} \partial_{\boldsymbol{\alpha}^1\boldsymbol{\beta}^1} f^{k_1} \cdots \partial_{\boldsymbol{\alpha}^m\boldsymbol{\beta}^m} f^{k_m} \langle [\bullet^{\boldsymbol{\alpha}}\,]_{(\boldsymbol{\beta})}, \widetilde{\boldsymbol{X}} \rangle && (4.63) \\
&= \sum_{|\gamma^1|,\ldots,|\gamma^m|>0} \frac{1}{|\gamma^1|!\cdots|\gamma^m|!} \langle \bullet^{\gamma^1 \ldots \gamma^m}, \boldsymbol{X} \rangle \\
&\quad + \sum_{\substack{|\gamma^1|,\ldots,|\gamma^m|>0 \\ \boldsymbol{\delta}^1,\ldots,\boldsymbol{\delta}^m \\ |\boldsymbol{\delta}|>0}} \frac{1}{|\gamma^1|!|\boldsymbol{\delta}^1|!\cdots|\gamma^m|!|\boldsymbol{\delta}^m|!} \partial_{\gamma^1\boldsymbol{\delta}^1} f^{k_1} \cdots \partial_{\gamma^m\boldsymbol{\delta}^m} f^{k_m} \langle [\bullet^{\boldsymbol{\delta}}\,]_{(\gamma)}, \widetilde{\boldsymbol{X}} \rangle
\end{aligned}$$

$$- \sum_{\substack{\alpha^1,\dots,\alpha^m \\ \beta^1,\dots,\beta^m \\ |\gamma^1|,\dots,|\gamma^m|>0}} \frac{1}{|\alpha^1|!|\beta^1|!\cdots|\alpha^m|!|\beta^m|!} \partial_{\gamma^1} f^{k_1}\cdots\partial_{\gamma^m} f^{k_m} \langle [\bullet^{\alpha}\ ]_{(\beta)}, \widetilde{X}\rangle \tag{4.64}$$

$$= \sum_{|\gamma^1|,\dots,|\gamma^m|>0} \frac{1}{|\gamma^1|!\cdots|\gamma^m|!} \Bigg[\partial_{\gamma^1} f^{k_1}\cdots\partial_{\gamma^m} f^{k_m} \langle \bullet^{\gamma} - \sum_{\substack{(\alpha,\beta)\in \mathrm{Sh}^{-1}(\gamma) \\ |\alpha|,|\beta|>0}} [\bullet^{\alpha}\ ]_{(\beta)}, \widetilde{X}\rangle$$

$$+ \sum_{\substack{\delta^1,\dots,\delta^m \\ |\delta|>0}} \frac{1}{|\delta^1|!\cdots|\delta^m|!} \partial_{\gamma^1\delta^1} f^{k_1}\cdots\partial_{\gamma^m\delta^m} f^{k_m} \langle [\bullet^{\delta}\ ]_{(\gamma)}, \widetilde{X}\rangle \Bigg] \tag{4.65}$$

$$= \sum_{|\gamma^1|,\dots,|\gamma^m|>0} \frac{1}{|\gamma^1|!\cdots|\gamma^m|!} \int \partial_{\gamma^m} f^{k_1}\cdots\partial_{\gamma^m} f^{k_m} \mathrm{d}\widetilde{X}^{(\gamma^1\cdots\gamma^m)} \tag{4.66}$$

$$= f_*\widetilde{X}^{(k_1\cdots k_m)}$$

We begin from the end, going up:

(4.66) is the definition in the statement.

(4.65) Here we have expanded the integral: on the first line we have written the zero-th order term in its expansion, and the sum in the second line, in which δ is the concatenation $\delta^1 \dots \delta^m$, is the sum of it's Gubinelli derivatives. Note that the δ^l's are allowed to be empty, as long as their concatenation is not. The missing step is

$$\sum_{\substack{|\delta|>0 \\ (\delta^1,\dots,\delta^m)\in \mathrm{Sh}^{-1}(\delta)}} \frac{1}{|\delta|!} \partial_{\gamma^1\delta^1} f^{k_1}\cdots\partial_{\gamma^m\delta^m} f^{k_m}(X_s) \langle [\bullet^{\delta}\ ]_{(\gamma)}, \widetilde{X}\rangle$$

$$= \sum_{\substack{\delta^1,\dots,\delta^m \\ |\delta|>0}} \binom{|\delta|}{|\delta^1|,\dots,|\delta^m|} \frac{1}{|\delta|!} \partial_{\gamma^1\delta^1} f^{k_1}\cdots\partial_{\gamma^m\delta^m} f^{k_m}(X_s) \langle [\bullet^{\delta}\ ]_{(\gamma)}, \widetilde{X}\rangle$$

where we went from summing on δ to summing on the individual tuples δ^l: the multinomial coefficient is the cardinality of $\mathrm{Sh}(\delta^1,\dots,\delta^m)$.

(4.64) Here we have separated the two summands in the first term, the second of which has become the sum that appears on the third line, with a negative sign. Since each $(\alpha,\beta) \in \mathrm{Sh}^{-1}(\gamma)$ restricts to a $(\alpha^l,\beta^l) \in \mathrm{Sh}^{-1}(\gamma^l)$, we may argue as above (this time we need m binomial coefficients) and go from summing over the γ^l's to the α^l,β^l's, with the condition that $\gamma^l := \alpha^l\beta^l$ be non-empty.

(4.63) is a consequence of the following observation: in the sum on the second line in the previous step (4.65) the γ^l's are all non-empty, while this is not the case for the β^l's in the last sum. The sum on the second line of (4.63) is given by this difference.

The remaining identities are best understood by starting at the top and going down.

(4.61) Is the definition of bracket polynomial.

(4.62) In this sum we are setting $h := |i|$ and expanding out the controlled path terms, using bracket consistency on the leaves as described in (4.47), and using (4.60) on the root.

(4.63) In this sum $\boldsymbol{\alpha} := \boldsymbol{\alpha}^1 \ldots \boldsymbol{\alpha}^h$ replace $\boldsymbol{\gamma}^1, \ldots \boldsymbol{\gamma}^h$, and the sum over $\boldsymbol{\varepsilon}$ and $(\boldsymbol{\varepsilon}^1, \ldots, \boldsymbol{\varepsilon}^{m-h}) \in \mathrm{Sh}^{-1}(\boldsymbol{\varepsilon})$ is replaced with one over $\boldsymbol{\alpha}^{h+1}, \ldots, \boldsymbol{\alpha}^m$, using the usual trick involving binomial coefficients. The sum over $\boldsymbol{\delta}$ and $\boldsymbol{\delta}^1, \ldots, \boldsymbol{\delta}^{m-h}$ s.t. $(\boldsymbol{\delta}^1, \ldots, \boldsymbol{\delta}^{m-h}) = \boldsymbol{\delta}$ is equivalent to a sum on $\boldsymbol{\delta}^1, \ldots, \boldsymbol{\delta}^{m-h}$, since each arrangement of such tuples appears exactly once, and is replaced with a sum over $\boldsymbol{\beta}^h, \ldots, \boldsymbol{\beta}^m$. Under this correspondence of old and new tuples, $\boldsymbol{\beta}^1, \ldots, \boldsymbol{\beta}^h$ are all empty, and since $h = |\boldsymbol{i}|$ was always positive, the condition in the new sum is satisfied.

This concludes the proof. ■

In light of the above, we can push forward not only rough paths, but also simple bracket extensions:

Definition 4.26 (Pushforward of a simple bracket extension)**.** We will call $f_* \widetilde{\boldsymbol{X}}$ the *pushforward* of the simple bracket extension $\widetilde{\boldsymbol{X}}$.

Example 4.27 (Bracket purshforward for $3 \leq p < 4$)**.** In this case the second and third levels of (4.59) read (sums on single indices)

$$f_* \widetilde{\boldsymbol{X}}^{(ij)}_{st} = \int_s^t \Big[\partial_\alpha f^i \partial_\beta f^j (X) \mathrm{d} \widetilde{\boldsymbol{X}}^{(\alpha\beta)} + \frac{1}{2} (\partial f^i_{\alpha\gamma} \partial f^j_\beta (X) + \partial f^i_\alpha \partial f^j_{\beta\gamma}(X)) \Big] \mathrm{d} \widetilde{\boldsymbol{X}}^{(\alpha\beta\gamma)}$$
$$f_* \widetilde{\boldsymbol{X}}^{(ijk)}_{st} = \int_s^t \partial_\alpha f^i \partial_\beta f^j \partial_\gamma f^k (X) \mathrm{d} \widetilde{\boldsymbol{X}}^{(\alpha\beta\gamma)}$$

The next lemma guarantees that "differentials and integrals cancel out". It is not stated in the most general terms, only in a way that will be strictly required in Section 4.4. It is checked by expanding the $(f_* \widetilde{\boldsymbol{X}})_{st}$ terms in the integral on the LHS by using (4.59) and (4.23) to express Gubinelli derivatives of the integrands, arguing as for the geometric case Theorem 3.27.

Lemma 4.28 (Associativity of the integral of one-forms against pushforwards)**.** *Given smooth functions* $g_{k_1,\ldots,k_m}$ *and* f, $\widetilde{\boldsymbol{X}}$ *as in Theorem 4.25,*

$$\begin{aligned} &\int g_{k_1,\ldots,k_m}(f(X)) \mathrm{d}(f_* \widetilde{\boldsymbol{X}})^{(k_1 \cdots k_m)} \\ &= \sum_{|\gamma^1|,\ldots,|\gamma^m|>0} \frac{1}{|\gamma^1|! \cdots |\gamma^m|!} \int (g_{k_1,\ldots,k_m} \circ f) \partial_{\gamma^1} f^{k_1} \cdots \partial_{\gamma^m} f^{k_m}(X) \mathrm{d} \widetilde{\boldsymbol{X}}^{(\gamma^1 \cdots \gamma^m)} \end{aligned} \tag{4.67}$$

This holds, in particular, when g does not depend on Y.

4.3 A characterisation of quasi-geometric rough paths

While most of the literature on rough paths distinguishes between geometric and branched rough paths, there is an intermediate type that is general enough to include Itô integration, but defined on a Hopf algebra that is simpler to describe than the Connes-Kreimer one: the *quasi-shuffle algebra*, original to [Hof00]. Rough paths defined on the quasi shuffle algebra are called *quasi-geometric*. Although the topic has been known about for some time [HK13], it has only recently appeared in the literature [Bel20].

In this section we show how quasi-geometric rough paths can be characterised as consistent bracket extensions of branched rough paths whose non-simple terms vanish. Geometric rough paths admit a similar characterisation (alternative to the one provided in [HK15, §4.1]), namely, they coincide with those branched rough paths that admit a consistent bracket extension that is trivial. The main reason for which we are interested in quasi-geometric rough paths is that the change of variable formula for RDE solutions simplifies into one that is analogous to the change of variable formula for functions, which could make it possible to adapt the transfer principle of Section 4.4 to define RDEs on manifolds.

We begin with a very brief review of the quasi-shuffle algebra of [Hof00]. While these are usually defined w.r.t. a bracket, we will treat the most general case of the free bracket; any rough path that is defined on the quasi-shuffle algebra w.r.t. a particular bracket can be defined on the free quasi-shuffle algebra, with the only drawback that we may have some redundant coordinates (e.g. if X is a d-dimensional Brownian motion and we want to define a rough path by Itô integration we have to set $X^{(\alpha\beta)} = [X]^{\alpha\beta} = 0$ when $\alpha \neq \beta$ instead of just setting $(\alpha\beta) = 0$). The advantage is that we can speak of *the* quasi-shuffle algebra, without having to specify a bracket. In this section the core components of our rough path will have trace valued in $\mathbb{R}^d$. All of this means that our quasi-shuffle algebra $T(\widetilde{\mathbb{R}}^d)_{\widetilde{\sqcup\!\sqcup}}$ has as its underlying set the tensor algebra not over $\mathbb{R}^d$, but over $\widetilde{\mathbb{R}}^d := \mathbb{R}^{\widetilde{[d]}}$ (notation for $\widetilde{[d]}$ as in (4.1),(4.33)). Note that $\widetilde{[d]}$ is a countably infinite set despite $[d]$ being a finite one; this will not be an issue once we are dealing with rough paths, since they will be defined on the algebra truncated at some order, considering that the weighting on $\widetilde{[d]}$ is given by cardinality of multisets (counting repetitions). We will use round brackets to denote multisets, unless the multiset only has one element, in which case brackets will be omitted. Generators (i.e. elementary tensors) of $T(\widetilde{\mathbb{R}}^d)_{\widetilde{\sqcup\!\sqcup}}$ are words of multisets, e.g. for the following words of weight 9

$$\alpha(\alpha\beta)\gamma\delta\varepsilon(\zeta\zeta\eta) = \alpha(\beta\alpha)\gamma\delta\varepsilon(\zeta\eta\zeta) \neq \alpha(\beta\alpha)\gamma\varepsilon\delta(\zeta\eta\zeta), \qquad \alpha,\ldots,\eta \in [d]$$

The quasi-shuffle product is defined recursively by declaring the empty word to be the identity element for it, and for w, z words in the alphabet $\widetilde{[d]}$ and $a, b \in \widetilde{[d]}$

$$wa \,\widetilde{\sqcup\!\sqcup}\, zb = \{wa \,\widetilde{\sqcup\!\sqcup}\, z\}b + \{w \,\widetilde{\sqcup\!\sqcup}\, zb\}a + \{w \,\widetilde{\sqcup\!\sqcup}\, z\}(ab) \tag{4.68}$$

where braces are used to specify the order of operations (quasi-shuffle and concatenation). Here $(ab) := a \cup b$ as multisets, e.g. if $a = (\alpha\alpha\beta)$, $b = (\beta\beta\gamma)$ we have $(ab) = (\alpha\alpha\beta\beta\beta\gamma)$; the same notation will be used for n-fold unions. The shuffle product admits the following non-recursive expression [EFMPW15, p.9]:

$$a_1 \ldots a_m \,\widetilde{\sqcup\!\sqcup}\, a_{m+1} \ldots a_{m+n} := \sum_{\substack{m \vee n \leq k \leq m+n \\ f\colon [m+n] \twoheadrightarrow [k] \\ f|_{[1,m]}, f|_{[m+1,n]} \nearrow}} \Big(\bigcup_{i \in f^{-1}(1)} a_i \Big) \ldots \Big(\bigcup_{i \in f^{-1}(k)} a_i \Big), \quad a_i \in \widetilde{[d]} \tag{4.69}$$

where we are summing over all surjections from the set with $m+n$ elements to the set of k elements, k ranging from $m \vee n$ to $m+n$, and s.t. $f(1) < \ldots < f(m)$ and $f(m+1) < \ldots < f(m+n)$. An example is

$$\begin{aligned}
\alpha_1(\alpha_2\alpha_3) \,\widetilde{\sqcup\!\sqcup}\, \beta_1\beta_2 =& \alpha_1(\alpha_2\alpha_3)\beta_1\beta_2 + \alpha_1\beta_1(\alpha_2\alpha_3)\beta_2 + \beta_1\alpha_1(\alpha_2\alpha_3)\beta_2 + \alpha_1\beta_1\beta_2(\alpha_2\alpha_3) \\
&+ \beta_1\alpha_1\beta_2(\alpha_2\alpha_3) + \beta_1\beta_2\alpha_1(\alpha_2\alpha_3) \\
&+ \alpha_1\beta_1(\alpha_2\alpha_3\beta_2) + \beta_1\alpha_1(\alpha_2\alpha_3\beta_2) + \alpha_1(\alpha_2\alpha_3\beta_1)\beta_2 + \beta_1(\beta_2\alpha_1)(\alpha_2\alpha_3) \\
&+ (\alpha_1\beta_1)(\alpha_2\alpha_3)\beta_2 + (\alpha_1\beta_1)\beta_2(\alpha_2\alpha_3)
\end{aligned}$$

$$+ (\alpha_1\beta_1)(\alpha_2\alpha_3\beta_2)$$

where we have used indentation to separate the sum by cardinality of the codomain's surjection: 4 (shuffles), 3 and 2.

The coproduct on $T(\widetilde{\mathbb{R}}^d)_{\widetilde{\sqcup\!\sqcup}}$ is identical to the one for the shuffle algebra, i.e. deconcatenation $\Delta_\otimes$, and so are the unit and counit. These turns $T(\widetilde{\mathbb{R}}^d)_{\widetilde{\sqcup\!\sqcup}}$ into a Hopf algebra, whose antipode is described explicitly in [Hofoo, Theorem 3.2], in which the following fundamental fact is proved: the quasi shuffle algebra is isomorphic, as a graded Hopf algebra, to the shuffle algebra over $\widetilde{\mathbb{R}}^d$. The isomorphism, *Hoffman's exponential*, is given explicitly by

$$\begin{aligned} \exp\colon T(\widetilde{\mathbb{R}}^d)_{\sqcup\!\sqcup} &\xrightarrow{\cong} T(\widetilde{\mathbb{R}}^d)_{\widetilde{\sqcup\!\sqcup}} \\ \boldsymbol{a} &\mapsto \sum_{\boldsymbol{a}^1\ldots\boldsymbol{a}^m=\boldsymbol{a}} \frac{1}{(\#\boldsymbol{a}^1)!\cdots(\#\boldsymbol{a}^m)!}(\boldsymbol{a}^1)\ldots(\boldsymbol{a}^m) \end{aligned} \tag{4.70}$$

with inverse

$$\begin{aligned} \log\colon T(\widetilde{\mathbb{R}}^d)_{\widetilde{\sqcup\!\sqcup}} &\xrightarrow{\cong} T(\widetilde{\mathbb{R}}^d)_{\sqcup\!\sqcup} \\ \boldsymbol{a} &\mapsto \sum_{\boldsymbol{a}^1\ldots\boldsymbol{a}^m=\boldsymbol{a}} \frac{(-1)^{|\boldsymbol{a}|-m}}{\#\boldsymbol{a}^1\cdots\#\boldsymbol{a}^m}(\boldsymbol{a}^1)\ldots(\boldsymbol{a}^m) \end{aligned} \tag{4.71}$$

Here $\#$ denotes the number of letters in the word (whose letters belong to the alphabet $\widetilde{[d]}$) and in both cases we are summing over all possible ways of expressing $\boldsymbol{a}$ as a concatenation $\boldsymbol{a}^1\ldots\boldsymbol{a}^m=\boldsymbol{a}$ with m variable. For words of length ≤ 3 these read

$$\begin{aligned} \exp(a)=a,\quad \exp(ab)=ab+\frac{1}{2}(ab),\quad \exp(abc)=abc+\frac{1}{2}\Big[(ab)c+a(bc)\Big]+\frac{1}{6}(abc) \\ \log(a)=a,\quad \log(ab)=ab-\frac{1}{2}(ab),\quad \log(abc)=abc-\frac{1}{2}\Big[(ab)c+a(bc)\Big]+\frac{1}{3}(abc) \end{aligned} \tag{4.72}$$

Note that these maps are (at least in this thesis) unrelated to the exponential and logarithm mapping the Lie algebra of primitives to the Lie group of grouplike elements in a Hopf algebra.

For a self-contained treatment of quasi-geometric rough paths we refer to [Bel20]; we will not, for the sake of conciseness, give yet another rough path definition. After all, it has been shown that rough paths can be defined in a uniform manner on a large class of Hopf algebras, with the core theorems remaining true [CEFMMK20, §4.2]; quasi-geometric rough paths are just another instance of this principle, with the quasi-shuffle algebra playing the same role as that of the shuffle or of the Connes-Kreimer Hopf algebra. Also, it will be convenient to keep thinking of quasi-geometric rough paths as branched rough paths, so that we do not have to switch settings, but may simply make use of the simplifications that the quasi-shuffle algebra makes possible. How this is done will become clear after we give brief survey of the maps between shuffle, quasi-shuffle and nonplanar forest Hopf algebras that appear in the following commutative diagram of graded Hopf algebra morphisms:

$$\begin{array}{ccccccc} T(\mathbb{R}^{\mathcal{T}^d})_{\sqcup\!\sqcup} & \xleftarrow{\ \psi\ } & \mathcal{H}^d_{\mathrm{CK}} & \hookrightarrow & \widetilde{\mathcal{H}}^d_{\mathrm{CK}} & \xrightarrow[\cong]{\mathrm{Aexp}} & \widetilde{\mathcal{H}}^d_{\mathrm{CK}} \\ \downarrow & \swarrow_{\phi} & & & \downarrow \phi & & \downarrow \widetilde{\phi} \\ T(\mathbb{R}^d)_{\sqcup\!\sqcup} & & \hookrightarrow & & T(\widetilde{\mathbb{R}}^d)_{\sqcup\!\sqcup} & \xrightarrow[\cong]{\exp} & T(\widetilde{\mathbb{R}}^d)_{\widetilde{\sqcup\!\sqcup}} \end{array} \tag{4.73}$$

The epimorphism ϕ on the left is used in [HK15, §4.1] to show how geometric rough paths canonically define branched rough paths: namely, given a geometric rough path $\boldsymbol{Z}$ we can define a branched one $\boldsymbol{X} := \phi^*\boldsymbol{Z}$, i.e. $\boldsymbol{X}^{f} := \boldsymbol{Z}^{\phi(f)}$; this amounts to expressing the branched components using integration by parts. In the following we will be using the convenient notation $\boldsymbol{Z}^{\phi} := \phi^*\boldsymbol{Z}$ and similar. ϕ is characterised as the unique algebra morphism s.t.

$$\phi([f]_\gamma) = \phi(f)\gamma, \qquad f \in \mathcal{F}^d \tag{4.74}$$

where $\phi(f)\gamma$ denotes the word obtained by juxtaposing the word $\phi(f)$ and γ. The other map labelled ϕ is defined in the same way on the enlarged alphabet. Intuitively, these maps sum over all ways of collapsing a forest onto the vertical axis in ways that preserve the ordering, and then reading off the labels from top to bottom to obtain a word. There is also an inclusion (not drawn) $\iota\colon T(\mathbb{R}^d)_{\sqcup\!\sqcup} \hookrightarrow \mathcal{H}^d_{\mathrm{CK}}$ (and a similar one $T(\widetilde{\mathbb{R}}^d)_{\sqcup\!\sqcup} \hookrightarrow \widetilde{\mathcal{H}}^d_{\mathrm{CK}}$), which maps the word $\gamma_1 \cdots \gamma_n$ to the ladder tree with vertices labelled $\gamma_1, \ldots, \gamma_n$ from top to bottom; ι is a right inverse to ϕ and a coalgebra morphism but not an algebra one. ι is used to check weather a branched rough path comes from a geometric one, which occurs if and only if $\boldsymbol{X} = \boldsymbol{X}^{\iota\circ\phi}$. The Hopf algebra monomorphism ψ is used in [HK15, §4.2] for the following purpose: given an $\mathbb{R}^d$-valued branched rough path $\boldsymbol{X}$, it is of interest to define a geometric rough path $\overline{\boldsymbol{X}}$ on the larger space $\mathbb{R}^{\mathcal{T}^d}$ (this means that its trace is indexed by $[d]$-labelled trees) with the property that $\boldsymbol{X}$-driven RDEs can be equivalently expressed as $\overline{\boldsymbol{X}}$-driven ones. ψ is used to formulate the condition, namely $\overline{\boldsymbol{X}}^{\psi(f)} = \boldsymbol{X}^{f}$, that $\overline{\boldsymbol{X}}$ must satisfy for $\overline{\boldsymbol{X}}$ to contain the data encoded in $\boldsymbol{X}$. It is characterised as the unique algebra morphism s.t.

$$\psi(t) = t + \sum_{(\widetilde{t})_{\mathrm{CK}}} \psi(t_{(1)}) \otimes t_{(2)}, \qquad t \in \mathcal{T}^d \tag{4.75}$$

where we emphasise that the coproduct is reduced. This map cannot be used to actually define $\overline{\boldsymbol{X}}$, a task first achieved through a recursive procedure (similar to the one used for defining the bracket $\widehat{\boldsymbol{X}}$) with calls to the non-constructive Lyons-Victoir extension theorem; a constructive (but still non-canonical) one was defined in [TZ20]. Another method to obtain a constructive Itô-Stratonovich formula was identified in [BC19], where the authors use the surprising fact, proved independently in [Foi02a, Cha10], that the Grossman-Larson Hopf algebra is free, i.e. isomorphic to the tensor algebra, over some set of trees. The explicit description of a (non-unique) free $\star$-basis of $\mathcal{H}^d_{\mathrm{GL}}$ is not given in the aforecited articles; this task is dual to that of finding a basis of the Lie algebra $\mathcal{P}^d_{\mathrm{CK}}$. These problems appear to lie beyond the current state of the art in algebraic combinatorics; moreover, it is to be expected that such bases, and therefore the resulting definition of $\overline{\boldsymbol{X}}$, will be non-canonical. A detailed comparison of the various Itô-Stratonovich formulae in branched rough path theory is performed in [Bru20]. Onto the right side of the diagram, the map $\widetilde{\phi}$ is the unique algebra morphism satisfying the same condition as (4.74), but is distinct from ϕ in that $\widetilde{\phi}(fg) = \phi(f)\widetilde{\sqcup\!\sqcup}\phi(g)$ with the quasi-shuffle product. The Hopf algebra automorphism Aexp that makes the square commute is the *arborified exponential* described [BCEF20, Theorem 2]. The significance of Hoffman's exponential (and of its arborified version) for rough paths will be discussed below. Arrows that have not been mentioned are obvious inclusions/projections.

We now transition to the main focus of this section: describing those branched rough paths that are (quasi-)geometric, in terms of their brackets. To preliminarily identify those branched rough paths that come from a quasi-geometric one we can proceed as done for geometric ones, but using the map $\widetilde{\phi}$ of (4.73): namely, a branched rough path $\widetilde{\boldsymbol{X}} \in \mathscr{C}^p_\omega([0,T], \widetilde{\mathbb{R}}^d)$ is *quasi-geometric* if $\widetilde{\boldsymbol{X}} = \widetilde{\boldsymbol{X}}^{\iota \circ \widetilde{\phi}}$. This is equivalent to the definition of [Bel20], i.e. we can define it as a functional on the quasi-shuffle Hopf algebra as $\widetilde{\boldsymbol{Z}} \coloneqq \widetilde{\boldsymbol{X}}^\iota$, and it holds that $\widetilde{\boldsymbol{X}} = \widetilde{\boldsymbol{X}}^{\iota \circ \widetilde{\phi}} = \widetilde{\boldsymbol{Z}}^{\widetilde{\phi}}$ ($\widetilde{\boldsymbol{X}}$ is determined by $\widetilde{\boldsymbol{Z}}$ by a quasi-integration by parts rule). Indeed, $\boldsymbol{Z}$ is group-valued since, for words w, z

$$\widetilde{\boldsymbol{Z}}^w \widetilde{\boldsymbol{Z}}^z = \widetilde{\boldsymbol{X}}^{\iota(w)} \widetilde{\boldsymbol{X}}^{\iota(z)} = \widetilde{\boldsymbol{X}}^{\iota(w)\iota(z)} = \widetilde{\boldsymbol{Z}}^{\phi(\iota(w)\iota(z))} = \widetilde{\boldsymbol{Z}}^{\phi \circ \iota(w) \widetilde{\sqcup\!\sqcup} \phi \circ \iota(z)} = \widetilde{\boldsymbol{Z}}^{w \widetilde{\sqcup\!\sqcup} z}$$

as ϕ is a left inverse to ι. $\widetilde{\boldsymbol{Z}}$ is multiplicative because the Connes-Kreimer coproduct on ladder trees corresponds to deconcatenation of the corresponding word. The next result states that a branched rough path defined on $\widetilde{\mathbb{R}}^d$ is quasi-geometric if and only if it is closed under taking the simple bracket extension, and admits trivial non-simple bracket extension. The statement will explain these assertions. We will preliminarily need to consider forests indexed by multisets of the set $\widetilde{[d]}$, i.e. $\widetilde{\widetilde{[d]}}$; an example of such a label is $(\alpha(\beta\gamma))\varepsilon$. The set of forests indexed by such labels may be denoted $\widetilde{\mathcal{F}}^{\widetilde{[d]}}$. Also, recall the notation (4.45).

Theorem 4.29 (Characterisation of (quasi-)geometric rough paths)**.** *The following are equivalent:*

1. $\widetilde{\boldsymbol{X}} \in \mathscr{C}^p_\omega([0,T], \widetilde{\mathbb{R}}^d)$ *is quasi-geometric;*
2. $\widetilde{\boldsymbol{X}} \in \mathscr{C}^p_\omega([0,T], \widetilde{\mathbb{R}}^d)$ *defines an element of* $\widehat{\mathscr{C}}^p_\omega([0,T], \mathbb{R}^d)$ *in the following way:* $\widehat{\boldsymbol{X}}^f = 0$ *for all* f *that have at least one label in* $\widehat{\widetilde{[d]}} \setminus \widetilde{\widetilde{[d]}}$*, and the simple bracket extension is given by joining labels, i.e. a label of the form* $(a_1 \cdots a_n)$ *with* $a_k = (\alpha^1_k \cdots \alpha^{m_k}_k)$*,* $\alpha^i_j \in [d]$*, are set to* $(\alpha^1_1 \cdots \alpha^{m_1}_1 \cdots \cdots \alpha^1_n \cdots \alpha^{m_n}_n)$*. Performing such substitutions at all vertices of* $f \in \widetilde{\mathcal{F}}^{\widetilde{[d]}}$ *yields a forest in* $\widetilde{\mathcal{F}}^d$*, against which* $\widetilde{\boldsymbol{X}}$ *can be evaluated.*

Similarly, the following are equivalent:

1. $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T], \mathbb{R}^d)$ *is geometric;*
2. *Setting* $\widehat{\boldsymbol{X}}^f = 0$ *for* f *having at least one label in* $\widehat{[d]} \setminus [d]$ *defines an element of* $\widehat{\mathscr{C}}^p_\omega([0,T], \mathbb{R}^d)$.

Proof. We will only give a proof of the characterisation of quasi-geometric rough paths; the characterisation of geometric ones follows an analogous, and simplified, procedure. We begin with 1 ⇒ 2. We must show the bracket relations, which in this case read

$$\langle f \curvearrowright_\nu g, \widehat{\boldsymbol{X}} \rangle = \sum_{C \in \mathrm{Cut}^\bullet(f)} \langle [\underline{f}_C]_{(\overline{f}_C)} \curvearrowright_\nu g, \widehat{\boldsymbol{X}} \rangle, \qquad f, g \in \widetilde{\mathcal{F}}^d,\ \nu \in f \text{ (or } -)$$

where, letting $f = s_1 \cdots s_n$ with $s_k \in \widetilde{\mathcal{T}}^d$, $\mathrm{Cut}^\bullet(f)$ denotes the elements of $\mathrm{Cut}(f)$ with the property that $\overline{f}_C$ is a non-empty product of single vertices: these are characterised as restricting to each s_k as either the total cut or the cut disconnecting the root from the rest of the tree, with at least one cut of the latter type overall. Moreover, the label $(\overline{f}_C)$ is defined by the label-joining rule expressed in the statement. By quasi-geometricity, these relations can be written as

$$\langle \iota \circ \widetilde{\phi}(\not{f} \frown_\nu t), \widehat{\boldsymbol{X}} \rangle = \sum_{C \in \mathrm{Cut}^\bullet(\not{f})} \langle \iota \circ \widetilde{\phi}([\not{f}_C]_{(\bar{\not{f}}_C)} \frown_\nu t), \widehat{\boldsymbol{X}} \rangle \tag{4.76}$$

and since, by the fact that $\widetilde{\phi}$ is linear in $\cdot \frown_\nu t$, the left and right hand sides are identical expressions in $\widetilde{\phi}(\not{f})$ and $\widetilde{\phi}\big(\sum_{C \in \mathrm{Cut}^\bullet(\not{f})} [\not{f}_C]_{(\bar{\not{f}}_C)} \big)$ respectively, it will suffice to show the first identity in

$$\widetilde{\phi}(\not{f}) = \sum_{C \in \mathrm{Cut}^\bullet(\not{f})} \widetilde{\phi}(\not{f}_C)(\bar{\not{f}}_C) = \widetilde{\phi}\Big(\sum_{C \in \mathrm{Cut}^\bullet(\not{f})} [\not{f}_C]_{(\bar{\not{f}}_C)} \Big) \tag{4.77}$$

the second of which follows from the definition of $\widetilde{\phi}$. Letting $\not{f} = [\not{g}_1]_{a_1} \cdots [\not{g}_n]_{a_n}$, for $\not{g}_k \in \widetilde{\mathcal{F}}^d$ and $a_k \in \widetilde{[d]}$, we have

$$\begin{aligned}
&\widetilde{\phi}([\not{g}_1]_{a_1} \cdots [\not{g}_n]_{a_n}) \\
&= \widetilde{\phi}(\not{g}_1) a_1 \,\widetilde{\sqcup\!\sqcup}\, \cdots \,\widetilde{\sqcup\!\sqcup}\, \widetilde{\phi}(\not{g}_n) a_n \\
&= \sum_{\substack{\{i_1,\ldots,i_k\} \sqcup \{j_1,\ldots,j_{n-k}\} = [n] \\ k<n}} \{ \widetilde{\phi}(\not{g}_{i_1}) a_{i_1} \,\widetilde{\sqcup\!\sqcup}\, \cdots \,\widetilde{\sqcup\!\sqcup}\, \widetilde{\phi}(\not{g}_{i_k}) a_{i_k} \,\widetilde{\sqcup\!\sqcup}\, \widetilde{\phi}(\not{g}_{j_1}) \,\widetilde{\sqcup\!\sqcup}\, \cdots \,\widetilde{\sqcup\!\sqcup}\, \widetilde{\phi}(\not{g}_{j_{n-k}}) \} (a_{j_1} \cdots a_{j_{n-k}}) \\
&= \sum_{\substack{\{i_1,\ldots,i_k\} \sqcup \{j_1,\ldots,j_{n-k}\} = [n] \\ k<n}} \widetilde{\phi}([\not{g}_{i_1}]_{a_{i_1}} \cdots [\not{g}_{i_k}]_{a_{i_k}} \not{g}_{j_1} \cdots \not{g}_{j_{n-k}})(a_{j_1} \cdots a_{j_{n-k}})
\end{aligned} \tag{4.78}$$

which uses an n-factor version of the recursive definition of $\widetilde{\sqcup\!\sqcup}$ (4.68) (easily shown by induction): this is precisely the identity needed in (4.77), expressed in terms of $\mathrm{Cut}^\bullet$. That this bracket extension is a rough path descends directly from the fact that $\widetilde{\boldsymbol{X}}$ is, and from the general fact that extending a rough path to a new alphabet trivially also preserves the rough path properties.

We now prove 2 ⇒ 1. We show

$$\langle \not{f} \frown_\nu \not{g}, \widetilde{\boldsymbol{X}} \rangle = \langle \iota \circ \widetilde{\phi}(\not{f}) \frown_\nu \not{g}, \widetilde{\boldsymbol{X}} \rangle, \qquad \not{f}, \not{g} \in \widetilde{\mathcal{F}}^d,\ \nu \in \not{g} \text{ or } -$$

by induction on the height of $\not{f} \in \widetilde{\mathcal{F}}$, i.e. the maximum number of edges connecting a leaf and the root. The statement of quasi-geometricity can be recovered by taking $\not{g} = \varnothing, \nu = -$. For height 0, $\not{f}$ is a single vertex and the assertion is obvious. For the inductive step

$$\begin{aligned}
\langle \not{f} \frown_\nu \not{g}, \widetilde{\boldsymbol{X}} \rangle &= \Big\langle \sum_{C \in \mathrm{Cut}^\bullet(\not{f})} [\not{f}_C]_{(\bar{\not{f}}_C)} \frown_\nu \not{g}, \widetilde{\boldsymbol{X}} \Big\rangle \\
&= \Big\langle \sum_{C \in \mathrm{Cut}^\bullet(\not{f})} [\iota \circ \widetilde{\phi}(\not{f}_C)]_{(\bar{\not{f}}_C)} \frown_\nu \not{g}, \widetilde{\boldsymbol{X}} \Big\rangle \\
&= \Big\langle \sum_{C \in \mathrm{Cut}^\bullet(\not{f})} \iota \circ \widetilde{\phi}([\not{f}_C]_{(\bar{\not{f}}_C)}) \frown_\nu \not{g}, \widetilde{\boldsymbol{X}} \Big\rangle \\
&= \langle \iota \circ \widetilde{\phi}(\not{f}) \frown_\nu \not{g}, \widetilde{\boldsymbol{X}} \rangle
\end{aligned}$$

where we have used the consistent bracket relations (4.76), the inductive hypothesis, and in the last step (4.78). ■

Example 4.30 (Quasi-geometricity for $3 \le p < 4$)**.** When $3 \le p < 4$ the only obstruction to quasi geometricity of a consistent full bracket extension is its evaluations on the labels $\bullet_{\gamma}\,\mathbf{I}^{\beta}_{\alpha}$: these vanish if and only if $\widetilde{\boldsymbol{X}}$ is quasi-geometric, by (4.48) and the above theorem.

As a consequence we have the next change of variable formula. This statement restricted to the case in which the RDE is driven by $\boldsymbol{X}$ is stated in [HK13, p.25], whereas a self-contained proof of the change of variable formula for functions of X can be found in [Bel20]. We will be considering equations

$$\mathrm{d}Y = \sum_{\gamma} \frac{1}{|\gamma|!} F_{(\gamma)}(Y) \mathrm{d}\widetilde{\boldsymbol{X}}^{(\gamma)} \tag{4.79}$$

driven by the whole of $\widetilde{\boldsymbol{X}}$, but F only depends on the trace X (not that of the bracket extensions). Note that we are summing on tuples, not multisets, and the $|\gamma|!^{-1}$ included for convenience. This is a particular case of (4.27) with $A = \widetilde{[d]}$.

Corollary 4.31 (Change of variable formula for quasi-geometric RDEs)**.** *Let $\widetilde{\boldsymbol{X}} \in \mathscr{C}^p_\omega([0,T], \mathbb{R}^d)$ restricting to $\boldsymbol{X} \in \mathscr{C}^p_\omega([0,T], \mathbb{R}^d)$ be quasi-geometric, and Y be a solution to (4.79). For $g \in C^\infty \mathbb{R}^e$ we have*

$$g(Y)_{st} = \sum_{|\gamma^1|,\ldots,|\gamma^m|>0} \frac{1}{n!|\gamma^1|! \cdots |\gamma^n|!} \int_s^t \partial_{k_1,\ldots,k_n} g(Y) F^{r_1}_{(\gamma^1)} \cdots F^{r_n}_{(\gamma^n)}(Y) \mathrm{d}\widetilde{\boldsymbol{X}}^{(\gamma^1 \ldots \gamma^n)} \tag{4.80}$$

Proof. This is a straightforward application of [Theorem 4.29, 1.⇒2.] to (4.40), where in the latter we are taking A to be $\widetilde{[d]}$. ■

The last theoretical topic of this section is the Itô-Stratonovich correction for quasi-geometric RDEs. This is made possible by Hoffman's exponential, which yields an Itô-Stratonovich conversion formula that, unlike the ones valid for general branched rough paths, is canonical. Given a quasi-geometric (branched) rough path $\widetilde{\boldsymbol{X}} \in \mathscr{C}^p_\omega([0,T], \widetilde{\mathbb{R}}^d)$, recall that $\widetilde{\boldsymbol{X}}^\iota$ denotes its restriction to the quasi-shuffle algebra. Since exp (see (4.73)) is a Hopf algebra isomorphism, $\widetilde{\boldsymbol{X}}^{\iota \circ \exp}$ defines a geometric one, still on the enlarged alphabet $\widetilde{[d]}$. The arborified exponential Aexp can be used to obtain the branched components of $\widetilde{\boldsymbol{X}}^{\iota \circ \exp}$ from $\widetilde{\boldsymbol{X}}$, i.e. it holds that $\widetilde{\boldsymbol{X}}^{\mathrm{Aexp}} = \widetilde{\boldsymbol{X}}^{\iota \circ \exp \circ \phi}$ (despite $\mathrm{Aexp} \neq \iota \circ \exp \circ \phi$ — the latter is not even an algebra morphism). Hoffman's exponential can be used, as stated in [HK13, p.21], to write RDEs driven by a quasi-geometric rough path $\boldsymbol{X}$ in terms of the geometric one $\widetilde{\boldsymbol{X}}^{\exp}$. We emphasize that $\widetilde{\boldsymbol{X}}$ must already be defined on the extended alphabet, and in general the new RDE will depend on this extension regardless of whether the old one does. The following type of result is likely known to experts but is not, as far as we can tell, present in the literature.

Theorem 4.32 (Itô-Stratonovich for quasi-geometric RDEs)**.** *The RDE*

$$\mathrm{d}Y = F_a(Y) \mathrm{d}\widetilde{\boldsymbol{X}}^a$$

driven by a quasi-geometric rough path $\widetilde{\boldsymbol{X}} \in \mathscr{C}^p_\omega([0,T], \widetilde{\mathbb{R}}^d)$ is equivalent to

$$\mathrm{d}Y = \frac{(-1)^{\#\boldsymbol{a}-1}}{\#\boldsymbol{a}} F_{\boldsymbol{a}}(Y) \mathrm{d}\overline{\boldsymbol{X}}^{(\boldsymbol{a})} \tag{4.81}$$

driven by the geometric rough path $\overline{\boldsymbol{X}} := \widetilde{\boldsymbol{X}}^{\exp}$*. If the original RDE was driven by* $\boldsymbol{X}$ *alone, this becomes*

$$\mathrm{d}Y = \frac{(-1)^{|\boldsymbol{\gamma}|-1}}{|\boldsymbol{\gamma}|} F_{\boldsymbol{\gamma}}(Y)\mathrm{d}\overline{\boldsymbol{X}}^{(\boldsymbol{\gamma})}$$

A few comments: $\boldsymbol{a}$ denotes a word in the alphabet $\widetilde{[d]}$, while $\boldsymbol{\gamma}$ is one using letters of $[d]$ (for which $\#$ and $|\cdot|$ agree). $(\boldsymbol{a})$ is, as usual, defined by joining labels. $F_{\boldsymbol{a}}$ is defined by composition of vector fields, i.e. $F_{a_1\ldots a_n} = F_{a_1}\ldots F_{a_n}$; in particular, note that $F_{\boldsymbol{a}}$ and $F_{(\boldsymbol{a})}$ (which does not appear) are distinct.

Proof of Theorem 4.32. The Davie expansion of (4.81) is

$$\begin{aligned} Y_{st} &\approx \sum_{\#\boldsymbol{a}^1,\ldots,\#\boldsymbol{a}^n>0} \frac{(-1)^{\#\boldsymbol{a}^1+\ldots+\#\boldsymbol{a}^n-n}}{\#\boldsymbol{a}^1\cdots\#\boldsymbol{a}^n} F_{\boldsymbol{a}^1}\ldots F_{\boldsymbol{a}^n}(Y_s)\overline{\boldsymbol{X}}^{(\boldsymbol{a}_1)\ldots(\boldsymbol{a}_n)}_{st} \\ &= \sum_{\substack{\#\boldsymbol{a}\geq 1 \\ \boldsymbol{a}^1\ldots\boldsymbol{a}^n=\boldsymbol{a} \\ \#\boldsymbol{a}^1,\ldots,\#\boldsymbol{a}^n>0}} \frac{(-1)^{\#\boldsymbol{a}-1}}{\#\boldsymbol{a}^1\cdots\#\boldsymbol{a}^n} F_{\boldsymbol{a}}(Y_s)\overline{\boldsymbol{X}}^{(\boldsymbol{a}_1)\ldots(\boldsymbol{a}_n)}_{st} \\ &= F_{\boldsymbol{a}}(Y_s)\overline{\boldsymbol{X}}^{\log(\boldsymbol{a})}_{st} \\ &= F_{\boldsymbol{a}}(Y_s)\widetilde{\boldsymbol{X}}^{\boldsymbol{a}}_{st} \end{aligned}$$

where we have used associativity of composition of vector fields. This is the Davie expansion of original RDE, where $\widetilde{\boldsymbol{X}}$ is intended as being defined directly on $T(\widetilde{\mathbb{R}}^d)_{\widetilde{\sqcup\!\sqcup}}$. ■

Example 4.33 (Theorem 4.32 for $p < 4$)**.** When $p < 4$ the above theorem reads

$$\begin{aligned} &\mathrm{d}Y = F_\gamma(Y)\mathrm{d}\boldsymbol{X}^\gamma + \frac{1}{2}F_{(\alpha\beta)}(Y)\mathrm{d}\widetilde{\boldsymbol{X}}^{(\alpha\beta)} + \frac{1}{6}F_{(\alpha\beta\gamma)}(Y)\mathrm{d}\widetilde{\boldsymbol{X}}^{(\alpha\beta\gamma)} \\ \iff\quad &\mathrm{d}Y = F_\gamma(Y)\mathrm{d}\overline{\boldsymbol{X}}^\gamma + \frac{1}{2}(F_{(\alpha\beta)} - F_\alpha F_\beta)(Y)\mathrm{d}\overline{\boldsymbol{X}}^{(\alpha\beta)} \\ &\qquad + \Big[\frac{1}{3}F_\alpha F_\beta F_\gamma - \frac{1}{4}\big(F_{(\alpha\beta)}F_\gamma + F_\gamma F_{(\alpha\beta)}\big) + \frac{1}{6}F_{(\alpha\beta\gamma)}\Big](Y)\mathrm{d}\overline{\boldsymbol{X}}^{(\alpha\beta\gamma)} \end{aligned}$$

where the sums, unlike that of (4.81), is on single indices (this is the reason for the presence of the factor $1/4$). Note that we have included factors in the initial RDE, since the resulting formula appears more natural with them. Truncated at order 2, this reduces to the usual Itô-Stratonovich formula valid for semimartingales.

We end this section with a few stochastic examples. The next two are meant as a very brief overview of some interesting cases of random quasi-geometric rough paths.

Example 4.34 ($2 \leq p < 3$)**.** We mention this example for the sake of completeness, but do not delve into the details, which have been described elsewhere (e.g. [Bel20, §3.3]). While Young integration against paths of bounded $2 > p$-variation is vacuously geometric, every $[2,3) \ni p$-rough path is quasi-geometric, since the terms $X^{(\alpha\beta)}$ can be defined canonically in terms of the X^γ's as in (4.43), with no extra lifts involved; this is the *bracket* of [FH20, Definition 5.5]. Iterated Itô integrals and their relationship with Stratonovich ones fit into this case; these topics were treated prior to the introduction of quasi-shuffles by using Wick products

[Gai94, Gai95]. Iterated Itô and Stratonovich integrals can be obtained from one another by using the exp and log, and similarly their branched counterparts by using the arborified versions of exp and log.

Example 4.35 (Itô formulae for the 1-dimensional heat equation, [Bel19, Ch. 4])**.** The cited PhD thesis considers the solution of the 1-dimensional stochastic heat equation with additive noise

$$\begin{cases} \partial_t u = \frac{1}{2}\partial_{xx} u + \xi \\ u(0,x) = 0 \end{cases} \qquad X_t := u(t,x), \qquad x \in \mathbb{R}$$

although it is stated that the same techniques would work for many other Gaussian processes, such as fractional Brownian motion with Hurst parameter $1/4$. The author is then able to reproduce and shed new light on Itô-type formulae for X present in the literature, by defining three distinct $4 = p$-quasi-geometric rough paths above X, in addition to the geometric rough path defined canonically by powers of X thanks to unidimensionality. $\widetilde{\boldsymbol{X}}_{\mathrm{BS}}$ is a quasi-geometric rough path above with $\widetilde{X}^1_{\mathrm{BS}} = X$ and $\widetilde{X}^2_{\mathrm{BS}} = B$, where B is a certain Brownian motion independent of X. Then (4.41) becomes the Burdzy-Swanson formula

$$f(X)_{st} = \int_s^t f'(X)\mathrm{d}\boldsymbol{X}^1_{\mathrm{BS}} + \frac{\kappa}{2}\int_s^t f''(X)\mathrm{d}B$$

where the second is an Itô integral and κ is a certain deterministic constant. Similarly, a quasi-geometric rough path $\widetilde{\boldsymbol{X}}_{\mathrm{CN}}$ with $\widetilde{X}^1_{\mathrm{CN}} = X$ and $\widetilde{X}^2_{\mathrm{CN}} = \sigma^2 := \mathbb{E}[X^2]$ reproduces the Cheridito-Nualart formula

$$f(X)_{st} = \int_s^t f'(X)\mathrm{d}\boldsymbol{X}^1_{\mathrm{CN}} + \frac{1}{2}\int_s^t f''(X)\mathrm{d}\sigma^2$$

where the second integral is intended in the sense of Young. These two change of variable formulae only have terms of order ≤ 2 because the rough integrals of orders 3 and 4 cancel out, despite them not vanishing individually. There is a quasi-geometric rough path $\widetilde{\boldsymbol{X}}_{\mathrm{QV}}$ with $\widetilde{X}^1_{\mathrm{QV}} = X$ and $\widetilde{X}^2_{\mathrm{QV}} = 6t/\pi$ the quartic variation of X. The corresponding change of variable is a quartic variation formula

$$f(X)_{st} = \int_s^t \Big[f'(X)\mathrm{d}\boldsymbol{X}_{\mathrm{QV}} + \frac{1}{2}f''(X)\mathrm{d}\widetilde{\boldsymbol{X}}^{(11)}_{\mathrm{QV}} + \frac{1}{6}f^{(3)}(X)\mathrm{d}\widetilde{\boldsymbol{X}}^{(111)}_{\mathrm{QV}}\Big] + \frac{1}{4\pi}\int_s^t f^{(4)}(X_u)\mathrm{d}u$$

where the second and third integrals do not vanish even though $\widetilde{X}^{(11)} = 0 = \widetilde{X}^{(111)}$. This example demonstrates the rich variety of identities that quasi-geometric rough paths are able to generate in stochastic analysis.

Not all stochastic integration theories, however, are compatible with the algebra of quasi-shuffles, as demonstrated by the following sketched example.

Example 4.36 (Processes of finite cubic variation)**.** These stochastic processes were studied in [ER03]. The possible link with this type of process was already spotted in [Kel12, p.121]. In this example we wish to make a further observation on its potential compatibility with branched rough paths; while a full reconciliation of the two integration theories lies beyond the scope of this chapter, at the end of this example we list the steps needed to make the connection precise. An $\mathbb{R}^d$-valued process X defined on $[0,T]$ is said to be of *finite cubic variation* if for $\alpha, \beta, \gamma = 1, \ldots, d$

$$\text{UCP}\lim_{h \to 0} \frac{1}{h} \int_0^{\cdot} X^\alpha_{s,s+h} X^\beta_{s,s+h} X^\gamma_{s,s+h} \mathrm{d}s =: [X]^{\alpha\beta\gamma} \qquad \text{exists, and}$$

$$\sup_{h > 0} \frac{1}{h} \int_0^{\cdot} |X^\alpha_{s,s+h} X^\beta_{s,s+h} X^\gamma_{s,s+h}| \mathrm{d}s < \infty \qquad \text{a.s.}$$

where the limit is taken uniformly on compacts in probability; if these conditions are fulfilled $[X]$ is an $(\mathbb{R}^d)^{\otimes 3}$-valued process of bounded variation. The *symmetric integral* [RV93] of another stochastic process H against X is defined as

$$\int_0^{\cdot} H \mathrm{d}^\circ X \coloneqq \text{UCP}\lim_{h \to 0} \frac{1}{2h} \int_0^{\cdot} H_s X_{s-h,s+h} \mathrm{d}s$$

whenever the UCP limit exists; when X is a continuous semimartingale this is the Stratonovich integral, but it is defined in greater generality. It is then proven in [ER03, Remark 3.6] that for $f \in C^\infty \mathbb{R}^d$

$$f(X)_{st} = \int_s^t \partial_\gamma f(X) \mathrm{d}^\circ X^\gamma - \frac{1}{12} \int_s^t \partial_{\alpha\beta\gamma} f(X) \mathrm{d}[X]^{\alpha\beta\gamma} \tag{4.82}$$

where the second integral is Stieltjes. The following integration by parts formulae are also shown [ER03, p.277,281]

$$(X^\alpha X^\beta)_{st} = \int_s^t X^\alpha \mathrm{d}^\circ X^\beta + \int_s^t X^\beta \mathrm{d}^\circ X^\alpha \tag{4.83}$$

$$[X]^{\alpha\beta\gamma}_{st} = 2\Big(\int_s^t X^\gamma X^\beta \mathrm{d}^\circ X^\alpha + \int_s^t X^\gamma X^\alpha \mathrm{d}^\circ X^\beta - \int_s^t X^\gamma \mathrm{d}^\circ (X^\alpha X^\beta) \Big) \tag{4.84}$$

whenever all integrals exist.

We now perform some formal calculations aimed at computing the values of the bracket terms of a hypothetical rough path above a finite cubic variation process which is defined via symmetric integration using (4.16); we later comment on what our calculations mean and on what is needed to make them rigorous. (4.83) is an order-2 integration by parts formula, equivalent to

$$X^\alpha_{st} X^\beta_{st} = \int_s^t X^\alpha_{su} \mathrm{d}^\circ X^\beta_u + \int_s^t X^\beta_{su} \mathrm{d}^\circ X^\alpha_u, \quad \text{or} \quad \widetilde{X}^{(\alpha\beta)}_{st} = \boldsymbol{X}^{\ll \alpha\beta \gg}_{st} = 0$$

in branched rough path notation. (4.84) in rough path notation reads, setting $f(x) \coloneqq x^\alpha x^\beta$ and using associativity Lemma 4.28

$$\begin{aligned}
-\frac{[X]^{\alpha\beta\gamma}_{st}}{2} &= \int_s^t X^\gamma_{su} \mathrm{d} f_* \boldsymbol{X}_u - \int_s^t X^\gamma_{su} X^\alpha_u \mathrm{d}\boldsymbol{X}^\beta_u - \int_s^t X^\gamma_{su} X^\beta_u \mathrm{d}\boldsymbol{X}^\alpha_u \\
&= \int_s^t X^\gamma_u \mathrm{d}\widetilde{\boldsymbol{X}}^{(\alpha\beta)}_u \\
&= \int_s^t X^\gamma_{su} \mathrm{d}\widetilde{\boldsymbol{X}}^{(\alpha\beta)}_u \\
&= \widetilde{\boldsymbol{X}}^{\bullet^{\gamma} \!-\! \bullet_{(\alpha\beta)}}_{st}
\end{aligned}$$

where we have used that $\widetilde{X}^{\alpha\beta} = 0$. Now, comparing (4.41) with (4.82) for the function $f(x) = x^\alpha x^\beta x^\gamma$ we obtain

$$
\begin{aligned}
&\int_s^t X^\alpha X^\beta \mathrm{d}\boldsymbol{X}^\gamma + X^\alpha X^\gamma \mathrm{d}\boldsymbol{X}^\beta + X^\beta X^\gamma \mathrm{d}\boldsymbol{X}^\alpha + X^\alpha \mathrm{d}\boldsymbol{X}^{(\beta\gamma)} + X^\beta \mathrm{d}\boldsymbol{X}^{(\alpha\gamma)} + X^\gamma \mathrm{d}\boldsymbol{X}^{(\alpha\beta)} + \mathrm{d}\boldsymbol{X}^{\alpha\beta\gamma} \\
&= (X^\alpha X^\beta X^\gamma)_{st} \\
&= \int_s^t X^\alpha X^\beta \mathrm{d}^\circ X^\gamma + X^\alpha X^\gamma \mathrm{d}^\circ X^\beta + X^\beta X^\gamma \mathrm{d}^\circ X^\alpha - \frac{6}{12}[X]_{st}^{(\alpha\beta\gamma)}
\end{aligned}
$$

Cancelling corresponding rough/symmetric integrals, and using the expression for $\widetilde{\boldsymbol{X}}_{st}^{[\bullet_\gamma]_{(\alpha\beta)}}$, we obtain

$$
-\frac{3}{2}[X]_{st}^{\alpha\beta\gamma} + X_{st}^{(\alpha\beta\gamma)} = -\frac{1}{2}[X]_{st}^{\alpha\beta\gamma}, \quad \text{or} \quad X_{st}^{(\alpha\beta\gamma)} = [X]_{st}^{\alpha\beta\gamma}
$$

The last term left to compute is $\widehat{X}^{(\bullet_\gamma [\bullet_\alpha]_\beta)}$, which by Example 4.30 is the only obstruction to quasi-geometricity for $(3, 4] \ni p$-branched rough paths. Since by (4.83)

$$
X_{st}^\gamma \int_s^t X_{su}^\alpha \mathrm{d}^\circ X_u^\beta = \boldsymbol{X}_{st}^{[[\bullet_\alpha]_\beta]_\gamma} + \int_s^t X_{su}^\gamma \mathrm{d}^\circ \textstyle\int_s^u X_{sr}^\beta \mathrm{d}^\circ X_r^\alpha
$$

and therefore by (4.43)

$$
\widehat{X}_{st}^{(\bullet_\gamma [\bullet_\alpha]_\beta)} = \int_s^t X_{su}^\gamma \mathrm{d}^\circ \textstyle\int_s^u X_{sr}^\alpha \mathrm{d}^\circ X_r^\beta - \boldsymbol{X}_{st}^{[\bullet_\alpha \bullet_\gamma]_\beta} - \widetilde{\boldsymbol{X}}_{st}^{[\bullet_\alpha]_{(\beta\gamma)}}
$$

While we did not see a way of computing this term fully without delving into the specifics of symmetric integration, it is not difficult to see that its symmetrisation in α, β does not vanish if X has non-trivial cubic variation: by additivity in the integrand and integrator and the previous formulae we have

$$
\begin{aligned}
&\langle \ll \bullet_\gamma [\bullet_\alpha]_\beta \gg + \ll \bullet_\gamma [\bullet_\beta]_\alpha \gg, \boldsymbol{X}_{st} \rangle \\
&= \langle \bullet_\gamma [\bullet_\alpha]_\beta + \bullet_\gamma [\bullet_\beta]_\alpha - [[\bullet_\alpha]_\beta]_\gamma - [[\bullet_\beta]_\alpha]_\gamma - [\bullet_\beta \bullet_\gamma]_\alpha - [\bullet_\alpha \bullet_\gamma]_\beta - [\bullet_\alpha]_{(\beta\gamma)} - [\bullet_\beta]_{(\alpha\gamma)}, \boldsymbol{X}_{st} \rangle \\
&= \langle \bullet_\alpha \bullet_\beta \bullet_\gamma - [\bullet_\beta \bullet_\gamma]_\alpha - [\bullet_\alpha \bullet_\gamma]_\beta - [\bullet_\beta \bullet_\alpha]_\gamma - [\bullet_\alpha]_{(\beta\gamma)} - [\bullet_\beta]_{(\alpha\gamma)}, \boldsymbol{X}_{st} \rangle \\
&= \widetilde{X}_{st}^{(\alpha\beta\gamma)} + \widetilde{\boldsymbol{X}}_{st}^{[\bullet_\gamma]_{(\alpha\beta)}} \\
&= \frac{[X]_{st}^{\alpha\beta\gamma}}{2}
\end{aligned}
$$

Here is a sketch of what needs to be shown in order to make this a precise example of a stochastic branched rough path, and what our calculations will imply once this is carried out:

- Currently, no examples of stochastic processes with finite, non-trivial cubic variation are known (the closest thing to this is that fractional Brownian motion with Hurst parameter $1/6$ has cubic variation converging in law, but not UCP [NRS10]). For this example to be interesting, one such process must be produced.

- It must be shown that the branched iterated integrals of such a process defined via symmetric integration define a branched rough path with bracket extension, with the only caveat that $\langle \overset{\gamma}{\bullet}\!\!\!\underset{(\alpha\beta)}{\bullet}, \widetilde{\boldsymbol{X}}_{st}\rangle$ is not defined by $\int_s^t X^\gamma_{su}\mathrm{d}\widetilde{X}^{(\alpha\beta)}_u = 0$, but by

$$\int_s^t X^\gamma_{su}\mathrm{d}^\circ(X^\alpha X^\beta)_u - \int_s^t X^\gamma_{su}X^\beta_{su}\mathrm{d}^\circ X^\alpha_u - \int_s^t X^\gamma_{su}X^\alpha_{su}\mathrm{d}^\circ X^\beta_u$$

  Note that while this failure of associativity may seem strange, it would neither contradict bracket consistency (which only implies $\langle \overset{(\alpha\beta)}{\bullet}\!\!\!\underset{\gamma}{\bullet}, \boldsymbol{X}_{st}\rangle = 0$, compatibly with the value of the corresponding symmetric integral), nor would it compromise associativity of integration against this rough path. Such behaviour is not even ruled out for geometric rough paths, for which it is perfectly possible for the rough path not to be zero above a zero component, and this does indeed happen in the third rough path in Example 4.35 above.

- One would also need to show that, for a suitable class of integrands that includes functions of X, rough integration against this rough path coincides a.s. with symmetric integration (i.e. the analogue of the statements [FH20, Proposition 5.1, Corollary 5.2] in the context of Itô and Stratonovich integration), which has been used in the formal calculations above.

- We will then have a stochastic branched rough path extended with a bracket satisfying

$$\widetilde{X}^{(\alpha\beta)} \equiv 0, \qquad X^{(\alpha\beta\gamma)}_{st} = [X]^{\alpha\beta\gamma}_{st}, \qquad \widetilde{\boldsymbol{X}}^{\overset{\gamma}{\bullet}\underset{(\alpha\beta)}{\bullet}}_{st} = \frac{[X]^{\alpha\beta\gamma}_{st}}{2}, \qquad \widehat{X}^{\frac{1}{2}\left(\,{}^{\gamma}_{\bullet}\overset{\alpha}{\bullet}\underset{\beta}{\bullet} + {}^{\gamma}_{\bullet}\overset{\beta}{\bullet}\underset{\alpha}{\bullet}\,\right)}_{st} = \frac{[X]^{\alpha\beta\gamma}_{st}}{4}$$

  the last of which implies that this rough path cannot be quasi-geometric if it has non-zero cubic variation.

4.4 Integration against branched rough paths on manifolds

The definition of branched rough paths on manifolds is almost identical to those given for non-geometric $3 > p$-rough paths in Chapter 2 and for geometric rough paths in Chapter 3. The only significant difference is that we must push forward not only the branched rough path, but also a consistent simple bracket extension, which is done using the augmented definition of pushforwards Definition 4.26. This is because, first of all, a definition that does not carry these data would not be possible, since the simple bracket is even necessary to define the non-bracketed pushforward; moreover, it will be needed when defining rough integration. We will call the set of such objects $\widetilde{\mathscr{C}}^p_\omega([0,T],M)$. It is helpful to continue to think of $\boldsymbol{X}$ as being manifold-valued (since its trace X literally is), with the additional data of the simple bracket $\widetilde{\boldsymbol{X}}$ (whose trace does not take values in the manifold) also being carried from one chart to the other.

Before discussing integration, we begin with some background and notation. From now on we assume M is endowed with a covariant derivative on its tangent bundle; we will not assume M to be Levi-Civita, or even torsion-free. The notions that we will need about covariant derivatives are not advanced and, unless otherwise specified, can all be found in [Lee97, Ch.4]; some care is, however, required, since we will often be considering the n^{th} iterated covariant derivative, something that is not done often. We adopt the convention that, if S, T

are tensor fields and V is a vector field

$$\langle \nabla_V T, S \rangle = \langle \nabla T, V \otimes S \rangle \tag{4.85}$$

i.e. the direction of covariant differentiation occupies the "first slot"; this is the only instance in which we differ from [Lee97, Ch.4] in terms of notation. Our choice works well with the convention that, in the expression of a tensor, the contravariant part (i.e. a tensor product of the tangent space TM) always comes before the covariant part. Furthermore, it has the benefit of not reversing the order of covariant differentiation when iterating ∇, e.g. for another vector field U we have

$$\langle \nabla^2 T, U \otimes V \otimes S \rangle \coloneqq \langle \nabla\nabla T, U \otimes V \otimes S \rangle = \langle \nabla_U \nabla T, V \otimes S \rangle$$

Note that this is not equal to $\langle \nabla_U \nabla_V T, S \rangle$: by the Leibniz rules for ∇ w.r.t. to the dual pairing and tensor product

$$\begin{aligned} \langle \nabla\nabla T, U \otimes V \otimes S \rangle &= \nabla_U \langle \nabla T, V \otimes S \rangle - \langle \nabla T, \nabla_U (V \otimes S) \rangle \\ &= \nabla_U \langle \nabla T, V \otimes S \rangle - \langle \nabla T, \nabla_U V \otimes S \rangle - \langle \nabla T, V \otimes \nabla_U S \rangle \end{aligned}$$

while

$$\langle \nabla_U \nabla_V T, S \rangle = \nabla_U \langle \nabla_V T, S \rangle - \langle \nabla_V T, \nabla_U S \rangle$$

so that

$$\langle \nabla_U \nabla_V T, S \rangle - \langle \nabla\nabla T, U \otimes V \otimes S \rangle = \langle \nabla_{\nabla_U V} T, S \rangle$$

For a tensor field T we define $\nabla^n T$ inductively by $\nabla\nabla^{n-1} T$; this is obviously associative, i.e. $\nabla\nabla^n T = \nabla^n \nabla T$. The most important case is when T is a function $f \in C^\infty M$, for which $\nabla^n f \in \Gamma(T^*M^{\otimes n})$, Γ denoting the $C^\infty M$-module of sections. While it is well known that $\nabla^2 f$ is a symmetric tensor if and only if ∇ is torsion-free, the same does not hold at higher orders: indeed, even assuming torsion-freeness, $\nabla^3 f$ is symmetric (for general f) if and only if ∇ is flat [Kum05, Theorem 2.3] (incidentally, simple examples show that the torsion may contribute to the symmetric part of $\nabla^3 f$). It is not possible, therefore, to assume symmetry of higher order covariant derivatives of functions, without restricting our attention to trivial cases. For this reason, while symmetry will play an important and delicate role in this section, torsion will not.

In keeping with the rest of this chapter, we will mostly compute things in coordinates, which on manifolds are local. Given a chart, we denote ∂_γ the basis elements of TM that they define at each point, and d^γ the elements of the dual basis. For a tuple $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)$ we will denote the operator

$$\nabla_{\boldsymbol{\gamma}} - \coloneqq \langle \nabla^n -, \partial_{\gamma_1} \otimes \cdots \otimes \partial_{\gamma_n} \rangle$$

Given two charts, we will denote the "new" coordinates using Latin indices. We will use the symbol ∂ to denote the basis vectors ∂_γ, to denote the Jacobian of the change of coordinates ∂^k_γ (so $\partial_i = \partial^\gamma_i \partial_\gamma$), and more generally $\partial_{\boldsymbol{\gamma}}$ will denote the operator consisting of partial differentiation according to the tuple $\boldsymbol{\gamma}$ in the given chart φ, with $\partial^\alpha_{\boldsymbol{\beta}} \coloneqq \partial_{\boldsymbol{\beta}} \varphi^\alpha$. We will often use manipulations such as $\partial_k \partial^\alpha_{\boldsymbol{\beta}} = \partial^\gamma_k \partial^\alpha_{\gamma\boldsymbol{\beta}}$.

We are now ready to discuss the transfer principle. Let $\widetilde{\boldsymbol{X}} \in \widetilde{\mathscr{C}}^p_\omega([0,T], M)$. We are looking for an expression in local coordinates

$$\mathrm{d}_\nabla \widetilde{\boldsymbol{X}}^{\boldsymbol{\alpha}} = \frac{|\boldsymbol{\alpha}|!}{|\boldsymbol{\beta}|!} S^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}}(X) \mathrm{d}\widetilde{\boldsymbol{X}}^{(\boldsymbol{\beta})} \tag{4.86}$$

for tuples $\boldsymbol{\alpha}$, where we are summing over the tuple $\boldsymbol{\beta}$, $S^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}}$ are locally-defined smooth functions on M and $\mathrm{d}\widetilde{\boldsymbol{X}}^{(\boldsymbol{\beta})}$ are the local expressions of the differentials of $\widetilde{\boldsymbol{X}}$. As in the rest of the chapter, in this section too we are only taking sums on tuples of length $\leq \lfloor p \rfloor$. (4.86) is meant to extend the usual differential of the simply bracketed $\widetilde{\boldsymbol{X}}$ to the curved setting. When $\boldsymbol{\gamma}$ is a single index γ, we will denote $\mathrm{d}_\nabla \widetilde{\boldsymbol{X}}^{\gamma} = \mathrm{d}_\nabla \boldsymbol{X}^{\gamma}$, while keeping in mind that this still depends on the simple bracket. The first thing to notice is that S is only determined up to symmetry in the bottom indices, since it is being evaluated against $\mathrm{d}\widetilde{\boldsymbol{X}}$. On the other hand, there is no particular reason to assume S to be symmetric in the top indices: this means that the $\mathrm{d}_\nabla \widetilde{\boldsymbol{X}}^{\boldsymbol{\gamma}}$'s may not be symmetric, which is why we have omitted the round brackets for the tuple $\boldsymbol{\gamma}$. While this may seem counterintuitive at first, one should remember that ∇^n is also generally not a symmetric operator when M is not flat. (4.86) will then locally define a rough path (by lifting the integrals of the RHS) whose trace is indexed not by $\widetilde{[d]}$ but by tuples in $[d]$ (of length $\leq \lfloor p \rfloor$).

We require two conditions of (4.86):

Itô-Kelly formula on manifolds. For $g \in C^\infty M$ the formula (4.41) holds with covariant differentiation replacing ordinary differentiation:

$$g(X)_{st} = \frac{1}{|\boldsymbol{\gamma}|!} \int_s^t \nabla_{\boldsymbol{\gamma}} g(X) \mathrm{d}_\nabla \widetilde{\boldsymbol{X}}^{(\boldsymbol{\gamma})} \tag{4.87}$$

Contravariance. $\mathrm{d}_\nabla \boldsymbol{X}$ (i.e. the differentials indexed by single indices) transforms like a tangent vector: the change of coordinates reads

$$\mathrm{d}_\nabla \boldsymbol{X}^k = \partial^k_\gamma \mathrm{d}_\nabla \boldsymbol{X}^\gamma \tag{4.88}$$

While both requirements may seem to fall under the category change of variable formulae, the second cannot in general be inferred from the first: a counterexample to the general case is provided in Example 4.41 below. Moreover, we see no way to solve the second condition for S, while doing so for the first will be the first step performed to obtain conditions on it. It should also be remarked that, although it would seem natural to require that the whole of $\mathrm{d}_\nabla \widetilde{\boldsymbol{X}}$ transform as a tensor, this is not in general compatible with the first requirement, also shown in Example 4.41 below. The above change of variable formula is not the main function of the bracket integrators $\mathrm{d}_\nabla \widetilde{\boldsymbol{X}}$ (in excess of $\mathrm{d}_\nabla \boldsymbol{X}$), which has to do with defining RDEs; this will not be explained fully, but is mentioned in Conclusions and further directions below.

We proceed by writing (4.87) by expanding the LHS using the ordinary change of variable formula (4.41), and the RHS by using the ansatz (4.86): we have

$$\frac{1}{|\boldsymbol{\beta}|!} \partial_{\boldsymbol{\beta}} g(X) \mathrm{d}\widetilde{\boldsymbol{X}}^{(\boldsymbol{\beta})} = \frac{1}{|\boldsymbol{\beta}|!} \nabla_{\boldsymbol{\alpha}} g(X) S^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}}(X) \mathrm{d}\widetilde{\boldsymbol{X}}^{(\boldsymbol{\beta})}$$

from which, matching coefficients up to symmetry of $\boldsymbol{\beta}$, and requiring the resulting identity to hold on the whole of M

$$\partial_{\boldsymbol{\beta}} g = \nabla_{\boldsymbol{\alpha}} g S^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}} \tag{4.89}$$

for all $g \in C^\infty M$. This identity has the benefit of not containing differentials, and is the natural condition on S

that guarantees (4.87) for arbitrary elements of $\widetilde{\mathscr{C}}^p_\omega([0,T],M)$. It immediately implies $S^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}} = 0$ for $|\boldsymbol{\alpha}| > |\boldsymbol{\beta}|$, since $\nabla_{\boldsymbol{\alpha}} g$ is a sum that does not contain derivatives of g of order higher than $|\boldsymbol{\alpha}|$: indeed, if $|\boldsymbol{\beta}| < |\boldsymbol{\alpha}| = \lfloor p \rfloor$ then $S^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}}$ must vanish, since otherwise the RHS of (4.89) would contain a derivative of order higher than $|\boldsymbol{\beta}|$; this implies that if $|\boldsymbol{\beta}| < |\boldsymbol{\alpha}| = \lfloor p \rfloor - 1$ again $S^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}} = 0$, and so on until we reach $|\boldsymbol{\beta}| = |\boldsymbol{\alpha}|$. Since $\nabla_{\boldsymbol{\gamma}} g$ contains exactly one term involving order-$|\boldsymbol{\gamma}|$ derivatives, and it is $\partial_{\boldsymbol{\gamma}} g$, $S^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}}$ is forced equal to $\delta^{\boldsymbol{\alpha}}_{(\boldsymbol{\beta})} = \delta^{(\boldsymbol{\alpha})}_{(\boldsymbol{\beta})}$ when $|\boldsymbol{\beta}| = |\boldsymbol{\alpha}|$, with round brackets denoting the multiset corresponding to the tuple (so "on diagonal" it actually is true that $S^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}}$ may be considered symmetric in the upper indices). As a nice consequence, we have that $\mathrm{d}_\nabla \widetilde{\boldsymbol{X}}^{(\boldsymbol{\beta})}$ will always be an integrator of regularity $|\boldsymbol{\gamma}|$. The terms $S^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}}$ with $|\boldsymbol{\alpha}| < |\boldsymbol{\beta}|$ are more complex to describe, and underdetermined without further assumptions. One idea might be to require them to be symmetric in the upper indices, which fixes them uniquely: indeed, if $|\boldsymbol{\alpha}| = |\boldsymbol{\beta}| - 1$, since $S^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}}$ multiplies the only remaining derivatives of order $|\boldsymbol{\alpha}|$ apart from the one present in $\nabla_{\boldsymbol{\beta}} g$, it is fixed, up to symmetry, by the requirement that none appear on the LHS, and this continues to hold inductively all the way down to $|\boldsymbol{\alpha}| = 1$. Another way of viewing all of this is to think of $\nabla_{\boldsymbol{\gamma}} g, \partial_{\boldsymbol{\gamma}} g$ as a row vectors,

$$\nabla_{\boldsymbol{\beta}} g = \partial_{\boldsymbol{\alpha}} g L^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}} \tag{4.90}$$

for a block matrix L (whose upper indices are only determined up to symmetry since they multiply coordinate partial derivatives): (4.89) now becomes $\partial_{\boldsymbol{\gamma}} g = \partial_{\boldsymbol{\alpha}} g L^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}} S^{\boldsymbol{\beta}}_{\boldsymbol{\gamma}}$ or

$$L^{(\boldsymbol{\alpha})}_{\boldsymbol{\beta}} S^{\boldsymbol{\beta}}_{(\boldsymbol{\gamma})} = \delta^{(\boldsymbol{\alpha})}_{(\boldsymbol{\gamma})} \tag{4.91}$$

In other words, S must be a left inverse to L (viewed as acting on row vectors - this reverses order of composition), which exists since L is upper triangular with identity maps on the diagonal and therefore injective. We will call the unique S's that are all symmetric in the upper indices the *symmetric solution*.

Remark 4.37. Note that the uniqueness statement about the symmetric solution does not mean that the general solution to (4.89) is determined up to symmetry in the upper indices: it is neither true that two solutions must have identical symmetric part, nor that adding arbitrary asymmetric functions to a solution will still result in a solution. This is because asymmetries in the upper indices of some order may affect the symmetric part at lower orders, as will be exemplified in Example 4.41 below. (Note: we have use the term *asymmetric* to mean "having zero symmetric part", which at order > 2 is different to "antisymmetric", which means alternating).

While the symmetric solution might seem like the most natural one, it will not work for our purposes, since it fails to transform like a vector, as shown below in Example 4.41; we will instead need the following

Definition 4.38 (Higher Christoffel symbols). In a chart define tensor fields $\Gamma^m_{\boldsymbol{\gamma}} \in \Gamma_{\mathrm{loc}} TM^{\otimes m}$ for $m \in \mathbb{N}$ and non-empty tuples of indices $\boldsymbol{\gamma}$ recursively by

$$\begin{cases} \Gamma^1_\gamma := \partial_\gamma \\ \Gamma^m_{\gamma_1,\ldots,\gamma_n} := 0 & m > n \vee m = 0 \\ \Gamma^m_{\gamma_1,\ldots,\gamma_n} := \nabla_{\gamma_1} \Gamma^m_{\gamma_2,\ldots,\gamma_n} + \partial_{\gamma_1} \otimes \Gamma^{m-1}_{\gamma_2,\ldots,\gamma_n} & 2 \le m \le n \end{cases} \tag{4.92}$$

Letting $\Gamma^m_{\boldsymbol{\beta}} =: \Gamma^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}} \partial_{\boldsymbol{\alpha}}$, we call the $\Gamma^{\boldsymbol{\alpha}}_{\boldsymbol{\beta}}$'s the *Higher Christoffel symbols*.

We will always be careful that using the superscript to denote tuples (or single indices) or orders does not introduce ambiguities. The most important of these symbols are the $\Gamma_{\mathrm{loc}}TM \ni \Gamma^1$'s, which are given by

$$\Gamma^1_{\beta_1 \ldots \beta_n} = \nabla_{\beta_1} \cdots \nabla_{\beta_{n-1}} \partial_{\beta_n} \tag{4.93}$$

from which it is clear that they restrict to the ordinary Christoffel symbols when $n = 2$. The recursive definition in coordinates reads

$$\Gamma^{\alpha_1,\ldots,\alpha_m}_{\beta_1,\ldots,\beta_n} = \partial_{\beta_1} \Gamma^{\alpha_1,\ldots,\alpha_m}_{\beta_2,\ldots,\beta_n} + \sum_{k=1}^{m} \Gamma^{\alpha_k}_{\beta_1\gamma} \Gamma^{\alpha_1,\ldots,\gamma,\ldots,\alpha_m}_{\beta_2,\ldots,\beta_n} + \delta^{\alpha_1}_{\beta_1} \Gamma^{\alpha_2,\ldots,\alpha_m}_{\beta_2,\ldots,\beta_n} \tag{4.94}$$

which when $m = 1$ reads

$$\Gamma^{\alpha}_{\beta_1,\ldots,\beta_n} = \partial_{\beta_1} \Gamma^{\alpha}_{\beta_2,\ldots,\beta_n} + \Gamma^{\alpha}_{\beta_1\gamma} \Gamma^{\gamma}_{\beta_2,\ldots,\beta_n} \tag{4.95}$$

Note that although they are defined as tensors, the Γ's are not tensorial in the bottom indices: this is already clear from the case $m = 1, n = 2$. Also, these coefficients are completely determined by the connection, and therefore by the ordinary Christoffel symbols. If we were to define the Γ's without referring to coordinates, they would take values in complicated jet bundles; for this reason we prefer the coordinate description.

Proposition 4.39 (Γ is a solution)**.** *The higher Christoffel symbols* (4.92) *solve* (4.89)*.*

Proof. The claim can be rewritten as

$$\sum_{k=1}^{n} \langle \nabla^k g, \Gamma^k_{\gamma_1,\ldots,\gamma_n} \rangle = \partial_{\gamma_1,\ldots,\gamma_n} g$$

for $n \geq 1$ and indices $\gamma_1, \ldots, \gamma_n$. We proceed by induction. For $n = 1$ there is nothing to show. Using the Leibniz rule for dual pairings, we have

$$\begin{aligned}
&\sum_{k=1}^{n+1} \langle \nabla^k g, \Gamma^k_{\gamma_1,\ldots,\gamma_{n+1}} \rangle \\
&= \langle \nabla^1 g, \nabla_{\gamma_1} \Gamma^1_{\gamma_2,\ldots,\gamma_{n+1}} \rangle + \sum_{k=2}^{n} \langle \nabla^k g, \nabla_{\gamma_1} \Gamma^k_{\gamma_2,\ldots,\gamma_{n+1}} + \partial_{\gamma_1} \otimes \Gamma^{k-1}_{\gamma_2,\ldots,\gamma_{n+1}} \rangle + \langle \nabla^{n+1} g, \partial_{\gamma_1} \otimes \Gamma^n_{\gamma_2,\ldots,\gamma_{n+1}} \rangle \\
&= \nabla_{\gamma_1} \sum_{k=1}^{n} \langle \nabla^k g, \Gamma^k_{\gamma_2,\ldots,\gamma_{n+1}} \rangle - \sum_{k=1}^{n} \langle \nabla_{\gamma_1} \nabla^k g, \Gamma^k_{\gamma_2,\ldots,\gamma_{n+1}} \rangle + \sum_{k=2}^{n+1} \langle \nabla^k g, \partial_{\gamma_1} \otimes \Gamma^{k-1}_{\gamma_2,\ldots,\gamma_{n+1}} \rangle \\
&= \nabla_{\gamma_1} \partial_{\gamma_2,\ldots,\gamma_{n+1}} g - \sum_{k=1}^{n} \langle \nabla^{k+1} g, \partial_{\gamma_1} \otimes \Gamma^k_{\gamma_2,\ldots,\gamma_{n+1}} \rangle + \sum_{k=2}^{n+1} \langle \nabla^k g, \partial_{\gamma_1} \otimes \Gamma^{k-1}_{\gamma_2,\ldots,\gamma_{n+1}} \rangle \\
&= \partial_{\gamma_1,\ldots,\gamma_{n+1}} g
\end{aligned}$$

This concludes the proof. ■

We now turn our attention to the contravariance condition: this is what will allow us to integrate in a way that does not depend on the system of local coordinates. As done before, we seek a formulation of it that does not involve rough paths. Substituting in the ansatz (4.86) and applying the formula for pushforwards of simple brackets (4.59) to the RHS

$$\begin{aligned}\frac{1}{|\boldsymbol{j}|!}S^i_{\boldsymbol{j}}(X)\mathrm{d}\widetilde{\boldsymbol{X}}^{(\boldsymbol{j})} &= \frac{1}{|\boldsymbol{\beta}|!}\partial^i_\alpha S^\alpha_{\boldsymbol{\beta}}(X)\mathrm{d}\widetilde{\boldsymbol{X}}^{(\boldsymbol{\beta})}\\ &= \frac{1}{|\boldsymbol{\beta}|!|\boldsymbol{j}^1|!\cdots|\boldsymbol{j}^m|!}\partial^i_\alpha S^\alpha_{\boldsymbol{\beta}}\partial^{\beta_1}_{\boldsymbol{j}^1}\cdots\partial^{\beta_m}_{\boldsymbol{j}^m}(X)\mathrm{d}\widetilde{\boldsymbol{X}}^{(\boldsymbol{j}^1\ldots\boldsymbol{j}^m)}\end{aligned}$$

The transformation law that must hold therefore is

$$S^i_{\boldsymbol{j}} \overset{(\boldsymbol{j})}{=} \sum_{\substack{\boldsymbol{j}^1\ldots\boldsymbol{j}^m=\boldsymbol{j}\\|\boldsymbol{j}^1|,\ldots,|\boldsymbol{j}^m|>0}} \frac{|\boldsymbol{j}|!}{|\boldsymbol{\beta}|!|\boldsymbol{j}^1|!\cdots|\boldsymbol{j}^m|!}\partial^i_\alpha S^\alpha_{\boldsymbol{\beta}}\partial^{\beta_1}_{\boldsymbol{j}^1}\cdots\partial^{\beta_m}_{\boldsymbol{j}^m} \tag{4.96}$$

where the $(\boldsymbol{j})$ is meant as a reminder that the identity is only considered up to symmetry in $\boldsymbol{j}$, which will be crucial in the next proof. Note, however, that since we are indexing S with tuples, we are summing over all non-empty tuples $\boldsymbol{j}^1,\ldots,\boldsymbol{j}^m$ whose concatenation (without taking multisets) equals $\boldsymbol{j}$.

Proposition 4.40. *The higher Christoffel symbols with single upper index satisfy the change of coordinates* (4.96).

Proof. We proceed by induction. For $|\boldsymbol{j}| = 1$ the assertion reduces to the obvious identity $\delta^i_j = \partial^i_\alpha\delta^\alpha_\beta\partial^\beta_j$. In all the following expressions we denote $\boldsymbol{j} = (j_0,\ldots,j_n)$, $\boldsymbol{j}^- = (j_1,\ldots,j_n)$ and the variable tuples $\boldsymbol{\beta} = (\beta_0,\ldots,\beta_m)$, $\boldsymbol{\beta}^- = (\beta_1,\ldots,\beta_m)$ over which we will be summing. Using (4.95), we have

$$\begin{aligned}\Gamma^i_{j_0\cdots j_n} &= \partial_{j_0}\Gamma^i_{j_1,\ldots,j_n} + \Gamma^i_{j_0k}\Gamma^k_{j_1,\ldots,j_n}\\ &= \sum_{\substack{\boldsymbol{j}^1\ldots\boldsymbol{j}^m=\boldsymbol{j}^-\\|\boldsymbol{j}^1|,\ldots,|\boldsymbol{j}^m|>0}}\frac{n!}{m!|\boldsymbol{j}^1|!\cdots|\boldsymbol{j}^m|!}\partial_{j_0}(\partial^i_\alpha\Gamma^\alpha_{\boldsymbol{\beta}^-}\partial^{\beta_1}_{\boldsymbol{j}^1}\cdots\partial^{\beta_m}_{\boldsymbol{j}^m})\\ &\quad + (\partial^i_\alpha\Gamma^\alpha_{\beta_0\gamma}\partial^{\beta_0}_{j_0}\partial^\gamma_k + \partial^i_\alpha\partial^\alpha_{j_0k})\cdot\sum_{\substack{\boldsymbol{j}^1\ldots\boldsymbol{j}^m=\boldsymbol{j}^-\\|\boldsymbol{j}^1|,\ldots,|\boldsymbol{j}^m|>0}}\frac{n!}{m!|\boldsymbol{j}^1|!\cdots|\boldsymbol{j}^m|!}\partial^k_\alpha\Gamma^\alpha_{\boldsymbol{\beta}^-}\partial^{\beta_1}_{\boldsymbol{j}^1}\cdots\partial^{\beta_m}_{\boldsymbol{j}^m}\\ &= \sum_{\substack{\boldsymbol{j}^1\ldots\boldsymbol{j}^m=\boldsymbol{j}^-\\|\boldsymbol{j}^1|,\ldots,|\boldsymbol{j}^m|>0}}\frac{n!}{m!|\boldsymbol{j}^1|!\cdots|\boldsymbol{j}^m|!}\Big[\partial^i_{\beta_0\alpha}\partial^{\beta_0}_{j_0}\Gamma^\alpha_{\boldsymbol{\beta}^-}\partial^{\beta_1}_{\boldsymbol{j}^1}\cdots\partial^{\beta_m}_{\boldsymbol{j}^m} + \partial^i_\alpha\partial_{\beta_0}\Gamma^\alpha_{\boldsymbol{\beta}^-}\partial^{\beta_0}_{j_0}\partial^{\beta_1}_{\boldsymbol{j}^1}\cdots\partial^{\beta_m}_{\boldsymbol{j}^m}\\ &\qquad + \partial^i_\alpha\Gamma^\alpha_{\boldsymbol{\beta}^-}\sum_{l=1}^m\partial^{\beta_1}_{\boldsymbol{j}^1}\cdots\partial^{\beta_l}_{j_0\boldsymbol{j}^l}\cdots\partial^{\beta_m}_{\boldsymbol{j}^m}\Big]\\ &\quad + \sum_{\substack{\boldsymbol{j}^1\ldots\boldsymbol{j}^m=\boldsymbol{j}^-\\|\boldsymbol{j}^1|,\ldots,|\boldsymbol{j}^m|>0}}\frac{n!}{m!|\boldsymbol{j}^1|!\cdots|\boldsymbol{j}^m|!}\Big[\partial^i_\alpha\Gamma^\alpha_{\beta_0\gamma}\Gamma^\gamma_{\boldsymbol{\beta}^-}\partial^{\beta_0}_{j_0}\partial^{\beta_1}_{\boldsymbol{j}^1}\cdots\partial^{\beta_m}_{\boldsymbol{j}^m} + \partial^i_\gamma\partial^\gamma_{j_0k}\partial^k_\alpha\Gamma^\alpha_{\boldsymbol{\beta}^-}\partial^{\beta_1}_{\boldsymbol{j}^1}\cdots\partial^{\beta_m}_{\boldsymbol{j}^m}\Big]\end{aligned}$$

Since $\partial^i_{\beta_0\alpha}\partial^{\beta_0}_{j_0} + \partial^i_\gamma\partial^\gamma_{j_0k}\partial^k_\alpha = \partial_\alpha(\partial^i_{\beta_0}\partial^{\beta_0}_{j_0}) = \partial_\alpha\delta^i_{j_0} = 0$ the first and last terms cancel out, and we continue, using (4.95) again

$$\begin{aligned}&= \sum_{\substack{\boldsymbol{j}^1\ldots\boldsymbol{j}^m=\boldsymbol{j}^-\\|\boldsymbol{j}^1|,\ldots,|\boldsymbol{j}^m|>0}}\frac{n!}{m!|\boldsymbol{j}^1|!\cdots|\boldsymbol{j}^m|!}\Big[\partial^i_\alpha\partial_{\beta_0}\Gamma^\alpha_{\boldsymbol{\beta}^-}\partial^{\beta_0}_{j_0}\partial^{\beta_1}_{\boldsymbol{j}^1}\cdots\partial^{\beta_m}_{\boldsymbol{j}^m} + \partial^i_\alpha\Gamma^\alpha_{\boldsymbol{\beta}^-}\sum_{l=1}^m\partial^{\beta_1}_{\boldsymbol{j}^1}\cdots\partial^{\beta_l}_{j_0\boldsymbol{j}^l}\cdots\partial^{\beta_m}_{\boldsymbol{j}^m}\\ &\qquad + \partial^i_\alpha\Gamma^\alpha_{\beta_0\gamma}\Gamma^\gamma_{\boldsymbol{\beta}^-}\partial^{\beta_0}_{j_0}\partial^{\beta_1}_{\boldsymbol{j}^1}\cdots\partial^{\beta_m}_{\boldsymbol{j}^m}\Big]\end{aligned}$$

$$= \sum_{\substack{\boldsymbol{j}^1 \dots \boldsymbol{j}^m = \boldsymbol{j}^- \\ |\boldsymbol{j}^1|,\dots,|\boldsymbol{j}^m|>0}} \frac{n!}{m!|\boldsymbol{j}^1|!\cdots|\boldsymbol{j}^m|!} \left[\partial^i_\alpha \Gamma^\alpha_{\boldsymbol{\beta}} \partial^{\beta_0}_{j_0} \partial^{\beta_1}_{\boldsymbol{j}^1} \cdots \partial^{\beta_m}_{\boldsymbol{j}^m} + \partial^i_\alpha \Gamma^\alpha_{\boldsymbol{\beta}^-} \sum_{l=1}^m \partial^{\beta_1}_{\boldsymbol{j}^1} \cdots \partial^{\beta_l}_{j_0 \boldsymbol{j}^l} \cdots \partial^{\beta_m}_{\boldsymbol{j}^m} \right]$$

Now, recall that the ways of writing a positive integer as a sum of other positive integers is called a *composition* when order of summands is taken into account, and *partition* when it is not. Since $\boldsymbol{j}$ is fixed, summing over $\boldsymbol{j}^1 \dots \boldsymbol{j}^m = \boldsymbol{j}^-$ is the same as summing over compositions of n. To proceed, however, we will need to make use of the fact that the above expression only needs to be considered up to symmetry in $\boldsymbol{j}$: this means that summands of one of the two types

$$\partial^i_\alpha \Gamma^\alpha_{\boldsymbol{\beta}} \partial^{\beta_0}_{j_0} \partial^{\beta_1}_{\boldsymbol{j}^1} \cdots \partial^{\beta_m}_{\boldsymbol{j}^m}, \qquad \partial^i_\alpha \Gamma^\alpha_{\boldsymbol{\beta}^-} \partial^{\beta_1}_{\boldsymbol{j}^1} \cdots \partial^{\beta_l}_{j_0 \boldsymbol{j}^l} \cdots \partial^{\beta_m}_{\boldsymbol{j}^m} \tag{4.97}$$

corresponding to distinct compositions of $n+1$ that define the same partition must be merged into the same summand.

Let $P(k)$ denote the set of integer partitions of k, $P_h(k)$ denote the set of those consisting of h summands, $P^1_h(k)$ those that have at least one summand that is 1, and $P^{>1}_h(k)$ those for which each summand is greater than 1. We proceed to express the expression obtained for $\Gamma^i_{j_0 \cdots j_n}$ by summing over $\rho \in P(n+1)$. This means that in each summand we will pick one composition whose corresponding partition is ρ (it does not matter which one), and we will multiply by the number of such compositions. In the following, ρ will be the partition of $n+1$ given by $\sum_{l=1}^q \lambda_l n_l$ with $n_1 < \ldots < n_q$ and $\lambda_l \in \mathbb{N}_*$ s.t. $\sum_{l=1}^q \lambda_l = m$. If $\rho \in P^1_m(n+1)$ then both types of term (4.97) contribute to the corresponding summand, while if $\rho \in P^{>1}_m(n+1)$ only the second does. We will imply that in each summand, $\boldsymbol{k}^1, \dots, \boldsymbol{k}^m$ realises the partition ρ as described above: we are not summing over the $\boldsymbol{k}^l$'s, but just picking one such collection of tuples in each summand. The expression obtained is (setting the variable tuple $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_m)$)

$$= \sum_{\substack{m=1,\dots,n+1 \\ \rho \in P^1_m(n+1)}} \left[\frac{(m-1)!}{(\lambda_1 - 1)!\lambda_2! \cdots \lambda_q!} \frac{n!}{(m-1)! n_1!^{\lambda_1} \cdots n_q!^{\lambda_q}} \right.$$
$$\left. + m \sum_{r=2}^q \frac{\lambda_r (m-1)!}{\lambda_1! \cdots \lambda_q!} \frac{n!}{m! n_1!^{\lambda_1} n_2!^{\lambda_2} \cdots (n_r - 1)! n_r!^{\lambda_r - 1} \cdots n_q!^{\lambda_q}} \right] \partial^i_\alpha \Gamma^\alpha_{\boldsymbol{\gamma}} \partial^{\gamma_1}_{\boldsymbol{k}^1} \cdots \partial^{\gamma_m}_{\boldsymbol{k}^m}$$

$$+ \sum_{\substack{m=1,\dots,n+1 \\ \rho \in P^{>1}_m(n+1)}} m \sum_{r=1}^q \frac{\lambda_r (m-1)!}{\lambda_1! \cdots \lambda_q!} \frac{n!}{m! n_1!^{\lambda_1} n_2!^{\lambda_2} \cdots (n_r - 1)! n_r!^{\lambda_r - 1} \cdots n_q!^{\lambda_q}} \partial^i_\alpha \Gamma^\alpha_{\boldsymbol{\gamma}} \partial^{\gamma_1}_{\boldsymbol{k}^1} \cdots \partial^{\gamma_m}_{\boldsymbol{k}^m}$$

The two summands in the square bracket correspond to the two types of terms in (4.97), while in the single term in the second sum corresponds to the second type. In each product of two fractions, the first is the number of multiset partitions realising ρ, and the second is the old term $\frac{n!}{m!|\boldsymbol{j}^1|!\cdots|\boldsymbol{j}^m|!}$ written with the updated parameters. The precise expressions are the result of standard counting arguments. We elaborate on the least trivial, the term on the second line; the explanations for the other two follow similar arguments. Here we are counting, given $\sum_{l=1}^q \lambda_l n_l = \rho \in P^1_m(n+1)$, how many times one of the old terms $\partial^i_\alpha \Gamma^\alpha_{\boldsymbol{\beta}^-} \partial^{\beta_1}_{\boldsymbol{j}^1} \cdots \partial^{\beta_l}_{j_0 \boldsymbol{j}^l} \cdots \partial^{\beta_m}_{\boldsymbol{j}^m}$, with $\boldsymbol{j}^1 \dots \boldsymbol{j}^m = \boldsymbol{j}^-$ represent the partition ρ. Again by symmetry, we may assume $l = 1$, i.e. $\partial^i_\alpha \Gamma^\alpha_{\boldsymbol{\beta}^-} \partial^{\beta_1}_{j_0 \boldsymbol{j}^1} \cdots \partial^{\beta_m}_{\boldsymbol{j}^m}$; however, since the previous expression had a sum $\sum_{l=1}^m$ we must multiply by a factor of m (immediately to the left of $\sum_{r=2}^{q}$ in the new expression). Now, the tuple $j_0\boldsymbol{j}^1$ has length ≥ 2: this means it cannot correspond to any of the first λ_1 summands of ρ, since $n_1 = 1$. We sum, therefore, over $l = \lambda_1 + 1, \ldots, m$, which is the same as summing over $r = 2, \ldots, q$ and multiplying each term by λ_r, which is in the numerator of the first fraction. The $(m-1)!$ in the same numerator reflects the fact that, once the tuple $j_0\boldsymbol{j}^1$ has been assigned length n_r, the remaining ones can be permuted, factoring out permutations between those of the same length — this explains the denominator $\lambda_1! \cdots \lambda_q!$. The reason for the terms in the second fraction is now obvious. The inductive step can now be concluded:

$$
\begin{aligned}
&= \sum_{\substack{m=1,\ldots,n+1\\ \rho\in P_m^1(n+1)}} \left[\frac{(m-1)!}{(\lambda_1-1)!\lambda_2!\cdots\lambda_q!}\frac{n!}{(m-1)!n_1!^{\lambda_1}\cdots n_q!^{\lambda_q}}\right.\\
&\qquad\qquad \left. + m\sum_{r=2}^{q}\frac{(m-1)!}{\lambda_1!\cdots\lambda_q!}\frac{\lambda_r n!}{m!n_1!^{\lambda_1}n_2!^{\lambda_2}\cdots(n_r-1)!n_r!^{\lambda_r-1}\cdots n_q!^{\lambda_q}}\right]\partial_\alpha^i\Gamma_{\boldsymbol{\gamma}}^\alpha\partial_{\boldsymbol{k}^1}^{\gamma_1}\cdots\partial_{\boldsymbol{k}^m}^{\gamma_m}\\
&\quad + \sum_{\substack{m=1,\ldots,n+1\\ \rho\in P_m^{>1}(n+1)}} m\sum_{r=1}^{q}\frac{\lambda_r n!}{m!n_1!^{\lambda_1}n_2!^{\lambda_2}\cdots(n_r-1)!n_r!^{\lambda_r-1}\cdots n_q!^{\lambda_q}}\frac{(m-1)!}{\lambda_1!\cdots\lambda_q!}\partial_\alpha^i\Gamma_{\boldsymbol{\gamma}}^\alpha\partial_{\boldsymbol{k}^1}^{\gamma_1}\cdots\partial_{\boldsymbol{k}^m}^{\gamma_m}\\
&= \sum_{\substack{m=1,\ldots,n+1\\ \rho\in P_m^1(n+1)}} \frac{n!\left(\lambda_1+\sum_{r=2}^{q}\lambda_r n_r\right)}{n_1!^{\lambda_1}\cdots n_q!^{\lambda_q}\lambda_1!\lambda_2!\cdots\lambda_q!}\partial_\alpha^i\Gamma_{\boldsymbol{\gamma}}^\alpha\partial_{\boldsymbol{k}^1}^{\gamma_1}\cdots\partial_{\boldsymbol{k}^m}^{\gamma_m}\\
&\quad + \sum_{\substack{m=1,\ldots,n+1\\ \rho\in P_m^{>1}(n+1)}} \frac{n!\sum_{r=1}^{q}\lambda_r n_r}{n_1!^{\lambda_1}\cdots n_q!^{\lambda_q}\lambda_1!\lambda_2!\cdots\lambda_q!}\partial_\alpha^i\Gamma_{\boldsymbol{\gamma}}^\alpha\partial_{\boldsymbol{k}^1}^{\gamma_1}\cdots\partial_{\boldsymbol{k}^m}^{\gamma_m}\\
&= \sum_{\substack{m=1,\ldots,n+1\\ \rho\in P_m(n+1)}} \frac{(n+1)!}{n_1!^{\lambda_1}\cdots n_q!^{\lambda_q}\lambda_1!\lambda_2!\cdots\lambda_q!}\partial_\alpha^i\Gamma_{\boldsymbol{\gamma}}^\alpha\partial_{\boldsymbol{k}^1}^{\gamma_1}\cdots\partial_{\boldsymbol{k}^m}^{\gamma_m}\\
&= \sum_{\boldsymbol{k}^1\ldots\boldsymbol{k}^m=\boldsymbol{j}} \frac{(n+1)!}{m!|\boldsymbol{k}^1|!\cdots|\boldsymbol{k}^m|!}\partial_\alpha^i\Gamma_{\boldsymbol{\gamma}}^\alpha\partial_{\boldsymbol{k}^1}^{\gamma_1}\cdots\partial_{\boldsymbol{k}^m}^{\gamma_m}
\end{aligned}
$$

concluding the proof. ■

Before stating the main theorem we lay out some of the above calculations in the first interesting case, which is already complex enough to illustrate many of the subtleties that have been brought up.

Example 4.41 (Christoffel symbols of order ≤ 3)**.** We solve (4.89) up to level 3, which reads

$$\partial_{\alpha\beta\gamma}g = \nabla_\lambda g S^\lambda_{\alpha\beta\gamma} + \nabla_{\mu\nu}g S^{\mu\nu}_{\alpha\beta\gamma} + \nabla_{\lambda\mu\nu}g S^{\lambda\mu\nu}_{\alpha\beta\gamma} \tag{4.98}$$

A couple of interesting observations will follow. For $g \in C^\infty M$ we compute

$$
\begin{aligned}
\nabla_\gamma g &= \partial_\gamma g\\
\nabla_{\alpha\beta} g &= \partial_{\alpha\beta} g - \partial_\delta g \Gamma^\delta_{\alpha\beta}\\
\nabla_{\alpha\beta\gamma} g &= \partial_{\alpha\beta\gamma} g - \partial_\delta g \partial_\alpha \Gamma^\delta_{\beta\gamma} - \Gamma^\delta_{\alpha\beta}\partial_{\gamma\delta} f - \Gamma^\delta_{\alpha\gamma}\partial_{\beta\delta} g - \Gamma^\delta_{\beta\gamma}\partial_{\alpha\delta} g + \partial_\varepsilon g \Gamma^\varepsilon_{\delta\gamma}\Gamma^\delta_{\alpha\beta} + \partial_\varepsilon g \Gamma^\varepsilon_{\beta\delta}\Gamma^\delta_{\alpha\gamma}
\end{aligned}
$$

The presence of the term $\partial_{\alpha\beta\gamma} g$ in the expression for $\nabla_{\alpha\beta\gamma}$ implies

$$\begin{aligned} S^{\lambda\mu\nu}_{\alpha\beta\gamma} &= \sum_{\sigma\in\mathfrak{S}_6} \epsilon_\sigma \delta^{\sigma_*(\lambda\mu\nu)}_{\alpha\beta\gamma} \qquad\qquad \text{with} \sum_{\sigma\in\mathfrak{S}_6} \epsilon_\sigma = 1 \\ &= \sum_{\sigma\in\mathfrak{S}_6} \epsilon_\sigma \delta^{\lambda\mu\nu}_{\sigma_*(\alpha\beta\gamma)} = \delta^{\lambda\mu\nu}_{\alpha\beta\gamma} \end{aligned}$$

up to symmetry in the bottom indices. Substituting into (4.98) we get

$$\begin{aligned} &\partial_\lambda g S^{\lambda}_{\alpha\beta\gamma} + (\partial_{\mu\nu} g - \partial_\lambda g \Gamma^\lambda_{\mu\nu})\Gamma^{\mu\nu}_{\alpha\beta\gamma} \\ &- \partial_\delta g \partial_\alpha \Gamma^\delta_{\beta\gamma} - \Gamma^\delta_{\alpha\beta}\partial_{\delta\gamma} g - \Gamma^\delta_{\alpha\gamma}\partial_{\beta\delta} g - \Gamma^\delta_{\beta\gamma}\partial_{\alpha\delta} g + \partial_\varepsilon g \Gamma^\varepsilon_{\delta\gamma}\Gamma^\delta_{\alpha\beta} + \partial_\varepsilon g \Gamma^\varepsilon_{\beta\delta}\Gamma^\delta_{\alpha\gamma} = 0 \end{aligned}$$

Setting the sum of all the second derivatives to zero implies

$$S^{\mu\nu}_{\alpha\beta\gamma} = c\Gamma^\mu_{\alpha\beta}\delta^\nu_\gamma + (3-c)\Gamma^\nu_{\alpha\beta}\delta^\mu_\gamma, \qquad c \in \mathbb{R}$$

up to symmetry in the bottom indices. The presence of the parameter c is due to the fact that this multiplies $\partial_{\mu\nu} g$, which is symmetric in μ, ν, while we are not assuming $\Gamma^{\mu\nu}_{\alpha\beta\gamma}$ to be symmetric in the upper indices. Resubstituting, we can solve for $\Gamma^\lambda_{\alpha\beta\gamma}$. The full, general solution is given by

$$\begin{pmatrix} S^\lambda_\gamma & S^\lambda_{\alpha\beta} & S^\lambda_{\alpha\beta\gamma} \\ S^{\mu\nu}_\gamma & S^{\mu\nu}_{\alpha\beta} & S^{\mu\nu}_{\alpha\beta\gamma} \\ S^{\lambda\mu\nu}_\gamma & S^{\lambda\mu\nu}_{\alpha\beta} & S^{\lambda\mu\nu}_{\alpha\beta\gamma} \end{pmatrix} = \begin{pmatrix} \delta^\lambda_\gamma & \Gamma^\lambda_{\alpha\beta} & \partial_\alpha\Gamma^\lambda_{\beta\gamma} + \Gamma^\lambda_{\gamma\sigma}\Gamma^\sigma_{\alpha\beta} + 2(c-1)\Gamma^\lambda_{[\sigma\gamma]}\Gamma^\sigma_{\alpha\beta} \\ 0 & \delta^{\mu\nu}_{\alpha\beta} & c\Gamma^\mu_{\alpha\beta}\delta^\nu_\gamma + (3-c)\Gamma^\nu_{\alpha\beta}\delta^\mu_\gamma \\ 0 & 0 & \delta^{\lambda\mu\nu}_{\alpha\beta\gamma} \end{pmatrix}, \qquad c \in \mathbb{R}$$

where the square brackets denote antisymmetrisation. Two important cases of the parameter are $c = 1$, corresponding to Definition 4.38 (as can be seen by computing terms with (4.94)), and $c = 3/2$, corresponding to the symmetric solution (since it makes $\Gamma^{\mu\nu}_{\alpha\beta\gamma}$ symmetric in the upper two indices). We emphasise once more that the above expression is considered up to symmetry in the lower indices α, β, γ. Since the term $\Gamma^\lambda_{[\mu\gamma]}\Gamma^\mu_{\alpha\beta}$ vanishes if and only if $c = 1$, Proposition 4.40 implies that $\Gamma^\lambda_{\alpha\beta\gamma} = \partial_\alpha\Gamma^\lambda_{\beta\gamma} + \Gamma^\lambda_{\gamma\sigma}\Gamma^\sigma_{\alpha\beta}$ has the correct transformation rule,

$$\Gamma^l_{ijk} = \partial^l_\lambda \Gamma^\lambda_{\alpha\beta\gamma}\partial^\alpha_i\partial^\beta_j\partial^\gamma_k + 2\partial^l_\lambda\Gamma^\lambda_{\alpha\beta}\partial^\alpha_i\partial^\beta_{jk} + \partial^l_\lambda\partial^\lambda_{ijk} \tag{4.99}$$

which in turn implies that, when $c \neq 1$, $S^\lambda_{\alpha\beta\gamma}$ has the same transformation rule if and only if $\Gamma^\lambda_{[\sigma\gamma]}\Gamma^\sigma_{\alpha\beta}$ does. We can check that this is not the case: recalling that torsion is a tensor, we have

$$\Gamma^l_{[hk]}\Gamma^h_{ij} = \partial^l_\lambda\Gamma^\lambda_{[\sigma\gamma]}\partial^\sigma_h\partial^\gamma_k(\partial^h_\tau\Gamma^\tau_{\alpha\beta}\partial^\alpha_i\partial^\beta_j + \partial^h_\delta\partial^\delta_{ij}) = \partial^l_\lambda\Gamma^\lambda_{[\sigma\gamma]}\Gamma^\sigma_{\alpha\beta}\partial^\alpha_i\partial^\beta_j\partial^\gamma_k + \partial^l_\lambda\Gamma^\lambda_{[\sigma\gamma]}\partial^\sigma_{ij}\partial^\gamma_k$$

This is clearly different from (4.99) (even after symmetrising), which has a third derivative that is not present here. This shows us that the only choice for S that satisfies both desiderata (4.87), (4.88) is Definition 4.38. In particular, the symmetric solution fails the latter condition. This counterexample continues to work when considering solutions to (4.89) up to arbitrary orders, since the procedure restricts to the above for the portion of the solution with 3 lower indices $S_{\alpha\beta\gamma}$.

Continuing with the example of Γ up to order 3, another interesting consequence is the following. A natural extension of the contravariance condition would be to require that not only $\mathrm{d}_\nabla \boldsymbol{X}^\gamma$ transform as a vector, but the whole of $\mathrm{d}_\nabla \widetilde{\boldsymbol{X}}$ as a tensor. This, however, cannot be the case: arguing as in (4.96) and using Example 4.27 we see that S^2_3 must satisfy the transformation rule

$$S^{hl}_{ijk} = \partial^h_\mu \partial^l_\nu S^{\mu\nu}_{\alpha\beta\gamma} \partial^\alpha_i \partial^\beta_j \partial^\gamma_k + \frac{3}{2} \partial^h_\mu \partial^l_\nu (\partial^\mu_{ij} \partial^\nu_k + \partial^\nu_{ij} \partial^\mu_k)$$

On the other hand,

$$c\Gamma^h_{ij}\delta^l_k + (3-c)\Gamma^l_{ij}\delta^h_k = \partial^h_\mu \partial^l_\nu \big(c\Gamma^\mu_{\alpha\beta}\delta^\nu_\gamma + (3-c)\Gamma^\nu_{\alpha\beta}\delta^\mu_\gamma\big)\partial^\alpha_i \partial^\beta_j \partial^\gamma_k + \partial^h_\mu \partial^l_\nu \big(c\partial^\mu_{ij}\partial^\nu_k + (3-c)\partial^\nu_{ij}\partial^\mu_k\big)$$

implying that for tensoriality to hold, we must be in the symmetric case $c = 3/2$, not the Christoffel case $c = 1$. Symmetrising in the bottom indices does not affect this assertion.

Proposition 4.39, Proposition 4.40 and one of the conclusions of the above example can be summarised in the following statement:

Theorem 4.42 (Existence and uniqueness of the transfer principle)**.** *The higher Christoffel symbols are the unique solution to the two conditions* (4.87) *and* (4.88)*.*

Rough integration can now be defined unambiguously. For simplicity, we will state it only for integrands given by one-forms, though it can be expected to hold for more general controlled integrands (as done in Chapter 2 in the $2 \leq p < 3$ case).

Theorem 4.43 (Integration against branched rough paths on manifolds)**.**
Let $\boldsymbol{X} \in \widetilde{\mathscr{C}}^p_\omega([0,T], M)$ and $f \in \Gamma T^*M$. The expression

$$\int f_\gamma(X)\mathrm{d}_\nabla \boldsymbol{X}^\gamma \coloneqq \frac{1}{|\boldsymbol{\beta}|!}\int f_\alpha \Gamma^\alpha_{\boldsymbol{\beta}}(X)\mathrm{d}\widetilde{\boldsymbol{X}}^{(\boldsymbol{\beta})}$$

(with sums on $\alpha, \boldsymbol{\beta}$) does not depend on the system of local coordinates. Moreover, for $g \in C^\infty M$,

$$g(X)_{st} = \frac{1}{|\boldsymbol{\gamma}|!}\int_s^t \nabla_{\boldsymbol{\gamma}} g(X)\mathrm{d}_\nabla \widetilde{\boldsymbol{X}}^{(\boldsymbol{\gamma})}$$

Proof. This is a straightforward conclusion of all the theory developed in this section; it should be pointed out that Lemma 4.28 is necessary when integrating against the transformed rough path. ∎

Conclusions and further directions

In this chapter we hope to have convinced the reader that non-geometric rough paths are not, after all, incompatible with differential geometry, even at low regularity. Nonetheless, the additional technical baggage needed to formulate a theory of branched rough paths on manifolds is certainly not light, and for this reason we have, admittedly, only formulated some of the results that one could envisage including in such a discussion.

A goal that appears achievable without too much additional effort would involve using our transfer principle to give meaning to RDEs driven by quasi-geometric rough paths, as done in Definition 2.42 in the $2 \leq p < 3$ case. This was actually our motivation for discussing quasi-geometric rough paths in the first place, and for proving their characterisation in terms of the bracket. The point is that the same transfer principle should apply for RDEs, since transformations of quasi-geometric RDEs only require the simple bracket. The only complicating factor in defining

$$\mathrm{d}_N \boldsymbol{Y} = F(Y, X)\mathrm{d}_M \boldsymbol{X} \tag{4.100}$$

for quasi-geometric $\widetilde{\boldsymbol{X}}$, is that its the local expression

$$\frac{1}{|\boldsymbol{j}|!}{}^N\Gamma^i_{\boldsymbol{j}}(Y)\mathrm{d}\widetilde{\boldsymbol{Y}}^{(\boldsymbol{j})} = \frac{1}{|\boldsymbol{\beta}|!}F^i_\alpha(Y,X){}^M\Gamma^\alpha_{\boldsymbol{\beta}}(X)\mathrm{d}\widetilde{\boldsymbol{X}}^{(\boldsymbol{\beta})}$$

is not a closed-form RDE. To turn it into one, $F \in \Gamma\mathrm{Hom}(TM, TN)$ must be extended to a field of maps $\widetilde{F}$ that locally drive an RDE for the whole of $\widetilde{\boldsymbol{Y}}$, and this must be done compatibly with the condition that $\widetilde{\boldsymbol{Y}}$ be a quasi-geometric bracket extension of $\boldsymbol{Y}$. Ideally, it should be shown that $\widetilde{F}$ is unique with this property. A few preparatory lemmas are needed to set up these results, e.g. a version of the bracket pushforwards Theorem 4.25 for quasi-geometric RDEs, which should also make clear that pushforwards preserve quasi-geometricity (this is necessary to make the term "quasi-geometric" apply in a well-defined way to rough paths on manifolds in the first place). Once this is achieved, one could additionally show an analogue of the Itô-Stratonovich conversion Theorem 4.32 in the manifold case; this would require first showing that pushing forward commutes with Hoffman's exponential.

Writing out a transfer principle for RDEs driven by general branched rough paths is a much more ambitious goal. As discussed in (4.20), even full bracket extensions are not closed under lifts of their controlled paths. For this reason, such a transfer principle would have to depend on an iterated bracket extension. A first step in this direction would be to work out the formula for $\mathrm{d}_\nabla\widehat{\boldsymbol{X}}$ when $3 \leq p < 4$: since the regularity is low enough, this actually only requires a single iteration of the bracket extension. Moreover, this case would be sufficient to give meaning to manifold-valued RDEs driven by the rough path of Chapter 6 below.

A Hopf algebra conspicuously absent from the discussion on page 175 is that of planar rooted trees. This comes in two flavours: the one defined independently in [Foi02b] and [Hol03], in which the product is given by juxtaposing forests, and the one introduced in [MKW08], where the product is instead defined by shuffling them. The latter, which has the advantage over the former of being commutative, is used in [CEFMMK20] to define a particular type of differential equation: given a Lie group G, with associated Lie algebra $\mathfrak{g} = T_{1_G}G$, a smooth manifold M acted on transitively by G, smooth maps $F_\gamma \colon M \to \mathfrak{g}$, and defining vector fields $\#F_\gamma \in \Gamma TM$ by

$$\#F_\gamma(x) \coloneqq \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_0 \exp(tF_\gamma(x)).x$$

where $\exp\colon \mathfrak{g} \to G$ is the Lie-algebraic exponential map, the authors are interested in giving meaning to

$$\mathrm{d}Y = \#F_\gamma(Y)\mathrm{d}\boldsymbol{X}^\gamma$$

where $\boldsymbol{X}$ is an $\mathbb{R}^d$-valued *planarly branched* rough path. It would be interesting, especially once our transfer principle is extended to RDEs as detailed above, to compare our definition of RDE with theirs, and in particular to see whether/how the additional data of the connection needed in our case corresponds to the planar structure required in theirs. It should be said, however, that we do not expect planar forests to be particularly helpful for the type of RDE (4.100) (which does not involve Lie group actions). This can already be seen from the case of $2 \leq p < 3$, in which the only difference between non-planar and planar labelled forests is that the latter distinguishes between $\bullet_\alpha \bullet_\beta$ and $\bullet_\beta \bullet_\alpha$; their symmetrisation $\bullet_\alpha \bullet_\beta + \bullet_\beta \bullet_\alpha$, the shuffle of $\bullet_\alpha$ and $\bullet_\beta$, would already be occupied by $X^\alpha X^\beta$. The only excess terms given by planar forests are therefore $\bullet_\alpha \bullet_\beta - \bullet_\beta \bullet_\alpha$, which cannot be of help when it comes to indexing the symmetric bracket term $X^{(\alpha\beta)}$.

The last idea we wish to mention, very briefly, is a little vague. It has to do with relating the bracket extension of [Kel12] with the geometric rough path of [HK15] that is built to drive RDEs equivalent to ones driven by a specified branched rough path. These accomplish distinct goals, but have some points in common: for example, when it comes to quasi-geometric rough paths, which are simple bracket extensions, the Itô-Stratonovich formula of [HK15] seems to reduce to Hoffman's exponential. The type of result that we have in mind is a canonical mapping sending the iterated bracket extension (in the sense of (4.20)) of a branched rough path to a geometric rough path, expressed algebraically as a map of Hopf algebras. This should then be compared with the isomorphism of [BC19], which does not need a bracket extension, but appears to be non-canonical, and possibly not invariant under permutations of indices.

5

Computing the Wiener chaos expansion of the signature of a Gaussian process

Project status. This chapter consists of a report on work in progress, carried out jointly with Thomas Cass. While the formal calculation is present in its entirety, there are non-trivial technical aspects that need further attention, before the project can be considered complete. These are highlighted in eight ***Assumption*** environments, in which we state what must be assumed for the computation to proceed, and often offer ideas on how we plan on tackling these problems. In at least one case, Assumption 2, significant progress towards the solution has already been made (not included in this thesis, since it is the work of my supervisor). Once these questions are resolved, and modulo some reorganisation, we believe this chapter will be ready for journal submission. The validity of our results is independently corroborated by several examples, in which we show how our results agree with those in the literature.

Introduction

The signature of a bounded variation path $X\colon [0,T] \to \mathbb{R}^d$, namely its formal series of iterated Riemann-Stieltjes integrals

$$\sum_{n=0}^{\infty} \int_{0<u_1<\ldots<u_n<T} \mathrm{d}X_{u_1} \otimes \cdots \otimes \mathrm{d}X_{u_n} \quad \in T((\mathbb{R}^d))$$

is an almost complete invariant of the path's trace: as shown in the landmark paper [HL10], later extended to rough paths in [BGLY16], it characterises $\mathrm{Im}(X)$ up to tree-like excursions, i.e. sections in which the path retraces itself (a type of behaviour that a "generic" path does not exhibit when $d > 1$). In [CL16] it was shown that this remarkable fact admits a probabilistic counterpart: under reasonably general assumptions, the expected signature of a stochastic process characterises its law. In other words, expected signatures can be viewed as the moments of the process.

The importance of expected signatures was already known earlier. The formula for the expected Stratonovich signature of Brownian motion is original to [Faw03], but can be computed in several ways [FS17]. Of most relevance to this chapter is the method of [Bau04], which consists of performing an iterated Stratonovich-Itô conversion, and using the martingale property and the Itô isometry to simplify the expression. In [BC07] the authors compute and study the asymptotics of the expected signature of fractional Brownian motion (fBm) with Hurst parameter $H > 1/2$ as a way of describing the geometry of differential equations driven by it for small times. The calculation is not performed through a conversion of integrals, but by a simple linear interpolation method. The result, though not as explicit as that of the Brownian case, is still a closed-form expression involving multiple integrals of products of fractional powers $(v-u)^{2H}$. When $1/4 < H \leq 1/2$, however, while the signature is still defined in the rough path sense and integrable (see [CL16, Example 6.7]), this linear approximation fails to converge. Through an alternative method, [BC07] manage to compute the expected signature up to level 4, but their formula does not appear to generalise straightforwardly to higher levels, nor to different types of Gaussian processes.

The terms of the signature of a path X, or more generally of a rough path $\boldsymbol{X}$, can be used to expand, at arbitrary order, the solution of a rough differential equation (RDE) driven by $\boldsymbol{X}$. In [CQ02] it was shown that $H < 1/4$-fBm admits a canonical rough path lift, and lifts of more general Gaussian processes were introduced in [LQ02, FV10a]. On the other hand, there is another method for decomposing a functional of Gaussian noise: the Wiener chaos expansion. The terms of this decomposition are orthogonal in L^2 and the zero-th Wiener chaos projection corresponds to taking the expectation. In the Brownian case, the canonical basis for this expansion is also given by iterated integrals, but ones taken in the sense of Itô, not Stratonovich. For more general Gaussian processes the elements of this basis are defined via the divergence operator of Malliavin calculus, or Skorokhod integral. This operator has the drawback of not being defined in a pathwise sense, but has the advantage of retaining some of the properties of Itô integration, most notably that it vanishes in mean. Given that a Gaussian Wiener functional admits two different expansions, one pathwise and one probabilistic, it is natural to ask how these two expansions fit together.

The planned purpose of this chapter, which is work in progress, is to compute the Wiener chaos expansion of the signature for a broad class of centred Gaussian processes lifted to rough paths, which include $1/3 < H$-fBm and the Brownian bridge returning to the origin. At level zero, we obtain a formula for the expected signature. Our results can also be conjectured to hold for $1/4 < H$-fBm, and for many other processes with paths of bounded $4 > p$-variation that admit rough path lifts, although additional subtleties emerge in this regime. The formula that we obtain for the expected signature is similar to that obtained by [BC07] for $1/2 < H$-fBm, but differs in a fundamental aspect, which ensures that the integrals in it converge, and our description of the higher chaos levels follows similar lines. Our strategy is quite different to those employed previously, although for Brownian motion, and more generally Gaussian martingales, it simplifies significantly and reduces to the Stratonovich-Itô method mentioned above. The key result which we use is the rough integral-Skorokhod integral conversion formula of [CL19], which expresses the rough integral of an RDE solution against the driving Gaussian rough path as the sum of the Skorokhod integral of Y and correction terms consisting of Young and 2D Young integrals against the variance and covariance functions of the process. An extension of our results to the $p \in [3,4)$ case would use the formally analogous formula of [CL20], in which, however, the 2D Young integral may not converge a.s. Relying on the framework of Malliavin calculus for Gaussian rough paths (introduced in Section 5.1), we apply this formula to the exponential RDE satisfied by the signature (Section 5.2).

This results in a recursive formula relating each level of the signature to Skorokhod integrals and Young integrals against the variance and covariance functions of the signature terms of lower order and their products (Section 5.3). We are then able to derive a non-recursive formula which expresses the signature as a sum of mixed Skorokhod/Young iterated integrals (Section 5.4). Next, we show how to take Malliavin derivatives and expectations of this identity, from which, after some intricate combinatorial manipulations of terms, it is possible obtain the main result by using Stroock's formula for the Wiener chaos projections (Section 5.5). Our mixed iterated integrals are represented using diagrams which play a role similar to that of rooted forests in branched rough path theory, and the operations of taking the expectation and Malliavin derivative are given a precise graphical description in terms of such diagrams. A crucial ingredient of this calculation involves proving an n-factor Itô-Skorokhod isometry formula. We conclude with some examples of computations of low levels of the signature, which include checking that our formula at level 4 coincides with that of [BC07] for $1/4 < H$-fBm, and in Conclusions and further directions offer some ideas on how to extend and apply our results.

While our method is quite lengthy and complicated, we believe it is justified for the following reasons. First and most simply, we do not see an alternative way to prove our identity, since all other methods for computing expected signatures use properties that are not available for $(1/3, 1/2] \ni H$-fBm. Indeed, it was only applying [CL19] that it was possible for us to even discover the correct formula for the expected signature (and its higher chaos), which was not known to us and does not seem to appear in the literature: the formula at level 4 of [BC07] has quite a different presentation, and showing equivalence with ours is not trivial. Moreover, our result applies to a general class of Gaussian processes, which need not satisfy properties such as stationarity of increments. This means that our results are interesting and novel even for Gaussian semimartingales that are not martingales, such as the Brownian bridge returning to the starting point. For such processes, Itô and Skorokhod integration do not agree, and it is the latter that results in the cleanest way of obtaining our end results, while the calculation performed using classical stochastic calculus would rely in an inessential way on the Bichteler-Dellacherie decomposition of the process. Finally, our proposed method for computing the expected signature extends without additional difficulty to the computation of the Wiener chaos expansion. As argued in the conclusions, this could lead to important results on numerical schemes for RDEs.

As already mentioned, this chapter does not consist of a finished paper. At present, it can only be said to consist of the formal calculation involved in the proof, since several key technical lemmas of Gaussian analysis remain unproven. However, the validity of our results is corroborated by the fact that our formula for the expected signature coincides with the one of [BC07] for (i) $1/2 < H$-fBm, for (ii) $1/4 < H$-fBm at level 2 and 4, and with (iii) the well-known Brownian case. The precise details of what is missing are clearly spelt out in a series of ***Assumption***s, in many of which we also give our idea of the proposed way forward.

5.1 Background on Malliavin calculus for Gaussian processes

In this section we give a brief overview of the tools of Malliavin calculus that apply to the study of multidimensional Gaussian processes. This is by no means meant as an exhaustive treatment of the topic. In this section we follow [Nua06, NP12] for the general Malliavin calculus framework and [CL19] for certain aspects that are specific to the case of 1-parameter Gaussian processes and their rough paths.

Throughout this chapter we will be working with a Gaussian process with i.i.d. components $X \colon \Omega \times [0,T] \to \mathbb{R}^d$ where $\Omega = C([0,T], \mathbb{R}^d)$, $X_t(\omega) \coloneqq \omega(t)$, $\mathcal{F}_t \coloneqq \sigma(X_s : 0 \leq s \leq t)$. In the following we will write Ω to denote the filtered probability space $(\Omega, \mathcal{F}_\bullet, \mathbb{P})$ where, by Gaussianity, the probability measure $\mathbb{P}$ on Ω is characterised by the covariance function of X

$$R\colon [0,T]^2 \to \mathbb{R}^d \otimes \mathbb{R}^d, \quad R(s,t) \coloneqq \mathbb{E}[X_s \otimes X_t] \tag{5.1}$$

We will denote

$$\begin{gathered} R(t) \coloneqq R(t,t), \qquad R(\Delta(s,t)) \coloneqq R(t) - R(s) \\ R(s,\Delta(u,v)) \coloneqq R(s,v) - R(s,u) = \mathbb{E}[X_s \otimes X_{uv}], \qquad R(\Delta(u,v),s) \coloneqq R(s,\Delta(u,v)) \\ R(\Delta(s,u),\Delta(t,v)) \coloneqq R\begin{pmatrix} s & u \\ t & v \end{pmatrix} \coloneqq R(s,t) + R(u,v) - R(t,u) - R(s,v) = \mathbb{E}[X_{st} \otimes X_{uv}] \end{gathered} \tag{5.2}$$

for $u,v,s,t \in [0,T]$. Note that $R(\Delta(s,t)) \neq R(\Delta(s,t),\Delta(s,t))$. Under the i.i.d. hypothesis we will use R to denote the scalar covariance function.

Remark 5.1. We have assumed the that X has i.i.d. components mostly because the results in the literature on Gaussian processes and rough paths are proven under this hypothesis. Nevertheless, we will make as many as possible of our statements using notation that preserves its meaning to the non-i.i.d. case, should the underlying theory for it become available. For instance, when we write $V_{\alpha\beta}(y)R^{\alpha\beta}(t)$ using the Einstein convention, the reader should be aware that this simplifies to $\sum_{\gamma=1}^d V_{\gamma\gamma}(y)R^{11}(t)$, since $R^{\alpha\beta} = \delta^{\alpha\beta}R^{11}$ in the present setting.

We need to impose regularity hypotheses on R, which are taken from [CL19]. We assume that one of the following two cases holds:

$1 \le p < 2$. There exist $M \ge 0,\ 0 < \varrho < 1$ s.t. for all $s,t \in [0,T]$

$$|R(\Delta(s,t),\Delta(s,t))| \le M|t-s|^{1/\varrho}$$

This implies X admits a $1/p$-Hölder modification for any $p > 2\varrho$, and in particular for some $p < 2$. Moreover, we require $\|R(\cdot)\|_{q\text{-var}} < \infty$ for some q s.t. $1/p + 1/q > 1$.

$2 \le p < 3$. There exist $M \ge 0,\ 1 \le \varrho < 3/2$ s.t. for all $s,t \in [0,T]$

$$\|R(t,\cdot) - R(s,\cdot)\|_{\varrho\text{-var}} \le M|t-s|^{1/\varrho}$$

This implies that X admits a $1/p$-Hölder modification and can be lifted to a p-rough path, for any $p > 2\varrho$ and in particular for some $p \in [2,3)$.

Furthermore, we introduce the following hypothesis, which we do not believe to be essential, but which simplifies matters considerably:

Smoothness. $R(\cdot) \in C^1((0,T],(\mathbb{R}^d)^{\otimes 2})$, and $R(\cdot,\cdot)$ restricts to an element of $C^2(\Delta^2[0,T],(\mathbb{R}^d)^{\otimes 2})$, $\Delta^2[0,T]$ the open 2-simplex $\{(s,t) \in [0,T]^2 \mid s < t\}$.

Note that we are not requiring smoothness of $R(\cdot,\cdot)$ on $[0,T]^2$, which would exclude all examples of interest, but only off-diagonal. **From now on we assume X satisfies these requirements.**

We move onto the treatment of Malliavin calculus for X. We let $\mathcal{H}$ be the Hilbert space given by the completion of the $\mathbb{R}$-linear span of functions $[0,T] \to \mathbb{R}^d$, or equivalently $[0,T] \times [d] \to \mathbb{R}$ ($[d]$ the set with d elements)

$$\{\mathbb{1}^\gamma_{[0,t)} \mid t \in [0,T],\ \gamma = 1,\ldots,d\} \tag{5.3}$$

w.r.t. the inner product

$$\langle \mathbb{1}^\alpha_{[0,s)}, \mathbb{1}^\beta_{[0,t)} \rangle_{\mathcal{H}} \coloneqq R^{\alpha\beta}(s,t) \tag{5.4}$$

Under the i.i.d. hypothesis, $\mathcal{H} = \mathcal{H}^1 \oplus \ldots \oplus \mathcal{H}^n$, with the $\mathcal{H}^\gamma$'s all canonically isomorphic. This framework allows us to view the process as an isometry

$$X \colon \mathcal{H} \to L^2\Omega, \quad \mathbb{1}^\gamma_{[0,t)} \mapsto X^\gamma_t \tag{5.5}$$

often called an isonormal Gaussian process. Although $\mathcal{H}$ is defined abstractly as the completion of a set of basis functions under an inner product, it is possible to identify a dense subspace $\widetilde{\mathcal{H}}$ of it that is constituted of functions $[0,T] \to \mathbb{R}^d$:

$$\widetilde{\mathcal{H}} \coloneqq \bigcup_{r < \frac{\varrho}{\varrho - 1}} C^{r\text{-var}}_{\mathrm{pw}}([0,T],\mathbb{R}^d) \tag{5.6}$$

where the ϱ is the parameter appearing in the conditions on X, and the subscript pw means "piecewise continuous". $\widetilde{\mathcal{H}}$ is shown to be continuously embedded into $\mathcal{H}$ when equipped with inner product

$$\langle f, g \rangle_{\widetilde{\mathcal{H}}} = \int_{[0,T]^2} f_\alpha(s) g_\beta(t) R^{\alpha\beta}(\mathrm{d}s, \mathrm{d}t) \tag{5.7}$$

after quotienting modulo the equivalence relation $f \sim g \Leftrightarrow \|f - g\|_{\widetilde{\mathcal{H}}} = 0$, which is just a.e. equality under a suitable non-degeneracy hypothesis on R. In practice, we will always be dealing with paths in $\widetilde{\mathcal{H}}$ as opposed to abstract elements of $\mathcal{H}$.

***Assumption* 1.** Throughout this chapter we are going to assume that the elements of tensor powers of $\mathcal{H}$ that we are dealing with in fact lie in tensor powers of $\widetilde{\mathcal{H}}$. In many cases this is not obvious and must be shown. Alternatively, it should be argued that these elements are/can be manipulated as functions, which is the reason we are proceeding with this assumption.

The integral in (5.7) is a 2D Young integral: for $F \colon [0,T]^2 \to \mathbb{R}$ the it is defined as the limit

$$\int_{[0,T]^2} F(s,t) R(\mathrm{d}s, \mathrm{d}t) \coloneqq \lim_{|\pi| \to 0} \sum_{[s,u],[t,v] \in \pi} F(s,t) R \begin{pmatrix} s & u \\ t & v \end{pmatrix} \tag{5.8}$$

when it exists; sufficient conditions for convergence can be found in [CL19]. Since we are assuming the covariance function to be C^2 off-diagonal, it holds that 2D Young and Lebesgue integrals coincide when both are well defined:

$$\int_{0<s<t<T} F(s,t) R(\mathrm{d}s, \mathrm{d}t) = \int_{0<s<t<T} F(s,t) \partial_{12} R(s,t) \mathrm{d}s\mathrm{d}t \tag{5.9}$$

where the integration domain $0 < s < t < T$ just means the integrand is multiplied by the indicator function

$\mathbb{1}_{s<t}$ in both cases. On the other hand, the relationship between

$$\int_{[0,T]^2} F(s,t)R(\mathrm{d}s,\mathrm{d}t), \qquad \int_{[0,T]^2\setminus\{s=t\}} F(s,t)\partial_{12}R(s,t)\mathrm{d}s\mathrm{d}t$$

is more complicated. For example, when X is a Brownian motion the first is equal to $\int_0^T F(t,t)R(\mathrm{d}t) = \int_0^T F(t,t)R'(t)\mathrm{d}t$, while the second vanishes. If $F = \mathbb{1}_{[a,b]\times[c,d]}$, the first is equal to $R(\Delta(a,b),\Delta(c,d))$ while the integrand of the second may fail to be in $L^1([0,T]\setminus\{s=t\})$, as is the case when X is a $1/2 > H$-fBm (see Example 5.7 below) and $c < b$.

Let V be a separable Hilbert space; in all cases considered here V will be Hilbert tensor powers of $\mathcal{H}$, often the zero-th power $\mathbb{R}$. The derivative operator $\mathcal{D}$ is defined by setting

$$\mathcal{D}^m F := \sum_{k_1,\ldots,k_m=1}^{n} \partial_{k_1,\ldots,k_m} f(X^{\gamma_1}_{t_1},\ldots,X^{\gamma_n}_{t_n})\mathbb{1}^{\gamma_{k_1}}_{[0,t_{k_1}]}\otimes\cdots\otimes\mathbb{1}^{\gamma_{k_n}}_{[0,t_{k_n}]} \qquad F = f(X^{\gamma_1}_{t_1},\ldots,X^{\gamma_n}_{t_n})$$

where $f\in C^\infty(\mathbb{R}^n)$ has all derivatives of polynomial growth; more in general, for $v_i\in V$ and the F_i's of the same form as F above we set

$$\mathcal{D}^m Z := \sum_{i=1}^{l}\mathcal{D}^m F_i\otimes v_i \in L^q(\Omega,\mathcal{H}^{\odot m}\otimes V), \qquad Z = \sum_{i=1}^{l} F_i v_i \in L^q(\Omega, V) \tag{5.10}$$

Let $\mathcal{S}_V$ denote the set of random variables of the form of Z above. $\mathcal{D}^m$ is shown to be a closable operator, i.e. we have that for every sequence of random variables Z_n of the form of Z above s.t. $Z_n\to 0$ in $L^q(\Omega,V)$ and $\mathcal{D}^m Z_n\to D$ in $L^q(\Omega,\mathcal{H}^{\odot m}\otimes V)$ it must be the case that $D=0$. Defining $\mathbb{D}^{m,q}(V)\subseteq L^q(\Omega,V)$ to be the closure of $\mathcal{S}_V$ w.r.t. to the norm

$$\|Z\|_{\mathbb{D}^{m,q}(V)} = \mathbb{E}\Big[\sum_{k=0}^{m}\|\mathcal{D}^k Z\|^q_{\mathcal{H}^{\odot k}\otimes V}\Big]^{1/q} \tag{5.11}$$

(where $\mathcal{D}^0 = \mathbb{1}_{L^q(\Omega,V)}$) which defines the m-th *Malliavin derivative* of V-valued random variables.

$$\mathcal{D}^m\colon \mathbb{D}^{m,q}(V)\to L^q(\Omega,\mathcal{H}^{\odot m}\otimes V) \tag{5.12}$$

When $m=1$ or $V=\mathbb{R}$ they will be omitted from the notation. We will almost always take $q=2$. $\mathcal{D}^m$ is, by construction, closed: this means that if $\mathbb{D}^{m,q}(V)\ni Z_n\to Z$ in $L^q(\Omega,V)$ and $\mathcal{D}^m Z_n\to D$ in $L^q(\Omega,\mathcal{H}^{\odot m}\otimes V)$ then $Z\in\mathbb{D}^{m,q}(V)$ and $\mathcal{D}^m Z = D$. Very often $\mathcal{D}^m Z$ will actually belong to $L^q(\Omega,\widetilde{\mathcal{H}}^m\otimes V)$, which means it will be a random variable valued in a space of symmetric functions $([0,T]\times[d])^m\to V$, i.e. a function of m indices and m times. $\mathcal{D}$ satisfies a Leibniz rule, proved on $\mathcal{S}_V$ and extended: for $Y\in\mathbb{D}^{1,q}$ and $Z\in\mathbb{D}^{1,q}(V)$ s.t. $YZ\in\mathbb{D}^{1,q}(V)$

$$\mathcal{D}(YZ) = Y\mathcal{D}Z + Z\mathcal{D}Y \tag{5.13}$$

Remark 5.2 (Cameron-Martin space). There is an alternative, more variational approach to Malliavin Calculus, which bases the definitions of derivative and divergence on the Cameron-Martin space $\mathscr{H}$ instead of on $\mathcal{H}$. The former is defined as the closure of the linear span of

$$\{R^\gamma(t,\cdot) \mid t \in [0,T],\ \gamma = 1,\ldots,d\} \quad \text{w.r.t.} \quad \langle R^\alpha(s,\cdot), R^\beta(t,\cdot)\rangle_{\mathscr{H}} \coloneqq R^{\alpha\beta}(s,t)$$

and is isometric to $\mathcal{H}$ by the obvious mapping of basis elements

$$R_*\colon \mathcal{H} \to \mathscr{H}, \quad \mathbb{1}^\gamma_{[0,t)} \mapsto R^\gamma(t,\cdot)$$

The main difference between $\mathcal{H}$ and $\mathscr{H}$ is that the former contains discontinuous functions, whereas the latter is densely and continuously embedded in Ω with the sup norm. Indeed, one advantage of the Cameron-Martin space approach is that it allows us to interpret the Malliavin derivative as a directional derivative, i.e. roughly speaking it holds that

$$\langle \mathcal{D}Z(\omega), f\rangle_{\mathcal{H}} = \frac{\mathrm{d}}{\mathrm{d}r}\bigg|_0 Z(\omega + rR_* f)$$

On the other hand the divergence operator, which is more important for our purposes, is more nicely interpreted in the isonormal framework adopted in this chapter, as explained below. For a more detailed comparison of the two approaches see [CL19, p.14-15].

The m-th order *divergence operator* or *Skorokhod integral* is defined as the adjoint of the unbounded operator $\mathcal{D}^m$. We will not define its domain explicitly, rather use the (non-trivial) fact that it is not only defined but even continuous on the corresponding domain of the derivative:

$$\delta^m\colon \mathbb{D}^{m,q}(\mathcal{H}^{\odot m} \otimes V) \to L^q(\Omega, V) \tag{5.14}$$

and extended to $\mathbb{D}^{m,q}(\mathcal{H}^{\otimes m} \otimes V)$ via the projection $\mathcal{H}^{\otimes m} \twoheadrightarrow \mathcal{H}^{\odot m}$; we take the norm on symmetric tensor spaces to be just the restriction of the Hilbert tensor norm, without any scaling factor. δ^m is characterised by the adjoint property

$$\forall Z \in \mathbb{D}^{m,q}(V) \quad E[\langle Z, \delta^m H\rangle_V] = \mathbb{E}[\langle \mathcal{D}^m Z, H\rangle_{\mathcal{H}^{\odot m}\otimes V}] \tag{5.15}$$

according to which $\delta^0 = \mathbb{1}$ and δ the isometry X (5.5). Both derivative and divergence are associative in the sense that $\mathcal{D}^i \circ \mathcal{D}^j = \mathcal{D}^{i+j}$ and $\delta^i \circ \delta^j = \delta^{i+j}$. Taking $Z = 1 \in \mathbb{R}$ in (5.15) immediately implies that δ (and therefore δ^m) satisfies the zero mean property

$$\mathbb{E} \circ \delta = 0 \tag{5.16}$$

which will be of fundamental importance in this chapter. Another consequence of (5.15) is the following expression, valid for $H \in \mathbb{D}^{1,2}(\mathcal{H} \otimes V)$ and $Z \in \mathbb{D}^{1,2}$ s.t. $ZH \in \mathbb{D}^{1,2}(\mathcal{H} \otimes V)$

$$\delta(ZH) = Z\delta H - \langle \mathcal{D}Z, H\rangle_{\mathcal{H}} \tag{5.17}$$

This is shown by testing against arbitrary $Y \in \mathbb{D}^{1,2}$:

$$\begin{aligned}\mathbb{E}[Y\delta(ZH)] &= \mathbb{E}[\langle \mathcal{D}Y, ZH\rangle_{\mathcal{H}}] \\ &= \mathbb{E}[\langle Z\mathcal{D}Y, H\rangle_{\mathcal{H}}]\end{aligned}$$

$$= \mathbb{E}[\langle \mathcal{D}(YZ) - Y\mathcal{D}Z, H\rangle_{\mathcal{H}}]$$
$$= \mathbb{E}[Y(Z\delta H - \langle \mathcal{D}Z, H\rangle_{\mathcal{H}})]$$

by the Leibniz rule for $\mathcal{D}$. (5.17) shows how random constants cannot be taken out of the Skorokhod integral without a non-trivial correction term being generated. For example, when $Z = X(f)$ and $H = g$ for $f, g \in \mathcal{H}$ the above identity reduces to $\delta(X(f)g) = X(f)X(g) - \langle f, g\rangle_{\mathcal{H}}$, with deterministic correction term. Another important identity is *Heisenberg's commutativity* relation

$$\mathcal{D}\delta H = H + \delta(\mathcal{D}H), \qquad H \in \mathbb{D}^{1,2}(\mathcal{H} \otimes V) \tag{5.18}$$

which in particular implies $\mathcal{D}X(f) = f$. It is instructive to compute the divergence of an "elementary adapted integrand": for $0 \leq u_1 \leq \ldots \leq u_n \leq s \leq t \leq T$ and $F = f(X^{\beta_1}_{u_1}, \ldots, X^{\beta_n}_{u_n}) \in \mathcal{S}$ by (5.17) we have

$$\begin{aligned}\delta(F\mathbb{1}^{\alpha}_{[s,t)}) &= FX^{\alpha}_{st} - \langle \mathcal{D}F, \mathbb{1}^{\alpha}_{[s,t)}\rangle_{\mathcal{H}} \\ &= FX^{\alpha}_{st} - \sum_{k=1}^{n} \partial_k f(X^{\beta_1}_{u_1}, \ldots, X^{\beta_n}_{u_n}) R^{\alpha\beta_k}\begin{pmatrix} 0 & u_k \\ s & t\end{pmatrix}\end{aligned}$$

If X is a Gaussian martingale (but not necessarily if it is only a semimartingale) the second term vanishes by orthogonality of increments over disjoint intervals, and in this case the divergence operator actually agrees with the Itô integral for adapted processes that are integrable in both senses. For non-trivial covariance structures, however, the correction above implies that δ does not behave like a true integration operator, since, for instance, one cannot take out random constants. Because we will compare it with other types of integrals, it will nevertheless be helpful to use integral notation: for $H \in \mathbb{D}^{1,2}(\mathcal{H}^{\gamma})$ we interchangeably denote

$$\int_s^t H_u \delta X^{\gamma}_u = \delta^{\gamma}_u(H_u) = \delta^{\gamma}(H) = \delta(H\mathbb{1}^{\gamma}_{[s,t)}) \tag{5.19}$$

and similar notation will also be used at higher orders of the divergence operator, through the use of multiple jointly ordered time and index sub/superscripts.

Remark 5.3 (Symmetry). It is important to note that, although $\mathcal{D}^m$ is symmetric tensor-valued and δ^m is invariant under symmetries of its integrand, these properties only hold when times and indices are permuted jointly. For instance, while it is true that $\delta^{\alpha\beta}_{uv}(\mathbb{1}_{\Delta[s,t]}) = \delta^{\beta\alpha}_{vu}(\mathbb{1}_{\Delta[s,t]})$, the difference

$$\delta^{\alpha\beta}_{uv}(\mathbb{1}_{\Delta[s,t]}) - \delta^{\beta\alpha}_{uv}(\mathbb{1}_{\Delta[s,t]}) = \int_{s<u<v<t} \left(\delta X^{\alpha}_u \delta X^{\beta}_v - \delta X^{\beta}_u \delta X^{\alpha}_v\right)$$

does not vanish, and equals the — equivalently Itô or Stratonovich — Lévy area increment $\mathbb{A}^{\alpha\beta}_{st}$ when X is a martingale, a statement we will later show to carry over to Gaussian rough paths.

A special class of Skorokhod integrand is given by deterministic, i.e. elements of $\mathcal{H}^{\otimes m} \otimes V \subseteq \mathbb{D}^{m,2}(\mathcal{H}^{\otimes m} \otimes V)$. For this class of integrands the Skorokhod integral is called the *multiple Wiener integral*. A remarkable feature of multiple Wiener integration is that for $f \in \mathcal{H}^{\odot m}$ and $g \in \mathcal{H}^{\odot n}$

$$\mathbb{E}[\delta^m(f)\delta^n(g)] = \delta^{mn} m! \langle f, g\rangle_{\mathcal{H}^{\otimes m}} \tag{5.20}$$

so if $f, g \in \widetilde{\mathcal{H}}^{\otimes m}$ are not necessarily symmetric

$$\mathbb{E}[\delta^m(f)\delta^m(g)] = \sum_{\sigma \in \mathfrak{S}_m} \int_{[0,T]^{2m}} f_{\alpha_1,\ldots,\alpha_m}(s_1,\ldots,s_m) g_{\beta_{\sigma(1)},\ldots,\beta_{\sigma(m)}}(t_{\sigma(1)},\ldots,t_{\sigma(m)}) \\ R^{\alpha_1\beta_{\sigma(1)}}(\mathrm{d}s_1, \mathrm{d}t_{\sigma(1)}) \cdots R^{\alpha_m\beta_{\sigma(m)}}(\mathrm{d}s_m, \mathrm{d}t_{\sigma(m)})$$

This tells us that multiple Wiener integration defines an isometry

$$\delta^\bullet \colon \bigoplus_{m=0}^{\infty} \mathcal{H}^{\odot m} \xrightarrow{\cong} L^2\Omega \tag{5.21}$$

where the source is given the rescaled inner product $(f, g) \mapsto m!\langle f, g\rangle_{\mathcal{H}^{\otimes m}}$ for f, g in the same $\mathcal{H}^{\odot m}$ and zero otherwise, and recall that $L^2\Omega$ is given the sigma algebra generated by the process X. The image of the m-th Wiener integral operator, the space of the random variables $\delta^m f$ with f ranging in $\mathcal{H}^{\odot m}$, is called the m-th *Wiener chaos* of X. We denote it $\mathscr{W}^m$ and the m-th Wiener chaos projection $w^m \colon L^2\Omega \twoheadrightarrow \mathscr{W}^m$. Note that $w^0 = \mathbb{E}$ with values in $\mathscr{W}^0 = \mathbb{R}$, while $\mathscr{W}^1$ is given by linear functions of X. We therefore have the Wiener chaos decomposition

$$L^2\Omega = \bigoplus_{m=0}^{\infty} \mathscr{W}^m \tag{5.22}$$

which means we may represent any $L^2\Omega$ random variable as an L^2-absolutely convergent series

$$L^2\Omega \ni Z = \sum_{m=0}^{\infty} w^m Z, \qquad \|Z\|^2_{L^2\Omega} = \sum_{m=0}^{\infty} \|w^m Z\|^2_{L^2\Omega} = \sum_{m=0}^{\infty} m!\|f^m\|^2_{\mathcal{H}^{\otimes m}} \tag{5.23}$$

where $f^m = (\delta^m)^{-1} \circ w^m(Z)$. This decomposition is useful to generate L^2-convergent numerical schemes. The map $(\delta^m)^{-1} \circ w^m$ admits an expression in terms of the Malliavin derivative: this is *Stroock's formula*, which states that for $Z \in \mathbb{D}^{m,2}$

$$(\delta^m)^{-1} \circ w^m(Z) = \frac{1}{m!}\mathbb{E}[\mathcal{D}^m Z] = \mathbb{E}[\mathbb{1}_{\Delta^m[0,T]}\mathcal{D}^m Z] \tag{5.24}$$

where $\Delta^m[0, T]$ is the m-simplex over the interval $[0, T]$. As a consequence, if $Z \in \mathbb{D}^{\infty,2} \coloneqq \bigcap_{m=0}^{\infty} \mathbb{D}^{m,2}$ we can write its Wiener chaos decomposition as the series

$$Z = \sum_{m=0}^{\infty} \delta^m \mathbb{E}[\mathbb{1}_{\Delta^m[0,T]}\mathcal{D}^m Z] \tag{5.25}$$

5.2 The rough-Skorokhod conversion formula for RDE solutions

In this section we state the main recent result which underlies the rest of this chapter, the rough-Skorokhod integral conversion formula [CL19, Theorem 6.1], or rather a slight generalisation of it. All assumptions and notation is as in Section 5.1, and we consider a matrix-valued RDE

$$\mathrm{d}Y^k_\alpha = V^k_{\alpha\beta}(Y)\mathrm{d}\boldsymbol{X}^\beta, \quad Y_0 = y_0 \tag{5.26}$$

Here X takes values in $\mathbb{R}^d$, Y in $\mathbb{R}^{e\times d}$, and we use Greek indices to denote the coordinates of $\mathbb{R}^d$ and Latin ones for $\mathbb{R}^e$. We will denote $V^{k,\alpha}_\beta := V^k_{\alpha\beta}$ when it is convenient in order to apply the Einstein summation convention correctly. The conversion formula then reads

$$\begin{aligned}\int_0^T Y^k_\gamma \mathrm{d}\boldsymbol{X}^\gamma = \int_0^T Y^k_\gamma \delta X^\gamma + \frac{1}{2}\int_0^T V^k_{\alpha\beta}(Y_t)R^{\alpha\beta}(\mathrm{d}t) \\ + \int_{0<s<t<T} \left((J_{ts}V)(Y_s) - V(Y_t)\right)^k_{\alpha\beta} R^{\alpha\beta}(\mathrm{d}s,\mathrm{d}t)\end{aligned} \tag{5.27}$$

Here the integral against δ is a Skorokhod integral, $R^{\alpha\beta}(\mathrm{d}t)$ is a Young integral and the one against $R^{\alpha\beta}(\mathrm{d}s,\mathrm{d}t)$ is a 2D Young integral. By the smoothness hypotheses on the variance and covariance functions and (5.9) we can write

$$R(\mathrm{d}s,\mathrm{d}t) = \partial_{12}R(s,t)\mathrm{d}s\mathrm{d}t, \quad R(\mathrm{d}t) = R'(t)\mathrm{d}t, \quad R(s,\mathrm{d}t) = \partial_2 R(s,t)\mathrm{d}t \tag{5.28}$$

where ∂_2 denotes partial differentiation w.r.t. to the second argument (not a second derivative) and ∂_1 partial differentiation w.r.t. to the first argument. $J_{ts}(Y_s)$ is the Jacobian of the flow of the RDE, evaluated at the solution with initial condition $Y_0 = y_0$, from time s to time t. It is defined by

$$J^{i,\alpha}_{j,\beta;ts} := \frac{\mathrm{d}}{\mathrm{d}r}\bigg|_{r=0} (\text{solution of } \mathrm{d}Y^i_\alpha = V^i_{\alpha\gamma}(Y)\mathrm{d}\boldsymbol{X}^\gamma \text{ at time } t\text{, started at } Y_s + r\partial_{j,\beta} \text{ at time } s)$$

with Y as in (5.26) and $\{\partial_{k,\gamma} \mid \gamma = 1,\dots,d;\ k = 1,\dots,e\}$ the canonical basis of $\mathbb{R}^{e\times d}$, and $(J_{ts}V)^k_{\alpha\beta}(y) := (J_{ts}V)^{k,\alpha}_\beta(y) := J^{k,\alpha}_{h,\gamma;ts}V^{h,\gamma}_\beta(y)$. All that is necessary to know about J_{ts} is that is satisfies the linear RDE [CL19, (42)]

$$\mathrm{d}_t J^{i,\alpha}_{j,\beta;ts}(Y_s) = \partial_{k,\gamma}V^{i,\alpha}_\delta(Y_t)J^{k,\gamma}_{j,\beta;ts}(Y_s)\mathrm{d}\boldsymbol{X}^\delta_t, \quad J^{i,\alpha}_{j,\beta;ss}(Y_s) = \delta^i_j\delta^\alpha_\beta$$

which implies $(J_{ts}V)^{i,\alpha}_\beta(Y_s)$ satisfies the linear RDE

$$\mathrm{d}_t A^{i,\alpha}_{\beta;ts} = \partial_{k,\gamma}V^{i,\alpha}_\delta(Y_t)A^{k,\gamma}_{\beta;ts}\mathrm{d}\boldsymbol{X}^\delta_t, \quad A^{i,\alpha}_{\beta;ss} = V^{i,\alpha}_\beta(Y_s) \tag{5.29}$$

Remark 5.4. It is important to stress that the 2D Young integral in (5.27) cannot in general be split as the difference of two integrals, as is shown in [CL19, p.53], as the integrand must vanish on the diagonal (at the correct order) for convergence to hold.

For some more regular processes, however, $R(\cdot,\cdot)$ is once differentiable on the diagonal: an important example is $1/2 < H$-fBm (see Example 5.7 below), for which

$$\partial_2 R(s,t) = H(s^{2H-1} - (t-s)^{2H-1}), \qquad s \le t \tag{5.30}$$

and therefore

$$R'(t) = \partial_1 R(t,t) + \partial_2 R(t,t) = 2\partial_2 R(t,t) \tag{5.31}$$

so that (using that $R(0,\cdot)=0$) Remark 5.4 does not apply and we may write

$$\begin{aligned}&\int_{0<s<t<T}\big((J_{ts}V)(Y_s)-V(Y_t)\big)^k_{\alpha\beta}R^{\alpha\beta}(\mathrm{d}s,\mathrm{d}t)\\ &=\int_{0<s<t<T}(J_{ts}V)^k_{\alpha\beta}(Y_s)R^{\alpha\beta}(\mathrm{d}s,\mathrm{d}t)-\int_0^T V^k_{\alpha\beta}(Y_t)R^{\alpha\beta}(t,\mathrm{d}t)\\ &=\int_{0<s<t<T}(J_{ts}V)^k_{\alpha\beta}(Y_s)R^{\alpha\beta}(\mathrm{d}s,\mathrm{d}t)-\frac{1}{2}\int_0^T V^k_{\alpha\beta}(Y_t)R^{\alpha\beta}(\mathrm{d}t)\end{aligned}$$

and (5.27) reduces to

$$\int_0^T Y^k_\gamma \mathrm{d}\boldsymbol{X}^\gamma=\int_0^T Y^k_\gamma \delta X^\gamma+\int_{0<s<t<T}(J_{ts}V)^k_{\alpha\beta}(Y_s)R^{\alpha\beta}(\mathrm{d}s,\mathrm{d}t) \tag{5.32}$$

Assumption 2. We will need to assume that (5.27) (which includes Skorokhod-integrability of the solution and (2D) Young integrability of the other terms) holds in a slightly more general case than is stated in [CL20]:

1. We need to allow for V to be linear (as opposed to only bounded);
2. We need to consider an integrand Y which is a piecewise solution to two RDEs, i.e.

$$\mathrm{d}Y^k_{\alpha;t}=\begin{cases}{}^1V^k_{\alpha\beta}(Y_t)\mathrm{d}\boldsymbol{X}^\beta_t,\ Y_0=y_0, & 0\le t\le S\\ {}^2V^k_{\alpha\beta}(Y_t)\mathrm{d}\boldsymbol{X}^\beta_t,\ Y_S=Y_S^-, & S\le t\le T\end{cases} \tag{5.33}$$

   where we intend (5.27) to hold with

$$V(Y_t)=\begin{cases}{}^1V(Y_t), & 0\le t\le S\\ {}^2V(Y_t), & S\le t\le T\end{cases},\quad J(Y_t)=\begin{cases}{}^1J_{ts}(Y_s), & 0\le s\le t\le S\\ {}^2J_{tS}(Y_S){}^1J_{Ss}(Y_s), & 0\le s\le S\le t\le T\\ {}^2J_{ts}(Y_s), & S\le s\le t\le T\end{cases} \tag{5.34}$$

   where ${}^1J$ and ${}^2J$ the Jacobians of the flows of the respective RDEs.

The first generalisation above will be needed to consider the RDE for the signature, and the second will be necessary to do so when the signature is started at some $S\ge 0$. Indeed, it is not possible to simply apply the original formula to an RDE started at $S>0$ since $X_{\cdot\ge S}\mid\mathcal{F}_S$ is no longer zero mean, as confirmed by the presence of the additional term in the next formula: this is obtained by subtracting (5.27) written from 0 to S from the same formula written from 0 to T, with V,J as in (5.34)

$$\begin{aligned}\int_S^T Y^k_\gamma \mathrm{d}\boldsymbol{X}^\gamma=&\int_S^T Y^k_\gamma\delta X^\gamma+\frac{1}{2}\int_S^T V^k_{\alpha\beta}(Y_t)R^{\alpha\beta}(\mathrm{d}t)\\ &+\int_{S<s<t<T}\big((J_{ts}V)(Y_s)-V(Y_t)\big)^k_{\alpha\beta}R^{\alpha\beta}(\mathrm{d}s,\mathrm{d}t)\\ &+\int_{0<s<S<t<T}\big((J_{ts}V)(Y_s)-V(Y_t)\big)^k_{\alpha\beta}R^{\alpha\beta}(\mathrm{d}s,\mathrm{d}t)\end{aligned} \tag{5.35}$$

Now, the last integral may be split, and we have

$$\int_{0<s<S<t<T} \big((J_{ts}V)(Y_s) - V(Y_t)\big)^k_{\alpha\beta} R^{\alpha\beta}(\mathrm{d}s, \mathrm{d}t) \tag{5.36}$$

$$= \int_{0<s<S<t<T} (J_{ts}V)^k_{\alpha\beta}(Y_s) R^{\alpha\beta}(\mathrm{d}s, \mathrm{d}t) - \int_{0<s<S<t<T} V^k_{\alpha\beta}(Y_t) R^{\alpha\beta}(S, \mathrm{d}t) \tag{5.37}$$

***Assumption* 3.** The simplification (5.32) is not mentioned in [CL19] and must shown. Also, it should be clarified whether it holds for all processes with a.a. paths of bounded $2 > p$-variation. This could also be handled directly for our final results and be omitted for the moment. Similarly, it we must show that (for the most general type of Gaussian process considered here) the separation of integrals performed in (5.36) is admissible. The point is that the integral being split is improper only at a point and not on a diagonal, which should be enough for it to converge. Examples involving $(1/3, 1/2) \ni H$-fBm confirm this: for instance, the integral $\int_{0<s<t<T} R(\mathrm{d}s, \mathrm{d}t) = \int_{0<s<t<T} (t-s)^{2H-2} \mathrm{d}s\mathrm{d}t$ does not converge, but

$$\int_{0<s<S<t<T} (t-s)^{2H-2} \mathrm{d}s\mathrm{d}t = (2H-1)^{-1} \int_0^S \big[(T-s)^{2H-1} - (S-s)^{2H-1}\big] \mathrm{d}s$$

does. On the other hand, when $H > 1/2$ the integral $\int_{0<s<t<T} (t-s)^{2H-2} \mathrm{d}s\mathrm{d}t$ does converge, which is what motivates our first assertion. See Remark 5.13 below for a proper introduction to these kinds of computation.

The second case of the following definition is motivated by (5.31), since in general the covariance function is not differentiable on the diagonal:

$$R(\Delta(u,v), \mathrm{d}w) := \begin{cases} R(v, \mathrm{d}w) - R(u, \mathrm{d}w) & \text{for } v \neq w \\ \frac{1}{2} R(\mathrm{d}w) - R(u, \mathrm{d}w) & \text{for } v = w \end{cases} \tag{5.38}$$

which, by smoothness and (5.28), is equal to $\partial_2 R(\Delta(u,v), w)\mathrm{d}w$ with

$$\partial_2 R(\Delta(u,v), w) := \begin{cases} \partial_2 R(v, w) - \partial_2 R(u, w) & \text{for } v \neq w \\ \frac{1}{2} R'(w) - \partial_2 R(u, w) & \text{for } v = w \end{cases} \tag{5.39}$$

We have shown

Theorem 5.5 (Rough-Skorokhod integral conversion formula from S to T)**.** *With V and J as in* (5.34) *we have*

$$\begin{aligned} \int_S^T Y^k_\gamma \mathrm{d}\boldsymbol{X}^\gamma = \int_S^T Y^k_\gamma \delta X^\gamma &+ \int_S^T V^k_{\alpha\beta}(Y_t) R^{\alpha\beta}(\Delta(S,t), \mathrm{d}t) \\ &+ \int_{S<s<t<T} ((J_{ts}V)(Y_s) - V(Y_t))^k_{\alpha\beta} R^{\alpha\beta}(\mathrm{d}s, \mathrm{d}t) \\ &+ \int_{0<s<S<t<T} (J_{ts}V)^k_{\alpha\beta}(Y_s) R^{\alpha\beta}(\mathrm{d}s, \mathrm{d}t) \end{aligned} \tag{5.40}$$

If the simplification discussed in Assumption 3 holds, this reduces to

$$\begin{aligned}\int_S^T Y_\gamma^k \mathrm{d}\boldsymbol{X}^\gamma = \int_S^T Y_\gamma^k \delta X^\gamma &+ \int_{S<s<t<T} (J_{ts}V)_{\alpha\beta}^k (Y_s)(\mathrm{d}s, \mathrm{d}t) \\ &+ \int_{0<s<S<t<T} (J_{ts}V)_{\alpha\beta}^k (Y_s) R^{\alpha\beta}(\mathrm{d}s, \mathrm{d}t)\end{aligned} \tag{5.41}$$

Moreover, if the first RDE is the trivial one ${}^1V = 0$*, the last integral in both expressions above vanishes.*

Example 5.6 (Gaussian semimartingales)**.** Assume X is additionally a semimartingale with decomposition $X = M + A$, M a continuous local martingale and A a process of bounded variation; in fact, by [BO10, Theorem 4.5] A and M are both Gaussian and the latter is a true martingale, and in particular a martingale in its own filtration $\mathcal{F}^M$. In general the increments of a martingale are orthogonal in L^2, and Gaussianity implies they are independent, so $M_{st} \perp\!\!\!\perp \mathcal{F}_s^M$ for $0 \leq s \leq t \leq T$. Therefore, denoting R_M the covariance function of M, we have

$$\begin{aligned}E[M_t^\alpha M_t^\beta - R_M^{\alpha\beta}(t))|\mathcal{F}_s^M] &= M_s^\alpha M_s^\beta + E[(M^\alpha M^\beta)_{st}|\mathcal{F}_s^M] - R_M^{\alpha\beta}(t) \\ &= M_s^\alpha M_s^\beta + E[(M^\alpha M^\beta)_{st}] - R_M^{\alpha\beta}(t) \\ &= M_s^\alpha M_s^\beta + (R_M^{\alpha\beta}(t) - R_M^{\alpha\beta}(s)) - R_M^{\alpha\beta}(t) \\ &= M_s^\alpha M_s^\beta - R_M^{\alpha\beta}(s)\end{aligned}$$

$R_M^{\alpha\beta}(\cdot)$ is thus the unique continuous process H s.t. $M_t^\alpha M_t^\beta - H_t$ is a local martingale, which implies $[X]_t = [M]_t = R_M(t)$. The classical Stratonovich-Itô conversion formula then reads

$$\int_S^T Y_\gamma^k \circ \mathrm{d}X^\gamma = \int_S^T Y_\gamma^k \mathrm{d}M^\gamma + \int_S^T Y_\gamma^k \mathrm{d}A^\gamma + \frac{1}{2}\int_S^T V_{\alpha\beta}^k(Y) R_M^{\alpha\beta}(\mathrm{d}t) \tag{5.42}$$

We compare this with (5.27): since $\boldsymbol{X}$ is now given by Stratonovich integration, it is well known that the LHSs are equal. While the individual terms on the right are not pairwise equal, in each equation the first term on the RHS has zero mean. If X is a (local) martingale then $X = M$, $A = 0$, $R_M = R$, and $R(\mathrm{d}s, \mathrm{d}t) = 0$ on $s < t$ again by L^2-orthogonality of martingale increments (the setup for Skorokhod integration is very similar to the classical white noise case [Nua06, §. 1.3.2]). This implies $\int Y \delta X = \int Y \mathrm{d}X = \int Y \mathrm{d}M$ (as can also be seen directly, for more general integrands), and the two conversion formulae are one and the same. For general centred Gaussian semimartingales, however, (5.40) and (5.42) yield distinct ways of representing the rough/Stratonovich integral of Y against X as the sum of a zero mean random variable and corrections given by pathwise-defined integrals. Even though the classical Itô-Stratonovich seems advantageous, both because it is more concise and because the zero-mean term is additionally a martingale, we will see that the rough-Skorokhod formula is better suited for out purposes.

Example 5.7 (Fractional Brownian motion)**.** Arguably the best known example of a stochastic process that is not semimartingale is *fractional Brownian motion* with Hurst parameter $H \in (0, 1)$ (H-fBm), introduced in [MN68]. It is a scalar centred Gaussian process with covariance function

$$R(s, t) = \frac{1}{2}(t^{2H} + s^{2H} - (t - s)^{2H}), \quad s \leq t \tag{5.43}$$

and, still for $s \leq t$ the integrators of interest are given by

$$R(\mathrm{d}s, \mathrm{d}t) = H(2H-1)(t-s)^{2H-2}\mathrm{d}s\mathrm{d}t, \quad R(\Delta(s,t), \mathrm{d}t) = H(t-s)^{2H-1}\mathrm{d}t \tag{5.44}$$

When $H \in (1/3, 1)$ fBm satisfies the hypotheses required in Section 5.1; in particular, H-fBm is a.s. α-Hölder regular for all $\alpha > H$. We call a d-dimensional H-fBm a vector of d independent scalar H-fBms.

Example 5.8 (The Riemann-Liouville process)**.** Another centred continuous Gaussian process, originally introduced in [Lév53] and subsequently in [MN68], is the *Riemann-Liouville process* with Hurst parameter $H \in (0, 1)$ (sometimes called type-II fBm), is a centred Gaussian process with covariance function [MR99, p.116-117]

$$\begin{aligned} R(s,t) &\underset{s<t}{=} \frac{1}{2}\left[t^{2H} + s^{2H} - 2H(t-s)^{2H}\left(\frac{1}{2H} + \int_0^{s/(t-s)} \left((1+u)^{H-1/2} - u^{H-1/2}\right)^2 \mathrm{d}u\right)\right] \\ R(t) &= t^{2H} \end{aligned} \tag{5.45}$$

Like fBm, this process specifies to Brownian motion when $H = 1/2$. Their main difference is that fBm has jointly stationary increments (see Example 5.9 below) while for the Riemann-Liouville process not even single increments are stationary. We were not able to find a concise expression for the derivatives of the covariance function of this process.

Example 5.9 (Stationarity and joint stationarity of increments)**.** X is stationary if and only if we may write

$$R(s,t) = \overline{R}(t-s) \tag{5.46}$$

for some function $\overline{R}\colon [0,T] \to \mathbb{R}^{d\times d}$. If $\overline{R}$ is smooth we may use (5.28) to write

$$\begin{aligned} & R(\mathrm{d}s, \mathrm{d}t) = -\overline{R}''(t-s)\mathrm{d}s\mathrm{d}t, \quad R(\mathrm{d}t) = 0, \quad R(s, \mathrm{d}t) = \overline{R}'(t-s)\mathrm{d}t \\ \Rightarrow \quad & R(\Delta(s,t), \mathrm{d}t) = -\overline{R}'(t-s)\mathrm{d}t \end{aligned} \tag{5.47}$$

An example of a centred stationary Gaussian process is the Ornstein-Uhlenbeck process $\exp(-t/2)W_{\exp(t)}$ where W is a Brownian motion and $t \in [0,T]$: its covariance function is $R(s,t) = \exp(-(t-s)/2)$ for $s \leq t$.

There is a weaker property that results in a similar simplification. We will say that a stochastic process X has *jointly stationary increments* if for all $s_1 \leq t_1, \ldots, s_n \leq t_n$ the distribution of the random vector of increments $(X_{s_1t_1}, \ldots, X_{s_nt_n})$ only depends on the differences $t_1 - s_1, \ldots, t_n - s_n$ and $s_2 - s_1, \ldots, s_n - s_{n-1}$ (if $n = 1$ the latter condition vanishes, and ordinary stationarity of increments is recovered). If X is Gaussian this need only be required for $n = 2$, and if it holds we may write

$$R\begin{pmatrix} s & u \\ t & v \end{pmatrix} = \mathbb{E}[X_{su} \otimes X_{tv}] = \widehat{R}(u-s, v-t, t-s) \tag{5.48}$$

for some function $\widehat{R}\colon [0,T]^3 \to \mathbb{R}^{d\times d}$. This property is satisfied by fBm, since if H is the Hurst parameter

we have

$$
\begin{aligned}
&R\begin{pmatrix} s & u \\ t & v \end{pmatrix} \\
&= \frac{1}{2}\big[(t-u)^{2H} + (v-s)^{2H} - (t-s)^{2H} - (v-u)^{2H}\big] \\
&= \frac{1}{2}\Big[\big((t-s)-(u-s)\big)^{2H} + \big((v-t)+(t-s)\big)^{2H} - \big(t-s\big)^{2H} - \big((v-t)+(t-s)-(u-s)\big)^{2H}\Big]
\end{aligned}
$$

If X has jointly stationary increments

$$
\partial_{12}R(s,t) = \lim_{\substack{u\to s \\ v\to t}} \frac{R\begin{pmatrix} s & u \\ t & v \end{pmatrix}}{(v-t)(u-s)} = \partial_{12}\widehat{R}(0,0,t-s) \tag{5.49}
$$

Although similar simplifications are not available for $\partial_2 R(s,t)$ and $R'(t)$ individually (as they are in the stationary case), they are once we consider the integrator $R(\Delta(s,t),\mathrm{d}t)$: indeed, we have

$$
\begin{aligned}
&R(\Delta(s,t),v) - R(\Delta(s,t),t) = R\begin{pmatrix} s & t \\ t & v \end{pmatrix} = \widehat{R}(t-s, v-t, t-s) \\
\implies\ &R(\Delta(s,t),\mathrm{d}t) = \partial_2\widehat{R}(t-s,0,t-s)\mathrm{d}t
\end{aligned}
$$

We therefore conclude that joint stationarity of increments, though a much more general property than stationarity, results in the same simplifications that are of relevance to this chapter, namely that $R(\mathrm{d}s,\mathrm{d}t)$ and $R(\Delta(s,t),\mathrm{d}t)$ only depend on $t-s$. This is because these are the only two deterministic integrators considered in all our end results.

5.3 The recursive formula

We are interested in applying the Skorokhod-rough integral formula to the RDE for the signature of $\boldsymbol{X}$ from S to T. We will be denoting the signature from S to T by $\boldsymbol{X}_{ST} \in T(\mathbb{R}^d)$, and its components $\boldsymbol{X}_{ST}^{\gamma_1,\ldots,\gamma_n} \coloneqq \langle \boldsymbol{X}_{ST}, \partial_{\gamma_1}\otimes\cdots\otimes\partial_{\gamma_n}\rangle$. Note that, although this is written as an RDE in an infinite-dimensional vector space, it can always be reduced to one in a finite-dimensional one by truncating the tensor algebra at an appropriately high order; for this reason, we will keep working with the whole tensor algebra, keeping in mind that all results pertaining to rough paths in finite-dimensional vector spaces apply. The first RDE will be the trivial one, i.e.

$$
{}^1V = 0,\ Y_0 = 1 \in T(\mathbb{R}^d)\otimes\mathbb{R}^d = \bigoplus_{n=1}^{\infty}(\mathbb{R}^d)^{\otimes n},\ J^{i,\alpha}_{j,\beta;ts} \equiv \delta^i_j\delta^\alpha_\beta \tag{5.50}
$$

Here δ is a Kronecker delta (note the difference between the characters δ, δ and δ, with the second being used for Skorokhod integration and with the third which will figure as an index). Now, since we have

$$
\boldsymbol{X}_{ST}^{\gamma_1,\ldots,\gamma_n} = \int_S^T \boldsymbol{X}_{St}^{\gamma_1,\ldots,\gamma_{n-1}}\mathrm{d}\boldsymbol{X}_t^{\gamma_n}
$$

we would like $Y^{\gamma_1,\ldots,\gamma_n}_{\gamma;t} = \boldsymbol{X}^{\gamma_1,\ldots,\gamma_{n-1}}_{St}\delta^{\gamma_n}_{\gamma}$, which solves

$$\begin{aligned} &\mathrm{d}_t Y^{\gamma_1,\ldots,\gamma_n}_{\gamma;St} = {}^2V^{(\gamma_1,\ldots,\gamma_n),\gamma}_{\delta}(Y_t)\mathrm{d}\boldsymbol{X}^{\delta}_t, \quad \text{with } {}^2V^{(\alpha_1,\ldots,\alpha_n),\alpha}_{\beta}(y) = y^{\alpha_1,\ldots,\alpha_{n-2},\alpha_n}_{\alpha}\delta^{\alpha_{n-1}}_{\beta} \\ &\text{i.e.} \quad {}^2V^{(\alpha_1,\ldots,\alpha_n)}_{\alpha\beta}(Y_t) = \boldsymbol{X}^{\alpha_1,\ldots,\alpha_{n-2}}_{St}\delta^{\alpha_{n-1}}_{\beta}\delta^{\alpha_n}_{\alpha} \end{aligned} \tag{5.51}$$

(5.29) now becomes, for $S \le s \le t \le T$

$$\begin{aligned} \mathrm{d}_t A^{(\alpha_1,\ldots,\alpha_n),\alpha}_{\beta;ts} &= \partial_{(\gamma_1,\ldots,\gamma_m),\gamma} V^{(\alpha_1,\ldots,\alpha_n),\alpha}_{\delta}(Y_t) A^{(\gamma_1,\ldots,\gamma_m),\gamma}_{\beta;ts}\mathrm{d}\boldsymbol{X}^{\delta}_t \\ &= \delta^{(\alpha_1,\ldots,\alpha_{n-2},\alpha_n)}_{(\gamma_1,\ldots,\gamma_m)}\delta^{\alpha}_{\gamma}\delta^{\alpha_{n-1}}_{\delta}A^{(\gamma_1,\ldots,\gamma_m),\gamma}_{\beta;ts}\mathrm{d}\boldsymbol{X}^{\delta}_t \\ &= A^{(\alpha_1,\ldots,\alpha_{n-2},\alpha_n),\alpha}_{\beta;ts}\mathrm{d}\boldsymbol{X}^{\alpha_{n-1}}_t, \quad A^{(\alpha_1,\ldots,\alpha_n),\alpha}_{\beta;ss} = \boldsymbol{X}^{\alpha_1,\ldots,\alpha_{n-2}}_{Ss}\delta^{\alpha_{n-1}}_{\beta}\delta^{\alpha_n,\alpha} \end{aligned}$$

We conclude that

$$({}^2J_{ts}V)^{(\alpha_1,\ldots,\alpha_n),\alpha}_{\beta}(Y_s) = \sum_{l=1}^{n-1}\boldsymbol{X}^{\alpha_1,\ldots,\alpha_{l-1}}_{Ss}\delta^{\alpha_l}_{\beta}\boldsymbol{X}^{\alpha_{l+1},\ldots,\alpha_{n-1}}_{st}\delta^{\alpha_n,\alpha} \tag{5.52}$$

by checking that it satisfies this RDE and initial condition. We have now proven the following theorem, which is the result of applying Theorem 5.5 to the signature (we replace s, t with u, v and S, T with s, t):

Theorem 5.10 (Recursive formula).

$$\begin{aligned} \boldsymbol{X}^{\gamma_1,\ldots,\gamma_n}_{st} = &\int_s^t \boldsymbol{X}^{\gamma_1,\ldots,\gamma_{n-1}}_{sv}\delta X^{\gamma_n}_v + \int_s^t \boldsymbol{X}^{\gamma_1,\ldots,\gamma_{n-2}}_{sv}R^{\gamma_{n-1}\gamma_n}(\Delta(s,v),\mathrm{d}v) \\ &+ \sum_{l=1}^{n-2}\int_{s<u<v<t}\boldsymbol{X}^{\gamma_1,\ldots,\gamma_{l-1}}_{su}\boldsymbol{X}^{\gamma_{l+1},\ldots,\gamma_{n-1}}_{uv}R^{\gamma_l,\gamma_n}(\mathrm{d}u,\mathrm{d}v) \\ &- \int_{s<u<v<t}(\boldsymbol{X}^{\gamma_1,\ldots,\gamma_{n-2}}_{sv} - \boldsymbol{X}^{\gamma_1,\ldots,\gamma_{n-2}}_{su})R^{\gamma_{n-1}\gamma_n}(\mathrm{d}u,\mathrm{d}v) \end{aligned} \tag{5.53}$$

If the simplification discussed in Assumption 3 holds, this reduces to

$$\boldsymbol{X}^{\gamma_1,\ldots,\gamma_n}_{st} = \int_s^t \boldsymbol{X}^{\gamma_1,\ldots,\gamma_{n-1}}_{sv}\delta X^{\gamma_n}_v + \sum_{l=1}^{n-1}\int_{s<u<v<t}\boldsymbol{X}^{\gamma_1,\ldots,\gamma_{l-1}}_{su}\boldsymbol{X}^{\gamma_{l+1},\ldots,\gamma_{n-1}}_{uv}R^{\gamma_l\gamma_n}(\mathrm{d}u,\mathrm{d}v) \tag{5.54}$$

***Assumption* 4.** In the above theorem we have taken finite sums out of the integral, following the only rule that, in all double integrals, the integrand must vanish on the diagonal $u = v$ (see Remark 5.4). This would not be the case if we separated the difference in the last integral, which is why it has been kept inside the integral sign, separately from the previous sum, in which all terms do vanish thanks to the fact that $\boldsymbol{X}^{\gamma_{l+1},\ldots,\gamma_{n-1}}_{uv}$ carries a positive number of indices. It must still be shown that all other integrals converge, since this does not follow automatically from the conversion formula. That this can be done can be understood by picking $\gamma_l = \gamma_n$ or $\gamma_{n-1} = \gamma_n$, and observing that the other terms must converge, but this argument must be made precise.

What we see from the above theorem is that the term

$$\int_{s<u<v<t}\boldsymbol{X}^{\gamma_1,\ldots,\gamma_{n-2}}_{su}R^{\gamma_l\gamma_n}(\mathrm{d}u,\mathrm{d}v)$$

which is present in the case of $p < 2$, but ill-defined in the general $p < 4$ case, is replaced with the term

$$\int_s^t \boldsymbol{X}_{sv}^{\gamma_1,\ldots,\gamma_{n-2}} R^{\gamma_{n-1}\gamma_n}(\Delta(s,v),\mathrm{d}v) - \int_{s<u<v<t} (\boldsymbol{X}_{sv}^{\gamma_1,\ldots,\gamma_{n-2}} - \boldsymbol{X}_{su}^{\gamma_1,\ldots,\gamma_{n-2}}) R^{\gamma_{n-1}\gamma_n}(\mathrm{d}u,\mathrm{d}v)$$

Remark 5.11 (An alternative recursive formula). It is tempting to try to improve (5.53) by applying the Chen identity to the last term:

$$\begin{aligned} \boldsymbol{X}_{st}^{\gamma_1,\ldots,\gamma_n} = &\int_s^t \boldsymbol{X}_{sv}^{\gamma_1,\ldots,\gamma_{n-1}} \delta X_v^{\gamma_n} + \int_s^t \boldsymbol{X}_{sv}^{\gamma_1,\ldots,\gamma_{n-2}} R^{\gamma_{n-1}\gamma_n}(\Delta(s,v),\mathrm{d}v) \\ &+ \sum_{l=1}^{n-2} \int_{s<u<v<t} \boldsymbol{X}_{su}^{\gamma_1,\ldots,\gamma_{l-1}} \boldsymbol{X}_{uv}^{\gamma_{l+1},\ldots,\gamma_{n-1}} R^{\gamma_l,\gamma_n}(\mathrm{d}u,\mathrm{d}v) \\ &- \sum_{l=0}^{n-3} \int_{s<u<v<t} \boldsymbol{X}_{su}^{\gamma_1,\ldots,\gamma_l} \boldsymbol{X}_{uv}^{\gamma_{l+1},\ldots,\gamma_{n-2}} R^{\gamma_{n-1}\gamma_n}(\mathrm{d}u,\mathrm{d}v) \end{aligned} \tag{5.55}$$

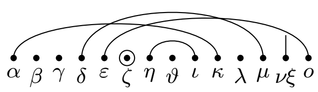
This formula has the advantage of not containing inseparable differences as integrands, but has the drawback of having more terms, some of which carry a negative sign.

Example 5.12 (Gaussian semimartingales)**.** The classical Stratonovich-Itô formula for Gaussian semimartingales (5.42) similarly yields an alternate formula (notation as in Example 5.6):

$$\boldsymbol{X}_{st}^{\gamma_1,\ldots,\gamma_n} = \int_s^t \boldsymbol{X}_{sv}^{\gamma_1,\ldots,\gamma_{n-1}} \mathrm{d}M_v^{\gamma_n} + \int_s^t \boldsymbol{X}_{sv}^{\gamma_1,\ldots,\gamma_{n-1}} \mathrm{d}A_v^{\gamma_n} + \frac{1}{2}\int_s^t \boldsymbol{X}_{sv}^{\gamma_1,\ldots,\gamma_{n-2}} R_M^{\gamma_{n-1}\gamma_n}(\mathrm{d}v) \tag{5.56}$$

5.4 The closed-form representation

Since solving the recursion of Theorem 5.10 will involve mixed iterated integrals, it will be helpful to develop some graphical notation that keeps track of all the different types of integrator. Consider diagrams formed by an ordered set of *nodes* with some *arcs* drawn (above the list of nodes) between pairs of nodes (the *left* and *right endpoints* of the arc), and straight lines (*edges*) beginning at some nodes (the *endpoint* of the edge) and going straight up. Each node may be the endpoint of at most one arc or edge (not both). We do not allow arcs between consecutive nodes: as is explained in Remark 5.13 below, this is what guarantees convergence, at least when $2 \leq p < 3$. Nodes that are not the endpoint of an arc or edge may be *circled*; if they are not they will be called *single*. Nodes are labelled with elements of $\{1,\ldots,d\}$: more precisely, an uncircled node has two labels if and only if it is the endpoint of an edge, and precisely one otherwise. We will refer to the last node of a diagram as its *terminal* node and say that the diagram *ends in/with* it, and if it is the endpoint of an arc or edge, we will call this the *leading* arc or edge, and say that the diagram *ends in/with* an arc/edge. The *degree* of a diagram is the number of labels (which is the same as the number of nodes plus the number of edges). We will call the set of all such diagrams $\mathcal{K}^d$, the set of those with m circled nodes $\mathcal{K}_m^d$, and $\mathcal{K}_m^d(\gamma_1,\ldots,\gamma_n)$ the subset of these of degree n labelled $\gamma_1,\ldots,\gamma_n$ from left to right (with double labels counted as two single labels). We also consider an empty diagram $\varnothing \in \mathcal{K}^d$. An example is



$$\in \mathcal{K}_1^d(\alpha,\ldots,o) \tag{5.57}$$

We now explain what these diagrams represent: single nodes stand for Skorokhod integration, arcs for double integration against the covariance function, edges for integration against $R(\Delta(u,v), \mathrm{d}v)$ (5.38), nodes that are circled for free $(\text{time variable}, \text{index})$ pairs (i.e. the resulting diagram will be a function of such times and indices) with the label standing for a Kronecker delta, and we are integrating over the simplex of order the number of nodes between two times $0 \le s \le t \le T$ to be specified. This approximate description, however, is not enough to give these random variables an unequivocal meaning: the problem is that mixed Skorokhod/Young integration on the simplex is not associative. We illustrate this with an example: omitting the labelling (or taking $d = 1$) for simplicity,

$$
\begin{aligned}
(\cdots)_{st} &= \int_{\Delta^4[s,t]} \partial_{12} R(u_2, u_4) \delta X_{u_1} \mathrm{d}u_2 \delta X_{u_3} \mathrm{d}u_4 \\
&= \int_s^t \int_s^{u_4} \int_{s<u_1<u_2<u_3} \partial_{12} R(u_2, u_4) \delta X_{u_1} \mathrm{d}u_2 \delta X_{u_3} \, \mathrm{d}u_4 \\
&= \int_s^t \delta_{u_3} \bigg(\mathbb{1}_{(s,u_4)}(u_3) \int_{s<u_1<u_2<u_3} \partial_{12} R(u_2, u_4) \delta X_{u_1} \mathrm{d}u_2 \bigg) \mathrm{d}u_4
\end{aligned}
$$

On the other hand, if we first resolve the Skorokhod integrals

$$
\begin{aligned}
(\cdots)_{st} &= \int_s^t \int_{s<u_2<u_4<t} \partial_{12} R(u_2, u_4) \int_s^{u_2} \delta X_{u_1} \int_{u_2}^{u_4} \delta X_{u_3} \, \mathrm{d}u_2 \mathrm{d}u_4 \\
&= \int_s^t \int_{s<u_2<u_4<t} \partial_{12} R(u_2, u_4) X_{su_2} X_{u_2u_4} \mathrm{d}u_2 \mathrm{d}u_4
\end{aligned}
$$

These two quantities cannot be equal, since taking their expectation we should expect to obtain

$$
\begin{aligned}
&\mathbb{E} \int_s^t \delta_{u_3} \bigg(\mathbb{1}_{(s,u_4)}(u_3) \int_{s<u_1<u_2<u_3} \partial_{12} R(u_2, u_4) \delta X_{u_1} \mathrm{d}u_2 \bigg) \mathrm{d}u_4 \\
&= \int_s^t \mathbb{E} \delta_{u_3} \bigg(\mathbb{1}_{(s,u_4)}(u_3) \int_{s<u_1<u_2<u_3} \partial_{12} R(u_2, u_4) \delta X_{u_1} \mathrm{d}u_2 \bigg) \mathrm{d}u_4 \\
&= 0 \\
&\neq \int_{\Delta^4[s,t]} \partial_{12} R(u_2, u_4) \partial_{12} R(u_1, u_3) \mathrm{d}u_1 \mathrm{d}u_2 \mathrm{d}u_3 \mathrm{d}u_4 \\
&= \mathbb{E} \int_s^t \int_{s<u_2<u_4<t} \partial_{12} R(u_2, u_4) X_{su_2} X_{u_2u_4} \mathrm{d}u_2 \mathrm{d}u_4
\end{aligned} \tag{5.58}
$$

with the $\neq$ is intended in general (when X is a martingale equality does actually hold, by orthogonality of its increments — this is precisely the case in which case Skorokhod integrals are Itô integrals, and mixed integration over the simplex is associative). As will be clear in a moment, the definition that we want to hold is the second one.

We now define a random variable C_{st} for $C \in \mathcal{K}^d$ and $0 \le s \le t \le T$. The definition we provide is a recursive one, not for C_{st} but for the differential $\text{đ}C_{st}$, which can be one of several types of differentials; we can then recover the former by

$$C_{st} := \begin{cases} 1 & C = \varnothing \\ \mathbb{1}^{\gamma}_{(s,t)}(\,\cdot\,)D_{s}. & C = D \underset{\gamma}{\odot}\ ,\ D \in \mathcal{K}^d \\ \displaystyle\int_s^t \text{đ}C_{sv} & \text{otherwise} \end{cases} \tag{5.59}$$

In the second case · is the time variable of which C_{st} is a function (and the index is evaluated against δ^{γ}). Note that, although $\text{đ}C_{sv}$ has no meaning on its own, it will always appear inside an integral as above. The meaning of đ depends on the type of terminal node of C:

1. $\text{đ}(C \underset{\gamma}{\bullet})_{sv} = C_{sv}\delta X^{\gamma}_{v}$;

2. $\displaystyle \text{đ}(\underset{\alpha}{\overset{\frown}{A\,\bullet}}\ \underset{\beta}{B\,\bullet})_{sv} := \int_{\substack{s<u<v \\ (w_1,\ldots,w_m)\in\Delta^m[s,u] \\ (z_1,\ldots,z_m)\in\Delta^m[u,v]}} \mathring{A}_{su}\mathring{B}_{uv}\prod_{k=1}^{m} R^{\gamma_k\delta_k}(\text{d}w_k,\text{d}z_k)R^{\alpha\beta}(\text{d}u,\text{d}v),$
   where $\mathring{A}$ and $\mathring{B}$ are the diagrams respectively given by taking A and B and circling all nodes that are endpoints of arcs (represented, in the initial diagram, by the dashed line) that intersect the leading arc. These nodes have associated (time variable, index) pairs $(w_1,\gamma_1),\ldots,(w_m,\gamma_m)$ for A and $(z_1,\delta_1),\ldots,(z_m,\delta_m)$ for B, in that order. All in all, this is an integral on $\Delta^{2m+1}[s,v]$ (along with the extra variable v, yet to be integrated);

3. $\text{đ}(C \underset{\alpha\beta}{\downarrow})_{sv} := \begin{cases} R^{\alpha\beta}(\Delta(s,v),\text{d}v) & C = \varnothing \\ \mathbb{1}^{\gamma}_{(s,v)}(\,\cdot\,)D_{s}.R(\Delta(\,\cdot\,,v),\text{d}v) & C = D \underset{\gamma}{\odot}\ ,\ D \in \mathcal{K}^d \\ \displaystyle\int_s^v \text{đ}C_{su}R^{\alpha\beta}(\Delta(u,v),\text{d}v) & \text{otherwise} \end{cases}$

Note how, in order to make sense of $\text{đ}C_{su}R^{\alpha\beta}(\Delta(u,v),\text{d}v)$, we can use smoothness and write $\partial_2 R^{\alpha\beta}(\Delta(u,v),v)\text{đ}C_{su}\text{d}v$, so that all instances of the variable u appear before $\text{d}u$; without the smoothness assumption this definition would be more difficult to make sense of. Following the assumptions of Assumption 1, we will continue by considering C_{st} elements of $\widetilde{\mathcal{H}}$ when $C \in \mathcal{K}^d$.

***Assumption* 5.** It must be shown that the integrals defined by (5.59) converge for the class of stochastic processes that we are considering. This task is closely related to that of Assumption 4, and the two should be handled together. Also, this should be accomplished in a way that makes it obvious that all manipulations of these integrals, e.g. those performed in Lemma 5.15 below, are sound. The next two remarks are meant to clarify some of these aspects, but the general task remains outstanding.

Remark 5.13. The main reason why this definition is so involved is the presence of edges, whose purpose is to replace arcs between consecutive nodes. The reason why these are disallowed is that, in general, they cause the integral not to converge. If it were possible to replace each endpoint of an edge with two nodes, and the edge with an arc between them, the above definition would not need to involve the multi-purpose differential đ and would be much simpler to state: we would only have to set

$$
\begin{aligned}
(\underset{\alpha}{\overset{A}{\bullet}}\!\frown\!\underset{\beta}{\overset{B}{\bullet}})_{st} &:= \int_{\substack{s<u<v<t\\(w_1,\ldots,w_m)\in\Delta^m[s,u]\\(z_1,\ldots,z_m)\in\Delta^m[u,v]}} \mathring{A}_{su}\mathring{B}_{uv}\prod_{k=1}^m R^{\gamma_k\delta_k}(\mathrm{d}w_k,\mathrm{d}z_k)R^{\alpha\beta}(\mathrm{d}u,\mathrm{d}v)\\
(C\underset{\gamma}{\bullet})_{st} &= \int_s^t C_{su}\delta X^\gamma_u\\
(D\underset{\gamma}{\odot})_{st} &= \mathbb{1}^\gamma_{(s,t)}(\,\cdot\,)D_s.
\end{aligned}
\tag{5.60}
$$

with all symbols as above. It is checked that, when these integrals converge, they coincide with the ones defined above, with arcs between consecutive nodes replaced with edges. This second definition is possible in certain cases, such as $1/2 < H$-fBm (see Assumption 3), but not for $1/2 \geq H$-fBm: for instance, we have (omitting labels)

$$
\begin{aligned}
(\bullet\!\frown\!\bullet)_{st} &= \int_{s<u<v<t} R(\mathrm{d}u,\mathrm{d}v)\\
&= H(2H-1)\int_{s<u<v<t}(v-u)^{2H-2}\mathrm{d}u\mathrm{d}v\\
&\overset{H\neq 1/2}{=} -H\int_s^t\Big[(v-u)^{2H-1}\Big]_{u=s}^{v}\mathrm{d}v
\end{aligned}
$$

in which the square bracket is only finite for $H > 1/2$, and the integral diverges to $-\infty$ when $H < 1/4$. On the other hand

$$
\begin{aligned}
(\overset{\downarrow}{\bullet})_{st} &= \int_s^t R(\Delta(s,u),\mathrm{d}u) = \frac{R(s)+R(t)}{2} - R(s,t)\\
&= H\int_s^t (u-s)^{2H-1}\mathrm{d}u = \frac{1}{2}(t-s)^{2H}
\end{aligned}
\tag{5.61}
$$

converges for $H > 1/4$ (and indeed for $H > 0$, but this is of little relevance, since the rough path is not defined for $H \leq 1/4$), and coincides with the above quantity when $H > 1/2$. At the threshold $H = 1/2$ both integrals converge, but to different values, and it is the latter that coincides with $\mathbb{E}\int_s^t X_{sv}\circ\mathrm{d}X_v = (t-s)/2$, while the first vanishes.

Remark 5.14 ($3 \leq p < 4$). It is well known that fBm can be lifted to a rough path even when $1/4 < H \leq 1/3$. Indeed, in [CL20] the same conversion formula as (5.27) is proved, with one caveat: the 2D integral is defined as a certain limit in L^2 and may not converge almost surely in the 2D Young sense (see [CL20, Remark 5.3]). A manifestation of this issue is that, when $1/4 < H \leq 1/3$, there is another problematic type of integral: using Hölder regularity of fBm (and a^- meaning $\forall b\; 0 < b < a$) the best we can do is

$$
\begin{aligned}
|(\bullet\!\frown\!\bullet\,\bullet)_{st}| &= \left|H(2H-1)\int_{s<u<v<t}X_{uv}(v-u)^{2H-2}\mathrm{d}u\mathrm{d}v\right|\\
&\lesssim \int_{s<u<v<t}(v-u)^{3H^--2}\mathrm{d}u\mathrm{d}v\\
&\lesssim \int_s^t\Big[(v-u)^{3H^--1}\Big]_{v=u}^{t}\mathrm{d}u
\end{aligned}
$$

and the square bracket is only finite for $H > 1/3$. A similar comments holds for the diagram ($\bullet\ \overset{\downarrow}{\bullet}$). We will say more about this problem at the end, since some of our final results actually seem not to have problems even for $1/4 < H \leq 1/3$, even though the arguments needed to reach them do.

We provide an example of an integral defined by (5.59). If C is the diagram (5.57), C_{st} is given by (recalling (5.39))

$$\int_{s<u_1<u_4<u_5<u_{10}<u_{12}<u_{14}} \partial_{12}R^{\alpha\kappa}(u_1,u_{10})\partial_{12}R^{\delta\mu}(u_4,u_{12})\partial_{12}R^{\varepsilon o}(\mathrm{d}u_5,\mathrm{d}u_{14})$$
$$(\odot \underset{\beta}{\bullet} \underset{\gamma}{\bullet} \odot)_{su_5}(\underset{\zeta}{\odot} \underset{\eta}{\bullet} \underset{\vartheta}{\bullet} \underset{\iota}{\bullet} \odot \underset{\lambda}{\bullet} \odot \underset{\nu\xi}{\overset{\downarrow}{\bullet}})_{u_5u_{14}}\mathrm{d}u_1\mathrm{d}u_4\mathrm{d}u_5\mathrm{d}u_{10}\mathrm{d}u_{12}\mathrm{d}u_{14}$$

with

$$(\odot \underset{\beta}{\bullet} \underset{\gamma}{\bullet} \odot)_{su_5} = \int_{s<u_1<u_2<u_3<u_4<u_5} \delta X^{\beta}_{u_2}\delta X^{\gamma}_{u_3},$$
$$(\underset{\zeta}{\odot} \underset{\eta}{\bullet} \underset{\vartheta}{\bullet} \underset{\iota}{\bullet} \odot \underset{\lambda}{\bullet} \odot \underset{\nu\xi}{\overset{\downarrow}{\bullet}})_{u_5u_{14}} = \delta^{\zeta}\int_{u_5<u_6<u_7<u_9<u_{10}<u_{12}<u_{13}<u_{14}} X^{\vartheta}_{u_7u_9}X^{\lambda}_{u_{10}u_{12}}$$
$$\partial_{12}R^{\eta\iota}(u_7,u_9)\partial_2R^{\nu\xi}(\Delta(u_{12},u_{13}),u_{13})\mathrm{d}u_7\mathrm{d}u_9\mathrm{d}u_{13}$$

where we have removed some labels from the circled nodes and instead placed them directly onto the R's. u_6 is the free time variable, with associated index to be evaluated against δ^{ζ}.

We now introduce the subset

$$\mathcal{I}^d \coloneqq \{C \in \mathcal{K}^d \mid \text{arcs do not intersect}\} \tag{5.62}$$

The notation regarding subscripts and indexing carries over. Before we proceed we will need the following technical lemma, in which we suppress the labelling (which does not interact with the statement).

Lemma 5.15. *For $C \in \mathcal{I}^d_0 = \mathcal{I}^d \cap \mathcal{K}^d_0$ we have*

$$\int_s^t C_{sv}R(\Delta(s,v),\mathrm{d}v) - \int_{s<u<v<t}(C_{sv}-C_{su})R(\mathrm{d}u,\mathrm{d}v) = \int_{s<u<v<t} \text{đ}C_{su}R(\Delta(u,v),\mathrm{d}v)$$

Proof. For $s<u<v<t$ we have

$$C_{sv}-C_{su} = \int_u^v \text{đ}C_{sw} \tag{5.63}$$

We check this by distinguishing the three cases for the last type of node in C and using standard additivity properties. For the case in which the last endpoint of C ends in an arc, using that this is not intersected by any other arcs (since $C \in \mathcal{I}^d$)

$$\begin{aligned}&(\,A \overset{\frown}{\bullet\, B\, \bullet}\,)_{sv} - (\,A \overset{\frown}{\bullet\, B\, \bullet}\,)_{su}\\ &= \int_{s<r<w<v} A_{sr}B_{rw}R(\mathrm{d}r,\mathrm{d}w) - \int_{s<r<w<u} A_{sr}B_{rw}R(\mathrm{d}r,\mathrm{d}w)\\ &= \int_{u<w<v}\int_{s<r<w} A_{sr}B_{rw}R(\mathrm{d}r,\mathrm{d}w)\end{aligned}$$

$$= \int_u^v \text{đ}(A \overset{\frown}{\bullet\; B \;\bullet})_{sw}$$

The case in which C ends in a node that is the endpoint of an edge is handled similarly:

$$\begin{aligned}
&(D \downarrow\!\bullet)_{sv} - (D \downarrow\!\bullet)_{su} \\
&= \int_s^v \int_s^w \text{đ} D_{sr} R(\Delta(r,w), \mathrm{d}w) - \int_s^u \int_s^w \text{đ} D_{sr} R(\Delta(r,w), \mathrm{d}w) \\
&= \int_{u<w<v} \int_{s<r<w} \text{đ} D_{sr} R(\Delta(r,w), \mathrm{d}w) \\
&= \int_u^v \text{đ}(D \downarrow\!\bullet)_{sw}
\end{aligned}$$

The case in which the last node is single follows trivially from additivity of the Skorokhod integral. Using (5.63), we have

$$\begin{aligned}
&\int_s^t C_{sv} R(\Delta(s,v), \mathrm{d}v) - \int_{s<u<v<t} (C_{sv} - C_{su}) R(\mathrm{d}u, \mathrm{d}v) \\
&= \frac{1}{2} \int_s^t C_{sv} R(\mathrm{d}v) - \int_s^t C_{sv} R(s, \mathrm{d}v) - \int_{s<u<w<v<t} \text{đ} C_{sw} R(\mathrm{d}u, \mathrm{d}v) \\
&= \frac{1}{2} \int_s^t C_{sv} R(\mathrm{d}v) - \int_s^t C_{sv} R(s, \mathrm{d}v) - \int_{s<w<v<t} \text{đ} C_{sw} R(\Delta(s,w), \mathrm{d}v) \\
&= \frac{1}{2} \int_s^t C_{sv} R(\mathrm{d}v) - \int_s^t C_{sv} R(s, \mathrm{d}v) + \int_{s<w<v<t} \text{đ} C_{sw} R(s, \mathrm{d}v) - \int_{s<w<v<t} \text{đ} C_{sw} R(w, \mathrm{d}v) \\
&= \frac{1}{2} \int_s^t C_{sv} R(\mathrm{d}v) - \int_{s<w<v<t} \text{đ} C_{sw} R(w, \mathrm{d}v)
\end{aligned}$$

concluding the proof. ■

We are now ready to prove the main theorem of this section.

Theorem 5.16 (Closed-form representation of the signature)**.**

$$\boldsymbol{X}_{st}^{\gamma_1,\ldots,\gamma_n} = \sum_{C \in \mathcal{I}_0^d(\gamma_1,\ldots,\gamma_n)} C_{st} \tag{5.64}$$

Proof. For the first statement we proceed by induction on n. For $n = 0, 1$ the statement is obvious ($\mathcal{I}_0^d(\gamma)$ is the singleton containing the diagram with a single node labelled γ). Assume the statement holds up to level $n-1$. By Theorem 5.10, the inductive hypothesis and Lemma 5.15 we have

$$\begin{aligned}
&\boldsymbol{X}_{st}^{\gamma_1,\ldots,\gamma_n} \\
&= \sum_{C \in \mathcal{I}_0^d(\gamma_1,\ldots,\gamma_{n-1})} \int_s^t C_{sv} \delta X_v^{\gamma_n} \\
&\quad + \sum_{\substack{l=1,\ldots,n-2 \\ A \in \mathcal{I}_0^d(\gamma_1,\ldots,\gamma_{l-1}) \\ B \in \mathcal{I}_0^d(\gamma_{l+1},\ldots,\gamma_{n-1})}} \int_{s<u<v<t} A_{su} B_{uv} R^{\gamma_l,\gamma_n}(\mathrm{d}u, \mathrm{d}v)
\end{aligned}$$

$$
+ \sum_{D \in \mathcal{I}_0^d(\gamma_1,\ldots,\gamma_{n-2})} \left[\int_s^t D_{sv} R^{\gamma_{n-1}\gamma_n}(\Delta(s,v), \mathrm{d}v) - \int_{s<u<v<t} (D_{sv} - D_{su}) R^{\gamma_{n-1}\gamma_n}(\mathrm{d}u, \mathrm{d}v) \right]
$$

$$
= \sum_{C \in \mathcal{I}_0^d(\gamma_1,\ldots,\gamma_{n-1})} (\, C \underset{\gamma_n}{\bullet} \,)_{st} + \sum_{\substack{l=1,\ldots,n-2 \\ A \in \mathcal{I}_0^d(\gamma_1,\ldots,\gamma_{l-1}) \\ B \in \mathcal{I}_0^d(\gamma_{l+1},\ldots,\gamma_{n-1})}} (\, A \, \overset{\frown}{\underset{\gamma_l}{\bullet} \, B \, \underset{\gamma_n}{\bullet}} \,)_{st} + \sum_{D \in \mathcal{I}_0^d(\gamma_1,\ldots,\gamma_{n-2})} (\, D \underset{\gamma_{n-1}\gamma_n}{\overset{\downarrow}{\bullet}} \,)_{st}
$$

$$
= \sum_{C \in \mathcal{I}_0^d(\gamma_1,\ldots,\gamma_n)} C_{st}
$$

concluding the proof. ■

Example 5.17 (Representation formulae for $n = 3$)**.** At level 3 the representation formula (5.64) reads

$$
\begin{aligned}
\boldsymbol{X}^{\alpha\beta\gamma}_{st} &= (\, \underset{\alpha}{\bullet} \, \underset{\beta}{\bullet} \, \underset{\gamma}{\bullet} \,)_{st} + (\, \underset{\alpha\beta}{\overset{\downarrow}{\bullet}} \, \underset{\gamma}{\bullet} \,)_{st} + (\, \underset{\alpha}{\bullet} \, \underset{\beta\gamma}{\overset{\downarrow}{\bullet}} \,)_{st} + (\, \overset{\frown}{\underset{\alpha}{\bullet} \, \underset{\beta}{\bullet} \, \underset{\gamma}{\bullet}} \,)_{st} \\
&= \int_{s<u<v<w<t} \delta X^\alpha_u \delta X^\beta_v \delta X^\gamma_w + \int_s^t \left(\frac{R^{\alpha\beta}(s) + R^{\alpha\beta}(u)}{2} - R^{\alpha\beta}(s,u) \right) \delta X^\gamma_u \\
&\quad + \int_{s<u<v<t} \partial_2 R^{\beta\gamma}(\Delta(u,v), \mathrm{d}u) \delta X^\alpha_{su} \mathrm{d}v + \int_{s<u<v<t} X^\beta_{uv} R^{\alpha\gamma}(\mathrm{d}u, \mathrm{d}v)
\end{aligned} \tag{5.65}
$$

where we have used the expression of $\int_s^v R(\Delta(s,u), \mathrm{d}u)$ as the difference between the average of the variances and the covariance (5.61).

In Remark 5.11 it was mentioned that a different method was available for obtaining a recursion that, similarly as done above, can be resolved to yield a closed-form expression of the signature. The symbols involved in this expression are a little easier to describe than the ones used, since we may simply take the modified definition (5.60), disallow arcs between consecutive nodes to ensure convergence, and introduce the additional symbol

$$
(\, \overline{\ulcorner C \, \underset{\alpha\beta}{\overset{\downarrow}{\bullet}}} \,)_{st} \coloneqq \int_s^t C_{sv} R^{\alpha\beta}(\Delta(s,v), \mathrm{d}v)
$$

that replaces edges. The resulting representation formula is analogous to (5.16) and easier to show (it does not require Lemma 5.15). For $n = 3$ it reads

$$
\begin{aligned}
\boldsymbol{X}^{\alpha\beta\gamma}_{st} &= (\, \underset{\alpha}{\bullet} \, \underset{\beta}{\bullet} \, \underset{\gamma}{\bullet} \,)_{st} + (\, \overline{\ulcorner \underset{\alpha\beta}{\overset{\downarrow}{\bullet}}} \, \underset{\gamma}{\bullet} \,)_{st} + (\, \overline{\ulcorner \underset{\alpha}{\bullet} \, \underset{\beta\gamma}{\overset{\downarrow}{\bullet}}} \,)_{st} + (\, \overset{\frown}{\underset{\alpha}{\bullet} \, \underset{\beta}{\bullet} \, \underset{\gamma}{\bullet}} \,)_{st} - (\, \overset{\frown}{\underset{\beta}{\bullet} \, \underset{\alpha}{\bullet} \, \underset{\gamma}{\bullet}} \,)_{st} \\
&= \int_{s<u<v<w<t} \delta X^\alpha_u \delta X^\beta_v \delta X^\gamma_w + \int_s^t \left(\frac{R^{\alpha\beta}(s) + R^{\alpha\beta}(u)}{2} - R^{\alpha\beta}(s,u) \right) \delta X^\gamma_u \\
&\quad + \int_s^t X^\alpha_{sv} R^{\beta\gamma}(s, \mathrm{d}v) + \int_{s<u<v<t} X^\beta_{uv} R^{\alpha\gamma}(\mathrm{d}u, \mathrm{d}v) - \int_{s<u<v<t} X^\alpha_{uv} R^{\beta\gamma}(\mathrm{d}u, \mathrm{d}v)
\end{aligned}
$$

(which can also be checked directly to be equal to (5.65)). The problem with this alternative formula is that it becomes quite complex for higher n, and while it is still possible to write it in closed form, we could not find a satisfactory way to collect all the terms generated when passing to the expectation in it. For this reason, we will continue to use Theorem 5.16, which is easier to work with for this purpose.

Example 5.18 (Gaussian semimartingales)**.** If X is a Gaussian semimartingale, we can alternatively use

Example 5.12 to obtain an expression for the signature in terms of iterated Itô integrals instead of Skorokhod integrals (notation as in Example 5.6):

$$\boldsymbol{X}_{st}^{\gamma_1,\ldots,\gamma_n} = \sum_{c_1\ldots c_m = \gamma_1\ldots\gamma_n} \int_{\Delta^m[s,t]} \mathrm{d}Z_{u_1}^{c_1}\cdots \mathrm{d}Z_{u_m}^{c_m} \tag{5.66}$$

where we define

$$\mathcal{C}^d \coloneqq (\{1,\ldots,d\}\times\{\{0\},\{1\}\}) \sqcup \{1,\ldots,d\}^2 \tag{5.67}$$

and

$$\text{for } c\in\mathcal{C}^d \quad Z_t^c \coloneqq \begin{cases} M_t^\gamma & c = (\gamma,\{0\}) \\ A_t^\gamma & c = (\gamma,\{1\}) \\ \frac{1}{2}R_M^{\alpha\beta}(t) & c = (\alpha,\beta) \end{cases} \tag{5.68}$$

and the identity $c_1\ldots c_m = \gamma_1\ldots\gamma_n$ means we are summing over all distinct words $c_1\ldots c_m$ over the alphabet $\mathcal{C}^d$ with the property that, stripping each c_i of its second factor if it belongs to the first term in the disjoint union of (5.67), or juxtaposing its two entries in $\{1,\ldots,d\}$ if it belongs to the second, we obtain the word $\gamma_1,\ldots,\gamma_n$. For example, $c_1\ldots c_m = \alpha\beta\gamma$ means we are considering the following 12 possibilities for $c_1\ldots c_m$:

$$\begin{aligned} &c_1 = (\alpha,i), c_2 = (\beta,j), c_3 = (\gamma,k) \text{ with } i,j,k\in\{0,1\}; \\ &c_1 = (\alpha,i), c_2 = \beta\gamma \text{ with } i\in\{0,1\}; \\ &c_1 = \alpha\beta, c_2 = (\gamma,j) \text{ with } j\in\{0,1\} \end{aligned}$$

When X is additionally a martingale ($A=0$) we obtain a formula analogous to that obtained for Brownian motion in [Bau04, Proposition 2.4] which is the same formula obtained by using Theorem 5.16, essentially because Skorokhod and Itô integration agree and $R_M = R$.

Example 5.19 (The Brownian bridge)**.** We illustrate the difference between the two representation formulae Theorem 5.16 and Example 5.18 with an example of a centred Gaussian semimartingale that is not a martingale: let X be a d-dimensional Brownian bridge returning to the origin at time 1. This is a centred Gaussian process with covariance function

$$R^{\alpha\beta}(s,t) = s(1-t)\delta^{\alpha\beta}, \quad 0\le s\le t\le 1 \tag{5.69}$$

The general formula for Gaussian processes should be therefore read by substituting in (5.64)

$$R^{\alpha\beta}(\mathrm{d}s,\mathrm{d}t) = -\delta^{\alpha\beta}\mathrm{d}s\mathrm{d}t, \quad R^{\alpha\beta}(\Delta(s,t),\mathrm{d}t) = \frac{1+2(s-t)}{2}\mathrm{d}t \tag{5.70}$$

Note that these are polynomials, and not difficult to integrate explicitly.

As for the formula for Gaussian semimartingales, we use the well-known expression for X as an Itô process:

$$X_t \coloneqq \begin{cases} W_t + \int_0^t\int_0^s \frac{\mathrm{d}W_u}{u-1}\mathrm{d}s & t\in[0,1) \\ 0 & t=1 \end{cases} \tag{5.71}$$

We can therefore substitute

$$dM_t^\gamma = dW_t^\gamma, \quad dA_t^\gamma = \int_0^t \frac{dW_u^\gamma}{u-1} dt, \quad R_M^{\alpha\beta}(dt) = \delta^{\alpha\beta} dt \tag{5.72}$$

in (5.66).

5.5 Passing to the expectation

The main ingredient that will be used in the computation of expectations is a formula that generalises the Itô-Skorokhod isometry, to the case of n factors. To do this, we introduce some notation that will be used in the proof of this result. We will use *placeholders* $\bullet, \star, \diamond \ldots$ to denote (time, index) pairs, that will be integrated away when passing to the scalar product on $\mathcal{H}$. The precise position of the placeholders will determine which variables are paired. For example, given $\mathbb{R}^d$-valued processes H, K, to denote

$$\int_{[0,T]^4} \mathbb{E}[\mathcal{D}_{\gamma;u} H_{\alpha;s} \mathcal{D}_{\delta;v} K_{\beta;t}] R^{\alpha\beta}(ds, dt) R^{\gamma\delta}(du, dv)$$

we will use the compact notation

$$\langle \mathbb{E}[\mathcal{D}_\bullet H_\star \mathcal{D}_\bullet K_\star] \rangle_{\bullet,\star}$$

The 2-factor Itô-Skorokhod isometry [CL19, Theorem 4.8] can then be written

$$\mathbb{E}[\delta H \delta K] = \langle \mathbb{E}[H_\bullet K_\bullet] \rangle_\bullet + \langle \mathbb{E}[\mathcal{D}_\bullet H_\star \mathcal{D}_\star K_\bullet] \rangle_{\bullet,\star} \tag{5.73}$$

Note the different pairing of the placeholders in the second term above and in the previous example. Sometimes, when we want to use a single symbol to denote multiple placeholders, we will use tuples of placeholders, denoted by underlying a symbol: for instance, when considering the order-m Malliavin derivative we may write $\mathcal{D}_{\underline{\star}}$ to mean $\mathcal{D}_{\star_1 \ldots \star_m}$. In this case we will not directly specify the order of the derivative, which is determined by the number of placeholders, in this case m. The next proposition expresses the expectation of a product of Skorokhod integrals as a multiple integral of the Malliavin derivatives of the integrands.

***Assumption* 6.** Here and below we pass to the expectation inside 2D Young integrals. That this is possible is non-trivial, and must be shown. In fact, even the referenced case of two factors (5.73) appears with the expectation outside the integral. While this is not essential for the proposition below (which could even be stated for abstract Wiener spaces), it will be necessary when the statement is applied. One option is to first establish that the 2D Young integrals are not only an a.s. limit, but a limit in L^2 as well (cf. [CL19, Proposition 4.10, Theorem 6.1], [CL20, Theorem 5.1]). Another is to use that all integrals avoid the diagonal and therefore are Lebesgue (5.9), so that one can try and apply Fubini's theorem.

Furthermore, the precise integrability hypotheses in the proposition below, which are similar to those of the 2-factor case [Nua06, Proposition 1.3.1], are likely required to be made more precise, especially in view of the above commutation requirement.

Proposition 5.20 (n-factor Itô-Skorokhod isometry)**.** *Given $H^1, \dots, H^n \in \mathbb{D}^{n,2}(\mathcal{H})$ it holds that*

$$\mathbb{E}[\delta H^1 \cdots \delta H^n] = \sum_{\substack{\bullet \frown \star \\ \bullet \frown \bullet}} \langle \mathbb{E}[\mathcal{D}_{\underline{\star}^1} H^1_{\bullet_1} \cdots \mathcal{D}_{\underline{\star}^n} H^n_{\bullet_n}] \rangle_{\underline{\bullet},\underline{\star}} \tag{5.74}$$

where $\underline{\star}^k$ is an unordered tuple $\star^k_1 \dots \star^k_{m_k}$ (of variable length m_k) for $k = 1, \dots, n$, $\bullet_1, \dots, \bullet_n$ are single placeholders, and we are summing over all possible combinations of pairings between placeholders (without taking the order of the $\underline{\star}^k$'s into account) s.t.

- *All placeholders are paired;*
- *Each pairing (denoted $\frown$) is of one of the following two types: $\star^i_l \frown \bullet_j$ with $i \neq j$ or $\bullet_h \frown \bullet_k$ with $h \neq k$.*

A few clarifying remarks are in order before beginning the proof. Note that there is a fixed number of $\bullet$ placeholders, n, while the number of $\star$ placeholders (or more precisely the length of each $\underline{\star}^k$) varies across terms in the sum. This number is bounded by n, since the $\star$'s may only be paired with the $\bullet$'s, and not with other $\star$'s (in particular, the sum is finite). A $\underline{\star}^k$ tuple may have length zero, in which case the corresponding term is just the undifferentiated $H^k_{\bullet_k}$. For $n = 2$ this identity reduces to (5.73), while for higher n it rapidly increases in complexity (see Example 5.21 below for the case of 3 factors, which involves 14 summands).

Proof of Proposition 5.20. We prove the following slight generalisation in which $Z \in \mathbb{D}^{n,2}$

$$\mathbb{E}[Z\delta H^1 \cdots \delta H^n] = \sum_{\substack{\bullet \frown \star \\ \bullet \frown \bullet}} \langle \mathbb{E}[\mathcal{D}_{\underline{\star}^0} Z \mathcal{D}_{\underline{\star}^1} H^1_{\bullet_1} \cdots \mathcal{D}_{\underline{\star}^n} H^n_{\bullet_n}] \rangle_{\underline{\bullet},\underline{\star}} \tag{5.75}$$

(the sum is defined similarly, with the inclusion of $\underline{\star}^0$ tuple) by induction on n; taking $Z = 1$ then yields the statement of the lemma. For $n = 0$ the statement is trivial. For the inductive step, using the adjoint property of the Skorokhod integral (5.15), Heisenberg's commutativity relation (5.18) and the inductive hypothesis applied with different choices of the Z term, we have

$$\begin{aligned}
&\mathbb{E}[Z\delta H^1 \cdots \delta H^{n+1}] \\
&= \langle \mathbb{E}[\mathcal{D}_\diamond Z \cdot \delta H^1 \cdots \delta H^n \cdot H^{n+1}_\diamond] \rangle_\diamond + \sum_{k=1}^n \langle \mathbb{E}[Z\delta H^1 \cdots \delta H^{k-1} \cdot \mathcal{D}_\diamond \delta H^k \cdot \delta H^{k+1} \cdots \delta H^n \cdot H^{n+1}_\diamond] \rangle_\diamond \\
&= \langle \mathbb{E}[(\mathcal{D}_\diamond Z H^{n+1}_\diamond)\delta H^1 \cdots \delta H^n] \rangle_\diamond \\
&\quad + \sum_{k=1}^n \langle \mathbb{E}[(Z H^{n+1}_\diamond)\delta H^1 \cdots \delta H^{k-1} \cdot \delta D_\diamond H^k \cdot \delta H^{k+1} \cdots \delta H^n] \rangle_\diamond \\
&\quad + \sum_{k=1}^n \langle \mathbb{E}[(Z H^k_\diamond H^{n+1}_\diamond)\delta H^1 \cdots \widehat{\delta H}^k \cdots \delta H^n] \rangle_\diamond \\
&= \sum_{\substack{\bullet \frown \star \\ \bullet \frown \bullet}} \langle \mathbb{E}[\mathcal{D}_{\underline{\star}^0}(\mathcal{D}_\diamond Z H^{n+1}_\diamond)\mathcal{D}_{\underline{\star}^1} H^1_{\bullet_1} \cdots \mathcal{D}_{\underline{\star}^n} H^n_{\bullet_n}] \rangle_{\diamond,\underline{\bullet},\underline{\star}} \\
&\quad + \sum_{\substack{k=1,\dots,n \\ \bullet \frown \star \\ \bullet \frown \bullet}} \langle \mathbb{E}[\mathcal{D}_{\underline{\star}^0}(Z H^{n+1}_\diamond)\mathcal{D}_{\underline{\star}^1} H^1_{\bullet_1} \cdots \mathcal{D}_{\underline{\star}^{k-1}} H^{k-1}_{\bullet_{k-1}} \cdot \mathcal{D}_{\underline{\star}^k,\diamond} H^k_{\bullet_k} \cdot \mathcal{D}_{\underline{\star}^{k+1}} H^{k+1}_{\bullet_{k+1}} \cdots \mathcal{D}_{\underline{\star}^n} H^n_{\bullet_n}] \rangle_{\diamond,\underline{\bullet},\underline{\star}}
\end{aligned}$$

$$+ \sum_{\substack{k=1,\dots,n \\ \bullet\frown\star \\ \bullet\frown\bullet}} \langle \mathbb{E}[\mathcal{D}_{\underline{\star}^0}(ZH^k_\diamond H^{n+1}_\diamond)\mathcal{D}_{\underline{\star}^1}H^1_{\bullet_1}\cdots\widehat{\mathcal{D}_{\underline{\star}^k}H^k_{\bullet_k}}\cdots\mathcal{D}_{\underline{\star}^n}H^n_{\bullet_n}]\rangle_{\diamond,\underline{\bullet},\underline{\star}}$$

$$= \sum_{\substack{\bullet\frown\star \\ \bullet\frown\bullet}} \langle \mathbb{E}[\mathcal{D}_{\underline{\star}^0,\diamond}Z\cdot\mathcal{D}_{\underline{\star}^{n+1}}H^{n+1}_\diamond\cdot\mathcal{D}_{\underline{\star}^1}H^1_{\bullet_1}\cdots\mathcal{D}_{\underline{\star}^n}H^n_{\bullet_n}]\rangle_{\diamond,\underline{\bullet},\underline{\star}}$$

$$+ \sum_{\substack{k=1,\dots,n \\ \bullet\frown\star \\ \bullet\frown\bullet}} \langle \mathbb{E}[\mathcal{D}_{\underline{\star}^0}Z\cdot\mathcal{D}_{\underline{\star}^{n+1}}H^{n+1}_\diamond\cdot\mathcal{D}_{\underline{\star}^1}H^1_{\bullet_1}\cdots\mathcal{D}_{\underline{\star}^{k-1}}H^{k-1}_{\bullet_{k-1}}\cdot\mathcal{D}_{\underline{\star}^k,\diamond}H^k_{\bullet_k}\cdot\mathcal{D}_{\underline{\star}^{k+1}}H^{k+1}_{\bullet_{k+1}}\cdots\mathcal{D}_{\underline{\star}^n}H^n_{\bullet_n}]\rangle_{\diamond,\underline{\bullet},\underline{\star}}$$

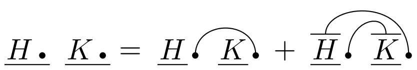
$$+ \sum_{\substack{k=1,\dots,n \\ \bullet\frown\star \\ \bullet\frown\bullet}} \langle \mathbb{E}[\mathcal{D}_{\underline{\star}^0}Z\cdot\mathcal{D}_{\underline{\star}^k}H^k_\diamond\cdot\mathcal{D}_{\underline{\star}^{n+1}}H^{n+1}_\diamond\cdot H^1_{\bullet_1}\cdots\widehat{\mathcal{D}_{\underline{\star}^k}H^k_{\bullet_k}}\cdots\mathcal{D}_{\underline{\star}^n}H^n_{\bullet_n}]\rangle_{\diamond,\underline{\bullet},\underline{\star}}$$

where in the last identity we have applied the iterated Leibniz rule for Malliavin derivatives (a consequence of the ordinary Leibniz rule (5.13))

$$\mathcal{D}_{\underline{\star}^0}(Y_1\cdots Y_n) = \sum_{(\underline{\star}^1,\dots,\underline{\star}^n)\in\mathrm{Sh}^{-1}(\underline{\star}^0)} \mathcal{D}_{\underline{\star}^1}Y_1\cdots\mathcal{D}_{\underline{\star}^n}Y_n$$

to the terms $\mathcal{D}_{\underline{\star}^0}(\mathcal{D}_\diamond ZH^{n+1}_\diamond)$, $\mathcal{D}_{\underline{\star}^0}(ZH^{n+1}_\diamond)$ and $\mathcal{D}_{\underline{\star}^0}(ZH^k_\diamond H^{n+1}_\diamond)$, relying on the initial sum of each line to regroup terms, with the sum now extending to pairings that include placeholders in $\underline{\star}^{n+1}$ (see Chapter 3 for the definition of unshuffles). Now the proof is concluded by a simple counting argument, which involves incorporating the $\diamond$ in the $\bullet$'s and $\star$'s. ■

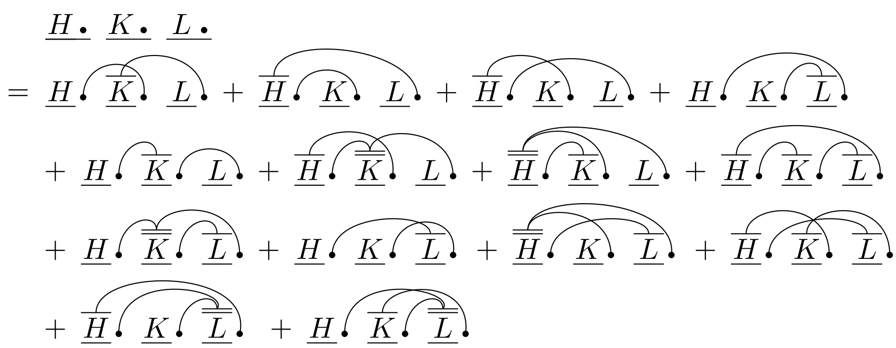
We will now transition back to the symbolic notation used in the previous section, consisting of nodes, arcs, edges, etc., to which we add two new elements. *Underlining* corresponds to taking the expectation, and when there are multiple terms that are underlined in the same expression, this means we are taking the expectation of their product. *Overlining* corresponds to taking Malliavin derivatives, with higher-order derivatives represented by multiple (commuting) overlines. For example, (5.73) becomes



(5.76)

Note how we have not labelled the two nodes: this reflects the fact that H and K are $\mathbb{R}^d$-valued. Somewhat more involved is the case of 3 factors:

Example 5.21 (3-factor Itô-Skorokhod isometry.). For H, K, L $\mathbb{R}^d$-valued processes, Proposition 5.20 reads



The LHS represents the expression $\mathbb{E}[\delta H \delta K \delta L]$, while the terms on the RHS represent double integrals of the expectation of products of order 0,1, and 2 Malliavin derivatives of H, K, L, with the pairings given by the arcs. For example, the second term on the third line (counting the LHS as the first line) represents the expression $\langle \mathbb{E}[(\mathcal{D}_\bullet H_\star)(\mathcal{D}_{\star\diamond} K_\bullet) L_\diamond] \rangle_{\bullet,\star,\diamond}$ in placeholder notation. An arc with endpoint on an overline represents a double integral where one $(\text{time}, \text{index})$ integration pair is the one belonging to the Malliavin derivative represented by the overline.

Since the formula of Proposition 5.20 generates expressions involving Malliavin derivatives, and since these are, in any case, present in Stroock's formula for the Wiener chaos expansion, we need to establish a formula for the Malliavin derivative of the random variables represented by elements of $\mathcal{K}^d$. For $C \in \mathcal{K}^d$, let $(C)_k$ denote the set of diagrams in $\mathcal{K}^d$ obtained by circling k single nodes that are single in C; this means $(C)_k \subseteq \mathcal{K}^d_{m+k}$.

Proposition 5.22 (Malliavin derivatives of diagrams)**.** *For $C \in \mathcal{K}^d$ we have*

$$\mathbb{1}_{\Delta^k[s,t]} \mathcal{D}^k C_{st} = \sum_{D \in (C)_k} D_{st} \tag{5.77}$$

In words, the k-the Malliavin derivative taken on the k-simplex is given by summing over all possible ways of circling k single nodes of C. Each resulting term will then be a function of $(u_1, \ldots, u_k) \in \Delta^k[s,t]$, respectively matched with k indices.

***Assumption 7*.** In the next proof and below, we use that it is possible to pass to the Malliavin derivative inside deterministic integrals. That this is possible is non-trivial, and must be shown. Our plan is to show that both the (2D-)Riemann-Stieljes approximations of the integrals and of their derivatives converge in L^2, and that indeed the latter (which are elementary to compute) converge to the integral of the Malliavin derivative. The statement will then follow from closedness of $\mathcal{D}$.

Proof of Proposition 5.22. To prove the case $k = 1$ we proceed by induction on the number of nodes of C. Once again, we omit the labelling, which is irrelevant to this task. The case $C = \varnothing$ is trivial. For the inductive step we distinguish cases with regards to the last node in C: passing to the Malliavin derivative inside deterministic integrals

$$\overline{C \ \downarrow \ldots \downarrow} = \overline{C} \ \downarrow \ldots \downarrow$$

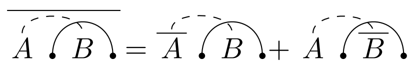
where there are m edges. More precisely, the diagram on the RHS, evaluated at $_{st}$, stands for

$$\int_{s<v_1<\ldots<v_m<t} \mathcal{D}\bigg(\int_s^{v_1} \partial_2 R(\Delta(u, v_1), v_1) \mathrm{d}C_{su} \bigg)$$
$$\partial_2 R(\Delta(v_1, v_2), v_2) \mathrm{d}v_1 \cdots \partial_2 R(\Delta(v_{m-1}, v_m), v_m) \mathrm{d}v_{m-1} \mathrm{d}v_m$$

if C ends in an uncircled node,

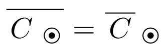
$$\int_{s<\cdot<v_1<\ldots<v_m<t} \mathcal{D}(B_{s\cdot}) \partial_2 R(\Delta(u, v_1), v_1)$$
$$\partial_2 R(\Delta(v_1, v_2), v_2) \mathrm{d}v_1 \cdots \partial_2 R(\Delta(v_{m-1}, v_m), v_m) \mathrm{d}v_{m-1} \mathrm{d}v_m$$

if $C = B \odot$ (with $\cdot$ standing for the free time variable corresponding to the circled node), and 0 if $C = \varnothing$. By the Leibniz rule for Malliavin derivatives (5.13) we have



(5.78)

where, as before, the dashed arc means we are allowing for arcs between nodes of A and nodes of B. The case of the terminal node being circled is similar:



(5.79)

For the case in which the terminal node is single, Heisenberg's commutativity relation (5.18) reads

(5.80)

In (5.78), (5.79) and (5.80) we must allow respectively B, C and C to be multiplied by a function $f(v)$, where v is the integration/free variable corresponding to the terminal node: this is because, in the induction, $f(v)$ must be be set to $\partial_2 R(\Delta(v,w),w)$ for some $w > v$. For example, the last case reads

$$\mathcal{D}_r \int_s^t C_{sv} f(v) \delta X_v = C_{sr} f(r) + \int_s^t \mathcal{D}_r C_{sv} f(v) \delta X_v$$

Using these facts and arguing as in the proof of Theorem 5.16 we complete the inductive step and obtain the formula for $k = 1$.

We now prove the general case by induction on k:

$$\begin{aligned} \mathcal{D}^{k+1} C_{st} &= [u_k < u_{k+1} < t] \mathcal{D}(\mathbb{1}_{\Delta^k[s,t]} \mathcal{D}^k C_{st}) \\ &= [u_k < u_{k+1} < t] \mathcal{D} \sum_{D \in (C)_k} D_{st} \\ &= [u_k < u_{k+1} < t] \sum_{\substack{D \in (C)_k \\ E \in (D)_1}} E_{st} \end{aligned}$$

where the square brackets denote a binary condition, u_k and u_{k+1} are the last two arguments of the indicator function $\mathbb{1}_{\Delta^{k+1}[s,t]}$, and in the last identity we have used the base case once again. The conclusion now follows from the fact that the condition $[u_k < u_{k+1} < t]$ results in us only counting terms corresponding to $E \in (D)_1$ where the newly circled node comes after the k that become circled in $D \in (C)_k$. ■

We now ask the question of what taking the expectation of a random variable C_{st}, with $C \in \mathcal{K}^d$, corresponds to in terms of symbolic operations on the diagram C. This will involve adding arcs, in a precise manner, between single nodes in C until there are no single nodes left. Here other examples below we continue to omit the labelling. For instance

$$\begin{aligned}
\mathbb{E}(\bullet\ \bullet\ \bullet\ \bullet)_{st} &= \mathbb{E}\int_{s<u_2<u_4<t} X_{su_2}X_{u_2u_4}R(\mathrm{d}u_2,\mathrm{d}u_4) \\
&= \int_{s<u_2<u_4<t} \mathbb{E}[X_{su_2}X_{u_2u_4}]R(\mathrm{d}u_2,\mathrm{d}u_4) \\
&= \int_{s<u_2<u_4<t}\int_{\substack{s<u_1<u_2\\ u_2<u_3<u_4}} R(\mathrm{d}u_1,\mathrm{d}u_3)\,R(\mathrm{d}u_2,\mathrm{d}u_4) \\
&= (\bullet\ \bullet\ \bullet\ \bullet)_{st}
\end{aligned} \tag{5.81}$$

In the following case, instead, the two single nodes do not get linked:

$$\mathbb{E}(\bullet\ \bullet\ \bullet\ \bullet)_{st} = \mathbb{E}\int_s^t (\bullet\ \bullet\ \bullet)_{su}\delta X_u = 0$$

For larger diagrams the linking procedure becomes more complex; in order to describe this we will require some more notation. We will be considering finite formal sums of diagrams $C = \sum_{k=1}^p C^k$, and we set $C_{st} \coloneqq \sum_{k=1}^p C_{st}^k$ and in the case of an empty sum $p = 0$ we set $C_{st} \coloneqq 0$ (not to be confused with the evaluation of the empty diagram, $\varnothing_{st} = 1$). We will use the notation $\underline{\mathcal{K}}^d$ to denote the set of diagrams in $\mathcal{K}^d$ where some nodes are underlined; the underline may be broken in several places, in which case we will call the collections of consecutive nodes that are all underlined the diagram's *underlined factors* (or simply *factors*). These represent products of random terms of whose product we are taking the expectation, and will always be placed in such a way that, when applying the definition (5.59) of the random variables associated to diagrams in $\mathcal{K}^d$, it will happen at some step of the recursion that the random variable will be expressed as a multiple deterministic integral of a product of underlined factors (with some arcs deleted and some nodes circled, as prescribed by item 2): this means we are taking the expectation of their product. For example

$$\begin{aligned}
(\underline{\bullet\ \bullet\ \bullet}\ \bullet\ \bullet\ \underline{\bullet\ \bullet\ \bullet\ \bullet\ \bullet}\ \bullet\ \downarrow\ \odot\ \bullet)_{st} = \int_{s<u_4<u_5<u_{11}<u_{12}<u_{13}<u_{14}<t} &\mathbb{E}[(\bullet\ \odot\ \bullet)_{su_4}(\bullet\ \bullet\ \odot\ \bullet\ \bullet)_{u_5u_{11}}] \\
&R(\mathrm{d}u_4,\mathrm{d}u_{11})R(\mathrm{d}u_5,\mathrm{d}u_{14})R(\Delta(u_{11},u_{12}),\mathrm{d}u_{12})R(\mathrm{d}u_2,\mathrm{d}u_8)
\end{aligned}$$

In particular, it will always be the case that all single nodes of an element of $\underline{\mathcal{K}}^d$ will belong to some underlined factor.

The procedure of taking the expectation of an element in $\mathcal{I}^d$ is described by the following recursive algorithm which we call `link`, whose crucial step is justified by the n-factor Itô-Skorokhod isometry. Its initial input will always be an element of $C \in \mathcal{I}^d$ which is underlined from start to finish, in which case we will write $\texttt{link}(C)$ to denote its output; it is necessary, however, to allow more general inputs in $\underline{\mathcal{K}}^d$ which will be considered in the algorithm's recursive calls. We introduce the subset

$$\mathcal{J}^d \coloneqq \{C \in \mathcal{K}^d \mid \text{no node of } C \text{ is single}\} \tag{5.82}$$

These diagrams represent deterministic quantities and will constitute the outputs of `link`; notations regarding the subscript and labelling carry over. More precisely, outputs of `link` will be formal sums of elements of $\mathcal{J}^d$, and we will use the symbol $\in$ to mean that a diagram is a summand in one of these sums.

Algorithm 1: `link`

input: $C \in \underline{\mathcal{K}}^d$.

while *not all of the underlined factors of* C *end with a single node* **do**

  for F *an underlined factor of* C **do**

    if *the terminal node* b *of* F *is circled or the endpoint of an edge* **then**

      shorten the underline so that b is no longer underlined;

    else if b *is the endpoint of an arc* **then**

      shorten the underline so that b is no longer underlined, and if the other endpoint of the arc was above the same underline, split it so that a is no longer underlined;

    end

  end

end

if $C \in \mathcal{J}^d$ **then**

  return C.

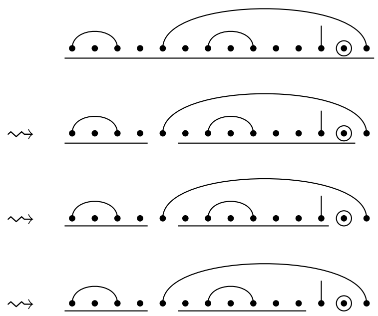
end

Consider all possible combinations of additional arcs with the following two properties: the rightmost node of each factor is linked to a single node belonging to a different factor, and these are the only new arcs introduced. Adding all such combinations of arcs yields a sum $\sum_k C_k$ with $C_k \in \underline{\mathcal{K}}^d$;

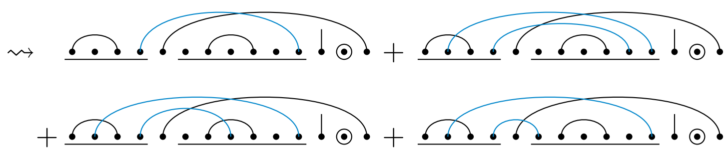
return $\sum_k$ `link`(C_k).

We illustrate `link`'s mechanism of action and elaborate on some of the steps in the following

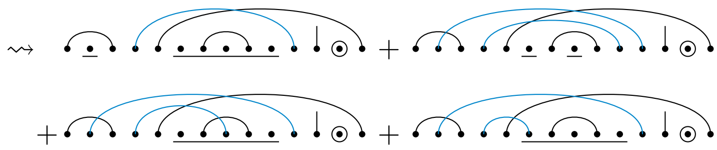
Example 5.23 (`link`)**.** We begin with a fully underlined diagram and walk through all the steps of the algorithm `link`, colouring light blue the arcs that are added in the process. For the purposes of conciseness, we "run the code in parallel" on all diagrams generated during execution. Note that each description refers to the step illustrated above it.



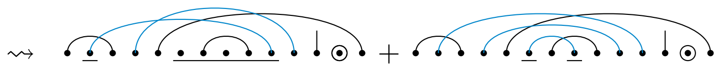
Up until here we have brought the underline inside arcs, edges and circled nodes; at this point we exit the **while** clause, as both underlined factors now end with a single node.



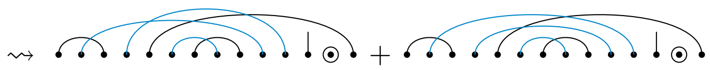
We have executed the last section of the algorithm, corresponding to applying the Itô-Skorokhod isometry for two factors (5.76).



Now in the second iteration of the algorithm, we have once again shortened the underlines of the factors.



Here we have again linked nodes: this means deleting the last two diagrams in the previous step, for which there are no valid combinations of links, as there is only one factor (this can be viewed as the Itô-Skorokhod isometry for one factor reducing to the zero mean property of δ).



The final output. For the first diagram (which lagged one step behind the second) we have deleted underlines, linked nodes, and deleted underlines again. For the second, which already had no single nodes left, we have simply deleted residual underlines. The algorithm now exits at the **if** $C \in \mathcal{J}^d$ clause.

Proposition 5.24. $\mathbb{E}C_{st} = \mathtt{link}(C)_{st}$ *for* $C \in \mathcal{I}^d$.

Proof. The idea is illustrated by the above diagram. Deleting underlines corresponds to passing to the expectation inside the integrals against $R(\mathrm{d}u, \mathrm{d}v)$ and $R(\Delta(u,v), \mathrm{d}v)$, which uses Assumption 6. The node linking procedure is justified by a combination of the n-factor Itô-Skorokhod isometry and the rule for Malliavin-differentiating random variables associated to diagrams. (In the description of the algorithm, these two are applied together: the graphical description of the intermediate step would involve drawing combinations of overlines above the underlined factors.) Indeed, linking the rightmost node of a factor to the rightmost node of another factor corresponds to the $\bullet \frown \bullet$-type pairing in the statement of Proposition 5.20. Linking the rightmost node of a factor to a single node in another factor that is not its rightmost node corresponds to the $\bullet \frown \star$-type pairing, and conversely every type of pairing of this type is realised, by Proposition 5.22. ■

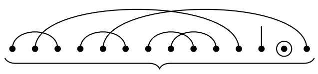
For the proof of the main result we will need to count the all the outputs of $\mathtt{link}$ as the input varies in $\mathcal{I}^d$. For this we need some extra notation and terminology. We will say that E is a *subdiagram* of $D \in \mathcal{K}^d$, written $E \leq D$, if it is comprised of an ordered set of nodes that are consecutive in D, along with additional data (labels, circles, arcs and edges) attached to those nodes: the only condition for a sequence of consecutive nodes of D to define a subdiagram is for there to be no arcs between such nodes and nodes elsewhere in D. We consider the position of $E \leq D$ in D to be part of the definition of the subdiagram (so that diagrams occurring at different positions of D always constitute different subdiagrams, even if they happen to be identical). When D ends with a node that is the endpoint of an arc, with endpoints a and b, and that is not intersected by other arcs, we will call the subdiagram of D whose nodes lie to the left of a the *left* subdiagram of D, and the subdiagram whose nodes lie between a and b the *right* subdiagram of D. If D ends with any other type of node c, we will call the subdiagram of D formed by all nodes to the left of c the *left* subdiagram of D (in these cases there is no right subdiagram). We will show that there is a recursive algorithm, which we call `unlink`, that is the inverse of `link` (viewed as a multivalued function). `unlink` takes as input a diagram in $\mathcal{K}^d$ together with a collection of its subdiagrams and outputs a single element of $\mathcal{I}^d$; we will abbreviate $\texttt{unlink}(D) := \texttt{unlink}(D, \{D\})$ when $D \in \mathcal{J}^d$, which will always be the case for the outermost input.

Algorithm 2: `unlink`

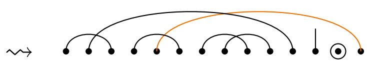
input: $(D, \mathcal{E})$ with $D \in \mathcal{K}^d$ and $\mathcal{E} \subseteq \{E \mid E \leq D\}$.
if $D \in \mathcal{I}^d$ **then**
  return D;
end
for $E \in \mathcal{E}$ *that contain arc intersections* **do**
  $\mathcal{E} \leftarrow \mathcal{E} \setminus \{E\}$;
  if E *ends with a node that is the endpoint of an arc* λ **then**
    $D \leftarrow D \setminus \{\text{all arcs that intersect } \lambda\}$;
    $\mathcal{E} \leftarrow \mathcal{E} \cup \{\text{the left and right subdiagrams of } E\}$;
  else
    $\mathcal{E} \leftarrow \mathcal{E} \cup \{\text{the left subdiagram of } E\}$;
end
return $\texttt{unlink}(D, \mathcal{E})$.

We illustrate the procedure with an example.

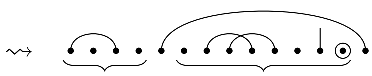
Example 5.25. We colour orange the arc λ (a local variable) of the above algorithm. The braces indicate the elements currently in $\mathcal{E}$ and that contain arc intersections (and that will therefore still be considered by the algorithm). As before, descriptions refer to the diagram above.



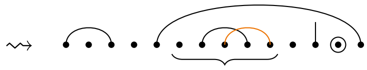
We begin with an element of $D \in \mathcal{J}^d$ and compute $\texttt{unlink}(D)$, which means $\mathcal{E}$ is now the singleton $\{D\}$.



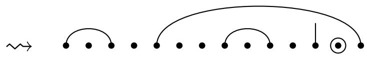
At this point we are at the **if** inside the **for**: the clause is satisfied, since the last node is the endpoint of an arc λ. We have removed the brace, since $\mathcal{E}$ is now empty, its only element having been removed.



Here we have deleted the two arcs that crossed λ (which we have re-coloured black). At the end of this iteration $\mathcal{E}$ is the set comprised of the left and right subdiagrams of the diagram above.



The left subdiagram is no longer considered by the algorithm, since it contains no intersections. The right subdiagram still does, so after 3 iterations which consist of skipping past nodes 13-11 from right to left (not illustrated), we land on the marked subdiagram, which ends in a node that is the right endpoint of an arc λ.



We delete the only arc that crosses the orange arc (now again black), and since there are no more intersections the algorithm returns the diagram above.

Let $D \in \mathcal{J}^d$ denote the diagram with which we started, $E = \texttt{unlink}(D) \in \mathcal{I}^d$ the output above, $C \in \mathcal{I}^d$ the input considered in Example 5.23 (without the underline) and $\texttt{link}(C) = A + B$ its two outputs. Note how, although D can be obtained from C by adding arcs, it coincides with neither A nor B, and indeed also $E \neq C$. The reader may check that applying link to E results in a sum of two diagrams, one of which is D. All of this makes sense in light of the following

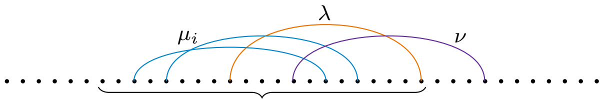
Lemma 5.26. *For* $D \in \mathcal{J}^d$ $\texttt{unlink}(D)$ *is the unique diagram* $C \in \mathcal{I}^d$ *s.t.* $D \in \texttt{link}(C)$, *and* $\texttt{link}(C)$ *contains precisely one copy of* D.

Proof. The fact that link(C) contains at most one copy of D is evident from the fact that link always outputs a sum whose terms are distinct elements of $\mathcal{J}^d$ (since, in the notation of Algorithm 1, at each iteration any two diagrams C_i and C_j on which the algorithm is recursively called will differ in that there must exist at least one node — the rightmost of a factor — which is linked to a different node in each).

Let $D \in \mathcal{J}^d$. There are two things to show: (i) $C \in \mathcal{I}^d,\ D \in \texttt{link}(C) \ \Rightarrow\ C = \texttt{unlink}(D)$, and (ii) $D \in \texttt{link}(\texttt{unlink}(D))$.

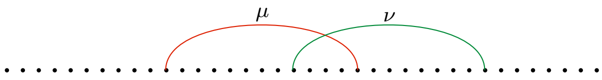
We begin with the former; let D, C be as in (i). We will say that an arc μ *left-intersects* an arc ν if the two arcs intersect, and the left endpoint of μ lies to the left of the left endpoint of ν (and therefore the same holds for the right endpoints). Every arc introduced by link, applied to $C \in \mathcal{I}^d$, must left-intersect an arc that was originally in C: this is because such an arc must have endpoints that, at some point in during execution, belonged to different underlined factors, which means (since link's input has the single underlined factor C) that between them there must be a node that is the left endpoint of an arc that was originally in C (since left endpoints of arcs originally in C are the only places at which an underline can be split, and new factors created). We now show, inductively, that at each iteration unlink removes arcs that do not belong to C. The arcs removed in the first iteration cannot belong to C, since they intersect the rightmost arc in D, which must be in C (since it cannot have been added by link). Now, referring to the notation in Algorithm 2, at each iteration the arc λ does not left-intersect any arc in the current state of D (since this would violate $E \leq D$):

this means that, if λ left-intersects an arc ν in the original input D, ν must have already been removed by `unlink`, and therefore, inductively, ν is not present in C. This means that λ does not left-intersect any arc in C, which implies that λ must be present in C, which in turn implies that all arcs μ_i that intersect λ (which are about to be deleted) do not belong to C. The arcs in this argument are illustrated in the following picture, in which the brace denotes the factor of which λ is the leading arc



This proves the claim, since when the algorithm terminates all arcs remaining must belong to C, since there are no intersections left.

We now prove existence; let $D \in \mathcal{J}^d$. We say that, given $A, B \in \mathcal{K}^d$ that can be both obtained by adding arcs to $C \in \mathcal{I}^d$, B is a *C-refinement* of A if B can be obtained from A by adding arcs that left-intersect arcs that were present in C. We claim that `link` admits the following non-recursive formulation: when called on $C \in \mathcal{I}^d$, it outputs the sum of all diagrams in $\mathcal{J}^d$ that are C-refinements of C. To this end, we show that, if $D \in \mathcal{J}^d$ is a C-refinement of $A \in \underline{\mathcal{K}}^d$, there exists one diagram in the sum $\sum_k C_k$ outputted by a single iteration of $\texttt{link}(A)$ of which D is still a C-refinement. Indeed, proceeding inductively, the only reason this could not be the case is that the rightmost node in one of A's factors is linked, in D, to a node inside the same factor. But this cannot be the case, since if any such arc μ left-intersects an arc ν in the current state of D, ν cannot be present in C and must have been added after: this is because it can never be the case, throughout the execution of $\texttt{link}(C)$, that an underlined factor contains the left endpoint, but not the right endpoint, of an arc that is in C.



Iterating, this implies that, if we begin with a $C \in \mathcal{I}^d$ of which D is a C-refinement, $\texttt{link}(C)$ will terminate with a sum of diagrams, one of which is D. But $\texttt{unlink}(D)$ is obtained from D by removing arcs that left-intersect arcs which are not considered in future iterations, and that will therefore still be present in $\texttt{unlink}(D)$: this implies $\texttt{unlink}(D)$ is an $\texttt{unlink}(D)$-refinement of D, and the conclusion follows. ∎

Theorem 5.27 (Wiener chaos expansion of the signature of a Gaussian process)**.**

$$w^m \boldsymbol{X}_{st}^{\gamma_1,\ldots,\gamma_n} = \sum_{E \in \mathcal{J}^d_m(\gamma_1,\ldots,\gamma_n)} \delta^m E_{st} \tag{5.83}$$

Proof. By Stroock's identity (5.24), the representation theorem Theorem 5.16, and the rule for Malliavin derivatives of diagrams Proposition 5.22

$$\begin{aligned} w^m \boldsymbol{X}_{st}^{\gamma_1,\ldots,\gamma_n} &= \delta^m \big(\mathbb{1}_{\Delta^m[0,T]} \mathbb{E}[\mathcal{D}^m \boldsymbol{X}_{st}^{\gamma_1,\ldots,\gamma_n}] \big) \\ &= \delta^m \Big(\sum_{C \in \mathcal{I}^d_0(\gamma_1,\ldots,\gamma_n)} \mathbb{E}[\mathbb{1}_{\Delta^m[0,T]} \mathcal{D}^m C_{st}] \Big) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{\substack{C \in \mathcal{I}_0^d(\gamma_1,\ldots,\gamma_n) \\ D \in (C)_m}} \delta^m \mathbb{E} D_{st} \\
&= \sum_{D \in \mathcal{I}_m^d(\gamma_1,\ldots,\gamma_n)} \delta^m \mathbb{E} D_{st} \\
&= \sum_{D \in \mathcal{I}_m^d(\gamma_1,\ldots,\gamma_n)} \delta^m \mathtt{link}(D)_{st}
\end{aligned}
$$

and for each $E \in \mathcal{J}_m^d(\gamma_1, \ldots, \gamma_n)$ the term E_{st} appears exactly once in the sum, precisely in the summand corresponding to $D = \mathtt{unlink}(E)$, by Lemma 5.26. ■

Corollary 5.28 (Expected signature of a Gaussian process)**.** *In particular*

$$
\mathbb{E}\boldsymbol{X}_{st}^{\gamma_1,\ldots,\gamma_n} = \sum_{E \in \mathcal{J}_0^d(\gamma_1,\ldots,\gamma_n)} E_{st} \tag{5.84}
$$

Remark 5.29. As a consequence we have confirmation of the well-known fact that $w^m \boldsymbol{X}_{st}^{(n)} = 0$ if $m > n$ (since in this case there are no diagrams of degree n in $\mathcal{J}_m^d$), and that the same holds if $n \not\equiv m \pmod 2$ (as there must be an even number of nodes to $\mathtt{link}$ for there to be an output). In particular $\mathbb{E}\boldsymbol{X}^{(n)} = 0$ for odd n, as can already be inferred by arguing that $\boldsymbol{X}$ and the signature of the Gaussian rough path lift of $-X$ are equal in law.

Example 5.30 (Expected signature of $1/2 < H$-fBm, cf. [BC07])**.** The expected signature of fractional Brownian motion with Hurst parameter $H > 1/2$ has already been calculated in [BC07, Theorem 31]. That Corollary 5.28 reduces to this result can be seen by Remark 5.13 and the following fact. Letting $n = 2k$, if $A := \{\{i_1, j_1\}, \ldots, \{i_k, j_k\}\}$ with the property that $\bigcup A = \{1, \ldots, n\}$ there exist precisely $k!2^k$ permutations $\sigma \in \mathfrak{S}_n$ with the property that $A = \{\{\sigma(1), \sigma(2)\}, \ldots, \{\sigma(n-1), \sigma(n)\}\}$ ($k!$ to order the pairs and 2^k to choose an ordering within each pair).

It is interesting to see why their proof, which consists of passing to the limit on a linear interpolation of X, does not work in the $1/4 < H < 1/2$ case. For the reader's convenience we reproduce the proof below, for simplicity in the case of the signature at orde $n = 2$, which already captures the essence of the problem (the case of higher n is treated using the Wick-Isserlis theorem on the higher moments of multivariate Gaussian distributions). Also, it is not limiting to confine ourselves to the scalar case $d = 1$, and time horizon $T = 1$. Let ${}^m X$ be the linear interpolation of X on a partition π_m of step size 2^{-m}. Then since ${}^m\boldsymbol{X} \to \boldsymbol{X}$ in L^2

$$
\begin{aligned}
\mathbb{E}\boldsymbol{X}_{01}^{(2)} &= \lim_{m\to\infty} \mathbb{E} \int_{0<u<v<1} {}^m\dot{X}_u {}^m\dot{X}_v \mathrm{d}u\mathrm{d}v \\
&= \lim_{m\to\infty} \int_{0<u<v<1} \mathbb{E}[{}^m\dot{X}_u {}^m\dot{X}_v] \mathrm{d}u\mathrm{d}v
\end{aligned}
$$

This is still true for $1/4 < H \leq 1/2$. We now split the integral in off- and on- (& "almost on"-) diagonal contributions:

$$
\begin{aligned}
&= \lim_{m\to\infty} \sum_{\substack{[a,b],[c,d]\in\pi_m \\ b<c}} \int_{[a,b]\times[c,d]} \mathbb{E}[{}^m\dot{X}_u {}^m\dot{X}_v]\mathrm{d}u\mathrm{d}v \\
&\quad + \lim_{m\to\infty} \sum_{\substack{[a,b],[c,d]\in\pi_m \\ c=b}} \int_{[a,b]\times[c,d]} \mathbb{E}[{}^m\dot{X}_u {}^m\dot{X}_v]\mathrm{d}u\mathrm{d}v \\
&\quad + \lim_{m\to\infty} \sum_{[a,b]\in\pi_m} \int_{\Delta^2[0,1]\cap[a,b]^2} \mathbb{E}[{}^m\dot{X}_u {}^m\dot{X}_v]\mathrm{d}u\mathrm{d}v
\end{aligned} \tag{5.85}
$$

Note that the 2D Young/Lebesgue integral $\int_{0<u<v<1} R(\mathrm{d}u,\mathrm{d}v) = \int_{0<u<v<1} \partial_{12}R(u,v)\mathrm{d}u\mathrm{d}v$ is equal to only the sum of the first two terms, since the on-diagonal terms do not get evaluated. If the indicator function had been $[u \le v]$, the last term would have been counted twice in the 2D Young integral (corresponding to the area of squares $[a,b]^2$ instead of triangles $\Delta^2[0,1]\cap[a,b]^2$). Now,

$$
\mathbb{E}[{}^m\dot{X}_u {}^m\dot{X}_v] = 2^{2m}\mathbb{E}[X_{ab}X_{cd}] \quad \text{with } u\in[a,b]\in\pi_m,\ v\in[c,d]\in\pi_m
$$

For $b<c$ we can write this as

$$
\mathbb{E}[X_{ab}X_{cd}] = H(2H-1)\int_{[a,b]\times[c,d]} (v-u)^{2H-2}\mathrm{d}u\mathrm{d}v
$$

and in all cases we have

$$
|\mathbb{E}[X_{ab}X_{cd}]| \le \sqrt{\mathbb{E}[X_{ab}^2]\mathbb{E}[X_{ab}^2]} = 2^{-2Hm}
$$

by Cauchy-Schwarz. It is therefore possible to control the last two terms in (5.85) by

$$
\begin{aligned}
&\sum_{\substack{[a,b],[c,d]\in\pi_m \\ c\in\{a,b\}}} \int_{\Delta^2[0,1]\cap([a,b]\times[c,d])} \mathbb{E}[{}^m\dot{X}_u {}^m\dot{X}_v]\mathrm{d}u\mathrm{d}v \\
\le &\sum_{\substack{[a,b],[c,d]\in\pi_m \\ c\in\{a,b\}}} \int_{[a,b]\times[c,d]} 2^{2m-2Hm}\mathrm{d}u\mathrm{d}v \\
= &\sum_{\substack{[a,b],[c,d]\in\pi_m \\ c\in\{a,b\}}} 2^{-2Hm} \\
\le &\ 2\cdot 2^{m-2Hm}
\end{aligned}
$$

This vanishes in the limit when $H>1/2$, and we conclude

$$
\mathbb{E}\boldsymbol{X}_{01}^{(2)} = H(2H-1)\int_{0<u<v<1} (v-u)^{2H-2}\mathrm{d}u\mathrm{d}v
$$

by dominated convergence. This idea extends to the case of $\mathbb{E}\boldsymbol{X}^{(n)}$ for higher n. For $H=1/2$ the proof can be readapted to account for the fact that the off-diagonal contributions vanish and the on-diagonal ones converge to $[X]_{01} = 1$. For $H<1/2$, however this method is fundamentally unsuited to the computation of the double integral, since the on- and off-diagonal contributions diverge to opposite infinities. For a determinate result to emerge one must approximate the triangle with something other than a square mesh, and we see no easy way of carrying out the calculation when n is arbitrary, since several pairs of linked consecutive nodes may appear in succession (in our notation corresponding to a sequence of consecutive edges), so that it is not possible to isolate each $\int_{s<u<v<t} R(\mathrm{d}u, \mathrm{d}v)$ integral and substitute in the above calculation. That it is not an easy task to do directly is confirmed by the fact that several technical lemmas [BC07, Theorem 34, Lemmas 35-38] are dedicated to the computation of $\mathbb{E}\boldsymbol{X}^{(2)}_{01}$ and $\mathbb{E}\boldsymbol{X}^{(4)}_{01}$, and that these do not result in a formula that generalises straightforwardly to higher n. Moreover, these arguments appear to be dependent on X being a fBm, while our method applies to $w^m \boldsymbol{X}^{(n)}_{st}$ for all $n, m \in \mathbb{N}$, $0 \leq s \leq t \leq T$ and a more general class of processes X that need not even have stationary increments. We believe these considerations justify our reliance on the more advanced tools of Malliavin calculus used in this chapter.

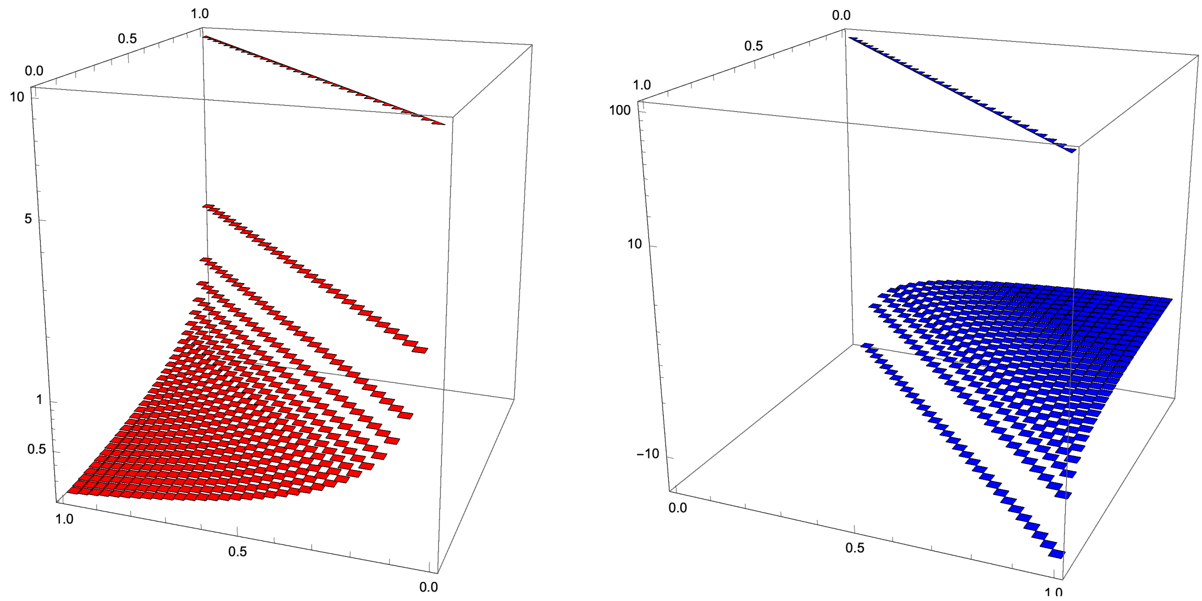


Figure 5.1: Here we compare the two behaviours, corresponding to $H > 1/2$ and $H < 1/2$, of $\int_{0<u<v<1} \dot{\mathbb{E}}[{}^m X_u {}^m \dot{X}_v] \mathrm{d}u\mathrm{d}v$ with ${}^m X$ the sequence piecewise linear interpolations of X on a partition of diadic step size. On the left we have chosen $H = 2/3$, and the sequence of integrals converges to a finite improper integral, whereas on the right $H = 1/3$ (any $1/4 < H < 1/2$ would have exhibited the same behaviour) and the on-diagonal contributions cannot be controlled, as refining the partition simply results in the issue repeating itself, in a self-similar fashion, on each smaller triangle. (The plots are oriented in different ways and the z-axis is scaled with the cube root function, both for improved visibility.)

Example 5.31 (The Wiener chaos decomposition of $\boldsymbol{X}^{(\leq 3)}$)**.** We give the explicit expression for the Wiener chaos expansion of the signature truncated at level 3. We represent each signature term as a sum of their Wiener chaos projections in ascending order; in particular the sum of all non-random terms constitutes the expectation of the LHS.

$$\boldsymbol{X}^{\varnothing}_{st} = \varnothing_{st} = 1, \quad \boldsymbol{X}^{\gamma}_{st} = \delta(\underset{\gamma}{\odot})_{st} = \delta(\mathbb{1}^{\gamma}_{(s,t)}) = X^{\gamma}_{st}$$

$$\boldsymbol{X}^{\alpha\beta}_{st} = \underset{\alpha\beta}{\downarrow} + \delta^2(\underset{\alpha}{\odot}\underset{\beta}{\odot})$$

$$= \frac{R^{\alpha\beta}(s) + R^{\alpha\beta}(t)}{2} - R^{\alpha\beta}(s,t) + \int_{s<u<v<t} \delta X^\alpha_u \delta X^\beta_v$$

$$\begin{aligned}
\boldsymbol{X}^{\alpha\beta\gamma}_{st} &= \delta(\overset{\downarrow}{\bullet}_{\alpha\beta} \odot_{\gamma})_{st} + \delta(\odot_{\alpha} \overset{\downarrow}{\bullet}_{\beta\gamma})_{st} + \delta(\overset{\frown}{\bullet_\alpha \odot_\beta \bullet_\gamma})_{st} + \delta^3(\odot_\alpha \odot_\beta \odot_\gamma)_{st} \\
&= \int_s^t \left(\frac{R^{\alpha\beta}(s) + R^{\alpha\beta}(u)}{2} - R^{\alpha\beta}(s,u) \right) \delta X^\gamma_u + \int_s^t \left(\frac{R^{\beta\gamma}(u) + R^{\beta\gamma}(t)}{2} - R^{\beta\gamma}(u,t) \right) \delta X^\alpha_u \\
&\quad + \int_s^t R^{\alpha\gamma} \begin{pmatrix} s & u \\ u & t \end{pmatrix} \delta X^\beta_u + \int_{s<u<v<w<t} \delta X^\alpha_u \delta X^\beta_v \delta X^\gamma_w
\end{aligned}$$

In particular, notice how the statement that “the Itô and Stratonovich Lévy areas are equal” carries over to the Skorokhod-rough setting, in the sense that

$$\frac{1}{2}(\boldsymbol{X}^{\alpha\beta}_{st} - \boldsymbol{X}^{\beta\alpha}_{st}) = \frac{1}{2} \int_{s<u<v<t} \delta X^\alpha_u \delta X^\beta_v - \delta X^\beta_u \delta X^\alpha_v \in \mathscr{W}^2 \tag{5.86}$$

by symmetry of the covariance function.

***Assumption* 8.** Note how the above expressions, in particular the one for $\boldsymbol{X}^{\alpha\beta\gamma}_{st}$, make sense even for $1/4 < H \leq 1/3$: for the purposes of a result in the next chapter, we conjecture that it continues to hold in this case.

Example 5.32 ($\mathbb{E}\boldsymbol{X}^{(4)}$)**.** The expected signature of level 4 is given by

$$\begin{aligned}
\mathbb{E}\boldsymbol{X}^{\alpha\beta\gamma\delta}_{st} &= (\overset{\downarrow}{\bullet}_{\alpha\beta} \overset{\downarrow}{\bullet}_{\gamma\delta})_{st} + (\overset{\frown}{\bullet_\alpha \overset{\downarrow}{\bullet}_{\beta\gamma} \bullet_\delta})_{st} + (\overset{\frown\frown}{\bullet_\alpha \bullet_\beta \bullet_\gamma \bullet_\delta})_{st} \\
&= \int_{s<u<v<t} R^{\alpha\beta}(\Delta(s,u), \mathrm{d}u) R^{\gamma\delta}(\Delta(u,v), \mathrm{d}v) \\
&\quad + \int_{s<u<v<w<t} R^{\beta\gamma}(\Delta(u,v), \mathrm{d}v) R^{\alpha\delta}(\mathrm{d}u, \mathrm{d}w) \\
&\quad + \int_{s<u<v<w<z<t} R^{\alpha\gamma}(\mathrm{d}u, \mathrm{d}w) R^{\beta\delta}(\mathrm{d}v, \mathrm{d}z)
\end{aligned} \tag{5.87}$$

Using a clever transformation, [BC07, Theorem 34] are able to compute $\mathbb{E}\boldsymbol{X}^{(2)}_{01}$ and $\mathbb{E}\boldsymbol{X}^{(4)}_{01}$ for $1/4 < H$-fBm. Their formulae are specific to the cases $n = 2, 4$, $s = 0$, $t = 1$ and not the same as ours, so it is sensible to check that the two coincide. This is immediate at level 2, using Example 5.31. We now check this at level 4 using (5.44): starting with the first integral above, we have

$$\begin{aligned}
&\int_{0<u<v<1} R(\Delta(s,u), \mathrm{d}u) R(\Delta(u,v), \mathrm{d}v) \\
&= H^2 \int_{0<u<v<1} u^{2H-1} (v-u)^{2H-1} \mathrm{d}u\mathrm{d}v \\
&= H^2 \int_0^1 u^{2H-1} \left[\frac{(v-u)^{2H-1}}{2H} \right]_{u=0}^1 \mathrm{d}u \\
&= \frac{H^2}{2} \int_0^1 u^{2H-1} \left[\frac{(v-u)^{2H}}{2H} \right]_{v=u}^1 \mathrm{d}u \\
&= \frac{H}{2} \int_0^1 u^{2H-1} (1-u)^{2H} \mathrm{d}u
\end{aligned}$$

$$= \frac{H}{4}\int_0^1 u^{2H-1}(1-u)^{2H-1}\mathrm{d}u$$

where the last identity can be verified by showing that the difference of the two integrands is odd about the point $u = 1/2$, which in turn is seen by observing that

$$\frac{H}{4}u^{2H-1}(1-u)^{2H-1} - \frac{H}{2}u^{2H-1}(1-u)^{2H} + \frac{H}{4}(1-u)^{2H-1}u^{2H-1} - \frac{H}{2}(1-u)^{2H-1}u^{2H}$$

has zero derivative and vanishes at $u = 1/2$. This shows equality with [BC07, coefficient of the first term of Γ_H^2 in Corollary 33]. We proceed with the second integral in Example 5.32:

$$\begin{aligned}
&\int_{0<u<v<w<1} R(\Delta(u,v),\mathrm{d}v)R(\mathrm{d}u,\mathrm{d}w)\\
&= H^2(2H-1)\int_{0<u<v<w<1}(w-u)^{2H-2}(v-u)^{2H-1}\mathrm{d}u\mathrm{d}v\mathrm{d}w\\
&= H^2\int_{0<u<v<1}\Big[(1-u)^{2H-1}(v-u)^{2H-1}-(v-u)^{4H-2}\Big]\mathrm{d}u\mathrm{d}v\\
&= \left(\frac{H}{2}-\frac{H^2}{4H-1}\right)\int_0^1(1-u)^{4H-1}\mathrm{d}u\\
&= \frac{2H-1}{8(4H-1)}
\end{aligned}$$

For the third integral we have

$$\begin{aligned}
&\int_{0<u<v<w<z<1} R(\mathrm{d}u,\mathrm{d}w)R(\mathrm{d}v,\mathrm{d}z)\\
&= H^2(2H-1)^2\int_{0<u<v<w<z<1}(w-u)^{2H-2}(z-v)^{2H-2}\mathrm{d}u\mathrm{d}v\mathrm{d}w\mathrm{d}z\\
&= H^2(2H-1)\int_{0<u<v<z<1}\Big[(z-u)^{2H-1}(z-v)^{2H-2}-(v-u)^{2H-1}(z-v)^{2H-2}\Big]\mathrm{d}u\mathrm{d}v\mathrm{d}z\\
&= \frac{H(2H-1)}{2}\int_{0<v<z<1}\Big[z^{2H}(z-v)^{2H-2}-(z-v)^{4H-2}-v^{2H}(z-v)^{2H-2}\Big]\mathrm{d}v\mathrm{d}z\\
&= \frac{H(2H-1)}{2}\int_{0<v<z<1}(z^{2H}-v^{2H})(z-v)^{2H-2}\mathrm{d}v\mathrm{d}z - \frac{H(2H-1)}{4H-1}\int_0^1(1-v)^{4H-1}\mathrm{d}v\\
&= \frac{H(2H-1)}{2}\int_{0<v<z<1}(z^{2H}-v^{2H})(z-v)^{2H-2}\mathrm{d}v\mathrm{d}z - \frac{2H-1}{4(4H-1)}\\
&= \frac{H}{2}\int_0^1(1-v^{2H})(1-v)^{2H-1}\mathrm{d}v - H^2\int_{0<v<z<1}z^{2H-1}(z-v)^{2H-1} - \frac{2H-1}{4(4H-1)}\\
&= \frac{H}{4(4H-1)} - \frac{H}{4}\int_0^t v^{2H-1}(1-v)^{2H-1}\mathrm{d}v
\end{aligned}$$

In the integration by parts we have used that $\lim_{z\to v^+}(z^{2H}-v^{2H})(z-v)^{2H-1} = 0$ which can be shown by using that for $1/4 < H < 1/2$

$$0 \le (z^{2H}-v^{2H})(z-v)^{2H-1} \le (z-v)^{4H-1} \xrightarrow{z\to v^+} 0$$

since $z^{2H} - v^{2H} < (z-v)^{2H}$ for $0 < v < z$ and $H < 1/2$. In the last identity we have used a similar symmetry argument as the one used in the first calculation, solved trivial integrals and rearranged terms. Note how this calculation would have been simpler if $H \geq 1/2$ since it would not have been necessary to integrate by parts to avoid integrating $(z-v)^{2H-2}$ (cf. [BC07, Lemma 32]).

It is interesting that showing these identities is non-trivial, which is to be expected given how different the two methods (Malliavin calculus and integration by substitution) are. Also note that we have only used $H > 1/4$ (in evaluating terms like $(v-u)^{4H-1}|_{u=v} = 0$), not $H > 1/3$. This shows, using [BC07, Theorem 34] (and scaling of fBm for the case of arbitrary $0 \leq s \leq t \leq T$) that (5.87) actually holds for $1/4 < H$-fBm.

Remark 5.33 (Expected signature of a Gaussian semimartingale). Our strategy to prove Theorem 5.27 relies on our being able to represent the signature using iterated Skorokhod integration and different types of deterministic integration, with the former satisfying the zero-mean property (and more in general the n-factor Itô-Skorokhod isometry) and the latter commuting with the expectation operator, as assumed in Assumption 6. The classical Itô-Stratonovich representation formula of Example 5.18 (of which we use the notation here) does not separate deterministic and stochastic integration to the same extent, since integration against $\mathrm{d}A$ is neither zero-mean nor deterministic. Though a classical proof specific to the Gaussian semimartingale case may be available, it is not trivial and would rely on the decomposition of X into martingale and bounded variation part, while our method only relies on the intrinsic Gaussian structure of X. The expected signature of the Brownian bridge of (5.70) can, for instance, be computed using Theorem 5.27, the expressions for the differentials in Example 5.19 (and without the representation (5.71)), and the integrals can be computed explicitly as polynomials in s, t. Therefore, while (5.66) may be more natural to work with in the context of semimartingales, we do view it as being suboptimal for the purpose of computing expectations. The exception that proves the rule is, of course, that of martingales, where the two methods coincide: this exact calculation is done in [Bau04, Proposition 1.3] for Brownian motion, which is already representative of the more general case of Gaussian martingales.

Conclusions and further directions

Once the missing technical details listed in the ***Assumption***s are clarified, we will have proven a formula for the expected signature (and Wiener chaos expansion) of a class of Gaussian processes for which it was not previously known. In clarifying these aspects, we intend to pay special attention to what happens in the case in which the Gaussian process is of bounded $[3,4) \ni p$-variation, the prime example being $(1/4, 1/3] \ni H$-fBm. This is because, as observed in Remark 5.14, Assumption 8 and Example 5.32, while the integrals in Theorem 5.16 may not converge a.s. the main theorem Theorem 5.27 can still be conjectured to hold. A careful reading of [CL20] is necessary to see whether the conversion formula in the $p \in [3,4)$ can be used as done for that of [CL19], or whether an entirely new approach is needed.

A task that was not considered in this chapter is the explicit calculation of the expected signature of the Brownian bridge returning to the starting point. This is possible in principle, since the integrals to compute are all polynomials, as seen in Example 5.19; the question is whether they can be collected is a nice formula, similar to Fawcett's one for the expected signature of Brownian motion.

Finally, we believe that Theorem 5.27 could have interesting numerical applications, specifically to compute the mean square error of the Euler approximation of an RDE. This would involve using the isometry property of the multiple Wiener integral (5.21) and estimating $\|D_{st}\|_{\mathcal{H}^m}$ for $D \in \mathcal{J}_m$, cf. [Pas20] for similar estimates pertaining to the signature of $1/2 < H$-fBm.

Such bounds could aid in generalising the results of Chapter 1, relating to the "L^2/weak-optimality for small times" of approximations of SDEs with ones intrinsic to submanifolds of $\mathbb{R}^d$, to the case of RDEs driven by fBm and a drift term given by t or t^{2H}. It might be, however, that for this purpose, the signature terms of X are not sufficient, and also the signature of X jointly with the drift term must be considered.

6

A branched rough path above $1/4 < H$-fractional Brownian motion

Project status. This chapter consists of a short project which relies on the background notions and notation established in Chapter 4 and Chapter 5; it should also be noted that Theorem 6.15 relies on the unproven Assumption 8 of the previous chapter. At the time of writing, I have not uploaded this material to arXiv or submitted for publication.

Introduction

In this last chapter we will combine the topics of Chapter 5 and Chapter 4 to build a branched rough path $\widehat{\boldsymbol{X}}$ above d-dimensional fractional Brownian motion (fBm) with Hurst parameter $H \in (1/4, 1)$. When $H \in (1/4, 1/3]$ this constitutes an example of a stochastic, multidimensional, non-geometric branched rough path, which may not be considered as taking values in the tensor algebra. Such an example, a similar one to which we could not find in the literature, arguably provides important support for the theory of branched rough paths, and opens up the possibility of studying a pathwise, non-geometric, stochastic integration theory of lower regularity and greater structural complexity than Itô's.

Our branched rough path $\widehat{\boldsymbol{X}}$ exhibits the following characteristics:

- It is canonically and explicitly defined solely in terms of X and its intrinsic properties as a Gaussian process, using Malliavin calculus.
- It is not geometric, or in general even quasi-geometric.
- At $H = 1/2$ it coincides with the Itô rough path. Moreover, for $(1/3, 1/2)$ we obtain the rough path defined in [QX18] in the context of arbitrage-free pricing of simple (state-dependent) contingent claims.

- The rough integral of a one-form coincides with its Skorokhod integral:

$$\int f(X)\mathrm{d}\widehat{\boldsymbol{X}} = \int f(X)\delta X$$

and in particular vanishes in mean. This does not hold for more general controlled integrands that are also Skorokhod integrable, e.g. solutions to $\widehat{\boldsymbol{X}}$-driven RDEs.

- While the proof of regularity is only carried out for the case of fBm, we believe the same construction to work for many other Gaussian processes of bounded $4 > p$-variation. This can be seen in the particular case of the Brownian bridge returning to the origin, for which $\widehat{\boldsymbol{X}}$ does not coincide with the Itô rough path.

- In the scalar case $d = 1$, $\widehat{\boldsymbol{X}}$ is quasi-geometric but not geometric, and its change of variable formula is analogous to that of Cheridito-Nualart [CN05].

- $\widehat{\boldsymbol{X}}$ is adapted to the filtration $\mathcal{F}$ generated by X, i.e. $\widehat{\boldsymbol{X}}_{st}$ is $\mathcal{F}_t$-measurable, and when $H > 1/3$ it is also the case that $\widehat{\boldsymbol{X}}_{st}$ is measurable w.r.t. the sigma-algebra generated by increments $\mathcal{F}_{st} \coloneqq \sigma(\{X_{su} \mid u \in [s,t]\})$. For $H \in (1/4, 1/3]$ the latter is no longer true, as one of the terms contains a path-dependency. However, it is still true that for a one-form f the integral $\int_s^t f(X)\mathrm{d}\widehat{\boldsymbol{X}}$ is measurable w.r.t. the sigma-algebra generated by the process between times s and t

$$\mathcal{F}_{s,t} \coloneqq \sigma(\{X_u \mid u \in [s,t]\}) \tag{6.1}$$

Since integrands that are not one-forms of X, e.g. RDE solutions, are typically path-dependent, this can be interpreted as meaning that rough integration against $\widehat{\boldsymbol{X}}$ does not introduce additional path dependency.

It is also worth mentioning that, even though Malliavin calculus can be carried out for rougher processes than the ones we consider, e.g. $(1, 1/4] \ni H$-fBm, the method described in this chapter cannot work for these. This is because their sample paths almost surely do not lie in the domain of the divergence operator [Nua06, p.301].

This chapter is organised as follows. There is no background section, and we refer to Chapter 4 and Chapter 5 for the notation. In Section 6.1 we define the grouplike multiplicative functional $\widehat{\boldsymbol{X}}$ and compute its terms. In Section 6.2 we show that they satisfy the regularity requirement that make $\widehat{\boldsymbol{X}}$ a rough path. In Section 6.3 we compute the bracket extension of $\widehat{\boldsymbol{X}}$, which is defined canonically: this yields a change of variable formula for functions of RDEs driven by X. In Section 6.4 we compare $\widehat{\boldsymbol{X}}$ to the canonical Stratonovich geometric rough path defined via smooth approximation, and compute the correction terms defind in [HK15] explicitly: this yields an Itô-Stratonovich formula for RDEs. Finally, in Conclusions and further directions we outline our plan for turning this short chapter into a paper, which involves a version of Hörmander's theorem for branched rough paths.

6.1 The multiplicative functional

Let X be a $(1/4, 1) \ni H$-fBm (of course, our construction will trivially reduce to Young integration when $H > 1/2$, so it is only interesting to assume $1/4 < H \leq 1/2$). We will assume the background introduced in

Section 5.1 and Section 4.1 and use notation therein. We will always state our results using notation that only assumes X to be a Gaussian process (e.g. our formulae will feature a generic covariance function R), but we will use that X is a fBm in several places, and H will denote its Hurst parameter throughout.

A very natural question that follows from Chapter 5 is whether one can define a branched rough path above a Gaussian process by Skorokhod integration, i.e.

$$\boldsymbol{X}_{st}^{[\ell]_\gamma} = \int_s^t \boldsymbol{X}_{su}^{\ell} \delta X_u^{\gamma}$$

and extending to all forests with products. This is immediately answered in the negative: indeed, using the formula for taking out random constants from the Skorokhod integral (5.17) we have, for $0 \leq r \leq s \leq t \leq T$

$$\begin{aligned}
&\boldsymbol{X}_{rt}^{\substack{\bullet\alpha \\ \bullet\beta}} \\
&= \int_r^t X_{ru}^{\alpha} \delta X_u^{\beta} \\
&= \int_r^t X_u^{\alpha} \delta X_u^{\beta} - \big(X_r^{\alpha} X_{rt}^{\beta} - R^{\alpha\beta}(r, \Delta(r,t))\big) \\
&= \int_r^s X_u^{\alpha} \delta X_u^{\beta} + \int_s^t X_u^{\alpha} \delta X_u^{\beta} - X_r^{\alpha} X_{rs}^{\beta} - X_s^{\alpha} X_{st}^{\beta} + X_{rs}^{\alpha} X_{st}^{\beta} \\
&\quad + R^{\alpha\beta}(r, \Delta(r,s)) + R^{\alpha\beta}(s, \Delta(s,t)) - R^{\alpha\beta}(\Delta(r,s), \Delta(s,t)) \\
&= \boldsymbol{X}_{rs}^{\substack{\bullet\alpha \\ \bullet\beta}} + \boldsymbol{X}_{st}^{\substack{\bullet\alpha \\ \bullet\beta}} + X_{rs}^{\alpha} X_{st}^{\beta} - R^{\alpha\beta}(\Delta(r,s), \Delta(s,t))
\end{aligned}$$

In other words, $\boldsymbol{X}$ fails the Chen identity because of the correction term in (5.17). Note $\boldsymbol{X}$ is not even almost multiplicative, since the best we can say is

$$|R^{\alpha\beta}(\Delta(r,s), \Delta(s,t))| \lesssim (t-r)^{2H}$$

which is not enough when $H < 1/2$, i.e. in all cases that matter.

We can, however, define a multiplicative functional in violation of (5.17), i.e. by taking out random constants without adding the correction term: for $t \in [0,T]$ set

$$\begin{aligned}
\widehat{\boldsymbol{X}}_{st}^{\substack{\bullet\alpha \\ \bullet\beta}} &\coloneqq \int_s^t X_u^{\alpha} \delta X_u^{\beta} - X_s^{\alpha} X_{st}^{\beta} \\
&= \int_s^t X_{su}^{\alpha} \delta X_u^{\beta} - R^{\alpha\beta}(s, \Delta(s,t))
\end{aligned} \tag{6.2}$$

This is no longer equal to $\int_s^t X_{su}^{\alpha} \delta X_u^{\beta}$, but is multiplicative by definition. This idea generalises to $\widehat{\boldsymbol{X}}$ evaluated on arbitrary forests. The following lemma can be viewed as a generalisation of [FH20, Exercise 2.4] to branched rough path of inhomogeneous regularity. Recall that in a given Hopf algebra the symbol $\mathcal{S}$ is used to denote the antipode and the symbol $\mathcal{G}$ is used to denote the group of elements in a Hopf algebra s.t. $\Delta x = x \otimes x$.

Lemma 6.1 (Defining $\boldsymbol{X}_{st}$ in terms of $\boldsymbol{X}_{0t}$). *Let A be a finite weighted alphabet and $\boldsymbol{X} \colon [0,T] \to \mathcal{H}_{\mathrm{GL}}^A$ be a map that takes values in $\mathcal{G}_{\mathrm{GL}}^A$. Then setting*

$$\boldsymbol{X}_{s0} \coloneqq \mathcal{S}_{\mathrm{GL}} \boldsymbol{X}_{0s}, \quad \boldsymbol{X}_{st} \coloneqq \boldsymbol{X}_{s0} \star \boldsymbol{X}_{0t} \tag{6.3}$$

defines a multiplicative functional $\boldsymbol{X} \colon \Delta_T \to \mathcal{G}^A_{\mathrm{GL}}$.

Proof. This is a direct consequence of the properties of Hopf algebras. $\boldsymbol{X}$ is multiplicative since

$$\boldsymbol{X}_{rs} \star \boldsymbol{X}_{st} = \boldsymbol{X}_{r0} \star \boldsymbol{X}_{0s} \star \mathcal{S}_{\mathrm{GL}} \boldsymbol{X}_{0s} \star \boldsymbol{X}_{0t} = \boldsymbol{X}_{r0} \star \boldsymbol{X}_{0t} = \boldsymbol{X}_{rt} \tag{6.4}$$

because $\mathcal{S}_{\mathrm{GL}}$ is inversion in $\mathcal{G}^A_{\mathrm{GL}}$. To show that $\boldsymbol{X}$ is group-valued we compute, for $f, g \in \mathcal{F}^A$

$$\begin{aligned}
\boldsymbol{X}^{fg}_{st} &= \langle fg, \mathcal{S}_{\mathrm{GL}} \boldsymbol{X}_{0s} \star \boldsymbol{X}_{0t} \rangle \\
&= \big\langle (\mathcal{S}_{\mathrm{CK}} \otimes \mathbb{1}) \circ \Delta_{\mathrm{CK}}(fg), \boldsymbol{X}_{0s} \otimes \boldsymbol{X}_{0t} \big\rangle \\
&= \big\langle (\mathcal{S}_{\mathrm{CK}} \otimes \mathbb{1})(\Delta_{\mathrm{CK}} f \cdot \Delta_{\mathrm{CK}} g), \boldsymbol{X}_{0s} \otimes \boldsymbol{X}_{0t} \big\rangle \\
&= \sum_{(f)_{\mathrm{CK}}, (g)_{\mathrm{CK}}} \langle \mathcal{S}_{\mathrm{CK}}(f_{(1)} g_{(1)}) \otimes f_{(2)} g_{(2)}, \boldsymbol{X}_{0s} \otimes \boldsymbol{X}_{0t} \rangle \\
&= \sum_{(f)_{\mathrm{CK}}, (g)_{\mathrm{CK}}} \langle \mathcal{S}_{\mathrm{CK}}(f_{(1)}) \mathcal{S}_{\mathrm{CK}}(g_{(1)}), \boldsymbol{X}_{0s} \rangle \langle f_{(2)} g_{(2)}, \boldsymbol{X}_{0t} \rangle \\
&= \sum_{(f)_{\mathrm{CK}}, (g)_{\mathrm{CK}}} \langle \mathcal{S}_{\mathrm{CK}}(f_{(1)}), \boldsymbol{X}_{0s} \rangle \boldsymbol{X}^{f_{(2)}}_{0t} \langle \mathcal{S}_{\mathrm{CK}}(g_{(1)}), \boldsymbol{X}_{0s} \rangle \boldsymbol{X}^{g_{(2)}}_{0t} \\
&= \sum_{(f)_{\mathrm{CK}}, (g)_{\mathrm{CK}}} \langle \mathcal{S}_{\mathrm{CK}}(f_{(1)}) \otimes f_{(2)}, \boldsymbol{X}_{0s} \otimes \boldsymbol{X}_{0t} \rangle \langle \mathcal{S}_{\mathrm{CK}}(g_{(1)}) \otimes g_{(2)}, \boldsymbol{X}_{0s} \otimes \boldsymbol{X}_{0t} \rangle \\
&= \big\langle (\mathcal{S}_{\mathrm{CK}} \otimes \mathbb{1}) \circ \Delta_{\mathrm{CK}}(f), \boldsymbol{X}_{0s} \otimes \boldsymbol{X}_{0t} \big\rangle \big\langle (\mathcal{S}_{\mathrm{CK}} \otimes \mathbb{1}) \circ \Delta_{\mathrm{CK}}(g), \boldsymbol{X}_{0s} \otimes \boldsymbol{X}_{0t} \big\rangle \\
&= \boldsymbol{X}^f_{st} \boldsymbol{X}^g_{st}
\end{aligned}$$

where we have used that antipodes are algebra antihomomorphisms, i.e. homomorphisms in the case of commutative Hopf algebras. ■

Definition 6.2 (The Skorokhod multiplicative functional $\widehat{\boldsymbol{X}}$)**.** We define $\widehat{\boldsymbol{X}}$ to be the group-valued multiplicative functional as in Lemma 6.1, with $A = [d]$ and

$$\widehat{\boldsymbol{X}}^{[f]_\gamma}_{0t} \coloneqq \int_0^t \boldsymbol{X}^f_{0u} \delta X^\gamma_u \tag{6.5}$$

for $t \in [0, T]$, $f \in \mathcal{F}^d$, $\gamma = 1, \ldots, d$.

Remark 6.3. Note that the above definition depends only on the intrinsic properties of the Gaussian process X, since the Malliavin calculus setup does.

Dualising (6.3) yields the explicit formula to compute the terms of $\widehat{\boldsymbol{X}}$:

$$\widehat{\boldsymbol{X}}^t_{st} = \sum_{(t)_{\mathrm{CK}}} \langle \mathcal{S}_{\mathrm{CK}} t_{(1)}, \widehat{\boldsymbol{X}}_{0s} \rangle \widehat{\boldsymbol{X}}^{t_{(2)}}_{0t} \tag{6.6}$$

and using the expression (4.12) for $\mathcal{S}_{\mathrm{CK}}$. We will only really need this formula when t is a tree, since forest terms can more simply be computed by taking products, and we will only be interested in forests of weight ≤ 3.

The next proposition contains the explicit expression of $\widehat{\boldsymbol{X}}^t$ with $t \in (\mathcal{T}^d)^{\leq 3}$, presented so that each summand has the needed regularity (as will be shown in the next section). The integrals are either Skorokhod or Young (and in some cases both — see Lemma 6.10 below), and will be shown to be well defined/convergent in Theorem 6.9. Note how the path dependency in the last term vanishes if $\alpha = \beta$ (always the case when $d = 1$).

Proposition 6.4 (Expression of $\widehat{\boldsymbol{X}}^{\leq 3}$).

$$
\begin{aligned}
\widehat{\boldsymbol{X}}^{\bullet_\gamma}_{st} &= X^\gamma_{st} \\
\widehat{\boldsymbol{X}}^{\mathrm{I}^{\alpha}_{\beta}}_{st} &= \int_s^t X^\alpha_{su}\delta X^\beta_u - R^{\alpha\beta}(s,\Delta(s,t)) \\
\widehat{\boldsymbol{X}}^{\mathrm{V}^{\alpha\beta}_{\gamma}}_{st} &= \int_s^t X^\alpha_{su}X^\beta_{su}\delta X^\gamma_u - \int_s^t X^\alpha_{su}R^{\beta\gamma}(s,\mathrm{d}u) - \int_s^t X^\beta_{su}R^{\alpha\gamma}(s,\mathrm{d}u) \\
\widehat{\boldsymbol{X}}^{\mathrm{I}^{\alpha}_{\beta\gamma}}_{st} &= \int_{s<r<u<v<t} \delta X^\alpha_r \delta X^\beta_u \delta X^\gamma_v - \int_s^t R^{\alpha\beta}(s,\Delta(s,u))\delta X^\gamma_u - \int_s^t X^\beta_{su}R^{\alpha\gamma}(s,\mathrm{d}u) \\
&\quad + \int_0^s X^\alpha_{us}R^{\beta\gamma}(\mathrm{d}u,\Delta(s,t)) - \int_0^s X^\beta_{us}R^{\alpha\gamma}(\mathrm{d}u,\Delta(s,t))
\end{aligned}
\tag{6.7}
$$

Proof. We use (6.6), the expression for the Connes-Kreimer coproduct (4.2), the expression for the Connes-Kreimer antipode (4.12) together with the fact that it is an algebra morphism, and the rule for Skorokhod-integrating a random constant times a process (5.17) to compute all terms. At order 1 we have

$$
\widehat{\boldsymbol{X}}^{\bullet_\gamma}_{st} = -\widehat{\boldsymbol{X}}^{\bullet_\gamma}_{0s} + \widehat{\boldsymbol{X}}^{\bullet_\gamma}_{0t} = -X^\gamma_s + X^\gamma_t = X^\gamma_{st}
$$

At order 2:

$$
\begin{aligned}
\widehat{\boldsymbol{X}}^{\mathrm{I}^{\alpha}_{\beta}}_{st} &= \langle \mathcal{S}_{\mathrm{CK}}\, \mathrm{I}^{\alpha}_{\beta}, \widehat{\boldsymbol{X}}_{0s}\rangle - X^\alpha_s X^\beta_t + \widehat{\boldsymbol{X}}^{\mathrm{I}^{\alpha}_{\beta}}_{0t} \\
&= -\widehat{\boldsymbol{X}}^{\mathrm{I}^{\alpha}_{\beta}}_{0s} + X^\alpha_s X^\beta_s - X^\alpha_s X^\beta_t + \widehat{\boldsymbol{X}}^{\mathrm{I}^{\alpha}_{\beta}}_{0t} \\
&= \int_s^t X^\alpha_u \delta X^\beta_u - X^\alpha_s X^\beta_{st} \\
&= \int_s^t X^\alpha_{su}\delta X^\beta_u - R^{\alpha\beta}(s,\Delta(s,t))
\end{aligned}
$$

At order 3:

$$
\begin{aligned}
&\widehat{\boldsymbol{X}}^{\mathrm{V}^{\alpha\beta}_{\gamma}}_{st} \\
&= \langle \mathcal{S}_{\mathrm{CK}}\, \mathrm{V}^{\alpha\beta}_{\gamma}, \widehat{\boldsymbol{X}}_{0s}\rangle + \langle \mathcal{S}_{\mathrm{CK}} \bullet^\alpha, \widehat{\boldsymbol{X}}_{0s}\rangle \widehat{\boldsymbol{X}}^{\mathrm{I}^{\beta}_{\gamma}}_{0t} + \langle \mathcal{S}_{\mathrm{CK}} \bullet^\beta, \widehat{\boldsymbol{X}}_{0s}\rangle \widehat{\boldsymbol{X}}^{\mathrm{I}^{\alpha}_{\gamma}}_{0t} + \langle \mathcal{S}_{\mathrm{CK}}(\bullet^\alpha\bullet^\beta), \widehat{\boldsymbol{X}}_{0s}\rangle X^\gamma_t + \widehat{\boldsymbol{X}}^{\mathrm{V}^{\alpha\beta}_{\gamma}}_{0t} \\
&= -\widehat{\boldsymbol{X}}^{\mathrm{V}^{\alpha\beta}_{\gamma}}_{0s} + \widehat{\boldsymbol{X}}^{\bullet^\alpha \mathrm{I}^{\beta}_{\gamma}}_{0s} + \widehat{\boldsymbol{X}}^{\bullet^\beta \mathrm{I}^{\alpha}_{\gamma}}_{0s} - X^\alpha_s X^\beta_s X^\gamma_s - X^\alpha_s \widehat{\boldsymbol{X}}^{\mathrm{I}^{\beta}_{\gamma}}_{0t} - X^\beta_s \widehat{\boldsymbol{X}}^{\mathrm{I}^{\alpha}_{\gamma}}_{0t} + X^\alpha_s X^\beta_s X^\gamma_t + \widehat{\boldsymbol{X}}^{\mathrm{V}^{\alpha\beta}_{\gamma}}_{0t} \\
&= -\int_0^s X^\alpha_u X^\beta_u \delta X^\gamma_u + X^\alpha_s \int_0^s X^\beta_u \delta X^\gamma_u + X^\beta_s \int_0^s X^\alpha_u \delta X^\gamma_u + X^\alpha_s X^\beta_s X^\gamma_{st} \\
&\quad - X^\alpha_s \int_0^t X^\beta_u \delta X^\gamma_u - X^\beta_s \int_0^t X^\alpha_u \delta X^\gamma_u + \int_0^t X^\alpha_u X^\beta_u \delta X^\gamma_u \\
&= \int_s^t X^\alpha_u X^\beta_u \delta X^\gamma_u - X^\alpha_s \int_s^t X^\beta_u \delta X^\gamma_u - X^\beta_s \int_s^t X^\alpha_u \delta X^\gamma_u + X^\alpha_s X^\beta_s X^\gamma_{st}
\end{aligned}
$$

$$
\begin{aligned}
&= \int_s^t X^\alpha_{su} X^\beta_{su} \delta X^\gamma_u + \int_s^t X^\alpha_u X^\beta_s \delta X^\gamma_u + \int_s^t X^\alpha_s X^\beta_u \delta X^\gamma_u - \int_s^t X^\alpha_s X^\beta_s \delta X^\gamma_u \\
&\quad - X^\alpha_s \int_s^t X^\beta_u \delta X^\gamma_u - X^\beta_s \int_s^t X^\alpha_u \delta X^\gamma_u + X^\alpha_s X^\beta_s X^\gamma_{st} \\
&= \int_s^t X^\alpha_{su} X^\beta_{su} \delta X^\gamma_u - \langle \mathbb{1}^\beta_{[0,s)}, X^\alpha \mathbb{1}^\gamma_{[s,t)} \rangle_{\mathcal{H}} - \langle \mathbb{1}^\alpha_{[0,s)}, X^\beta \mathbb{1}^\gamma_{[s,t)} \rangle_{\mathcal{H}} - X^\alpha_s X^\beta_s X^\gamma_{st} \\
&\quad + \langle X^\alpha_s \mathbb{1}^\beta_{[0,s)} + X^\beta_s \mathbb{1}^\alpha_{[0,s)}, \mathbb{1}^\gamma_{[s,t)} \rangle_{\mathcal{H}} + X^\alpha_s X^\beta_s X^\gamma_{st} \\
&= \int_s^t X^\alpha_{su} X^\beta_{su} \delta X^\gamma_u - \int_s^t X^\alpha_{su} R^{\beta\gamma}(s, \mathrm{d}u) - \int_s^t X^\beta_{su} R^{\alpha\gamma}(s, \mathrm{d}u)
\end{aligned}
$$

For the last term,

$$
\begin{aligned}
&\widehat{\boldsymbol{X}}^{\begin{smallmatrix}\bullet_\alpha\\ |\\ \bullet_\beta\\ |\\ \bullet_\gamma\end{smallmatrix}}_{st} \\
&= \langle \mathcal{S}_{\mathrm{CK}} \begin{smallmatrix}\bullet_\alpha\\ |\\ \bullet_\beta\\ |\\ \bullet_\gamma\end{smallmatrix}, \widehat{\boldsymbol{X}}_{0s} \rangle + \langle \mathcal{S}_{\mathrm{CK}} \begin{smallmatrix}\bullet_\alpha\\ |\\ \bullet_\beta\end{smallmatrix}, \widehat{\boldsymbol{X}}_{0s} \rangle X^\gamma_t + \langle \mathcal{S}_{\mathrm{CK}} \bullet^\alpha, \widehat{\boldsymbol{X}}_{0s} \rangle \widehat{\boldsymbol{X}}^{\begin{smallmatrix}\bullet_\beta\\ |\\ \bullet_\gamma\end{smallmatrix}}_{0t} + \widehat{\boldsymbol{X}}^{\begin{smallmatrix}\bullet_\alpha\\ |\\ \bullet_\beta\\ |\\ \bullet_\gamma\end{smallmatrix}}_{0t} \\
&= \langle - \begin{smallmatrix}\bullet_\alpha\\ |\\ \bullet_\beta\\ |\\ \bullet_\gamma\end{smallmatrix} + \overset{\gamma}{\bullet} \begin{smallmatrix}\bullet_\alpha\\ |\\ \bullet_\beta\end{smallmatrix} + \overset{\alpha}{\bullet} \begin{smallmatrix}\bullet_\beta\\ |\\ \bullet_\gamma\end{smallmatrix} - \overset{\alpha}{\bullet}\overset{\beta}{\bullet}\overset{\gamma}{\bullet}, \widehat{\boldsymbol{X}}_{0s} \rangle + \langle - \begin{smallmatrix}\bullet_\alpha\\ |\\ \bullet_\beta\end{smallmatrix} + \overset{\alpha}{\bullet}\overset{\beta}{\bullet}, \widehat{\boldsymbol{X}}_{0s} \rangle X^\gamma_t - X^\alpha_s \widehat{\boldsymbol{X}}^{\begin{smallmatrix}\bullet_\beta\\ |\\ \bullet_\gamma\end{smallmatrix}}_{0t} + \widehat{\boldsymbol{X}}^{\begin{smallmatrix}\bullet_\alpha\\ |\\ \bullet_\beta\\ |\\ \bullet_\gamma\end{smallmatrix}}_{0t} \\
&= - \int_{0<u<v<s} X^\alpha_u \delta X^\beta_u \delta X^\gamma_v + X^\gamma_s \int_0^s X^\alpha_u \delta X^\beta_u + X^\alpha_s \int_0^s X^\beta_u \delta X^\gamma_u - X^\alpha_s X^\beta_s X^\gamma_s \\
&\quad - X^\gamma_t \int_0^s X^\alpha_u \delta X^\beta_u + X^\alpha_s X^\beta_s X^\gamma_t - X^\alpha_s \int_0^t X^\beta_u \delta X^\gamma_u + \int_{0<u<v<t} X^\alpha_u \delta X^\beta_u \delta X^\gamma_v \\
&= \int_{s<u<v<t} X^\alpha_u \delta X^\beta_u \delta X^\gamma_v + \int_{0<u<s<v<t} X^\alpha_u \delta X^\beta_u \delta X^\gamma_v - X^\gamma_{st} \int_0^s X^\alpha_u \delta X^\beta_u - X^\alpha_s \int_s^t X^\beta_u \delta X^\gamma_u + X^\alpha_s X^\beta_s X^\gamma_{st}
\end{aligned}
$$

Now,

$$
\begin{aligned}
&\int_{s<u<v<t} X^\alpha_u \delta X^\beta_u \delta X^\gamma_v \\
&= \int_{s<r<u<v<t} \delta X^\alpha_r \delta X^\beta_u \delta X^\gamma_v + \int_{s<u<v<t} X^\alpha_s \delta X^\beta_u \delta X^\gamma_v \\
&= \int_{s<r<u<v<t} \delta X^\alpha_r \delta X^\beta_u \delta X^\gamma_v + \int_s^t \big(X^\alpha_s X^\beta_{sv} - R^{\alpha\beta}(s, \Delta(s,v)) \big) \delta X^\gamma_v \\
&= \int_{s<r<u<v<t} \delta X^\alpha_r \delta X^\beta_u \delta X^\gamma_v - \int_s^t R^{\alpha\beta}(s, \Delta(s,u)) \delta X^\gamma_u + X^\alpha_s \int_s^t X^\beta_{su} \delta X^\gamma_u - \int_s^t X^\beta_{su} R^{\alpha\gamma}(s, \mathrm{d}u)
\end{aligned}
$$

Also, using (5.18)

$$
\begin{aligned}
&\int_{0<u<s<v<t} X^\alpha_u \delta X^\beta_u \delta X^\gamma_v \\
&= \delta^{\beta;\gamma}_{u,v} \big(X^\alpha_u \mathbb{1}_{[0,s)}(u) \mathbb{1}_{[s,t)}(v) \big) \\
&= \delta^\gamma_v \big(\delta^\beta_u (X^\alpha_u \mathbb{1}_{[0,s)}(u)) \mathbb{1}_{[s,t)}(v) \big) \\
&= X^\gamma_{st} \int_0^s X^\alpha_u \delta X^\beta_u - \langle \mathcal{D} \delta^\beta_u (X^\alpha_u \mathbb{1}_{[0,s)}(u)), \mathbb{1}^\gamma_{[s,t)} \rangle_{\mathcal{H}} \\
&= X^\gamma_{st} \int_0^s X^\alpha_u \delta X^\beta_u - \langle X^\alpha \mathbb{1}^\beta_{[0,s)} + \delta^\beta_u (\mathbb{1}^\alpha_{[0,u)} \mathbb{1}_{[0,s)}(u)), \mathbb{1}^\gamma_{[s,t)} \rangle_{\mathcal{H}}
\end{aligned}
$$

$$= X^{\gamma}_{st}\int_0^s X^{\alpha}_u \delta X^{\beta}_u - \int_0^s X^{\alpha}_u R^{\beta\gamma}(\mathrm{d}u, \Delta(s,t)) - \int_0^s X^{\beta}_{us} R^{\alpha\gamma}(\mathrm{d}u, \Delta(s,t))$$

where in the last identity we have used that $\delta^{\beta}_u(\mathbb{1}^{\alpha}_{[0,u)}(r)\mathbb{1}_{[0,s)}(u)) = X^{\beta}_{rs}\mathbb{1}^{\alpha}_{[0,s)}(r)$. The last substitution we need to perform is

$$X^{\alpha}_s \int_s^t X^{\beta}_u \delta X^{\gamma}_u = X^{\alpha}_s \int_s^t X^{\beta}_{su}\delta X^{\gamma}_u + X^{\alpha}_s X^{\beta}_s X^{\gamma}_{st} - X^{\alpha}_s R^{\beta\gamma}(s, \Delta(s,t))$$

Putting everything together

$$\begin{aligned}
&\widehat{\boldsymbol{X}}^{\substack{\bullet\alpha\\ \bullet\beta\\ \bullet\gamma}}_{st} \\
&= \int_{s<r<u<v<t} \delta X^{\alpha}_r \delta X^{\beta}_u \delta X^{\gamma}_v - \int_s^t R^{\alpha\beta}(s, \Delta(s,u))\delta X^{\gamma}_u + X^{\alpha}_s \int_s^t X^{\beta}_{su}\delta X^{\gamma}_u - \int_s^t X^{\beta}_{su} R^{\alpha\gamma}(s, \mathrm{d}u) \\
&\quad + X^{\gamma}_{st}\int_0^s X^{\alpha}_u \delta X^{\beta}_u - \int_0^s X^{\alpha}_u R^{\beta\gamma}(\mathrm{d}u, \Delta(s,t)) - \int_0^s X^{\beta}_{us} R^{\alpha\gamma}(\mathrm{d}u, \Delta(s,t)) \\
&\quad - X^{\gamma}_{st}\int_0^s X^{\alpha}_u \delta X^{\beta}_u - X^{\alpha}_s \int_s^t X^{\beta}_{su}\delta X^{\gamma}_u - X^{\alpha}_s X^{\beta}_s X^{\gamma}_{st} + X^{\alpha}_s R^{\beta\gamma}(s, \Delta(s,t)) + X^{\alpha}_s X^{\beta}_s X^{\gamma}_{st} \\
&= \int_{s<r<u<v<t} \delta X^{\alpha}_r \delta X^{\beta}_u \delta X^{\gamma}_v - \int_s^t R^{\alpha\beta}(s, \Delta(s,u))\delta X^{\gamma}_u - \int_s^t X^{\beta}_{su} R^{\alpha\gamma}(s, \mathrm{d}u) \\
&\quad + \int_0^s X^{\alpha}_{us} R^{\beta\gamma}(\mathrm{d}u, \Delta(s,t)) - \int_0^s X^{\beta}_{us} R^{\alpha\gamma}(\mathrm{d}u, \Delta(s,t))
\end{aligned}$$

which completes the calculation. ■

6.2 The rough path

In this section we will show that $\widehat{\boldsymbol{X}}$ actually defines a rough paths, and show some of its properties.

Remark 6.5. The first thing to note is that, once we prove that $\widehat{\boldsymbol{X}}$ is a rough path, the formula for computing its terms of order $\leq \lfloor p \rfloor$ does not extend to its signature terms of higher order, which are instead determined by Lyons's extension theorem (i.e. by taking Young integrals). This can already be seen in the case $p < 2$, in which all signature terms of order ≥ 2 are given by Young integration, and are not given by (6.5), e.g. the signature term $\widehat{\boldsymbol{X}}^{\substack{\bullet\alpha\\ \bullet\beta}}_{0t}$ is

$$\int_0^t X^{\alpha}_u \mathrm{d}X^{\beta}_u = \int_0^t X^{\alpha}_u \delta X^{\beta}_u + \frac{1}{2}R^{\alpha\beta}(t)$$

by [CL19, CL20].

The only thing left that is needed to show that $\widehat{\boldsymbol{X}}$ is a rough path is that its sample paths and higher order terms are regular enough. One of the main ingredients that we will need is Kolmogorov's continuity criterion [RY99, Theorem 2.1], which we state in the precise form in which it will be applied.

Theorem 6.6 (Kolmogorov's continuity criterion for 2-parameter processes)**.**
Let $Z\colon [0,T]^2 \times \Omega \to \mathbb{R}$ *be a stochastic process taking two time variables and with the property that* $Z_{ss} = 0$

for $s \in [0,T]$. *Assume there exist* $a, b, K > 0$ *s.t. for all* $u, v \in [0,T]$

$$\mathbb{E}[|Z_{st}|^a] \leq K|t-s|^{2+b} \tag{6.8}$$

Then for every $0 < c < b/a$*, there exists a random variable* $J = J(K,b,c) \in L^a\Omega$*, with the property that*

$$\sup_{0 \leq s < t \leq T} |Z_{st}| \leq J(t-s)^c \qquad \textit{a.s.} \tag{6.9}$$

Here we have have used that Z_{st} can be written as $Z_{st} - Z_{ss}$, so that $(t,s) - (s,s) = (t-s, 0)$ in the referenced statement.

In order to proceed we will need to derive explicit expressions for multiple Wiener integrals of products. While it is well known that $\delta^n(f^{\otimes n})$ for $f \in \mathcal{H}$ has a representation in terms of a degree-n polynomial in $X(f) = \delta(f)$, we will need to deal with the slightly more general task of representing $\delta^n(f^1 \otimes \cdots \otimes f^n)$ for $f^k \in \mathcal{H}$. The fact that the integrand is a product still makes it possible to represent this as a polynomial, but one in several variables, since the factors are different. For this purpose we define, given $r^{ij} \in \mathbb{R}$ for $i,j = 1,\ldots,n,\ i \neq j$, the *multivariate Hermite polynomials* $\mathcal{H}^n_r \in \mathbb{R}[x^1,\ldots,x^n]$ recursively by

$$\begin{aligned} \mathcal{H}^{-1}_r &\coloneqq 0, \qquad \mathcal{H}^0_r \coloneqq 1 \\ \mathcal{H}^{n+1}_r(x^1,\ldots,x^{n+1}) &\coloneqq \mathcal{H}^n_r(x^1,\ldots,x^n)x^{n+1} - \sum_{k=1}^n r^{k,n+1}\mathcal{H}^{n-1}(x^1,\ldots,\widehat{x}^k,\ldots,x^n) \end{aligned} \tag{6.10}$$

The first three are

$$\begin{aligned} \mathcal{H}^1_r(x^1) &= x^1 \\ \mathcal{H}^2_r(x^1,x^2) &= x^1x^2 - r^{12} \\ \mathcal{H}^3_r(x^1,x^2,x^3) &= x^1x^2x^3 - r^{23}x^1 - r^{13}x^2 - r^{12}x^3 \end{aligned} \tag{6.11}$$

The next proposition implies the corresponding one for the ordinary Hermite polynomials [NP12, 2.7.7] (note that the convention used to define the Hermite polynomials is different to the one adopted in [Nua06]), since for $f^1 = \cdots = f^n$ with $\|f\|_{\mathcal{H}} = 1$ we can pick $r \equiv 1$ and $\mathcal{H}^n_1(x,\ldots,x)$ becomes the ordinary Hermite polynomial in the single variable x.

Proposition 6.7 (Polynomial representation of multiple Wiener integrals of products)**.** *Let* $f^1,\ldots,f^n \in \mathcal{H}$. *We then have*

$$\delta^n(f^1 \otimes \cdots \otimes f^{n+1}) = \mathcal{H}^n_{\langle f^\cdot, f^\cdot\rangle_{\mathcal{H}}}(X(f^1),\ldots,X(f^n)) \tag{6.12}$$

with $\langle f^\cdot, f^\cdot\rangle^{ij}_{\mathcal{H}} \coloneqq \langle f^i, f^j\rangle_{\mathcal{H}}$. *In particular, if* $0 \leq s_k < t_k \leq T$ *and* $\gamma_k \in \{1,\ldots,d\}$ *for* $k = 1,\ldots,n$ *we have*

$$\int_{[s_1,t_1]\times\cdots\times[s_n,t_n]} \delta X^{\gamma_1}_{u_1}\cdots\delta X^{\gamma_n}_{u_n} = \mathcal{H}^n_{R(\Delta(s,t),\Delta(s,t))}(X^{\gamma_1}_{s_1t_1},\ldots,X^{\gamma_n}_{s_nt_n}) \tag{6.13}$$

where $R(\Delta(s,t),\Delta(s,t))^{ij} \coloneqq R^{\gamma_i\gamma_j}(\Delta(s_i,t_i),\Delta(s_j,t_j))$ *for* $i,j = 1,\ldots,n$.

Proof. Proceed by induction; the case $n = 1$ is obvious.

$$\delta^{n+1}(f^1 \otimes \cdots \otimes f^{n+1})$$

$$
\begin{aligned}
&= \delta(\delta^n(f^1 \otimes \cdots \otimes f^n) f^{n+1}) \\
&= \mathcal{H}_r^n(X(f^1), \ldots, X(f^n)) X(f^{n+1}) - \langle \mathcal{D}\delta^n(f^1 \otimes \cdots \otimes f^n), f^{n+1} \rangle_{\mathcal{H}^1} \\
&= \mathcal{H}_r^n(X(f^1), \ldots, X(f^n)) X(f^{n+1}) - \sum_{k=0}^{n} \delta^{n-1}(f^1 \otimes \cdots \otimes \widehat{f^k} \otimes \cdots \otimes f^n) \langle f^k, f^{n+1} \rangle_{\mathcal{H}} \\
&= \mathcal{H}_r^n(X(f^1), \ldots, X(f^n)) X(f^{n+1}) - \sum_{k=1}^{n} r^{k,n+1} \mathcal{H}_r^{n-1}(X(f^1), \ldots, \widehat{X}(f^k), \ldots, X(f^n)) \\
&= \mathcal{H}_r^{n+1}(X(f^1), \ldots, X(f^{n+1}))
\end{aligned}
$$

where the rule for Malliavin-differentiating the multiple Wiener integral is a simple induction on Heisenberg's commutativity relation (5.18). The second assertion follows from the first by picking $f^k := \mathbb{1}^{\gamma_k}_{[s_k, t_k]}$. ■

Next we discuss a slight generalisation of the scaling and increment stationarity properties of fBm. Recall that for $\lambda > 0$ and $s, t \in [0, t]$

$$R(\lambda s, \lambda t) = \lambda^{2H} R(s,t), \qquad X_{\lambda t} \sim \lambda^H X_t \tag{6.14}$$

(with $\sim$ denoting identity in law), and

$$R\begin{pmatrix} s & t \\ s & t \end{pmatrix} = R\begin{pmatrix} 0 & t-s \\ 0 & t-s \end{pmatrix}, \qquad X_{st} \sim X_{t-s} \tag{6.15}$$

The next lemma will extend these properties to the multidimensional Wiener integral operator. For $0 \leq s < t < T$ we define $\mathcal{H}_{st} \subset \mathcal{H}$ to be the span in $\mathcal{H}$ of $\{\mathbb{1}^{\gamma}_{[s,u)} \mid u \in [s,t], \ \gamma = 1, \ldots, d\}$ ($\mathcal{H}_{st}^{\otimes m}$ should be interpreted as the subspace of elements of $\mathcal{H}^{\otimes m}$ that are supported in the box $[s,t)^m$). Let

$$\ell_{st} \colon \mathcal{H}_{st} \to \mathcal{H}, \qquad \mathbb{1}^{\gamma}_{[s,v)} \mapsto \frac{(t-s)^H}{T^H} \mathbb{1}^{\gamma}_{[0, \frac{T}{t-s}(v-s))}$$

for $u \in [s,t]$. For functions $f \colon [s,t] \to \mathbb{R}^d$, this is the transformation $\ell_{st} f(u) = \frac{(t-s)^H}{T^H} f(s + \frac{t-s}{T} u)$ for $u \in [0,T]$.

Lemma 6.8 (Scaling & translation invariance of the fractional multiple Wiener integral)**.**
Let X be an $(1/4, 1/2) \ni H$-fBm defined on $[0,T]$. For all $0 < s < t < T$ the diagram

$$
\begin{array}{ccc}
\mathcal{H}_{st}^{\otimes m} & \xrightarrow{\ \ell_{st}^{\otimes m}\ } & \mathcal{H}^{\otimes m} \\
& {\scriptstyle \delta^n} \searrow \quad \swarrow {\scriptstyle \delta^n} & \\
& L^2\Omega &
\end{array}
\tag{6.16}
$$

commutes in law.

Proof. We only need to show the identity in law on basis elements $\mathbb{1}^{\gamma_1, \ldots, \gamma_n}_{[s,v_1) \times \cdots \times [s, v_n)}$ for $v_1, \ldots, v_n \in [s,t]$ (the conclusion will follow once again by continuity of δ^m). This can be shown by (6.13) and the fact that the corresponding Hermite polynomials are homogeneous of degree nH, as is seen by induction on (6.10), using (6.15) and (6.14) with $\lambda = t - s$. ■

Theorem 6.9. *If X is a $1/4 < H$-fBm then $\widehat{\boldsymbol{X}}$ of Definition 6.2 is a.s. a p-rough path for $p > 1/H$.*

Proof. We will show that each individual summand in the expressions of Proposition 6.4 is of the needed regularity. More precisely, we will consider each to be a process in the t parameter started at s, and show Hölder bounds that hold uniformly in $s \in [0, T]$. We will suppress indices in all bounds, and the Hurst parameter H will be considered fixed (all constants will implicitly depend on it). It is well known (and a classical application of the Kolmogorov criterion) that X has sample paths that are a.s. c-Hölder continuous for $c < H$, or H^- Hölder continuous, as we will write from now on. We will write $\lesssim$ when the constant of proportionality only depends on the process, i.e. the Hurst parameter H, and the time horizon T (and shall generally omit mention of such dependencies); all other dependencies will be added as subscripts, e.g. $\lesssim_n$ means that the constant depends on n. For terms of higher order we have, for $n \in \mathbb{N}^*$

$$\int_s^t X_{su}\delta X_u = \delta^2(\mathbb{1}_{\Delta^2[s,t]}) \sim \frac{(t-s)^{2nH}}{T^{2nH}}\delta^2(\mathbb{1}_{\Delta^2[0,T]})$$

and therefore

$$\mathbb{E}\left[\left|\int_s^t X_{su}\delta X_u\right|^n\right] = \frac{(t-s)^{2nH}}{T^{2nH}}\mathbb{E}\left[|\delta^2(\mathbb{1}_{\Delta^2[0,T]})|^n\right] \lesssim_n (t-s)^{2nH}$$

since multiple Wiener integrals have moments of all orders, e.g. by Meyer's inequalities [Nua06, Proposition 1.5.7]. Then by Kolmogorov's criterion with $a = n, b = 2nH - 2$ and K the above constant in (6.8), we have that for all $n \in \mathbb{N}$ and all $0 < c < (2nH-2)/n$, and therefore for all $0 < c < 2H$, a.s.

$$\left|\int_s^t X_{su}\delta X_u\right| \le J|t-s|^c, \qquad s < t < T$$

with $J = J(n) \in L^n\Omega$. As for the other-2 summand

$$|R(s,\Delta(s,t))| \le |t^{2H} - s^{2H} - (t-s)^{2H}| \lesssim (t-s)^{2H}$$

since

$$t^{2H} - s^{2H} \le (t-s)^{2H} \qquad \text{for } H < 1/2 \text{ and } s \le t \tag{6.17}$$

We now tackle the order-3 terms one by one. $\delta^3(\mathbb{1}_{\Delta^3[s,t]})$ is handled in the same way as $\delta^2(\mathbb{1}_{\Delta^2[s,t]})$. Using that

$$X_{st}X_{st} = \delta(\mathbb{1}_{[s,t]}\delta(\mathbb{1}_{[s,t]})) - \langle \mathbb{1}_{[s,t]}, \mathcal{D}X_{st}\rangle = \delta^2(\mathbb{1}_{\Delta[s,t]}) - R(\Delta(s,t),\Delta(s,t))$$

we compute

$$\int_s^t X_{sw}X_{sw}\delta X_w = \delta^3(\mathbb{1}_{s<u,v<w<t}) + \int_s^t R(\Delta(s,w),\Delta(s,w))\delta X_w \tag{6.18}$$

Again, the first term is handled which are both handled in the same way as $\delta^2(\mathbb{1}_{\Delta^2[s,t]})$, and so can the second:

$$\int_s^t (w-s)^{2H}\delta X_w \sim \frac{(t-s)^{3H}}{T^{3H}}\int_0^T u^{2H}\delta X_u$$

where we have again used by scaling/translation invariance, with the conclusion now following, as before, by taking the n-th moment and applying Kolmogorov's criterion. (Alternatively, we could have used Lemma 6.10 below and the Young-Lóeve estimates.) The term $\int_s^t R(s, \Delta(s,u))\delta X_u$ is handled similarly, using (6.17), which is also helpful to control another term:

$$\begin{aligned}
&\left|\int_s^t X_{su} R(s, \mathrm{d}u)\right| \\
&\lesssim \int_s^t |X_{su}|((u-s)^{2H-1} - u^{2H-1})\mathrm{d}u \\
&\lesssim (t-s)^{H^-}((t-s)^{2H} - (t^{2H} - s^{2H})) \\
&\lesssim (t-s)^{3H^-}
\end{aligned}$$

where a^- stands for "$\forall b < a$".

The last term that we must consider is the one that involves a path dependency, $\int_0^s X_{us} R(\mathrm{d}u, \Delta(s,t))$. The methods used up to now do not work here, since the interval $[s,t]$ is only counted twice, and the presence of the integral over $[0,s]$ must be used. If $0 < a < b < T$ and the integral ranged from 0 to a, it would be $O(t-s)$ uniformly for $b \leq s \leq T$; this cannot be said in our case since the integrator has a singularity at $u = s$; what can, however, be exploited is that, as $u \nearrow s$, $|X_{us}| \to 0$ at order $(s-u)^{H^-}$. By making the substitution $u = s - (t-s)v$ we proceed to bound

$$\begin{aligned}
\left|\int_0^s X_{us} R(\mathrm{d}u, \Delta(s,t))\right| &\lesssim \int_0^s (s-u)^{H^-}((s-u)^{2H-1} - (t-u)^{2H-1})\mathrm{d}u \\
&= (t-s)^{3H^-} \int_0^{s/(t-s)} v^{H^-}(v^{2H-1} - (v+1)^{2H-1})\mathrm{d}v \\
&\leq (t-s)^{3H^-} \int_0^{+\infty} v^{H^-}(v^{2H-1} - (v+1)^{2H-1})\mathrm{d}v
\end{aligned}$$

and the integral on the last line is obviously convergent at 0, and is also at $+\infty$ since $H < 1/3$ and

$$v^{2H-1} - (v+1)^{2H-1} = (1-2H)\eta^{2H-2} \leq (1-2H)v^{2H-2}$$

for some $v < \eta < v+1$ by Lagrange's theorem.

To conclude, we have shown that for all $p > 1/H$ there exists a random constant $C = C(p,T,H)$ s.t.

$$|\widehat{\boldsymbol{X}}^{\bullet\gamma}_{st}| \leq C(t-s)^{1/p}, \quad |\widehat{\boldsymbol{X}}^{[\bullet\alpha]\bullet\beta}_{st}| \leq C(t-s)^{2/p}, \quad |\widehat{\boldsymbol{X}}^{[\bullet\alpha\,\bullet\beta]\bullet\gamma}_{st}| \leq C(t-s)^{3/p}, \quad |\widehat{\boldsymbol{X}}^{[[\bullet\alpha]\bullet\beta]\bullet\gamma}_{st}| \leq C(t-s)^{3/p}$$

for all $0 \leq s < t \leq 0$. ■

6.3 The Itô formula

In this section we will show that it is possible to define the bracket extension of $\widehat{\boldsymbol{X}}$, and we will compute all bracket terms of order ≤ 3. In particular, we will show that $\widehat{\boldsymbol{X}}$ is not quasi-geometric.

We will need the following simple observation; note that one of the summands in the order-3 ladder term in (6.4) can be reinterpreted correspondingly.

Lemma 6.10. *If $f\colon [0,T] \to \mathbb{R}^d$ is piecewise continuous and of bounded q-variation for some q s.t. $q^{-1}+H > 1$. Assume additionally f is λ-Hölder for some positive λ. Then f is both the Wiener- and Young-integrable, and the integrals coincide:*

$$\int_0^T f(u)\delta X_u = \int_0^T f(u)\mathrm{d}X_u \tag{6.19}$$

Proof. The Hölder assumption and piecewise continuity and the results of [CL20, §3.1] guarantee that $f \in \mathcal{H}$ and that, for a sequence of partitions with vanishing mesh size π_n, $\sum_{[s,t]\in\pi_n} f(s)\mathbb{1}_{[s,t)} \to f$ in $\mathcal{H}$. By continuity of $\delta\colon \mathcal{H} \subseteq \mathbb{D}^{1,2}(\mathcal{H}) \to L^2\Omega$

$$\delta(f) = L^2\lim_{n\to\infty} \delta\Big(\sum_{[s,t]\in\pi_n} f(s)\mathbb{1}_{[s,t)}\Big) = L^2\lim_{n\to\infty} \sum_{[s,t]\in\pi_n} f(s)X_{st}$$

By the finite q-variation assumption and piecewise continuity, f is also (piecewise-)Young integrable against X. This implies the same Riemann-Stieltjes sums converge a.s., necessarily to the same limit. ■

Proposition 6.11 (Bracket extension of $\widehat{\boldsymbol{X}}$)**.** $\widehat{\boldsymbol{X}}$ *admits the following consistent (as defined in Section 4.2) bracket extension:*

$$\begin{aligned}
\widehat{X}^{(\alpha\beta)}_{st} &= R^{\alpha\beta}(\Delta(s,t)) \\
\widehat{\boldsymbol{X}}^{\substack{\bullet\gamma\\ |\\ \bullet(\alpha\beta)}}_{st} &= \int_s^t X^{\gamma}_{su} R^{\alpha\beta}(\mathrm{d}u) \\
\widehat{\boldsymbol{X}}^{\substack{\bullet(\alpha\beta)\\ |\\ \bullet\gamma}}_{st} &= \int_s^t R^{\alpha\beta}(\Delta(s,u))\mathrm{d}X^{\gamma}_u \\
\widehat{X}^{(\alpha\beta\gamma)}_{st} &= 0 \\
\widehat{X}^{\left(\bullet_\gamma\, \substack{\bullet\alpha\\ |\\ \bullet\beta}\right)}_{st} &= \int_s^t \Big(R^{\alpha\gamma}(u,t)\mathrm{d}X^{\beta}_u + X^{\alpha}_u R^{\beta\gamma}(\mathrm{d}u,t)\Big) - \int_s^t \Big(R^{\alpha\gamma}(u)\mathrm{d}X^{\beta}_u + X^{\alpha}_u R^{\beta\gamma}(\mathrm{d}u)\Big) \\
&\quad + \int_0^s \Big(R^{\alpha\gamma}(u,\Delta(s,t))\mathrm{d}X^{\beta}_u + X^{\alpha}_u R^{\beta\gamma}(\mathrm{d}u,\Delta(s,t))\Big)
\end{aligned} \tag{6.20}$$

In particular, it holds a.s. that $\widehat{\boldsymbol{X}}$ is quasi-geometric if and only if $d = 1$.

Proof. By Proposition 6.4, (6.13) and (6.11) we have

$$\begin{aligned}
\widehat{X}^{(\alpha\beta)}_{st} &= \Big\langle \bullet_{\alpha}\bullet_{\beta} - \substack{\bullet\beta\\ |\\ \bullet\alpha} - \substack{\bullet\alpha\\ |\\ \bullet\beta}\,, \widehat{\boldsymbol{X}}_{st}\Big\rangle \\
&= X^{\alpha}_{st}X^{\beta}_{st} - \int_s^t X^{\alpha}_{su}\delta X^{\beta}_u - \int_s^t X^{\beta}_{su}\delta X^{\alpha}_u + 2R^{\alpha\beta}(s,\Delta(s,t)) \\
&= X^{\alpha}_{st}X^{\beta}_{st} - \int_{[s,t]^2} \delta X^{\alpha}_u \delta X^{\beta}_v + 2R^{\alpha\beta}(s,\Delta(s,t)) \\
&= R^{\alpha\beta}(\Delta(s,t),\Delta(s,t)) + 2R^{\alpha\beta}(s,\Delta(s,t)) \\
&= R^{\alpha\beta}(\Delta(s,t))
\end{aligned}$$

The obvious choices for the two lifted terms are

$$\widehat{\boldsymbol{X}}^{\substack{\bullet\gamma\\ |\\ \bullet(\alpha\beta)}}_{st} = \int_s^t X^{\gamma}_{su} R^{\alpha\beta}(\mathrm{d}u)$$

and using Lemma 6.10

$$\widehat{\boldsymbol{X}}^{[(\alpha\beta)]_\gamma}_{st} = \int_s^t R^{\alpha\beta}(\Delta(s,u))\mathrm{d}X^\gamma_u = \int_s^t R^{\alpha\beta}(\Delta(s,u))\delta X^\gamma_u = \langle\, [\alpha\,\beta]_\gamma - [[\alpha]_\beta]_\gamma - [[\beta]_\alpha]_\gamma \,, \widehat{\boldsymbol{X}}_{st}\rangle$$

which guarantees consistency, by Example 4.18. The Chen identity is trivial to check in the first case and automatic in the second. Both have the needed regularity, as can be inferred through the same methods used in the proof of Theorem 6.9.

We now focus on the order-3 terms. Using (6.18), Lemma 6.10 (and integration by parts for Young integrals) and Proposition 6.7, we have

$$\begin{aligned}
&\widehat{X}^{(\alpha\beta\gamma)}_t \\
&= \langle\, \bullet_\alpha\bullet_\beta\bullet_\gamma - [\beta\,\gamma]_\alpha - [\alpha\,\gamma]_\beta - [\alpha\,\beta]_\gamma - [\alpha]_{(\beta\gamma)} - [\beta]_{(\alpha\gamma)} - [\gamma]_{(\alpha\beta)} \,, \widehat{\boldsymbol{X}}_{st}\rangle \\
&= X^\alpha_t X^\beta_t X^\gamma_t - \int_{0<u,v<w<t} \Big[\delta X^\beta_u \delta X^\gamma_v \delta X^\alpha_w + \delta X^\alpha_u \delta X^\gamma_v \delta X^\beta_w + \delta X^\alpha_u \delta X^\beta_v \delta X^\gamma_w\Big] \\
&\quad - \int_0^t \Big[R^{\alpha\beta}(u)\mathrm{d}X^\gamma_u + R^{\alpha\gamma}(u)\mathrm{d}X^\beta_u + R^{\beta\gamma}(u)\mathrm{d}X^\alpha_u + X^\gamma_u R^{\alpha\beta}(\mathrm{d}u) + X^\beta_u R^{\alpha\gamma}(\mathrm{d}u) + X^\alpha_u R^{\beta\gamma}(\mathrm{d}u)\Big] \\
&= R^{\alpha\beta}(t)X^\gamma_t + R^{\alpha\gamma}(t)X^\beta_t + R^{\beta\gamma}(t)X^\alpha_t - \Big[R^{\alpha\beta}(t)X^\gamma_t + R^{\alpha\gamma}(t)X^\beta_t + R^{\beta\gamma}(t)X^\alpha_t\Big] \\
&= 0
\end{aligned}$$

As for the last term, we have

$$\begin{aligned}
&\widehat{X}^{(\bullet_\gamma [\alpha]_\beta)}_t \\
&= \langle\, \bullet_\gamma [\alpha]_\beta - [[\alpha]_\beta]_\gamma - [\alpha\,\gamma]_\beta - [\alpha]_{(\beta\gamma)} \,, \widehat{\boldsymbol{X}}_{0t}\rangle \\
&= X^\gamma_t \int_0^t X^\alpha_u \delta X^\beta_u - \int_{0<u<v<w<t} \delta X^\alpha_u \delta X^\beta_v \delta X^\gamma_w - \int_0^t X^\alpha_u X^\gamma_u \delta X^\beta_u - \int_0^t X^\alpha_u R^{\beta\gamma}(\mathrm{d}u) \overset{*}{=}
\end{aligned}$$

Now,

$$\begin{aligned}
&X^\gamma_t \int_0^t X^\alpha_u \delta X^\beta_u \\
&= \int_0^t X^\gamma_t X^\alpha_u \delta X^\beta_u + \int_0^t X^\alpha_u R^{\beta\gamma}(t,\mathrm{d}u) \\
&= \int_0^t \delta(\delta(\mathbb{1}^\alpha_{[0,u)})\mathbb{1}^\gamma_{[0,t)})\delta X^\beta_u + \int_0^t R^{\alpha\gamma}(u,t)\delta X^\beta_u + \int_0^t X^\alpha_u R^{\beta\gamma}(t,\mathrm{d}u) \\
&= \delta^{\alpha,\beta,\gamma}_{r,u,v}(\mathbb{1}_{\Delta^2[0,t]}(r,u)\mathbb{1}_{[0,t]}(v)) + \int_0^t R^{\alpha\gamma}(u,t)\mathrm{d}X^\beta_u + \int_0^t X^\alpha_u R^{\beta\gamma}(t,\mathrm{d}u) \\
&= \int_{0<r<u<v<t} \delta X^\alpha_r \delta X^\beta_u \delta X^\gamma_v + \int_{0<r<v<u<t} \delta X^\alpha_r \delta X^\gamma_v \delta X^\beta_u + \int_{0<v<r<u<t} \delta X^\gamma_v \delta X^\alpha_r \delta X^\beta_u \\
&\quad + \int_0^t R^{\alpha\gamma}(u,t)\mathrm{d}X^\beta_u + \int_0^t X^\alpha_u R^{\beta\gamma}(t,\mathrm{d}u)
\end{aligned}$$

so that

$$
\begin{aligned}
&\overset{*}{=} \int_{0<r<v<u<t} \delta X^\gamma_v \delta X^\alpha_r \delta X^\beta_u + \int_{0<v<r<u<t} \delta X^\alpha_r \delta X^\gamma_v \delta X^\beta_u + \int_0^t R^{\alpha\gamma}(u,t)\delta X^\beta_u + \int_0^t X^\alpha_u R^{\beta\gamma}(t,\mathrm{d}u) \\
&\quad - \int_0^t \int_{0<r,u<v<t} \delta X^\alpha_r \delta X^\gamma_v \delta X^\beta_u - \int_0^t R^{\alpha\gamma}(u)\mathrm{d}X^\beta_u - \int_0^t X^\alpha_u R^{\beta\gamma}(\mathrm{d}u) \\
&= \int_0^t \big(R^{\alpha\gamma}(u,t) - R^{\alpha\gamma}(u)\big)\mathrm{d}X^\beta_u + \int_0^t X^\alpha_u \big(R^{\beta\gamma}(\mathrm{d}u,t) - R^{\beta\gamma}(\mathrm{d}u)\big)
\end{aligned}
$$

The statement is then just obtained by taking differences. This is clearly a.s. non-zero when $d > 1$: take for instance $\alpha = \gamma \neq \beta$ and X to be $(4,3] \ni H$-fBm: the above expression reduces to a non-trivial Wiener integral. On the other hand, when $d = 1$ the above expression vanishes thanks to integration by parts. Example 4.30 then implies the statement on quasi-geometricity. ∎

The following is an immediate consequence of the above and Kelly's simple change of variable formula (4.42). Note that the change of variable formula for RDE solutions, which we do not state but that can be written directly using (4.40), is more complex, as it contains the non-simple term.

Corollary 6.12 (Simple Itô formula for $\widehat{\boldsymbol{X}}$)**.** *For $f \in C^\infty\mathbb{R}^d$ it a.s. holds that*

$$
f(X_t) - f(X_s) = \int_s^t \partial_\gamma f(X_u)\mathrm{d}\widehat{\boldsymbol{X}}^\gamma_u + \frac{1}{2}\int_s^t \partial_{\alpha\beta} f(X_u) R^{\alpha\beta}(\mathrm{d}u)
$$

Example 6.13 (The scalar case)**.** When $d = 1$ the indices are suppressed, which results in significant simplifications. Indeed, the path dependency in last term in (6.7) cancels out; moreover the last term in (6.20) also vanishes, since taking $s = 0$ we have, integrating by parts

$$
\widehat{X}_t^{(\bullet\,\mathbf{I})} = \int_0^t \big(R(u,t) - R(u)\big)\mathrm{d}X^\beta_u + \int_0^t X_u\big(R(\mathrm{d}u,t) - R(\mathrm{d}u)\big) = X_t R(t) - X_t R(t) = 0
$$

This implies the scalar rough path is quasi-geometric, and its terms have the simplified expressions (using tuple notation)

$$
\begin{aligned}
&\widehat{\boldsymbol{X}}^1_{st} = X_{st}, \qquad \widehat{\boldsymbol{X}}^{11}_{st} = \frac{1}{2}\big((X_{st})^2 - R(\Delta(s,t))\big), \qquad \widehat{\boldsymbol{X}}^{1(11)}_{st} = \int_s^t X_{su} R(\mathrm{d}u) \\
&\widehat{\boldsymbol{X}}^{(11)1}_{st} = \int_s^t R(\Delta(s,u))\mathrm{d}X_u, \qquad \widehat{\boldsymbol{X}}^{111}_{st} = \frac{1}{6}(X_{st})^3 - \frac{1}{2}X_{st}R(\Delta(s,t)), \qquad \widehat{\boldsymbol{X}}^{(111)} = 0
\end{aligned}
$$

which can be deduced from Proposition 6.4 by applying Proposition 6.7, Lemma 6.10 and integration by parts. It is simple to check that these terms satisfy the quasi-geometric relations. A 4-rough path that coincides with this up to level 3 has already been studied in [Bel19, Proposition 4.3.12] (already discussed in Example 4.35), where it is shown that its change of variable formula replicates the Cheridito-Nualart formula of [CN05]. There and in [LN05] the authors investigate an extension of the domain of the fractional Skorokhod integral which includes a.a. sample paths of the fBm even when $H \leq 1/4$. This suggests that this scalar rough path could be extended to the regime $H \leq 1/4$, and the fact that the Cheridito-Nualart formula is the same as Corollary 6.12 suggests that all further bracket relations should vanish in this case as well. It should be noted that this framework does not carry over to the multidimensional case, since it is apparent from [LN05, Theorem 3.2] that

$\int_{\Delta^2[s,t]} \delta X^\alpha \delta X^\beta$ fails to be defined even in the extended sense when $\alpha \neq \beta$.

6.4 Itô vs. Stratonovich

To conclude this thesis, we derive the formulae that convert RDEs driven by $\widehat{\boldsymbol{X}}$ into ones driven by the canonical geometric rough path, which from now on we denote $\boldsymbol{X}$. Note that we will use tuple notation for geometric rough paths.

We begin with the case $2 \leq p < 3$, corresponding to $1/3 < H \leq 1/2$ for fBm, since this will also be useful to treat the more general case. The algebra here is very simple, and the only thing that has to be computed is the difference

$$\boldsymbol{X}^{\alpha\beta}_{st} - \widehat{\boldsymbol{X}}^{\substack{\bullet\alpha\\ \bullet\beta}}_{st} = \frac{1}{2} R^{\alpha\beta}(\Delta(s,t)) \tag{6.21}$$

which is done by comparing (6.7) and the Wiener chaos expansion of the signature Example 5.31. This rough path, which closely resembles Itô integration, has already been defined (specifically for fBm) in [QX18]. As a result we obtain that if $\boldsymbol{H}$ is a controlled integrand — a condition that is only defined in terms of X — the usual Itô-Stratonovich formulae hold:

$$\int \boldsymbol{H} \mathrm{d}\boldsymbol{X} - \int \boldsymbol{H} \mathrm{d}\widehat{\boldsymbol{X}} = \frac{1}{2} \int H'_{\alpha\beta} \mathrm{d}R^{\alpha\beta}(\,\cdot\,) \tag{6.22}$$

$$\mathrm{d}Y^k = V^k_\gamma(Y)\mathrm{d}\boldsymbol{X}^\gamma \iff \mathrm{d}Y^k = V^k_\gamma(Y)\mathrm{d}\widehat{\boldsymbol{X}}^\gamma + \frac{1}{2} V_\alpha V_\beta \mathbb{1}^k(Y)\mathrm{d}R^{\alpha\beta}(\,\cdot\,) \tag{6.23}$$

where $V_\alpha V_\beta$ denotes composition of vector fields and $\mathbb{1}^k$ is the projection on the k^{th} coordinate.

Example 6.14 (Gaussian semimartingales)**.** Although we have assumed up to now that X is a fBm, this hypothesis is only really used, in addition to the properties of X as a Gaussian process, when $1/4 < H \leq 1/3$; when $p < 3$ it is not difficult to see that all statements (restricted to the case $p < 3$) carry over. For example, to check regularity of $\delta^2(\mathbb{1}_{\Delta^2[s,t]})$ one can use Example 5.31 together with regularity of $\boldsymbol{X}_{st}$ and an appropriate estimate of $\frac{1}{2}(R(s) + R(t)) - R(s,t)$. Semimartingales are particularly of interest in this regard, since there is a third type of integral, the Itô integral, that can be considered. Let $X = M + A$ be a continuous Gaussian semimartingale with M a martingale (see Example 5.6) and A of bounded variation. As usual, R will denote the covariance function of X, while R_M will denote that of M, R_A that of A, and $R_{M,A}(s,t) = \mathbb{E}[M_s A_t]$ with $R_{M,A}(t) \coloneqq R_{M,A}(t,t)$. Using (6.22), the fact that integration against $\boldsymbol{X}$ is the same as Stratonovich integration, and the classical Itô-Stratonovich formula we compute

$$\int \boldsymbol{H} \mathrm{d}\widehat{\boldsymbol{X}} = \int H \mathrm{d}X - \frac{1}{2} \int H'_{\alpha\beta;t}(R - R_M)(\mathrm{d}t) = \int H \mathrm{d}X - \frac{1}{2} \int H'_{\alpha\beta;t}(R_A + 2R_{M,A})(\mathrm{d}t)$$

As will be seen in a moment, when $H = f(X)$ is a one-form, it is integration against $\widehat{\boldsymbol{X}}$ that vanishes in expectation, while the Itô integral does not. Of course, when X is a martingale the two integration theories coincide. An example of a Gaussian semimartingale that is not a martingale, already explored in Chapter 5, is the Brownian bridge returning to the origin at time 1 Example 5.19, for which the above becomes

$$\int \boldsymbol{H} \mathrm{d}\widehat{\boldsymbol{X}} = \int H \mathrm{d}X + \sum_{\gamma=1}^{d} \int t H'_{\gamma\gamma;t} \mathrm{d}t$$

Note how, in light of (6.27) below, this implies $\mathbb{E}\int f(X)\mathrm{d}X = -\sum_\gamma \int \mathbb{E}[\partial_\gamma f_\gamma(X_u)]u\mathrm{d}u$ for suitably regular f, which is not evident from elementary semimartingale theory.

We proceed to consider the general case $p < 4$. Our main goal is to apply the framework of [HK15] to write an RDE driven by $\widehat{\boldsymbol{X}}$ in terms of one driven by an extension $\overline{\boldsymbol{X}}$ of $\boldsymbol{X}$. Unlike their main result, however, $\overline{\boldsymbol{X}}$ will be defined canonically, without using the Lyons-Victoir extension theorem: this is because we already have a geometric rough path $\boldsymbol{X}$ that we want to compare $\widehat{\boldsymbol{X}}$ to, and do not have to build one from scratch. Recall that $\overline{\boldsymbol{X}}$ is a rough path with trace indexed by trees, which must satisfy $\overline{\boldsymbol{X}}^{\psi(t)} = \widehat{\boldsymbol{X}}^t$, where ψ is the map (4.75). To index the higher order components we will use tuple notation for trees: for instance $\bullet^\alpha \bullet^\beta$ should no longer be considered a nonplanar forest, but the 2-tuple $(\bullet^\alpha, \bullet^\beta)$. Since we will never consider the branched components of $\overline{\boldsymbol{X}}$, there will be no ambiguity. We may thus define the terms

$$
\begin{aligned}
&\overline{\boldsymbol{X}}^{\bullet_\gamma} := \boldsymbol{X}^\gamma, \quad \overline{\boldsymbol{X}}^{(\bullet_\alpha,\bullet_\beta)} := \boldsymbol{X}^{\alpha\beta}, \quad \overline{\boldsymbol{X}}^{(\bullet_\alpha,\bullet_\beta,\bullet_\gamma)} := \boldsymbol{X}^{\alpha\beta\gamma}, \quad \overline{\boldsymbol{X}}^{[\bullet_\alpha]_\beta} := -R^{\alpha\beta}(\,\cdot\,) \\
&\overline{\boldsymbol{X}}_{st}^{(\bullet_\gamma,[\bullet_\alpha]_\beta)} := -\frac{1}{2}\int_s^t X_{su}^\gamma R^{\alpha\beta}(\mathrm{d}u), \quad \overline{\boldsymbol{X}}_{st}^{([\bullet_\alpha]_\beta,\bullet_\gamma)} := -\frac{1}{2}\int_s^t R^{\alpha\beta}(\Delta(s,u))\mathrm{d}X_u^\gamma
\end{aligned}
\tag{6.24}
$$

where we have used (6.21), and defined the last two terms canonically via Young integration (which satisfy the regularity condition by the same arguments used in Theorem 6.9). The calculation for the order-3 requires the expression for the Wiener chaos representation of $\boldsymbol{X}^{\alpha\beta\gamma}$ contained in Example 5.31, conjectured in Assumption 8 to hold for $p \in [3,4)$; note that three of the integrals therein can be considered Young integrals using Lemma 6.10. We do not reproduce the calculation here, which is long and uses the same techniques used in Proposition 6.4 and Proposition 6.11. The expressions obtained are:

$$
\begin{aligned}
\overline{\boldsymbol{X}}^{[\bullet_\alpha\bullet_\beta]_\gamma} &= \widehat{\boldsymbol{X}}^{[\bullet_\alpha\bullet_\beta]_\gamma} - \overline{\boldsymbol{X}}^{(\bullet_\alpha,[\bullet_\beta]_\gamma)} - \overline{\boldsymbol{X}}^{(\bullet_\beta,[\bullet_\alpha]_\gamma)} - \overline{\boldsymbol{X}}^{(\bullet_\alpha,\bullet_\beta,\bullet_\gamma)} - \overline{\boldsymbol{X}}^{(\bullet_\beta,\bullet_\alpha,\bullet_\gamma)} = 0 \\
\overline{\boldsymbol{X}}_{st}^{[[\bullet_\alpha]_\beta]_\gamma} &= \widehat{\boldsymbol{X}}_{st}^{[[\bullet_\alpha]_\beta]_\gamma} - \overline{\boldsymbol{X}}_{st}^{(\bullet_\alpha,[\bullet_\beta]_\gamma)} - \overline{\boldsymbol{X}}_{st}^{([\bullet_\alpha]_\beta,\bullet_\gamma)} - \overline{\boldsymbol{X}}_{st}^{(\bullet_\alpha,\bullet_\beta,\bullet_\gamma)} \\
&= \int_s^t R^{\beta\gamma}(u,\Delta(u,t))\mathrm{d}X_u^\alpha - \int_s^t R^{\alpha\gamma}(u,\Delta(u,t))\mathrm{d}X_u^\beta \\
&\quad + \int_0^s R^{\beta\gamma}(u,\Delta(s,t))\mathrm{d}X_u^\alpha - \int_0^s R^{\alpha\gamma}(u,\Delta(s,t))\mathrm{d}X_u^\beta
\end{aligned}
\tag{6.25}
$$

Note how, similarly to what happens in the case of the bracket extension, the above term containing the path-dependency vanishes in the scalar case and we re-obtain the formula (6.23). These calculations entered into [HK15, Theorem 5.8] yield

Theorem 6.15 (Hairer-Kelly formula for $\widehat{\boldsymbol{X}}$). *The RDEs*

$$
\begin{aligned}
&\mathrm{d}Y = V_\gamma(Y)\mathrm{d}\widehat{\boldsymbol{X}}^\gamma, \quad \textit{and} \\
&\mathrm{d}Y_t = V_\gamma(Y_t)\mathrm{d}\boldsymbol{X}_t^\gamma - \frac{1}{2}V_\alpha V_\beta(Y_t)R^{\alpha\beta}(\mathrm{d}t) + V_\alpha V_\beta V_\gamma(Y_t)\mathrm{d}\overline{\boldsymbol{X}}_t^{[[\bullet_\alpha]_\beta]_\gamma}
\end{aligned}
$$

where the last term is defined in (6.25)*, are equivalent. When* $\mathrm{d} = 1$ *the last term vanishes and the formula reduces to the one obtained via Hoffman's exponential* Theorem 4.32*.*

We end with a discussion of the following special case.

Example 6.16 (Integrals of one-forms)**.** The Hairer-Kelly formula is interesting even when applied to the case of one-forms. Given $f \in C^\infty_\mathrm{b}(\mathbb{R}^d, \mathbb{R}^{e\times d})$, with the usual trick of doubling the variables, it is possible to view the integral as an RDE solution jointly with the trivial RDE for X. Theorem 6.15 then yields the usual Itô-Stratonovich formula

$$\int f(X)\mathrm{d}\widehat{\boldsymbol{X}} = \int f(X)\mathrm{d}\boldsymbol{X} - \frac{1}{2}\int \partial_\alpha f_\beta(X)\mathrm{d}R^{\alpha\beta}(\,\cdot\,) \tag{6.26}$$

since the term in (6.25) is evaluated against the symmetry in α, β. Note that this holds only for one-forms, not more general integrands as (6.22) does, but is valid for $H > 1/4$. Also, this could not have been inferred from the very similar-looking Itô formula Corollary 6.12, which only works for exact one-forms.

A couple of interesting consequences follow. The conversion formula of [CL19, CL20] (see (5.27) above) simplifies, since the Jacobian of the flow of the explicit RDE $\mathrm{d}Y = f(X)\mathrm{d}\boldsymbol{X}$ is the identity. This means that the 2D integral vanishes, and we are left with the usual Itô-Stratonovich formula, this time with the Skorokhod integral playing the role of "Itô". Matching terms with (6.26) implies that integration against $\mathrm{d}\widehat{\boldsymbol{X}}$ and Skorokhod integration coincide for one-forms:

$$\int f(X)\mathrm{d}\widehat{\boldsymbol{X}} = \int f(X)\mathrm{d}\boldsymbol{X} - \frac{1}{2}\int \partial_\alpha f_\beta(X)\mathrm{d}R^{\alpha\beta}(\,\cdot\,) = \int f(X)\delta X \tag{6.27}$$

Of course, this relationship does not hold for more general integrands. Consider the constant integrand X_r (with zero Gubinelli derivatives): even assuming X scalar, $r \leq s < t$, we have

$$\int_s^t X_r \mathrm{d}\widehat{\boldsymbol{X}}_u - \int_s^t X_r \delta X_u = R(r, \Delta(s,t))$$

(6.27) immediately implies that integrals of one-forms against $\widehat{\boldsymbol{X}}$ vanish in expectation, a result already obtained by [QX18] for $(1/3, 1/2] \ni H$-fBm. It also implies that $\int_s^t f(X)\mathrm{d}\widehat{\boldsymbol{X}} = \int_s^t f(X)\delta X$ is measurable w.r.t. the sigma-algebra generated by the process between times s, t (6.1), since the middle term in (6.27) has this property.

Conclusions and further directions

The main purpose of this chapter was to give an example of a multidimensional branched rough path that is not (quasi-)geometric, of regularity $p \geq 3$ (and that therefore cannot be viewed as a functional on the tensor algebra), and that is defined above the paths of a stochastic process in a manner that takes probability theory into account. Along the way, we have shown that our example actually has some interesting features: its change of variable formula, its relationship to other integration theories, and its behaviour w.r.t. the filtered probability space.

It would be interesting to study the interaction between the algebra and stochastics pertaining to this example, at a deeper level. The main instance of this that we have in mind concerns a Hörmander condition for branched rough paths.

The original version of Hörmander's theorem, which concerns the existence of a smooth density for solutions to Stratonovich SDEs, can be deduced from the celebrated article [Hör67]. The search for a probabilistic proof of the same result [Mal78] lead to the development of Malliavin calculus, and much later it was realised that similar statements held true for Young differential equations [BH07] and for rough ones [CF10, CHLT15] driven by non-Markovian processes.

It is clear that the same Hörmander condition cannot work for branched rough paths: its interpretation as the negation of Frobenius's theorem [Haill, Introduction] no longer applies, since it is not true that if the vector fields of an RDE driven by a branched rough path are tangent to a submanifold, the solution will remain on the submanifold (as can already be seen from Itô calculus). The question of whether an RDE solution still admits a smooth density, however, still makes perfect sense. Optimally, the Hörmander condition should be written in a purely algebraic way, perhaps using the Lie algebra of Grossman-Larson primitives, in a way that it is conjecturable to imply existence and smoothness of the density for fairly general stochastic branched rough paths (in particular geometric ones). Once this is achieved, the idea would be to test the condition by proving a version of Hörmander's theorem for the branched rough path of this chapter. An interpretation of this in terms of the solution not remaining locally in any (time- or otherwise-dependent) positive-codimension subamifold would require the transfer principle of Chapter 4 to be extended to RDEs and reformulated in extrinsic terms.

A more challenging goal would be to study Hörmander's condition and theorem for branched rough paths *on manifolds*, where equations are given meanings by using transfer principles.

References

[AB16] John Armstrong and Damiano Brigo. Optimal approximation of SDEs on submanifolds: the Ito-vector and Ito-jet projections. arXiv:1610.03887v2, 2017 (2016). https://arxiv.org/abs/1610.03887v2.

[AB18] John Armstrong and Damiano Brigo. Intrinsic stochastic differential equations as jets. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 474(2210), 2018.

[ABCRF20] John Armstrong, Damiano Brigo, Thomas Cass, and Emilio Rossi Ferrucci. Non-geometric rough paths on manifolds. arXiv:2007.06970, 2020. https://arxiv.org/abs/2007.06970.

[ABF18] John Armstrong, Damiano Brigo, and Emilio Rossi Ferrucci. Projections of SDEs onto submanifolds. arXiv:1810.03923, 2018. https://arxiv.org/abs/1810.03923.

[ABRF19] John Armstrong, Damiano Brigo, and Emilio Rossi Ferrucci. Optimal approximation of SDEs on submanifolds: the Itô-vector and Itô-jet projections. *Proc. Lond. Math. Soc. (3)*, 119(1):176–213, 2019.

[AT03] Marc Arnaudon and Anton Thalmaier. Horizontal martingales in vector bundles. In *Séminaire de Probabilités, XXXVI*, volume 1801 of *Lecture Notes in Math.*, pages 419–456. Springer, Berlin, 2003.

[Bai19] Ismael Bailleul. Rough integrators on Banach manifolds. *Bull. Sci. Math.*, 151:51–65, 2019.

[Bau04] Fabrice Baudoin. *An Introduction to the Geometry of Stochastic Flows*. Published by Imperial College Press and distributed by World Scientific Publishing Co., 2004.

[BC07] Fabrice Baudoin and Laure Coutin. Operators associated with a stochastic differential equation driven by fractional Brownian motions. *Stochastic Processes and their Applications*, 117(5):550 – 574, 2007.

[BC19] Horatio Boedihardjo and Ilya Chevyrev. An isomorphism between branched and geometric rough paths. *Ann. Inst. H. Poincaré Probab. Statist.*, 55(2):1131–1148, 05 2019.

[BCEF20] Yvain Bruned, Charles Curry, and Kurusch Ebrahimi-Fard. Quasi-shuffle algebras and renormalisation of rough differential equations. *Bulletin of the London Mathematical Society*, 52(1):43–63, 2020.

[BD90] Ya. I. Belopolskaya and Yu. L. Dalecky. *Stochastic Equations and Differential Geometry*. Kluwer Academic Publishers, Dordrecht, 1990.

[Bel19] Carlo Bellingeri. *Ito formulae for the stochastic heat equation via the theories of rough paths and regularity structures*. Theses, Sorbonne Université, July 2019.

[Bel20] Carlo Bellingeri. Quasi-geometric rough paths and rough change of variable formula. arXiv:2009.00903, 2020. https://arxiv.org/abs/2009.00903.

[BGLY16] Horatio Boedihardjo, Xi Geng, Terry Lyons, and Danyu Yang. The signature of a rough path: Uniqueness. *Advances in mathematics (New York. 1965)*, 293:720–737, 2016.

[BH07] Fabrice Baudoin and Martin Hairer. A version of hörmander's theorem for the fractional brownian motion. *Probability theory and related fields*, 139(3):373–395, 2007.

[BL15] Youness Boutaib and Terry Lyons. A new definition of rough paths on manifolds. arXiv:1510.07833v2, 2015. https://arxiv.org/abs/1510.07833v2.

[BO10] Andreas Basse-O'Connor. Representation of Gaussian semimartingales with applications to the covariance function. *Stochastics*, 82(4):381–401, 2010.

[Bru20] Yvain Bruned. Renormalisation from non-geometric to geometric rough paths. arXiv:2007.14385, 2020. https://arxiv.org/abs/2007.14385.

[CC09] Ovidiu Calin and Der-Chen Chang. *Sub-Riemannian geometry*, volume 126 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2009. General theory and examples.

[CDL15] Thomas Cass, Bruce K. Driver, and Christian Litterer. Constrained rough paths. *Proc. Lond. Math. Soc. (3)*, 111(6):1471–1518, 2015.

[CDLL16] Thomas Cass, Bruce K. Driver, Nengli Lim, and Christian Litterer. On the integration of weakly geometric rough paths. *J. Math. Soc. Japan*, 68(4):1505–1524, 10 2016.

[CEFMMK20] Charles Curry, Kurusch Ebrahimi-Fard, Dominique Manchon, and Hans Z. Munthe-Kaas. Planarly branched rough paths and rough differential equations on homogeneous spaces. *Journal of Differential Equations*, 269(11):9740–9782, 2020.

[CF10] Thomas Cass and Peter Friz. Densities for rough differential equations under Hörmander's condition. *Ann. of Math. (2)*, 171(3):2115–2141, 2010.

[Cha10] Frédéric Chapoton. Free pre-lie algebras are free as lie algebras. *Canadian mathematical bulletin*, 53(3):425–437, 2010.

[Che77] Kuo-Tsai Chen. Iterated path integrals. *Bulletin of the American Mathematical Society*, 83(5):831 – 879, 1977.

[CHLT15] Thomas Cass, Martin Hairer, Christian Litterer, and Samy Tindel. Smoothness of the density for solutions to Gaussian rough differential equations. *The Annals of Probability*, 43(1):188 – 239, 2015.

[CL16] Ilya Chevyrev and Terry Lyons. Characteristic functions of measures on geometric rough paths. *The Annals of probability*, 44(6):4049–4082, 2016.

[CL19] Thomas Cass and Nengli Lim. A Stratonovich-Skorohod integral formula for Gaussian rough paths. *Ann. Probab.*, 47(1):1–60, 2019.

[CL20] Thomas Cass and Nengli Lim. Skorohod and rough integration for stochastic differential equations driven by Volterra processes. *L'Institut Henri Poincare, Annales B: Probabilites et Statistiques*, 2020.

[CLL12] Thomas Cass, Christian Litterer, and Terry Lyons. Rough paths on manifolds. In *New trends in stochastic analysis and related topics*, volume 12 of *Interdiscip. Math. Sci.*, pages 33–88. World Sci. Publ., Hackensack, NJ, 2012.

[CN05] Patrick Cheridito and David Nualart. Stochastic integral of divergence type with respect to fractional brownian motion with hurst parameter $h \in (0, 1/2)$. *Annales de l'I.H.P. Probabilités et statistiques*, 41(6):1049–1081, 2005.

[CQ02] Laure Coutin and Zhongmin Qian. Stochastic analysis, rough path analysis and fractional brownian motions. *Probability theory and related fields*, 122(1):108–140, 2002.

[DG78] Daniela Dohrn and Francesco Guerra. Nelson's stochastic mechanics on Riemannian manifolds. *Lett. Nuovo Cimento (2)*, 22(4):121–127, 1978.

[Dri04] Bruce K. Driver. Curved wiener space analysis. arXiv:math/0403073, 2004. https://arxiv.org/abs/math/0403073.

[Dri18] Bruce K. Driver. Global existence of geometric rough flows. arXiv:1810.03708v1, 2018. https://arxiv.org/abs/1810.03708.

[DS17] Bruce K. Driver and Jeremy S. Semko. Controlled rough paths on manifolds I. *Rev. Mat. Iberoam.*, 33(3):885–950, 2017.

[É89] Michel Émery. *Stochastic calculus in manifolds*. Universitext. Springer-Verlag, Berlin, 1989. With an appendix by Paul-André Meyer.

[É90] Michel Émery. On two transfer principles in stochastic differential geometry. In *Séminaire de Probabilités, XXIV, 1988/89*, volume 1426 of *Lecture Notes in Math.*, pages 407–441. Springer, Berlin, 1990.

[EE71] James Eells and K. David Elworthy. Wiener integration on certain manifolds. *Problems in non-linear analysis*, page 67–94, 1971.

[EFMPW15] Kurusch Ebrahimi-Fard, Simon J. A. Malham, Frédéric Patras, and Anke Wiese. Flows and stochastic taylor series in itô calculus. *Journal of Physics A: Mathematical and Theoretical*, 48(49):495202, nov 2015.

[EFP15] Kurusch Ebrahimi-Fard and Frédéric Patras. Cumulants, free cumulants and half-shuffles. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2176):20140843, 2015.

[ER03] Mohammed Errami and Francesco Russo. n-covariation, generalized Dirichlet processes and calculus with respect to finite cubic variation processes. *Stochastic Process. Appl.*, 104(2):259–299, 2003.

[Faw03] Thomas Fawcett. *Problems in stochastic analysis: connections between rough paths and non-commutative harmonic analysis*. PhD thesis, University of Oxford, 2003.

[FH14] Peter K. Friz and Martin Hairer. *A course on rough paths*. Universitext. Springer, Cham, 2014. With an introduction to regularity structures.

[FH20] Peter K. Friz and Martin Hairer. *A Course on Rough Paths: With an Introduction to Regularity Structures*. Universitext. Springer International Publishing AG, Cham, 2020.

[Fis63] Donald Fisk. *Quasi-martingales and stochastic integrals*. PhD thesis, Michigan State University, 1963.

[Foi02a] Loïc Foissy. Finite dimensional comodules over the Hopf algebra of rooted trees. *Journal of Algebra*, 255(1):89 – 120, 2002.

[Foi02b] Loïc Foissy. Les algèbres de Hopf des arbres enracinés décorés, i. *Bulletin des Sciences Mathématiques*, 126(3):193–239, 2002.

[Foi13] Loïc Foissy. An introduction to hopf algebras of trees. 2013.

[Föl81] Hans Föllmer. Calcul d'ito sans probabilités. *Séminaire de probabilités de Strasbourg*, 15:143–150, 1981.

[FS17] Peter K. Friz and Atul Shekhar. General rough integration, Lévy rough paths and a Lévy–Kintchine-type formula. *The Annals of Probability*, 45(4):2707 – 2765, 2017.

[FV10a] Peter Friz and Nicolas Victoir. Differential equations driven by Gaussian signals. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 46(2):369 – 413, 2010.

[FV10b] Peter K. Friz and Nicolas B. Victoir. *Multidimensional stochastic processes as rough paths*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Theory and applications.

[Gai94] J. G. Gaines. The algebra of iterated stochastic integrals. *Stochastics and stochastics reports*, 49(3-4):169–179, 1994.

[Gai95] J.G. Gaines. A basis for iterated stochastic integrals. *Mathematics and Computers in Simulation*, 38(1):7–11, 1995.

[Gli11] Yuri E. Gliklikh. *Global and Stochastic Analysis with Applications to Mathematical Physics*. Springer-Verlag, London, 2011.

[Gub04] Massimiliano Gubinelli. Controlling rough paths. *J. Funct. Anal.*, 216(1):86–140, 2004.

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

[Hai14] Martin Hairer. A theory of regularity structures. *Inventiones mathematicae*, 198(2):269–504, 2014.

[Haill] Martin Hairer. Advanced stochastic analysis, 2021 . http://www.hairer.org/notes/Malliavin.pdf.

[Hat02] Allen Hatcher. *Algebraic topology*. Cambridge University Press, Cambridge, 2002. http://pi.math.cornell.edu/~hatcher/AT/AT.pdf.

[HK13] Martin Hairer and David Kelly. Itô's formula via rough paths. presentation for the Stochastic Analysis Seminar 2013, Oxford, https://cims.nyu.edu/~dtkelly/slides/quasi_oxford.pdf, 2013.

[HK15] Martin Hairer and David Kelly. Geometric versus non-geometric rough paths. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(1):207–251, 2015.

[HL10] Ben Hambly and Terry Lyons. Uniqueness for the signature of a path of bounded variation and the reduced path group. *Annals of mathematics*, 171(1):109–167, 2010.

[Hof00] Michael E. Hoffman. Quasi-shuffle products. *Kluwer Academic journals*, 11(1), 2000.

[Hof03] Michael E. Hoffman. Combinatorics of rooted trees and Hopf algebras. *Transactions of the American Mathematical Society*, 355(9):3795–3811, 2003.

[Hol03] Ralf Holtkamp. Comparison of Hopf algebras on trees. *Archiv der Mathematik*, 80(4):368–383, 2003.

[Hör67] Lars Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.

[Hsu02] Elton P. Hsu. *Stochastic analysis on manifolds*, volume 38 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002.

[Itô44] Kiyosi Itô. Stochastic integral. *Proceedings of the Imperial Academy*, 20(8):519 – 524, 1944.

[Itô46] Kiyosi Itô. On a stochastic integral equation. *Proceedings of the Japan Academy*, 22(2):32 – 35, 1946.

[Itô50] Kiyosi Itô. Stochastic differential equations in a differentiable manifold. *Nagoya mathematical journal*, 1:35–47, 1950.

[Itô51] Kiyosi Itô. On a formula concerning stochastic differentials. *Nagoya Mathematical Journal*, 3(none):55 – 65, 1951.

[Jos05] Jürgen Jost. *Riemannian geometry and geometric analysis*. Universitext. Springer-Verlag, Berlin, fourth edition, 2005.

[Kel12] David Kelly. *Itô corrections in stochastic equations*. PhD thesis, University of Warwick, 2012.

[KN96] Shoshichi Kobayashi and Katsumi Nomizu. *Foundations of differential geometry. Vol. I*. Wiley Classics Library. John Wiley & Sons, Inc., New York, 1996. Reprint of the 1963 original, A Wiley-Interscience Publication.

[Kum05] David Kumar. Higher order hessian structures on manifolds. *Proceedings of the Indian Academy of Sciences. Mathematical sciences*, 115(3):259–277, 2005.

[LCL07] Terry Lyons, Michael Caruana, and Thierry Lévy. Differential equations driven by rough paths ecole d'eté de probabilités de saint-flour xxxiv-2004, 2007.

[Lee97] John M. Lee. *Riemannian manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997. An introduction to curvature.

[Lév53] Paul Lévy. *Random Functions: General Theory with Special Reference to Laplacian Random Functions*. University of California publications in statistics. University of California Press, 1953.

[LN05] Jorge A. León and David Nualart. An extension of the divergence operator for gaussian processes. *Stochastic Processes and their Applications*, 115(3):481–492, 2005.

[LQ98] Terry Lyons and Zhongmin Qian. Flow of diffeomorphisms induced by a geometric multiplicative functional. *Probability theory and related fields*, 112(1):91–119, 1998.

[LQ02] Terry Lyons and Zhongmin Qian. *System control and rough paths*. Oxford mathematical monographs. Clarendon Press, Oxford, 2002.

[LS06] Terry Lyons and Nadia Sidorova. On the radius of convergence of the logarithmic signature. *Illinois J. Math.*, 50(1-4):763–790, 2006.

[Lyo94] Terry Lyons. Differential equations driven by rough signals (I): an extension of an inequality of l. c. young. *Mathematical research letters*, 1(4):451–464, 1994.

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

[Mal78] Paul Malliavin. Stochastic calculus of variation and hypoelliptic operators. In *Proceedings of the International Symposium on Stochastic Differential Equations (Res. Inst. Math. Sci., Kyoto Univ., Kyoto, 1976)*, pages 195–263. Wiley, New York-Chichester-Brisbane, 1978.

[Man06] Dominique Manchon. Hopf algebras, from basics to applications to renormalization. arXiv:math/0408405v2, 2006. https://arxiv.org/abs/math/0408405.

[Mey81] Paul-André Meyer. Géométrie stochastique sans larmes, I. *Séminaire de probabilités de Strasbourg*, 15:44–102, 1981.

[Mey82] Paul-André Meyer. Géometrie différentielle stochastique. II. In *Seminar on Probability, XVI, Supplement*, volume 921 of *Lecture Notes in Math.*, pages 165–207. Springer, Berlin-New York, 1982.

[MKW08] Hans Z. Munthe-Kaas and Will M. Wright. On the Hopf algebraic structure of lie group integrators. *Foundations of computational mathematics*, 8(2):227–257, 2008.

[MM65] John W. Milnor and John C. Moore. On the structure of hopf algebras. *Annals of Mathematics*, 81(2):211–264, 1965.

[MN68] Benoit B. Mandelbrot and John W. Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968.

[MR99] Domenico Marinucci and Peter M. Robinson. Alternative forms of fractional brownian motion. *Journal of Statistical Planning and Inference*, 80(1):111 – 122, 1999.

[MT88] Emilio Musso and Franco Tricerri. Riemannian metrics on tangent bundles. *Ann. Mat. Pura Appl. (4)*, 150:1–19, 1988.

[Nak03] Mikio Nakahara. *Geometry, topology and physics*. Graduate Student Series in Physics. Institute of Physics, Bristol, second edition, 2003.

[Nic12] Liviu Nicolaescu. Random Morse functions and spectral geometry. arXiv:1209.0639v3, 2014 (2012). https://arxiv.org/abs/1209.0639v3.

[NP12] Ivan Nourdin and Giovanni Peccati. *Normal approximations with Malliavin calculus: from Stein's method to universality*. Cambridge tracts in mathematics; 192. Cambridge University Press, Cambridge [England]; New York, 2012.

[NRS10] Ivan Nourdin, Anthony Réveillac, and Jason Swanson. The weak Stratonovich integral with respect to fractional Brownian motion with Hurst parameter 1/6. *Electronic Journal of Probability*, 15(none):2117 – 2162, 2010.

[Nua06] David Nualart. *The Malliavin Calculus and Related Topics*. Probability and Its Applications. 2nd ed. edition, 2006.

[PAGG+18] Imanol Perez Arribas, Guy M. Goodwin, John R. Geddes, Terry Lyons, and Kate E. A. Saunders. A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder. *Translational psychiatry*, 8(1):274–7, 2018.

[Pas20] Riccardo Passeggeri. On the signature and cubature of the fractional brownian motion for h>12. *Stochastic Processes and their Applications*, 130(3):1226–1257, 2020.

[Pet06] Peter Petersen. *Riemannian geometry*, volume 171 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2006.

[Pro05] Philip E. Protter. *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2005. Second edition. Version 2.1, Corrected third printing.

[QX18] Zhongmin Qian and Xingcheng Xu. *Itô integrals for fractional Brownian motion and applications to option pricing*. arXiv:1803.00335, 2018. https://arxiv.org/abs/1803.00335.

[Rei17] Jeremy Reizenstein. Calculation of iterated-integral signatures and log signatures. arXiv:1712.02757, 2017. https://arxiv.org/abs/1712.02757.

[Reu93] Christophe. Reutenauer. *Free lie algebras*. London Mathematical Society monographs. New series ; 7. Clarendon, Oxford, 1993.

[RV93] Francesco Russo and Pierre Vallois. Forward, backward and symmetric stochastic integration. *Probability theory and related fields*, 97(3):403–421, 1993.

[RW00] L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Itô calculus, Reprint of the second (1994) edition.

[RY99] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*. Grundlehren der mathematischen Wissenschaften, 293. Springer, Berlin, [Germany] ;, 1999 - 1999.

[SA11] Arif A. Salimov and Filiz Agca. Some properties of sasakian metrics in cotangent bundles. *Mediterranean Journal of Mathematics*, 8(2):243–255, Jun 2011.

[Sas58] Shigeo Sasaki. On the differential geometry of tangent bundles of Riemannian manifolds. *Tôhoku Math. J. (2)*, 10:338–354, 1958.

[Sch82] Laurent Schwartz. Géométrie différentielle du 2ème ordre, semi-martingales et équations différentielles stochastiques sur une variété différentielle. *Séminaire de probabilités de Strasbourg*, S16:1–148, 1982.

[Str66] Ruslan L. Stratonovich. A new representation for stochastic integrals and equations. *SIAM journal on control*, 4(2):362–371, 1966.

[TZ20] Nikolas Tapia and Lorenzo Zambotti. The geometry of the space of branched rough paths. *Proceedings of the London Mathematical Society*, 121(2):220–251, 2020.

[ua] user anonymous. What is torsion in differential geometry intuitively? MathOverflow. https://mathoverflow.net/q/20510 (version: 2010-04-06).

[Wei18] Martin Weidner. *A geometric view on rough differential equations*. PhD thesis, Imperial College London, 2018. http://spiral.imperial.ac.uk/handle/10044/1/62658.

[WZ65a] Eugene Wong and Moshe Zakai. On the Convergence of Ordinary Integrals to Stochastic Integrals. *The Annals of Mathematical Statistics*, 36(5):1560 – 1564, 1965.

[WZ65b] Eugene Wong and Moshe Zakai. On the relation between ordinary and stochastic differential equations. *International Journal of Engineering Science*, 3(2):213–229, 1965.

[YI73] Kentaro Yano and Shigeru Ishihara. *Tangent and cotangent bundles: differential geometry*. Marcel Dekker, Inc., New York, 1973. Pure and Applied Mathematics, No. 16.

[You36] Laurence C. Young. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Mathematica*, 67(none):251 – 282, 1936.