Peter Friz and Martin Hairer

A Course on Rough Paths

With an introduction to regularity structures

June 2014

Errata (last update: April 2015)

Springer

To Waltraud and Rudolf Friz

and

To Xue-Mei

Preface

Since its original development in the mid-nineties by Terry Lyons, culminating in the landmark paper [Lyo98], the theory of rough paths has grown into a mature and widely applicable mathematical theory, and there are by now several monographs dedicated to the subject, notably Lyons–Qian [LQ02], Lyons et al [LCL07] and Friz–Victoir [FV10b]. So why do we believe that there is room for yet another book on this matter? Our reasons for writing this book are twofold.

First, the theory of rough paths has gathered the reputation of being difficult to access for "mainstream" probabilists because it relies on some non-trivial algebraic and / or geometric machinery. It is true that if one wishes to apply it to signals of arbitrary roughness, the general theory relies on several objects (in particular on the Hopf-algebraic properties of the free tensor algebra and the free nilpotent group embedded in it) that are unfamiliar to most probabilists. However, in our opinion, some of the most interesting applications of the theory arise in the context of stochastic differential equations, where the driving signal is Brownian motion. In this case, the theory simplifies dramatically and essentially no non-trivial algebraic or geometric objects are required at all. This simplification is certainly not novel. Indeed, early notes by Lyons, and then of Davie and Gubinelli, all took place in this simpler setting (which allows to incorporate Brownian motion and Lévy's area). However, it does appear to us that all these ideas can nowadays be put together in unprecedented simplicity, and we made a conscious choice to restrict ourselves to this simpler case throughout most of this book.

The second and main *raison d'être* of this book is that the scope of the theory has expanded dramatically over the past few years and that, in this process, the point of view has slightly shifted from the one exposed in the aforementioned monographs. While Lyons' theory was built on the integration of 1-forms, Gubinelli gave a natural extension to the integration of so-called "controlled rough paths". As a benefit, differential equations driven by rough paths can now be solved by fixed point arguments in *linear* Banach spaces which contain a sufficiently accurate (second order) local description of the solution.

This shift in perspective has first enabled the use of rough paths to provide solution theories for a number of classically ill-posed stochastic partial differential equations with *one-dimensional* spatial variables, including equations of Burgers type and the KPZ equation. More recently, the perspective which emphasises linear spaces containing sufficiently accurate local descriptions modelled on some (rough) input, spurred the development of the theory of “regularity structures” which allows to give consistent interpretations for a number of ill-posed equations, also in higher dimensions. It can be viewed as an extension of the theory of controlled rough paths, although its formulation is somewhat different. In the last chapters of this book, we give a short and rather informal (i.e. very few proofs) introduction to that theory, which in particular also sheds new light on some of the definitions of the theory of rough paths.

This book does not have the ambition to provide an exhaustive description of the theory of rough paths, but rather to complement the existing literature on the subject. As a consequence, there are a number of aspects that we chose not to touch, or to do so only barely. One omission is the study of rough paths of arbitrarily low regularity: we do provide hints at the general theory at the end of several chapters, but these are self-contained and can be skipped without impacting the understanding of the rest of the book. Another serious omission concerns the systematic study of *signatures*, that is the collection of all iterated integrals over a fixed interval associated to a sufficiently regular path, providing an intriguing nonlinear characterisation.

We have used several parts of this book for lectures and mini-courses. In particular, over the last years, the material on rough paths was given repeatedly by the first author at TU Berlin (Chapters 1-12, in the form of a 4h/week, full semester lecture for an audience of beginning graduate students in stochastics) and in some mini-courses (Vienna, Columbia, Rennes, Toulouse; e.g. Chapters 1-5 with a selection of further topics). The material of Chapters 13-15 originates in a number of minicourses by the second author (Bonn, ETHZ, Toulouse, Columbia, XVII Brazilian School of Probability, 44th St. Flour School of Probability, etc). The “KPZ and rough paths” summer school in Rennes (2013) was a particularly good opportunity to try out much of the material here in joint mini-course form – we are very grateful to the organisers for their efforts. Chapters 13-15 are, arguably, a little harder to present in a classroom. Jointly with Paul Gassiat, the first author gave this material as full lecture at TU Berlin (with examples classes run by Joscha Diehl, and more background material on Schwartz distributions, Hölder spaces and wavelet theory than what is found in this book); we also started to use consistently colours on our handouts. We felt the resulting improvement in readability was significant enough to try it out also in the present book and take the opportunity to thank Jörg Sixt from Springer for making this possible, aside from his professional assistance concerning all other aspects of this book project. We are very grateful for all the feedback we received from participants at all theses courses. Furthermore, we would like to thank Bruce Driver, Paul Gassiat, Massimilliano Gubinelli, Terry Lyons, Etienne Pardoux, Jeremy Quastel and Hendrik Weber for many interesting discussions on how to present this material. In addition, Khalil Chouk, Joscha Diehl and Sebastian Riedel kindly offered to partially proofread the final manuscript.

At last, we would like to acknowledge financial support: PKF was supported by the European Research Council under the European Union’s Seventh Framework

Programme (FP7/2007-2013) / ERC grant agreement nr. 258237 and DFG, SPP 1324. MH was supported by the Leverhulme trust through a leadership award and by the Royal Society through a Wolfson research award.

Berlin and Coventry, *Peter K. Friz*
June 2014 *Martin Hairer*

Contents

Chapter 1
Introduction

Abstract We give a short overview of the scopes of both the theory of rough paths and the theory of regularity structures. The main ideas are introduced and we point out some analogies with other branches of mathematics.

1.1 Controlled differential equations

Differential equations are omnipresent in modern pure and applied mathematics; many "pure" disciplines in fact originate in attempts to analyse differential equations from various application areas. Classical ordinary differential equations (ODEs) are of the form $\dot{Y}_t = f(Y_t, t)$; an important sub-class is given by *controlled ODEs* of the form

$$\dot{Y}_t = f_0(Y_t) + f(Y_t)\dot{X}_t \;, \tag{1.1}$$

where X models the input (taking values in $\mathbf{R}^d$, say), and Y is the output (in $\mathbf{R}^e$, say) of some system modelled by nonlinear functions f_0 and f, and by the initial state Y_0. The need for a non-smooth theory arises naturally when the system is subject to white noise, which can be understood as the scaling limit as $h \to 0$ of the discrete evolution equation

$$Y_{i+1} = Y_i + hf_0(Y_i) + \sqrt{h}f(Y_i)\xi_{i+1} \;, \tag{1.2}$$

where the (ξ_i) are i.i.d. standard Gaussian random variables. Based on martingale theory, Itô's *stochastic differential equations* (SDEs) have provided a rigorous and extremely useful mathematical framework for all this. And yet, stability is lost in the passage to continuous time: while it is trivial to solve (1.2) for a fixed realisation of $\xi_i(\omega)$, after all $(\xi_{1,} \dots \xi_T; Y_0) \mapsto Y_i$ is surely a continuous map, the continuity of the solution as a function of the driving noise is lost in the limit.

Taking $\dot{X} = \xi$ to be white noise in time (which amounts to say that X is a Brownian motion, say B), the solution map $S \colon B \mapsto Y$ to (1.1), known as *Itô map*, is a measurable map which in general lacks continuity, whatever norm one uses to equip the space of realisations of B. [1] Actually, one can show the following negative result (see [Lyo91, LCL07] as well as Exercise 5.21 below):

Proposition 1.1. *There exists no separable Banach space $\mathcal{B} \subset \mathcal{C}([0,1])$ with the following properties:*

1. *Sample paths of Brownian motions lie in $\mathcal{B}$ almost surely.*
2. *The map $(f,g) \mapsto \int_0^\cdot f(t)\dot g(t)\,dt$ defined on smooth functions extends to a continuous map from $\mathcal{B} \times \mathcal{B}$ into the space of continuous functions on $[0,1]$.*

Since, for any two distinct indices i and j, the map

$$B \mapsto \int_0^\cdot B^i(t)\,\dot B^j(t)\,dt\;, \tag{1.3}$$

is itself the solution of one of the simplest possible differential equations driven by B (take $Y \in \mathbf{R}^2$ solving $\dot Y^1 = \dot B^i$ and $\dot Y^2 = Y^1\,\dot B^j$), this shows that it takes very little for S to lack continuity. In this sense, solving SDEs is an analytically ill-posed task! On the other hand, there are well-known probabilistic well-posedness results for SDEs of the form [2]

$$dY_t = f_0(Y_t)dt + f(Y_t) \circ dB_t\;, \tag{1.4}$$

(see e.g. [INY78, Thm 4.1]), which imply for instance

Theorem 1.2. *Let $\xi_\varepsilon = \delta_\varepsilon * \xi$ denote the regularisation of white noise in time with a compactly supported smooth mollifier δ_ε. Denote by Y^ε the solutions to (1.1) driven by $\dot X = \xi_\varepsilon$. Then Y^ε converges in probability (uniformly on compact sets). The limiting process does not depend on the choice of mollifier δ_ε, and in fact is the Stratonovich solution to (1.4).*

There are many variations on such “Wong–Zakai” results, another popular choice being $\xi_\varepsilon = \dot B^{(\varepsilon)}$ where $B^{(\varepsilon)}$ is a piecewise linear approximation (of mesh size $\sim \varepsilon$) to Brownian motion. However, as consequence of the aforementioned lack of continuity of the Itô-map, there are also reasonable approximations to white noise for which the above convergence fails. (We shall see an explicit example in Section 3.4.)

Perhaps rather surprisingly, it turns out that well-posedness is restored via the iterated integrals (1.3) which are in fact the only data that is missing to turn S into a continuous map. The role of (1.3) was already appreciated in [INY78, Thm 4.1] and related works in the seventies, but statements at the time were probabilistic in nature, such as Theorem 1.2 above. *Rough path analysis* introduced by Terry Lyons in the seminal article [Lyo98] and by now exposed in several monographs [LQ02, LCL07, FV10b], provides the following remarkable insight: Itô's solution map can be factorised into a measurable “universal” map Ψ and a “nice” solution map $\hat S$ as

[1] This lack of regularity is the raison d'être for *Malliavin calculus,* a Sobolev type theory of $\mathcal{C}([0,T])$ equipped with *Wiener measure*, the law of Brownian motion.

[2] For the purpose of this introduction, all coefficients are assumed to be sufficiently nice.

$$B(\omega) \overset{\Psi}{\mapsto} (B,\mathbb{B})(\omega) \overset{\hat{S}}{\mapsto} Y(\omega). \tag{1.5}$$

The map Ψ is universal in the sense that it depends neither on the initial condition, nor on the vector fields driving the stochastic differential equation, but merely consists of *enhancing* Brownian motion with iterated integrals of the form

$$\mathbb{B}^{i,j}(s,t) = \int_s^t \big(B^i(r) - B^i(s)\big)\, dB^j(r)\ . \tag{1.6}$$

At this stage, the choice of stochastic integration in (1.6) (e.g. Itô or Stratonovich) *does* matter and probabilistic techniques are required for the construction of Ψ. Indeed, the map Ψ is only measurable and usually requires the use of some sort of stochastic integration theory (or some equivalent construction, see for example Section 10 below for a general construction in a Gaussian, non-semimartingale context).

The solution map $\hat{S}$ on the other hand, the solution map to a *rough differential equation (RDE)*, also known as *Itô–Lyons map* and discussed in Chapter 8.1, is purely deterministic and only makes use of analytical constructions. More precisely, it allows input signals to be arbitrary *rough paths* which, as discussed in Chapter 2, are objects (thought of as *enhanced paths*) of the form $(X,\mathbb{X})$, defined via certain algebraic properties (which mimic the interplay between a path and its iterated integrals) and certain analytical, Hölder-type regularity conditions. In Chapter 3 these conditions will be seen to hold true a.s. for $(B,\mathbb{B})$; a typical realisation is thus called *Brownian rough path.*

The Itô–Lyons map turns out, cf. Section 8.6, to be "nice" in the sense that it is a continuous map of both its initial condition and the driving noise $(X,\mathbb{X})$, provided that the dependency on the latter is measured in a suitable "rough path" metric. In other words, rough path analysis allows for a pathwise solution theory for SDEs i.e. for a fixed realisation of the Brownian rough path. The solution map $\hat{S}$ is however a much richer object than the original Itô map, since its construction is completely independent of the choice of stochastic integral and even of the knowledge that the driving path is Brownian. For example, if we denote by Ψ^I (resp. Ψ^S) the maps $B \mapsto (B,\mathbb{B})$ obtained by Itô (resp. Stratonovich) integration, then we have the almost sure identities

$$S^I = \hat{S} \circ \Psi^I\ , \qquad S^S = \hat{S} \circ \Psi^S\ ,$$

where S^I (resp. S^S) denotes the solution to (1.4) interpreted in the Itô (resp. Stratonovich) sense. Returning to Theorem 1.2, we see that the convergence there is really a deterministic consequence of the probabilistic question whether or not $\Psi^S(B^\varepsilon) \to \Psi^S(B)$ in probability and rough path topology, with $\dot{B}^\varepsilon = \xi^\epsilon$. This can be shown to hold in the case of mollifier, piecewise linear, and many other approximations.

So how is this Itô–Lyons map $\hat{S}$ built? In order to solve (1.1), we need to be able to make sense of the expression

$$\int_0^t f(Y_s)\, dX_s \,, \tag{1.7}$$

where Y is itself the as yet unknown solution. Here is where the usual pathwise approach breaks down: as we have seen in Proposition 1.1 it is in general impossible, even in the simplest cases, to find a Banach space of functions containing Brownian sample paths and in which (1.7) makes sense. Actually, if we measure regularity in terms of Hölder exponents, then (1.7) makes sense as a limit of Riemann sums for X and Y that are arbitrary α-Hölder continuous functions if and only if $\alpha > \frac{1}{2}$. The keyword here is *arbitrary*: in our case the function Y is anything but arbitrary! Actually, since the function Y solves (1.1), one would expect the small-scale fluctuations of Y to look exactly like the small-scale fluctuations of X in the sense that one would expect that

$$Y_{s,t} = f(Y_s) X_{s,t} + R_{s,t}$$

where, for any path F with values in a linear space, we set $F_{s,t} = F_t - F_s$, and where $R_{s,t}$ is some remainder that one would expect to be "of higher order".

Suppose now that X is a "rough path", which is to say that it has been "enhanced" with a two-parameter function $\mathbb{X}$ which should be interpreted as giving the values for

$$\mathbb{X}^{i,j}(s,t) = \int_s^t X^i_{s,r}\, dX^j_r \,. \tag{1.8}$$

Note here that this identity should be read in the reverse order from what one may be used to: it is the right hand side that is *defined* by the left hand side and not the other way around! The idea here is that if X is too rough, then we do not a priori know how to define the integral of X against itself, so we simply postulate its values. Of course, $\mathbb{X}$ cannot just be anything, but should satisfy a number of natural algebraic identities and analytical bounds, see Chapter 2 below.

Anyway, assuming that we are provided with the data $(X, \mathbb{X})$, then we know how to give meaning to the integral of components of X against other components of X: this is precisely what $\mathbb{X}$ encodes. Intuitively, this suggests that if we similarly encode the fact that Y "looks like X at small scales", then one should be able to extend the definition of (1.7) to a large enough class of integrands to include solutions to (1.1), even when $\alpha < \frac{1}{2}$. One of the achievements of rough path theory is to make this intuition precise. Indeed, in the framework of rough integration sketched here and made precise in Chapter 4, the barrier $\alpha = \frac{1}{2}$ can be lowered to $\alpha = \frac{1}{3}$. In principle, this can be lowered further by further enhancing X with iterated integrals of higher-order, but we decided to focus on the first non-trivial case for the sake of simplicity and because it already covers the most important case when X is given by a Brownian motion, or a stochastic process with properties similar to those of Brownian motion. We do however indicate very briefly in Sections 2.4, 4.5 and 7.6 how the theory can be modified to cover the case $\alpha \le \frac{1}{3}$, at least in the "geometric" case when X is a limit of smooth paths.

The simplest way for Y to "look like X" is when $Y = G(X)$ for some sufficiently regular function G. Despite what one might guess, it turns out that this particular class of functions Y is already sufficiently rich so that knowing how to define integrals of the form $\int_0^t G(X_s)\,dX_s$ for (non-gradient) functions G allows to give a meaning to equations of the type (1.1), which is the approach originally developed in [Lyo98]. More recently, Gubinelli realised in [Gub04] that, in order to be able to give a meaning to $\int_0^t Y_s\,dX_s$ given the data $(X, \mathbb{X})$, it is sufficient that Y admits a "derivative" Y' such that

$$Y_{s,t} = Y'_s X_{s,t} + R_{s,t} \;,$$

with a remainder satisfying $R_{s,t} = \mathrm{O}(|t-s|^{2\alpha})$. This extension of the original theory turns out to be quite convenient, especially when applying it to problems other than the resolution of evolution equations of the type (1.1).

An intriguing question is to what extent rough path theory, essentially a theory of controlled ordinary differential equations, can be extended to partial differential equations. In the case of finite-dimensional noise, and very loosely stated, one has for instance a statement of the following type. (See [CF09, CFO11, FO14, GT10, Tei11, DGT12] as well as Section 12.1 below.)

Theorem 1.3. *Classes of SPDEs of the form $du = F[u]\,dt + H[u] \circ dB$, with second and first order differential operators F and H, respectively, and driven by finite-dimensional noise, with the Zakai equation from filtering and stochastic Hamilton–Jacobi–Bellman (HJB) equations as examples, can be solved pathwise, i.e. for a fixed realisation of the Brownian rough path. As in the SDE case, the SPDE solution map factorises as $S^S = \hat{S} \circ \Psi^S$ where $\hat{S}$, the solution map to a rough partial differential equation (RPDE) is continuous in the rough path topology.*

As a consequence, if $\xi_\varepsilon = \delta_\varepsilon * \xi$ denotes the regularisation of white noise in time with a compactly supported smooth mollifier δ_ε that is scaled by ε, and if u^ε denotes the random PDE solutions driven by $\xi_\varepsilon dt$ (instead of $\circ dB$) then u^ε converges in probability. The limiting process does not depend on the choice of mollifier δ_ε, and is viewed as Stratonovich SPDE solution. The same conclusion holds whenever $\Psi^S(B^\varepsilon) \to \Psi^S(B)$ in probability and rough path topology.

The case of SPDEs driven by infinite-dimensional noise poses entirely different problems. Already the stochastic heat equation in space dimension one has not enough spatial regularity for additional nonlinearities of the type $g(u)\partial_x u$ (which arises in applications from path sampling [Hai11b, HW13]) or $(\partial_x u)^2$ (the Kardar–Parisi–Zhang equation) to be well-defined. In space dimension one, "spatial" rough paths indexed by x, rather than t, have proved useful here and the quest to handle dimension larger than one led to the general theory of regularity structures, see Section 1.3 below.

Rather than trying to survey all applications to date of rough paths to stochastics, let us say that the past few years have seen an explosion of results made possible by the use of rough paths theory. New stimulus to the field was given by its use in rather diverse mathematical fields, including for example quantum field theory [GL09], nonlinear PDEs [Gub12], Malliavin calculus [CFV09], non-Markovian Hörmander and ergodic theory, [CF10, HP13, CHLT12] and the analysis of chaotic behaviour in fast-slow systems [KM14].

In view of these developments, we believe that it is an opportune time to try to summarise some of the main results of the theory in a way that is as elementary as possible, yet sufficiently precise to provide a technical working knowledge of the theory. We therefore include elementary but essentially complete proofs of several of the main results, including the continuity and definition of the Itô–Lyons map, the lifting of a class of Gaussian processes to the space of rough paths, etc. In contrast to the available textbook literature [LQ02, LCL07, FV10b], we emphasize Gubinelli's view on rough integration [Gub04, Gub10] which allows to linearise many considerations and to simplify the exposition. That said, the resulting theory of rough differential equations is (immediately) seen to be equivalent to Davie's definition [Dav08] and, generally, we have tried to give a good idea what other perspectives one can take on what amounts to essentially the same objects.

1.2 Analogies with other branches of mathematics

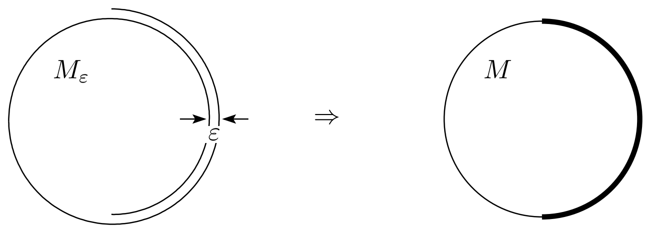
As we have just seen, the main idea of the theory of rough paths is to "enhance" a path X with some additional data $\mathbb{X}$, namely the integral of X against itself, in order to restore continuity of the Itô map. The general idea of building a larger object containing additional information in order to restore the continuity of some nonlinear transformation is of course very old and there are several other theories that have a similar "flavour" to the theory of rough paths, one of them being the theory of *Young measures* (see for example the notes [Bal00]) where the value of a function is replaced by a probability measure, thus allowing to describe limits of highly oscillatory functions.

Nevertheless, when first confronted with some of the notions just outlined, the first reaction of the reader might be that simply postulating the values for the right hand side of (1.8) makes no sense. Indeed, if X is smooth, then we "know" that there is only one "reasonable" choice for the integral $\mathbb{X}$ of X against itself, and this is the Riemann integral. How could this be replaced by something else and how can one expect to still get a consistent theory with a natural interpretation? These questions will of course be fully answered in these notes.

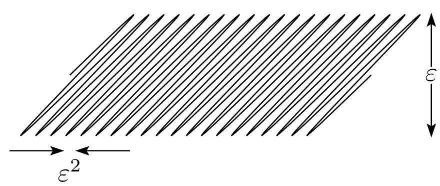
For the moment, let us draw an analogy with a very well established branch of geometric measure theory, namely the theory of *varifolds* [Alm66, LY02].

Varifolds arise as natural extensions of submanifolds in the context of certain variational problems. We are not going into details here, but loosely speaking a k-dimensional varifold in $\mathbf{R}^n$ is a (Radon) measure $\mathbf{v}$ on $\mathbf{R}^n \times \mathcal{G}(k,n)$, where $\mathcal{G}(k,n)$ denotes the space of all k-dimensional subspaces of $\mathbf{R}^n$. Here, one should interpret $\mathcal{G}(k,n)$ as the space of all possible tangent spaces at any given point for a k-dimensional submanifold of $\mathbf{R}^n$. The projection of $\mathbf{v}$ onto $\mathbf{R}^n$ should then be interpreted as a generalisation of the natural "surface measure" of a submanifold, while the conditional (probability) measure on $\mathcal{G}(k,n)$ induced at almost every point by disintegration should be interpreted as selecting a (possibly random) tangent space at each point. Why is this a reasonable extension of the notion of submanifold?

Consider the following sequence M_ε of one-dimensional submanifolds of $\mathbf{R}^2$:



It is intuitively clear that, as $\varepsilon \to 0$, this converges to a circle, but the right half has twice as much "weight" as the left half so that, if we were to describe the limit M simply as a manifold, we would have lost some information about the convergence of the surface measures in the process. More dramatically, there are situations where one has a sequence of smooth manifolds such that the limit is again a smooth manifold, but with a limiting "tangent space" which has nothing to do with the actual tangent space of the limit! Indeed, consider the sequence of one-dimensional submanifolds of $\mathbf{R}^2$ given by



This time, the limit is a piece of straight line, which is in principle a perfectly nice smooth submanifold, but the limiting tangent space is deterministic and makes a 45° angle with the canonical tangent space associated to the limit.

The situation here is philosophically very similar to that of the theory of rough paths: a subset $M \subset \mathbf{R}^n$ may be sufficiently "rough" so that there is no way of canonically associating to it either a k-dimensional Riemannian volume element, or a k-dimensional tangent space, so we simply *postulate* them. The two examples given above show that even in situations where M *is* a nice smooth manifold, it still makes sense to associate to it a volume element and / or tangent space that are different from the ones that one would construct canonically. A similar situation arises in the theory of rough paths. Indeed, it may so happen that X is actually given by a smooth function. Even so, this does *not* automatically mean that the right hand side of (1.8) is given by the usual Riemann integral of X against itself. An explicit example illustrating this fact is given in Exercise 2.17 below. Similarly to the examples of "non-canonical" varifolds given above, "non-canonical" rough paths can also be constructed as limits of ordinary smooth paths (with the second-order term $\mathbb{X}$ defined by (1.8) where the integral *is* the usual Riemann integral), provided that one takes limits in a suitably weak topology.

1.3 Regularity structures

Very recently, a new theory of "regularity structures" was introduced [Hai14c], unifying various flavours of the theory of rough paths (including Gubinelli's controlled rough paths [Gub04], as well as his branched rough paths [Gub10]), as well as the usual Taylor expansions. While it has its roots in the theory of rough paths, the main advantage of this new theory is that it is no longer tied to the one-dimensionality of the time parameter, which makes it also suitable for the description of solutions to stochastic *partial* differential equations, rather than just stochastic ordinary differential equations.

The main achievement of the theory of regularity structures is that it allows to give a (pathwise!) meaning to ill-posed stochastic PDEs that arise naturally when trying to describe the macroscopic behaviour of models from statistical mechanics near criticality. One example of such an equation is the KPZ equation arising as a natural model for one-dimensional interface motion [KPZ86, BG97, Hai13]:

$$\partial_t h = \partial_x^2 h + (\partial_x h)^2 - C + \xi \;. \tag{1.9}$$

The problem with this equation is that, if anything, one has $(\partial_x h)^2 = +\infty$ (a consequence of the roughness of $(1+1)$-dimensional space-time white noise) and one would have to compensate with $C = +\infty$. It has become custom to *define* the solution of the KPZ equation as the logarithm of the (multiplicative) stochastic heat equation $\partial_t u = \partial_x^2 u + u\xi$, essentially ignoring the (infinite) Itô-correction term.[3] The so-constructed solutions are called *Hopf–Cole solutions* and, to cite J. Quastel [Qua11],

> The evidence for the Hopf–Cole solutions is now overwhelming. Whatever the physicists mean by KPZ, it is them.

It should emphasised that previous to [Hai13], to be discussed in Chapter 15, no direct mathematical meaning had been given to the actual KPZ equation.

Another example is the dynamical Φ^4_3 model arising for example in the stochastic quantisation of Euclidean quantum field theory [PW81, JLM85, AR91, DPD03, Hai14c], as well as a universal model for phase coexistence near the critical point [GLP99]:

$$\partial_t \Phi = \Delta\Phi + C\Phi - \Phi^3 + \xi \;. \tag{1.10}$$

Here, ξ denotes $(3+1)$-dimensional space-time white noise. In contrast to the KPZ equation where the Hopf–Cole solution is a Hölder continuous random field, here Φ is at best a random Schwartz distribution, making the term Φ^3 ill-defined. Again, one formally needs to set $C = \infty$ to create suitable cancellations and so, again, the stochastic partial differential equation (1.10) has no "naïve" mathematical meaning.

Loosely speaking, the type of well-posedness results that can be proven with the help of the theory of regularity structures can be formulated as follows.

[3] This requires one of course to know that solutions to $\partial_t u = \partial_x^2 u + u\xi$ stay strictly positive with probability one, provided $u_0 > 0$ a.s., but this turns out to be the case.

Theorem 1.4. *Consider KPZ and Φ_3^4 on a bounded square spatial domain with periodic boundary conditions. Let $\xi_\varepsilon = \delta_\varepsilon * \xi$ denote the regularisation of space-time white noise with a compactly supported smooth mollifier δ_ε that is scaled by ε in the spatial direction(s) and by ε^2 in the time direction. Denote by h_ε and Φ_ε the solutions to*

$$\begin{aligned} \partial_t h_\varepsilon &= \partial_x^2 h_\varepsilon + (\partial_x h_\varepsilon)^2 - C_\varepsilon + \xi_\varepsilon \,, \\ \partial_t \Phi_\varepsilon &= \Delta \Phi_\varepsilon + \tilde C_\varepsilon \Phi_\varepsilon - \Phi_\varepsilon^3 + \xi_\varepsilon \,. \end{aligned}$$

Then, there exist choices of constants C_ε and $\tilde C_\varepsilon$ diverging as $\varepsilon \to 0$, as well as processes h and Φ such that $h_\varepsilon \to h$ and $\Phi_\varepsilon \to \Phi$ in probability. Furthermore, while the constants C_ε and $\tilde C_\varepsilon$ do *depend crucially on the choice of mollifiers δ_ε, the limits h and Φ do* not *depend on them.*

In the case of the KPZ equation, the topology in which one obtains convergence is that of convergence in probability in a suitable space of space-time Hölder continuous functions. Let us also emphasise that in this case the resulting *renormalised* solutions coincide indeed with the Hopf–Cole solutions.

In the case of the dynamical Φ_3^4 model, convergence takes place instead in some space of space-time distributions. One caveat that also has to be dealt with in the latter case is that the limiting process Φ may in principle explode in finite time for some instances of the driving noise. (Although this is of course not expected.)

The penultimate sections of this book gives a short and mostly self-contained introduction to the theory of regularity structures and the last section shows how it can be used to provide a robust solution theory for the KPZ equation. The material in these sections differs significantly in presentation from the remainder of the book. Indeed, since a detailed and rigorous exposition of this material would require an entire book by itself (see the rather lengthy articles [Hai13] and [Hai14c]), we made a conscious decision to keep the exposition mostly at an intuitive level. We therefore omit virtually all proofs (with the notable exception of the proof of the reconstruction theorem, Theorem 13.12, which is the fundamental result on which the theory builds) and instead give short glimpses of the main ideas involved.

1.4 Frequently used notations

We shall deal with paths with values in, as well as maps between, Banach spaces V, W. It will be important to consider tensor products of such Banach spaces. Assume at first that V, W are finite-dimensional, $V \cong \mathbf{R}^m, W \cong \mathbf{R}^n$. In this case the tensor product $V \otimes W$ can be identified with the matrix space $\mathbf{R}^{m \times n}$. Indeed, if $(e_i : 1 \le i \le m)$ [resp. $(f_j : 1 \le j \le n)$] is a basis of V [resp. W], then $(e_i \otimes f_j : 1 \le i \le m, 1 \le j \le n)$ is a basis of $V \otimes W$. If (e_i) and (f_j) are *orthonormal* bases it is natural to define a Euclidean structure on $V \otimes W$ by declaring the $(e_i \otimes f_j)$ to be orthonormal. This induces a norm on $V \otimes W$ which is *compatible* in the sense $|v \otimes w| \le |v| \cdot |w| \;\forall v \in V, w \in W$. When applied to $V \otimes V$ we also have the (permutation) *invariance* property, $|u \otimes v| = |v \otimes u|\ \forall u, v \in V$. A well-known and useful feature of tensor product spaces is their ability to linearise bilinear maps,[4]

$$\mathcal{L}(V \times \bar{V}, W) \cong \mathcal{L}(V \otimes \bar{V}, W).$$

In coordinates, this identification is almost trivial: any $A \in \mathcal{L}(V \times \bar{V}, W)$, i.e. any bilinear map from $V \times \bar{V}$ into W, can be expressed in terms of a 3-tensor $(A^j_{i,k})$ such that A maps $v = v^i e_i, \bar{v} = \bar{v}^k \bar{e}_k$ into $v^i \bar{v}^k A^j_{i,k} f_j \in W$. The same 3-tensor gives rise to $\bar{A} \in \mathcal{L}(V \otimes \bar{V}, W)$. Indeed, any $M = M^{i,k}(e_i \otimes \bar{e}_k) \in V \otimes \bar{V}$ is mapped linearly into $M^{i,k} A^j_{i,k} f_j \in W$. (A brief discussion how these things are adapted in an infinite-dimensional Banach setting is given in the following subsection.)

It will also be important to consider nonlinear maps between Banach spaces. Generically, we write $\mathcal{C}^n_b$ for the space of bounded continuous function, say $F : V \to W$, say, with up to n bounded, continuous derivatives in Fréchet sense, i.e. such that

$$\|F\|_{\mathcal{C}^n_b} \equiv \|F\|_\infty + \|DF\|_\infty + \ldots + \|D^n F\|_\infty < \infty$$

whenever $F \in \mathcal{C}^n_b$; recall $DF(v) \in \mathcal{L}(V, W), D^2 F \in \mathcal{L}(V, \mathcal{L}(V, W)) \cong \mathcal{L}(V \times V, W)$ and so on.

The notation $\mathcal{C}^\alpha$, for $\alpha \le 1$ is reserved for paths, such as $X, Y, \ldots$ defined on $[0, T]$, with values in some Banach space, Hölder continuous of exponent α (short: α-Hölder). For $X \in \mathcal{C}^\alpha$, the usual α-Hölder semi-norm is given by

$$\|X\|_\alpha \stackrel{\text{def}}{=} \sup_{s,t \in [0,T]} \frac{|X_{s,t}|}{|t-s|^\alpha} < \infty ,$$

where we define the *path increment* $X_{s,t} \stackrel{\text{def}}{=} X_t - X_s$ (and also use the convention $0/0 \stackrel{\text{def}}{=} 0$). As is well known, $\mathcal{C}^\alpha$ is a Banach space when equipped with the norm $X \mapsto |X_0| + \|X\|_\alpha$. When working with paths starting at the origin, the term $|X_0|$ can be omitted, i.e. we can work with directly with $\| \cdot \|_\alpha$. The same is true if we are only interested in the α-Hölder distance between two paths started at the same point $\xi \in V$. Often we shall work with partitions or dissections of $[0, T]$; since every *dissection* $\mathcal{D} = \{0 = t_0 < t_1 < \cdots < t_n = T\} \subset [0, T]$ can be thought of as a *partition* of $[0, T]$ into (essentially) disjoint intervals, $\mathcal{P} = \{[t_{i-1}, t_i] : i = 1, \ldots n\}$, and vice-versa, we shall use whatever is (notationally) more convenient. We recall that $\lim_{|\mathcal{P}| \to 0}$, typically defined via *nets*, means convergence along any sequence $(\mathcal{P}_k)$ with mesh $|\mathcal{P}_k| \to 0$, with identical limit along each such sequence. Here, the mesh $|\mathcal{P}|$ of a partition $\mathcal{P}$ is the length of its largest element, i.e. $|\mathcal{P}| = \sup_{k \in \{1,\ldots,n\}} |t_k - t_{k-1}|$ if $\mathcal{P}$ is as above.

We will frequently deal with functions $\varXi$ mapping $(s, t) \in [0, T]^2$ continuously into some Banach space and which enjoy some sort of "on-diagonal" α-Hölder regularity. More precisely, we write $\varXi(s, t) \stackrel{\text{def}}{=} \varXi_{s,t} \in \mathcal{C}^\alpha_2$ if there exists a constant C such that $|\varXi_{s,t}| \le C|t - s|^\alpha$ for all $(s, t) \in [0, T]^2$. The smallest such constant is

---

[4] This will arise naturally, with $\bar{V} = V$, when pairing the second Fréchet derivatives (of some $F : V \to W$) with second iterated integrals with values in $V \otimes V$.

then given by

$$\|\Xi\|_\alpha \stackrel{\text{def}}{=} \sup_{s,t\in[0,T]} \frac{|\Xi_{s,t}|}{|t-s|^\alpha}.$$

In particular, if X is a function defined on $[0,T]$ that is α-Hölder continuous in the usual sense, then its increments $(s,t) \mapsto X_{s,t}$ belong to $\mathcal{C}_2^\alpha$. For any such (non-trivial) path increment, one has necessarily $\alpha \le 1$, for otherwise $\dot{X} = 0$ and then $X_{s,t} = \int_s^t \dot{X} \equiv 0$. In general, however, one has non-trivial elements $\Xi \in \mathcal{C}_2^\alpha$ also for $\alpha > 1$ and indeed this is a crucial property whenever $\Xi_{s,t}$ represents some error term, since, in this case, whenever $\mathcal{P}$ is a partition of the interval $[0,T]$, one has $\sum_{[s,t]\in\mathcal{P}} |\Xi_{s,t}| \le CT|\mathcal{P}|^{\alpha-1}$, which goes to 0 with the mesh of $\mathcal{P}$.

As usual, we will use the notation $A = \mathrm{O}(x)$ if there exists a constant C such that the bound $|A| \le Cx$ holds for every $x \le 1$ (or every $x \ge 1$, depending on the context). Similarly, we write $A = \mathrm{o}(x)$ if the constant C can be made arbitrarily small as $x \to 0$ (or as $x \to \infty$, depending on the context). We will also occasionally write C for a generic constant that only depends on the data of the problem under consideration and which can change value from one line to the other without further notice.

At last, let us note that the symbols $\mathscr{C}^\alpha$, $\mathscr{D}_X^\alpha$ etc. refer to spaces of rough paths and controlled rough paths, respectively. (Both are introduced in details in the relevant sections below.)

1.5 Rough path theory works in infinite dimensions

Unless explicitly otherwise stated, all rough path results in this book are valid (with no complications in the arguments!) in a general Banach setting. Linear (or bilinear) maps are now assumed to be continuous and we still use $\mathcal{L}(...)$ for the class of such maps. What is a little more involved is the (classical) construction of a tensor product as Banach space: one completes the algebraic tensor product, $V \otimes_{\mathrm{a}} \bar{V}$, under a compatible tensor norm upon which the resulting space $V \otimes \bar{V}$ depends. What one would like, as above, is

$$\mathcal{L}(V \times \bar{V}, W) \cong \mathcal{L}(V \otimes \bar{V}, W)$$

so that for every $A \in \mathcal{L}(V \times \bar{V}, W)$ there exists a unique $\bar{A} \in \mathcal{L}(V \otimes \bar{V}, W)$ satisfying $\tilde{A}(v \otimes \bar{v}) = A(v, \bar{v})$, and such that $A \leftrightarrow \bar{A}$ is an isometric isomorphism between the Banach spaces $\mathcal{L}(V \times \bar{V}, W)$ and $\mathcal{L}(V \otimes \bar{V}, W)$. This is known to be true (e.g. [Rya02, Thm 2.9]) when $V \otimes \bar{V} = V \otimes_{\text{proj}} \bar{V}$, i.e. the closure of $V \otimes_{\mathrm{a}} \bar{V}$ under the so-called projective tensor norm.

In fact, a continuous embedding (or "canonical injection") of the form

$$\mathcal{L}(V, \mathcal{L}(V, W)) \hookrightarrow \mathcal{L}(V \otimes V, W)$$

will be enough for our purposes. For the rest of this text we shall thus make the *standing assumption* that $V \otimes V$ has been equipped with a compatible tensor norm that has this property. In many situations of interest the space V is just a copy of $\mathbf{R}^m$ and then this is trivially true. In the existing literature, such aspects are discussed in [LCL07, p19-20], [LQ02, p28,111].

Chapter 2
The space of rough paths

Abstract We define the space of (Hölder continuous) rough paths, as well as the subspace of "geometric" rough paths which preserve the usual rules of calculus. The latter can be interpreted in a natural way as paths with values in a certain nilpotent Lie group. At the end of the chapter, we give a short discussion showing how these definitions should be generalized to treat paths of arbitrarily low regularity.

2.1 Basic definitions

In this section, we give a practical definition of the space of Hölder continuous rough paths. Our choice of Hölder spaces is chiefly motivated by our hope that most readers will already be familiar with the classical Hölder spaces from real analysis. We could in the sequel have replaced "α-Hölder continuous" by "finite p-variation" for $p = 1/\alpha$ in many statements. This choice would also have been quite natural, due to the fact that one of our primary goals will be to give meaning to integrals of the form $\int f(X)\,dX$ or solutions to controlled differential equations of the form $dY = f(Y)\,dX$ for rough paths X. The value of such an integral / solution does not depend on the parametrisation of X, which dovetails nicely with the fact that the p-variation of a function is also independent of its parametrisation. This motivated its choice in the original development of the theory. In some other applications however (like the solution theory to rough stochastic *partial* differential equations developed in [Hai11b, HW13, Hai13] and more generally the theory of regularity structures [Hai14c]), parametrisation-independence is lost and the choice of Hölder norms is more natural.

A rough path on an interval $[0,T]$ with values in a Banach space V then consists of a continuous function $X\colon [0,T] \to V$, as well as a continuous "second order process" $\mathbb{X}\colon [0,T]^2 \to V \otimes V$, subject to certain algebraic and analytical conditions. Regarding the former, the behaviour of iterated integrals, such as (2.2) below, suggests to impose the algebraic relation ("*Chen's relation*"),

$$\mathbb{X}_{s,t} - \mathbb{X}_{s,u} - \mathbb{X}_{u,t} = X_{s,u} \otimes X_{u,t} \;, \tag{2.1}$$

which we assume to hold for every triple of times (s, u, t). Since $X_{t,t} = 0$, it immediately follows (take $s = u = t$) that we also have $\mathbb{X}_{t,t} = 0$ for every t. As already mentioned in the introduction, one should think of $\mathbb{X}$ as *postulating* the value of the quantity

$$\int_s^t X_{s,r} \otimes dX_r \stackrel{\text{def}}{=} \mathbb{X}_{s,t} \;, \tag{2.2}$$

where we take the right hand side as a *definition* for the left hand side. (And not the other way around!) We insist (cf. Exercise 2.7 below) that as a consequence of (2.1), knowledge of the path $t \mapsto (X_{0,t}, \mathbb{X}_{0,t})$ already determines the entire second order process $\mathbb{X}$. In this sense, the pair $(X, \mathbb{X})$ is indeed a path, and not some two-parameter object, although it is often more convenient to consider it as one. If X is a smooth function and we read (2.2) from right to left, then it is straightforward to verify (see Exercise 2.6 below) that the relation (2.1) does indeed hold. Furthermore, one can convince oneself that if $f \mapsto \int f\, dX$ denotes any form of "integration" which is linear in f, has the property that $\int_s^t dX_r = X_{s,t}$, and is such that $\int_s^t f(r)\, dX_r + \int_t^u f(r)\, dX_r = \int_s^u f(r)\, dX_r$ for any admissible integrand f, and if we use such a notion of "integral" to define $\mathbb{X}$ via (2.2), then (2.1) does automatically hold. This makes it a very natural postulate in our setting.

Note that the algebraic relations (2.1) are by themselves not sufficient to determine $\mathbb{X}$ as a function of X. Indeed, for any $V \otimes V$-valued function F, the substitution $\mathbb{X}_{s,t} \mapsto \mathbb{X}_{s,t} + F_t - F_s$ leaves the left hand side of (2.1) invariant. We will see later on how one should interpret such a substitution. It remains to discuss what are the natural analytical conditions one should impose for $\mathbb{X}$. We are going to assume that the path X itself is α-Hölder continuous, so that $|X_{s,t}| \lesssim |t - s|^\alpha$. The archetype of an α-Hölder continuous function is one which is self-similar with index α, so that $X_{\lambda s, \lambda t} \sim \lambda^\alpha X_{s,t}$.

(We intentionally do not give any mathematical definition of self-similarity here, just think of $\sim$ as having the vague meaning of "looks like".) Given (2.2), it is then very natural to expect $\mathbb{X}$ to also be self-similar, but with $\mathbb{X}_{\lambda s, \lambda t} \sim \lambda^{2\alpha} \mathbb{X}_{s,t}$. This discussion motivates the following definition of our basic spaces of rough paths.

Definition 2.1. For $\alpha \in (\frac{1}{3}, \frac{1}{2}]$, define the space of α-Hölder rough paths (over V), in symbols $\mathscr{C}^\alpha([0,T], V)$, as those pairs $(X, \mathbb{X}) =: \mathbf{X}$ such that

$$\|X\|_\alpha \stackrel{\text{def}}{=} \sup_{s \neq t \in [0,T]} \frac{|X_{s,t}|}{|t-s|^\alpha} < \infty \;, \qquad \|\mathbb{X}\|_{2\alpha} \stackrel{\text{def}}{=} \sup_{s \neq t \in [0,T]} \frac{|\mathbb{X}_{s,t}|}{|t-s|^{2\alpha}} < \infty \;, \tag{2.3}$$

and such that the algebraic constraint (2.1) is satisfied.

Remark 2.2. Given an arbitrary path $X \in \mathcal{C}^\alpha$ with values in some Banach space V it is far from obvious that this path can indeed be lifted to a rough path $(X, \mathbb{X}) \in \mathscr{C}^\alpha$. The Lyons–Victoir extension theorem [LV07] asserts that this can always be done provided $\alpha \in (\frac{1}{3}, \frac{1}{2})$, with an infinite dimensional counter example given in the case

$\alpha = 1/2$. When $\dim V < \infty$, there is no such restriction, see Proposition 13.23 below. In typical applications to stochastic processes, a "canonical" lift is constructed via probability and one does not rely on the extension theorem.

If one ignores the nonlinear constraint (2.1), there is a natural way to think of $(X, \mathbb{X})$ as an element in the Banach space $\mathcal{C}^\alpha \oplus \mathcal{C}_2^{2\alpha}$ of such maps with (semi-)norm $\|X\|_\alpha + \|\mathbb{X}\|_{2\alpha}$. However, taking into account (2.1) we see that $\mathscr{C}^\alpha$ is not a linear space, although it is a closed subset of the aforementioned Banach space. We will need (some sort of) a norm and metric on $\mathscr{C}^\alpha$. The induced "natural" norm on $\mathscr{C}^\alpha$ given by $\|X\|_\alpha + \|\mathbb{X}\|_{2\alpha}$ fails to respect the structure of (2.1) which is homogeneous with respect to a natural dilatation on $\mathscr{C}^\alpha$, given by $(X, \mathbb{X}) \mapsto (\lambda X, \lambda^2 \mathbb{X})$. This suggests to introduce the α-Hölder (homogeneous) rough path norm

$$\|\!|\mathbf{X}|\!\|_\alpha \stackrel{\text{def}}{=} \|X\|_\alpha + \sqrt{\|\mathbb{X}\|_{2\alpha}}, \tag{2.4}$$

which, although not a norm in the usual sense of normed linear spaces, is the adequate concept for the rough path $\mathbf{X} = (X, \mathbb{X})$.

Note also that the quantities defined in (2.3) are merely seminorms since they vanish for constants. Most importantly, (2.3) leads to a notation of rough path metric (and then rough path topology).

Definition 2.3. Given rough paths $\mathbf{X}, \mathbf{Y} \in \mathscr{C}^\alpha([0,T], V)$, we define the (inhomogeneous) α-Hölder rough path metric [1]

$$\varrho_\alpha(\mathbf{X}, \mathbf{Y}) := \sup_{s \neq t \in [0,T]} \frac{|X_{s,t} - Y_{s,t}|}{|t-s|^\alpha} + \sup_{s \neq t \in [0,T]} \frac{|\mathbb{X}_{s,t} - \mathbb{Y}_{s,t}|}{|t-s|^{2\alpha}}.$$

The perhaps cheapest way to show convergence with respect to this rough path metric is based on interpolation: in essence, it is enough to establish pointwise convergence in conjunction with uniform "rough path" bounds of the form (2.3); see Exercise 2.9. Let us also note that $\mathscr{C}^\alpha([0,T], V)$ so becomes a complete, metric space; the reader is asked to work out the details in Exercise 2.11.

We conclude this part with two important remarks. First, we can ask ourselves up to which point the relations (2.1) are already sufficient to determine $\mathbb{X}$. Assume that we can associate to a given function X two different second order processes $\mathbb{X}$ and $\bar{\mathbb{X}}$, and set $G_{s,t} = \mathbb{X}_{s,t} - \bar{\mathbb{X}}_{s,t}$. It then follows immediately from (2.1) that

$$G_{s,t} = G_{u,t} + G_{s,u} \;,$$

so that in particular $G_{s,t} = G_{0,t} - G_{0,s}$. Since, conversely, we already noted that setting $\bar{\mathbb{X}}_{s,t} = \mathbb{X}_{s,t} + F_t - F_s$ for an arbitrary continuous function F does not change the left hand side of (2.1), we conclude that $\mathbb{X}$ is in general determined

[1] As was already emphasised, $\mathscr{C}^\alpha$ is not a linear space but is naturally embedded in the normed space of maps $X, \mathbb{X}$; the definition of ϱ_α makes use of this. While this may not appear intrinsic (the situation is somewhat similar to using the (restricted) Euclidean metric on $\mathbf{R}^3$ on the 2-sphere), the ultimate justification is that the Itô map will turn out to be locally Lipschitz continuous in ϱ_α.

only up to the increments of some function $F \in \mathcal{C}^{2\alpha}(V \otimes V)$. The choice of F *does* usually matter and there is in general no obvious canonical choice. However, there are important examples where such a canonical choice exists and we will see in Section 10 below that such examples are provided by a large class of Gaussian processes that in particular include Brownian motion, and more generally fractional Brownian motion for every Hurst parameter $H > \frac{1}{4}$.

The second remark is that this construction can possibly be useful only if $\alpha \leq \frac{1}{2}$. Indeed, if $\alpha > \frac{1}{2}$, then a canonical choice of $\mathbb{X}$ is given by reading (2.2) from right to left and interpreting the left hand side by a simple Young integral [You36]. Furthermore, it is clear in this case that $\mathbb{X}$ must be unique, since any additional increment should be 2α-Hölder continuous by (2.3), which is of course only possible if $\alpha \leq \frac{1}{2}$. Let us stress once more however that this is *not* to say that $\mathbb{X}$ is uniquely determined by X if the latter is smooth, when it is interpreted as an element of $\mathscr{C}^\alpha$ for some $\alpha \leq \frac{1}{2}$. Indeed, if $\alpha \leq \frac{1}{2}$, F is any 2α-Hölder continuous function with values in $V \otimes V$ and $\mathbb{X}_{s,t} = F_t - F_s$, then the path $(0, \mathbb{X})$ is a perfectly "legal" element of $\mathscr{C}^\alpha$, even though one cannot get any smoother than the function 0. The impact of perturbing $\mathbb{X}$ by some $F \in \mathcal{C}^{2\alpha}$ in the context of integration is considered in Example 4.13 below. In Section 5, we shall use this for a (rough-path) understanding of how exactly Itô and Stratonovich integrals differ.

2.2 The space of geometric rough paths

While (2.1) does capture the most basic (additivity) property that one expects any decent theory of integration to respect, it does *not* imply any form of integration by parts / chain rule. Now, if one looks for a first order calculus setting, such as is valid in the context of smooth paths or the Stratonovich stochastic calculus, then for any pair e_i^*, e_j^* of elements in V^*, writing $X_t^i = e_i^*(X_t)$ and $\mathbb{X}_{s,t}^{ij} = (e_i^* \otimes e_j^*)(\mathbb{X}_{s,t})$, one would expect to have the identity

$$\begin{aligned}
\mathbb{X}_{s,t}^{ij} + \mathbb{X}_{s,t}^{ji} \text{ ``$=$'' } &\int_s^t X_{s,r}^i \, dX_r^j + \int_s^t X_{s,r}^i \, dX_r^j \\
&= \int_s^t d(X^i X^j)_r - X_s^i \, X_{s,t}^j - X_s^j \, X_{s,t}^i \\
&= (X^i X^j)_{s,t} - X_s^i \, X_{s,t}^j - X_s^j \, X_{s,t}^i = X_{s,t}^i \, X_{s,t}^j \;,
\end{aligned}$$

so that the symmetric part of $\mathbb{X}$ is determined by X. In other words, for all times s, t we have the "first order calculus" condition

$$\mathrm{Sym}(\mathbb{X}_{s,t}) = \frac{1}{2} X_{s,t} \otimes X_{s,t} \;. \tag{2.5}$$

However, if we take X to be an n-dimensional Brownian path and define $\mathbb{X}$ by Itô integration, then (2.1) still holds, but (2.5) certainly does not.

There are two natural ways to define a set of "geometric" rough paths for which (2.5) holds. On the one hand, we can define a subspace $\mathscr{C}_g^\alpha \subset \mathscr{C}^\alpha$ by stipulating that $(X, \mathbb{X}) \in \mathscr{C}_g^\alpha$ if and only if $(X, \mathbb{X}) \in \mathscr{C}^\alpha$ and (2.5) holds for every s, t. Note that $\mathscr{C}_g^\alpha$ is a closed subset of $\mathscr{C}^\alpha$. On the other hand, we have already seen that every smooth path can be lifted canonically to an element of $\mathscr{C}^\alpha$ by reading the definition (2.2) from right to left. This choice of $\mathbb{X}$ then obviously satisfies (2.5) and we can define $\mathscr{C}_g^{0,\alpha}$ as the closure of lifts of smooth paths in $\mathscr{C}^\alpha$. We leave it as exercise to the reader to see that *smooth paths* in the definition of $\mathscr{C}_g^{0,\alpha}$ may be replaced by *piecewise smooth paths* or *(piecewise)* $\mathcal{C}^1$ *paths* without changing the resulting space of geometric rough paths; see also Exercise 2.12.

One has the obvious inclusion $\mathscr{C}_g^{0,\alpha} \subset \mathscr{C}_g^\alpha$, which turns out to be strict [FV06a]. The situation is similar to the classical situation of the set of α-Hölder continuous functions being strictly larger than the closure of smooth functions under the α-Hölder norm. (Or the set of bounded measurable functions being strictly larger than $\mathcal{C}$, the closure of smooth functions under the supremum norm.) Also similar to the case of classical Hölder spaces, one has the converse inclusion $\mathscr{C}_g^\beta \subset \mathscr{C}_{g,0}^\alpha$ whenever $\beta > \alpha$, see Exercise 2.14. Let us finally mention that non-geometric rough paths can always be embedded in a space of geometric rough paths at the expense of adding new components; this is made precise in Exercise 2.14 and was systematically explored in [HK12].

2.3 Rough paths as Lie-group valued paths

We now present a very fruitful interpretation of rough paths, at least in finite dimensions, say $V = \mathbf{R}^d$. To this end, consider $X : [0,T] \to \mathbf{R}^d$, $\mathbb{X} : [0,T]^2 \to \mathbf{R}^d \otimes \mathbf{R}^d$ subject to (2.1) and define (with $X_{s,t} = X_t - X_s$ as usual)

$$\mathbf{X}_{s,t} := (1, X_{s,t}, \mathbb{X}_{s,t}) \in \mathbf{R} \oplus \mathbf{R}^d \oplus \big(\mathbf{R}^d \otimes \mathbf{R}^d\big) \stackrel{\text{def}}{=} T^{(2)}\big(\mathbf{R}^d\big). \tag{2.6}$$

The space $T^{(2)}\big(\mathbf{R}^d\big)$ has an obvious ("component-wise") vector space structure. More interestingly, for our purposes, it is a non-commutative algebra with unit element $(1,0,0)$ under

$$(a,b,c) \otimes (a',b',c') \stackrel{\text{def}}{=} (aa', ab' + a'b, ac' + a'c + b \otimes b')\;,$$

also known as *truncated tensor algebra*. This multiplicative structure is very well adapted to our needs since (2.1), combined with the obvious identity $X_{s,t} = X_{s,u} + X_{u,t}$, means precisely that (again, called "Chen's relation")

$$\mathbf{X}_{s,t} = \mathbf{X}_{s,u} \otimes \mathbf{X}_{u,t}\;.$$

Set $T_a^{(2)}\big(\mathbf{R}^d\big) = \big\{(a,b,c)\colon b \in \mathbf{R}^d,\, c \in \mathbf{R}^d \otimes \mathbf{R}^d\big\}$. As suggested in (2.6), the (affine) subspace $T_1^{(2)}\big(\mathbf{R}^d\big)$ will play a special role for us. We remark that each of its elements has an explicit inverse given by

$$(1,b,c)\otimes(1,-b,-c+b\otimes b)=(1,-b,-c+b\otimes b)\otimes(1,b,c)=(1,0,0)\ , \tag{2.7}$$

so that $T_1^{(2)}\big(\mathbf{R}^d\big)$ is a Lie group. It follows that $\mathbf{X}_{s,t}=\mathbf{X}_{0,s}^{-1}\otimes\mathbf{X}_{0,t}$ are the natural increments of the *group-valued path* $t\mapsto\mathbf{X}_{0,t}$.

Identifying $1,b,c$ with elements $(1,0,0),(0,b,0),(0,0,c)\in T^{(2)}\big(\mathbf{R}^d\big)$, we may write $(1,b,c)=1+b+c$. Computations using "formal power series" are then possible by considering the standard basis $\{e_i : 1\le i\le d\}\subset\mathbf{R}^d$ as non-commutative variables. The usual power series $(1+x)^{-1}=1-x+x^2-\dots$ then leads to

$$\begin{aligned}(1+b+c)^{-1}&=1-(b+c)+(b+c)\otimes(b+c)\\&=1-b-c+b\otimes b\ ,\end{aligned}$$

and confirms the inverse of $1+b+c$ given in (2.7). The usual power-series also suggest

$$\begin{aligned}\log(1+b+c)&\stackrel{\text{def}}{=}b+c-\frac{1}{2}b\otimes b\\ \exp(b+c)&\stackrel{\text{def}}{=}1+b+c+\frac{1}{2}b\otimes b\end{aligned} \tag{2.8}$$

and effectively allow to identify $T_0^{(2)}\big(\mathbf{R}^d\big)\cong\mathbf{R}^d\oplus\mathbf{R}^{d\times d}$, with $T_1^{(2)}\big(\mathbf{R}^d\big)=\exp\big(\mathbf{R}^d\oplus\mathbf{R}^{d\times d}\big)$. A Lie algebra structure is defined on $T_0^{(2)}\big(\mathbf{R}^d\big)$ by

$$[b+c,b'+c']=b\otimes b'-b'\otimes b\ ,$$

which is nothing but the commutator associated to the non-commutative product $\otimes$. Denote by $\mathfrak{g}^{(2)}\subset T_0^{(2)}\big(\mathbf{R}^d\big)$ the sub-algebra generated by elements of the form $(0,b,0)$. One can check that, as a Lie algebra, $\mathfrak{g}^{(2)}=\mathbf{R}^d\oplus\mathfrak{so}(d)$, i.e. the linear span of $(e_i : 1\le i\le d)$ and $(e_{ij} : 1\le i<j\le d)$, where $e_{ij}=[e_i,e_j]$. The Lie bracket of e_{ij} with any other element in $\mathfrak{g}^{(2)}$ vanishes. Since $\mathfrak{g}^{(2)}$ is closed under the operation $[\cdot,\cdot]$, its image under the exponential map, $G^{(2)}\big(\mathbf{R}^d\big):=\exp\big(\mathfrak{g}^{(2)}\big)$, is a Lie subgroup of $T_1^{(2)}\big(\mathbf{R}^d\big)$.

We call $G^{(2)}(\mathbf{R}^d)$ the step-2 nilpotent Lie group (with d generators). The algebraic constraint (2.5) then translates precisely to the statement that the path $t\mapsto\mathbf{X}_{0,t}$ (and then the increments $\mathbf{X}_{s,t}$) takes values in $G^{(2)}(\mathbf{R}^d)$.

Without going into too much details here, $G^{(2)}(\mathbf{R}^d)$ admits a natural *homogeneous "Carnot-Carathéodory norm"* $\|\cdot\|_{\mathrm{C}}$ with the property, for $\mathbf{x}=\exp(b+c)$,

$$\|\mathbf{x}\|_{\mathrm{C}}\asymp|b|+|c|^{1/2}\ , \tag{2.9}$$

where $\asymp$ indicates Lipschitz equivalence (with constants that may depend on the dimension d). A left-invariant metric d_{C}, known as the *Carnot-Carathéodory metric*, is induced by $\|\cdot\|_{\mathrm{C}}$ so that

$$d_{\mathrm{C}}(\mathbf{X}_s, \mathbf{X}_t) = \|\mathbf{X}_{s,t}\|_{\mathrm{C}} \asymp |X_{s,t}| + |\mathbb{X}_{s,t}|^{1/2}. \tag{2.10}$$

As a matter of fact, defining the "truncated signature" of a smooth path $\gamma\colon [0,1] \to \mathbf{R}^d$ by

$$G^{(2)}(\mathbf{R}^d) \ni S^{(2)}(\gamma) = \Big(1, \int_0^1 d\gamma(t), \int_0^1 \int_0^t d\gamma(s) \otimes d\gamma(t)\Big) ,$$

we have the identity

$$\|\mathbf{x}\|_{\mathrm{C}} \stackrel{\text{def}}{=} \inf\Big\{\int_0^1 |\dot\gamma(t)|\, dt \,:\, \gamma \in \mathcal{C}^1([0,1], \mathbf{R}^d) \,, \quad S^{(2)}(\gamma) = \mathbf{x}\Big\} .$$

Using the homogeneous rough path norm introduced in (2.4), taking into account (2.3), we thus have

$$\|\!|\mathbf{X}|\!\|_{\alpha;[0,T]} \asymp \sup_{s,t\in[0,T]} \frac{d_{\mathrm{C}}(\mathbf{X}_s, \mathbf{X}_t)}{|t-s|^{\alpha}} ,$$

and in particular the following appealing characterisation of geometric rough paths.

Proposition 2.4. *Let $\alpha \in \big(\frac{1}{3}, \frac{1}{2}\big]$. The following two statements are equivalent:*

1. *One has $(X, \mathbb{X}) \in \mathscr{C}_g^{\alpha}$, i.e. it satisfies (2.1), (2.3) and (2.5).*
2. *The path $t \mapsto \mathbf{X}_t = 1 + X_{0,t} + \mathbb{X}_{0,t}$ takes values in $G^{(2)}(\mathbf{R}^d)$ and is α-Hölder continuous with respect to the distance d_{C}.*

Without going into full detail, the above proposition, combined with the *geodesic* nature of the space $G^{(2)}(\mathbf{R}^d)$, shows that geometric rough paths are essentially limits of smooths paths ("geodesic approximations" in the terminology of [FV10b]) in the rough path metric.

Proposition 2.5. *Let $\beta \in \big(\frac{1}{3}, \frac{1}{2}\big]$. For every $(X, \mathbb{X}) \in \mathscr{C}_g^{\beta}\big([0,T], \mathbf{R}^d\big)$, there exists a sequence of smooth paths $X^n : [0,T] \to \mathbf{R}^d$ such that*

$$(X^n, \mathbb{X}^n) \stackrel{\text{def}}{=} \Big(X^n, \int_0^{\cdot} X^n_{0,t} \otimes dX^n_t\Big) \to (X, \mathbb{X}) \text{ uniformly on } [0,T]$$

with uniform rough path bounds $\sup_n \|X^n\|_{\beta} + \|\mathbb{X}^n\|_{2\beta} < \infty$. By interpolation, convergence holds in α-Hölder rough path metric for any $\alpha \in \big(\frac{1}{3}, \beta\big)$, namely $\lim_{n\to\infty} \varrho_{\alpha}((X^n, \mathbb{X}^n), (X, \mathbb{X})) = 0$.

2.4 Geometric rough paths of low regularity

The interpretation given above gives a strong hint on how to construct geometric rough paths with α-Hölder regularity for $\alpha \le \frac{1}{3}$: setting $p = \lfloor 1/\alpha \rfloor$, one defines the p-step truncated tensor algebra $T^{(p)}(\mathbf{R}^d)$ by

$$T^{(p)}(\mathbf{R}^d) \stackrel{\text{def}}{=} \mathbf{R} \oplus \bigoplus_{n=1}^{p} (\mathbf{R}^d)^{\otimes n} .$$

We can construct a Lie group $G^{(p)}(\mathbf{R}^d) \subset T^{(p)}(\mathbf{R}^d)$ as before, by setting $G^{(p)} = \exp(\mathfrak{g}^{(p)})$, where $\mathfrak{g}^{(p)} \subset T_0^{(p)}(\mathbf{R}^d)$ is the Lie algebra spanned by elements of the form $(1, b, 0, \ldots, 0)$. Again, one can construct a "homogeneous Carnot-Carathéodory metric" on $G^{(p)}$, with a property similar to (2.9), but with the contribution coming from the kth level scaling like $|\cdot|^{1/k}$.

A *geometric* α-Hölder rough path for arbitrary $\alpha \in \left(0, \frac{1}{2}\right]$ is then given by a function $t \mapsto \mathbf{X}_t \in G^{(p)}(\mathbf{R}^d)$ with $p = \lfloor 1/\alpha \rfloor$, which is α-Hölder continuous with respect to the corresponding distance d_{C}. It is actually also possible to extend this construction to the non-geometric setting. This is algebraically somewhat more involved and requires to keep track of more than just the "iterated integrals" of the rough path X, see [Gub10]. Again, as in Exercise 2.14, it is possible to embed spaces of non-geometric rough paths of low regularity into a suitable space of geometric rough paths. This construction however is also much more involved in the case of very low regularities and can be found in [HK12].

2.5 Exercises

Exercise 2.6. Let X be a smooth V-valued path and let $\mathbb{X}$ be given by the left hand side of (2.2), namely

$$\mathbb{X}_{s,t} = \int_s^t X_{s,r} \otimes \dot{X}_r \, dr .$$

a) Show that $\mathbb{X}$ does indeed satisfy Chen's relation (2.1).
b) Consider the collection of all iterated integrals over $[s,t]$, viewed as element in the tensor algebra over V, say

$$\mathbf{X}_{s,t} := \left(1, X_{s,t}, \mathbb{X}_{s,t}, , \int_{s<u_1<u_2<u_3<t} dX_{u_1} \otimes dX_{u_2} \otimes dX_{u_3}, \ldots \right) \in T((V)). \tag{2.11}$$

and show that the following general form of Chen's relation holds,

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

Hint: It suffices to consider the projection of $\mathbf{X}_{s,t}$ to $V^{\otimes n}$, for an arbitrary integer n, given by the n-fold integral of $dX_{u_1} \otimes \cdots \otimes dX_{u_n}$ over the simplex $\{s < u_1 < \cdots < u_n < t\}$.

Exercise 2.7. It is common to define $\mathbb{X}$ on $\Delta_{0,T} := \{(s,t) \,:\, 0 \le s \le t \le T\}$ rather than $[0,T]^2$. There is no difference however: if $\mathbb{X}_{s,t}$ is only defined for $s \le t$, show that the relation (2.1) already determines the values of $\mathbb{X}_{s,t}$ for $s > t$ and give an explicit formula. In fact, show that knowledge of the path $t \mapsto (X_{0,t}, \mathbb{X}_{0,t})$ already determines the entire second order process $\mathbb{X}$. In this sense $(X, \mathbb{X})$ is indeed a path, and not some two-parameter object.

Exercise 2.8. Consider $s \equiv \tau_0 < \tau_1 < \cdots < \tau_N \equiv t$. Show that (2.1) implies

$$\begin{aligned}\mathbb{X}_{s,t} &= \sum_{0\le i<N} \mathbb{X}_{\tau_i,\tau_{i+1}} + \sum_{0\le j<i<N} X_{\tau_j,\tau_{j+1}} \otimes X_{\tau_i,\tau_{i+1}} \\ &= \sum_{i=0}^{N-1} \big(\mathbb{X}_{\tau_i,\tau_{i+1}} + X_{s,\tau_i} \otimes X_{\tau_i,\tau_{i+1}}\big). \end{aligned} \tag{2.12}$$

Exercise 2.9 (Interpolation). Assume that $\mathbf{X}^n \in \mathscr{C}^\beta$, for $1/3 < \alpha < \beta$, with uniform bounds

$$\sup_n \|X^n\|_\beta < \infty \qquad \text{and} \qquad \sup_n \|\mathbb{X}^n\|_{2\beta} < \infty$$

and uniform convergence $X^n_{s,t} \to X_{s,t}$ and $\mathbb{X}^n_{s,t} \to \mathbb{X}_{s,t}$, i.e. uniformly over $s, t \in [0, T]$. Show that this implies $\mathbf{X} \in \mathscr{C}^\beta$ and

$$\varrho_\alpha(\mathbf{X}^n, \mathbf{X}) \to 0.$$

Show furthermore that the assumption of uniform convergence can be weakened to pointwise convergence:

$$\forall t \in [0, T]: \quad X^n_{0,t} \to X_{0,t} \quad \text{and} \quad \mathbb{X}^n_{0,t} \to \mathbb{X}_{0,t}\ .$$

Solution 2.10. Using the uniform bounds and *pointwise* convergence, there exists C such that uniformly in s, t

$$|X_{s,t}| = \lim_n \big|X^n_{s,t}\big| \le C|t-s|^\beta\ , \qquad |\mathbb{X}_{s,t}| = \lim_n \big|\mathbb{X}^n_{s,t}\big| \le C|t-s|^{2\beta}\ .$$

It readily follows that $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\beta$. In combination with the assumed uniform convergence, there exists $\varepsilon_n \to 0$, such that, uniformly in s, t,

$$\begin{aligned} |X_{s,t} - X^n_{s,t}| \le \varepsilon_n\ , &\qquad |X_{s,t} - X^n_{s,t}| \le 2C|t-s|^\beta\ , \\ |\mathbb{X}^n_{s,t} - \mathbb{X}_{s,t}| \le \varepsilon_n\ , &\qquad |\mathbb{X}^n_{s,t} - \mathbb{X}_{s,t}| \le 2C|t-s|^{2\beta}\ . \end{aligned}$$

By geometric interpolation ($a \wedge b \le a^{1-\theta} b^\theta$ when $a, b > 0$ and $0 < \theta < 1$) with $\theta = \alpha/\beta$ we have

$$|X_{s,t} - X^n_{s,t}| \lesssim \varepsilon_n^{1-\alpha/\beta} |t-s|^\alpha\ , \qquad |\mathbb{X}^n_{s,t} - \mathbb{X}_{s,t}| \lesssim \varepsilon_n^{1-\alpha/\beta} |t-s|^{2\alpha}\ ,$$

and the desired ϱ_α-convergence follows.

It remains to weaken the assumption to pointwise convergence. By Chen's relation, pointwise convergence of $\mathbf{X}^n_{0,t}$ for all t actually implies pointwise convergence of $\mathbf{X}^n_{s,t}$ for all s, t. We claim that, thanks to the uniform Hölder bounds, this implies uniform convergence. Indeed, given $\varepsilon > 0$, pick a (finite) dissection D of $[0,T]$ with small enough mesh so that $C|D|^\beta < \varepsilon/8$. Given $s,t \in [0,T]$ write $\hat{s},\hat{t}$ for the nearest points in D and note that

$$\begin{aligned}|X_{s,t} - X^n_{s,t}| &\le |X_{\hat{s},\hat{t}} - X^n_{\hat{s},\hat{t}}| + |X_{s,\hat{s}}| + |X^n_{s,\hat{s}}| + |X_{t,\hat{t}}| + |X^n_{t,\hat{t}}| \\ &\le |X_{\hat{s},\hat{t}} - X^n_{\hat{s},\hat{t}}| + \varepsilon/2 \;.\end{aligned}$$

By picking n large enough, $|X_{\hat{s},\hat{t}} - X^n_{\hat{s},\hat{t}}|$ can also be bounded by $\varepsilon/2$, uniformly over the (finitely many!) points in D, so that $X^n \to X$ uniformly. Although the second level is handled similarly, the non-additivity of $(s,t) \mapsto \mathbb{X}_{s,t}$ requires some extra care, (2.1). For simplicity of notation only, we assume $s < \hat{s} < t = \hat{t}$ so that

$$|\mathbb{X}_{s,t} - \mathbb{X}^n_{s,t}| \le |\mathbb{X}_{s,\hat{s}} - \mathbb{X}^n_{\hat{s},t}| + |\mathbb{X}_{\hat{s},t}| + |X_{s,\hat{s}} \otimes X_{\hat{s},t} - X^n_{s,\hat{s}} \otimes X^n_{\hat{s},t}|.$$

It remains to write the last summand as $|X_{s,\hat{s}} \otimes (X_{\hat{s},t} - X^n_{\hat{s},t}) - (X^n_{s,\hat{s}} - X_{s,\hat{s}}) \otimes X^n_{\hat{s},t}|$ and to repeat the same reasoning as in the first level.

Exercise 2.11. Check that $\mathscr{C}^\alpha([0,T],V)$ is a complete metric space under the metric $|X_0 - Y_0| + \varrho_\alpha(\mathbf{X},\mathbf{Y})$.

Assuming that $\dim V \ge 1$ to avoid trivialities, show that $\mathscr{C}^\alpha([0,T],V)$ is not separable. **Hint:** Reduce to the case of scalar Hölder paths on $[0,1]$; non-separability of such spaces is well known.

Exercise 2.12. a) Define the space of geometric (α-Hölder) rough paths

$$\mathscr{C}^{0,\alpha}_g([0,T],V) \subset \mathscr{C}^\alpha([0,T],V)$$

as the ϱ_α-closure of smooth paths (enhanced with their iterated Riemann integrals) in $\mathscr{C}^\alpha([0,T],V)$. Assuming that V is separable, show that $\mathscr{C}^{0,\alpha}_g([0,T],V)$ is also separable.

b) Show that for every geometric $1/2$-Hölder rough path, $\mathbf{X} \in \mathscr{C}^{0,1/2}_g$, $\mathbb{X}$ is necessarily the iterated Riemann-Stieltjes integral of the underlying path $X \in \mathcal{C}^{0,1/2}$. (Attention, this does not mean that for every $X \in \mathcal{C}^{0,1/2}$ the iterated Riemann-Stieltjes integral exist! A counterexample is found in [FV10b, Ex.9.14 (iii)].)

Solution 2.13. Let Q be a countable, dense subset of V and consider the space Λ_n of paths which are piecewise linear between level-n dyadic rationals $\mathbb{D}^n := \{kT/2^n : 0 \le k \le 2^n\}$, and, at level-$n$ dyadic points, take values in Q. Clearly $\Lambda = \cup\Lambda_n$ is countable for each Λ_n is in one-to-one correspondence with the (2^n+1)-fold Cartesian product of Q. It is easy to see that each smooth X is the limit in $\mathcal{C}^1$ of some sequence $(X^n) \subset \Lambda$. Indeed, one can take X^n to be the piecewise linear dyadic approximation, modified such that $X^n|_{\mathbb{D}^n}$ takes values in Q and such that $|(X^n - X)|_{\mathbb{D}^n}| < 1/n$. By continuity of the map $X \in \mathcal{C}^1 \mapsto \big(X, \int X \otimes dX\big) \in \mathscr{C}^\alpha$ in the respective topologies (we could even take $\alpha = 1$), we have more than enough to assert that every lifted smooth path, $\big(X, \int X \otimes dX\big)$, is the ϱ_α-limit of lifted paths in Λ. It is then easy to see that every ϱ_α limit point of lifted smooth path is also the ϱ_α-limit of lifted paths in Λ.

Turning to the second part of the question, it is not hard to see that

$$\mathscr{C}_g^{0,\alpha} \subset \left\{ \mathbf{X} \in \mathscr{C}^\alpha : \sup_{s,t:|t-s|<\varepsilon} \frac{|X_{s,t}|}{|t-s|^\alpha} \to 0, \sup_{s,t:|t-s|<\varepsilon} \frac{|\mathbb{X}_{s,t}|}{|t-s|^{2\alpha}} \to 0 \text{ as } \varepsilon \to 0 \right\}.$$

Consider now the case $\alpha = 1/2$ and a dissection $\{s = \tau_0 < \tau_1 < \cdots < \tau_N = t\}$ with mesh $\leq \varepsilon$. It follows from Chen's relation (2.1) that

$$\begin{aligned} \Big|\mathbb{X}_{s,t} - \sum_{0\leq i<n} X_{s,\tau_i} \otimes X_{\tau_i,\tau_{i+1}}\Big| &= \Big|\sum_{0\leq i<n} \mathbb{X}_{\tau_i,\tau_{i+1}}\Big| \\ &\leq C(\varepsilon) \sum_{0\leq i<n} |\tau_{i+1} - \tau_i|^{2\alpha} = TC(\varepsilon). \end{aligned}$$

It follows that $\mathbb{X}_{s,t}$ is the limit of the above Riemann-Stieltjes sum.

Exercise 2.14. One can also consider "non-geometric" separable subspaces of $\mathscr{C}^\alpha$. Consider $1/3 < \alpha < 1/2$ (in view of the previous exercise there is no point in taking $\alpha = 1/2$ here) and define

$$\mathscr{C}^{0,\alpha}([0,T],V) \subset \mathscr{C}^\alpha([0,T],V)$$

as the ϱ_α-closure of smooth paths and their iterated integrals plus smooth $V \otimes V$-valued path increments. Show that

$$\mathscr{C}^{0,\alpha}([0,T],V) \cong \mathscr{C}_g^{0,\alpha}([0,T],V) \oplus \mathcal{C}^{0,2\alpha}([0,T],V \otimes V).$$

Define the (non-separable) space of weak geometric α-Hölder rough paths, $\mathscr{C}_g^\alpha$ as those elements $\mathbf{X} \in \mathscr{C}^\alpha$ for which $2\,\mathrm{Sym}\,(\mathbb{X}) = X \otimes X$. Show that $\mathscr{C}_g^{0,\alpha}$ is a closed subspace of $\mathscr{C}_g^\alpha$ and that

$$\mathscr{C}^\alpha([0,T],V) \cong \mathscr{C}_g^\alpha([0,T],V) \oplus \mathcal{C}^{2\alpha}([0,T],V \otimes V).$$

The point of this exercise is that non-geometric rough path spaces can effectively be embedded in geometric rough path spaces.

Exercise 2.15. At least when $\dim V < \infty$, there is not much difference between $\mathscr{C}_g^{0,\alpha} \subset \mathscr{C}_g^\alpha$ in the following sense. Let $\frac{1}{3} < \alpha < \beta \leq \frac{1}{2}$. By using the (non-trivial!) fact that every $\mathbf{X} \in \mathscr{C}_g^\beta$ can be approximated uniformly by smooth paths, with uniform β-Hölder rough path bounds, use interpolation to see that $\mathbf{X} \in \mathscr{C}_g^{0,\alpha}$, in fact show that one has the compact embedding

$$\mathscr{C}_g^\beta \hookrightarrow \mathscr{C}_g^{0,\alpha}.$$

Show a similar statement for non-geometric rough path spaces.

Solution 2.16. $\mathscr{C}^\beta \subset \mathscr{C}^{0,\alpha}$ (and in fact a continuous embedding) is obvious from the interpolation exercise above. The compactness of the embedding is a consequence of Arzela-Ascoli (use $\dim V < \infty$). At last the extension to non-geometric rough path spaces, is fairly straightforward using the embedding into geometric rough path spaces.

Exercise 2.17 (Pure area rough path). Identify $\mathbf{R}^2$ with the complex numbers and consider

$$[0,1] \ni t \mapsto n^{-1} \exp\left(2\pi i n^2 t\right) \equiv X^n.$$

a) Set $\mathbb{X}^n_{s,t} := \int_s^t X^n_{s,r} \otimes dX^n_r$. Show that, for fixed $s < t$,

$$X^n_{s,t} \to 0, \qquad \mathbb{X}^n_{s,t} \to \pi(t-s)\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{2.13}$$

b) Establish the uniform bounds $\sup_n \|X^n\|_{1/2} < \infty$ and $\sup_n \|\mathbb{X}^n\|_1 < \infty$.

c) Conclude by interpolation that (2.13) takes place in α-Hölder rough path metric ϱ_α for any $1/3 < \alpha < 1/2$.

Solution 2.18. a) Obviously, $X^n_{s,t} = \mathrm{O}(1/n) \to 0$ uniformly in s,t. Then

$$\mathbb{X}^n_{s,t} = \frac{1}{2} X^n_{s,t} \otimes X^n_{s,t} + A^n_{s,t} = \mathrm{O}\big(1/n^2\big) + A^n_{s,t}$$

where $A^n_{s,t} \in \mathfrak{so}(2)$ is the antisymmetric part of $\mathbb{X}^n_{s,t}$. To avoid cumbersome notation, we identify

$$\begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix} \in \mathfrak{so}(2) \leftrightarrow a \in \mathbf{R}.$$

$A^n_{s,t}$ then represents the signed area between the curve $(X^n_r : s \le r \le t)$ and the straight chord from X^n_t to X^n_s. (This is a simple consequence of Stokes theorem: the exterior derivative of the 1-form $\frac{1}{2}(x\,dy - y\,dx)$ which vanishes along straight chords, is the volume form $dx \wedge dy$.) With $s < t$, $(X^n_r : s \le r \le t)$ makes $\lfloor n^2(t-s) \rfloor$ full spins around the origin, at radius $1/n$. Each full spin contributes area $\pi(1/n)^2$, while the final incomplete spin contributes some area less than $\pi(1/n)^2$. The total signed area, with multiplicity, is thus

$$A^n_{s,t} = \big(n^2(t-s) + \mathrm{O}(1)\big)\frac{\pi}{n^2} = \pi(t-s) + \frac{C_{s,t}}{n^2}\ ,$$

where $|C_{s,t}| \le \pi$ uniformly in s,t. It follows that

$$\mathbb{X}^n_{s,t} = \pi(t-s)\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + \mathrm{O}\big(1/n^2\big) \tag{2.14}$$

and the claimed uniform convergence follows.

b) The following two estimates for path increments of $n^{-1}\exp\left(2\pi i n^2 t\right) \equiv X^n_t$ hold true:

$$\left|X^n_{s,t}\right| \le \left|\dot X^n\right|_\infty |t-s| \le n|t-s| \;, \qquad \left|X^n_{s,t}\right| \le 2|X^n|_\infty = 2/n \;.$$

Since $a \wedge b \le \sqrt{ab}$, it immediately follows that

$$\left|X^n_{s,t}\right| \le \sqrt{2|t-s|} \;,$$

uniformly in n, s, t. In other words, $\sup_n \|X^n\|_{1/2} < \infty$. The argument for the uniform bounds on $\mathbb{X}_{s,t}$ is similar. On the one hand, we have the bound (2.14). On the other hand, we also have

$$\left|\mathbb{X}^n_{s,t}\right| = \left|\int\int_{s<u<v<t} \dot X^n_u \otimes \dot X^n_v \, du\, dv\right| \le \left|\dot X^n\right|^2_\infty \frac{|t-s|^2}{2} \le \frac{n^2}{2}|t-s|^2 \;.$$

The required uniform bound on $\|\mathbb{X}\|_1$ follows by using (2.14) for $n^2|t-s| > 1$ and the above bound for $n^2|t-s| \le 1$.

c) The interpolation argument is left to the reader.

Exercise 2.19 (Translation of rough paths). Fix $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ and $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha\big([0,T], \mathbf{R}^d\big)$. For sufficiently smooth $h : [0,T] \to \mathbf{R}^d$, the *translation* of $\mathbf{X}$ in direction h is given by

$$T_h(\mathbf{X}) \stackrel{\text{def}}{=} \left(X^h, \mathbb{X}^h\right) \;,$$

where $X^h := X + h$ and

$$\mathbb{X}^h_{s,t} := \mathbb{X}_{s,t} + \int_s^t h_{s,r} \otimes dX_r + \int_s^t X_{s,r} \otimes dh_r + \int_s^t h_{s,r} \otimes dh_r \;. \tag{2.15}$$

a) Assume h is Lipschitz. (In particular, the last three integrals above are well-defined Riemann–Stieltjes integrals.) Show that for fixed h, the translation operator $T_h : \mathbf{X} \mapsto T_h(\mathbf{X})$ is a continuous map from $\mathscr{C}^\alpha$ into itself.
b) The above (Lipschitz) assumption on h is equivalently expressed by saying that $h \in W^{1,\infty}$, where $W^{1,q}$ denotes the space of absolutely continuous paths h with derivative $\dot h \in L^q$. Weaken the assumption on h by only requiring $\dot h \in L^q$, for suitable $q = q(\alpha)$. Show that $q = 2$ ("Cameron–Martin paths of Brownian motion") works for all $\alpha \le 1/2$. As a matter of fact, the integrals appearing in (2.15) make sense for every $q \ge 1$, but the resulting translated "rough path" would not necessarily lie in $\mathscr{C}^\alpha$.

2.6 Comments

The notion of rough path is due to Lyons and was introduced in [Lyo98]. Rather than using Hölder-type norms, the original article introduced rough paths in the p-variation sense for any $p \in [1,\infty)$. For $p \ge 3$ (corresponding to $\alpha < \frac{1}{3}$), this requires additional $[p]^{th}$ order information. Various notes by Lyons preceding [Lyo98] already dealt with α-Hölder rough paths for $\alpha \in \left(\frac{1}{3}, \frac{1}{2}\right]$.

In the recent literature, elements in $\mathscr{C}_g^\alpha$ are actually called *weakly geometric (α-Hölder) rough paths*. In contrast, the space of *geometric rough paths* $\mathscr{C}_g^{0,\alpha}$ is, by definition, obtained via completion of smooth paths in ϱ_α. We do not insist on this terminology here and indeed, by Proposition 2.5 there is not much difference. In the early literature the two concepts were somewhat blurred, matters were clarified in [FV06a].

Chapter 3
Brownian motion as a rough path

Abstract In this chapter, we consider the most important example of a rough path, which is the one associated to Brownian motion. We discuss the difference, at the level of rough paths, between Itô and Stratonovich Brownian motion. We also provide a natural example of approximation to Brownian motion which converges to neither of them.

3.1 Kolmogorov criterion for rough paths

Consider random $X(\omega) : [0,T] \to V$ and $\mathbb{X}(\omega) : [0,T]^2 \to V \otimes V$, subject to (2.1). Equivalently, following Exercise 2.7, we can think of

$$\mathbf{X}(\omega) \equiv (X, \mathbb{X})(\omega) : [0,T] \to V \oplus (V \otimes V)$$

as a (random) path. The basic example, of course, is that of d-dimensional standard Brownian motion B enhanced with

$$\mathbb{B}_{s,t} \stackrel{\text{def}}{=} \int_s^t B_{s,r} \otimes dB_r \in \mathbf{R}^d \otimes \mathbf{R}^d \cong \mathbf{R}^{d \times d}. \tag{3.1}$$

The integration here is understood either in Itô- or Stratonovich sense (in the latter case, we would write $\circ dB$); sometimes we indicate this by writing $\mathbb{B}^{\text{Itô}}$ resp. $\mathbb{B}^{\text{Strat}}$. It should be noted that the antisymmetric part of $\mathbb{B}$, which is nothing but *Lévy's stochastic area* and takes values in $\mathfrak{so}(d)$, is not affected by the choice of stochastic integration. Condition (2.1) is seen to be valid with either choice, while condition (2.5) only holds in the Stratonovich case. We now address the question of α- resp. 2α-Hölder regularity of X resp. $\mathbb{X}$ by a suitable extension of the classical Kolmogorov criterion; the application to Brownian motion is then carried out in detail in the following subsection.

Recalling that $B \in \mathcal{C}^\alpha$, a.s. for any $\alpha < 1/2$, we now address the question of 2α-Hölder regularity for $\mathbb{B}$.

Using Brownian scaling and exponential integrability of $\mathbb{B}_{0,1}$, which is an immediate consequence of the integrability properties of the second Wiener chaos, the following result applies with $\beta = 1/2$ and all $q < \infty$. It gives the desired 2α-Hölder regularity for $\mathbb{B}$, a.s. for any $\alpha < 1/2$. As a consequence, $(B, \mathbb{B}) \in \mathscr{C}^\alpha$ almost surely, where we may take any $\alpha \in \left(\frac{1}{3}, \frac{1}{2}\right)$ and $\mathbf{B} \equiv (B, \mathbb{B})$ is known as *Brownian rough path* or *enhanced Brownian motion*. In the Stratonovich case, thanks to (2.5), we obtain a *geometric* rough path, i.e. $(B, \mathbb{B}^{\text{Strat}}) \in \mathscr{C}_g^\alpha$.

Theorem 3.1 (Kolmogorov criterion for rough paths). *Let $q \geq 2, \beta > 1/q$. Assume, for all s, t in $[0, T]$*

$$|X_{s,t}|_{L^q} \leq C|t-s|^\beta , \qquad |\mathbb{X}_{s,t}|_{L^{q/2}} \leq C|t-s|^{2\beta} , \tag{3.2}$$

for some constant $C < \infty$. Then, for all $\alpha \in [0, \beta - 1/q)$, there exists a modification of $(X, \mathbb{X})$ (also denoted by $(X, \mathbb{X})$) and random variables $K_\alpha \in L^q, \mathbb{K}_\alpha \in L^{q/2}$ such that, for all s, t in $[0, T]$

$$|X_{s,t}| \leq K_\alpha(\omega)|t-s|^\alpha , \qquad |\mathbb{X}_{s,t}| \leq \mathbb{K}_\alpha(\omega)|t-s|^{2\alpha} . \tag{3.3}$$

In particular, if $\beta - \frac{1}{q} > \frac{1}{3}$ then, for every $\alpha \in \left(\frac{1}{3}, \beta - \frac{1}{q}\right)$, we have $(X, \mathbb{X}) \in \mathscr{C}^\alpha$ a.s.

Proof. The proof is almost the same as the classical proof of Kolmogorov's continuity criterion, as exposed for example in [RY91]. Without loss of generality take $T = 1$ and let D_n denote the set of integer multiples of 2^{-n} in $[0, 1)$. As in the usual criterion, it suffices to consider $s, t \in \bigcup_n D_n$, with the values at the remaining times filled in using continuity. (This is why in general one ends up with a modification.) Note that the number of elements in D_n is given by $\#D_n = 1/|D_n| = 2^n$. Set

$$K_n = \sup_{t \in D_n} \left|X_{t,t+2^{-n}}\right| , \qquad \mathbb{K}_n = \sup_{t \in D_n} \left|\mathbb{X}_{t,t+2^{-n}}\right| .$$

It follows from (3.2) that

$$\begin{aligned}
\mathbf{E}\big(K_n^q\big) &\leq \mathbf{E} \sum_{t \in D_n} \left|X_{t,t+2^{-n}}\right|^q \leq \frac{1}{|D_n|} C^q |D_n|^{\beta q} = C^q |D_n|^{\beta q - 1}, \\
\mathbf{E}\big(\mathbb{K}_n^{q/2}\big) &\leq \mathbf{E} \sum_{t \in D_n} \left|\mathbb{X}_{t,t+2^{-n}}\right|^{q/2} \leq \frac{1}{|D_n|} C^{q/2} |D_n|^{2\beta q/2} = C^{q/2} |D_n|^{\beta q - 1}.
\end{aligned}$$

Fix $s < t$ in $\bigcup_n D_n$ and choose $m : |D_{m+1}| < t - s \leq |D_m|$. The interval $[s, t)$ can be expressed as the finite disjoint union of intervals of the form $[u, v) \in D_n$ with $n \geq m + 1$ and where no three intervals have the same length. In other words, we have a partition of $[s, t)$ of the form

$$s = \tau_0 < \tau_1 < \cdots < \tau_N = t ,$$

where $(\tau_i, \tau_{i+1}) \in D_n$ some $n \geq m+1$, and for each fixed $n \geq m+1$ there are at most two such intervals taken from D_n. It follows that

$$|X_{s,t}| \leq \max_{0 \leq i < N} \left|X_{s,\tau_{i+1}}\right| \leq \sum_{i=0}^{N-1} \left|X_{\tau_i,\tau_{i+1}}\right| \leq 2 \sum_{n \geq m+1} K_n \ ,$$

and similarly,

$$\begin{aligned}
|\mathbb{X}_{s,t}| &= \left|\sum_{i=0}^{N-1} \mathbb{X}_{\tau_i,\tau_{i+1}} + X_{s,\tau_i} \otimes X_{\tau_i,\tau_{i+1}}\right| \leq \sum_{i=0}^{N-1} \left(\left|\mathbb{X}_{\tau_i,\tau_{i+1}}\right| + |X_{s,\tau_i}|\left|X_{\tau_i,\tau_{i+1}}\right|\right) \\
&\leq \sum_{i=0}^{N-1} \left|\mathbb{X}_{\tau_i,\tau_{i+1}}\right| + \max_{0 \leq i < N} \left|X_{s,\tau_{i+1}}\right| \sum_{j=0}^{N-1} \left|X_{\tau_j,\tau_{j+1}}\right| \\
&\leq 2 \sum_{n \geq m+1} \mathbb{K}_n + \Big(2 \sum_{n \geq m+1} K_n\Big)^2 .
\end{aligned}$$

We thus obtain

$$\frac{|X_{s,t}|}{|t-s|^\alpha} \leq \sum_{n \geq m+1} \frac{1}{|D_{m+1}|^\alpha} 2K_n \leq \sum_{n \geq m+1} \frac{2K_n}{|D_n|^\alpha} \leq K_\alpha \ ,$$

where $K_\alpha := 2\sum_{n\geq 0} K_n/|D_n|^\alpha$ is in L^q. Indeed, since $\alpha < \beta - 1/q$ by assumption and $|D_n|$ to any positive power is summable, we have

$$\|K_\alpha\|_{L^q} \leq \sum_{n \geq 0} \frac{2}{|D_n|^\alpha} |\mathbf{E}(K_n^q)|^{1/q} \leq \sum_{n \geq 0} \frac{2C}{|D_n|^\alpha} |D_n|^{\beta - 1/q} < \infty \ .$$

Similarly,

$$\frac{|\mathbb{X}_{s,t}|}{|t-s|^{2\alpha}} \leq \sum_{n \geq m+1} \frac{1}{|D_{m+1}|^{2\alpha}} 2\mathbb{K}_n + \Big(\sum_{n \geq m+1} \frac{1}{|D_{m+1}|^\alpha} 2K_n\Big)^2 \leq \mathbb{K}_\alpha + K_\alpha^2 \ ,$$

where $\mathbb{K}_\alpha := 2\sum_{n\geq 0} \mathbb{K}_n/|D_n|^{2\alpha}$ is in $L^{q/2}$. Indeed,

$$\|\mathbb{K}_\alpha\|_{L^{q/2}} \leq \sum_{n \geq 0} \frac{2}{|D_n|^{2\alpha}} \left|\mathbf{E}\Big(\mathbb{K}_n^{q/2}\Big)\right|^{2/q} \leq \sum_{n \geq 0} \frac{2C}{|D_n|^{2\alpha}} |D_n|^{2\beta - 2/q} < \infty \ ,$$

thus concluding the proof. $\square$

The reader will notice that the classical Kolmogorov criterion (KC) is contained in the above proof and theorem by simply ignoring all considerations related to the second-order process $\mathbb{X}$. Let us also note in this context that the classical KC works for processes $(\mathbf{X}_t : 0 \leq t \leq 1)$ with values in an arbitrary (separable) metric space (it suffices to replace $|X_{s,t}|$ by $d(\mathbf{X}_s, \mathbf{X}_t)$ in the argument). This observation actually gives an alternative and immediate proof of Theorem 3.1, at least for "geometric" $(X, \mathbb{X})$, i.e. in presence of the algebraic constraint (2.5), and at the price of some Lie group language. The key observation, as discussed in Section 2.3, is that $t \mapsto \mathbf{X}_t := (1, X_{0,t}, \mathbb{X}_{0,t})$ takes values in the step-2 nilpotent group with d generators, $\big(G^{(2)}\big(\mathbf{R}^d\big), \otimes\big)$, endowed with the Carnot-Carathéodory metric

$$d_{\mathrm{C}}(\mathbf{X}_s, \mathbf{X}_t) \asymp |X_{s,t}| + |\mathbb{X}_{s,t}|^{1/2}.$$

The assumptions of Theorem 3.1 then translate precisely to $|d_{\mathrm{C}}(\mathbf{X}_s, \mathbf{X}_t)|_{L^q} \leq C|t-s|^{\beta}$, and the same conclusion, $d_{\mathrm{C}}(\mathbf{X}_s, \mathbf{X}_t) \leq K_\alpha(\omega)|t-s|^{\alpha}$, for $K_\alpha \in L^q$, is obtained from the classical Kolmogorov criterion.

Remark 3.2 (Warning). It is not possible to obtain (3.3) by applying the classical KC to the $(V \otimes V)$-valued process $(\mathbb{X}_{0,t} : 0 \leq t \leq T)$. Doing so only gives $|\mathbb{X}_{s,t}| = \mathrm{O}(|t-s|^{\alpha})$ a.s. since one misses a crucial cancellation inherent in (cf. (2.1))

$$\mathbb{X}_{s,t} = \mathbb{X}_{0,t} - \mathbb{X}_{0,s} - X_{0,s} \otimes X_{s,t}.$$

That said, it is possible (but tedious) to use a 2-parameter version of the KC to see that $(s,t) \mapsto \mathbb{X}_{s,t}/|t-s|^{2\alpha}$ admits a continuous modification. In particular, this then implies that $\|\mathbb{X}\|_{2\alpha}$ is finite almost surely. In the Brownian setting, this was carried out in [Fri05].

Here is a similar result for rough path distances, say between $\mathbf{X}$ and $\tilde{\mathbf{X}}$. Note that, due to the nonlinear structure of rough path spaces, one cannot simply apply Theorem 3.1 to the "difference" of two rough paths. Indeed, $\tilde{\mathbf{X}} - \mathbf{X}$ is not defined in general for, formally, one misses the information about the mixed integrals in

$$\tilde{\mathbf{X}} - \mathbf{X} = \left(\tilde{X} - X, \int \big(\tilde{X} - X\big) \otimes d\big(\tilde{X} - X\big)\right).$$

Even when all of these expressions are well-defined, say when $\tilde{X}$ is smooth, convergence of the right-hand side above to zero is *different* from saying that

$$\tilde{X} \to X, \qquad \int \tilde{X} \otimes d\tilde{X} \to \int X \otimes dX$$

and it is this type of convergence (in suitable Hölder-type norms) which our rough path metric ϱ_α expresses.

Theorem 3.3 (Kolmogorov criterion for rough path distance). *Let α, β, q be as above in Kolmogorov's criterion (KC), Theorem 3.1. Assume that both $\tilde{\mathbf{X}} = (\tilde{X}, \tilde{\mathbb{X}})$ and $\mathbf{X} = (X, \mathbb{X})$ satisfy the moment condition in the statement of KC with some constant C. Set*

$$\Delta X := \tilde{X} - X\,, \qquad \Delta \mathbb{X} := \tilde{\mathbb{X}} - \mathbb{X}\,,$$

and assume that for some $\varepsilon > 0$ and all $s, t \in [0, T]$

$$|\Delta X_{s,t}|_{L^q} \le C\varepsilon |t-s|^{\beta}\,, \qquad |\Delta \mathbb{X}_{s,t}|_{L^{q/2}} \le C\varepsilon |t-s|^{2\beta}\,.$$

Then there exists M, depending increasingly on C, so that $|\|\Delta X\|_{\alpha}|_{L^q} \le M\varepsilon$ and $|\|\Delta \mathbb{X}\|_{2\alpha}|_{L^{q/2}} \le M\varepsilon$. In particular, if $\beta - \frac{1}{q} > \frac{1}{3}$ then, for every $\alpha \in \left(\frac{1}{3}, \beta - \frac{1}{q}\right)$ we have $\|\tilde{\mathbf{X}}\|_{\alpha}, \|\mathbf{X}\|_{\alpha} \in L^q$ and

$$|\varrho_{\alpha}(\tilde{\mathbf{X}}, \mathbf{X})|_{L^q} \le M\varepsilon.$$

Proof. The proof is a straightforward modification of the proof of Theorem 3.1 and is left as an exercise to the reader. □

In applications of this theorem one typically has a a family

$$\{\mathbf{X}^n \equiv (X^n, \mathbb{X}^n) : 1 \le n \le \infty\}\,,$$

such that the moment conditions in the statement of KC hold with with a constant C, uniformly over $1 \le n \le \infty$. Application of the above with $\varepsilon = \varepsilon_n$ then gives L^q-rates of convergence,

$$|\varrho_{\alpha}(\mathbf{X}^n, \mathbf{X})|_{L^q} \lesssim \varepsilon_n.$$

Of course, when ε_n decays sufficiently fast, a Borel–Cantelli argument also gives almost-sure convergence with suitable rates.

3.2 Itô Brownian motion

Consider a d-dimensional standard Brownian motion B enhanced with its iterated integrals

$$\mathbb{B}_{s,t} \stackrel{\text{def}}{=} \int_s^t B_{s,r} \otimes dB_r \in \mathbf{R}^d \otimes \mathbf{R}^d \cong \mathbf{R}^{d\times d}\,, \tag{3.4}$$

where the stochastic integration is understood in the sense of Itô. The antisymmetric part of $\mathbb{B}$ is known as "*Lévy's stochastic area*". Sometimes we indicate this by writing $\mathbb{B}^{\text{Itô}}$. We shall assume straight away that B_t and $\mathbb{B}_{s,t}$ are continuous in t and s,t respectively, with probability one. For instance, if one takes as granted that almost surely Brownian motion and indefinite Itô integrals against Brownian motion (such as $\mathbb{B}_{0,\cdot}$) are continuous, then it suffices to (re)define the second order increments as $\mathbb{B}_{s,t} = \mathbb{B}_{0,t} - \mathbb{B}_{0,s} - B_s \otimes B_{s,t}$. Of course, by additivity of the Itô integral, this coincides a.s. with the earlier definition. En passant, the so-defined $\mathbf{B}_{s,t} = (B_{s,t}, \mathbb{B}_{s,t})$ immediately satisfies (2.1), for all times, on a common set of probability one.

Proposition 3.4. *For any $a \in (1/3, 1/2)$ and $T > 0$ with probability one,*

$$\mathbf{B} = (B, \mathbb{B}^{\text{Itô}}) \in \mathscr{C}^{\alpha}([0,T], \mathbf{R}^d)\,.$$

Proof. Using Brownian scaling and finite moments of $\mathbb{B}_{0,1}$, which are immediate from integrability properties of the (homogeneous) second Wiener–Itô chaos, the KC for rough paths applies with $\beta = 1/2$ and all $q < \infty$. (As exercise, the reader may want to show finite moments of $\mathbb{B}_{0,1}$ without chaos arguments; an elementary way to do so is via conditioning, Itô isometry, and reflection principle.) □

Observe that Brownian motion enhanced with its iterated Itô integrals (2nd order calculus!) yields a (random) rough path but not a *geometric* rough path which is, by definition, an object with hardwired first order behaviour. Indeed, Itô formula yields the identity

$$d(B^i B^j) = B^i dB^j + B^j dB^i + \langle B^i, B^j \rangle \, dt \;, \qquad i,j = 1,\dots,d \;,$$

so that, writing I for the identity matrix in d dimensions, we have for $s < t$,

$$\mathrm{Sym}\left(\mathbb{B}^{\text{Itô}}_{s,t}\right) = \frac{1}{2} B_{s,t} \otimes B_{s,t} - \frac{1}{2} I(t-s) \neq \frac{1}{2} B_{s,t} \otimes B_{s,t} \;,$$

in contradiction with (2.5).

Let us also mention that Brownian motion with values in infinite-dimensional spaces can also be lifted to rough paths, see the exercise section.

3.3 Stratonovich Brownian motion

In the previous section we defined $\mathbb{B}^{\text{Itô}}$ by *Itô-integration* of d-dimensional Brownian B motion against itself. Now, for (scalar) continuous semimartingales, M, N say, the *Stratonovich integral* is defined as

$$\int_0^t M \circ dN := \int_0^t M dN + \frac{1}{2}[M,N]_t$$

and has the advantage of a first order calculus. For instance, one has the *first order product rule*

$$d(MN) = M \circ dN + N \circ dM.$$

One can then define $\mathbb{B}^{\text{Strat}}$ by (component-wise) *Stratonovich-integration* of Brownian motion against itself. Using basic results on quadratic variation between Brownian motions ($d\big[B^i, B^j\big]_t = \delta^{i,j} dt$ where $\delta^{i,j} = 1$ if $i = j$, zero else), we see that

$$\mathbb{B}^{\text{Strat}}_{s,t} = \mathbb{B}^{\text{Itô}}_{s,t} + \frac{1}{2} I(t-s) \tag{3.5}$$

where I stands for the identity matrix. Note that the difference between $\mathbb{B}^{\text{Strat}}$ and $\mathbb{B}^{\text{Itô}}$ is symmetric, so that the antisymmetric parts of the two processes (Lévy's stochastic area) are identical.

Proposition 3.5. *For any $a \in (1/2, 1/3)$ and with probability one,*

$$\mathbf{B} = (B, \mathbb{B}^{\text{Strat}}) \in \mathscr{C}_g^{\alpha}([0,T], \mathbf{R}^d).$$

A typical realisation $\mathbf{B}(\omega)$ is called Brownian rough path, $\mathbf{B} = \mathbf{B}^{\text{Strat}}$ as a process is called (Stratonovich) enhanced Brownian motion.

Proof. Using (3.5) rough path regularity of $\mathbf{B}$ is immediately reduced to the already establish Itô-case. (Alternatively, one can use again the Kolmogorov criterion for rough paths; the only – insignificant – difference is that now $\mathbb{B}_{0,1}^{\text{Strat}}$ takes values in the inhomogeneous second chaos, due to the deterministic part $I/2$.) At last, $\mathbf{B}(\omega)$ is *geometric* since

$$\text{Sym}\left(\mathbb{B}_{s,t}^{\text{Strat}}\right) = \frac{1}{2} B_{s,t} \otimes B_{s,t},$$

an immediate consequence of the first order product rule. □

It is a *deterministic* feature of every geometric rough path $(X, \mathbb{X})$ that it can be approximated – in the precise sense of Proposition 2.5 – by smooth paths in the rough path topology. Such approximations require knowledge not only of the underlying path X, but of the entire *rough path*, including the second order information $\mathbb{X}$.

In contrast, one has the *probabilistic* statement that piecewise linear (and in fact: many other "obvious") approximations still converge in rough path sense. More specifically, in the present context of d-dimensional standard Brownian motion, we now give an elegant proof of this based on (discrete-time!) martingale arguments.

Proposition 3.6. *Consider dyadic piecewise-linear approximations $\left(B^{(n)}\right)$ to B on $[0,T]$. That is, $B_t^{(n)} = B_t$ whenever $t = iT/2^n$ for some integer i, and linearly interplolated on intervals $[iT/2^n, (i+1)T/2^n]$. Then, with probability one,*

$$\left(B^{(n)}, \int_0^{\cdot} B^{(n)} \otimes dB^{(n)}\right) \to (B, \mathbb{B}^{\text{Strat}}) \quad \textit{in} \quad \mathscr{C}_g^{\alpha}\ .$$

(The integral on the left-hand side is understood as classical Riemann–Stieltjes integral.)

Remark 3.7. With Theorem 3.3, one can see rough path convergence (in probability, and actually L^q, any $q < \infty$) of piecewise linear approximation along any sequence of dissections with mesh tending to zero. Moreover, this approach will give the rate θ, any $\theta < 1/2 - \alpha$.

Proof. It is easy to check that B gives $B^{(n)}$ via conditioning on B at dyadic times,

$$B^{(n)} = \mathbf{E}(B | \sigma\{B_{kT2^{-n}} : 0 \le k \le 2^n\}).$$

By independence of the components B^i, B^j for $i \neq j$, the same holds for $\mathbb{B}^{\text{Strat}}$ *off-diagonal*; the on-diagonal terms require no further attention since $\mathbb{B}_{s,t}^{\text{Strat};i,i} = \frac{1}{2}\left(B_{s,t}^i\right)^2$. Almost sure pointwise convergence then readily follows from martingale convergence. Furthermore, Theorem 3.1 implies

$$\left|B^i_{s,t}\right| \le K_\alpha(\omega)|t-s|^\alpha , \qquad \left|\mathbb{B}^{\mathrm{Strat};i,j}_{s,t}\right| \le \mathbb{K}_\alpha(\omega)|t-s|^{2\alpha} ,$$

and upon conditioning with respect to $\sigma\{B_{kT2^{-n}} : 0 \le k \le 2^n\}$, the same bounds hold for $B^{(n);i}$ and for $\int_0^\cdot B^{(n);i} dB^{(n);j}$. In fact, $K_\alpha, \mathbb{K}_\alpha$ have (more than enough) integrability to apply Doob's maximal inequality. This leads, with probability one, to the bound

$$\sup_n \left\| B^{(n)}, \int_0^\cdot B^{(n)} \otimes dB^{(n)} \right\|_{2\alpha} < \infty .$$

Together with a.s. pointwise convergence, a (deterministic) interpolation argument shows a.s. convergence with respect to the α-Hölder rough path metric ϱ_α. □

The reader should be warned that there are perfectly smooth and uniform approximations to Brownian motion, which do not converge to Stratonovich enhanced Brownian motion, but instead to some different geometric (random) rough path, such as

$$\bar{\mathbf{B}} = \left(B, \bar{\mathbb{B}}\right) , \quad \text{where} \quad \bar{\mathbb{B}}_{s,t} = \mathbb{B}^{\mathrm{Strat}}_{s,t} + (t-s)A , \qquad A \in \mathfrak{so}(d) .$$

Note that the difference between $\bar{\mathbb{B}}$ and $\mathbb{B}^{\mathrm{Strat}}$ is now *antisymmetric*, i.e. $\bar{\mathbf{B}}$ has a stochastic area that is different from Lévy's area. To construct such approximations, it suffices to include oscillations (at small scales) such as to create the desired effect in the area, while they do no affect the limiting path, see Exercise 2.17. (In the context of Brownian motion and SDEs driven by Brownian motion such approximations were studied by McShean, Ikeda–Watanabe and others, see [McS72, IW89].) Although such "twisted" approximations do not seem to be the most obvious way to approximate Brownian motion, they also arise naturally in some perfectly reasonable situations.

3.4 Brownian motion in a magnetic field

Newton's second law for a particle in $\mathbf{R}^3$ with mass m, and position $x = x(t)$, (for simplicity: constant) frictions $\alpha_1, \alpha_2, \alpha_3 > 0$ in orthonormal directions, subject to a (3-dimensional) white noise in time, i.e. the distributional derivative of a 3-dimensional Brownian motion B, reads

$$m\ddot{x} = -M\dot{x} + \dot{B}, \tag{3.6}$$

assuming M symmetric with spectrum $\alpha_1, \alpha_2, \alpha_3$. The process $x(t)$ describes what is known as *physical Brownian motion*. It is well known that in small mass regime, $m \ll 1$, of obvious physical relevance when dealing with particles, a good approximation is given by (mathematical) *Brownian motion* (with non-standard covariance). To see this formally, it suffices to take $m = 0$ in (3.6) in which case $x = M^{-1}B$.

Let us now assume that our particle (with position x and momentum $m\dot{x}$) carries a non-zero electric charge and moves in a magnetic field which we assume to be constant. Recall that such a particle experiences a sideways force ("Lorentz force") that is proportional to the strength of the magnetic field, the component of the velocity that is perpendicular to the magnetic field and the charge of the particle. In terms of our assumptions, this simply means that a non-zero antisymmetric component is added to M. We shall hence drop the assumption of symmetry, and instead consider for M a general square matrix with

$$\mathrm{Real}\{\sigma(M)\} \subset (0,\infty).$$

Note that these second order dynamics can be rewritten as evolution equation for the *momentum* $p(t) = m\dot{x}(t)$,

$$\dot{p} = -M\dot{x} + \dot{B} = -\frac{1}{m}M\dot{p} + \dot{B}.$$

As we shall see $X = X^m$, indexed by "mass" m, converges in a quite non-trivial way to Brownian motion on the level of rough paths. In fact, the correct limit in rough path sense is $\bar{\mathbf{B}} = (B, \bar{\mathbb{B}})$, where

$$\bar{\mathbb{B}}_{s,t} = \mathbb{B}^{\mathrm{Strat}}_{s,t} + (t-s)A, \tag{3.7}$$

in terms of an *antisymmetric* matrix A; written explicitly as $A = \frac{1}{2}(M\Sigma - \Sigma M^*) \in \mathfrak{so}(d)$, where

$$\Sigma = \int_0^\infty e^{-Ms}e^{-M^*s}ds.$$

When M is normal, i.e. $M^*M = MM^*$, it is an exercise in linear algebra to show that this expression simplifies to

$$A = \frac{1}{2}\,\mathrm{Anti}(M)\,\mathrm{Sym}(M)^{-1}\ ,$$

where $\mathrm{Anti}(M)$ denotes the antisymmetric part of a matrix and $\mathrm{Sym}(M)$ its symmetric part. We can now state the result in full detail.

Theorem 3.8. *Let $M \in \mathbf{R}^{d\times d}$ be a square matrix in dimension d such that all its eigenvalues have strictly positive real part. Let B be a d-dimensional standard Brownian motion, $m > 0$, and consider the stochastic differential equations*

$$dX = \frac{1}{m}P\,dt\ , \qquad dP = -\frac{1}{m}MP\,dt + dB\ .$$

with zero initial position X and momentum P. Then, for any $q \geq 1$ and $\alpha \in (1/3, 1/2)$, as mass $m \to 0$,

$$\left(MX, \int MX \otimes d(MX)\right) \to \bar{\mathbf{B}} \quad \text{in } \mathscr{C}^\alpha \text{ and } L^q.$$

Proof. Step 1. (Pointwise convergence in L^q.) In order to exploit Brownian scaling, it is convenient to set $m = \varepsilon^2$ and then Y^ε as rescaled momentum,

$$Y^\varepsilon = P/\varepsilon.$$

We shall also write $X^\varepsilon = X$, to emphasize dependence on ε. We then have

$$dY^\varepsilon = -\varepsilon^{-2} M Y^\varepsilon dt + \varepsilon^{-1} dB\;, \qquad dX^\varepsilon = \varepsilon^{-1} Y^\varepsilon dt\;.$$

By assumption, there exists $\lambda > 0$ such that the real part of every eigenvalue of M is (strictly) bigger than λ. For later reference, we note that this implies the estimate $|\exp(-\tau M)| = \mathrm{O}(\exp(-\lambda\tau))$ as $\tau \to \infty$. For fixed ε, define the Brownian motion $\tilde{B}_t = \varepsilon B_{\varepsilon^{-2}t}$, and consider the SDEs

$$d\tilde{Y} = -M\tilde{Y}\,dt + d\tilde{B}\;, \qquad d\tilde{X} = \tilde{Y}\,dt\;.$$

Note that the law of the solutions does not depend on ε. Furthermore, when solved with identical initial data, we have pathwise equality

$$\big(Y^\varepsilon_t, \varepsilon^{-1} X^\varepsilon_t\big) = \big(\tilde{Y}_{\varepsilon^{-2}t}, \tilde{X}_{\varepsilon^{-2}t}\big)\;. \tag{3.8}$$

Thanks to our assumption on M, $\tilde{Y}$ is ergodic; the stationary solution has (zero mean, Gaussian) law $\nu \sim \mathcal{N}(0,\Sigma)$ for some covariance matrix Σ. To compute it, write down the stationary solution

$$\tilde{Y}^{\text{stat}}_t = \int_{-\infty}^t e^{-M(t-s)} dB_s\;.$$

For each t (and in particular for $t = 0$), the law of $\tilde{Y}^{\text{stat}}_t$ is precisely ν. We then see that

$$\Sigma = \mathbf{E}\Big(\tilde{Y}^{\text{stat}}_0 \otimes \tilde{Y}^{\text{stat}}_0\Big) = \int_{-\infty}^0 e^{-M(-s)} e^{-M^*(-s)} ds = \int_0^\infty e^{-Ms} e^{-M^* s} ds.$$

Since $\sup_{0\le t<\infty} \mathbf{E}|\tilde{Y}^2_t| < \infty$, it is clear that $\varepsilon\tilde{Y}_{\varepsilon^{-2}t} = \varepsilon Y^\varepsilon_t \to 0$ in L^2 uniformly in t (and hence in L^q for any $q < \infty$). Noting that $MX^\varepsilon_t = B_t - \varepsilon Y^\varepsilon_{0,t}$, the first part of the proposition is now obvious. Moreover, by the ergodic theorem[1],

$$\int_0^t f(Y^\varepsilon_t)\,dt \to t \int f(y)\nu(dy)\;, \qquad \text{in } L^q \text{ for any } q < \infty, \tag{3.9}$$

for all reasonable test functions f; we shall only use it for quadratics. Using $dX^\varepsilon = \varepsilon^{-1} Y^\varepsilon dt$ we can then write

$$\int_0^t MX^\varepsilon_s \otimes d(MX^\varepsilon)_s = \int_0^t MX^\varepsilon_s \otimes dB_s - \varepsilon \int_0^t MX^\varepsilon_s \otimes dY^\varepsilon_s$$

[1] As found e.g. in textbooks by Stroock [Str11] or Kallenberg [Kal02]. Test functions are usually assumed to be bounded, but by a truncation argument in our setting, this is easily extended to quadratics.

$$
\begin{aligned}
&= \int_0^t MX_s^\varepsilon \otimes dB_s - MX_t^\varepsilon \otimes (\varepsilon Y_t^\varepsilon) + \varepsilon \int_0^t d(MX^\varepsilon)_s \otimes Y_s^\varepsilon \\
&= \int_0^t MX_s^\varepsilon \otimes dB_s - MX_t^\varepsilon \otimes (\varepsilon Y_t^\varepsilon) + \int_0^t MY_s^\varepsilon \otimes Y_s^\varepsilon ds \\
&\to \int_0^t B_s \otimes dB_s - 0 + t \int (My \otimes y)\, \nu(dy) \\
&= \int_0^t B_s \otimes dB_s + tM\Sigma = \mathbb{B}_{0,t} + t\Big(M\Sigma - \frac{1}{2} I\Big) \;,
\end{aligned}
$$

where the convergence is in L^q for any $q \geq 2$. By considering the symmetric part of the above equation,

$$
\frac{1}{2}(MX_t^\varepsilon) \otimes (MX_t^\varepsilon) \to \frac{1}{2} B_t \otimes B_t + \mathrm{Sym}\left(M\Sigma - \frac{1}{2} I\right),
$$

we see that $M\Sigma - \frac{1}{2}I$ has symmetric part 0, i.e. is antisymmetric, and hence also equals $\frac{1}{2}(M\Sigma - \Sigma M^*)$. This settles pointwise convergence, in the sense that

$$
S(MX^\varepsilon)_t := \left(MX_t^\varepsilon, \int_0^t MX_s^\varepsilon \otimes d(MX^\varepsilon)_s \right) \to \big(B_t, \bar{\mathbb{B}}_{0,t}\big).
$$

Step 2. (Uniform rough path bounds in L^q.) We claim that, for any $q < \infty$,

$$
\sup_{\varepsilon \in (0,1]} \mathbf{E}[\|MX^\varepsilon\|_\alpha^q] < \infty \;, \qquad \sup_{\varepsilon \in (0,1]} \mathbf{E}\left[\left\| \int MX^\varepsilon \otimes d(MX^\varepsilon) \right\|_{2\alpha}^q \right] < \infty \;,
$$

which, in view of Theorem 3.1, is an immediate consequence of the bounds

$$
\sup_{\varepsilon \in (0,1]} \mathbf{E}\big[|X_{s,t}^\varepsilon|^q \big] \lesssim |t-s|^{\frac{q}{2}} \;, \qquad \sup_{\varepsilon \in (0,1]} \mathbf{E}\left[\left| \int_s^t X_{s,\cdot}^\varepsilon \otimes dX^\varepsilon \right|^q \right] \lesssim |t-s|^q \;.
$$

Since X is Gaussian, it follows from integrability properties of the first two Wiener–Itô chaoses that it is enough to show these bounds for $q = 2$. Furthermore, we note that the desired estimates are a consequence of the bounds

$$
\mathbf{E}\Big[\big|\tilde{X}_{s,t}\big|^2 \Big] \lesssim |t-s| \;, \tag{3.10}
$$

$$
\mathbf{E}\left[\left| \int_s^t \tilde{X}_{s,u}^\varepsilon \otimes d\tilde{X}_u \right|^2 \right] \lesssim |t-s|^2 \;, \tag{3.11}
$$

where the implied proportionality constants are uniform over $t, s \in (0, \infty)$. Indeed, this follows directly from writing

$$
\mathbf{E}\Big[\big|X_{s,t}^\varepsilon\big|^2 \Big] = \mathbf{E}\Big[\big|\varepsilon \tilde{X}_{\varepsilon^{-2}s, \varepsilon^{-2}t}\big|^2 \Big] \lesssim \varepsilon^2 \big|\varepsilon^{-2} t - \varepsilon^{-2} s\big| = |t-s| \;,
$$

(note the uniformity in ε), and similarly for the second moment of the iterated integral.

In order to check (3.10), it is enough to note that $M\tilde{X}_{s,t} = \tilde{B}_{s,t} - \tilde{Y}_{s,t}$, combined with the estimate

$$\mathbf{E}\Big[|\tilde{Y}_{s,t}|^2\Big] = \mathbf{E}\Big[\Big|(e^{-M(t-s)} - I)\tilde{Y}_s\Big|^2\Big] + \int_s^t \mathrm{Tr}(e^{-Mu}e^{-M^*u})\,du \lesssim |t-s|\;,$$

where we used the fact that Real$\{\sigma(M)\} \subset (0,\infty)$ to get a uniform bound. In order to control (3.11), we consider one of the components and write

$$\begin{aligned}
\mathbf{E}\Big[\Big|\int_s^t \tilde{X}^i_{s,u} d\tilde{X}^j_u\Big|^2\Big] &= \mathbf{E}\Big[\Big|\int_s^t \int_s^u \tilde{Y}^i_r \tilde{Y}^j_u dr\,du\Big|^2\Big] \\
&= \int_{[s,t]^4} \mathbf{E}\Big[\tilde{Y}^i_r\tilde{Y}^j_u\tilde{Y}^i_q\tilde{Y}^j_v\Big]\mathbf{1}_{\{r\le u; q\le v\}} dr\,du\,dq\,dv \\
&\le \int_{[s,t]^4} \Big(\Big|\mathbf{E}\Big[\tilde{Y}^i_r\tilde{Y}^j_u\Big]\Big|\Big|\mathbf{E}\Big[\tilde{Y}^i_q\tilde{Y}^j_v\Big]\Big| + \Big|\mathbf{E}\Big[\tilde{Y}^i_r\tilde{Y}^i_q\Big]\Big|\Big|\mathbf{E}\Big[\tilde{Y}^j_u\tilde{Y}^j_v\Big]\Big| \\
&\qquad + \Big|\mathbf{E}\Big[\tilde{Y}^i_r\tilde{Y}^j_v\Big]\Big|\Big|\mathbf{E}\Big[\tilde{Y}^j_u\tilde{Y}^i_q\Big]\Big|\Big) dr\,du\,dq\,dv \\
&\lesssim \Big(\int_{[s,t]^2} \Big|\mathbf{E}\Big[\tilde{Y}_r \otimes \tilde{Y}_u\Big]\Big| dr\,du\Big)^2 \\
&\lesssim \Big(\int_{[s,t]^2} \Big|\mathbf{E}\Big[\tilde{Y}_r \otimes \tilde{Y}_u\Big]\Big|\mathbf{1}_{\{r\le u\}} dr\,du\Big)^2\;,
\end{aligned}$$

where we have used the fact that $\tilde{Y}$ is Gaussian (which yields Wick's formula for the expectation of products) in order to get the bound on the third line. But for $r \le u$, $\mathbf{E}\big[\tilde{Y}_u\big|\tilde{Y}_r\big] = e^{-M(u-r)}\tilde{Y}_r$, so that

$$\begin{aligned}
\int_{[s,t]^2} \Big|\mathbf{E}\Big[\tilde{Y}_r \otimes \tilde{Y}_u\Big]\Big|\mathbf{1}_{\{r\le u\}}\,dr\,du &= \int_{[s,t]^2} \Big|\mathbf{E}\Big[\tilde{Y}_r \otimes e^{-M(u-r)}\tilde{Y}_r\Big]\Big|\mathbf{1}_{\{r\le u\}} dr\,du \\
&\lesssim \int_s^t \Big(\int_r^t e^{-\lambda(u-r)} du\Big)\mathbf{E}\big[\big|\tilde{Y}_r\big|^2\big]\,dr \lesssim |t-s|\;.
\end{aligned}$$

It now suffices to recall that $|\exp(-\tau M)| = \mathrm{O}(\exp(-\lambda\tau))$ to conclude the proof of (3.11).

Step 3. (Rough path convergence in L^q.) The remainder of the proof is an easy application of interpolation, along the lines of Exercise 2.9. $\square$

3.5 Cubature on Wiener Space

Quadrature rules replace Lebesgue measure λ on $[0,1]$ by a finite, convex linear combination of point masses, say $\mu = \sum a_i \delta_{x_i}$, where weights (a_i) and points (x_i) are chosen such that all monomials (and hence all polynomials) up to degree N are correctly evaluated. In other words, one first computes the moments of λ, namely

$$\int_0^1 x^n d\lambda(x) = \frac{1}{n+1} \ ,$$

for all $n \geq 0$. One then looks for a measure μ such that $\int_0^1 x^n d\mu(x) = 1/(n+1)$ for all $n \in \{0, 1, \ldots, N\}$. The same can be done on Wiener space: the monomial x^n is then replaced by the n-fold iterated integrals (in the sense of Stratonovich), integration is on $\mathcal{C}\big([0,T], \mathbf{R}^d\big)$ against standard d-dimensional Wiener measure. In order to find such cubature formulae, the mandatory first step, on which we focus here, is the computation of the expectations of the n-fold iterated integrals[2]

$$\mathbf{E}\bigg(\int_{0<t_1<\ldots t_n<T} \circ dB \otimes \cdots \otimes \circ dB\bigg) \ .$$

Let us combine all of these integrals into one single object (also called the "signature" of Brownian motion) by writing

$$S(B)_{0,T} = 1 + \sum_{n \geq 1} \int_{0<t_1<\ldots t_n<T} \circ dB \otimes \cdots \otimes \circ dB \ .$$

The signature $S(B)_{0,T}$ naturally takes values in the tensor algebra $T\big(\big(\mathbf{R}^d\big)\big) = \bigoplus_{n\geq 0} \big(\mathbf{R}^d\big)^{\otimes n}$. It turns out that in the case of Brownian motion, the expected signature can be expressed in a particularly concise and elegant form.

Theorem 3.9 (Fawcett). *Consider $S(B)_{0,T}$ as above as a $T\big(\big(\mathbf{R}^d\big)\big)$-valued random variable. Then*

$$\mathbf{E}S(B)_{0,T} = \exp\left(\frac{T}{2}\sum_{i=1}^d e_i \otimes e_i\right).$$

Proof. (Shekhar) Set $\varphi_t := \mathbf{E}S(B)_{0,t}$. (It is not hard to see, by Wiener–Itô chaos integrability or otherwise, that all involved iterated integrals are integrable so that φ is well-defined.) By Chen's formula (in its general form, see Exercise 2.6) and the independence of Brownian increments, one has the identity

$$\varphi_{t+s} = \varphi_t \otimes \varphi_s \ .$$

Since $\varphi_t \otimes \varphi_s = \varphi_s \otimes \varphi_t$, we have $[\varphi_s, \varphi_t] = 0$, so that

[2] We remark that all n-fold iterated Stratonovich integrals can be obtained from the "level-2" rough path $(B(\omega), \mathbb{B}^{\text{Strat}}(\omega)) \in \mathscr{C}_g^\alpha$ by a continuous map. In fact, this so-called *Lyons lift*, allows to view any geometric rough path as a "level-n" rough path for arbitrary $n \geq 2$.

$$\log \varphi_{t+s} = \log \varphi_t + \log \varphi_s \ .$$

For integers m, n we have $\log \varphi_m = n \log \varphi_{m/n}$ and $\log \varphi_m = m \log \varphi_1$. It follows that

$$\log \varphi_t = t \log \varphi_1 \ ,$$

first for $t = \frac{m}{n} \in \mathbf{Q}$, then for any real t by continuity. On the other hand, for $t > 0$, Brownian scaling implies that $\varphi_t = \delta_{\sqrt{t}} \varphi_1$ where δ_λ is the dilatation operator, which acts by multiplication with λ^n on the n^{th} tensor level, $(\mathbf{R}^d)^{\otimes n}$. Since δ_λ commutes with $\otimes$ (and thus also with log, defined as power series),

$$\log \varphi_t = \delta_{\sqrt{t}} \log \varphi_1$$

and it follows that one necessarily has

$$\log \varphi_1 \in \left(\mathbf{R}^d\right)^{\otimes 2} \ .$$

It remains to identify $\log \varphi_1$ with $\frac{1}{2} \sum_{i=1}^d e_i \otimes e_i$. To this end it suffices to compute the expected signature up to level two, which yields

$$\mathbf{E} S^{(2)}(B) = \mathbf{E}\left(1 + B_{0,1} + \int_0^1 B \otimes \circ dB\right) = 1 + \frac{1}{2} \sum_{i=1}^d e_i \otimes e_i \ .$$

Recall that in this expression, “1” is identified with $(1, 0, 0)$ in the truncated tensor algebra, and similarly for the other summands, and addition also takes place in $T^{(2)}(\mathbf{R}^d)$. Taking the logarithm (in the tensor algebra truncated beyond level 2; in this case $\log(1 + a + b) = a + \left(b - \frac{1}{2} a \otimes a\right)$ if a is a 1-tensor, b a 2-tensor) then immediately gives the desired identification. □

The (constructive) existence of *cubature formulae*, a finite family of piecewise smooth paths with associated probabilities, such as to mimic the behaviour of the expected signature up to a given level is not a trivial problem (although much has been achieved to date), the reader can explore a simple case in Exercise 3.24 below.

3.6 Scaling limits of random walks

Consider a family of continuous processes $\mathbf{X}^n = (X^n, \mathbb{X}^n)$, with values in values in $V \oplus (V \otimes V)$ where $\dim V < \infty$. Assume $\mathbf{X}_0^n = (0, 0)$ for all n. We leave the proof of the following result as exercise.

Theorem 3.10 (Kolmogorov tightness criterion for rough paths). *Let $q \geq 2, \beta > 1/q$. Assume, for all s, t in $[0, T]$*

$$\mathbf{E}_n \left|X_{s,t}^n\right|^q \leq C|t-s|^{\beta q} \ , \qquad \mathbf{E}_n \left|\mathbb{X}_{s,t}^n\right|^{q/2} \leq C|t-s|^{\beta q} \ , \tag{3.12}$$

for some constant $C < \infty$. Assume $\beta - \frac{1}{q} > \frac{1}{3}$. Then for every $\alpha \in \left(\frac{1}{3}, \beta - \frac{1}{q}\right)$, the $\mathbf{X}^n$'s are tight in $\mathscr{C}^{0,\alpha}$.

In typical applications, the X^n are only defined for discrete times, such as $s = j/n, t = k/n$ for integers j, k. The non-trivial work then consists, for a suitable choice of $\mathbb{X}^n$, in checking the following discrete tightness estimates,

$$\mathbf{E}_n \left| X^n_{\frac{j}{n},\frac{k}{n}} \right|^q \le C \left| \frac{j-k}{n} \right|^{\beta q}, \qquad \mathbf{E}_n \left| \mathbb{X}^n_{\frac{j}{n},\frac{k}{n}} \right|^{q/2} \le C \left| \frac{j-k}{n} \right|^{\beta q}. \tag{3.13}$$

The analogous continuous tightness estimates are typically obtained by suitable extension of $\mathbf{X}^n$ to continuous times (e.g. piecewise geodesic).

Proposition 3.11. *Consider a d-dimensional random walk $(X_j : j \in \mathbf{N})$, with i.i.d. increments of zero mean, finite moments of any order $q < \infty$, and unit covariance matrix. Extend the rescaled random walk*

$$X^n_{\frac{j}{n}} := \frac{1}{\sqrt{n}} X_j ,$$

defined on discrete times only, by piecewise linear interpolation to all times and construct to $\mathbf{X}^n = (X^n, \mathbb{X}^n)$ by iterated (Riemann–Stieltjes) integration. Then the tightness estimates in Theorem 3.10 hold with $\beta = 1/2$ and all $q < \infty$.

Proof. The iterated integrals of a linear (or affine) path with increment $v \in \mathbf{R}^d$ takes the simple form $\exp(v)$ in terms of the tensor exponential introduced in (2.8). Chen's relation then implies

$$\mathbf{X}^n_{\frac{j}{n},\frac{k}{n}} = \exp(X^n_{\frac{j}{n},\frac{j+1}{n}}) \otimes \cdots \otimes \exp(X^n_{\frac{k-1}{n},\frac{k}{n}}). \tag{3.14}$$

The simple calculus on the level-2 tensor algebra $T^{(2)}\left(\mathbf{R}^d\right)$ leads to an explicit expression for $\mathbb{X}^n_{\frac{j}{n},\frac{k}{n}}$, to which one can apply the (discrete) Burkholder–Davis–Gundy inequality in order to get the discrete tightness estimates (3.13). The extension to all times is straight-forward. Details are left to the reader (see e.g. [BF13]). An alternative argument, not restricted to level 2, is found in [BFH09]. □

Note that $\mathbf{X}^n$, as constructed above, is a (random) geometric rough path. Recall that such rough paths can be viewed as genuine paths with values in the Lie group $G^{(2)}\left(\mathbf{R}^d\right) \subset T^{(2)}\left(\mathbf{R}^d\right)$. On the other hand, from (3.14), we see that $\mathbf{X}^n$ restricted to discrete times $\{\frac{j}{n} : j \in \mathbf{N}\}$ is a Lie group valued random walk, rescaled with the aid of the dilatation operator. By using central limit theorems available on such Lie groups, one can see that $\mathbf{X}^n$ at unit time converges weakly to Brownian motion, enhanced with its iterated integrals in the Stratonovich sense. Under the additional assumption that $\mathbf{E}(X \otimes X) = I$, the identity matrix, this Brownian motion is in fact a standard Brownian motion. This is enough to characterise the finite-dimensional distributions of any weak limit point and one has the following "Donsker" type result.

Theorem 3.12. *In the rescaled random walk setting of Proposition 3.11, and under the additional assumption that* $\mathbf{E}(X \otimes X) = I$, *we have the weak convergence*

$$\mathbf{X}^n \implies \mathbf{B}^{\text{Strat}}$$

in the rough path space $\mathscr{C}^\alpha([0,T], \mathbf{R}^d)$, *any* $\alpha < 1/2$.

Recall that, by definition, weak convergence is stable under push-forward by continuous maps. The interest in this result is therefore clearly given by the fact that stochastic integrals and the Itô map can be viewed as continuous maps on rough path spaces, as will be discussed in later chapters.

3.7 Exercises

Exercise 3.13. Complete the proof of Theorem 3.3.

Exercise 3.14. Bypass the use of Wiener–Itô chaos integrability in Proposition 3.4 by showing directly that the matrix-valued random variable $\mathbb{B}^{\text{Itô}}_{0,1}$ has moments of all orders. **Hint:** this is trivial for the on-diagonal entries, for the off-diagonal entries one can argue via conditioning, Itô isometry, and reflection principle.

Exercise 3.15. Show that d-dimensional Brownian motion B enhanced with Lévy's stochastic area is a degenerate diffusion process and find its generator.

Exercise 3.16 (Q-Wiener process as rough path). Consider a separable Hilbert space H with orthonormal basis (e_k), $(\lambda_k) \in l^1$, $\lambda_k > 0$ for all k, and a countable sequence $\left(\beta^k\right)$ of independent standard Brownian motions. Then the limit

$$X_t := \sum_{k=1}^{\infty} \lambda_k^{1/2} \beta_t^k e_k$$

exists a.s. and in L^2, uniformly on compacts. This defines a Q-Wiener process in the sense of [DPZ92], where $Q = \sum_k \lambda_k \langle e_k, \cdot \rangle e_k$ is symmetric, non-negative and trace-class; conversely, any such operator Q on H can be written in this form and thus gives rise to a Q-Wiener process. Show that

$$\mathbb{X}_{s,t} := \sum_{j,k=1}^{\infty} \lambda_j^{1/2} \lambda_k^{1/2} \int_s^t \beta_s^j \, d\beta_s^k \, e_j \otimes e_k$$

exists a.s. and in L^2, uniformly on compacts and so defines $\mathbb{X}$ with values in $H \otimes_{\text{HS}} H$, the closure of the algebraic tensor product $H \otimes_a H$ under the Hilbert–Schmidt norm. Consider both the case of Itô and Stratonovich integration and verify that with either choice, $(X, \mathbb{X}) \in \mathscr{C}^\alpha$ a.s. for any $\alpha < 1/2$.

Exercise 3.17 (Banach-valued Brownian motion as rough path [LLQ02]). Consider a separable Banach space V equipped with a centred Gaussian measure μ. By a standard construction (cf. [Led96]) this gives rise to a so-called abstract Wiener space (V, H, μ), with $H \subset V$ the Cameron–Martin space of μ. (Examples to have in mind are $V = H = \mathbf{R}^d$ with $\mu = N(0, I)$, or the usual Wiener space $V = \mathcal{C}([0,1])$ equipped with Wiener measure, H is then the space of absolutely continuous paths starting at zero with L^2-derivative.) There then exists a V-valued Brownian motion $(B_t : t \in [0,T])$ such that

- $B_0 = 0$,
- B has independent increments,
- $\langle B_{s,t}, v^* \rangle \sim N\big(0, (t-s)\big|v^*\big|_H^2\big)$ whenever $0 \le s < t \le T$ and $v^* \in V^* \hookrightarrow H^* \cong H$.

We assume that $V \otimes V$ is equipped with an exact tensor norm (with respect to μ) in the sense that there exists $\gamma \in [1/2, 1)$ and a constant $C > 0$ such that for any sequence $\{G_k \otimes \tilde{G}_k : k \ge 1\}$ of independent V-valued Gaussian random variables with identical distribution μ,

$$\mathbf{E}\left(\left|\sum_{k=1}^{N} G_k \otimes \tilde{G}_k\right|_{V \otimes V}^2\right) \le C N^{2\gamma} = \mathrm{o}(N).$$

a) Verify that exactness holds with $\gamma = 1/2$ whenever $\dim V < \infty$. (More generally, exactness with $\gamma = 1/2$ always holds true if one works with the injective tensor product space, $V \otimes_{\text{inj}} V$, the injective norm being the smallest possible. For the largest possible norm, the projective norm, the $\mathrm{o}(N)$-estimate remains true but can be as slow as one wishes; exactness may then fail; cf. [LLQ02]. Exactness of the usual Wiener-space, with uniform or Hölder norm, is also known to be true.)

b) Fix $\alpha < 1/2$. Show that dyadic piecewise linear approximations B^n, enhanced with $\mathbb{B}^n = \int B^n \otimes dB^n$, converge in α-Hölder rough path metric to a limit $\mathbf{B}$ in $\mathscr{C}^\alpha([0,T], V)$. More precisely, use the previous exercise to show that the sequence $\mathbf{B}^n = (B^n, \mathbb{B}^n)$ is Cauchy in the sense that

$$|\varrho_a(\mathbf{B}^n, \mathbf{B}^m)|_{L^q} \to 0 \quad \text{with} \quad n, m \to \infty\ .$$

Conclude that $\mathbf{B}^n$ converges in $\mathscr{C}^\alpha$ and L^q to some limit $\mathbf{B} \in \mathscr{C}^\alpha([0,T], V)$ a.s.

c) Show that $\mathbf{B}$ is the L^q-limit in α-Hölder rough path metric for all piecewise linear approximations, say B^{D_n}, as long as mesh $|D_n| \to 0$ with $n \to \infty$. Show that the convergence is almost sure if $|D_n| \sim 2^{-n}$ and also $|D_n| \sim 1/n$.

Solution 3.18. We only sketch the main step in the proof of b). Without loss of generality, we set $T = 1$. The crux of the matter is to show that $\mathbb{B}^n_{0,1}$ converges in $V \otimes V$. The rest follows from scaling and equivalence of moments in the first two Wiener chaoses. Set $t^n_k = k/2^n$. Then

$$
\begin{aligned}
\left|\mathbb{B}^{n+1}_{0,1}-\mathbb{B}^{n}_{0,1}\right|^2_{L^2} &\sim \mathbf{E}\left|\sum_{k=1}^{2^n} B_{t^{n+1}_{2k-2},t^{n+1}_{2k-1}}\otimes B_{t^{n+1}_{2k-1},t^{n+1}_{2k}}\right|^2_{V\otimes V}\\
&\sim \frac{1}{2^{2n+2}}\mathbf{E}\left|\sum_{k=1}^{2^n} 2^{\frac{n+1}{2}} B_{t^{n+1}_{2k-2},t^{n+1}_{2k-1}}\otimes 2^{\frac{n+1}{2}} B_{t^{n+1}_{2k-1},t^{n+1}_{2k}}\right|^2_{V\otimes V}\\
&\sim 2^{-2n-2}\mathbf{E}\left|\sum_{k=1}^{2^n} G_k\otimes \tilde{G}_k\right|^2_{V\otimes V}\\
&\lesssim 2^{-2n-2}2^{2\gamma n}\\
&\sim 2^{-2n(1-\gamma)}\;,
\end{aligned}
$$

where the penultimate bound was obtained by exactness. By definition of exactness $1-\gamma>0$ and so $\mathbb{B}^n_{0,1}$ is Cauchy in the L^2-space of $V\otimes V$-valued random variables.

Exercise 3.19. In the context of Theorem 3.8, assume M normal and show that the Lévy area correction takes the form

$$A=\frac{1}{2}\,\mathrm{Anti}(M)\,\mathrm{Sym}(M)^{-1}$$

and conclude that the correction is zero if and only if M is symmetric. Is this also true without the assumption that M is normal?

Exercise 3.20. In the context of Theorem 3.8, show that "physical Brownian motion with mass m" converges as $m\to 0$, in ϱ_α and $L^q, \alpha\in(1/2,1/3)$ and $q<\infty$, with rate

$$\mathrm{O}\left(\frac{1}{m^\theta}\right),\ \text{any } \theta<1/2-\alpha.$$

Hint: Use Theorem 3.3 to show rough path convergence. (The computations are a little longer, but of similar type, with the additional feature that the use of the ergodic theorem can be avoided.)

Exercise 3.21. Consider physical Brownian motion in dimension $d=2$, with

$$M=I-\alpha\begin{pmatrix}0&-1\\1&0\end{pmatrix},\ \alpha\in\mathbf{R}.$$

Show that the area correction of X^m, in the (small mass) limit $m\to 0$ limit, is given by

$$\frac{\alpha}{2(1+\alpha^2)}\begin{pmatrix}0&-1\\1&0\end{pmatrix}.$$

(This correction is computed by multiscale / homogenisation techniques in the book [PS08]).

Exercise 3.22. Consider $X_t=bt+\sigma B_t$ where $b\in\mathbf{R}^d$, $a=\sigma\sigma^*\in(\mathbf{R}^d)^{\otimes 2}$. In other words, X is a Lévy process with triplet $(a,b,0)$. Show that the expected signature of

X over $[0,T]$ is given by

$$\mathbf{E}S(X)_{0,T} = \exp T\left(b + \frac{1}{2}a\right).$$

Here, the exponential should be interpreted as the exponential in the tensor algebra, i.e.

$$\exp(u) = 1 + u + \frac{1}{2!}u \otimes u + \frac{1}{3!}u \otimes u \otimes u + \dots$$

Exercise 3.23 (Expected signature for Lévy processes [FS12b]). Consider a compound Poisson process Y with intensity λ and jumps distributed like $J = J(\omega) \sim \nu$. in other words, Y is Lévy with triplet $(0,0,K)$ where the Lévy measure is given by $K = \lambda\nu$. A sample path of Y gives rise to piecewise linear, continuous path; simply by connecting J_1, $J_1 + J_2$ etc. Show that, under a suitable integrability condition for J,

$$\mathbf{E}S(Y)_{0,T} = \exp T\lambda\mathbf{E}(e^J - 1).$$

Can you handle the case of a general Lévy process?

Exercise 3.24 (Level-3 cubature formula). Define a measure μ on $\mathcal{C}\big([0,1],\mathbf{R}^d\big)$ by assigning equal weight 2^{-d} to each of the paths

$$t \mapsto t\begin{pmatrix} \pm 1 \\ \pm 1 \\ \dots \\ \pm 1 \end{pmatrix} \in \mathbf{R}^d.$$

Call the resulting process $(X_t(\omega) : t \in [0,1])$ and compute the expected signature up to level 3, that is

$$\mathbf{E}\Big(1, X_{0,1}, \int_{0<t_1<t_2<1} dX_{t_1} \otimes dX_{t_2}, \int_{0<t_1<t_2<t_3<1} dX_{t_1} \otimes dX_{t_2} \otimes dX_{t_3}\Big).$$

Compare with expected signature of Brownian motion, the tensor exponential $\exp(\frac{1}{2}I)$, projected to the first 3 levels.

Solution 3.25. $X_t(\omega) = t\sum_i Z_i(\omega)e_i$ with i.i.d. random variables Z_i taking values $+1,-1$ with equal probability. Clearly,

$$\mathbf{E}\int_{0<t_1<1} dX_{t_1} = \mathbf{E}X_{t_1} = 0.$$

Then,

$$\int_{0<t_1<t_2<1} dX_{t_1} dX_{t_2} = \frac{1}{2}\sum_{i,j} Z_i Z_j e_i \otimes e_j = \frac{1}{2}I + (\text{zero mean})$$

and so the expected value at level 2 matches $\pi_2\big(\exp(\frac{1}{2}I)\big) = \frac{1}{2}I$. A similar expansion on level 3 shows that every summand either contains, for some i, a factor $\mathbf{E}Z^i_{t_1} = 0$ or

$\mathbf{E}\big(Z^i_{t_1}\big)^3 = 0$. In other words, the expected signature at level 3 is zero, in agreement with $\pi_3\big(\exp(\frac{1}{2}I)\big) = 0$. We conclude that the expected signatures, of μ on the one hand and Wiener measure on the other hand, agree up to level 3.

Exercise 3.26. Prove the Kolmogorov tightness criterion, Theorem 3.10.

3.8 Comments

The modification of Kolmogorov's criterion for rough paths (Theorem 3.1) is a minor variation on a rather well-known theme. Rough path regularity of Brownian motion was first established in the thesis of Sipiläinen, [Sip93].

For extensions to infinite dimensional Wiener processes (and also convergence of piecewise linear approximations in rough path sense) see Ledoux, Lyons and Qian [LLQ02] and Dereich [Der10]; much of the interest here is to go beyond the Hilbert space setting. The resulting stochastic integration theory against Banach-space valued Brownian motion, which in essence cannot be done by classical methods, has proven crucial in some recent applications (cf. the works of Kawabi–Inahama [IK06], Dereich [Der10]).

Early proofs of Brownian rough path regularity were typically established by convergence of dyadic piecewise linear approximations to $(B, \mathbb{B}^{\text{Strat}})$ in (p-variation) rough path metric; see e.g. Lyons–Qian [LQ02]. Many other "obvious" (but as we have seen: not all reasonable) approximations are seen to yield the same Brownian rough path limit. The discussion of Brownian motion in a magnetic field follows closely Friz, Gassiat and Lyons [FGL13]. Continuous semi-martingales and large classes of multidimensional Gaussian – and Markovian – processes lift to random rough paths; convergence of piecewise linear approximation in rough path topology is also known to hold true to hold in great generality. See e.g. Friz–Victoir [FV10b] and the references therein. The expected signature of Brownian motion was first established in the thesis of Fawcett [Faw04]; different proofs were then given by Lyons–Victoir, Baudoin and Friz–Shekhar, [LV04, Bau04, FS12b]. Fawcett's formula is central to the Kusuoka–Lyons–Victoir cubature method ([Kus01, LV04]). More generally, expected signatures capture important aspects of the law of a stochastic process. See Chevyrev [Che13]. The extension to Lévy processes, Exercise 3.23, is taken from Friz–Shekhar [FS12b]. The computation of expected signatures of large classes of stochastic processes including stopped Brownian motion and stochastic Löwner equations is presently pursued by a number of people including Lyons–Ni [LN11], Werness [Wer12] and Boedihardjo–Qian [BNQ13]. The Donsker type theorem, Theorem 3.12, in uniform topology, is a consequence of Stroock–Varadhan [SV73]; the rough path case is due to Breuillard, Friz, and Huesmann [BFH09]]. Applications to cubature are discussed in [BF13].

Chapter 4
Integration against rough paths

Abstract The aim of this section is to give a meaning to the expression $\int Y_t \, dX_t$ for a suitable class of integrands Y, integrated against a rough path X. We first discuss the case originally studied by Lyons where $Y = F(X)$. We then introduce the notion of a controlled rough path and show that this forms a natural class of integrands.

4.1 Introduction

The aim of this chapter is to give a meaning to the expression $\int Y_t \, dX_t$, for $X \in \mathscr{C}^\alpha([0,T], V)$ and Y some continuous function with values in $\mathcal{L}(V, W)$, the space of bounded linear operators from V into some other Banach space W. Of course, such an integral cannot be defined for arbitrary continuous functions Y, especially if we want the map $(X, Y) \mapsto \int Y \, dX$ to be continuous in the relevant topologies. We therefore also want to identify a "good" class of integrands Y for the rough path X.

A natural approach would be to try to define the integral as a limit of Riemann–Stieltjes sums, that is

$$\int_0^1 Y_t \, dX_t = \lim_{|\mathcal{P}| \to 0} \sum_{[s,t] \in \mathcal{P}} Y_s \, X_{s,t} \;, \tag{4.1}$$

where $\mathcal{P}$ denotes a partition of $[0,1]$ (interpreted as a finite collection of essentially disjoint intervals such that $\bigcup \mathcal{P} = [0,1]$) and $|\mathcal{P}|$ denotes the length of the largest element of $\mathcal{P}$. Such a definition - the *Young integral* - has been studied in detail in the seminal paper by Young [You36], where it was shown that such a sum converges if $X \in \mathcal{C}^\alpha$ and $Y \in \mathcal{C}^\beta$, provided $\alpha + \beta > 1$, and that the resulting bilinear map is continuous. This result is sharp in the sense that one can construct sequences of smooth functions Y^n and X^n such that $Y^n \to 0$ and $X^n \to 0$ in $\mathcal{C}^{1/2}([0,1], \mathbf{R})$, but such that $\int Y^n \, dX^n \to \infty$.

As a consequence of Young's inequality [You36], one has the bound

$$\left|\int_0^1 (Y_r - Y_0)\, dX_r\right| \le C\|Y\|_{\beta;[0,1]}\|X\|_{\alpha;[0,1]} \;, \tag{4.2}$$

with C depending on $\alpha+\beta>1$. Given paths X, Y defined on $[s,t]$ rather than $[0,1]$ it is an easy consequence of the scaling properties of Hölder semi-norms, that

$$\left|\int_s^t Y_r dX_r - Y_s X_{s,t}\right| \le C\|Y\|_\beta \|X\|_\alpha |t-s|^{\alpha+\beta} \;. \tag{4.3}$$

In particular, when $\alpha=\beta>1/2$, the right hand side is proportional to $|t-s|^{2\alpha} = \mathrm{o}(|t-s|)$ which is to be compared with the estimate (4.20) below.

The main insight of the theory of rough paths is that this seemingly unsurmountable barrier of $\alpha+\beta>1/2$ (which reduces to $\alpha>1/2$ in the case $\alpha=\beta$ which is our main interest[1]) can be broken by adding additional structure to the problem. Indeed, for a rough path X, we *postulate* the values $\mathbb{X}_{s,t}$ of the integral of X against itself, see (2.2). It is then intuitively clear that one should be able to define $\int Y\,dX$ in a consistent way, provided that Y "looks like X", at least on very small scales (in the precise sense of (4.16) below). The easiest way for a function Y to "look like X" is to have $Y_t = F(X_t)$ for some sufficiently smooth $F\colon V \to \mathcal{L}(V,W)$, called a 1-*form*.

4.2 Integration of 1-forms

We aim to integrate $Y=F(X)$ against $\mathbf{X}=(X,\mathbb{X})\in\mathscr{C}^\alpha$. When $F\colon V\to\mathcal{L}(V,W)$ is in $\mathcal{C}^1$, or better, a Taylor approximation gives

$$F(X_r) \approx F(X_s) + DF(X_s)X_{s,r}, \tag{4.4}$$

for r in some (small) interval $[s,t]$, say. Recall (see sections 1.4 and1.5 concerning the infinite-dimensional case) that[2]

$$\mathcal{L}(V,\mathcal{L}(V,W)) \cong \mathcal{L}(V\otimes V, W) \;,$$

so that $DF(X_s)$ may be regarded as element in $\mathcal{L}(V\otimes V,W)$. Since the Young integral defined in (4.1), when applied to $Y=F(X)$, is effectively based on the approximation $F(X_r)\approx F(X_s)$, for $r\in[s,t]$, it is natural to hope, with a motivating look at (2.2), that the *compensated Riemann-Stieltjes sum* appearing at the right-hand

[1] .... but see Exercise 4.25.

[2] In coordinates, when $\dim V, \dim W < \infty$, $G = DF(X_s)$ takes the form of a $(1,2)$-tensor $(G^k_{i,j})$ and the identification amounts to

$$v \mapsto \Big(\tilde v \mapsto \Big(\sum_{i,j} G^k_{i,j} v^i \tilde v^j\Big)_k\Big) \quad \text{versus} \quad M \mapsto \Big(\sum_{i,j} G^k_{i,j} M^{i,j}\Big)_k \;.$$

side of

$$\int_0^1 F(X_s)\, d\mathbf{X}_s \approx \sum_{[s,t]\in\mathcal{P}} \big(F(X_s)X_{s,t} + DF(X_s)\mathbb{X}_{s,t}\big) \,, \tag{4.5}$$

provides a good enough approximation (say, is Cauchy as $|\mathcal{P}| \to 0$) even when X ceases to have α-Hölder regularity for $\alpha > 1/2$ (as required by Young theory), but assuming instead $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$, $\alpha \in \big(\frac{1}{3}, \frac{1}{2}\big]$. Why should this be good enough? The intuition is as follows: given $\alpha \in \big(\frac{1}{3}, \frac{1}{2}\big]$ neither $|X_{s,t}| \sim |t-s|^\alpha$ nor $|\mathbb{X}_{s,t}| \sim |t-s|^{2\alpha}$ in the above sum will be negligible as $|\mathcal{P}| \to 0$. Continuing in the same fashion, one expects (in fact one can show it) that the third iterated integral $\mathbb{X}^{(3)}_{s,t}$ is of order $\mathbb{X}^{(3)}_{s,t} \sim |t-s|^{3\alpha} = \mathrm{o}(|t-s|)$, so that adding a third term of the form $D^2F(X_s)\mathbb{X}^{(3)}_{s,t}$ in the sum of (4.5), at the very least, will not affect any limit, should it exist. In the following, we will see that this limit,[3]

$$\int_0^1 F(X_s)\, d\mathbf{X}_s = \lim_{|\mathcal{P}|\to 0} \sum_{[s,t]\in\mathcal{P}} \big(F(X_s)X_{s,t} + DF(X_s)\mathbb{X}_{s,t}\big), \tag{4.6}$$

does exists and call it *rough integral*.[4] In fact, in this section we shall construct the (indefinite) rough integral $Z = \int_0^\cdot F(X)d\mathbf{X}$ as element in $\mathcal{C}^\alpha$, i.e. as *path*, similar to the construction of stochastic integrals *as processes* rather than random variables. Even this may not be sufficient in applications - one often wants to have an extended meaning of the rough integral, such as $(Z, \mathbf{Z}) \in \mathscr{C}^\alpha$, point of view emphasised in [Lyo98, LQ02, LCL07], or something similar (such as "Z controlled by X" in the sense of Definition 4.6 below, to be discussed in the next section).

Lemma 4.1. *Let $F\colon V \to \mathcal{L}(V,W)$ be a $\mathcal{C}^2_b$ function and let $(X,\mathbb{X}) \in \mathscr{C}^\alpha$ for some $\alpha > \frac{1}{3}$. Set $Y_s := F(X_s)$, $Y'_s := DF(X_s)$ and $R^Y_{s,t} := Y_{s,t} - Y'_s X_{s,t}$. Then*

$$Y, Y' \in \mathcal{C}^\alpha \quad \text{and} \quad R^Y \in \mathcal{C}^{2\alpha}. \tag{4.7}$$

(In the terminology of the forthcoming Definition 4.6: "Y is controlled by X with Gubinelli derivative Y'; in symbols $(Y,Y') \in \mathscr{D}^{2\alpha}_X$".) More precisely, we have the estimates

$$\begin{aligned} \|Y\|_\alpha &\le \|DF\|_\infty \|X\|_\alpha, \\ \|Y'\|_\alpha &\le \big\|D^2F\big\|_\infty \|X\|_\alpha, \\ \big\|R^Y\big\|_{2\alpha} &\le \frac{1}{2}\big\|D^2F\big\|_\infty \|X\|^2_\alpha. \end{aligned}$$

---

[3] Recall that $\lim_{|\mathcal{P}|\to 0}$ means convergence along any sequence $(\mathcal{P}_n)$ with mesh $|\mathcal{P}_n| \to 0$, with identical limit along each such sequence. In particular, it is not enough to establish convergence along a particular sequence $(\mathcal{P}_n)$, although a particular sequence may be used to identify the limit.

[4] Of course, we can and will consider intervals other than $[0,1]$. Without further notice, $\mathcal{P}$ always denotes a partition of the interval under consideration.

Proof. $\mathcal{C}_b^2$ regularity of F implies that F and DF are both Lipschitz continuous with Lipschitz constants $\|DF\|_\infty$ and $\|D^2F\|_\infty$ respectively. The α-Hölder bounds on Y and Y' are then immediate. For the remainder term, consider the function

$$[0,1] \ni \xi \mapsto F(X_s + \xi X_{s,t}) \; .$$

A Taylor expansion, with intermediate value remainder, yields $\xi \in (0,1)$ such that

$$R^Y_{s,t} = F(X_t) - F(X_s) - DF(X_s)X_{s,t} = \frac{1}{2}D^2F(X_s + \xi X_{s,t})(X_{s,t}, X_{s,t}) \; .$$

The claimed 2α-Hölder estimate, in the sense that $|R^Y_{s,t}| \lesssim |t-s|^{2\alpha}$, then follows at once. $\square$

Before we prove that the rough integral (4.6) exists, we discuss some sort of abstract Riemann integration. In what follows, at first reading, one may have in mind the construction of a Riemann-Stieltjes (or Young) integral $Z_t := \int_0^t Y_r dX_r$. From Young's inequality (4.3), one has (with $Z_{s,t} = Z_t - Z_s$ as usual)

$$Z_{s,t} = Y_s X_{s,t} + \mathrm{o}(|t-s|)$$

and $\varXi_{s,t} := Y_s X_{s,t}$ is a sufficiently good local approximation in the sense that it fully determines the integral Z via the limiting procedure given in (4.1)). In this sense $Z = \mathcal{I}\varXi$ is the well-defined image of $\varXi$ under some *abstract integration map* $\mathcal{I}$. Note that $Z_{s,t} = Z_{s,u} + Z_{u,t}$, i.e. increments are additive (or "multiplicative" if one regards $+$ as group operation[5]) whereas a similar property fails for $\varXi$. In the language of [Lyo98], such a $\varXi$ corresponds to a "almost multiplicative functional" and it is a key result in the theory that there is a unique associated "multiplicative functional" (here: $Z = \mathcal{I}\varXi$). Following [Gub04, FdLP06] we call "sewing" the step from a (good enough) local approximation $\varXi$ to some (abstract) integral $\mathcal{I}\varXi$; the concrete estimate which quantifies how well $\mathcal{I}\varXi$ is approximated by $\varXi$ will be called "sewing lemma". It plays an analogous role to "Davie's lemma" (cf. section 8.7) in the context of (rough) differential equations.

We now formalize what we mean by $\varXi$ being a good enough local approximations. For this, we introduce the space $\mathcal{C}_2^{\alpha,\beta}([0,T],W)$ of functions $\varXi$ from the simplex $0 \le s \le t \le T$ into W such that $\varXi_{t,t} = 0$ and such that

$$\|\varXi\|_{\alpha,\beta} \stackrel{\text{def}}{=} \|\varXi\|_\alpha + \|\delta\varXi\|_\beta < \infty \; , \tag{4.8}$$

where $\|\varXi\|_\alpha = \sup_{s<t} \frac{|\varXi_{s,t}|}{|t-s|^\alpha}$ as usual, and also

$$\delta\varXi_{s,u,t} = \varXi_{s,t} - \varXi_{s,u} - \varXi_{u,t}, \quad \|\delta\varXi\|_\beta \stackrel{\text{def}}{=} \sup_{s<u<t} \frac{|\delta\varXi_{s,u,t}|}{|t-s|^\beta} \; .$$

[5] This terminology becomes natural if one considers Z together with its iterated integrals as group-valued path, increments of which satisfy Chen's "multiplicative" relation, see (2.3).

Provided that $\beta > 1$, it turns out that such functions are "almost" of the form $\Xi_{s,t} = F_t - F_s$, for some α-Hölder continuous function F (they would be if and only if $\delta\Xi = 0$). Indeed, it is possible to construct in a canonical way a function $\hat{\Xi}$ with $\delta\hat{\Xi} = 0$ and such that $\hat{\Xi}_{s,t} \approx \Xi_{s,t}$ for $|t-s| \ll 1$:

Lemma 4.2 (Sewing lemma). *Let α and β be such that $0 < \alpha \le 1 < \beta$. Then, there exists a (unique) continuous map $\mathcal{I}\colon \mathcal{C}_2^{\alpha,\beta}([0,T],W) \to \mathcal{C}^\alpha([0,T],W)$ such that $(\mathcal{I}\Xi)_0 = 0$ and*

$$\big|(\mathcal{I}\Xi)_{s,t} - \Xi_{s,t}\big| \le C|t-s|^\beta \;. \tag{4.9}$$

where C only depends on β and $\|\delta\Xi\|_\beta$. (The α-Hölder norm of $\mathcal{I}\Xi$ also depends on $\|\Xi\|_\alpha$ and hence on $\|\Xi\|_{\alpha,\beta}$.)

Proof. Note first that $\mathcal{I}$ will be built as a linear map, so that its continuity is an immediate consequence of its boundedness. Uniqueness of $\mathcal{I}$ is also immediate. Indeed, assume by contradiction that, for a given Ξ, there are two candidates F and $\bar{F}$ for $\mathcal{I}\Xi$. Since both of these functions have to satisfy the bound (4.9), the function $F - \bar{F}$ satisfies $(F-\bar{F})_0 = 0$ and $(F-\bar{F})_{s,t} \lesssim |t-s|^\beta$. Since $\beta > 1$ by assumption, it follows immediately that $F - \bar{F}$ vanishes identically.

It remains to find the map $\mathcal{I}$. It is very natural to make the guess

$$\big(\mathcal{I}\Xi\big)_{s,t} = \lim_{|\mathcal{P}|\to 0} \sum_{[u,v]\in\mathcal{P}} \Xi_{u,v} \;, \tag{4.10}$$

where $\mathcal{P}$ denotes a partition of $[s,t]$ and $|\mathcal{P}|$ denotes its mesh, i.e. the length of its largest element. The remainder of the proof shows that this expression is well-defined and that (4.9) holds.

Why is (4.10) well-defined? Because of its importance we give *two* (independent but related) arguments. The first argument is based on successive (dyadic) refinement, i.e. one starts by identifying the integral as limit of Riemann type sums, along a particular sequence $(\mathcal{P}_n)$. This is followed by checking that the limit is indeed independent of the choice of partitions. More precisely, for a given interval $[s,t]$, we start with the trivial partition $\mathcal{P}_0 = \{[s,t]\}$ and we set $\big(\mathcal{I}^0\Xi\big)_{s,t} = \Xi_{s,t}$. We then define recursively

$$\mathcal{P}_{n+1} = \bigcup_{[u,v]\in\mathcal{P}_n} \big\{\,[u,m],\,[m,v]\,\big\} \;,$$

with $m := m(u,v) := (u+v)/2$ so that $\mathcal{P}_n$, the level-n dyadic partion of $[s,t]$ contains 2^n intervals, each of length $2^{-n}|t-s|$. We then set

$$\big(\mathcal{I}^{n+1}\Xi\big)_{s,t} \stackrel{\text{def}}{=} \sum_{[u,v]\in\mathcal{P}_{n+1}} \Xi_{u,v} = \big(\mathcal{I}^n\Xi\big)_{s,t} - \sum_{[u,v]\in\mathcal{P}_n} \delta\Xi_{u,m,v} \;,$$

where it is a straightforward exercise to check that the second equality holds. It then follows immediately from the definition of $\|\cdot\|_{\alpha,\beta}$ that

$$\big|\big(\mathcal{I}^{n+1}\Xi\big)_{s,t} - \big(\mathcal{I}^n\Xi\big)_{s,t}\big| \le 2^{n(1-\beta)}|t-s|^\beta\|\delta\Xi\|_\beta \;.$$

Since $\beta > 1$, these terms are summable and we conclude immediately that the sequence $(\mathcal{I}^n \Xi)_{s,t}$ is Cauchy. It thus admits a limit $(\mathcal{I}\Xi)_{s,t}$ such that, by summing up the bound above, one has

$$\big|\big(\mathcal{I}\Xi\big)_{s,t} - \Xi_{s,t}\big| \le \sum_{n\ge 0} \big|\big(\mathcal{I}^{n+1}\Xi\big)_{s,t} - \big(\mathcal{I}^{n}\Xi\big)_{s,t}\big| \le C\|\delta\Xi\|_\beta |t-s|^\beta\;, \quad (4.11)$$

for some universal constant C depending only on β, which is precisely the required bound (4.9). It remains to see that the limit just constructed is independent of the choice of partitions. *Once one has shown that* $\delta\mathcal{I}\Xi = 0$, which is equivalent to $(\mathcal{I}\Xi)_{0,t} = (\mathcal{I}\Xi)_{0,s} + (\mathcal{I}\Xi)_{s,t}$ for all pairs s,t, this is not too difficult. Indeed, if $\mathcal{P}$ denotes an arbitrary partition of $[s,t]$ and we introduce

$$\int_{\mathcal{P}} \Xi := \sum_{[u,v]\in\mathcal{P}} \Xi_{u,v}\,,$$

then the difference between $(\mathcal{I}\Xi)_{s,t}$ and this approximation can the be estimated, thanks to (4.11) as

$$\Big|\sum_{[u,v]\in\mathcal{P}} \big((\mathcal{I}\Xi)_{u,v} - \Xi_{u,v}\big)\Big| = \mathrm{O}\big(|\mathcal{P}|^{\beta-1}\big)\;.$$

Since $\beta > 1$, this is enough to show that $(\mathcal{I}\Xi)_{s,t}$ is the limit along any sequence $\mathcal{P}_n$ with mesh tending to zero. What remains to be shown is $\delta\mathcal{I}\Xi = 0$. In general this is not obvious (but see Remark 4.3) and indeed, writing $\mathcal{P}_n^{s,t}$ for the level-n dyadic partition relative to $[s,t]$, as used above, this is quite tedious since $\mathcal{P}_n^{0,t}$ is not equal to the partition of $[0,t]$ given by $\mathcal{P}_n^{0,s} \cup \mathcal{P}_n^{s,t}$, even though both have mesh tending to zero with $n \to \infty$. In fact, one is better off to define the integral over $[s,t]$ as the limit of $\sum_{[u,v]\in\mathcal{P}_n^{0,T},[u,v]\subset[s.t]} \Xi_{u,v}$. In Exercise 4.21, the reader is invited to work out the remaining details.

The *second* argument, which is essentially due to Young, yields immediately convergence as $|\mathcal{P}| \to 0$, i.e. the same limit is obtained along any sequence $\mathcal{P}_n$ with mesh tending to zero. (As an immediate consequence, without any details left to the reader, $\delta\mathcal{I}\Xi = 0$. Another advantage of Young's construction is that it works under a weaker $1/\alpha$-variation assumption on $(X,\mathbb{X})$.) Consider a partition $\mathcal{P}$ of $[s,t]$ and let $r \ge 1$ be the number of intervals in $\mathcal{P}$. When $r \ge 2$ there exists $u \in [s,t]$ such that $[u_-,u],[u,u_+] \in \mathcal{P}$ and

$$|u_+ - u_-| \le \frac{2}{r-1}|t-s|.$$

Indeed, assuming otherwise gives the contradiction $2|t-s| \ge \sum_{u\in\mathcal{P}^\circ} |u_+ - u_-| > 2|t-s|$. Hence, $|\int_{\mathcal{P}\setminus\{u\}} \Xi - \int_{\mathcal{P}} \Xi| = |\delta\Xi_{u_-,u,u_+}| \le \|\delta\Xi\|_\beta\,(2|t-s|/(r-1))^\beta$ and by iterating this procedure until the partition is reduced to $\mathcal{P} = \{[s,t]\}$, we arrive at the *maximal inequality*,

$$\sup_{\mathcal{P}\subset[s,t]}\Big|\varXi_{s,t}-\int_{\mathcal{P}}\varXi\Big|\le 2^{\beta}\|\delta\varXi\|_{\beta}\zeta(\beta)|t-s|^{\beta}\ ,$$

where ζ denotes the classical ζ function. It then remains to show that

$$\sup_{|\mathcal{P}|\vee|\mathcal{P}'|<\varepsilon}\Big|\int_{\mathcal{P}}\varXi-\int_{\mathcal{P}'}\varXi\Big|\to 0\quad \text{as } \varepsilon\downarrow 0, \tag{4.12}$$

which implies existence of $\mathcal{I}\varXi$ as the limit $\lim_{|\mathcal{P}|\to 0}\int_{\mathcal{P}}\varXi$. To this end, at the price of adding / subtracting $\mathcal{P}\cup\mathcal{P}'$, we can assume without loss of generality that $\mathcal{P}'$ refines $\mathcal{P}$. In particular, then $|\mathcal{P}|\vee|\mathcal{P}'|=|\mathcal{P}|$ and

$$\int_{\mathcal{P}}\varXi-\int_{\mathcal{P}'}\varXi=\sum_{[u,v]\in\mathcal{P}}\Big(\varXi_{u,v}-\int_{\mathcal{P}'\cap[u,v]}\varXi\Big)\ .$$

But then, for any $\mathcal{P}$ with $|\mathcal{P}|\le\varepsilon$ we can use the maximal inequality to see that

$$\Big|\int_{\mathcal{P}}\varXi-\int_{\mathcal{P}'}\varXi\Big|\le 2^{\beta}\zeta(\beta)\|\delta\varXi\|_{\beta}\sum_{[u,v]\in\mathcal{P}}|v-u|^{\beta}=\mathrm{O}\big(|\mathcal{P}|^{\beta-1}\big)=\mathrm{O}(\varepsilon^{\beta-1}).$$

This concludes the Young argument (with no hidden tedium left to the reader). □

Remark 4.3. The first argument ultimately suffered from the tedium of checking the additivity property $\delta\mathcal{I}\varXi=0$. In some cases, however, this addivity property of $\mathcal{I}\varXi$ can be immediate. Imagine $X:[0,T]\to V$ is smooth, $\mathbb{X}=\int X\otimes dX$, and one is only interested in an error estimate for second order approximations of Riemann-Stieltjes integrals, of the form

$$\Big|\int_s^t F(X_r)\,dX_r-F(X_s)X_{s,t}-DF(X_s)\mathbb{X}_{s,t}\Big|\le\ \text{right-hand side of (4.13).}$$

(This is still a highly non-trivial estimate since the right-hand side is uniform over all (smooth) paths, as long as their α-rough path norms remain bounded!) In the context of the above proof, this estimate is contained in the first step, applied with

$$\varXi_{s,t}=F(X_s\,)X_{s,t}+DF(X_s)\mathbb{X}_{s,t}.$$

But here it is clear from classical Riemann-Stieltjes theory, or in fact just Riemann integration theory, that $\mathcal{I}\varXi_{s,t}$, constructed as limit of dyadic partitions of $[s,t]$, is precisely the Riemann-Stieltjes integral $\int_s^t F(X_r)\,dX_r$ and therefore additive. (The contribution of $DF(X)\mathbb{X}$ in the approximations disappears in the limit; indeed, it suffices to remark that $\mathbb{X}_{u,v}\sim|v-u|^2$, thanks to smoothness of X.)

We now apply the sewing lemma to the construction of (4.6). We have the following.

Theorem 4.4 (Lyons). *Let* $\mathbf{X}=(X,\mathbb{X})\in\mathscr{C}^{\alpha}([0,T],V)$ *for some* $T>0$ *and* $\alpha>\frac{1}{3}$, *and let* $F\colon V\to\mathcal{L}(V,W)$ *be a* $\mathcal{C}_b^2$ *function. Then, the rough integral defined* *in (4.6) exists and one has the bound*

$$\Big|\int_s^t F(X_r)\,d\mathbf{X}_r - F(X_s)X_{s,t} - DF(X_s)\mathbb{X}_{s,t}\Big|$$
$$\lesssim \|F\|_{\mathcal{C}_b^2}\Big(\|X\|_\alpha^3 + \|X\|_\alpha \|\mathbb{X}\|_{2\alpha}\Big)|t-s|^{3\alpha}\;, \quad (4.13)$$

where the proportionality constant depends only on α. Furthermore, the indefinite rough integral is α-Hölder continuous on $[0,T]$ and we have the following quantitative estimate,

$$\Big\|\int_0^\cdot F(X)\,d\mathbf{X}\Big\|_\alpha \le C\|F\|_{\mathcal{C}_b^2}\Big(|\!|\!|\mathbf{X}|\!|\!|_\alpha \vee |\!|\!|\mathbf{X}|\!|\!|_\alpha^{1/\alpha}\Big), \quad (4.14)$$

where the constant C only depends on T and α and can be chosen uniformly in $T \le 1$. Furthermore, $|\!|\!|\mathbf{X}|\!|\!|_\alpha = \|X\|_\alpha + \sqrt{\|\mathbb{X}\|_{2\alpha}}$ denotes again the homogeneous α-Hölder rough path norm.

Remark 4.5. We will see in Section 4.4 that the map $(X,\mathbb{X}) \in \mathscr{C}^\alpha \mapsto \int_0^\cdot F(X)\,d\mathbf{X} \in \mathcal{C}^\alpha$ is continuous in α-Hölder rough path metric.

Proof. Let us stress the fact that the argument given here only relies on the properties of the integrand $Y = F(X)$ collected in Lemma 4.1 above. In particular, the generalisation to "extended" integrands (Y, Y'), which replace $(F(X), DF(X))$, subject to (4.7), will be immediate. (We shall develop this "Gubinelli" point of view further in Section 4.3 below.)

The result follows as a consequence of Lemma 4.2. With the notation that we just introduced, the classical Young integral [You36] can be defined as the usual limit of Riemann sums by

$$\int_s^t Y_r\,dX_r = \big(\mathcal{I}\Xi\big)_{s,t}\;, \qquad \Xi_{s,t} = Y_s\,X_{s,t}\;.$$

Unfortunately, this definition satisfies the identity

$$\delta\Xi_{s,u,t} = -Y_{s,u}\,X_{u,t}\;,$$

so that, except in trivial cases, the required bound (4.8) is satisfied only if Y and X are Hölder continuous with Hölder exponents adding up to $\beta > 1$. In order to be able to cover the situation $\alpha < \frac{1}{2}$, it follows that we need to consider a better approximation to the Riemann sums, as discussed above. To this end, we use the notation from Lemma 4.1, namely

$$Y_s := F(X_s)\,, \quad Y'_s := DF(X_s) \quad \text{and} \quad R^Y_{s,t} := Y_{s,t} - Y'_s X_{s,t},$$

and then set $\Xi_{s,t} = Y_s\,X_{s,t} + Y'_s\,\mathbb{X}_{s,t}$. Note that, for any $u \in (s,t)$, we have the identity

$$\delta\Xi_{s,u,t} = -R^Y_{s,u}\,X_{u,t} - Y'_{s,u}\mathbb{X}_{u,t}\;.$$

Thanks to the α-Hölder regularity of X, Y' and the 2α-regularity of $R, \mathbb{X}$, the triangle inequality shows that (4.8) holds true with the given $\alpha > 1/3$ and $\beta := 3\alpha > 1$. The fact that the integral is well-defined, and the bound

$$\left| \int_s^t Y d\mathbf{X} - Y_s X_{s,t} - Y'_s \mathbb{X}_{s,t} \right| \lesssim \left(\|X\|_\alpha \|R^Y\|_{2\alpha} + \|\mathbb{X}\|_{2\alpha} \|Y'\|_\alpha \right) |t-s|^{3\alpha} \tag{4.15}$$

then follow immediately from (4.11). Upon substituting the estimate obtained in Lemma 4.1, we obtain (4.13).

We now turn to the proof of (4.14). Writing $Z = \int F(X) d\mathbf{X}$ and using the triangle inequality in (4.13) gives

$$\begin{aligned} |Z_{s,t}| &\le \|F\|_\infty |X_{s,t}| + \|DF\|_\infty |\mathbb{X}_{s,t}| \\ &\quad + C\|F\|_{\mathcal{C}_b^2} \Big(\|X\|_\alpha^3 + \|X\|_\alpha \|\mathbb{X}\|_{2\alpha} \Big) |t-s|^{3\alpha} \\ &\le C\|F\|_{\mathcal{C}_b^2} \Big[\mathrm{A}_1 |t-s|^\alpha + \mathrm{A}_2 |t-s|^{2\alpha} + \mathrm{A}_3 |t-s|^{3\alpha} \Big] , \end{aligned}$$

with $\mathrm{A}_i \le \|\!|\mathbf{X}|\!\|_\alpha$, for $1 \le i \le 3$. Allowing C to change, this already implies

$$\|Z\|_\alpha \le C\|F\|_{\mathcal{C}_b^2} \left(\|\!|\mathbf{X}|\!\|_\alpha \vee \|\!|\mathbf{X}|\!\|_\alpha^3 \right) ,$$

which is the claimed estimate (4.14) in the limit $\alpha \downarrow 1/3$. However, one can do better by realising that the above estimate is best for $|t-s|$ small, whereas for $t-s$ large it is better to split up $|Z_{s,t}|$ into the sum of small increments. To make this more precise, set $\varrho := \|\!|\mathbf{X}|\!\|_\alpha$ and write (hide factor $C = C(\alpha, T)$ in $\lesssim$ below)

$$\begin{aligned} |Z_{s,t}| &\lesssim \varrho |t-s|^\alpha + \varrho^2 |t-s|^{2\alpha} + \varrho^3 |t-s|^{3\alpha} \\ &\le 3\varrho |t-s|^\alpha \text{ for } \varrho^{1/\alpha} |t-s| \le 1. \end{aligned}$$

Increments of Z over $[s,t]$ with length greater than $h := \varrho^{-1/\alpha}$ are handled by cutting them into pieces of length h. More precisely (cf. Exercise 4.24) we have $\|Z\|_{\alpha;h} \le 3\varrho$ which entails

$$\|Z\|_\alpha \le 3\varrho \Big(1 \vee 2h^{-(1-\alpha)} \Big) \le 6 \Big(\varrho \vee \varrho^{1/\alpha} \Big).$$

At last, we note that $C = C(\alpha, T)$ can be chosen uniformly in $T \le 1$. □

4.3 Integration of controlled rough paths

Motivated by Lemma 4.7 and the observation that rough integration essentially relies on the properties (4.7) we introduce the notion of a *controlled* path Y, relative to some "reference" path X, due to Gubinelli [Gub04]. For the sake of the following definition we assume that Y takes values in some Banach space, say $\bar{W}$. When it comes to the definition of a rough integral we typically take $\bar{W} = \mathcal{L}(V, W)$; although other choices can be useful (see e.g. remark 4.11). In the context of rough differential equations, with solutions in $\bar{W} = W$, we actually need to integrate $f(Y)$, which will be seen to be controlled by X for sufficiently smooth coefficients $f : W \to \mathcal{L}(V, W)$.

Definition 4.6. Given a path $X \in \mathcal{C}^\alpha([0,T], V)$, we say that $Y \in \mathcal{C}^\alpha([0,T], \bar{W})$ is *controlled* by X if there exists $Y' \in \mathcal{C}^\alpha([0,T], \mathcal{L}(V, \bar{W}))$ so that the remainder term R^Y given implicitly through the relation

$$Y_{s,t} = Y'_s \, X_{s,t} + R^Y_{s,t} \;, \tag{4.16}$$

satisfies $\|R^Y\|_{2\alpha} < \infty$. This defines the space of *controlled rough paths*,

$$(Y, Y') \in \mathscr{D}^{2\alpha}_X([0,T], \bar{W}).$$

Although Y' is not, in general, uniquely determined from Y (cf. Remark 4.7 and Section 6 below) we call any such Y' the *Gubinelli derivative* of Y (with respect to X).

Here, $R^Y_{s,t}$ takes values in $\bar{W}$, and the norm $\| \cdot \|_{2\alpha}$ for a function with two arguments is given by (2.3) as before. We endow the space $\mathscr{D}^{2\alpha}_X$ with the semi-norm

$$\|Y, Y'\|_{X,2\alpha} \stackrel{\text{def}}{=} \|Y'\|_\alpha + \|R^Y\|_{2\alpha} \;. \tag{4.17}$$

As in the case of classical Hölder spaces, $\mathscr{D}^{2\alpha}_X$ is a Banach space under the norm $(Y, Y') \mapsto |Y_0| + |Y'_0| + \|Y, Y'\|_{X,2\alpha}$. This quantity also controls the α-Hölder regularity of Y since, for fixed X,

$$\|Y\|_\alpha \le \|R^Y\|_\alpha + \|Y'\|_\infty \|X\|_\alpha \le C(1 + \|X\|_\alpha)\,(|Y'_0| + \|Y, Y'\|_{X,2\alpha}) \;, \tag{4.18}$$

where the constant C only depends on T and α and in fact can be chosen uniformly over $T \in (0,1]$.

Remark 4.7. Since we only assume that $\|Y\|_\alpha < \infty$, but then impose that $\|R^Y\|_{2\alpha} < \infty$, it is in general the case that a genuine cancellation takes place in (4.16). The question arises to what extent Y determines Y'. Somewhat contrary to the classical situation, where a smooth function has a unique derivative, too much regularity of the underlying rough path **X** leads to less information about Y'. For instance, if Y is smooth, or in fact in $\mathcal{C}^{2\alpha}$, and the underlying rough path **X** happens to have a path component X that is also $\mathcal{C}^{2\alpha}$, then we may take $Y' = 0$, but as a matter of fact any continuous path Y' would satisfy (4.16) with $\|R\|_{2\alpha} < \infty$. On the other hand, if X is far from smooth, i.e. genuinely rough on all (small) scales, uniformly in all directions, then Y' is uniquely determined by Y, cf. Section 6 below.

Remark 4.8. It is important to note that while the space of rough paths $\mathscr{C}^\alpha$ is not even a vector space, the space $\mathscr{D}^{2\alpha}_X$ is a perfectly normal Banach space for any given $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$. The twist of course is that the space in question depends in a crucial way on the choice of $\mathbf{X}$. The set of all pairs $(\mathbf{X}; (Y, Y'))$ gives rise to the total space

$$\mathscr{C}^\alpha \ltimes \mathscr{D}^{2\alpha} \stackrel{\text{def}}{=} \bigsqcup_{\mathbf{X} \in \mathscr{C}^\alpha} \{\mathbf{X}\} \times \mathscr{D}_X^{2\alpha},$$

with base space $\mathscr{C}^\alpha$ and "fibres" $\mathscr{D}_X^{2\alpha}$. While this looks reminiscent of fibre-bundles like the tangent bundles of a smooth manifold, it is quite different in the sense that the fibre spaces are in general *not* isomorphic. Loosely speaking, the rougher the underlying path X, the "smaller" is $\mathscr{D}_X^{2\alpha}$, see Chapter 6.

Remark 4.9. While the notion of "controlled rough path" has many appealing features, it does not come with a natural approximation theory. To wit, consider $\big(X, \mathbb{X}\big) \in \mathscr{C}_g^\alpha\big([0,T], \mathbf{R}^d\big)$ as limit of smooth paths $X_n : [0,T] \to \mathbf{R}^d$ in the sense of Proposition 2.5. Then it is natural to approximate $Y = F(X)$ by the $Y_n = F(X_n)$, which is again smooth (to the extent that F permits). On the other hand, there are no obvious approximations $(Y_n, Y_n') \in \mathscr{D}_{X^n}^{2\alpha}$ for an arbitrary controlled rough path $(Y, Y') \in \mathscr{D}_X^{2\alpha}$.

We are now ready to extend Young's integral to that of a path controlled by X against $\mathbf{X} = (X, \mathbb{X})$. Recall from Lemma 4.1 that $Y = F(X)$, with $Y' = DF(X)$, is somewhat the prototype of a controlled rough path. The definition of the rough integral $\int F(X) d\mathbf{X}$ in terms of compensated Riemann sums, cf. (4.6), then immediately suggests to define the integral of Y against $\mathbf{X}$ by[6]

$$\int_0^1 Y \, d\mathbf{X} \stackrel{\text{def}}{=} \lim_{|\mathcal{P}| \to 0} \sum_{[s,t] \in \mathcal{P}} \big(Y_s \, X_{s,t} + Y_s' \, \mathbb{X}_{s,t}\big) \,, \tag{4.19}$$

where we took $\bar{W} = \mathcal{L}(V, W)$ and used the canonical injection $\mathcal{L}(V, \mathcal{L}(V, W)) \hookrightarrow \mathcal{L}(V \otimes V, W)$ in writing $Y_s' \mathbb{X}_{s,t}$. With these notations, the resulting integral takes values in W.

With these notations at hand, it is now straight-forward to prove the following result, which is a slight reformulation of [Gub04, Prop 1]:

Theorem 4.10 (Gubinelli). *Let $T > 0$, let $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha([0,T], V)$ for some $\alpha > \frac{1}{3}$, and let $(Y, Y') \in \mathscr{D}_X^{2\alpha}\big([0,T], \mathcal{L}(V, W)\big)$. Then there exists a constant C depending only on T and α (and C can be chosen uniformly over $T \in (0,1]$) such that*

a) The integral defined in (4.19) exists and, for every pair s, t, one has the bound

$$\Big| \int_s^t Y_r \, d\mathbf{X}_r - Y_s X_{s,t} - Y_s' \, \mathbb{X}_{s,t} \Big| \le C \big(\|X\|_\alpha \|R^Y\|_{2\alpha} + \|\mathbb{X}\|_{2\alpha} \|Y'\|_\alpha \big) |t - s|^{3\alpha}. \tag{4.20}$$

b) The map from $\mathscr{D}_X^{2\alpha}\big([0,T], \mathcal{L}(V, W)\big)$ to $\mathscr{D}_X^{2\alpha}\big([0,T], W\big)$ given by

---

[6] Note the abuse of notation: we hide dependence on Y' which in general affects the limit but is usually clear from the context.

$$(Y, Y') \mapsto (Z, Z') := \left(\int_0^{\cdot} Y_t \, d\mathbf{X}_t, Y \right), \tag{4.21}$$

is a continuous linear map between Banach spaces and one has the bound

$$\|Z, Z'\|_{X,2\alpha} \le \|Y\|_\alpha + \|Y'\|_{L^\infty} \|\mathbb{X}\|_{2\alpha} + C\big(\|X\|_\alpha \|R^Y\|_{2\alpha} + \|\mathbb{X}\|_{2\alpha}\|Y'\|_\alpha\big) .$$

Proof. Part a) is an immediate consequence of Lemma 4.2, as already pointed out in the proof of Theorem 4.4. The estimate (4.20) was pointed out explicitly in (4.15).

The continuity is a consequence of the continuity of $\mathcal{I}$ in Lemma 4.2, and will be discussed in full detail in Section 4.4 below. It remains to show the bound on $\|Z, Z'\|_{X,2\alpha}$. Splitting up the left hand side of (4.20) after the first term, using the triangle inequality, gives immediately an α Hölder estimate on $\int_s^t Y_r dX_r = Z_{s,t}$, so that $Z \in \mathcal{C}^\alpha$. ($Z' = Y \in \mathcal{C}^\alpha$ is trivial, by the very nature of Y.) Similarly, splitting up the left hand side of (4.20) after the second term, gives an 2α-Hölder type estimate estimate on $\int_s^t Y_r dX_r - Y_s X_{s,t} = Z_{s,t} - Z'_s X_{s,t} =: R^Z_{s,t}$, i.e. on the remainder term in the sense of (4.16). The explicit estimate for $\|Z, Z'\|_{X,2\alpha} = \|Y\|_\alpha + \|R^Z\|_{2\alpha}$ is then obvious. □

Remark 4.11. As in the above theorem, assume that $(X, \mathbb{X}) \in \mathscr{C}^\alpha([0,T], V)$ and consider Y and Z two paths controlled by X. More precisely, we assume $(Y, Y') \in \mathscr{D}_X^{2\alpha}([0,T], \mathcal{L}(\bar{V}, W))$ and $(Z, Z') \in \mathscr{D}_X^{2\alpha}([0,T], \bar{V})$, where of course $V, \bar{V}, W$ are all Banach spaces. Then, in terms of the abstract integration map $\mathcal{I}$ (cf. the sewing lemma) we may define the integral of Y against Z, with values in W, as follows,

$$\int_s^t Y_u \, dZ_u \stackrel{\text{def}}{=} (\mathcal{I}\varXi)_{s,t} \ , \quad \varXi_{u,v} = Y_u \, Z_{u,v} + Y'_u Z'_u \mathbb{X}_{u,v} \ . \tag{4.22}$$

Here, we use the fact that $Z'_u \in \mathcal{L}(V, \bar{V})$ can be canonically identified with an operator in $\mathcal{L}(V \otimes V, V \otimes \bar{V})$ by acting only on the second factor, and $Y'_u \in \mathcal{L}(V, \mathcal{L}(\bar{V}, W))$ is identified as before with an operator in $\mathcal{L}(V \otimes \bar{V}, W)$. The reader may be helped to see this spelled out in coordinates, assuming finite dimensions: using indices i, j in $W, \bar{V}$ respectively, and then k, l in V:

$$(\varXi_{u,v})^i = (Y_u)^i_j (Z_{u,v})^j + (Y'_u)^i_{k,j} (Z'_u)^j_l (\mathbb{X}_{u,v})^{k,l}.$$

Note that, relative to the definition of $\varXi$ in the previous proof, it suffices to replace X by Z and Y' by $Y'Z'$. Making this substitution in $\delta\varXi$, as it appears in the aforementioned proof, then gives

$$\delta\varXi_{s,u,t} = -R^Z_{s,u} X_{u,t} - (Y'Z')_{s,u} \mathbb{X}_{u,t}$$

in the present situation. Clearly $Y'Z' \in \mathcal{C}^\alpha$ and so $\|\delta\varXi\|_\beta$ is finite which allows the proof to go through mutatis mutandis. In particular, (4.20) is valid, with the above substitution, and reads

$$\Big|\int_s^t Y_r\, dZ_r - Y_s Z_{s,t} - Y_s' Z_s' \mathbb{X}_{s,t}\Big| \le C\big(\|X\|_\alpha \|R^Z\|_{2\alpha} + \|\mathbb{X}\|_{2\alpha}\|Y'Z'\|_\alpha\big)|t-s|^{3\alpha} \;. \tag{4.23}$$

If $Z = X$ and Z' is the identity operator, then this coincides with the definition (4.19). Furthermore, in the smooth case, one can check that we again recover the usual Riemann / Young integral.

Remark 4.12. If, in the notation of the proof of Theorem 4.4, Ξ and $\tilde{\Xi}$ are such that $\Xi - \tilde{\Xi} \in \mathcal{C}_2^\beta$ for some $\beta > 1$, i.e.

$$|\Xi_{s,t} - \tilde{\Xi}_{s,t}| = \mathrm{O}(|t-s|^\beta) \;,$$

then $\mathcal{I}\Xi = \mathcal{I}\tilde{\Xi}$. Indeed, it is immediate that

$$\sum_{[u,v]\in\mathcal{P}} |\Xi_{u,v} - \tilde{\Xi}_{u,v}| = \mathrm{O}(|\mathcal{P}|^{\beta-1}) \;,$$

which converges to 0 as $|\mathcal{P}| \to 0$. (This remains true if $\mathrm{O}(|t-s|^\beta)$ with $\beta > 1$ is replaced by $\mathrm{o}(|t-s|)$.)

This also shows that, if X and Y are smooth functions and $\mathbb{X}$ is defined by (2.2), the integral that we just defined does coincide with the usual Riemann–Stieltjes integral. However, if we change $\mathbb{X}$, then the resulting integral *does* change, as will be seen in the next example.

Example 4.13. Let f be a 2α-Hölder continuous function and let $\mathbf{X} = (X, \mathbb{X})$ and $\bar{\mathbf{X}} = (\bar{X}, \bar{\mathbb{X}})$ be two rough paths such that

$$\bar{X}_t = X_t \;, \qquad \bar{\mathbb{X}}_{s,t} = \mathbb{X}_{s,t} + f(t) - f(s) \;.$$

Let furthermore $(Y, Y') \in \mathscr{D}_X^{2\alpha}$ as above. Then also $(\bar{Y}, \bar{Y}') := (Y, Y') \in \mathscr{D}_{\bar{X}}^{2\alpha}$. However, it follows immediately from (4.19) that

$$\int_s^t \bar{Y}_r\, d\bar{\mathbf{X}}_r = \int_s^t Y_r\, d\mathbf{X}_r + \int_s^t Y_r'\, df(r) \;. \tag{4.24}$$

Here, the second term on the right hand side is a simple Young integral, which is well-defined since $\alpha + 2\alpha > 1$ by assumption.

Remark 4.14. As we will see below, (4.24) can be interpreted as a generalisation of the usual expression relating Itô integrals to Stratonovich integrals.

Remark 4.15. The bound (4.20) does behave in a very natural way under dilatations. Indeed, the integral is invariant under the transformation

$$(Y, Y', X, \mathbb{X}) \mapsto (\lambda^{-1} Y, \lambda^{-2} Y', \lambda X, \lambda^2 \mathbb{X}) \;. \tag{4.25}$$

The same is true for the right hand side of (4.20), since under this dilatation, we also have $R^Y \mapsto \lambda^{-1} R^Y$.

4.4 Stability I: rough integration

Consider $\mathbf{X} = (X, \mathbb{X})$, $\tilde{\mathbf{X}} = (\tilde{X}, \tilde{\mathbb{X}}) \in \mathscr{C}^\alpha$ with $(Y, Y') \in \mathscr{D}_X^{2\alpha}$, $(\tilde{Y}, \tilde{Y}') \in \mathscr{D}_{\tilde{X}}^{2\alpha}$. Although (Y, Y') and $(\tilde{Y}, \tilde{Y}')$ live, in general, in different Banach spaces, the "distance"

$$d_{X,\tilde{X},2\alpha}\Big(Y, Y'; \tilde{Y}, \tilde{Y}'\Big) \stackrel{\text{def}}{=} \big\|Y' - \tilde{Y}'\big\|_\alpha + \big\|R^Y - R^{\tilde{Y}}\big\|_{2\alpha}$$

will be useful. Even when $X = \tilde{X}$, it is not a proper metric for it fails to separate (Y, Y') and $(Y + cX + \bar{c}, Y' + c)$ for any two constants c and $\bar{c}$. When $X \neq \tilde{X}$, the assertion "zero distance implies $\big(Y, Y'\big) = \big(\tilde{Y}, \tilde{Y}'\big)$" does not even make sense. (The two objects live in completely different spaces!) That said, for every fixed $(X, \mathbb{X}) \in \mathscr{C}^\alpha$, one has (with $R^Y_{s,t} = Y_{s,t} - Y'_s X_{s,t}$ as usual), a canonical map

$$\iota_X : \big(Y, Y'\big) \in \mathcal{C}_X^\alpha \mapsto \big(Y', R^Y\big) \in \mathcal{C}^\alpha \oplus \mathcal{C}_2^{2\alpha} .$$

Given $Y_0 = \xi$, this map is injective since one can reconstruct Y by $Y_t = \xi + Y'_0 X_{0,t} + R^Y_{0,t}$. From this point of view, one simply has

$$d_{X,\tilde{X},2\alpha} = \|\iota_X(.) - \iota_{\tilde{X}}(.)\|_{\alpha,2\alpha} ,$$

and one is back in a normal Banach setting, where $\|\cdot,\cdot\|_{\alpha,2\alpha} = \|\cdot\|_\alpha + \|\cdot\|_{2\alpha}$ is a natural semi-norm on $\mathcal{C}^\alpha \oplus \mathcal{C}_2^{2\alpha}$. (In fact, it is a norm if one only considers elements in $\mathcal{C}^\alpha$ started at 0.) Elementary estimates of the form

$$\big|ab - \tilde{a}\tilde{b}\big| \le \big|a\big|\,\big|b - \tilde{b}\big| + \big|a - \tilde{a}\big|\,\big|\tilde{b}\big| \tag{4.26}$$

then lead to

$$\begin{aligned}\big|Y_{s,t} - \tilde{Y}_{s,t}\big| &= \Big|\big(Y'_{0,s} + Y'_0\big)X_{s,t} + \big(\tilde{Y}_{0,s} + \tilde{Y}_0\big)\tilde{X}_{s,t} + R^Y_{s,t} - R^{\tilde{Y}}_{s,t}\Big| \\ &\le C|t-s|^\alpha\Big(\big|Y'_0 - \tilde{Y}'_0\big| + \big\|Y' - \tilde{Y}'\big\|_\alpha + \big\|X - \tilde{X}\big\|_\alpha + \big\|R^Y - R^{\tilde{Y}}\big\|_{2\alpha}\Big) ,\end{aligned}$$

with a constant $C = C(R, T)$, provided $|Y'_0|$, $\|X\|_\alpha$, $\|Y'\|_\alpha$ and similarly for the same quantities with tilde, all have their norms bounded by R. (As usual, C can be taken uniform in $T \le 1$ since in this case $\|\cdot\|_{\alpha;[0,T]} \le \|\cdot\|_{2\alpha;[0,T]}$.) It follows that

$$\big\|Y - \tilde{Y}\big\|_\alpha \le C\Big(\big\|X - \tilde{X}\big\|_\alpha + \big|Y'_0 - \tilde{Y}'_0\big| + d_{X,\tilde{X},2\alpha}\big(Y, Y'; \tilde{Y}, \tilde{Y}'\big)\Big) . \tag{4.27}$$

An estimate of the proper α-Hölder norm of $Y - \tilde{Y}$ (rather than its semi-norm) is obtained by adding $\big|Y_0 - \tilde{Y}_0\big|$ to both sides.

Theorem 4.16 (Stability of rough integration). *Consider* $\mathbf{X} = \big(X, \mathbb{X}\big)$, $\tilde{\mathbf{X}} = \big(\tilde{X}, \tilde{\mathbb{X}}\big) \in \mathscr{C}^\alpha$, $\big(Y, Y'\big) \in \mathscr{D}_X^{2\alpha}$, $\big(\tilde{Y}, \tilde{Y}'\big) \in \mathscr{D}_{\tilde{X}}^{2\alpha}$ *in a bounded set, in the sense*

$$|Y'_0| + \|Y, Y'\|_{X,2\alpha} \le M, \quad \varrho_\alpha(0, \mathbf{X}) \equiv ||X||_\alpha + \|\mathbb{X}\|_{2\alpha} \le M,$$

with identical bounds for $\big(\tilde{X},\tilde{\mathbb{X}}\big)$, $\big(\tilde{Y},\tilde{Y}'\big)$, *for some* $M<\infty$. *Define*

$$(Z,Z') := \left(\int_0^{\cdot} Y\,d\mathbf{X}, Y\right) \in \mathscr{D}_X^{2\alpha}\,,$$

and similarly for $\big(\tilde{Z},\tilde{Z}'\big)$. *Then the following (local) Lipschitz estimates holds true,*

$$d_{X,\tilde{X},2\alpha}\big(Z,Z';\tilde{Z},\tilde{Z}'\big) \le C_M\Big(\varrho_\alpha\big(\mathbf{X},\tilde{\mathbf{X}}\big) + \big|Y_0'-\tilde{Y}_0'\big| + d_{X,\tilde{X},2\alpha}\big(Y,Y';\tilde{Y},\tilde{Y}'\big)\Big), \tag{4.28}$$

and also

$$\big\|Z-\tilde{Z}\big\|_\alpha \le C_M\Big(\varrho_\alpha\big(\mathbf{X},\tilde{\mathbf{X}}\big) + \big|Y_0-\tilde{Y}_0\big| + \big|Y_0'-\tilde{Y}_0'\big| + d_{X,\tilde{X},2\alpha}\big(Y,Y';\tilde{Y},\tilde{Y}'\big)\Big)\,, \tag{4.29}$$

where $C_M = C(M,T,\alpha)$ *is a suitable constant.*

Proof. (The reader is advised to review the proofs of Theorems 4.4, 4.10.) We first note that (4.27) applied to $Z,\tilde{Z}$ (note: $Z_0'-\tilde{Z}_0 = Y_0-\tilde{Y}$) shows that (4.29) is an immediate consequence of the first estimate (4.28). Thus, we only need to discuss the first estimate. By definition of $d_{X,\tilde{X},2\alpha}$, we need to estimate

$$\big\|Z'-\tilde{Z}'\big\|_\alpha + \|R^Z - R^{\tilde{Z}}\|_{2\alpha} = \big\|Y-\tilde{Y}\big\|_\alpha + \big\|R^Z - R^{\tilde{Z}}\big\|_{2\alpha}.$$

Thanks to (4.27), the first summand is clearly bounded by the right-hand side of (4.28). For the second summand we recall

$$R^Z_{s,t} = Z_{s,t} - Z_s'X_{s,t} = \int_s^t Y\,d\mathbf{X} - Y_sX_{s,t} = (\mathcal{I}\Xi)_{s,t} - \Xi_{s,t} + Y_s'\mathbb{X}_{s,t}$$

where $\Xi_{s,t} = Y_sX_{s,t} + Y_s'\mathbb{X}_{s,t}$ and similar for $R^{\tilde{Z}}$. Setting $\Delta = \Xi - \tilde{\Xi}$, we use (4.11) with $\beta = 3\alpha$ and Ξ replaced by Δ, so that

$$\begin{aligned}\big|R^Z_{s,t} - R^{\tilde{Z}}_{s,t}\big| &= \big|\big(\mathcal{I}\Delta\big)_{s,t} - \Delta_{s,t}\big| + \big|Y_s'\mathbb{X}_{s,t} - \tilde{Y}_s'\tilde{\mathbb{X}}_{s,t}\big| \\ &\le C\|\delta\Delta\|_{3\alpha}|t-s|^{3\alpha} + \big|Y_s'\mathbb{X}_{s,t} - \tilde{Y}_s'\tilde{\mathbb{X}}_{s,t}\big|\,,\end{aligned}$$

where $\delta\Delta_{s,u,t} = R^{\tilde{Y}}_{s,u}\tilde{X}_{u,t} - R^Y_{s,u}X_{u,t} + \tilde{Y}'_{s,u}\tilde{\mathbb{X}}_{u,t} - Y'_{s,u}\mathbb{X}_{u,t}$. We then conclude with some elementary estimates of the type (4.26), noting that all involved quantities stay bounded. $\square$

4.5 Controlled rough paths of lower regularity

Recall that we showed in Section 2.3 how an α-Hölder rough path $\mathbf{X}$ could be defined as a path with values in the p-step nilpotent Lie group $G^{(p)}(\mathbf{R}^d) \subset T^{(p)}(\mathbf{R}^d)$, with $p = \lfloor 1/\alpha \rfloor$. It does not seem obvious at all a priori how one would define a controlled rough path in this context. One way of interpreting Definition 4.6 is as a kind of local “Taylor expansion” up to order 2α. It seems natural in the light of the previous subsections that if $\alpha < \frac{1}{3}$, a controlled rough path should have a kind of “Taylor expansion” up to order $p\alpha$.

As a consequence, if we expand $\mathbf{X}_{s,t} \stackrel{\text{def}}{=} \mathbf{X}_s^{-1} \otimes \mathbf{X}_t$ as

$$\mathbf{X}_{s,t} = \sum_{|w| \le p} \mathbf{X}_{s,t}^w \, e_w \;,$$

where $|w|$ denotes the length of the word w, one would expect that a controlled rough path should have an expansion of the form

$$\delta Y_{s,t} = \sum_{|w| \le p-1} Y_s^w \, \mathbf{X}_{s,t}^w + R_{s,t}^Y \;, \tag{4.30}$$

with $|R_{s,t}^Y| \lesssim |t-s|^{p\alpha}$. Recall however that in Definition 4.6 we also needed a regularity condition on the “derivative process” Y'. The equivalent statement in the present context is that the Y_s^w should themselves be described by a local “Taylor expansion”, but this time only up to order $(p-|w|)\alpha$. A neat way of packaging this into a compact statement is to view Y as a $T^{(p-1)}(\mathbf{R}^d)$-valued function and to introduce a scalar product on $T^{(p)}(\mathbf{R}^d)$ by postulating that $\langle e_w, e_{\bar w}\rangle = \delta_{w,\bar w}$ for any two words w and $\bar w$. One then has the following extension of Definition 4.6 (see Exercise 4.26).

Definition 4.17. A controlled rough path is a $T^{(p-1)}(\mathbf{R}^d)$-valued function Y such that, for every word w with $|w| \le p-1$, one has the bound

$$\bigl|\langle e_w, Y_t\rangle - \langle \mathbf{X}_{s,t} \otimes e_w, Y_s\rangle\bigr| \le C|t-s|^{(p-|w|)\alpha} \;. \tag{4.31}$$

Given such a controlled rough path Y, it is then natural to define its integral against any component X^i by

$$Z_t = \int_0^t Y_s \, dX_s^i \stackrel{\text{def}}{=} \lim_{|\mathcal{P}|\to 0} \sum_{[r,s]\in\mathcal{P}} \sum_{|w|\le p-1} Y_r^w \, \langle e_w \otimes e_i, \mathbf{X}_{r,s}\rangle \;,$$

where e_i is the unit vector associated to the word consisting of the single letter i. It turns out [Gub10] that Z is again a controlled rough path in the sense of Definition 4.17 provided that we lift it to $T^{(p-1)}(\mathbf{R}^d)$ by imposing that

$$\langle e_w \otimes e_i, Z_t\rangle \stackrel{\text{def}}{=} Y_r^w \;,$$

and by setting $Z_t^w = 0$ for all non-empty words that do not terminate with the letter i.

4.6 Exercises

Exercise 4.18. a) Deduce (4.3) from (4.2).

b) Show that there is a constant C depending only on $T > 0$ and $\alpha + \beta > 1$ such that

$$\Big\| \int_0^{\cdot} Y dX \Big\|_{\alpha;[0,T]} \le C\Big(|Y_0| + \|Y\|_{\beta;[0,T]}\Big)\|X\|_{\alpha;[0,T]}. \tag{4.32}$$

In fact, show that C can be chosen uniformly over $T \in (0, 1]$.

Solution 4.19. a) Given X on $[s,t]$, define $\tilde{X} : [0,1] \ni u \mapsto X(s + u(t-s))$ and verify $\|\tilde{X}\|_{\alpha;[0,1]} = |t-s|^{\beta}\|X\|_{\beta;[s,t]}$. Proceeding similarly for Y, applying (4.2) to $\tilde{X}$, $\tilde{Y}$ then gives (4.3).

b) Write Z for the indefinite integral. From (4.3), for every $0 \le s < t \le T$,

$$\begin{aligned}
|Z_{s,t}| &\le |Y_s||X_{s,t}| + C\|Y\|_{\beta;[s,t]}\|X\|_{\alpha;[s,t]}|t-s|^{\alpha+\beta} \\
&\le \Big(|Y_0| + \|Y\|_{\beta;[0,T]}T^{\beta}\Big)|X_{s,t}| + C\|Y\|_{\beta;[0,T]}\|X\|_{\alpha;[0,T]}T^{\beta}|t-s|^{\alpha} \\
&\le \Big[|Y_0| + \|Y\|_{\beta;[0,T]}T^{\beta}(1+C)\Big]\|X\|_{\alpha;[0,T]}|t-s|^{\alpha}. \\
&\le (1 \vee T)^{\beta}(1+C)\Big[|Y_0| + \|Y\|_{\beta;[0,T]}\Big]\|X\|_{\alpha;[0,T]}|t-s|^{\alpha}\ ,
\end{aligned}$$

and this entails the claimed estimates.

Exercise 4.20. Let $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^{\alpha}([0,T], V)$ and assume that $F : V \to \mathcal{L}(V, W)$ is of gradient form, i.e. $F = DG$ where $G : V \to W$ is sufficiently smooth, say $\mathcal{C}_b^3$. Show that the relation

$$\int_s^t F(X) d\mathbf{X} = G(X_t) - G(X_s)\ ,$$

holds true whenever $\mathbf{X}$ is a geometric rough path. (Hence, from a rough path perspective, integration of gradient 1-forms against geometric rough paths is trivial for the outcome does not depend on $\mathbb{X}$.) What about non-geometric rough paths?

Exercise 4.21. Complete the “first argument” in the proof of Theorem 4.4.

Solution 4.22. Let $\mathcal{P}_n$ by the dyadic partitions of $[0,T]$, so that $\#\mathcal{P}_n = 2^n$ and mesh $|\mathcal{P}_n| = T/2^n$. Call elements of $\mathcal{P}_n$ dyadic intervals (of level-n). Given an interval $[s,t] \subset [0,T]$ there exists $m \ge 0$, such that P^m is the coarsest dyadic partition which contains a dyadic interval $\subset [s,t]$. Note that $|t-s| \sim T/2^m$. We then define, for $n \ge m$, and a general interval $[s,t]$,

$$I^n_{s,t} := \sum_{\substack{[u,v] \in \mathcal{P}_n : \\ [u,v] \subset [s,t]}} \Xi_{u,v}.$$

Note $I^n_{s,t} = \Xi_{s,t}$ if $[s,t] \in \mathcal{P}_n$. Write $s^{(n)}_+$ (resp. $t^{(n)}_-$) for the closest right (resp. left) level-n dyadic neighbour of s (resp. t) so that

$$s \le s^{(n)}_+ < t^{(n)}_- \le t.$$

Note that if s is a level-m dyadic (i.e. $s = kT/2^m$ for some integer k) then $s^{(n)}_+ = s$ for all $n \ge m$, and similar for t. We have

$$\begin{aligned}\left|I^{n+1}_{s,t} - I^n_{s,t}\right| &\le \sum_{\substack{[u,v]\in\mathcal{P}_n:\\ [u,v]\subset[s,t]}} \left|\delta\Xi_{u,\frac{u+v}{2},v}\right| + \left|\Xi_{s^{(n+1)},s^{(n)}}\right| + \left|\Xi_{t^{(n)},t^{(n+1)}}\right| \\ &\lesssim \frac{|t^{(n)} - s^{(n)}|}{2^n}\left(\frac{1}{2^n}\right)^\beta + 2^{-(n+1)\alpha} + 2^{-(n+1)\alpha} .\end{aligned}$$

Plainly, these estimates imply, for general $[s,t] \subset [0,T]$, that $\left(I^n_{s,t} : n\right)$ is Cauchy and we call the limit $I_{s,t}$. In fact, I is additive in the sense that $\delta I \equiv 0$. Indeed, for general $s < u < t$ in $[0,T]$, if u_- (resp. u_+) denotes the closest left (resp. right) level-n dyadic neighbour, then

$$I^n_{s,t} = I^n_{s,u} + I^n_{u,v} + \Xi_{u_-,u_+} ,$$

and since $\left|\Xi_{u_-,u_+}\right| \lesssim |u_+ - u_-|^\alpha \sim (1/2^n)^\alpha$, additivity of the limit $I = \lim I^n$ follows at once. Another immediate consequence of the above estimates, if applied to a *dyadic* interval $[s,t]$, is the estimate

$$|I_{s,t} - \Xi_{s,t}| \lesssim |t-s|^\beta . \tag{4.33}$$

Indeed, in this case $m \ge 0$ is determined from $|t-s| = T/2^m$ so that $[s,t] \in \mathcal{P}_m$ and since $I^m_{s,t} = \Xi_{s,t}$ we have

$$|I_{s,t} - \Xi_{s,t}| = \left|I_{s,t} - I^m_{s,t}\right| \lesssim |t-s| \sum_{n\ge m} 2^{n(1-\beta)} \sim |t-s| 2^{m(1-\beta)} \sim |t-s|^\beta.$$

We claim that the estimate (4.33) is valid for all intervals $[s,t] \subset [0,T]$. By continuity, it will be enough to consider $s < t$ in $\cup_n \mathcal{P}_n$. As in the proof of the Kolmogorov criterion, Theorem 3.1, we consider a (finite) partition $P = (\tau_i)$ of $[s,t]$, which "efficiently" exhausts $[s,t]$ with dyadic intervals of length $\sim 2^{-n}, n \ge m$, in the sense that no three intervals have the same length. Note that $|P| \equiv \max\{|v-u| : [v,u] \in P\} = 2^{-m} \le |t-s|$ (and in fact $\sim |t-s|$ due to minimal choice of m). Thanks to additiviy of I and (4.33) for dyadic intervals,

$$\begin{aligned}|I_{s,t} - \Xi_{s,t}| &= \Big|\sum_{[u,v]\in P} (I_{u,v} - \Xi_{u,v}) - \Big(\Xi_{s,t} - \sum_{[u,v]\in P} \Xi_{u,v}\Big)\Big| \\ &\lesssim \sum_{[u,v]\in P} |v-u|^\beta + \Big(\Xi_{s,t} - \sum_{[u,v]\in P} \Xi_{u,v}\Big).\end{aligned}$$

$$\leq |t-s|^{\beta} + \sum_{i=0}^{\infty} \left|\delta\Xi_{s,\tau_{-(i+1)},\tau_{-i}} + \delta\Xi_{\tau_i,\tau_{i+1},t}\right| ,$$

where the sum is actually finite. Possibly allowing equality (“$\tau_i = \tau_{i+1}$” for some i), we may assume $|\tau_{i+1} - \tau_i| = \left|\tau_{-i} - \tau_{-(i+1)}\right| \lesssim 1/2^{m+i}$, so that

$$|t - \tau_i| = \sum_{j=i}^{\infty} |\tau_{j+1} - \tau_j| \lesssim \sum_{j=i}^{\infty} 1/2^{m+j} \sim 1/2^{m+i} ,$$

and similarly, $|\tau_{-i} - s| \lesssim 1/2^{m+i}$. As a consequence, one obtains

$$\sum_{i=0}^{\infty} \left|\delta\Xi_{s,\tau_{-(i+1)},\tau_{-i}} + \delta\Xi_{\tau_i,\tau_{i+1},t}\right| \lesssim 2 \sum_{n\geq m} (1/2^n)^{\beta} \sim 1/2^{m\beta} \sim |t-s|^{\beta} ,$$

so that $|I_{s,t} - \Xi_{s,t}| \lesssim |t-s|^{\beta}$, as required

Exercise 4.23. Adapt the proof of Theorem 4.4 such as to obtain Young’s estimate (4.3).

Exercise 4.24. Fix $\alpha \in (0,1], h > 0$ and $M > 0$. Consider a path $Z : [0,T] \to V$ and show that

$$\|Z\|_{\alpha;h} \equiv \sup_{\substack{0\leq s<t\leq T \\ t-s\leq h}} \frac{|Z_{s,t}|}{|t-s|^{\alpha}} \leq M \implies \|Z\|_{\alpha} \leq M\Big(1 \vee 2h^{-(1-\alpha)}\Big).$$

(Here, as usual, $\|Z\|_{\alpha} \equiv \sup_{0\leq s<t\leq T} |Z_{s,t}|/|t-s|^{\alpha}$.)

Proof. By scaling it suffices to consider $M = 1$. Fix $0 \leq s < t \leq T$, we need to show $|Z_{s,t}|/|t-s|^{\alpha}$ is bounded by $1 \vee 2h^{1/\alpha-1}$. There is nothing to show for $|t-s| \leq h$. We therefore assume $h \leq |t-s|$ and define $t_i = (s+ih) \wedge t$, for $i = 0, 1, \ldots$ noting that $t_N = t$ for $N \geq |t-s|/h$ and also $t_{i+1} - t_i \leq h$ for all i. But then

$$\begin{aligned} |Z_{s,t}| &\leq \sum_{0\leq i<|t-s|/h} \left|Z_{t_i,t_{i+1}}\right| \\ &\leq h^{\alpha}(1 + |t-s|/h) = h^{\alpha-1}(h + |t-s|) \leq 2h^{\alpha-1}|t-s|. \end{aligned}$$

and we are done. □

Exercise 4.25. Show the assumption on $Y \in \mathscr{D}_X^{2\alpha}$ can be weakend to $Y \in \mathscr{D}_X^{2\alpha'}$, $\alpha' < \alpha$, provided $\alpha + 2\alpha' > 1$, and reformulate Theorem 4.10 accordingly. In particular, show that the estimate (4.20) holds upon replacing the final factor $|t-s|^{3\alpha}$ by $|t-s|^{\alpha+2\alpha'}$, and $\|Y'\|_{\alpha}$ (resp. $\|R^Y\|_{2\alpha}$) by $\|Y'\|_{\alpha'}$ (resp. $\|R^Y\|_{2\alpha'}$).

Exercise 4.26. Check that this definition 4.17 is consistent with definition 4.6 in the case when $\alpha \in \left(\frac{1}{3}, \frac{1}{2}\right]$. Check also that if one takes $w = \phi$, the empty word, then (4.31) reduces to (4.30) with $|R^Y_{s,t}| \lesssim |t-s|^{p\alpha}$.

4.7 Comments

The notion of integration of 1-forms against "general" p-variation geometric rough paths, for any $p \in [1, \infty)$, was developed by Lyons [Lyo98]; see also [Lyo98, LQ02, LCL07]. A general estimate of the form (4.14) appears in [FV10b, Thm 10.47], at least in the finite-dimensional setting of that book. Rough integration against "controlled paths" is due to Gubinelli, see [Gub04] where it is developed in a α-Hölder setting, $\alpha > \frac{1}{3}$. Loosely speaking, it allows to "linearise" many considerations (the space of controlled paths is a Banach space, while a typical space of rough paths is not). This point of view has been generalized to arbitrary α (both in the geometric and the non-geometric setting) in [Gub10].

We will see in Chapter 13 that this point of view can be pushed even further and, as a matter of fact, the theory of regularity structures provides a unified framework in which the Gubinelli derivative and the regular derivatives are but two examples of a more general theory of objects behaving "like Taylor expansions" and allowing to describe the small-scale structure of a function and / or distribution in terms of "known" objects (polynomials in the case of Taylor expansions, the underlying rough path in the case of controlled paths).

Chapter 5
Stochastic integration and Itô's formula

Abstract In this chapter, we compare the integration theory developed in the previous chapter to the usual theories of stochastic integration, be it in the Itô or the Stratonovich sense.

5.1 Itô integration

Recall from Section 3 that Brownian motion B can be enhanced to a (random) rough path $\mathbf{B} = (B, \mathbb{B})$. Presently our focus is the case when $\mathbb{B}$ is given by the iterated Itô integral [1]

$$\mathbb{B}_{s,t} = \mathbb{B}^{\text{Itô}}_{s,t} \stackrel{\text{def}}{=} \int_s^t B_{s,u} \otimes dB_u$$

and the so enhanced Brownian motion has almost surely (non-geometric) α-Hölder rough sample paths, for any $\alpha \in \left(\frac{1}{3}, \frac{1}{2}\right)$. That is, $\mathbf{B}(\omega) = (B(\omega), \mathbb{B}(\omega)) \in \mathscr{C}^\alpha$ for every $\omega \in N_1^c$ where, here and in the sequel, $N_i, i = 1, 2, ...$ denote suitable null sets. We now show that rough integrals (against $\mathbf{B} = \mathbf{B}^{\text{Itô}}$) and Itô integrals, whenever both are well-defined, coincide.

Proposition 5.1. *Assume $(Y(\omega), Y'(\omega)) \in \mathscr{D}^{2\alpha}_{B(\omega)}$ for every $\omega \in N_2^c$. Set $N_3 = N_1 \cup N_2$. Then the rough integral*

$$\int_0^T Y d\mathbf{B} = \lim_{n\to\infty} \sum_{[u,v]\in\mathcal{P}_n} (Y_u B_{u,v} + Y'_u \mathbb{B}_{u,v})$$

exists, for each fixed $\omega \in N_3^c$, along any sequence $(\mathcal{P}_n)$ with mesh $|\mathcal{P}_n| \downarrow 0$. If Y, Y' are adapted then, almost surely,

$$\int_0^T Y d\mathbf{B} = \int_0^T Y dB \; .$$

[1] The case when $\mathbb{B}$ is given via iterated Stratonovich integration is left to Section 5.2 below.

Proof. Without loss of generality $T = 1$. The existence of the rough integral for $\omega \in N_3^c$ under the stated assumptions is immediate from Theorem 4.10, applied to $Y(\omega)$, controlled by $B(\omega)$, for $\omega \in N_2^c$ fixed. Recall (e.g. [RY91]) that for any continuous, adapted process Y the Itô integral against Brownian motion has the representation

$$\int_0^1 Y dB = \lim_{n\to\infty} \sum_{[u,v]\in\mathcal{P}_n} Y_u B_{u,v} \quad \text{(in probability)}$$

along any sequence $(\mathcal{P}_n)$ with mesh $|\mathcal{P}_n| \downarrow 0$. By switching to a subsequence, if necessary, we can assume that the convergence holds almost surely, say on N_4^c. Set $N_5 := N_3 \cup N_4$. We shall complete the proof under the assumption that there exists a (deterministic) constant $M > 0$ such that

$$\sup_{\omega\in N_5^c} |Y'(\omega)|_\infty \le M \;.$$

(This is the case in the "model" situation $Y = F(X), Y' = DF(X)$ where F was in particular assumed to have bounded derivatives; the general case is obtained by localisation and left to Exercise 5.13.)

The claim is that the rough and Itô integral coincide on N_5^c. With a look at the respective Riemann-sums, convergent away from N_5, basic analysis tells us that

$$\forall\omega \in N_5^c :\ \exists \lim_n \sum_{[u,v]\in\mathcal{P}_n} Y'_u \mathbb{B}_{u,v} \;,$$

and that this limit equals the difference of rough and Itô integrals (on N_5^c, a set of full measure). Of course, $|\mathcal{P}_n| \downarrow 0$, and to see that the above limit is indeed zero (at least on a set of full measure), it will be enough to show that

$$\Bigg\| \sum_{[u,v]\in\mathcal{P}} Y'_u \mathbb{B}_{u,v} \Bigg\|_{L^2}^2 = \mathrm{O}(|\mathcal{P}|) \;. \tag{5.1}$$

To this end, assume the partition is of the form $\mathcal{P} = \{0 = \tau_0 < \cdots < \tau_N = 1\}$ and define a (discrete-time) martingale started at $S_0 := 0$ with increments $S_{k+1} - S_k = Y'_{\tau_k} \mathbb{B}_{\tau_k,\tau_{k+1}}$. Since $|\mathbb{B}_{\tau_k,\tau_{k+1}}|_{L^2}^2$ is proportional to $|\tau_{k+1} - \tau_k|^2$, as may be seen from Brownian scaling, we then have

$$\begin{aligned} \Bigg| \sum_{[u,v]\in\mathcal{P}} Y'_u \mathbb{B}_{u,v} \Bigg|_{L^2}^2 &= \Bigg| \sum_{k=0}^{N-1} (S_{k+1} - S_k) \Bigg|_{L^2}^2 = \sum_{k=0}^{N-1} |S_{k+1} - S_k|_{L^2}^2 \\ &\le M^2 \sum_{k=0}^{N-1} \big|\mathbb{B}_{\tau_k,\tau_{k+1}}\big|_{L^2}^2 = \mathrm{O}(|\mathcal{P}|) \;, \end{aligned}$$

as desired. $\square$

5.2 Stratonovich integration

We could equally well have enhanced Brownian motion with

$$\mathbb{B}^{\text{Strat}}_{s,t} := \int_s^t B_{s,u} \otimes \circ dB_u = \mathbb{B}^{\text{Itô}}_{s,t} + \frac{1}{2}(t-s)I \ .$$

Almost surely, this construction then yields *geometric* α-Hölder rough sample paths, for any $\alpha \in \left(\frac{1}{3}, \frac{1}{2}\right)$. Recall that, by definition, the Stratonovich integral is given by

$$\int_0^T Y \circ dB \overset{\text{def}}{=} \int_0^T Y dB + \frac{1}{2}[Y, B]_T$$

whenever the Itô integral is well-defined and the *quadratic covariation* of Y and B exists in the sense that $[Y, B]_T := \lim_{|\mathcal{P}|\to 0} \sum_{[u,v]\in\mathcal{P}} Y_{u,v} B_{u,v}$ exists as limit in probability.

In complete analogy to the Itô case, we now show that rough integration against *Stratonovich* enhanced Brownian motion coincides with usual Stratonovich integration against Brownian motion under some natural assumptions guaranteeing that both notions of integral are well-defined.

Corollary 5.2. *As above, assume $Y = Y(\omega) \in \mathcal{C}^{\alpha}_{B(\omega)}$ for every $\omega \in N_2^c$. Set $N_3 = N_1 \cup N_2$. Then the rough integral of Y against $\mathbf{B} = \mathbf{B}^{\text{Strat}}$ exists,*

$$\int_0^T Y d\mathbf{B} = \lim_{n\to\infty} \sum_{[u,v]\in\mathcal{P}_n} \left(Y_u B_{u,v} + Y'_u \mathbb{B}^{\text{Strat}}_{u,v}\right).$$

Moreover, if Y, Y' are adapted, the quadratic covariation of Y and B exists and, almost surely,

$$\int_0^T Y d\mathbf{B} = \int_0^T Y \circ dB.$$

Proof. $\mathbb{B}^{\text{Strat}}_{s,t} = \mathbb{B}^{\text{Itô}}_{s,t} + f_{s,t}$ where $f(t) = (1/2)t\, I$, where $I \in \mathbf{R}^d \otimes \mathbf{R}^d$ denotes the identity matrix. This entails, as was discussed in Example 4.13,

$$\int_0^1 Y d\mathbf{B}^{\text{Strat}} = \int_0^1 Y d\mathbf{B}^{\text{Itô}} + \int_0^1 Y' df.$$

Thanks to Proposition 5.1, it only remains to identify $2\int_0^1 Y' df = \int_0^1 Y'_t\, dt$ with $[Y, B]_1$. To see this, write

$$\begin{aligned}\sum_{[u,v]\in\mathcal{P}} Y_{u,v} B_{u,v} &= \sum_{[u,v]\in\mathcal{P}} \left(\left(Y'_{u,v} B_{u,v}\right) B_{u,v} + R_{u,v} B_{u,v}\right) \\ &= \left(\sum_{[u,v]\in\mathcal{P}} Y'_{u,v}(B_{u,v} \otimes B_{u,v})\right) + \mathrm{O}\left(|\mathcal{P}|^{3\alpha-1}\right) ,\end{aligned}$$

where we used that $\sum R_{u,v} B_{u,v} = \mathrm{O}\big(|\mathcal{P}|^{3\alpha-1}\big)$ thanks to $R \in \mathcal{C}_2^{2\alpha}$ and $B \in \mathcal{C}^{\alpha}$. Note that

$$B_{u,v} \otimes B_{u,v} = 2 \ \mathrm{Sym}\left(\mathbb{B}^{\mathrm{Strat}}_{u,v}\right) = 2 \ \mathrm{Sym}\left(\mathbb{B}^{\mathrm{Itô}}_{u,v}\right) + (v-u)I.$$

We have seen in the proof of Proposition 5.1 that any limit (in probability, say) of

$$\sum_{[u,v]\in\mathcal{P}} Y'_{u,v} \mathbb{B}^{\mathrm{Itô}}_{u,v}$$

must be zero. In fact, a look at the argument reveals that this remains true with $\mathbb{B}^{\mathrm{Itô}}_{u,v}$ replaced by $\mathrm{Sym}\left(\mathbb{B}^{\mathrm{Itô}}_{u,v}\right)$. It follows that

$$\lim_{|\mathcal{P}|\to 0} \sum_{[u,v]\in\mathcal{P}} Y_{u,v} B_{u,v} = \lim_{|\mathcal{P}|\to 0} \Bigg(\sum_{[u,v]\in\mathcal{P}} Y'_{u,v}\,(v-u) \Bigg) = \int_0^1 Y'_t \, dt \ ,$$

thus concluding the proof. $\square$

5.3 Itô's formula and Föllmer

Given a smooth path $X : [0,T] \to V$ and a map $F : V \to W$ in $\mathcal{C}^1_b$, where V, W are Banach spaces as usual, the chain rule from classical "first oder" calculus tells us that

$$F(X_t) = F(X_0) + \int_0^t DF(X_s) dX_s, \quad 0 \le t \le T.$$

Unsurprisingly, the same change of variables formula holds for *geometric* rough paths $\mathbf{X} = (X, \mathbb{X})$, which are essentially limits of smooth paths, and it is not hard to figure out, in view of Example 4.13, that a "second order" correction, involving D^2F, appears in the non-geometric case. In other words, one can write down Itô formulae for rough paths.

Before doing so, however, an important preliminary discussion is in order. Namely, much of our effort so far was devoted to the understanding of (rough) integration against 1-forms, say $G = G(X)$ and indeed we found

$$\int G(X) d\mathbf{X} \approx \sum_{[s,t]\in\mathcal{P}} \langle G(X_s), X_{s,t}\rangle + \langle DG(X_s), \mathbb{X}_{s,t}\rangle$$

in the sense that the *compensated* Riemann-Stieltjes sums appearing on the right-hand side converge with mesh $|\mathcal{P}| \to 0$. Let us split $\mathbb{X}$ into symmetric part, $\mathbb{S}_{s,t} := \mathrm{Sym}\,(\mathbb{X}_{s,t})$, and antisymmetric ("area") part, $\mathrm{Anti}\,(\mathbb{X}_{s,t}) := \mathbb{A}_{s,t}$. Then

$$\langle DG(X_s), \mathbb{X}_{s,t}\rangle = \langle DG(X_s), \mathbb{S}_{s,t}\rangle + \langle DG(X_s), \mathbb{A}_{s,t}\rangle$$

and the final term disappears in the gradient case, i.e. when $G = DF$. Indeed, the contraction of a symmetric tensor (here: D^2F) with an antisymmetric tensor (here: $\mathbb{A}$) always vanishes. In other words, area matters very much for general integrals of 1-forms but not at all for gradient 1-forms. Note also that, contrary to $\mathbb{A}$, the symmetric part $\mathbb{S}$ is a nice function of the underlying path X. For instance, for Itô enhanced Brownian motion in $\mathbf{R}^d$, one has the identity

$$\mathbb{S}^{i,j}_{s,t} = \int_s^t B^i_{s,r} dB^j_r = \frac{1}{2}\Big(B^i_{s,t}B^j_{s,t} - \delta^{ij}(t-s)\Big) , \qquad 1 \le i,j \le d .$$

These considerations suggest that the following definition encapsulates all the data required for the integration of gradient 1-forms.

Definition 5.3. We call $\mathbf{X} = (X, \mathbb{S})$ a *reduced rough path*, in symbols $\mathbf{X} \in \mathscr{C}^\alpha_r([0,T],V)$, if $X = X_t$ takes values in a Banach space V, $\mathbb{S} = \mathbb{S}_{s,t}$ takes values in Sym $(V \otimes V)$, and the following hold:

i) a "reduced" Chen relation

$$\mathbb{S}_{s,t} - \mathbb{S}_{s,u} - \mathbb{S}_{u,t} = \mathrm{Sym}\,(X_{s,u} \otimes X_{u,t}) , \qquad 0 \le s,t,u \le T ,$$

ii) the usual analytical conditions, $X_{s,t} = \mathrm{O}(|t-s|^\alpha)$, $\mathbb{S}_{s,t} = \mathrm{O}\big(|t-s|^{2\alpha}\big)$, for some $\alpha > 1/3$.

Clearly, any $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha([0,T],V)$ induces a reduced rough path by ignoring its area $\mathbb{A} = \mathrm{Anti}\,(\mathbb{X})$. More importantly, and in stark contrast to the general rough path case, the lift of a path $X \in \mathcal{C}^\alpha$ to a reduced path is essentially trivial by setting $\mathbb{S}_{s,t} = \frac{1}{2}X_{s,t} \otimes X_{s,t}$. Indeed, we have the following result.

Lemma 5.4. *Given $X \in \mathcal{C}^\alpha$, $\alpha \in (1/3, 1/2]$, the "geometric" choice $\bar{\mathbb{S}}_{s,t} = \frac{1}{2}X_{s,t} \otimes X_{s,t}$ yields a reduced rough path, i.e. $\big(X, \bar{\mathbb{S}}\big) \in \mathscr{C}^\alpha_r$. Moreover, for any 2α-Hölder path γ (with values in* Sym $(V \otimes V)$*, the perturbation*

$$\mathbb{S}_{s,t} = \bar{\mathbb{S}}_{s,t} + \frac{1}{2}(\gamma_t - \gamma_s) = \frac{1}{2}(X_{s,t} \otimes X_{s,t} + \gamma_{s,t})$$

also yields a reduced rough path $(X, \mathbb{S})$. Finally, all reduced rough paths over X are obtained in this fashion.

Proof. A simple exercise for the reader. □

The previous lemma gives in particular a one-one correspondence between $\mathbb{S}$ and γ. We thus formalize the role of γ.

Definition 5.5 (Bracket of a reduced rough path). Given $\mathbf{X} = (X, \mathbb{S}) \in \mathscr{C}^\alpha_r(V)$, we define the bracket

$$\begin{aligned} [\mathbf{X}] : [0,T] &\to \mathrm{Sym}\,(V \otimes V) \\ t &\mapsto [\mathbf{X}]_t \stackrel{\text{def}}{=} X_{0,t} \otimes X_{0,t} - 2\mathbb{S}_{0,t} \ . \end{aligned}$$

Note that, as consequence of the previous lemma, $[\mathbf{X}] \in \mathcal{C}^{2\alpha}$. Furthermore, if one defines

$$[\mathbf{X}]_{s,t} \stackrel{\text{def}}{=} X_{s,t} \otimes X_{s,t} - 2\mathbb{S}_{s,t}\;,$$

then one has the identity $[\mathbf{X}]_{s,t} = [\mathbf{X}]_{0,t} - [\mathbf{X}]_{0,s}$ for any two times s,t.

Proposition 5.6 (Itô formula for reduced rough paths). *Let $F : V \to W$ be of class $\mathcal{C}^3_b$ and let $\mathbf{X} = (X, \mathbb{S}) \in \mathscr{C}^\alpha_r([0,T], V)$ with $\alpha > 1/3$. Then*

$$F(X_t) = F(X_0) + \int_0^t DF(X_s)d\mathbf{X}_s + \frac{1}{2}\int_0^t D^2F(X_s)d[\mathbf{X}]_s, \quad 0 \le t \le T.$$

Here, writing $\mathcal{P}$ for partitions of $[0,t]$, the first integral is given by[2]

$$\int_0^t DF(X_s)d\mathbf{X}_s \stackrel{\text{def}}{=} \lim_{|\mathcal{P}|\to 0} \sum_{[u,v]\in\mathcal{P}} \left(\langle DF(X_u), X_{u,v}\rangle + \left\langle D^2F(X_u), \mathbb{S}_{u,v}\right\rangle\right)\;, \tag{5.2}$$

while the second integral is a well-defined Young integral.

Proof. Consider first the geometric case, $\mathbb{S} = \bar{\mathbb{S}}$, in which case the bracket is zero. The proof is straightforward. Indeed, thanks to α-Hölder regularity of X with $\alpha > 1/3$, we obtain

$$\begin{aligned}
F(X_T) - F(X_0) &= \sum_{[u,v]\in\mathcal{P}} F(X_v) - F(X_u) \\
&= \sum_{[u,v]\in\mathcal{P}} \langle DF(X_u), X_{u,v}\rangle + \frac{1}{2}\left\langle D^2F(X_u), X_{u,v}\otimes X_{u,v}\right\rangle \\
&\quad + \mathrm{o}(|v-u|) \\
&= \sum_{[u,v]\in\mathcal{P}} \langle DF(X_u), X_{u,v}\rangle + \left\langle D^2F(X_u), \bar{\mathbb{S}}_{u,v}\right\rangle + \mathrm{o}(|v-u|)\;.
\end{aligned}$$

We conclude by taking the limit $|\mathcal{P}| \to 0$, also noting that $\sum_{[u,v]\in\mathcal{P}} \mathrm{o}(|v-u|) \to 0$. For the non-geometric situation, just substitute

$$\bar{\mathbb{S}}_{u,v} = \mathbb{S}_{u,v} + \frac{1}{2}[\mathbf{X}]_{u,v}.$$

Since D^2F is Lipschitz, $D^2F(X_\cdot) \in \mathcal{C}^\alpha$ and we can split-up the "bracket" term and note that

$$\sum_{[u,v]\in\mathcal{P}} \left\langle D^2F(X_u), [\mathbf{X}]_{u,v}\right\rangle \to \int_0^t D^2F(X_u)d[\mathbf{X}]_u\;,$$

where the convergence to the Young integral follows from $[X] \in \mathcal{C}^{2\alpha}$. The rest is now obvious. $\square$

---

[2] Note consistency with the rough integral when $\mathbf{X} \in \mathscr{C}^\alpha$.

Example 5.7. Consider the case when $\mathbf{X} = \mathbf{B}$, Itô enhanced Brownian motion. Then $\mathbb{X}$ is given by the iterated Itô integrals, with twice its symmetric part given by

$$2\mathbb{S}^{i,j}_{0,t} = \int_0^t B^i dB^j + B^j dB^i = B^i_t B^j_t - \langle B^i, B^j \rangle_t.$$

The usual Itô formula is then recovered from the fact that

$$[\mathbf{B}]^{i,j}_t = B^i_{0,t} B^j_{0,t} - 2\mathbb{S}^{i,j}_{0,t} = \langle B^i, B^j \rangle_{0,t} = \delta^{i,j} t \ .$$

We conclude this section with a short discussion on Föllmer's *calcul d'Itô sans probabilités* [Föl81]. For simplicity of notation, we take $V = \mathbf{R}^d, W = \mathbf{R}^e$ in what follows. With regard to (5.2), let us insist that the compensation is necessary and one cannot, in general, separate the sum into two convergent sums. On the other hand, we can combine the converging sums and write

$$\begin{aligned} F(X)_{0,t} = \lim_{|\mathcal{P}|\to 0} \sum_{[u,v]\in\mathcal{P}} \Big(&\langle DF(X_u), X_{u,v}\rangle + \big\langle D^2F(X_u), \mathbb{S}_{u,v} \big\rangle \\ &+ \frac{1}{2} \sum_{[u,v]\in\mathcal{P}} D^2F(X_u)[\mathbf{X}]_{u,v} \Big) \\ = \lim_{|\mathcal{P}|\to 0} \sum_{[u,v]\in\mathcal{P}} \Big(&\langle DF(X_u), X_{u,v}\rangle + \frac{1}{2} \big\langle D^2F(X_u), X_{u,v}\otimes X_{u,v} \big\rangle \Big) \ . \end{aligned} \tag{5.3}$$

We now put forward an assumption that allows to break up the sum above.

Definition 5.8. Let $(\mathcal{P}_n)$ be a sequence of partitions of $[0,T]$ with mesh $|\mathcal{P}_n| \to 0$. We say that $X : [0,T] \to \mathbf{R}^d$ has finite quadratic variation in the sense of Föllmer along $(\mathcal{P}_n)$ if and only if, for every $t \in [0,T]$ and $1 \le i,j \le d$ the limit

$$\big[X^i, X^j\big]_t := \lim_{n\to\infty} \sum_{[u,v]\in\mathcal{P}_n} \big(X^i_{v\wedge t} - X^i_{u\wedge t}\big)\big(X^j_{v\wedge t} - X^j_{u\wedge t}\big)$$

exists. Write $[X,X]$ for the resulting path with values in $\mathrm{Sym}\big(\mathbf{R}^d \otimes \mathbf{R}^d\big)$, i.e. the space of symmetric $d \times d$ matrices.

Lemma 5.9. *Assume $X : [0,T] \to \mathbf{R}^d$ has finite quadratic variation in the sense of Föllmer, along $(\mathcal{P}_n)$. Then the map $t \mapsto [X,X]_t$ is of bounded variation on $[0,T]$.*

Assume furthermore that $t \mapsto [X,X]_t$ is continuous. Then, for any continuous $G : [0,T] \to \mathcal{L}\big(\mathbf{R}^d \otimes \mathbf{R}^d, \mathbf{R}^e\big)$,

$$\lim_{n\to\infty} \sum_{\substack{[u,v]\in\mathcal{P}_n \\ u<t}} \langle G(u), X_{u,v}\otimes X_{u,v}\rangle = \int_0^t G(u) d[X,X]_u \in \mathbf{R}^e.$$

Proof. For the first statement, it is enough to argue component by component. Set $\big[X^i\big] := [X^i, X^i]$. By polarisation,

$$\left[X^i, X^j\right]_t = \frac{1}{2}\left[X^i + X^j\right]_t - \left[X^i\right]_t - \left[X^j\right]_t.$$

Since each term on the right-hand side is monotone in t, we see that $t \mapsto \left[X^i, X^j\right]_t$ is indeed of bounded variation.

Regarding the second statement, it is enough to check that, for continuous $g : [0,T] \to \mathbf{R}$ and Y of finite quadratic variation, with continuous bracket $t \mapsto [Y]_t$,

$$\lim_{n\to\infty} \sum_{\substack{[u,v]\in\mathcal{P}_n \\ u<t}} g(u) Y_{u,v}^2 = \int_0^t g(u) d[Y]_u. \tag{5.4}$$

Indeed, we can apply this for each component, with $g = G_{i,j}^k$ and

$$Y \in \left\{\left(X^i + X^j\right), X^i, X^j\right\},$$

which then also gives, by polarisation,

$$\sum_{\substack{[u,v]\in\mathcal{P}_n \\ u<t}} G_{i,j}^k(u) X_{u,v}^i X_{u,v}^j \to \int_0^t G_{i,j}^k(u) d\left[X^i, X^j\right]_u.$$

To see that (5.4) holds, write $\sum_{[u,v]\in\mathcal{P}_n, u<t} g(u) Y_{u,v}^2 = \int_{[0,t)} g(u) d\mu_n(u)$ with

$$\mu_n = \sum_{[u,v]\in\mathcal{P}_n, u<t} \delta_u Y_{u,v}^2$$

Note that the μ_n is a (finite) measure on $[0,t)$ with distribution function

$$F_n(s) := \mu_n([0,s]) = \sum_{\substack{[u,v]\in\mathcal{P}_n \\ u\le s}} Y_{u,v}^2.$$

As $n \to \infty$, $F_n(s) \to [Y]_s$ for any $s \le t$ by continuity of Y. Since $[Y]_s$ is a continuous function of s, convergence of the distribution functions implies weak convergence of the measures μ_n to the measure $d[Y]$ on $[0,t)$, with distribution function $[Y]$. Since $g|_{[0,t)}$ is continuous, (5.4) follows and the proof is finished. □

Combination of the above lemma with (5.3) gives the *Itô–Föllmer formula*,

$$F(X_t) = F(X_0) + \int_0^t DF(X_s) dX + \frac{1}{2}\int_0^t D^2 F(X_s) d[X,X]_t, \quad 0 \le t \le T \tag{5.5}$$

where the middle integral is given by the (now existent) limit of left-point Riemann-Stieltjes approximations

$$\lim_{n\to\infty} \sum_{[u,v]\in\mathcal{P}_n} \langle DF(X_u), X_{u,v}\rangle =: \int_0^t DF(X)dX.$$

In fact, we encourage the reader to verify as an exercise that this formula is valid whenever $X : [0,T] \to \mathbf{R}^d$ is continuous, of finite quadratic variation, with $t \mapsto [X,X]_t$ continuous. Note, however, that Föllmer's notion of quadratic variation (and the above integral) can and will depend in general on the sequence $(\mathcal{P}_n)$.

5.4 Backward integration

Let us recall that the *backward Itô-integral* is defined as

$$\int_0^T f_t \overleftarrow{dB}_t = \lim_n \sum_{[s,t]\in\mathcal{P}^n} f_t B_{s,t}$$

whenever this limit exists, in probability and uniformly on compact time intervals, and does not depend on the sequence of partitions (as long as their meshes tend to zero). For instance,

$$\int_0^t B_s \overleftarrow{dB}_s = \frac{1}{2}B_t^2 + \frac{t}{2}\,.$$

In many applications one encounters integrand f that are "backward adapted" in the sense that f_t is $\mathcal{F}_t^T$-measurable with $\mathcal{F}_s^t := \sigma(B_{u,v} : s \le u \le v \le t)$. For example

$$\int_0^t (B_t - B_s)\,\overleftarrow{dB}_s = B_t^2 - \frac{1}{2}B_t^2 - \frac{t}{2} = \frac{1}{2}B_t^2 - \frac{t}{2}$$

and we note (in contrast to the previous example) the zero mean property, which of course comes from a backward martingale structure. By analogy with its forward counterpart, the *backward Stratonovich integral* is defined as the backward Itô integral, minus $1/2$ times the quadratic variation of the integrand.

The purpose of this section is to understand backward integration as rough integration. To this end, recall that the rough integral of $(Y,Y') \in \mathscr{D}_X^{2\alpha}$ against $\mathbf{X} = (X,\mathbb{X})$ was defined by

$$\int_0^T Y d\mathbf{X} = \lim_{|P|\downarrow 0} \sum_{[s,t]\in P} Y_s X_{s,t} + Y_s'\mathbb{X}_{s,t}$$

where P are partitions of $[0,T]$ with mesh-size $|P|$. Clearly, some sort of "left-point" evaluation has been hard-wired into our definition of rough integral. On the other hand, one can expect that feeding in explicit second order information makes this choice somewhat less important than in the case of classical stochastic integration.

The next proposition, purely deterministic, answers the questions to what extent one can replace left-point by right-point evaluation.

Proposition 5.10 (Backward representation of rough integral). *Given a rough path $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ with $\alpha > 1/3$ and $(Y, Y') \in \mathscr{D}_X^{2\alpha}$ we have*

$$\int_0^T Y d\mathbf{X} = \lim_{|\mathcal{P}|\downarrow 0} \sum_{[s,t]\in\mathcal{P}} \bigl(Y_t X_{s,t} + Y_t'(\mathbb{X}_{s,t} - X_{s,t}\otimes X_{s,t})\bigr)\ . \tag{5.6}$$

Proof. We know that the rough integral is given as (compensated) Riemann–Stieltjes limit

$$\int_0^T Y d\mathbf{X} = \lim_{|\mathcal{P}|\downarrow 0} \sum_{[s,t]\in\mathcal{P}} \bigl(Y_s X_{s,t} + Y_s'\mathbb{X}_{s,t} + (*)_{s,t}\bigr)$$

whenever $(*)_{s,t} \approx 0$ in the sense that $(*)_{s,t} = \mathrm{O}\bigl(|t-s|^{3\alpha}\bigr) = \mathrm{o}(|t-s|)$, so that it does not contribute to the limit. (Recall (4.19) and Lemma 4.2.) But then

$$\begin{aligned} Y_s X_{s,t} + Y_s'\mathbb{X}_{s,t} &= Y_t X_{s,t} - Y_{s,t}X_{s,t} + Y_s'\mathbb{X}_{s,t} \\ &\approx Y_t X_{s,t} - Y_s' X_{s,t}\otimes X_{s,t} + Y_s'\mathbb{X}_{s,t} \\ &\approx Y_t X_{s,t} + Y_t'(\mathbb{X}_{s,t} - X_{s,t}\otimes X_{s,t})\ , \end{aligned}$$

and the claimed backward compensated Riemann–Stieltjes representation holds. □

Remark 5.11. Note that another way of writing (5.6) is the somewhat more suggestive

$$\int_0^T Y d\mathbf{X} = -\lim_{|\mathcal{P}|\downarrow 0} \sum_{[s,t]\in\mathcal{P}} Y_t X_{t,s} + Y_t'\mathbb{X}_{t,s}\ .$$

It is worth noting that a naively defined **backward rough integral**, by replacing left-point-evaluation (Y_s, Y_s') in the usual definition of the rough integral by right-point evaluation (Y_s, Y_s'), say

$$\lim_{|\mathcal{P}|\downarrow 0} \sum_{[s,t]\in\mathcal{P}} Y_t X_{s,t} + Y_t'\mathbb{X}_{s,t}$$

is, in general, not well-defined. In fact, in view of the above proposition, existence of this limit is equivalent to existence of (either)

$$\lim_{|\mathcal{P}|\downarrow 0} \sum_{[s,t]\in\mathcal{P}} Y_t' X_{s,t}\otimes X_{s,t} = \lim_{|\mathcal{P}|\downarrow 0} \sum_{[s,t]\in\mathcal{P}} Y_s' X_{s,t}\otimes X_{s,t}.$$

There is no reason why, for a general path $X \in \mathcal{C}^\alpha$, the above limits will exists. On the other hand, we already considered such sums in the context of the Itô–Föllmer formula, cf. Lemma 5.9. The appropriate condition for X was seen to be "quadratic variation (in the sense of Föllmer, along some $(\mathcal{P}_n)$)". And under this assumption,

$$\sum_{[s,t]\in\mathcal{P}^n} Y_s' X_{s,t}\otimes X_{s,t} \to \int_0^T Y_s' d[X]_s. \tag{5.7}$$

Of course, with probability one, d-dimensional standard Brownian motion has quadratic variation in the sense of Föllmer, along dyadic partitions, for instance, with $[B,B]_t = I\,t$, where I is the identity matrix. These remarks are crucial for proving

Theorem 5.12. *Define the random rough paths* $\mathbf{B}^{\text{Strat}} = (B, \mathbb{B}^{\text{Strat}})$ *and* $\mathbf{B}^{\text{back}} \stackrel{def}{=} (B, \mathbb{B}^{\text{back}})$ *by*

$$\mathbb{B}^{\text{Strat}}_{s,t} \stackrel{def}{=} \int_s^t B_{s,r} \circ dB_r = \mathbb{B}^{\text{Itô}}_{s,t} + \frac{1}{2} I(t-s) \;,$$
$$\mathbb{B}^{\text{back}}_{s,t} \stackrel{def}{=} \int_s^t B_{s,r}\, \overleftarrow{dB}_r = \mathbb{B}^{\text{Itô}}_{s,t} + I(t-s) \;.$$

Then, the following statements hold.

i) Assume $(Y(\omega), Y'(\omega)) \in \mathscr{D}^{2\alpha}_{B(\omega)}$ *a.s. and* Y, Y' *are adapted as processes. Then, with probability one, for all* $t \in [0,T]$,

$$\int_0^t Y\, d\mathbf{B}^{\text{Strat}} = \int_0^t Y_s dB_s + \frac{1}{2}\int_0^t Y'_s I\, ds = \int_0^t Y_s \circ dB_s \;,$$
$$\int_0^t Y\, d\mathbf{B}^{\text{back}} = \int_0^t Y_s dB_s + \int_0^t Y'_s I\, ds \;.$$

ii) Assume $(Y(\omega), Y'(\omega)) \in \mathscr{D}^{2\alpha}_{B(\omega)}$ *a.s. and* Y_t, Y'_t *are* $\mathcal{F}^T_t$*-measurable for all* $t < T$*. Then with probability one, for all* $r \in [0,T]$,

$$\int_r^T Y\, d\mathbf{B}^{\text{Strat}} = \int_r^T Y_t\, \overleftarrow{dB}_t - \frac{1}{2}\int_r^T Y'_t I\, dt = \int_r^T Y_s \circ \overleftarrow{dB}_s \;,$$
$$\int_r^T Y\, d\mathbf{B}^{\text{back}} = \int_r^T Y_t\, \overleftarrow{dB}_t \;.$$

Proof. Regarding point i), it follows from the definition of the rough integral (see also Example 4.13) that

$$\int_0^t Y\, d\mathbf{B}^{\text{back}} = \int_0^t Y\, d\mathbf{B}^{\text{Itô}} + \int_0^t Y' I\, ds \;.$$

The claim then follows from Proposition 5.1. The Stratonovich case is similar, now using Corollary 5.2.

We now turn to point ii). Thanks to the backward presentation established in Proposition 5.10,

$$\begin{aligned}\int_r^T Y\, d\mathbf{B}^{\text{back}} &= \lim_{n\to\infty} \sum_{[s,t]\in\mathcal{P}^n} Y_t B_{s,t} + Y'_t\big(\mathbb{B}^{\text{Itô}}_{s,t} + I(t-s) - B_{s,t}\otimes B_{s,t}\big) \\ &= \lim_{n\to\infty} \sum_{[s,t]\in\mathcal{P}^n} Y_t B_{s,t} + Y'_t \mathbb{B}^{\text{Itô}}_{s,t} - Y'_s\big(B_{s,t}\otimes B_{s,t} - I(t-s)\big) \;,\end{aligned}$$

using $Y'_{s,t}(X_{s,t} \otimes X_{s,t}) \approx 0$ and $Y'_{s,t}I(t-s) \approx 0$. (As before $(*)_{s,t} \approx 0$ means $(*)_{s,t} = \mathrm{o}(|t-s|)$.) Now we know that with probability 1, $B(\omega)$ has finite quadratic variation $[B]_t = It$, in the sense of Föllmer along some sequence $(\mathcal{P}^n)$. As a purely deterministic consequence, cf. (5.7), on the same set of full measure,

$$\lim_{n\to\infty} \sum_{[s,t]\in\mathcal{P}^n} Y'_s B_{s,t} \otimes B_{s,t} = \int_0^T Y'_s d[B]_s = \lim_{n\to\infty} \sum_{[s,t]\in\mathcal{P}^n} Y'_s I(t-s).$$

It follows at once that

$$\int_r^T Y \, d\mathbf{B}^{\text{back}}(\omega) = \lim_{n\to\infty} \sum_{[s,t]\in\mathcal{P}^n} Y_t B_{s,t} + Y'_t \mathbb{B}^{\text{Itô}}_{s,t}.$$

Since $\mathbb{B}^{\text{Itô}}_{s,t}$ is independent from $\mathcal{F}_t^T$ and Y_t, Y'_t are $\mathcal{F}_t^T$-measurable, a (backward) martingale argument shows that

$$\lim_{n\to\infty} \sum_{[s,t]\in\mathcal{P}^n} Y'_t \mathbb{B}^{\text{Itô}}_{s,t} = 0.$$

As a consequence, with probability one,

$$\int_r^T Y \, d\mathbf{B}^{\text{back}}(\omega) = \lim_{n\to\infty} \sum_{[s,t]\in\mathcal{P}^n} Y_t B_{s,t} = \int_r^T Y \, \overleftarrow{dB} \, .$$

The (backward) Stratonovich case is then treated as simple perturbation,

$$\begin{aligned}\int_r^T Y \, d\mathbf{B}^{\text{Strat}} &= \lim_{n\to\infty} \sum_{[s,t]\in\mathcal{P}^n} \Big(Y_t B_{s,t} + Y'_t \big(\mathbb{B}^{\text{Itô}}_{s,t} + I(t-s) - B_{s,t} \otimes B_{s,t} \big) \\ &\qquad - \frac{1}{2} Y'_t I(t-s) \Big) \\ &= \int_0^T Y_t \, \overleftarrow{dB}_t - \frac{1}{2} \int_0^T Y'_t I \, dt \, ,\end{aligned}$$

thus concluding the proof. $\square$

5.5 Exercises

Exercise 5.13. Complete the proof of Proposition 5.1 in the case the of unbounded Y'.

Solution 5.14. It suffices to show the convergence of (5.1) in probability; to this end, we introduce stopping times

$$\tau_M \stackrel{\text{def}}{=} \max\Big\{t \in \mathcal{P} : |Y'_t| < M\Big\} \in [0,T] \cup \{+\infty\}$$

and note that $\lim_{M\to\infty} \tau_M = \infty$ almost surely. The stopped process $S^{\tau_M}_\cdot$ is also a martingale, and we see as above that, for every fixed $M > 0$,

$$\Big| \sum_{\substack{[u,v]\in\mathcal{P} \\ u\le\tau_M}} Y'_u \mathbb{B}_{u,v} \Big|^2_{L^2} = \mathrm{O}(|\mathcal{P}|).$$

The proof is then easily finished by sending M to infinity.

Exercise 5.15 (Applications to statistics, see [DFM13]). Let B be a d-dimensional Brownian motion. Consider a $d \times d$ matrix A, a non-degenerate volatility matrix σ of the same dimension and a sufficiently nice map $h : \mathbf{R}^d \to \mathbf{R}^d$ so that the Itô stochastic differential equation

$$dY_t = A\, h(Y_t)dt + \sigma dB_t \tag{5.8}$$

has a unique solution, starting from any $Y_0 = y_0 \in \mathbf{R}^d$. (As a matter of fact, this SDE can be solved pathwise by considering the random ODE for $Z_t = Y_t - \sigma B_t$.) We are interested in the maximum likelihood estimation of the drift parameter A over a fixed time horizon $[0,T]$, given some observation path $Y = Y(\omega)$. Recall that this estimator, $\hat{A}_T(\omega)$, is based on the Radon–Nikodym density on pathspace, as given by Girsanov's theorem, relative to the drift free diffusion.

a) Let $d = 1$, $h(y) = y$. Show that the estimator $\hat{A}$ can be "robustified" in the sense that $\hat{A}_T(\omega) = \tilde{A}_T(Y(\omega))$ where

$$\tilde{A}_T(Y) = \frac{Y_T^2 - y_0^2 - \sigma^2 T}{2\int_0^T Y_t^2\, dt}. \tag{5.9}$$

is defined deterministically for any non-zero $Y \in \mathcal{C}([0,T],\mathbf{R}^d)$, and continuous with respect to uniform topology.

b) Take again $h(x) = x$, but now in dimension $d > 1$. Show that $\hat{A}$ admits a robust representation on rough path space, i.e. one has $\hat{A}_T(\omega) = \tilde{A}_T(\mathbf{Y}(\omega))$ where $\tilde{A}_T = \tilde{A}_T(\mathbf{Y})$ is deterministically defined and continuous with respect to α-Hölder rough path topology for any fixed $\alpha \in (1/3, 1/2)$. Here, $\mathbf{Y}(\omega)$ is the geometric rough path constructed from Y by iterated Stratonovich integration. Explain why there cannot be a robust representation on path space (as was the case when $d = 1$). What about more general h?

Exercise 5.16 (Skorokhod anticipating integration versus rough integration). We have seen that Itô integration coincides with rough integration against $\mathbf{B}^{\text{Itô}}(\omega)$, subject to natural conditions (in particular: adaptedness of (Y, Y') which guarantee that both are well-defined). A well-known extension of the Itô-integral to non-adapted integrands is given by the *Skorokhod* integral; details of which are found in most textbooks on Malliavin calculus, see for example [Nua06].

a) Let B denote one-dimensional Brownian motion on $[0,T]$. Show that the Skorokhod integral of B_T against B over $[0,T]$, in symbols $\int_0^T B_T \delta B_t$, is given by $B_T^2 - T$.
b) Set $Y_t(\omega) := B_T(\omega)$, with (zero) increments (trivially) controlled by B with $Y' := 0$. (In view of true roughness of Brownian motion, cf. Section 6, there is no other choice for Y'). Show that the rough integral of Y against Brownian motion over $[0,T]$ equals B_T^2. Conclude that Skorokhod and rough integrals (against Itô enhanced Brownian motion) do *not* coincide beyond adapted integrals.

Exercise 5.17 (Stratonovich anticipating integration versus rough integration; [CFV07]). In the spirit of Nualart–Pardoux [NP88], define the Stratonovich anticipating stochastic integral by

$$\int_0^t u(s,\omega) \circ dB_s(\omega) \stackrel{\text{def}}{=} \lim_{n\to\infty} \int_0^t u(s,\omega) \frac{dB_s^n(\omega)}{ds}\, ds,$$

where B^n is a dyadic piecewise-linear approximation to the Brownian motion B, whenever this limit exists in probability and uniformly on compacts. Consider (possibly anticipating) random 1-forms, $u(s,\omega) = F_\omega(B_s) \in \mathcal{C}_b^2$, for a.e. ω. Show that with probability one,

$$\int_0^\cdot F_\omega(B_s) d\mathbf{B}^{\text{Strat}}(\omega) \equiv \lim_{n\to\infty} \int_0^\cdot F_\omega(B_s) \frac{dB_s^n(\omega)}{ds}\, ds \ .$$

where the limit on the right hand side exists in almost-sure sense. Conclude that in this case rough integration against $\mathbf{B}^{\text{Strat}}$ coincides almost surely with Stratonovich anticipating stochastic integration, i.e.

$$\int_0^\cdot F_\omega(B_s) d\mathbf{B}^{\text{Strat}}(\omega) \equiv \int_0^\cdot F_\omega(B_s) \circ dB_s(\omega).$$

Hint: It is useful consider the pair $(\mathbf{B}^{\text{Strat}}, B^n)$, canonically viewed as (geometric) rough paths over $\mathbf{R}^{2d}$, followed by its rough path convergence to the "doubled" rough path $(\mathbf{B}^{\text{Strat}}, \mathbf{B}^{\text{Strat}})$ (which needs to be defined rigorously).

Remark. Nualart–Pardoux actually define their integral in terms of arbitrary deterministic (not necessarily dyadic) piecewise linear approximations and demand that the limit does not depend on the choice of the sequence of partitions. At the price of giving up the martingale argument, which made dyadic approximations easy (Proposition 3.6), everything can also be done in the general case; see exercises 10.13 and 10.14 below.

Exercise 5.18. Consider the Itô-Föllmer integral given by

$$\lim_{n\to\infty} \sum_{[u,v]\in\mathcal{P}_n} \langle DF(X_u), X_{u,v}\rangle =: \int_0^t DF(X) dX$$

whenever this limit exists along some a sequence of dissections $(\mathcal{P}_n) \subset [0,t]$ with mesh $|\mathcal{P}_n| \to 0$. Show that this limit does *not* exist, in general, when $X = B^H$, a d-dimensional fractional Brownian motion with Hurst parameter $H < 1/2$. **Hint:** Consider the simplest possible non-trival case, namely $d = 1$ and $F(x) = x^2$.

Solution 5.19. Assume convergence in probability say along some $(\mathcal{P}_n)$ for the approximating (left-point) sum,

$$\sum_{[u,v] \in \mathcal{P}_n} X_u X_{u,v}.$$

We look for a contradiction. Elementary "calculus for sums" implies that the mid-point sum converges, i.e. where X_u above is replaced by $X_u + X_{u,v}/2$. It follows that convergence of the left-point sums is equivalent to to existence of quadratic variation, i.e. existence of

$$\lim_{n\to\infty} \sum_{[u,v] \in \mathcal{P}_n} |X_{u,v}|^2.$$

Note that $\mathbf{E}|X_{u,v}|^2 = (1/2^n)^{2H}$ so that the expectation of this sum equals $2^{n(1-2H)}$, which diverges when $H < 1/2$. In particular, quadratic variation does not exist as L^1 limit. But is also cannot exists as a limit in probability, for both types of convergence are equivalent on any finite Wiener–Itô chaos.

Exercise 5.20. In Proposition 5.6, replace the assumption that $\mathbf{X} = (X, \mathbb{S}) \in \mathscr{C}_r^\alpha([0,T], V)$ with $\alpha > 1/3$, by a suitable p-variation assumption with $p < 3$. Show that $[\mathbf{X}]$ has finite $p/2$-variation and that $\int D^2 F(X) d[\mathbf{X}]$, as it appears in Itô's formula for reduced rough paths, remains a Young integral.

Exercise 5.21. Prove Proposition 1.1.

Solution 5.22. Without loss of generality, we consider the problem on the interval $[0, 2\pi]$. Assume by contradiction that there is a space $\mathcal{B} \subset \mathcal{C}([0, 2\pi])$ which carries the law μ of Brownian motion and such that $(f, g) \mapsto \int f\, dg$ is continuous on $\mathcal{B}$. By definition, the Cameron–Martin space of μ is $\mathcal{H} = W_0^{1,2}([0,1])$, which has an orthonormal basis $\{e_n\}_{n\in\mathbf{Z}}$ given by

$$e_0(t) = \frac{t}{\sqrt{2\pi}}\;, \quad e_k(t) = \frac{\sin kt}{k\sqrt{\pi}}\;, \quad e_{-k}(t) = \frac{1 - \cos kt}{k\sqrt{\pi}}\;,$$

for $k > 0$. It follows from standard Gaussian measure theory [Bog98] that, given a sequence ξ_n of i.i.d. normal Gaussian random variables, the sequence $X_N = \sum_{n=-N}^N e_n\, \xi_n$ converges almost surely in $\mathcal{B}$ to a limit X such that the law of X is μ. Write now $Y_N = \sum_{n=-N}^N \operatorname{sign}(n) e_n\, \xi_n$, so that one also has $Y_N \to Y$ with law of Y given by μ.

This immediately leads to a contradiction: on the one hand, assuming that $(f, g) \mapsto \int f\, dg$ is continuous on $\mathcal{B}$, this implies that $\int_0^{2\pi} X_N(t)\, dY_N(t)$ converges to some finite (random) real number. On the other hand, an explicit calculation yields

$$\int_0^{2\pi} X_N(t)\, dY_N(t) = \frac{\xi_0^2}{2} + \sum_{n=1}^{N} \frac{\xi_n^2 + \xi_{-n}^2}{n} \;.$$

It is now straightforward to verify that this diverges logarithmically, thus concluding the proof.

5.6 Comments

Rough integrals of 1-forms against the Brownian rough path (and also continuous semi-martingales enhanced to rough paths) are well known to coincide with stochastic integrals, see [LQ02, FV10b] for instance, but the extensions presented in this section seem to be new. Our Itô formula for reduced rough paths also appears to be new.

Chapter 6
Doob–Meyer type decomposition for rough paths

Abstract A deterministic Doob–Meyer type decomposition is established. It is closely related to question to what extent Y' is determined from Y, given that $(Y, Y') \in \mathscr{D}_X^{2\alpha}$. The crucial property is true roughness of X, a deterministic property that guarantees that X varies in all directions, all the time.

6.1 Motivation from stochastic analysis

Consider a continuous semi-martingale $(S_t : t \geq 0)$. By definition (e.g. [RY91, Ch. IV]) this means that $S = M + A$ where $M \in \mathcal{M}$, the space of continuous local martingales, and $A \in \mathcal{V}$, the space of continuous adapted process of finite variation. Then it is well known that the decomposition $S = M + A$ is unique in the following sense.

Proposition 6.1. *Assume $M, \tilde{M} \in \mathcal{M}$, vanishing at zero, and $A, \tilde{A} \in \mathcal{V}$ such that $M + A \equiv \tilde{M} + \tilde{A}$ (i.e. the respective processes are indistinguishable). Then*

$$M \equiv \tilde{M} \quad \textit{and} \quad A \equiv \tilde{A} \; .$$

Furthermore, if $S = M + A \equiv 0$ on some random interval $[0, \tau)$ where τ is a stopping time, then $[M] \equiv 0$ on $[0, \tau)$ and $A \equiv 0$ on $[0, \tau)$.

Proof. Assume $M + A \equiv \tilde{M} + \tilde{A}$. Then $M - \tilde{M} \in \mathcal{V}$, and null at zero. By a standard result in martingale theory, see for example [RY91, IV, Prop 1.2], this entails that $M - \tilde{M} \equiv 0$. But then $A \equiv \tilde{A}$ and the proof is complete.

Regarding the second statement, consider the stopped semi-martingale, $S^\tau = M^\tau + A^\tau$ where $M_t^\tau = M_{t \wedge \tau}$ and similarly for A. By assumption $S^\tau \equiv 0$ and hence, by the first part, $M^\tau, A^\tau \equiv 0$. This also implies that the quadratic variation of M^τ, denoted by $[M^\tau]$, vanishes. Since $[M^\tau] = [M]^\tau$ (see e.g. [RY91, Ch. IV]) it indeed follows that $[M] \equiv 0$ on $[0, \tau)$. $\square$

The above proposition applies in particular when M is given as multidimensional (say $\mathbf{R}^e$-valued) stochastic integral of a suitable $L(\mathbf{R}^d, \mathbf{R}^e)$-valued integrand Y (continuous and adapted will do) against d-dimensional Brownian motion B, while A is the indefinite integral of some suitable $\mathbf{R}^e$-valued process Z (again, continuous and adapted will do). We then have

Corollary 6.2. *Let B be a d-dimensional Brownian motion and let Y, Z, $\tilde{Y}$, $\tilde{Z}$ be continuous stochastic processes adapted to the filtration generated by B. Assume, in the sense of indistinguishability of left- and right-hand sides, that*

$$\int_0^\cdot Y dB + \int_0^\cdot Z dt \equiv \int_0^\cdot \tilde{Y} dB + \int_0^\cdot \tilde{Z}\, dt \qquad \textit{on } [0,T]. \tag{6.1}$$

Then $Y \equiv \tilde{Y}$ and $Z \equiv \tilde{Z}$ on $[0,T]$.

Proof. We may take dimension $e = 1$, otherwise argue componentwise. Also, by linearity, it suffices to consider the case $\tilde{Y} = 0$, $\tilde{Z} = 0$. By the second part of the previous proposition

$$\left[\int_0^\cdot Y dB\right]_\cdot \equiv \left[\sum_{k=1}^d \int_0^\cdot Y_k dB^k\right]_\cdot \equiv 0 \text{ on } [0,T].$$

On the other hand, due to $d\left[B^k, B^l\right]_t = dt$ if $k = l$, and zero else,

$$\left[\sum_{k=1}^d \int_0^\cdot Y_k dB^k\right]_\cdot \equiv \sum_{k,l=1}^d \int_0^\cdot Y_k Y_l d\left[B^k, B^l\right] = \sum_{k=1}^d \int_0^\cdot Y_k^2 dt.$$

It follows that $Y \equiv 0$ as claimed. By differentiation, it then follows that also $Z \equiv 0$. □

Clearly, the martingale and quadratic (co-)variation – i.e. *probabilistic* – properties of B play a key role in the proof of Corollary 6.2. It is worth noting that, with β a scalar Brownian motion and $B^1 = B^2 = \beta$ the conclusion fails; try non-zero $Y^1 \equiv -Y^2$, $Z \equiv 0$. It is crucial that d-dimensional standard Brownian motion "moves in all directions", captured through the non-degeneracy of the quadratic covariation matrix $\left[B^k, B^l\right]_t$.

Surprisingly perhaps, one can formulate a purely *deterministic* decomposition of the form (6.1): the stochastic integrals will be replaced by rough integrals, the relevant probabilistic properties of B by certain conditions ("roughness from below[1], in all directions") on the sample path.

[1] As opposed to Hölder regularity which quantifies "roughness from above", in the sense of an upper estimate of the increment.

6.2 Uniqueness of the Gubinelli derivative and Doob–Meyer

Here and in the sequel of this section we fix $\alpha \in (\frac{1}{3}, \frac{1}{2}]$, a rough path $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha([0,T], V)$ and a controlled rough path $(Y, Y') \in \mathscr{D}_X^{2\alpha}$. We first address the question to what extent X and Y determine the Gubinelli derivative Y'. As it turns out, Y' is uniquely determined, provided that X is sufficiently "rough from below, in all directions". A Doob–Meyer type decomposition will then follow as a corollary.

Let us first consider the case when X is scalar, i.e. with values in $V = \mathbf{R}$. Assume that for some given $s \in [0,T)$, there exists a sequence of times $t_n \downarrow s$ such that $|X_{s,t_n}|/|t_n - s|^{2\alpha} \to \infty$, i.e.

$$\overline{\lim_{t \downarrow s}} \frac{|X_{s,t}|}{|t-s|^{2\alpha}} = +\infty .$$

Then Y'_s is uniquely determined from Y by (4.16) and the condition that $\|R^Y\|_{2\alpha} < \infty$. In fact, one necessarily has $X_{s,t_n} \in \mathbf{R} \setminus \{0\}$ for n large enough and so, from the very definition of R^Y,

$$Y'_s = \frac{Y_{s,t_n}}{X_{s,t_n}} - \frac{R^Y_{s,t_n}}{|t_n - s|^{2\alpha}} \frac{|t_n - s|^{2\alpha}}{X_{s,t_n}}$$

which implies that $\lim_{n\to\infty} Y_{s,t_n}/X_{s,t_n}$ exists and equals Y'_s. The multidimensional case is not that different, and the above consideration suggests the following definition.

Definition 6.3. For fixed $s \in [0,T)$ we call $X \in \mathcal{C}^\alpha([0,T], V)$ "*rough at time s*" if

$$\forall v^* \in V^* \backslash \{0\} : \quad \overline{\lim_{t \downarrow s}} \frac{|\langle v^*, X_{s,t}\rangle|}{|t-s|^{2\alpha}} = \infty .$$

If X is rough on some dense set of $[0,T]$, we call it *truly rough*.

This definition is vindicated by the following result.

Proposition 6.4 (Uniqueness of Y'). *Let $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$, $(Y, Y') \in \mathscr{D}_X^{2\alpha}$, so that the rough integral $\int Y \, d\mathbf{X}$ exists. Assume X is rough at some time $s \in [0,T)$. Then*

$$Y_{s,t} = \mathrm{O}\big(|t-s|^{2\alpha}\big) \quad \textit{as } t \downarrow s \qquad \Rightarrow \qquad Y'_s = 0 . \tag{6.2}$$

As a consequence, if X is truly rough and $(Y, \tilde{Y}') \in \mathscr{D}_X^{2\alpha}$ is another controlled rough path (with respect to X) then $Y' \equiv \tilde{Y}'$.

Proof. From the definition of $(Y, \tilde{Y}') \in \mathscr{D}_X^{2\alpha}$, we have

$$Y_{s,t} = Y'_s X_{s,t} + \mathrm{O}\big(|t-s|^{2\alpha}\big) .$$

Hence, for $t \in (s, s+\varepsilon)$,

$$\frac{Y'_s X_{s,t}}{|t-s|^{2\alpha}} = \frac{Y_{s,t}}{|t-s|^{2\alpha}} + \mathrm{O}(1) = \mathrm{O}(1)\;,$$

where the second equality follows from the assumption made in (6.2). Now, $Y'_s X_{s,t}$ takes values in $\bar{W}$, the same Banach space in which Y takes its values. For every $w^* \in \bar{W}^*$, the map $V \ni v \mapsto \langle w^*, Y'_s v\rangle$ defines an element $v^* \in V^*$ so that

$$\frac{|\langle v^*, X_{s,t}\rangle|}{|t-s|^{2\alpha}} = \left|\left\langle w^*, \frac{Y'_s X_{s,t}}{|t-s|^{2\alpha}}\right\rangle\right| = \mathrm{O}(1) \text{ as } t \downarrow s;$$

Unless $v^* = 0$, the assumption that "X is rough at time s" implies that, along some sequence $t_n \downarrow s$, we have the divergent behaviour $|\langle v^*, X_{s,t_n}\rangle|/|t_n - s|^{2\alpha} \to \infty$, which contradicts that the same expression is $\mathrm{O}(1)$ as $t_n \downarrow s$. We thus conclude that $v^* = 0$. In other words,

$$\forall w^* \in W^*, v \in V : \qquad \langle w^*, Y'_s v\rangle = 0\;,$$

and this clearly implies $Y'_s = 0$. This finishes the proof of the implication stated in (6.2). □

The following result should be compared with Corollary 6.2.

Theorem 6.5 (Doob–Meyer for rough paths). *Assume that X is rough at some time $s \in [0,T)$ and let $(Y,Y') \in \mathscr{D}_X^{2\alpha}$. Then*

$$\int_s^t Y\,d\mathbf{X} = \mathrm{O}\big(|t-s|^{2\alpha}\big) \quad \text{as } t \downarrow s \qquad \Rightarrow \qquad Y_s = 0\;. \tag{6.3}$$

As a consequence, if X is truly rough, $\big(\tilde{Y}, \tilde{Y}'\big) \in \mathscr{D}_X^{2\alpha}$ and $Z, \tilde{Z} \in \mathcal{C}([0,T],W)$, then the identity

$$\int_0^\cdot Y\,d\mathbf{X} + \int_0^\cdot Z\,dt \equiv \int_0^\cdot \tilde{Y}\,d\mathbf{X} + \int_0^\cdot \tilde{Z}\,dt \tag{6.4}$$

on $[0,T]$ implies that $(Y,Y') \equiv (\tilde{Y},\tilde{Y}')$ and $Z \equiv \tilde{Z}$ on $[0,T]$.

Proof. Recall from Theorem 4.10 that $(I,I') := \big(\int Y\,d\mathbf{X}, Y\big)$ is controlled by X, i.e. $(I,I') \in \mathscr{D}_X^{2\alpha}$. The statement (6.3) is then an immediate consequence of (6.2).

The claim is now straightforward. Pick any $s \in [0,T)$ such that X is rough at time s. From (6.4), and for all $0 \le s \le t \le T$,

$$\int_s^t \big(Y - \tilde{Y}\big)\,d\mathbf{X} = \int_s^t \big(Z_r - \tilde{Z}_r\big)\,dr = \mathrm{O}(|t-s|) = \mathrm{O}\big(|t-s|^{2\alpha}\big)\;,$$

where the last inequality is just the statement that $|t-s| = \mathrm{O}\big(|t-s|^{2\alpha}\big)$ as $t \downarrow s$, thanks to $\alpha \le 1/2$. We then conclude using (6.3) that $Y_s = \tilde{Y}_s$. If we now assume true roughness of X, this conclusion holds for a dense set of times s and hence, by continuity of Y and $\tilde{Y}$, we have $Y \equiv \tilde{Y}$. But then, by Proposition 6.4, we also have $Y' \equiv \tilde{Y}'$ and so

$$\int_0^{\cdot} Y \, d\mathbf{X} \equiv \int_0^{\cdot} \tilde{Y} \, d\mathbf{X} \, .$$

(Attention that the above notation "hides" the dependence on Y' resp. $\tilde{Y}'$.) But then (6.4) implies

$$\int_0^t Z_r \, dr \equiv \int_0^t \tilde{Z}_r \, dr \qquad \text{for } t \in [0,T],$$

and we conclude by differentiation with respect to t. □

6.3 Brownian motion is truly rough

Recall that (say, d-dimensional standard) Brownian motion satisfies the so-called *(Khintchine) law of the iterated logarithm*, that is

$$\forall t \geq 0 : \quad \mathbf{P}\left[\overline{\lim_{h \downarrow 0}} \frac{|B_{t,t+h}|}{h^{\frac{1}{2}} (\ln \ln 1/h)^{1/2}} = \sqrt{2} \right] = 1. \tag{6.5}$$

See [McK69, p.18] or [RY91, Ch. II] for instance, typically proved with exponential martingales. Remark that it is enough to consider $t = 0$ since $(B_{t,t+h} : h \geq 0)$ is also a Brownian motion.

Theorem 6.6. *With probability one, Brownian motion on $V = \mathbf{R}^d$ is truly rough, relative to any Hölder exponent $\alpha \in [1/4, 1/2)$.*

Proof. It is enough to show that, for fixed time s, and any $\theta \in [1/2, 1)$,

$$\mathbf{P}\left[\forall \varphi \in V^*, |\varphi| = 1 : \overline{\lim_{t \downarrow s}} \frac{|\varphi(B_{s,t})|}{|t-s|^{\theta}} = +\infty \right] = 1.$$

(Then take $s \in \mathbf{Q}$ and conclude that the above event holds true, simultanously for all such s, with probability one.)

To this end, set $h^{\frac{1}{2}} (\ln \ln 1/h)^{1/2} \equiv \psi(h)$. We need the following two consequences of (6.5). There exists $c > 0$ (here $c = \sqrt{2}$) such that for every every fixed unit dual vector $\varphi \in V^* = \left(\mathbf{R}^d\right)^*$ and every fixed $s \in [0,T)$

$$\mathbb{P}\left[\overline{\lim_{t \downarrow s}} |\langle \varphi, B_{s,t} \rangle| / \psi(t-s) \geq c \right] = 1,$$

$$\mathbb{P}\left[\overline{\lim_{t \downarrow s}} \frac{|B_{s,t}|}{\psi(t-s)} < \infty \right] = 1.$$

Take $K \subset V^*$ to be any dense, countable set of dual unit vectors. Since K is countable, the set on which the first condition holds simultanously for all $\varphi \in K$ has full measure,

$$\mathbb{P}\left[\forall \varphi \in K : \overline{\lim_{t \downarrow s}} |\varphi(B_{s,t})|/\psi(t-s) \geq c\right] = 1$$

On the other hand, every unit dual vector $\varphi \in V^*$ is the limit of some $(\varphi_n) \subset K$. Then

$$\frac{|\langle \varphi_n, B_{s,t} \rangle|}{\psi(t-s)} \leq \frac{|\langle \varphi, B_{s,t} \rangle|}{\psi(t-s)} + |\varphi_n - \varphi|_{V^*} \frac{|B_{s,t}|}{\psi(t-s)}$$

so that, using $\overline{\lim}\,(|a|+|b|) \leq \overline{\lim}\,(|a|) + \overline{\lim}\,(|b|)$, and restricting to the above set of full measure,

$$c \leq \overline{\lim_{t \downarrow s}} \frac{|\langle \varphi_n, B_{s,t} \rangle|}{\psi(t-s)} \leq \overline{\lim_{t \downarrow s}} \frac{|\langle \varphi, B_{s,t} \rangle|}{\psi(t-s)} + |\varphi_n - \varphi|_{V^*} \overline{\lim_{t \downarrow s}} \frac{|B_{s,t}|_V}{\psi(t-s)}.$$

Sending $n \to \infty$ gives, with probability one,

$$0 < c \leq \overline{\lim_{t \downarrow s}} \frac{|\langle \varphi, B_{s,t} \rangle|}{\psi(t-s)}.$$

Hence, for a.e. sample $B = B(\omega)$ we can pick a sequence (t_n) converging to s such that $|\langle \varphi, B_{s,t_n} \rangle|/\psi(t_n - s) \geq c - 1/n$. On the other hand, for any $\theta \geq 1/2$ we have

$$\begin{aligned} \frac{|\langle \varphi, B_{s,t_n}(\omega) \rangle|}{|t_n - s|^\theta} &= \frac{|\langle \varphi, B_{s,t_n}(\omega) \rangle|}{\psi(t_n - s)} \frac{\psi(t_n - s)}{|t_n - s|^\theta} \\ &\geq (c - 1/n)|t_n - s|^{\frac{1}{2} - \theta} L(t_n - s) \to \infty \ , \end{aligned}$$

where in the borderline case $\theta = 1/2$ (which corresponds to $\alpha = 1/4$) this divergence is only logarithmic, $L(\tau) = (\ln \ln 1/\tau)^{1/2}$. $\square$

6.4 A deterministic Norris' lemma

We now turn our attention to a quantitative version of true roughness. In essence, we now replace 2α in definition 6.3 by θ *and quantify the divergence, uniformly over all directions.*

Definition 6.7. A path $X : [0,T] \to V$ with values in a Banach space V is said to be θ-*Hölder rough* for $\theta \in (0,1)$, on scale (smaller than) $\varepsilon_0 > 0$, if there exists a constant $L := L_\theta(X) := L(\theta, \varepsilon_0, T; X) > 0$ such that for every $\varphi \in V^*$, $s \in [0,T]$ and $\varepsilon \in (0, \varepsilon_0]$ there exists $t \in [0,T]$ such that

$$|t - s| < \varepsilon \ , \quad \text{and} \quad |\varphi(X_{s,t})| \geq L_\theta(X)\, \varepsilon^\theta |\varphi| \ . \tag{6.6}$$

the largest such value of L is called the *modulus of* θ-*Hölder roughness* of X.

Observe that, indeed, any element in $\mathcal{C}^\alpha$ which is θ-Hölder rough for $\theta < 2\alpha$ is truly rough. (We shall see in the next section that multidimensional Brownian motion is θ-Hölder rough for any $\theta > 1/2$.) The following result can be viewed as quantitative version of Proposition 6.4.

Proposition 6.8. *Let $(X, \mathbb{X}) \in \mathscr{C}^\alpha\big([0,T], V\big)$ be such that X is θ-Hölder rough for some $\theta \in (0,1]$. Then, for every controlled rough path $(Y, Y') \in \mathscr{D}_X^{2\alpha}\big([0,T], W\big)$ one has,*

$$\forall \varepsilon \in (0, \varepsilon_0] : L\varepsilon^\theta \|Y'\|_\infty \leq \operatorname{osc}(Y, \varepsilon) + \big\|R^Y\big\|_{2\alpha} \varepsilon^{2\alpha} \; . \tag{6.7}$$

As immediate consequence, if $\theta < 2\alpha$, Y' is uniquely determined from Y, i.e. if $\big(Y, Y'\big)$ and $\big(\tilde{Y}, \tilde{Y}'\big)$ both belong to $\mathscr{D}_X^{2\alpha}$ and $Y \equiv \tilde{Y}$, then $Y' \equiv \tilde{Y}'$.

Proof. Let us start with the consequence: apply estimate (6.7) with Y replaced by $Y - \tilde{Y} = 0$ and similarly Y' replaced by $Y' - \tilde{Y}'$. Thanks to $L > 0$ it follows that

$$\big\|Y' - \tilde{Y}'\big\|_\infty = \mathrm{O}\big(\varepsilon^{2\alpha-\theta}\big)$$

and we send $\varepsilon \to 0$ to conclude $Y' = \tilde{Y}'$. The remainder of the proof is devoted to establish (6.7). Fix $s \in [0,T]$ and $\varepsilon \in (0, \varepsilon_0]$. From the definition of the remainder R^Y in (4.16), it then follows that

$$\sup_{|t-s|\leq\varepsilon} |Y_s' X_{s,t}| \leq \sup_{|t-s|\leq\varepsilon} \big(|Y_{s,t}| + |R_{s,t}^Y|\big) \leq \operatorname{osc}(Y, \varepsilon) + \|R^Y\|_{2\alpha}\, \varepsilon^{2\alpha} \; . \tag{6.8}$$

Let now $\varphi \in W^*$ be such that $|\varphi| = 1$. Since X is θ-Hölder rough by assumption, there exists $v = v(\varphi)$ with $|v - s| < \varepsilon$ such that

$$\big|\varphi(Y_s' X_{s,v})\big| = \big|\big((Y_s')^*\varphi\big)(X_{s,v})\big| > L\, \varepsilon^\theta |(Y_s')^*\varphi| \; . \tag{6.9}$$

(Note that one has indeed $(Y_s')^* \colon W^* \to V^*$.) Combining both (6.8) and (6.9), we thus obtain that

$$L\, \varepsilon^\theta\, |(Y_s')^*\varphi| \leq \operatorname{osc}(X, \varepsilon) + \|R^Y\|_{2\alpha}\, \varepsilon^{2\alpha} \; .$$

Taking the supremum over all such $\varphi \in W^*$ of unit length,[2] and using the fact that the norm of a linear operator is equal to the norm of its adjoint, we obtain

$$L\, \varepsilon^\theta\, |Y_s'| \leq \operatorname{osc}(Y, \varepsilon) + \|R^Y\|_{2\alpha}\, \varepsilon^{2\alpha} \; .$$

Since s was also arbitrary, the stated bound follows at once. □

Remark 6.9. Even though the argument presented above is independent of the dimension of V, we are not aware of any example where $L(\theta, X) > 0$ and $\dim V = \infty$. The reason why this definition works well only in the finite-dimensional case will be apparent in the proof of Proposition 6.11 below.

[2] Note that $|Y_s'^*|$ denotes the operator norm, by definition equal to $\sup_{|\varphi|=1} |Y_s'^*\varphi|$.

This leads us to the folllowing quantitative version our previous Doob–Meyer result for rough paths, Theorem 6.5. As usual, we assume that $\alpha \in (1/3, 1/2)$.

Theorem 6.10 (Norris lemma for rough paths). *Let* $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha\big([0,T], V\big)$ *be such that* X *is* θ*-Hölder rough with* $\theta < 2\alpha$*. Let* $(Y, Y') \in \mathscr{D}_X^{2\alpha}\big([0,T], \mathcal{L}(V,W)\big)$ *and* $Z \in \mathcal{C}^\alpha\big([0,T], W\big)$ *and set*

$$I_t = \int_0^t Y_s d\mathbf{X}_s + \int_0^t Z_s ds.$$

Then there exist constants $r > 0$ *and* $q > 0$ *such that, setting*

$$\mathcal{R} := 1 + L_\theta(X)^{-1} + \|\mathbf{X}\|_\alpha + \|Y, Y'\|_{X;2\alpha} + |Y_0| + |Y_0'| + \|Z\|_\alpha + |Z_0|$$

one has the bound

$$\|Y\|_\infty + \|Z\|_\infty \leq M \mathcal{R}^q \|I\|_\infty^r ,$$

for a constant M *depending only on* α*,* θ*, and the final time* T*.*

Proof. We leave the details of the proof as an exercise, see [HP13], and only sketch its broad lines.

First, we conclude from Proposition 6.8 that I small in the supremum norm implies that $\|Y\|_\infty$ is also small. Then, we use interpolation to conclude from this that (Y, Y') is small when viewed as an element of $\mathscr{D}^{2\bar{\alpha}}$ for $\bar{\alpha} < \alpha$, thus implying that $\int Y \, d\mathbf{X}$ is necessarily small. This implies that $\int Z \, ds$ is itself small from which, using again interpolation, we finally conclude that Z itself must be small in the supremum norm. $\square$

6.5 Brownian motion is Hölder rough

We now turn to Hölder-roughness of Brownian motion. Our focus will be on the unit interval $T = 1$, and we consider scale up to $\varepsilon_0 = 1/2$ for the sake of argument.

Proposition 6.11. *Let* B *be a standard Brownian motion on* $[0,1]$ *taking values in* $\mathbf{R}^d$*. Then, for every* $\theta > \frac{1}{2}$*, the sample paths of* B *are almost surely* θ*-Hölder rough. Moreover, with scale* $\varepsilon_0 = 1/2$ *and writing* $L_\theta(B)$ *for the modulus of* θ*-Hölder roughness, there exists constants* M *and* c *such that*

$$\mathbf{P}(L_\theta(B) < \varepsilon) \leq M \, \exp\big(-c\varepsilon^{-2}\big) ,$$

for all $\varepsilon \in (0,1)$*.*

The proof of Proposition 6.11 relies on the following variation of the standard small ball estimate for Brownian motion:

Lemma 6.12. *Let B be a d-dimensional standard Brownian motion. Then, there exist constants $c > 0$ and $C > 0$ such that*

$$\mathbf{P}\Big(\inf_{|\varphi|=1} \sup_{t\in[0,\delta]} |\langle \varphi, B(t)\rangle| \le \varepsilon\Big) \le C\exp(-c\delta\varepsilon^{-2}) . \tag{6.10}$$

Proof. The standard small ball estimate for Brownian motion (see for example [LS01]) yields the bound

$$\sup_{|\varphi|=1} \mathbf{P}\Big(\sup_{t\in[0,\delta]} |\langle \varphi, B(t)\rangle| \le \varepsilon\Big) \le C\exp(-c\delta\varepsilon^{-2}) . \tag{6.11}$$

The required estimate then follows from a standard chaining argument, as in [Nor86, p. 127]: cover the sphere $|\varphi| = 1$ with $\varepsilon^{-2(d-1)}$ balls of radius ε^2, say, centred at φ_i. We then use the fact that, since the supremum of B has Gaussian tails, if $\sup_{t\in[0,\delta]} |\langle \varphi_i, B(t)\rangle| \le \varepsilon$, then the same bound, but with ε replaced by 2ε holds with probability exponentially close to 1 uniformly over all φ in the ball of radius ε^2 centred at φ_i. Since there are only polynomially many such balls required to cover the whole sphere, (6.10) follows. Note that this chaining argument uses in a crucial way that the number of balls or radius ε^2 required to cover the sphere $\|\varphi\| = 1$ grows only polynomially with ε^{-1}.

It is clear that bounds of the type (6.10) break down in infinite dimensions: if we consider a cylindrical Wiener process, then (6.11) still holds, but the unit sphere of a Hilbert space cannot be covered by a finite number of small balls anymore. If on the other hand, we consider a process with a non-trivial covariance, then we can get the chaining argument to work, but the bound (6.11) would break down due to the fact that $\langle \varphi, B(t)\rangle$ can then have arbitrarily small variance. □

Proof (Proposition 6.11). With $T = 1, \varepsilon_0 = 1/2$, a different way of formulating Definition 6.7 is given by

$$L_\theta(X) = \inf \sup_{t:|t-s|\le\varepsilon} \frac{1}{\varepsilon^\theta} |\langle \varphi, X_{s,t}\rangle|.$$

where the inf is taken over $|\varphi| = 1, s \in [0, 1]$ and $\varepsilon \in (0, 1/2]$. We then define the "discrete analog" $D_\theta(X)$ of $L_\theta(X)$ to be given by

$$D_\theta(X) = \inf \sup_{s,t\in I_{k,n}} 2^{n\theta} |\langle \varphi, X_{s,t}\rangle| ,$$

where $I_{k,n} = [\frac{k-1}{2^n}, \frac{k}{2^n}]$ and the inf is taken over $|\varphi| = 1$, $n \ge 1$ and $k \le 2^n$. We first claim that

$$L_\theta(X) \ge \frac{1}{2}\frac{1}{2^\theta} D_\theta(X). \tag{6.12}$$

To this end, fix a unit vector $\varphi \in V^*$, $s \in [0, 1]$ and $\varepsilon \in (0, 1/2]$. Pick $n \ge 1 :$ $\varepsilon/2 < 2^{-n} \le \varepsilon$. It follows that there exists some k such that $I_{k,n}$ is included in the set $\{t : |t - s| \le \varepsilon\}$. Then, by definition of D_θ, for any unit vector φ there exist two points $t_1, t_2 \in I_{k,n}$ such that

$$|\langle \varphi, X_{t_1,t_2}\rangle| \geq 2^{-n\theta} D_\theta(X).$$

Therefore, by the triangle inequality, we conclude that the magnitude of the difference between $\langle \varphi, X_s\rangle$ and one of the two terms $\langle \varphi, X_{t_i}\rangle, i = 1, 2$ (say t_1) is at least

$$|\langle \varphi, X_{s,t_1}\rangle| \geq \frac{1}{2} 2^{-n\theta} D_\theta(X)$$

and therefore

$$\frac{|\langle \varphi, X_{s,t_1}\rangle|}{\varepsilon^\theta} \geq \frac{1}{2}\frac{2^{-n\theta}}{\varepsilon^\theta} D_\theta(X) \geq \frac{1}{2}\frac{1}{2^\theta} D_\theta(X).$$

Since s, ε and φ were chosen arbitrarily, the claim (6.12) follows.

Applying this to Brownian sample paths, $X = B(\omega)$, it follows that it is sufficient to obtain the requested bound on $\mathbf{P}(D_\theta(B) < \varepsilon)$. We have the straightforward bound

$$\begin{aligned}\mathbf{P}(D_\theta(B) < \varepsilon) &\leq \mathbf{P}\Big(\inf_{\|\varphi\|=1} \inf_{n\geq 1} \inf_{k \leq 2^n} \sup_{s,t\in I_{k,n}} \frac{|\langle \varphi, B_{s,t}\rangle|}{2^{-n\theta}} < \varepsilon\Big) \\ &\leq \sum_{n=1}^\infty \sum_{k=1}^{2^n} \mathbf{P}\Big(\inf_{\|\varphi\|=1} \sup_{s,t\in I_{k,n}} |\langle \varphi, B_{s,t}\rangle| < 2^{-n\theta}\varepsilon\Big) .\end{aligned}$$

Trivially $\sup_{s,t\in I_{k,n}} |\langle \varphi, B_{s,t}\rangle| \geq \sup_{t\in I_{k,n}} |\langle \varphi, B_{r,t}\rangle|$, where r is the left boundary of the interval $I_{k,n}$, we can bound this by applying Lemma 6.12. Noting that the bound obtained in this way is independent of k, we conclude that

$$\mathbf{P}(D_\theta(B) < \varepsilon) \leq M \sum_{n=1}^\infty 2^n \exp\big(-c 2^{(2\theta-1)n}\varepsilon^{-2}\big) \leq \tilde M \sum_{n=1}^\infty \exp\big(-\tilde c n \varepsilon^{-2}\big) .$$

Here, we used the fact that as soon as $\theta > \frac{1}{2}$, we can find constants K and $\tilde c$ such that

$$n \log 2 - c 2^{(2\theta-1)n}\varepsilon^{-2} \leq K - \tilde c n \varepsilon^{-2} ,$$

uniformly over all $\varepsilon < 1$ and all $n \geq 1$. (Consider separately the cases $\varepsilon^2 \in (0, 1/n)$ and $\varepsilon^2 \in [1/n, 1)$.) We deduce from this the bound

$$\mathbf{P}(D_\theta(B) < \varepsilon) \leq M\Big(e^{-\tilde c \varepsilon^{-2}} + \int_1^\infty \exp\big(-\tilde c \varepsilon^{-2} x\big)\, dx\Big) ,$$

which immediately implies the result. □

Note that the proof given above is quite robust. In particular, we did not really make use of the fact that B has independent increments. In fact, it transpires that all that is required in order to prove the Hölder roughness of sample paths of a Gaussian process W with stationary increments is a small ball estimate of the type

$$\mathbf{P}\Big(\sup_{t\in[0,\delta]}|W_t - W_0| \le \varepsilon\Big) \le C\exp(-c\delta^\alpha\varepsilon^{-\beta})\ ,$$

for some exponents $\alpha, \beta > 0$. These kinds of estimates are available for example for fractional Brownian motion with arbitrary Hurst parameter $H \in (0,1)$.

6.6 Exercises

Exercise 6.13. Show that the Q-Wiener process (as introduced in Exercise 3.16) is truly rough.

Exercise 6.14. Prove and state precisely: multidimensional fractional Brownian motion B^H, $H \in (1/3, 1/2]$, is truly rough. **Hint:** A *law of iterated logarithm* for fractional Brownian motion of the form

$$\mathbf{P}\left[\overline{\lim_{h\downarrow 0}}\frac{\left|B^H_{t,t+h}\right|}{h^H(\ln\ln 1/h)^{1/2}} = \sqrt{2}\right] = 1$$

holds, cf. for example [MR06, Thm 7.2.15].

Exercise 6.15. In (6.7), estimate $\mathrm{osc}(Z,\varepsilon)$ by $2\|Y\|_\infty$ (or alternatively by $\|Y\|_\alpha\varepsilon^\alpha$) and deduce the estimate

$$\|Z'\|_\infty \le \frac{1}{L}\inf_{\varepsilon\in(0,\varepsilon_0]}\big(2\varepsilon^{-\theta}\|Y\|_\infty + \big\|R^Z\big\|_{2\alpha}\varepsilon^{2\alpha-\theta}\big).$$

Carry out the elementary optimization, e.g. when $\varepsilon_0 = T/2$, to see that

$$\|Z'\|_\infty \le \frac{4\|Y\|_\infty}{L(\theta,X)}\Big(\big\|R^Z\big\|_{2\alpha}^{\frac{\theta}{2\alpha}}\|Y\|_\infty^{-\frac{\theta}{2\alpha}} \vee T^{-\theta}\Big).$$

Exercise 6.16 (Norris lemma for rough paths; [HP13]). Give a complete proof of Theorem 6.10.

6.7 Comments

The notion of θ-roughness was first introduced in Hairer–Pillai [HP13], which also contains Proposition 6.8, although some of the ideas underlying the concepts presented here were already apparent in Baudoin–Hairer and Hairer–Mattingly [BH07, HM11]. A version of this “Norris lemma” in the context of SDEs driven by fractional Brownian motion was proposed independently by Hu–Tindel [HT13]. The simplified condition of “true” roughness (which may be verified in infinite dimensions), targeted directly at a Doob–Meyer decomposition, is taken from Friz–Shekhar

[FS12a]; the quantitative “Norris lemma” is taken from Cass, Litterer, Hairer and Tindel [CHLT12]. These results also hold in “rougher” situations, i.e. when $\alpha \leq 1/3$, [FS12a, CHLT12].

Chapter 7
Operations on controlled rough paths

Abstract At first sight, the notation $\int Y\,dX$ introduced in Chapter 4 is ambiguous since the resulting controlled rough path depends in general on the choices of both the second-order process $\mathbb{X}$ and the derivative process Y'. Fortunately, this "lack of completeness" in our notations is mitigated by the fact that in virtually all situations of interest, Y is constructed by using a small number of elementary operations described in this chapter. For all of these operations, it turns out to be intuitively rather clear how the corresponding derivative process is constructed.

7.1 Relation between rough paths and controlled rough paths

Consider $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha([0,T], V)$. It is easy to see that X itself can be interpreted as a path "controlled by X". Indeed, we can identify X with the element $(X, I) \in \mathscr{D}_X^{2\alpha}$, where I is the identity matrix (more precisely: the constant path with value I for all times). Conversely, an element $(Y, Y') \in \mathscr{D}_X^{2\alpha}([0,T], W)$ can itself be interpreted as a rough path again, say $\mathbf{Y} = (Y, \mathbb{Y})$. Indeed, with the interpretation of the integral in the sense of (4.22), below fully spelled out for the reader's convenience, we can set

$$\mathbb{Y}_{s,t} = \int_s^t Y_{s,r} \otimes dY_r \stackrel{\text{def}}{=} \lim_{|\mathcal{P}|\to 0} \int_{\mathcal{P}} \Xi \;, \quad \Xi_{u,v} = Y_u \otimes Y_{u,v} + Y'_u \otimes Y'_u \mathbb{X}_{u,v} \;.$$

where $Y'_u \otimes Y'_u \in \mathcal{L}(V \otimes V, W \otimes W)$ is given by $(Y'_u \otimes Y'_u)(v \otimes \tilde{v}) = (Y'_u(v)) \otimes (Y'_u(\tilde{v}))$. The fact that $\|\mathbb{Y}\|_{2\alpha}$ is finite is then a consequence of (4.23). On the other hand, the algebraic relations (2.1) already hold for the "Riemann sum" approximations to the three integrals, provided that the partition used for the approximation of $\mathbb{Y}_{s,t}$ is the union of the one used for the approximation of $\mathbb{Y}_{s,u}$ with the one used for $\mathbb{Y}_{u,t}$.

We summarise the above consideration in saying that for every fixed $\mathbf{X} \in \mathscr{C}^\alpha([0,T], V)$, we have a continuous canonical injection

$$\mathscr{D}_X^{2\alpha}([0,T],W) \hookrightarrow \mathscr{C}^\alpha([0,T],W) \;.$$

Furthermore, this interpretation of elements of $\mathscr{D}_X^{2\alpha}$ as elements of $\mathscr{C}^\alpha$ is coherent in terms of the theory of integration constructed in the previous section, as can be seen by the following result:

Proposition 7.1. *Let $(X,\mathbb{X}) \in \mathscr{C}^\alpha$, let $(Y,Y') \in \mathscr{D}_X^{2\alpha}$, and let $\mathbf{Y} = (Y,\mathbb{Y}) \in \mathscr{C}^\alpha$ be the associated rough path constructed as above. If $(\tilde Z, \tilde Z') \in \mathscr{D}_Y^{2\alpha}$, then $(Z,Z') \in \mathscr{D}_X^{2\alpha}$, where $Z_t = \tilde Z_t$ and $Z'_t = \tilde Z'_t Y'_t$. Furthermore, one has the identity*

$$\int_0^t Z_s dY_s = \int_0^t \tilde Z_s d\mathbf{Y}_s \;. \tag{7.1}$$

Here, the left hand side uses (4.22) to define the integral of two controlled rough paths against each other and the right hand side uses the original definition (4.19) of the integral of a controlled rough path against its reference path.

Proof. By assumption, one has $Y_{s,t} = Y'_s X_{s,t} + \mathrm{O}(|t-s|^{2\alpha})$ and $\tilde Z_{s,t} = Z'_s Y_{s,t} + \mathrm{O}(|t-s|^{2\alpha})$. Combining these identities, it follows immediately that

$$Z_{s,t} = \tilde Z'_s Y'_s X_{s,t} + \mathrm{O}(|t-s|^{2\alpha}) = Z'_s X_{s,t} + \mathrm{O}(|t-s|^{2\alpha}) \;,$$

so that $(Z,Z') \in \mathscr{D}_X^{2\alpha}$ as required. Now the left hand side of (7.1) is given by $\mathcal{I}\Xi_{0,t}$ with $\Xi_{s,t} = Z_s Y_{s,t} + Z'_s Y'_s \mathbb{X}_{s,t}$, whereas the right hand side is given by $\mathcal{I}\tilde\Xi_{0,t}$, where we set $\tilde\Xi_{s,t} = \tilde Z_s \tilde Y_{s,t} + \tilde Z'_s \mathbb{Y}_{s,t}$. Since $|\mathbb{Y}_{s,t} - Y'_s Y'_s \mathbb{X}_{s,t}| \le C|t-s|^{3\alpha}$ by (4.20), the claim now follows from Remark 4.12. □

Remark 7.2. It is straightforward to see that if $\frac{1}{3} < \beta < \alpha$, then $\mathscr{C}^\alpha \hookrightarrow \mathscr{C}^\beta$ and, for every $X \in \mathscr{C}^\alpha$, we have a canonical embedding $\mathscr{D}_X^{2\alpha} \hookrightarrow \mathscr{D}_X^{2\beta}$. Furthermore, in view of the definition (4.10) of $\mathcal{I}$, the values of the integrals defined above do not depend on the interpretation of the integrand and integrator as elements of one or the other space.

7.2 Lifting of regular paths.

There is a canonical embedding $\iota\colon \mathcal{C}^{2\alpha} \hookrightarrow \mathscr{D}_X^{2\alpha}$ given by $\iota Y = (Y,0)$, since in this case $R_{s,t} = Y_{s,t}$ does indeed satisfy $\|R\|_{2\alpha} < \infty$. Recall that we are only interested in the case $\alpha \le \frac{1}{2}$. After all, if $Y_{s,t} = \mathrm{O}(|t-s|^{2\alpha})$ with $\alpha > \frac{1}{2}$, then Y has a vanishing derivative and must be constant.

7.3 Composition with regular functions.

Let W and $\bar{W}$ be two Banach spaces and let $\varphi\colon W \to \bar{W}$ be a function in $\mathcal{C}_b^2$. Let furthermore $(Y, Y') \in \mathscr{D}_X^{2\alpha}([0,T], W)$ for some $X \in \mathcal{C}^\alpha$. (In applications X will be part of some $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ but this is irrelevant here.) Then one can define a (candidate) controlled rough path $(\varphi(Y), \varphi(Y)') \in \mathscr{D}_X^{2\alpha}([0,T], \bar{W})$ by

$$\varphi(Y)_t = \varphi(Y_t) , \qquad \varphi(Y)'_t = D\varphi(Y_t)Y'_t . \tag{7.2}$$

It is straightforward to check that the corresponding remainder term does indeed satisfy the required bound. It is also straightforward to check that, as a consequence of the chain rule, this definition is consistent in the sense that $(\varphi \circ \psi)(Y, Y') = \varphi(\psi(Y, Y'))$. We have

Lemma 7.3. *Let $\varphi \in \mathcal{C}_b^2, (Y, Y') \in \mathscr{D}_X^{2\alpha}([0,T], W)$ for some $X \in \mathcal{C}^\alpha$ with $|Y'_0| + \|Y, Y'\|_{X,2\alpha} \le M \in [1, \infty)$. Let $(\varphi(Y), \varphi(Y)') \in \mathscr{D}_X^{2\alpha}([0,T], \bar{W})$ be given by (7.2). Then, there exists a constant C depending only on $T > 0$ and $\alpha > \frac{1}{3}$ such that one has the bound*

$$\big\|\varphi(Y), \varphi(Y)'\big\|_{X,2\alpha} \le C_{\alpha,T} M \|\varphi\|_{\mathcal{C}_b^2} (1 + \|X\|_\alpha)^2 \Big(|Y'_0| + \|Y, Y'\|_{X,2\alpha}\Big).$$

At last, C can be chosen uniformly over $T \in (0,1]$.

Proof. We have $\big(\varphi(Y), \varphi(Y)'\big) = (\varphi(Y_\cdot), D\varphi(Y_\cdot)Y'_\cdot) \in \mathscr{D}_X^{2\alpha}$. Indeed,

$$\begin{aligned}
\|\varphi(Y_\cdot)\|_\alpha &\le \|D\varphi\|_\infty \|Y_\cdot\|_\alpha \\
\big\|\varphi(Y)'\big\|_\alpha &\le \|D\varphi(Y_\cdot)\|_\infty \|Y'_\cdot\|_\alpha + \|Y'_\cdot\|_\infty \|D\varphi(Y_\cdot)\|_\alpha \\
&\le \|D\varphi(Y_\cdot)\|_\infty \|Y'_\cdot\|_\alpha + \|Y'_\cdot\|_\infty \big\|D^2\varphi(Y_\cdot)\big\|_\infty \|Y_\cdot\|_\alpha ,
\end{aligned}$$

which shows that $\varphi(Y), \varphi(Y)' \in \mathcal{C}^\alpha$. Furthermore, $R^\varphi \equiv R^{\varphi(Y)}$ is given by

$$\begin{aligned}
R^\varphi_{s,t} &= \varphi(Y_t) - \varphi(Y_s) - D\varphi(Y_s)Y'_s X_{s,t} \\
&= \varphi(Y_t) - \varphi(Y_s) - D\varphi(Y_s)Y_{s,t} + D\varphi(Y_s)R^Y_{s,t}
\end{aligned}$$

so that,

$$\|R^\varphi\|_{2\alpha} \le \frac{1}{2}\big|D^2\varphi\big|_\infty \|Y\|_\alpha^2 + |D\varphi|_\infty \big\|R^Y\big\|_{2\alpha}.$$

It follows that

$$\begin{aligned}
\big\|\varphi(Y), \varphi(Y)'\big\|_{X,2\alpha} &\le \|D\varphi(Y_\cdot)\|_\infty \|Y'_\cdot\|_\alpha + \|Y'_\cdot\|_\infty \big\|D^2\varphi(Y_\cdot)\big\|_\infty \|Y_\cdot\|_\alpha \\
&\quad + \frac{1}{2}\big|D^2\varphi\big|_\infty \|Y\|_\alpha^2 + |D\varphi|_\infty \big\|R^Y\big\|_{2\alpha} \\
&\le \|\varphi\|_{\mathcal{C}_b^2} \Big(\|Y'_\cdot\|_\alpha + \|Y'_\cdot\|_\infty \|Y_\cdot\|_\alpha + \|Y\|_\alpha^2 + \big\|R^Y\big\|_{2\alpha}\Big) \\
&\le C_{\alpha,T} \|\varphi\|_{\mathcal{C}_b^2} (1 + \|X\|_\alpha)^2 \Big(1 + |Y'_0| + \|Y, Y'\|_{X,2\alpha}\Big)
\end{aligned}$$

$$\times \Big(|Y'_0| + \|Y,Y'\|_{X,2\alpha} \Big) ,$$

where we used in particular (4.18). □

It follows immediately that one has the following "Leibniz rule", the proof of which is left to the reader:

Corollary 7.4. *Let (Y,Y') and (Z,Z') be two controlled paths in $\mathscr{D}_X^{2\alpha}$ for some $X \in \mathcal{C}^\alpha$. Then the path $U = YZ$, with Gubinelli derivative $U' = YZ' + ZY'$ also belongs to $\mathscr{D}_X^{2\alpha}$.*

7.4 Stability II: Regular functions of controlled rough paths

We now investigate the continuity properties of the controlled rough path constructed in Lemma 7.3. In doing so, we shall use notation previously introduced in Section 4.4.

Theorem 7.5 (Stability of composition). *Let $\mathbf{X} = \big(X, \mathbb{X}\big)$, $\tilde{\mathbf{X}} = \big(\tilde{X}, \tilde{\mathbb{X}}\big) \in \mathscr{C}^\alpha$, $\big(Y,Y'\big) \in \mathscr{D}_X^{2\alpha}$, $\big(\tilde{Y},\tilde{Y}'\big) \in \mathscr{D}_{\tilde{X}}^{2\alpha}$. For $\varphi \in \mathcal{C}_b^3$ define*

$$(Z,Z') := (\varphi(Y), D\varphi(Y)Y') \in \mathscr{D}_X^{2\alpha} \tag{7.3}$$

and $\big(\tilde{Z},\tilde{Z}'\big)$ similarly. Then, one has the local Lipschitz estimates

$$\begin{aligned} d_{X,\tilde{X},2\alpha}\big(Z,Z';\tilde{Z},\tilde{Z}'\big) \le C_M \Big(&\varrho_\alpha\big(\mathbf{X},\tilde{\mathbf{X}}\big) + \big|Y_0 - \tilde{Y}_0\big| + \big|Y'_0 - \tilde{Y}'_0\big| \\ &+ d_{X,\tilde{X},2\alpha}\big(Y,Y';\tilde{Y},\tilde{Y}'\big)\Big) , \end{aligned} \tag{7.4}$$

as well as

$$\big\|Z - \tilde{Z}\big\|_\alpha \le C_M \Big(\varrho_\alpha\big(\mathbf{X},\tilde{\mathbf{X}}\big) + \big|Y_0 - \tilde{Y}_0\big| + \big|Y'_0 - \tilde{Y}'_0\big| + d_{X,\tilde{X},2\alpha}\big(Y,Y';\tilde{Y},\tilde{Y}'\big)\Big) , \tag{7.5}$$

for a suitable constant $C_M = C(M,T,\alpha,\varphi)$.

Proof. (The reader is urged to revisit Lemma 7.3 where the composition (7.3) was seen to be well-defined for $\varphi \in \mathcal{C}_b^2$.) Similar as in the previous proof, noting that

$$\big|Z'_0 - \tilde{Z}'_0\big| = \big|D\varphi(Y_0)Y'_0 - D\varphi\big(\tilde{Y}_0\big)\tilde{Y}'_0\big| \le C_M \big(\big|Y_0 - \tilde{Y}_0\big| + \big|Y'_0 - \tilde{Y}'_0\big| \big)$$

it suffices to establish the first estimate, for (7.5) is an immediate consequence of (7.4) and (4.27). In order to establish the first estimate we need to bound

$$\big\|D\varphi(Y)Y' - D\varphi\big(\tilde{Y}\big)\tilde{Y}'\big\|_\alpha + \big\|R^Z - R^{\tilde{Z}}\big\|_{2\alpha}.$$

Write $C_M(\varepsilon_X + \varepsilon_0 + \varepsilon'_0 + \varepsilon)$ for the right hand side of (7.4). Note that with this notation, from (4.27),

$$\|Y - \tilde{Y}\|_\alpha \lesssim \varepsilon_X + \varepsilon_0' + \varepsilon =: \varepsilon_Y ,$$

and also $\|Y - \tilde{Y}\|_{\infty;[0,T]} \lesssim \varepsilon_0 + \varepsilon_Y$ (uniformly over $T \le 1$). Since $D\varphi \in \mathcal{C}_b^2$, we know from Lemma 8.2 that

$$\begin{aligned}\|D\varphi(\tilde{Y}) - D\varphi(Y)\|_{\mathcal{C}^\alpha} &= |D\varphi(\tilde{Y}_0) - D\varphi(Y_0)| + \|D\varphi(\tilde{Y}) - D\varphi(Y)\|_\alpha \\ &\le C(\varepsilon_0 + \varepsilon_Y)\end{aligned}$$

where C depends on the $\mathcal{C}_b^3$-norm of φ. Also, $\|Y' - \tilde{Y}'\|_{\mathcal{C}^\alpha} \le \varepsilon_0' + \varepsilon$. Clearly then ($\mathcal{C}^\alpha$ is a Banach algebra under pointwise multiplication), we have, for a constant C_M,

$$\begin{aligned}\|D\varphi(Y)Y' - D\varphi(\tilde{Y})\tilde{Y}'\|_\alpha &\le C_M(\varepsilon_0 + \varepsilon_Y + \varepsilon_0' + \varepsilon) \\ &\lesssim C_M(\varepsilon_X + \varepsilon_0 + \varepsilon_0' + \varepsilon) .\end{aligned}$$

To deal with $R^Z - R^{\tilde{Z}}$, write

$$\begin{aligned}R^Z_{s,t} &= \varphi(Y_t) - \varphi(Y_s) - D\varphi(Y_s)Y_s'X_{s,t} \\ &= \varphi(Y_t) - \varphi(Y_s) - D\varphi(Y_s)Y_{s,t} + D\varphi(Y_s)R^Y_{s,t}.\end{aligned}$$

Taking the difference with $R^{\tilde{Z}}$ (replace Y, Y', R^Y above by $\tilde{Y}, \tilde{Y}', R^{\tilde{Y}}$) leads to the bound $|R^Z_{s,t} - R^{\tilde{Z}}_{s,t}| \le T_1 + T_2$ where

$$\begin{aligned}T_1 &:= \varphi(Y_t) - \varphi(Y_s) - D\varphi(Y_s)Y_{s,t} - \big(\varphi(\tilde{Y}_t) - \varphi(\tilde{Y}_s) - D\varphi(\tilde{Y}_s)\tilde{Y}_{s,t}\big) \\ &= \int_0^1 \Big(D^2\varphi(Y_s + \theta Y_{s,t})(Y_{s,t}, Y_{s,t}) - D^2\varphi(\tilde{Y}_s + \theta\tilde{Y}_{s,t})(\tilde{Y}_{s,t}, \tilde{Y}_{s,t})\Big)(1-\theta)d\theta \\ T_2 &:= D\varphi(Y_s)R^Y_{s,t} - D\varphi(\tilde{Y}_s)R^{\tilde{Y}}_{s,t} .\end{aligned}$$

As for the second term, we know $R^Y_{s,t} - R^{\tilde{Y}}_{s,t} \le (\varepsilon_0' + \varepsilon)|t-s|^{2\alpha}$, for all s, t, while

$$|D\varphi(\tilde{Y}_s) - D\varphi(Y_s)| \le \|D^2\varphi\|_\infty |\tilde{Y}_s - Y_s| \le \|D^2\varphi\|_\infty (\varepsilon_0 + \varepsilon_Y).$$

By elementary estimates of the form $|ab - \tilde{a}\tilde{b}| \le |a|\,|b - \tilde{b}| + |a - \tilde{a}|\,|\tilde{b}|$ it then follows immediately that one has $T_2 \le C(\varepsilon_X + \varepsilon_0 + \varepsilon_0' + \varepsilon)|t-s|^{2\alpha}$.

One argues similarly for the first term. This time, we consider the expression under the above integral $\int(\ldots)(1-\theta)d\theta$ for fixed integration variable $\theta \in [0,1]$. Using $Y^n \to Y$ in α-Hölder norm, we obtain

$$\begin{aligned}|D^2\varphi(\tilde{Y}_s + \theta\tilde{Y}_{s,t}) - D^2\varphi(Y_s + \theta Y_{s,t})| &\le \|D^3\varphi\|_\infty\big(|\tilde{Y}_s - Y_s| + |\tilde{Y}_{s,t} - Y_{s,t}|\big) \\ &\le 3\|D^3\varphi\|_\infty \|\tilde{Y} - Y\|_\infty \lesssim \varepsilon_0 + \varepsilon_Y ,\end{aligned}$$

noting that this estimate is uniform in $s, t \in [0,T]$ and $\theta \in [0,1]$. It then suffices to insert/subtract $D^2\varphi(Y_s + \theta Y_{s,t})(\tilde{Y}_{s,t}, \tilde{Y}_{s,t})$ under the integral $\int \ldots (1-\theta)d\theta$ appearing in the definition of T_1 and conclude with the triangle inequality and some simple estimates, keeping in mind that $\|Y - \tilde{Y}\|_\alpha \le \varepsilon_Y$ and $\|Y\|_\alpha, \|\tilde{Y}\|_\alpha \lesssim C_M$. $\square$

7.5 Itô's formula revisited

Let $F : V \to W$ in $\mathcal{C}_b^3$, $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ and $(Y, Y') \in \mathscr{D}_X^{2\alpha}$ a controlled rough path of the form

$$Y_t = \int_0^t Y'_s \, d\mathbf{X}_s + \Gamma_t \;, \tag{7.6}$$

for some controlled rough path $(Y', Y'') \in \mathscr{D}_X^{2\alpha}$ and some path $\Gamma \in \mathcal{C}^{2\alpha}$. This is the case for rough integrals of 1-forms, cf. Section 4.2, and also if Y is the solution to a rough differential equation driven by $\mathbf{X}$, to be discussed in Section 8.1.

In this situation, in analogy with the "usual" Itô formula, we would expect that

$$\begin{aligned} F(Y_t) = F(Y_0) + \int_0^t DF(Y_s) Y'_s d\mathbf{X}_s + \int_0^t DF(Y_s)\, d\Gamma_s \\ + \frac{1}{2} \int_0^t D^2 F(Y_s)\big(Y'_s, Y'_s\big)\, d[\mathbf{X}]_s \;, \end{aligned} \tag{7.7}$$

which is meaningful if we interpret the last two integrals as Young integrals. To show that this is indeed the case, note first that a consequence of (7.6) and Theorem 4.10, the increments of Y are of the form

$$Y_{s,t} = Y'_s X_{s,t} + Y''_s \mathbb{X}_{s,t} + \Gamma_{s,t} + \mathrm{o}(|t-s|) \;. \tag{7.8}$$

Furthermore, by Lemma 7.3 and Corollary 7.4, the path $G := DF(Y)Y'$ is controlled by X, with $G' = D^2F(Y)(Y', Y') + DF(Y)Y''$, so that the rough integral

$$\int_0^t DF(Y_s) Y'_s d\mathbf{X}_s = \int_0^t G_s d\mathbf{X}_s = \lim_{|P| \to 0} \sum_{[u,v] \in P} G_u X_{u,v} + G'_u \mathbb{X}_{u,v} \;, \tag{7.9}$$

which is the first term in the above identity, makes sense as a rough integral. Note that, if $\mathbf{X} = \mathbf{B}$, Itô enhanced Brownian motion, and Y, Y', Y'' are all adapted, then so is G and the integral is identified, by Proposition 5.1, as a classical Itô integral.

Proposition 7.6. *Under the assumption (7.8), the Itô formula (7.7) holds true.*

Proof. By the (previous) Itô formula, we know that $F(Y_t) - F(Y_0)$ equals

$$\lim_{|\mathcal{D}| \to 0} \sum_{[u,v] \in \mathcal{D}} \big(DF(Y_u) Y_{u,v} + D^2 F(Y_u) \mathbb{Y}_{u,v} \big) + \lim_{|\mathcal{D}| \to 0} \sum_{[u,v] \in \mathcal{D}} D^2 F(Y_u) [\mathbf{Y}]_{u,v} \tag{7.10}$$

where $\mathbb{Y}_{u,v} = \int_u^v Y_{u,\cdot} \otimes dY$ in the sense of remark 4.11, noting that $\mathbb{Y}_{u,v} = Y'_u Y'_u \mathbb{X}_{u,v} + \mathrm{o}(|v-u|)$. Also,

$$
\begin{aligned}
[\mathbf{Y}]_{u,v} &= Y_{u,v} \otimes Y_{u,v} - 2\,\mathrm{Sym}\,(\mathbb{Y}_{u,v}) \\
&= Y'_u Y'_u (X_{u,v} \otimes X_{u,v} - 2\,\mathrm{Sym}\,(\mathbb{X}_{u,v})) + \mathrm{o}(|v-u|) \\
&= Y'_u Y'_u [\mathbf{X}]_{u,v} + \mathrm{o}(|v-u|).
\end{aligned}
$$

Let us also subtract/add $DF(Y_u)Y''_u \mathbb{X}_{u,v}$ from (7.10). Then $F(Y_t) - F(Y_0)$ equals

$$
\begin{aligned}
&\lim_{|\mathcal{D}|\to 0} \sum_{[u,v]\in\mathcal{D}} \big(DF(Y_u)(Y_{u,v} - Y''_u \mathbb{X}_{u,v}) + DF(Y_u)Y''_u \mathbb{X}_{u,v} + D^2F(Y_u)Y'_u Y'_u \mathbb{X}_{u,v}\big) \\
&\qquad + \lim_{|\mathcal{D}|\to 0} \sum_{[u,v]\in\mathcal{D}} D^2F(Y_u)Y'_u Y'_u [\mathbf{X}]_{u,v} \\
&\quad = \lim_{|\mathcal{D}|\to 0} \sum_{[u,v]\in\mathcal{D}} DF(Y_u)Y'_u X_{u,v} + \big(DF(Y_u)Y''_u + D^2F(Y_u)Y'_u Y'_u\big)\mathbb{X}_{u,v} \\
&\qquad + \lim_{|\mathcal{D}|\to 0} \sum_{[u,v]\in\mathcal{D}} DF(Y_u)\Gamma_{u,v} + \int_0^t D^2F(Y_u)Y'Y'_u d[\mathbf{X}]_u \,.
\end{aligned}
$$

In view of (7.9), also noting the appearance of two Young integrals in the last line, the proof is complete. □

7.6 Controlled rough paths of low regularity

Let us conclude this section by showing how these canonical operations can be lifted to the case of controlled rough paths of low regularity, i.e. when $\alpha < \frac{1}{3}$. Recall from Section 4.5 that in this case we view a controlled rough path Y as a $T^{(p-1)}(\mathbf{R}^d)$-valued function, which is controlled by increments of **X** in the sense of Definition 4.17.

This suggests that, in order to define the product of two controlled rough paths Y and $\bar{Y}$, we should first ask ourselves how a product of the type $\mathbf{X}^w_{s,t}\mathbf{X}^{\bar{w}}_{s,t}$ for two different words w a $\bar{w}$ can be rewritten as a linear combination of the increments of **X**. It was realised by Chen [Che54] that such a product is described by the *shuffle product*. Recall that, for any alphabet $\mathcal{A}$, the shuffle product $\sqcup\!\sqcup$ is defined on the free algebra over $\mathcal{A}$ by considering all possible ways of interleaving two words in ways that preserve the original order of the letters. For example, if a, b and c are letters in $\mathcal{A}$, one has the identity

$$
ab \sqcup\!\sqcup\, ac = abac + 2aabc + 2aacb + acab \,.
$$

In our case, the choice of basis described earlier then defines a natural algebra homomorphism $w \mapsto e_w$ from the free algebra over $\{1, \ldots, d\}$ into $T^{(p)}(\mathbf{R}^d)$, and we denote by $\star$ the corresponding product. In other words, we have the identity

$$
e_w \star e_{\bar{w}} = e_{w \sqcup\!\sqcup \bar{w}} \,,
$$

where, if w is a linear combination of words, e_w is the corresponding linear combination of basis vectors.

With this definition at hand, it turns out that any geometric rough path **X** satisfies the identity

$$\mathbf{X}^w_{s,t}\,\mathbf{X}^{\bar w}_{s,t} = \mathbf{X}^{w \sqcup\mathchoice{}{}{}{}\sqcup \bar w}_{s,t}\;.$$

This strongly suggests that the "correct" way of multiplying two controlled rough paths Y and $\bar Y$ is to define their product Z by

$$Z_t = Y_t \star \bar Y_t\;.$$

It is possible to check that Z is indeed again a controlled rough path. Similarly, if F is a smooth function and Y is a controlled rough path, we define $F(Y)$ by

$$F(Y)_t \stackrel{\text{def}}{=} F(Y_t^\phi) + \sum_{k=1}^{p-1} \frac{1}{k!} F^{(k)}(Y_t^\phi)\,\tilde Y_t^{\star k}\;,$$

where $F^{(k)}$ denotes the kth derivative of F and $\tilde Y_t \stackrel{\text{def}}{=} Y_t - Y_t^\phi$ is the part describing the "local fluctuations" of Y.

It is again possible to show that $F(Y)$ is a controlled rough path if Y is a controlled rough path and F is sufficiently smooth. (It should be of class $\mathcal{C}^p_b$.) See [Hai14c] for an extremely general setting in which a similar calculus is still useful.

7.7 Exercises

Exercise 7.7. Verify that $\mathbb{X}_{s,t} = \int_s^t X_{s,r} \otimes dX_r$ where the integral is to be interpreted in the sense of (4.22), taking (Y, Y') to be (X, I). In fact, check that this holds not only in the limit $|\mathcal{P}| \to 0$ but in fact for every fixed $|\mathcal{P}|$, i.e. $\mathbb{X}_{s,t} = \int_{\mathcal{P}} \Xi$. Compare this with formula (2.12), obtained in Exercise 2.7.

Exercise 7.8. Let $\varphi\colon W \times [0,T] \to \bar W$ be a function which is uniformly $\mathcal{C}^2$ in its first argument (i.e. φ is bounded and both $D_y\varphi$ and $D_y^2\varphi$ are bounded, where D_y denotes the Fréchet derivative with respect to the first argument) and uniformly $\mathcal{C}^{2\alpha}$ in its second argument. Let furthermore $(Y, Y') \in \mathscr{D}^{2\alpha}_X([0,T], W)$. Show that

$$\varphi(Y)_t = \varphi(Y_t, t)\;, \qquad \varphi(Y)'_t = D_y\varphi(Y_t, t) Y'_t\;.$$

defines an element $(\varphi(Y), \varphi(Y)') \in \mathscr{D}^{2\alpha}_X([0,T], \bar W)$. In fact, show that there exists a constant C, depending only on T, such that one has the bound

$$\|\varphi(Y)\|_{X,2\alpha} \le C\big(\|D_y^2\varphi\|_\infty + \|\varphi\|_\infty + \|\varphi\|_{2\alpha;t}\big)\big(1 + \|X\|_\alpha\big)^2\big(1 + \|Y\|_{X,2\alpha}\big)^2\;,$$

where we denote by $\|\varphi\|_{2\alpha;t}$ the supremum over y of the 2α-Hölder norm of $\varphi(y,\cdot)$.

Exercise 7.9. Convince yourself that in the case $p = 2$, the definitions given in Section 7.6 coincide with the definitions given earlier in this section.

Chapter 8
Solutions to rough differential equations

Abstract We show how to solve differential equations driven by rough paths by a simple Picard iteration argument. This yields a pathwise solution theory mimicking the standard solution theory for ordinary differential equations. We start with the simple case of differential equations driven by a signal that is sufficiently regular for Young's theory of integration to apply and then proceed to the case of more general rough signals.

8.1 Introduction

We now turn our attention to (rough) differential equations of the form

$$dY_t = f(Y_t)\, dX_t\ , \qquad Y_0 = \xi \in W\ . \tag{8.1}$$

Here, $X : [0,T] \to V$ is the driving or input signal, while $Y : [0,T] \to W$ is the output signal. As usual V and W are Banach spaces, and $f : W \to \mathcal{L}(V,W)$. When $\dim V = d < \infty$, one may think of f as a collection of vector fields $(f_1, \ldots, f_d)$ on W. As usual, the reader is welcome to think $V = \mathbf{R}^d$ and $W = \mathbf{R}^n$ but there is really no difference in the argument. Such equations are familiar from the theory of ODEs, and more specifically, control theory, where X is typically assumed to be absolutely continuous so that $dX_t = \dot{X}_t\, dt$. The case of SDEs, *stochastic differential equations*, with dX interpreted as Itô or Stratonovich differential of Brownian motion, is also well known. Both cases will be seen as special examples of RDEs, *rough differential equations*.

We may consider (8.1) on the unit time interval. Indeed, equation (8.1) is invariant under time-reparametrization so that any (finite) time horizon may be rescaled to $[0,1]$. Alternatively, global solutions on a larger time horizon are constructed successively, i.e. by concatenating $Y|_{[0,1]}$ (started at Y_0) with $Y|_{[1,2]}$ (started at Y_1) and so on. As a matter of fact, we shall construct solutions by a variation of the classical *Picard iteration* on intervals $[0,T]$, where $T \in (0,1]$ will be chosen sufficiently small to guarantee invariance of suitable balls and the contraction property. Our key ingredients are estimates for rough integrals (cf. Theorem 4.10) and the composition of controlled paths with smooth maps (Lemma 7.3). Recall that, for rather trivial reasons (of the sort $|t-s|^{2\alpha} \le |t-s|$, when $0 \le s \le t \le T \le 1$), all constants in these estimates were seen to be uniform in $T \in (0,1]$.

8.2 Review of the Young case: a priori estimates

Let us postulate that there exists a solution to a differential equation in Young's sense and let us derive an a-priori estimate. (In finite dimension, this can actually be used to prove the existence of solutions. Note that the regularity requirement here is "one degree less" than what is needed for the corresponding uniqueness result.)

Proposition 8.1. *Assume $X, Y \in \mathcal{C}^\beta([0,1], V)$ for some $\beta \in (1/2, 1]$ such that, given $\xi \in W$, $f \in \mathcal{C}^1_b(W, \mathcal{L}(V, W))$, we have*

$$dY_t = f(Y_t)dX_t\,, \qquad Y_0 = \xi\,,$$

in the sense of a Young integral equation. Then

$$\|Y\|_\beta \le C\Big[\Big(\|f\|_{\mathcal{C}^1_b}\|X\|_\beta\Big) \vee \Big(\|f\|_{\mathcal{C}^1_b}\|X\|_\beta\Big)^{1/\beta}\Big].$$

Proof. By assumption, for $0 \le s < t \le 1$, $Y_{s,t} = \int_s^t f(Y_r)dX_r$. Using Young's inequality (4.3), with $C = C(\beta)$,

$$\begin{aligned}|Y_{s,t} - f(Y_s)X_{s,t}| &= \left|\int_s^t (f(Y_r) - f(Y_s))dX_r\right| \\ &\le C\|Df\|_\infty \|Y\|_{\beta;[s,t]}\|X\|_{\beta;[s,t]}|t-s|^{2\beta}\end{aligned}$$

so that

$$|Y_{s,t}|/|t-s|^\beta \le \|f\|_\infty\|X\|_\beta + C\|Df\|_\infty\|Y\|_{\beta;[s,t]}\|X\|_{\beta;[s,t]}|t-s|^\beta.$$

Write $\|Y\|_{\beta;h} \equiv \sup |Y_{s,t}|/|t-s|^\beta$ where the sup is restricted to times $s,t \in [0,1]$ for which $t - s \le h$. Clearly then,

$$\|Y\|_{\beta;h} \le \|f\|_\infty\|X\|_\beta + C\|Df\|_\infty\|Y\|_{\beta;h}\|X\|_\beta h^\beta$$

and upon taking h small enough, s.t. $\delta h^\beta \asymp 1$, with $\delta = \|X\|_\beta$, more precisely s.t.

$$C\|Df\|_\infty\|X\|_\beta h^\beta \le C\Big(1 + \|f\|_{\mathcal{C}^1_b}\Big)\|X\|_\beta h^\beta \le 1/2$$

(we will take h such that the second $\leq$ becomes an equality; adding 1 avoids trouble when $f \equiv 0$)

$$\frac{1}{2}\|Y\|_{\beta;h} \leq \|f\|_\infty \|X\|_\beta .$$

It then follows from Exercise 4.24 that, with $h \propto \|X\|_\beta^{-1/\beta}$,

$$\begin{aligned}\|Y\|_\beta &\leq \|Y\|_{\beta;h}\Big(1 \vee h^{-(1-\beta)}\Big) \leq C\|X\|_\beta\Big(1 \vee h^{-(1-\beta)}\Big)\\ &= C\Big(\|X\|_\beta \vee \|X\|_\beta^{1/\beta}\Big) .\end{aligned}$$

Here, we have absorbed the dependence on $f \in \mathcal{C}_b^1$ into the constants. By scaling (any non-zero f may be normalised to $\|f\|_{\mathcal{C}_b^1} = 1$ at the price of replacing X by $\|f\|_{\mathcal{C}_b^1} \times X$) we then get immediately the claimed estimate. $\square$

8.3 Review of the Young case: Picard iteration

The reader may be helped by first reviewing the classical Picard argument in a Young setting, i.e. when $\beta \in (1/2, 1]$. Given $\xi \in W$, $f \in \mathcal{C}_b^2(W, \mathcal{L}(V, W))$, $X \in \mathcal{C}^\beta([0,1], V)$ and $Y : [0,T] \to W$ of suitable Hölder regularity, $T \in (0,1]$, one defines the map $\mathcal{M}_T$ by

$$\mathcal{M}_T(Y) := \left(\xi + \int_0^t f(Y_s)dX_s : t \in [0,T]\right) .$$

Following a classical pattern of proof, we shall establish invariance of suitable balls, and then a contraction property upon taking $T = T_0$ small enough. The resulting unique fixed point is then obviously the unique solution to (8.1) on $[0, T_0]$. The unique solution on $[0, 1]$ is then constructed successively, i.e. by concatenating the solution Y on $[0, T_0]$, started at $Y_0 = \xi$, with the solution Y on $[T_0, 2T_0]$ started at Y_{T_0} and so on. Care is necessary to ensure that T_0 can be chosen uniformly; for instance, if f were only $\mathcal{C}^2$ (without the boundedness assumption) one can still obtain local existence on $[0, T_1]$, and then $[T_1, T_2]$, etc, but explosion may happen at some finite time $\lim_n T_n$ within our time horizon $[0, 1]$. The situation here is completely analogous to what we are familiar with from the usual theory of nonlinear ODEs.

We will need the Hölder norm of X over $[0, T]$ to tend to zero as $T \downarrow 0$. Now, as the example of the $t \mapsto t$ and $\beta = 1$ shows, this may not be true relative to the β-Hölder norm; the (cheap) trick is to take $\alpha \in (1/2, \beta)$ and to view $\mathcal{M}_T$ as map from the Banach space $\mathcal{C}^\alpha([0,T], W)$, rather than $\mathcal{C}^\beta([0,T], W)$, into itself. Young's inequality is still applicable since all paths involved will be (at least) α-Hölder continuous with $\alpha > 1/2$. On the other hand,

$$\|X\|_{\alpha;[0,T]} \leq T^{\beta-\alpha}\|X\|_{\beta;[0,T]} ,$$

and so the α-Hölder norm of X has the desired behaviour. As previously, when no confusion is possible, we write $\|\cdot\|_\alpha \equiv \|\cdot\|_{\alpha;[0,T]}$.

To avoid norm versus semi-norm considerations, it is convenient to work on the space of paths started at ξ, namely $\{Y \in \mathcal{C}^\alpha([0,T],W) : Y_0 = \xi\}$. This affine subspace is a complete metric space under $\big(Y,\tilde{Y}\big) \mapsto \big\|Y-\tilde{Y}\big\|_\alpha$ and so is the closed unit ball

$$\mathcal{B}_T = \{Y \in \mathcal{C}^\alpha([0,T],W) : Y_0 = \xi, \|Y\|_\alpha \le 1\} .$$

Young's inequality (4.32) shows that there is a constant C which only depends on α (thanks to $T \le 1$) such that for every $Y \in \mathcal{B}_T$,

$$\begin{aligned}\|\mathcal{M}_T(Y)\|_\alpha &\le C(|f(Y_0)| + \|f(Y)\|_\alpha)\|X\|_\alpha \\ &\le C(|f(\xi)| + \|Df\|_\infty \|Y\|_\alpha)\|X\|_\alpha \\ &\le C(|f|_\infty + \|Df\|_\infty)\|X\|_\alpha \le C|f|_{\mathcal{C}_b^1}\|X\|_\beta T^{\beta-\alpha} .\end{aligned}$$

Similarly, for $Y,\tilde{Y} \in \mathcal{B}_T$, using Young, $f\big(Y_0\big) = f\big(\tilde{Y}_0\big)$ and Lemma 8.2 below (with $K=1$)

$$\begin{aligned}\Big\|\mathcal{M}_T\big(Y\big) - \mathcal{M}_T\big(\tilde{Y}\big)\Big\|_\alpha &= \left\|\int_0^\cdot f\big(Y_s\big)dX_s - \int_0^\cdot f\big(\tilde{Y}_s\big)dX_s\right\|_\alpha \\ &\le C\Big(\big|f\big(Y_0\big) - f\big(\tilde{Y}_0\big)\big| + \big\|f\big(Y\big) - f\big(\tilde{Y}\big)\big\|_\alpha\Big)\|X\|_\alpha \\ &\le C\|f\|_{\mathcal{C}_b^2}\|X\|_\beta T^{\beta-\alpha}\big\|Y-\tilde{Y}\big\|_\alpha .\end{aligned}$$

It is clear from the previous estimates that a small enough $T_0 = T_0(f,\alpha,\beta,X) \le 1$ can be found such that $\mathcal{M}_{T_0}(\mathcal{B}_{T_0}) \subset \mathcal{B}_{T_0}$ and, for all $Y,\tilde{Y} \in \mathcal{B}_{T_0}$,

$$\big\|\mathcal{M}_{T_0}\big(Y\big) - \mathcal{M}_{T_0}\big(\tilde{Y}\big)\big\|_{\alpha;[0,T_0]} \le \frac{1}{2}\big\|Y-\tilde{Y}\big\|_{\alpha;[0,T_0]} .$$

Therefore, $\mathcal{M}_{T_0}(\cdot)$ admits a unique fixed point $Y \in \mathcal{B}_{T_0}$ which is the unique solution Y to (8.1) on the (small) interval $[0,T_0]$. Noting that the choice $T_0 = T_0(f,\alpha,\beta,X)$ can indeed be done uniformly (in particular it does not change when the starting point ξ is replaced by Y_{T_0}), the unique solution on $[0,1]$ is then constructed iteratively, as explained in the beginning.

Lemma 8.2. *Assume $f \in \mathcal{C}_b^2(W,\bar{W})$ and $T \le 1$. Then there exists a $C_{\alpha,K}$ such that for all $X,Y \in \mathcal{C}^\alpha$ with $\|X\|_{\alpha;[0,T]}, \|Y\|_{\alpha;[0,T]} \le K \in [1,\infty)$*

$$\|f(X) - f(Y)\|_{\alpha;[0,T]} \le C_{\alpha,K}\|f\|_{\mathcal{C}_b^2}\Big(|X_0 - Y_0| + \|X-Y\|_{\alpha;[0,T]}\Big).$$

Proof. Consider the difference

$$f(X)_{s,t} - f(Y)_{s,t} = (f(X_t) - f(Y_t)) - (f(X_s) - f(Y_s)).$$

The idea is to use a *division property* of sufficiently smooth functions. In the present context, this simply means that one has

$$f(x) - f(y) = g(x,y)(x-y) \quad \text{with} \quad g(x,y) := \int_0^1 Df(tx + (1-t)y)\, dt\,,$$

where $g : W \times W \to \mathcal{L}(W, \bar{W})$ is obviously bounded by $|Df|_\infty$ and in fact Lipschitz with $|g|_{Lip} \le C|D^2 f|_\infty$ for some constant $C \ge 1$ relative to any product norm on $W \times W$, such as $|(x,y)|_{W\times W} = |x| + |y|$. It follows that

$$|(g(x,y) - g(\tilde{x},\tilde{y}))| \le |g|_{Lip}|(x-\tilde{x}, y-\tilde{y})| \le C|D^2 f|_\infty(|x-\tilde{x}| + |y-\tilde{y}|).$$

Setting $\Delta_t = X_t - Y_t$ then allows to write

$$\begin{aligned}
|f(X)_{s,t} - f(Y)_{s,t}| &= |g(X_t,Y_t)\Delta_t - g(X_s,Y_s)\Delta_s| \\
&= |g(X_t,Y_t)(\Delta_t - \Delta_s) + (g(X_t,Y_t) - g(X_s,Y_s))\Delta_s| \\
&\le |g|_\infty |X_{s,t} - Y_{s,t}| + |g|_{Lip}|(X_{s,t},Y_{s,t})|_{W\times W}|X_s - Y_s| \\
&\le |Df|_\infty |X_{s,t} - Y_{s,t}| + C|D^2 f|_\infty(|X_{s,t}| + |Y_{s,t}|)\|X-Y\|_{\infty;[0,T]} \\
&\lesssim |t-s|^\alpha\big(|Df|_\infty\|X-Y\|_\alpha + K|D^2 f|_\infty\|X-Y\|_{\infty;[0,T]}\big)\,.
\end{aligned}$$

Since $T \le 1$ we can also estimate $\|X-Y\|_{\infty;[0,T]} \le |X_0 - Y_0| + \|X-Y\|_{\alpha;[0,T]}$ and the claimed estimate on $f(X) - f(Y)$ follows immediately. $\square$

8.4 Rough differential equations: a priori estimates

We now consider a priori estimates for rough differential equations, similar to Section 8.2. Recall that the homogeneous rough path norm $\|\!|\mathbf{X}|\!\|_\alpha$ was introduced in (2.4).

Proposition 8.3. *Let $\xi \in W$, $f \in \mathcal{C}_b^2(W, \mathcal{L}(V,W))$ and a rough path $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ with $\alpha \in (1/3, 1/2]$ and assume that $(Y, Y') = (Y, f(Y)) \in \mathscr{D}_X^{2\alpha}$ is a RDE solution to $dY = f(Y)\, d\mathbf{X}$ started at $Y_0 = \xi \in W$. That is, for all $t \in [0,T]$,*

$$Y_t = \xi + \int_0^t f(Y_s)\, d\mathbf{X}_s\,, \tag{8.2}$$

where the integral is interpreted in the sense of Theorem 4.10 and $f(Y) \in \mathscr{D}_X^{2\beta}$ is built from Y by Lemma 7.3. (Thanks to $\mathcal{C}_b^2$-regularity of f and Lemma 7.3 the above rough integral equation (8.2) is well-defined.[1])

Then the following (a priori) estimate holds true

$$\|Y\|_\alpha \le C\Big[\Big(\|f\|_{\mathcal{C}_b^2}\|\!|\mathbf{X}|\!\|_\alpha\Big) \vee \Big(\|f\|_{\mathcal{C}_b^2}\|\!|\mathbf{X}|\!\|_\alpha\Big)^{1/\alpha}\Big]$$

[1] Later we will establish existence and uniqueness under $\mathcal{C}_b^3$-regularity.

where $C = C(\alpha)$ is a suitable constant.

Proof. Consider an interval $I := [s,t]$ so that, using basic estimates for rough integrals (cf. Theorem 4.10),

$$\begin{aligned} \left|R^Y_{s,t}\right| &= |Y_{s,t} - f(Y_s)X_{s,t}| \\ &\le \left|\int_s^t f(Y)dX - f(Y_s)X_{s,t} - Df(Y_s)f(Y_s)\mathbb{X}_{s,t}\right| + |Df(Y_s)f(Y_s)\mathbb{X}_{s,t}| \\ &\lesssim \Big(\|X\|_{\alpha;I}\big\|R^{f(Y)}\big\|_{2\alpha;I} + \|\mathbb{X}\|_{2\alpha;I}\|f(Y)\|_{\alpha;I}\Big)|t-s|^{3\alpha} \\ &\quad + \|\mathbb{X}\|_{2\alpha;I}|t-s|^{2\alpha}. \end{aligned} \tag{8.3}$$

Recall that $\|\cdot\|_\alpha$ is the usual Hölder semi-norm over $[0,T]$, while $\|\cdot\|_{\alpha;I}$ denotes the same norm, but over $I \subset [0,T]$, so that trivially $\|X\|_{\alpha;I} \le \|X\|_\alpha$. Whenever notationally convenient, multiplicative constants depending on α and f are absorbed in $\lesssim$, at the very end we can use scaling to make the f dependence reappear. We will also write $\|\cdot\|_{\alpha;h}$ for the supremum of $\|\cdot\|_{\alpha;I}$ over all intervals $I \subset [0,T]$ with length $|I| \le h$. Again, one trivially has $\|X\|_{\alpha;I} \le \|X\|_{\alpha;h}$ whenever $|I| \le h$. Using this notation, we conclude from (8.3) that

$$\big\|R^Y\big\|_{2\alpha;h} \lesssim \|\mathbb{X}\|_{2\alpha;h} + \Big(\|X\|_{\alpha;h}\big\|R^{f(Y)}\big\|_{2\alpha;h} + \|\mathbb{X}\|_{2\alpha;h}\|f(Y)\|_{\alpha;h}\Big)h^\alpha .$$

We would now like to relate $R^{f(Y)}$ to R^Y. As in the proof of Lemma 7.3, we obtain the bound

$$\begin{aligned} R^{f(Y)}_{s,t} &= f(Y_t) - f(Y_s) - Df(Y_s)Y'_s X_{s,t} \\ &= f(Y_t) - f(Y_s) - Df(Y_s)Y_{s,t} + Df(Y_s)R^Y_{s,t} \end{aligned}$$

so that,

$$\begin{aligned} \big\|R^{f(Y)}\big\|_{2\alpha;h} &\le \frac{1}{2}\big|D^2 f\big|_\infty \|Y\|^2_{\alpha;h} + |Df|_\infty \big\|R^Y\big\|_{2\alpha;h} \\ &\lesssim \|Y\|^2_{\alpha;h} + \big\|R^Y\big\|_{2\alpha;h}. \end{aligned}$$

Hence, also using $\|f(Y)\|_{\alpha;h} \lesssim \|Y\|_{\alpha;h}$, there exists $c_1 > 0$, not dependent on $\mathbf{X}$ or Y, such that

$$\begin{aligned} \big\|R^Y\big\|_{2\alpha;h} &\le c_1\|\mathbb{X}\|_{2\alpha;h} + c_1\|X\|_{\alpha;h}h^\alpha\|Y\|^2_{\alpha;h} \\ &\quad + c_1\|X\|_{\alpha;h}h^\alpha\big\|R^Y\big\|_{2\alpha;h} + c_1\|\mathbb{X}\|_{2\alpha;h}h^\alpha\|Y\|_{\alpha;h} \,. \end{aligned} \tag{8.4}$$

We now restrict ourselves to h small enough so that $\|\!|\mathbf{X}|\!\|_\alpha h^\alpha \ll 1$. More precisely, we choose it such that

$$c_1\|X\|_\alpha h^\alpha \le \frac{1}{2}\,, \qquad c_1\|\mathbb{X}\|^{1/2}_{2\alpha} h^\alpha \le \frac{1}{2}\,.$$

Inserting this bound into (8.4), we conclude that

$$\|R^Y\|_{2\alpha;h} \le c_1\|\mathbb{X}\|_{2\alpha;h} + \frac{1}{2}\|Y\|^2_{\alpha;h} + \frac{1}{2}\|R^Y\|_{2\alpha;h} + \|\mathbb{X}\|^{1/2}_{2\alpha;h}\|Y\|_{\alpha;h} \ .$$

This in turn yields the bound

$$\begin{aligned} \|R^Y\|_{2\alpha;h} &\le 2c_1\|\mathbb{X}\|_{2\alpha;h} + \|Y\|^2_{\alpha;h} + 2\|\mathbb{X}\|^{1/2}_{2\alpha;h}\|Y\|_{\alpha;h} \\ &\le c_2\|\mathbb{X}\|_{2\alpha;h} + 2\|Y\|^2_{\alpha;h} \ , \end{aligned} \tag{8.5}$$

with $c_2 = (2c_1 + 1)$. On the other hand, since $Y_{s,t} = f(Y_s)X_{s,t} - R^Y_{s,t}$ and f is bounded, we have the bound

$$\|Y\|_{\alpha;h} \lesssim \|X\|_\alpha + \|R^Y\|_{2\alpha;h}h^\alpha \ .$$

Combining this bound with (8.5) yields

$$\begin{aligned} \|Y\|_{\alpha;h} &\le c_3\|X\|_\alpha + c_3\|\mathbb{X}\|_{2\alpha;h}h^\alpha + c_3\|Y\|^2_{\alpha;h}h^\alpha \\ &\le c_3\|X\|_\alpha + c_4\|\mathbb{X}\|^{1/2}_{2\alpha;h} + c_3\|Y\|^2_{\alpha;h}h^\alpha \ , \end{aligned}$$

for some constant c_3. Multiplication with $c_3 h^\alpha$ then yields, with $\psi_h := c_3\|Y\|_{\alpha;h}h^\alpha$ and $\lambda_h := c_5 |||\mathbf{X}|||_\alpha h^\alpha \to 0$ as $h \to 0$,

$$\psi_h \le \lambda_h + \psi_h^2.$$

Clearly, for all h small enough depending on Y (so that $\psi_h \le 1/2$) $\psi_h \le \lambda_h + \psi_h/2$ implies $\psi_h \le 2\lambda_h$ and so

$$\|Y\|_{\alpha;h} \le c_6 |||\mathbf{X}|||_\alpha.$$

To see that this is true for all h small enough without dependence on Y, pick h_0 small enough so that $\lambda_{h_0} < 1/4$. It then follows that for each $h \le h_0$, one of the following two estimates must hold true

$$\begin{aligned} \psi_h &\ge \psi_+ \equiv \frac{1}{2} + \sqrt{\frac{1}{4} - \lambda_h} \ge \frac{1}{2} \\ \psi_h &\le \psi_- \equiv \frac{1}{2} - \sqrt{\frac{1}{4} - \lambda_h} = \frac{1}{2}\Big(1 - \sqrt{1 - 4\lambda_h}\Big) \sim \lambda_h \text{ as } h \downarrow 0. \end{aligned}$$

(In fact, for reasons that will become apparent shortly, we may decrease h_0 further to guarantee that for $h < h_0$ we have not only $\psi_h < 1/2$ but $\psi_h < 1/6$.) We already know that we are in the regime of the second estimate above as $h \downarrow 0$. Noting that $\psi_h(< 1/6) < 1/2$ in the second regime, the only reason that could prevent us from being in the second regime for all $h < h_0$ is an (upwards) jump of the (increasing) function $(0, h_0] \ni h \mapsto \psi_h$. But $\psi_h \le 3\lim_{g\uparrow h}\psi_g$, as seen from

$$\|Y\|_{\alpha;h} \le 3\|Y\|_{\alpha;h/3} \le 3\lim_{g\uparrow h}\|Y\|_{\alpha;g} \ ,$$

(and similarly: $\lim_{g\downarrow h}\psi_g \le 3\psi_h$) which rules out any jumps of relative jump size greater than 3. However, given that $\psi_h \ge 1/2$ in the first regime and $\psi_h < 1/6$ in the second, we can never jump from the second into the first regime, as h increases (from zero). And so, we indeed must be in the second regime for all $h \le h_0$. Elementary estimates on ψ_-, as function of λ_h then show that

$$\|Y\|_{\alpha;h} \le c_6 \|\mathbf{X}\|_\alpha \,,$$

for all $h \le h_0 \sim \|\mathbf{X}\|^{-1/\alpha}$. We conclude with Exercise 4.24, arguing exactly as in the Young case, Proposition 8.1. $\square$

8.5 Rough differential equations

The aim of this section is to show that if f is regular enough and $(X,\mathbb{X}) \in \mathscr{C}^\beta$ with $\beta > \frac{1}{3}$, then we can solve differential equations driven by the rough path $\mathbf{X} = (X,\mathbb{X})$ of the type

$$dY = f(Y)\,d\mathbf{X}\,.$$

Such an equation will yield solutions in $\mathscr{D}_X^{2\alpha}$ and will be interpreted in the corresponding integral formulation, where the integral of $f(Y)$ against X is defined using Lemma 7.3 and Theorem 4.10. More precisely, one has the following result:

Theorem 8.4. *Given $\xi \in W$, $f \in \mathcal{C}^3(W,\mathcal{L}(V,W))$ and $\mathbf{X} = (X,\mathbb{X}) \in \mathscr{C}^\beta(\mathbf{R}_+,V)$ with $\beta \in (\frac{1}{3},\frac{1}{2})$, there exists a unique element $(Y,Y') \in \mathscr{D}_X^{2\beta}([0,1],W)$ such that*

$$Y_t = \xi + \int_0^t f(Y_s)\,d\mathbf{X}_s\,, \qquad t < \tau\,, \tag{8.6}$$

for some $\tau > 0$. Here, the integral is interpreted in the sense of Theorem 4.10 and $f(Y) \in \mathscr{D}_X^{2\beta}$ is built from Y by Lemma 7.3. Furthermore, one has $Y' = f(Y)$ and, if $f \in \mathcal{C}_b^3$, solutions are global in time.

Proof. With $\mathbf{X} = (X,\mathbb{X}) \in \mathscr{C}^\beta \subset \mathscr{C}^\alpha$, $\frac{1}{3} < \alpha < \beta$ and $(Y,Y') \in \mathscr{D}_X^{2\alpha}$ we know from Lemma 7.3 that

$$(\Xi,\Xi') := \big(f(Y), f(Y)'\big) := (f(Y), Df(Y)Y') \in \mathscr{D}_X^{2\alpha}\,.$$

Restricting from $[0,1]$ to $[0,T]$, any $T \le 1$, Theorem 4.10 allows to define the map

$$\mathcal{M}_T(Y,Y') \stackrel{\text{def}}{=} \left(\xi + \int_0^\cdot \Xi_s d\mathbf{X}_s, \Xi\right) \in \mathscr{D}_X^{2\alpha}\,.$$

The RDE solution on $[0,T]$ we are looking for is a fixed point of this map. Strictly speaking, this would only yield a solution (Y,Y') in $\mathscr{D}_X^{2\alpha}$. But since $\mathbf{X} \in \mathscr{C}^\beta$, it turns out that this solution is automatically an element of $\mathscr{D}_X^{2\beta}$. Indeed, $|Y_{s,t}| \le$

$|Y'|_\infty |X_{s,t}| + \|R^Y\|_{2\alpha} |t-s|^{2\alpha}$, so that $Y \in \mathcal{C}^\beta$. From the fixed point property it then follows that $Y' = f(Y) \in \mathcal{C}^\beta$ and also $R^Y \in \mathcal{C}_2^{2\beta}$, since $\mathbb{X} \in \mathcal{C}_2^{2\beta}$ and

$$\begin{aligned}\left|R^Y_{s,t}\right| = \left|Y_{s,t} - Y'_s X_{s,t}\right| &= \left|\int_s^t (f(Y_r) - f(Y_s)) d\mathbf{X}_t\right| \\ &\le |Y'|_\infty |\mathbb{X}_{s,t}| + \mathrm{O}\big(|t-s|^{3\alpha}\big)\ .\end{aligned}$$

Note that if (Y, Y') is such that $(Y_0, Y'_0) = (\xi, f(\xi))$, then the same is true for $\mathcal{M}_T(Y, Y')$. Therefore, $\mathcal{M}_T$ can be viewed as map on the space of controlled paths started at $(\xi, f(\xi))$, i.e.

$$\big\{(Y, Y') \in \mathscr{D}_X^{2\alpha}([0,T], W) : Y_0 = \xi, Y'_0 = f(\xi)\big\}\ .$$

Since $\mathscr{D}_X^{2\alpha}$ is a Banach space (under the norm $(Y, Y') \mapsto |Y_0| + |Y'_0| + \|Y, Y'\|_{X,2\alpha}$) the above (affine) subspace is a complete metric space under the induced metric. This is also true for the (closed) unit ball $\mathcal{B}_T$ centred at, say

$$t \mapsto (\xi + f(\xi) X_{0,t}, f(\xi)).$$

(Note here that the apparently simpler choice $t \mapsto \big(\xi, f(\xi)\big)$ does in general not belong to $\mathscr{D}_X^{2\alpha}$.) In other words, $\mathcal{B}_T$ is the set of all $(Y, Y') \in \mathscr{D}_X^{2\alpha}([0,T], W)$: $Y_0 = \xi, Y'_0 = f(\xi)$ and

$$\begin{aligned}|Y_0 - \xi| + |Y'_0 - f(\xi)| + \|(Y - (\xi + f(\xi) X_{0,\cdot}), Y'_\cdot - f(\xi))\|_{X,2\alpha}& \\ = \|(Y - f(\xi) X_{0,\cdot}, Y'_\cdot - f(\xi))\|_{X,2\alpha} &\le 1.\end{aligned}$$

In fact, $\|(Y - f(\xi) X_{0,\cdot}, Y'_\cdot - f(\xi))\|_{X,2\alpha} = \|Y, Y'_\cdot\|_{X,2\alpha}$ as a consequence of the triangle inequality and $\|(f(\xi) X_{0,\cdot}, f(\xi))\|_{X,2\alpha} = \|f(\xi)\|_\alpha + \|0\|_{2\alpha} = 0$, so that

$$\mathcal{B}_T = \Big\{(Y, Y') \in \mathscr{D}_X^{2\alpha}([0,T], W) : Y_0 = \xi, Y'_0 = f(\xi) : \|(Y, Y'_\cdot)\|_{X,2\alpha} \le 1\Big\}.$$

Let us also note that, for all $(Y, Y') \in \mathcal{B}_T$, one has the bound

$$\left|Y'_0\right| + \|(Y, Y')\|_{X,2\alpha} \le |f|_\infty + 1 =: M \in [1, \infty). \tag{8.7}$$

We now show that, for T small enough, $\mathcal{M}_T$ leaves $\mathcal{B}_T$ invariant and in fact is contracting. Constants below are denoted by C, may change from line to line and may depend on $\alpha, \beta, X, \mathbb{X}$ without special indication. They are, however, uniform in $T \in (0,1]$ and we prefer to be explicit (enough) with respect to f such as to see where $\mathcal{C}_b^3$-regularity is used. With these conventions, we recall the following estimates, direct consequences from Lemma 7.3 and Theorem 4.10 , respectively,

$$\|\Xi, \Xi'\|_{X,2\alpha} \le CM \|f\|_{\mathcal{C}_b^2} \Big(|Y'_0| + \|Y, Y'\|_{X,2\alpha}\Big)$$

$$
\begin{aligned}
\left\| \int_0^{\cdot} \Xi_s d\mathbf{X}_s, \Xi \right\|_{X,2\alpha} &\leq \|\Xi\|_\alpha + \|\Xi'\|_\infty \|\mathbb{X}\|_{2\alpha} \\
&\quad + C\big(\|X\|_\alpha \big\|R^\Xi\big\|_{2\alpha} + \|\mathbb{X}\|_{2\alpha}\|\Xi'\|_\alpha\big) \\
&\leq \|\Xi\|_\alpha + C\Big(|\Xi'_0| + \|\Xi, \Xi'\|_{X,2\alpha}\Big)(\|X\|_\alpha + \|\mathbb{X}\|_{2\alpha}) \\
&\leq \|\Xi\|_\alpha + C\Big(|\Xi'_0| + \|\Xi, \Xi'\|_{X,2\alpha}\Big)T^{\beta-\alpha}.
\end{aligned}
$$

Invariance: For $(Y, Y') \in \mathcal{B}_T$, noting that $\|\Xi\|_\alpha = \|f(Y)\|_\alpha \leq \|f\|_{\mathcal{C}_b^1}\|Y\|_\alpha$ and that $|\Xi'_0| = |Df(Y_0)Y'_0| \leq \|f\|^2_{\mathcal{C}_b^1}$, we obtain the bound

$$
\begin{aligned}
\big\|\mathcal{M}_T(Y, Y')\big\|_{X,2\alpha} &= \left\| \int_0^{\cdot} \Xi_s d\mathbf{X}_s, \Xi \right\|_{X,2\alpha} \\
&\leq \|\Xi\|_\alpha + C\Big(|\Xi'_0| + \|\Xi, \Xi'\|_{X,2\alpha}\Big)T^{\beta-\alpha} \\
&\leq \|f\|_{\mathcal{C}_b^1}\|Y\|_\alpha + C\Big(\|f\|^2_{\mathcal{C}_b^1} + CM\|f\|_{\mathcal{C}_b^2}\Big(|Y'_0| + \|Y, Y'\|_{X,2\alpha}\Big)\Big)T^{\beta-\alpha} \\
&\leq \|f\|_{\mathcal{C}_b^1}(\|f\|_\infty + 1)T^{\beta-\alpha} + CM\Big(\|f\|^2_{\mathcal{C}_b^1} + \|f\|_{\mathcal{C}_b^2}(\|f\|_\infty + 1)\Big)T^{\beta-\alpha} ,
\end{aligned}
$$

where in the last step we used (8.7) and also $\|Y\|_{\alpha;[0,T]} \leq C_f T^{\beta-\alpha}$, seen from

$$
\begin{aligned}
|Y_{s,t}| &\leq |Y'|_\infty |X_{s,t}| + \big\|R^Y\big\|_{2\alpha}|t-s|^{2\alpha} \\
&\leq (|Y'_0| + \|Y'\|_\alpha)\|X\|_\beta |t-s|^\beta + \big\|R^Y\big\|_{2\alpha}|t-s|^{2\alpha} .
\end{aligned}
$$

Then, using $T^\alpha \leq T^{\beta-\alpha}$ and $\big\|R^Y\big\|_{2\alpha} \leq \|Y, Y'\|_{X,2\alpha} \leq 1$, we obtain the bound

$$
\begin{aligned}
\|Y\|_{\alpha;[0,T]} &\leq \Big(|Y'_0| + \|Y, Y'\|_{X,2\alpha}\Big)\|X\|_\beta T^{\beta-\alpha} + \big\|R^Y\big\|_{2\alpha}T^{\beta-\alpha} \qquad (8.8) \\
&\leq \Big((\|f\|_\infty + 1)\|X\|_\beta + 1\Big)T^{\beta-\alpha} .
\end{aligned}
$$

In other words, $\|\mathcal{M}_T(Y, Y')\|_{X,2\alpha} = \|\mathcal{M}_T(Y, Y')\|_{X,2\alpha;[0,T]} = \mathrm{O}\big(T^{\beta-\alpha}\big)$ with constant only depending on α, β, **X** and $f \in \mathcal{C}_b^2$. By chosing $T = T_0$ small enough, we obtain the bound $\|\mathcal{M}_{T_0}(Y, Y')\|_{X,2\alpha;[0,T_0]} \leq 1$ so that $\mathcal{M}_{T_0}$ leaves $\mathcal{B}_{T_0}$ invariant, as desired.

Contraction: Setting $\Delta_s = f(Y_s) - f\big(\tilde{Y}_s\big)$ as a shorthand, we have the bound

$$
\begin{aligned}
\big\|\mathcal{M}_T\big(Y, Y'\big) - \mathcal{M}_T\big(\tilde{Y}, \tilde{Y}'\big)\big\|_{X,2\alpha} &= \left\| \int_0^{\cdot} \Delta_s d\mathbf{X}_s, \Delta \right\|_{X,2\alpha} \\
&\leq \|\Delta\|_\alpha + C\Big(|\Delta'_0| + \|\Delta, \Delta'\|_{X,2\alpha}\Big)T^{\beta-\alpha} \\
&\leq C\|f\|_{\mathcal{C}_b^2}\big\|Y - \tilde{Y}\big\|_\alpha + C\|\Delta, \Delta'\|_{X,2\alpha}T^{\beta-\alpha} .
\end{aligned}
$$

The contraction property is obvious, provided that we can establish the following two estimates:

$$\big\|Y - \tilde{Y}\big\|_\alpha \le CT^{\beta-\alpha}\big\|Y - \tilde{Y}, Y' - \tilde{Y}'\big\|_{X,2\alpha}\,, \tag{8.9}$$

$$\big\|\Delta, \Delta'\big\|_{X,2\alpha} \le C\big\|Y - \tilde{Y}, Y' - \tilde{Y}'\big\|_{X,2\alpha}\,. \tag{8.10}$$

To obtain (8.9), replace Y by $Y - \tilde{Y}$ in (8.8), noting $Y_0' - \tilde{Y}_0' = 0$, shows that

$$\begin{aligned}\big\|Y - \tilde{Y}\big\|_\alpha &\le \big\|Y' - \tilde{Y}'\big\|_\alpha \|X\|_\beta T^{\beta-\alpha} + \big\|R^Y - R^{\tilde{Y}}\big\|_{2\alpha} T^{\beta-\alpha}\\ &\le CT^{\beta-\alpha}\big\|Y - \tilde{Y}, Y' - \tilde{Y}'\big\|_{X,2\alpha}.\end{aligned}$$

We now turn to (8.10). Similar to the proof of Lemma 8.2, $f \in \mathcal{C}^3$ allows to write $\Delta_s = G_s H_s$ where

$$G_s := g\big(Y_s, \tilde{Y}_s\big)\;,\quad H_s := Y_s - \tilde{Y}_s\;,$$

and $g \in \mathcal{C}_b^2$ with $\|g\|_{\mathcal{C}_b^2} \le C\|f\|_{\mathcal{C}_b^3}$. Lemma 7.3 tells us that $(G, G') \in \mathscr{D}_X^{2\alpha}$ (with $G' = (D_Y g)Y' + (D_{\tilde{Y}} g)\tilde{Y}'$) and in fact immediately yields an estimate of the form

$$\|G, G'\|_{X,2\alpha} \le C\|f\|_{\mathcal{C}_b^3}\;,$$

uniformly over $\big(Y, Y'\big), \big(\tilde{Y}, \tilde{Y}'\big) \in \mathcal{B}_T$ and $T \le 1$. On the other hand, $\mathscr{D}_X^{2\alpha}$ is an algebra in the sense that $(GH, (GH)') \in \mathscr{D}_X^{2\alpha}$ with $(GH)' = G'H + GH'$. In fact, we leave it as easy exercise to the reader to check that

$$\begin{aligned}\|GH, (GH)'\|_{X,2\alpha} \lesssim &\Big(|G_0| + |G_0'| + \|G, G'\|_{X,2\alpha}\Big)\\ &\times \Big(|H_0| + |H_0'| + \|H, H'\|_{X,2\alpha}\Big)\,.\end{aligned}$$

In our situation, $H_0 = Y_0 - \tilde{Y}_0 = \xi - \xi = 0$, and similarly $H_0' = 0$, so that, for all $\big(Y, Y'\big), \big(\tilde{Y}, \tilde{Y}'\big) \in \mathcal{B}_T$, we have

$$\begin{aligned}\big\|\Delta, \Delta'\big\|_{X,2\alpha} &\lesssim \Big(|G_0| + |G_0'| + \|G, G'\|_{X,2\alpha}\Big)\|H, H'\|_{X,2\alpha}\\ &\lesssim \Big(\|g\|_\infty + \|g\|_{\mathcal{C}_b^1}\big(\big|Y_0'\big| + \big|\tilde{Y}_0'\big|\big) + C\|f\|_{\mathcal{C}_b^3}\Big)\big\|Y - \tilde{Y}, Y' - \tilde{Y}'\big\|_{X,2\alpha}\\ &\lesssim \big\|Y - \tilde{Y}, Y' - \tilde{Y}'\big\|_{X,2\alpha}\,,\end{aligned}$$

where we made use of $\|g\|_\infty, \|g\|_{\mathcal{C}_b^1} \lesssim \|f\|_{\mathcal{C}_b^3}$ and $|Y_0'| = \big|\tilde{Y}_0'\big| = |f(\xi)| \le |f|_\infty$.

The argument from here on is identical to the Young case: the previous estimates allow for a small enough $T_0 \le 1$ such that $\mathcal{M}_{T_0}(\mathcal{B}_{T_0}) \subset \mathcal{B}_{T_0}$ and for all $\big(Y, Y'\big), \big(\tilde{Y}, \tilde{Y}'\big) \in \mathcal{B}_{T_0}$:

$$\left\| \mathcal{M}_{T_0}(Y, Y') - \mathcal{M}_{T_0}(\tilde{Y}, \tilde{Y}') \right\|_{X,2\alpha} \le \frac{1}{2} \| Y - \tilde{Y}, Y' - \tilde{Y}' \|_{X,2\alpha}$$

and so $\mathcal{M}_{T_0}(\cdot)$ admits a unique fixed point $(Y, Y') \in \mathcal{B}_{T_0}$, which is then the unique solution Y to (8.1) on the (possibly rather small) interval $[0, T_0]$. Noting that the choice of T_0 can again be done uniformly in the starting point, the solution on $[0, 1]$ is then constructed iteratively as before. □

In many situations, one is interested in solutions to an equation of the type

$$dY = f_0(Y, t)\, dt + f(Y, t)\, d\mathbf{X}_t \,, \tag{8.11}$$

instead of (8.6). On the one hand, it is possible to recast (8.11) in the form (8.6) by writing it as an RDE for $\hat{Y}_t = (Y_t, t)$ driven by $\hat{\mathbf{X}}_t = (\hat{X}, \hat{\mathbb{X}})$ where $\hat{X} = (X_t, t)$ and $\hat{\mathbb{X}}$ is given by $\mathbb{X}$ and the "remaining cross integrals" of X_t and t, given by usual Riemann-Stieltjes integration. However, it is possible to exploit the structure of (8.11) to obtain somewhat better bounds on the solutions. See [FV10b, Ch. 12].

8.6 Stability III: Continuity of the Itô–Lyons map

We now obtain continuity of solutions to rough differential equations as function of their (rough) driving signals.

Theorem 8.5 (Rough path stability of the Itô–Lyons map). *Let $f \in \mathcal{C}_b^3$ and let $(Y, f(Y)) \in \mathscr{D}_X^{2\alpha}$ be the (unique) RDE solution given by Theorem 8.4 to*

$$dY = f(Y)\, d\mathbf{X}, \quad Y_0 = \xi \in W;$$

similarly, let $(\tilde{Y}, f(\tilde{Y}))$ be the RDE solution driven by $\tilde{\mathbf{X}}$ and started at ξ where $\mathbf{X}, \tilde{\mathbf{X}} \in \mathscr{C}^\beta$ and $\alpha < \beta$. Assuming

$$\|\mathbf{X}\|_\beta, \|\tilde{\mathbf{X}}\|_\beta \le M < \infty$$

we have the local Lipschitz estimates

$$d_{X,\tilde{X},2\alpha}(Y, f(Y); \tilde{Y}, f(\tilde{Y})) \le C_M \Big(|\xi - \tilde{\xi}| + \varrho_\beta(\mathbf{X}, \tilde{\mathbf{X}}) \Big) \,,$$

and also

$$\|Y - \tilde{Y}\|_\alpha \le C_M \Big(|\xi - \tilde{\xi}| + \varrho_\beta(\mathbf{X}, \tilde{\mathbf{X}}) \Big) \,,$$

where $C_M = C(M, \alpha, \beta, f)$ is a suitable constant.

Remark 8.6. The "loss" of Hölder regularity (the fact that we have two exponents satisfying $\alpha < \beta$) is not really necessary, but it allows for a quick proof.

Proof. Recall that, for given $\mathbf{X} \in \mathscr{C}^\beta$, th RDE solution $(Y, f(Y)) \in \mathscr{D}_X^{2\alpha}$ was constructed as fixed point of

$$\mathcal{M}_T(Y,Y') := (Z,Z') := \left(\xi + \int_0^{\cdot} f(Y_s)d\mathbf{X}_s, f(Y_\cdot)\right) \in \mathscr{D}_X^{2\alpha}$$

and similarly for $\tilde{\mathcal{M}}_T\big(\tilde{Y}, f\big(\tilde{Y}\big)\big) \in \mathcal{C}^{\alpha}_{\tilde{X}}$. Then, thanks to the fixed point property

$$(Y, f(Y)) = (Y, Y') = (Z, Z') = (Z, f(Y)),$$

(similarly with tilde) and the local Lipschitz estimate for rough integration (uniform in $T \le 1$) writing $(\Xi, \Xi') := \big(f(Y), f(Y)'\big)$ for the integrand,

$$\begin{aligned} d_{X,\tilde{X},2\alpha}\big(Y, f(Y); \tilde{Y}, f\big(\tilde{Y}\big)\big) &= d_{X,\tilde{X},2\alpha}\big(Z, Z'; \tilde{Z}, \tilde{Z}'\big) \\ &\lesssim \varrho_\alpha\big(\mathbf{X}, \tilde{\mathbf{X}}\big) + \big|\xi - \tilde{\xi}\big| + d_{X,\tilde{X},2\alpha}\big(\Xi, \Xi'; \tilde{\Xi}, \tilde{\Xi}'\big) \\ &\le \varrho_\beta\big(\mathbf{X}, \tilde{\mathbf{X}}\big) + \big|\xi - \tilde{\xi}\big| + d_{X,\tilde{X},2\beta}\big(\Xi, \Xi'; \tilde{\Xi}, \tilde{\Xi}'\big) \,, \end{aligned}$$

where we used $\alpha < \beta$ and $T \le 1$ in the last step. Thanks to the local Lipschitz estimate for composition (also uniform over $T \le 1$)

$$\begin{aligned} d_{X,\tilde{X},2\beta}\big(\Xi, \Xi'; \tilde{\Xi}, \tilde{\Xi}'\big) &\lesssim \varrho_\beta\big(\mathbf{X}, \tilde{\mathbf{X}}\big) + \big|\xi - \tilde{\xi}\big| + d_{X,\tilde{X},\beta}\big(Y, f(Y); \tilde{Y}, f\big(\tilde{Y}\big)\big) \\ &\le \varrho_\beta\big(\mathbf{X}, \tilde{\mathbf{X}}\big) + \big|\xi - \tilde{\xi}\big| + d_{X,\tilde{X},2\alpha}\big(Y, f(Y); \tilde{Y}, f\big(\tilde{Y}\big)\big)\, T^{\beta-\alpha} \,. \end{aligned}$$

In summary, for some constant $C = C(\alpha, \beta, f, M)$, we have the bound

$$\begin{aligned} d_{X,\tilde{X},2\alpha}\big(Y, f(Y); \tilde{Y}, f\big(\tilde{Y}\big)\big) \le C\big(&\varrho_\beta\big(\mathbf{X}, \tilde{\mathbf{X}}\big) + \big|\xi - \tilde{\xi}\big| \\ &+ d_{X,\tilde{X},2\alpha}\big(Y, f(Y); \tilde{Y}, f\big(\tilde{Y}\big)\big)T^{\beta-\alpha}\big) \,. \end{aligned}$$

By taking $T = T_0(M, \alpha, \beta, f)$ smaller, if necessary, we may assume that $CT^{\beta-\alpha} \le 1/2$, from which it follows that

$$d_{X,\tilde{X},2\alpha}\big(Y, f(Y); \tilde{Y}, f\big(\tilde{Y}\big)\big) \le 2C\big(\varrho_\beta\big(\mathbf{X}, \tilde{\mathbf{X}}\big) + \big|\xi - \tilde{\xi}\big|\big) \,,$$

which is precisely the required bound. $\square$

8.7 Davie's definition and numerical schemes

Fix $f \in \mathcal{C}_b^2(W, \mathcal{L}(V, W))$ and $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\beta([0,T], V)$ with $\beta > \frac{1}{3}$. Under these assumptions, the rough differential equation $dY = f(Y)d\mathbf{X}$ makes sense as well-defined integral equation. (In Theorem 8.4 we used additional regularity, namely $\mathcal{C}_b^3$, to establish existence of a *unique* solution on $[0,T]$.) By the very definition of an RDE solution, unique or not, $(Y, f(Y)) \in \mathscr{D}_X^{2\beta}$ i.e.

$$Y_{s,t} = f(Y_s)X_{s,t} + \mathrm{O}\big(|t-s|^{2\beta}\big)$$

and we recognise a step of *first-order Euler* approximation, $Y_{s,t} \approx f(Y_s)X_{s,t}$, started from Y_s. Clearly $\mathrm{O}\big(|t-s|^{2\beta}\big) = \mathrm{o}(|t-s|)$ if and only if $\beta > 1/2$ and one can show that iteration of such steps along a partition $\mathcal{P}$ of $[0,T]$ yields a convergent "Euler" scheme as $|\mathcal{P}| \downarrow 0$, see [Dav08] or [FV10b].

In the case $\beta \in \left(\frac{1}{3}, \frac{1}{2}\right]$ we have to exploit that we know more than just $(Y, f(Y)) \in \mathscr{D}_X^{2\beta}$. Indeed, since $Y_{s,t} = \int_s^t f(Y)dX$, estimate (4.20) for rough integrals tells us that, for all pairs s,t

$$Y_{s,t} = f(Y_s)X_{s,t} + (f(Y))'_s\mathbb{X}_{s,t} + \mathrm{O}\big(|t-s|^{3\beta}\big) \,. \tag{8.12}$$

Using the identity $f(Y)' = Df(Y)Y' = Df(Y)f(Y)$, this can be spelled out further to

$$Y_{s,t} = f(Y_s)X_{s,t} + Df(Y_s)f(Y_s)\mathbb{X}_{s,t} + \mathrm{o}(|t-s|) \tag{8.13}$$

and, omitting the small remainder term, we recognise a step of a *second-order Euler* or *Milstein* approximation. Again, one can show that iteration of such steps along a partition $\mathcal{P}$ of $[0,T]$ yields a convergent "Euler" scheme as $|\mathcal{P}| \downarrow 0$; see [Dav08] or [FV10b].

Remark 8.7. This schemes can be understood from simple Taylor expansions based on the differential equation $dY = f(Y)dX$, at least when X is smooth (enough), or via Itô's formula in a semi-martingale setting. With focus on the smooth case, the Euler approximation is obtained by a "left-point freezing" approximation $f(Y_\cdot) \approx f(Y_s)$ over $[s,t]$ in the integral equation,

$$Y_{s,t} = \int_s^t f(Y_r)dX_r \approx f(Y_s)X_{s,t}$$

whereas the Milstein scheme, with $\mathbb{X}_{s,t} = \int_s^t X_{s,r}dX_r$ for smooth paths, is obtained from the next-best approximation

$$\begin{aligned} f(Y_r) &\approx f(Y_s) + Df(Y_s)Y_{s,r} \\ &\approx f(Y_s) + Df(Y_s)f(Y_s)X_{s,r} \,. \end{aligned}$$

It turns out that the description (8.13) is actually a formulation that is equivalent to the RDE solution built previously in the following sense.

Proposition 8.8. *The following two statements are equivalent*

i) $(Y, f(Y))$ *is a RDE solution to (8.6), as constructed in Theorem 8.4.*
ii) $Y \in \mathcal{C}([0,T], W)$ *is an "RDE solution in the sense of Davie", i.e. in the sense of (8.13).*

Proof. We already discussed how (8.13) is obtained from an RDE solution to (8.6). Conversely, (8.13) implies immediately $Y_{s,t} = f(Y_s)X_{s,t} + \mathrm{O}\big(|t-s|^{2\beta}\big)$ which shows that $Y \in \mathcal{C}^\beta$ and also $Y' := f(Y) \in \mathcal{C}^\beta$, thanks to $f \in \mathcal{C}_b^2$, so that $(Y, f(Y)) \in \mathscr{D}_X^{2\beta}$. It remains to see, in the notation of the proof of Theorem 4.10, that $Y_{s,t} = (\mathcal{I}\Xi)_{s,t}$ with

$$\Xi_{s,t} = f(Y_s)X_{s,t} + (f(Y))'_s\mathbb{X}_{s,t} = f(Y_s)X_{s,t} + Df(Y_s)f(Y_s)\mathbb{X}_{s,t} \ .$$

To see this, we note that trivially $Y_{s,t} = (\mathcal{I}\tilde{\Xi})_{s,t}$ with $\tilde{\Xi}_{s,t} := Y_{s,t}$. But $\tilde{\Xi}_{s,t} = \Xi_{s,t} + o(|t-s|)$ and one sees as in Remark 4.12 that $\mathcal{I}\tilde{\Xi} = \mathcal{I}\Xi$. $\square$

8.8 Lyons' original definition

A slightly different notion of solution was originally introduced in [Lyo98] by Lyons.[2] This notion only uses the spaces $\mathscr{C}^\alpha$, without ever requiring the use of the spaces $\mathscr{D}_X^{2\alpha}$ of "controlled rough paths". Indeed, for $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha([0,T], V)$ and $F \in \mathcal{C}_b^2(V, \mathcal{L}(V,W))$ we can define an element $\mathbf{Z} = (Z, \mathbb{Z}) = I_F(X) \in \mathscr{C}^\alpha([0,T], W)$ directly by

$$\begin{aligned} Z_t &\stackrel{\text{def}}{=} \big(\mathcal{I}\Xi\big)_{0,t} \ , \quad \Xi_{s,t} = F(X_s)\, X_{s,t} + DF(X_s)\mathbb{X}_{s,t} \ , \\ \mathbb{Z}_{s,t} &\stackrel{\text{def}}{=} \big(\mathcal{I}\bar{\Xi}^s\big)_{s,t} \ , \quad \bar{\Xi}^s_{u,v} = Z_{s,u}\, Z_{u,v} + \big(F(X_u) \otimes F(X_u)\big)\mathbb{X}_{u,v} \ . \end{aligned}$$

It is possible to check that $\bar{\Xi}^s \in \mathcal{C}_2^{\alpha,3\alpha}$ for every fixed s (see the proof of Theorem 4.10) so that the second line makes sense. It is also straightforward to check that $(Z, \mathbb{Z})$ satisfies (2.1), so that it does indeed belong to $\mathscr{C}^\alpha$. Actually, one can see that

$$Z_t = \int_0^t F(X_s)\, d\mathbf{X}_s \ , \qquad \mathbb{Z}_{s,t} = \int_s^t Z_{s,r} \otimes dZ_r \ ,$$

where the integrals are defined as in the previous sections, where $F(X) \in \mathscr{D}_X^{2\alpha}$ as in Section 7.3.

We can now define solutions to (8.6) in the following way.

Definition 8.9. A rough path $\mathbf{Y} = (Y, \mathbb{Y}) \in \mathscr{C}^\alpha([0,T], W)$ is a solution in the sense of Lyons to (8.6) if there exists $\mathbf{Z} = (Z, \mathbb{Z}) \in \mathscr{C}^\alpha(V \oplus W)$ such that the projection of $(Z, \mathbb{Z})$ onto $\mathscr{C}^\alpha(V)$ is equal to $(X, \mathbb{X})$, the projection onto $\mathscr{C}^\alpha(W)$ is equal to $(Y, \mathbb{Y})$, and $Z = I_F(Z)$ where

$$F(x,y) = \begin{pmatrix} I & 0 \\ f(y) & 0 \end{pmatrix} .$$

It is straightforward to see that if $(Y, Y') \in \mathscr{D}_X^{2\alpha}(W)$ is a solution to (8.6) in the sense of the previous section, then the path $Z = X \oplus Y$ is controlled by X. As seen in Section 7.1, it can therefore be interpreted as an element of $\mathscr{C}^\alpha$. It follows immediately from the definitions that it is then also a solution in the sense of Lyons. Conversely, if $(Y, \mathbb{Y})$ is a solution in the sense of Lyons, then one can check that one

[2] As always, we only consider the step-2 α-Hölder case, i.e. $\alpha > \frac{1}{3}$, whereas Lyons' theory is valid for every Hölder-exponent $\alpha \in (0,1]$ (or: variation parameter $p \geq 1$) at the complication of heaving to deal with $\lfloor p \rfloor$ levels.

necessarily has $(Y, f(Y)) \in \mathscr{D}_X^{2\alpha}(W)$ and that this is a solution in the sense of the previous section. We leave the verification of this fact as an exercise to the reader.

8.9 Stability IV: Flows

We briefly state, without proof, a result concerning regularity of flows associated to rough differential equations, as well as local Lipschitz estimates of the Itô–Lyons maps on the level of such flows. More precisely, given a geometric rough path $\mathbf{X} \in \mathscr{C}_g^\alpha([0,T], \mathbf{R}^d)$, we saw in Theorem 8.4 that, for $\mathcal{C}_b^3$ vector fields $f = (f_1, \ldots, f_d)$ on $\mathbf{R}^e$, there is a unique global solution to the rough integral equation

$$Y_t = y + \int_0^t f(Y_s)\, d\mathbf{X}_s\ , \qquad t \geq 0\ . \tag{8.14}$$

Write $\pi_{(f)}(0, y; \mathbf{X}) = Y$ for this solution. Note that the inverse flow exists trivially, by following the RDE driven by $\mathbf{X}(. - t)$,

$$\pi_{(f)}(0, y; \mathbf{X})_t^{-1} = \pi_{(f)}(0, y; \mathbf{X}(. - t)_t.$$

We call the map $y \mapsto \pi_{(f)}(0, y; \mathbf{X})$ the flow associated to the above RDE. Moreover, if X^ϵ is a smooth approximation to $\mathbf{X}$ (in rough path metric), then the corresponding ODE solution Y^ϵ is close to Y, with a local Lipschitz estimate as given in Section 8.6.

It is natural to ask if the flow depends smoothly on y. Given a multi-index $k = (k_1, \ldots, k_e) \in \mathbf{N}^e$, write D^k for the partial derivative with respect to $y^1, \ldots, y^e$. The proof of the following statement is an easy consequence of [FV10b, Chapter 12].

Theorem 8.10. *Let $\alpha \in (1/3, 1/2]$ and $\mathbf{X}, \tilde{\mathbf{X}} \in \mathscr{C}_g^\alpha$. Assume $f \in \mathcal{C}_b^{3+n}$ for some integer n. Then the associated flow is of regularity $\mathcal{C}^{n+1}$ in y, as is its inverse flow. The resulting family of partial derivatives, $\{D^k \pi_{(f)}(0, \xi; \mathbf{X}), |k| \leq n\}$ satisfies the RDE obtained by formally differentiating $dY = f(Y) d\mathbf{X}$.*

At last, for every $M > 0$ there exist C, K depending on M and the norm of f such that, whenever $\|\!|\mathbf{X}|\!\|_\alpha, \|\!|\tilde{\mathbf{X}}|\!\|_\alpha \leq M < \infty$ and $|k| \leq n$,

$$\begin{aligned}
\sup_{\xi \in \mathbf{R}^e} \big| D^k \pi_{(f)}(0, \xi; \mathbf{X}) - D^k \pi_{(f)}(0, \xi; \tilde{\mathbf{X}}) \big|_{\alpha;[0,t]} &\leq C \varrho_\alpha(\mathbf{X}, \tilde{\mathbf{X}}), \\
\sup_{\xi \in \mathbf{R}^e} \big| D^k \pi_{(f)}(0, \xi; \mathbf{X})^{-1} - D^k \pi_{(f)}(0, \xi; \tilde{\mathbf{X}})^{-1} \big|_{\alpha;[0,t]} &\leq C \varrho_\alpha(\mathbf{X}, \tilde{\mathbf{X}}), \\
\sup_{\xi \in \mathbf{R}^e} \big| D^k \pi_{(f)}(0, \xi; \mathbf{X}) \big|_{\alpha;[0,t]} &\leq K, \\
\sup_{\xi \in \mathbf{R}^e} \big| D^k \pi_{(f)}(0, \xi; \mathbf{X})^{-1} \big|_{\alpha;[0,t]} &\leq K.
\end{aligned}$$

8.10 Exercises

Exercise 8.11. a) Consider the case of a smooth, one-dimensional driving signal $X : [0,T] \to \mathbf{R}$. Show that the solution map to the (ordinary) differential equation $dY = f(Y)dX$, for sufficiently nice f (say bounded with bounded derivatives) and started at some fixed point $Y_0 = \xi$, is locally Lipschitz continuous with respect to the driving signal in the supremum norm on $[0,T]$. Conclude that it admits a unique continuous extension to every *continuous* driving signal X.

b) Show by an example that, in general, no such statement holds for multi-dimensional driving signals.

c) Formulate a condition on f under which the statement does still hold for multidimensional driving signals.

Exercise 8.12 (Linear RDEs). Consider $f \in \mathcal{L}(W, \mathcal{L}(V,W))$. Given an a priori estimate for solutions to $dY = f(Y)d\mathbf{X}$. Conclude with a (global) existence and uniqueness results for such linear RDEs.

Exercise 8.13 (Explicit solution, Chen–Strichartz formula). View

$$f = (f_1, \ldots, f_d) \in \mathcal{C}_b^\infty\big(\mathbf{R}^e, \mathcal{L}\big(\mathbf{R}^d, \mathbf{R}^e\big)\big) ,$$

as collection of d (smooth, bounded with bounded derivatives of all orders) vector fields on $\mathbf{R}^e$. Assume that f is step-2 nilpotent in the sense that $[f_i,[f_j,f_k]] \equiv 0$ for any triple of indices $i,j,k \in \{1,\ldots,d\}$. Here, $[\cdot,\cdot]$ denotes the Lie bracket between two vector fields. Let $(Y, f(Y))$ be the RDE solution to $dY = f(Y)d\mathbf{X}$ started at some $\xi \in \mathbf{R}^e$ and assume that the rough path $\mathbf{X}$ is geometric. Give an explicit formula of the type $Y_t = \exp(...)\xi$ where exp denotes the unit time solution flow along a vector field (...) which you should write down explicitly.

Exercise 8.14 (Explosion along linear-growth vector fields). Give an example of smooth f with linear growth, and $\mathbf{X} \in \mathscr{C}^\alpha$ so that $dY = f(Y)d\mathbf{X}$ started at some ξ fails to have a global solution.

Exercise 8.15. Establish existence, continuity and stability for rough differential equations with drift (cf. (8.6)),

$$dY_t = f_0(Y_t)\,dt + f(Y_t)\,d\mathbf{X}_t \; . \tag{8.15}$$

You may assume $f_0 \in \mathcal{C}_b^3$ (although one can do much better and f_0 Lipschitz is enough). **Hint:** Under this assumption, one solves $dY = \bar{f}(Y)\bar{\mathbf{X}}$ with $\bar{f} = (f, f_0)$ and a $\bar{\mathbf{X}}$ a "space-time" rough path extension of $\mathbf{X}$.

Exercise 8.16. Let $f \in \mathcal{C}_b^2$ and assume $(Y, f(Y))$ is a RDE solution to (8.6), as constructed in Theorem 8.4. Show that the o-term in Davie's definition, (8.13), can be chosen uniformly over $(X,\mathbb{X}) \in B_R$, any $R < \infty$, where

$$B_R := \Big\{(X,\mathbb{X}) \in \mathscr{C}^\beta : \|X\|_\beta + \|\mathbb{X}\|_{2\beta} \le R\Big\}, \text{ any } R < \infty.$$

Show also that RDE solutions are β-Hölder, uniformly over $(X, \mathbb{X}) \in B_R$, any $R < \infty$.

Exercise 8.17. Show that $d_{X,X^n,2\alpha}((Y, f(Y)), (Y^n, f(Y^n))) \to 0$, together with $\mathbf{X} \to \mathbf{X}^n$ in $\mathscr{C}^\beta$ implies that also $(Y^n, \mathbb{Y}^n) \to (Y, \mathbb{Y})$ in $\mathscr{C}^\alpha$. Since, at the price of replacing f by F, cf. Definition 8.9, there is no loss of generality in solving for the controlled rough path $Z = X \oplus Y$, conclude that continuity of the RDE solution map (Itô–Lyons map) also holds with Lyons' definition of a solution.

8.11 Comments

ODEs driven by not too rough paths, i.e. paths that are α-Hölder continuous for some $\alpha > 1/2$ or of finite p-variation with $p < 2$, understood in the (Young) integral sense were first studied by Lyons in [Lyo94]; nonetheless, the terminology Young-ODEs is now widely used. Existence and uniqueness for such equations via Picard iterations is by now classical, our discussion in Section 8.3 is a mild variation of [LCL07, p.22] where also the *division property* (cf. proof of Lemma 8.2) is emphasised. Existence and uniqueness of solutions to RDEs via Picard iteration in the (Banach!) space of controlled rough paths originates in [Gub04] for regularity $\alpha \in (\frac{1}{3}, \frac{1}{2})$. This approach also allows to treat arbitrary regularities, see [Gub10, Hai14c].

The continuity result of Theorem 8.5 is due to T. Lyons; proofs of uniform continuity on bounded sets were given in [Lyo98, LQ02, LCL07]. Local Lipschitz estimates were pointed out subsequently and in different settings by various authors including Lyons–Qian [LQ02], Gubinelli [Gub04], Friz–Victoir [FV10b], Inahama [Ina10], Deya et al. [DNT12a].

The name *universal limit theorem* was suggested by P. Malliavin, meaning continuity of the Itô–Lyons map in rough path metrics. As we tried to emphasise, the stability in rough path metrics is seen at all levels of the theory.

Lyons' original argument (for arbitrary regularity) also involves a Picard iteration, see e.g. [LCL07, p.88]. For regularity $\alpha > 1/3$, Davie [Dav08] proves existence and uniqueness for Young resp. rough differential equations via discrete Euler resp. Milstein approximations. Using Lie group techniques, Davie's argument was adapted to arbitrary values of α by Friz–Victoir [FV10b]. Let us also note that the regularity assumption in Theorem 8.4 ($f \in \mathcal{C}_b^3$) is not sharp; it is fairly straightforward to push the argument to γ-Lipschitz (in the sense of Stein) regularity, for any $\gamma > 1/\alpha$. It is less straightforward [Dav08, FV10b] to show that uniqueness also holds for $\gamma \geq 1/\alpha$ and this is optimal, with counter-examples constructed in [Dav08]. Existence results on the other hand are available for $\gamma > (1/\alpha) - 1$. Setting $\alpha = 1$, this is consistent with the theory of ODEs where it is well known that, modulo possible logarithmic divergencies, Lipschitz continuity of the coefficients is required for the uniqueness of local solutions, but continuity is sufficient for their existence.

Chapter 9
Stochastic differential equations

Abstract We identify the solution to a rough differential equation driven by the Itô or Stratonovich lift of Brownian motion with the solution to the corresponding stochastic differential equation. In combination with continuity of the Itô–Lyons maps, a quick proof of the Wong–Zakai theorem is given. Applications to Stroock–Varadhan support theory and Freidlin–Wentzell large deviations are briefly discussed.

9.1 Itô and Stratonovich equations

We saw in Section 3 that d-dimensional Brownian motion lifts in an essentially canonical way to $\mathbf{B} = (B, \mathbb{B}) \in \mathscr{C}^\alpha\big([0,T], \mathbf{R}^d\big)$ almost surely, for any $\alpha \in \big(\frac{1}{3}, \frac{1}{2}\big)$. In particular, we may use almost every realisation of $(B, \mathbb{B})$ as the driving signal of a rough differential equation. This RDE is then solved "pathwise" i.e. for a *fixed* realisation of $(B(\omega), \mathbb{B}(\omega))$. Recall that the choice of $\mathbb{B}$ is never unique: two important choices are the Itô and the Stratonovich lift, we write $\mathbf{B}^{\text{Itô}}$ and $\mathbf{B}^{\text{Strat}}$, where $\mathbb{B}$ is defined as $\int B \otimes dB$ and $\int B \otimes \circ dB$ respectively. We now discuss the interplay with classical stochastic differential equations (SDEs).

Theorem 9.1. *Let $f \in \mathcal{C}_b^3\big(\mathbf{R}^e, \mathcal{L}\big(\mathbf{R}^d, \mathbf{R}^e\big)\big)$, let $f_0 : \mathbf{R}^e \to \mathbf{R}^e$ be Lipschitz continuous, and let $\xi \in \mathbf{R}^e$. Then,*

i) With probability one, $\mathbf{B}^{\text{Itô}}(\omega) \in \mathscr{C}^\alpha$, any $\alpha \in (1/3, 1/2)$ and there is a unique RDE solution $(Y(\omega), f(Y(\omega))) \in \mathscr{D}^{2\alpha}_{B(\omega)}$ to

$$dY = f_0(Y)dt + f(Y)\, d\mathbf{B}^{\text{Itô}}\,, \quad Y_0 = \xi.$$

Moreover, $Y = (Y_t(\omega))$ is a strong solution to the Itô SDE $dY = f_0(Y)dt + f(Y)dB$ started at $Y_0 = \xi$.

ii) Similarly, the RDE solution driven by $\mathbf{B}^{\text{Strat}}$ yields a strong solution to the Stratonovich SDE $dY = f_0(Y)dt + f(Y) \circ dB$ started at $Y_0 = \xi$.

Proof. We assume zero drift f_0, but see Exercise 8.15. The map

$$B|_{[0,t]} \mapsto (B, \mathbb{B}^{\text{Strat}})|_{[0,t]} \in \mathscr{C}_g^{0,\alpha}\big([0,t], \mathbf{R}^d\big)$$

is measurable, where $\mathscr{C}_g^{0,\alpha}$ denotes the (separable, hence Polish) subspace of $\mathscr{C}^\alpha$ obtained by taking the closure, in α-Hölder rough path metric, of piecewise smooth paths. This follows, for instance, from Proposition 3.6. By the continuity of the Itô–Lyons map (adding a drift vector field is left as an easy exercise) the RDE solution $Y_t \in \mathbf{R}^e$ is the continuous image of the driving signal $(B, \mathbb{B}^{\text{Strat}})|_{[0,t]} \in \mathscr{C}_g^{0,\alpha}\big([0,t], \mathbf{R}^d\big)$. It follows that Y_t is adapted to

$$\sigma\{B_{r,s}, \mathbb{B}_{r,s} : 0 \le r \le s \le t\} = \sigma\{B_s : 0 \le s \le t\} \;,$$

and it suffices to apply Corollary 5.2. Since $\mathbb{B}^{\text{Itô}}_{s,t} = \mathbb{B}^{\text{Strat}}_{s,t} - \frac{1}{2}(t-s)I$, measurability is also guaranteed and we conclude with the same argument, using Proposition 5.1. □

Remark 9.2. In contrast to standard SDE theory, the present solution constructed via RDEs is immediately well-defined as a *flow*, i.e. for all ξ on a common set of probability one. The price to pay is that of $\mathcal{C}^3$ regularity of f, as opposed to the mere Lipschitz regularity required for the standard theory.

9.2 The Wong–Zakai theorem

A classical result (e.g. [IW89, p.392]) asserts that SDE approximations based on piecewise linear approximations to the driving Brownian motions converge to the solution of the Stratonovich equation. Using the machinery built in the previous sections, we can now give a simple proof of this by combining Proposition 3.6, Theorem 8.5 and the understanding that RDEs driven by $\mathbf{B}^{\text{Strat}}$ yield solutions to the Stratonovich equation (Theorem 9.1).

Theorem 9.3 (Wong–Zakai, Clark, Stroock–Varadhan). *Let f, f_0, ξ be as in Theorem 9.1 above. Let $\alpha < 1/2$. Consider dyadic piecewise-linear approximations (B^n) to B on $[0,T]$, as defined in Proposition 3.6. Write Y^n for the (random) ODE solutions to $dY^n = f_0(Y^n)dt + f(Y^n)dB^n$ and Y for the Stratonovich SDE solution to $dY = f_0(Y)dt + f(Y) \circ dB$, all started at ξ. Then the Wong–Zakai approximations converge a.s. to the Stratonovich solution. More precisely, with probability one,*

$$\|Y - Y^n\|_{\alpha;[0,T]} \to 0.$$

The only reason for *dyadic* piecewise-linear approximations in the above statement is the formulation of the martingale-based Proposition 3.6. In Section 10 we shall present a direct analysis (going far beyond the setting of Brownian drivers) which easily entails quantitative convergence (in probability and L^q, any $q < \infty$) for all piecewise-linear approximations towards a (Gaussian) rough path.

In the forthcoming Exercise 10.14 it will be seen that (non-dyadic) piecewise linear approximations (meshsize $\sim 1/n$), viewed canonically as rough paths, converge a.s. in $\mathscr{C}^\alpha$ with rate anything less than $1/2-\alpha$. As long as $\alpha > 1/3$, it then follows from (local) Lipschitzness of the Itô–Lyons map that Wong–Zakai approximations also converge with rate $1/2-\alpha-$. Note that the "best" rate one obtains in this way is $1/2-1/3- = 1/6-$; the reason being that rate is measured in some Hölder space with exponent $1/3+$, rather than the uniform norm. The (well known) almost sure "strong" rate $1/2-$ can be obtained from rough path theory at the price of working in rough path spaces of (much) lower regularity; see [FR14].

9.3 Support theorem and large deviations

We briefly discuss two fundamental results in diffusion theory and explain how the theory of rough paths provides elegant proofs, reducing a question for general diffusion to one for Brownian motion and its Lévy area.

The results discussed in this section were among the very first applications of rough path theory to stochastic analysis, see Ledoux et al. [LQZ02]. Much more on these topics is found [FV10b], so we shall be brief. The first result, due to Stroock–Varadhan, concerns the support of diffusion processes.

Theorem 9.4 (Stroock–Varadhan support theorem). *Let f, f_0, ξ be as in Theorem 9.1 above. Let $\alpha < 1/2$, B be a d-dimensional Brownian motion and consider the unique Stratonovich SDE solution Y on $[0,T]$ to*

$$dY = f_0(Y)dt + \sum_{i=1}^{d} f_i(Y) \circ dB^i \tag{9.1}$$

started at $Y_0 = \xi \in \mathbf{R}^e$. Write y^h for ODE solution obtained by replacing $\circ dB$ with $dh \equiv \dot{h}\,dt$, whenever $h \in \mathcal{H} = W_0^{1,2}$, i.e. absolutely continuous, $h(0) = 0$ and $\dot{h} \in L^2([0,T],\mathbf{R}^d)$. Then, for every $\delta > 0$,

$$\lim_{\varepsilon\to 0} \mathbf{P}\Big(\|Y - Y^h\|_{\alpha;[0,T]} < \delta \;\Big|\; \|B-h\|_{\infty;[0,T]} < \varepsilon \Big) = 1 \tag{9.2}$$

(where Euclidean norm is used for the conditioning $\|B-h\|_{\infty,[0,T]} < \varepsilon$). As a consequence, the support of the law of Y, viewed as measure on the pathspace $\mathcal{C}^{0,\alpha}([0,T],\mathbf{R}^e)$, is precisely the α-Hölder closure of $\{y^h : \dot{h} \in L^2([0,T],\mathbf{R}^d)\}$.

Proof. Using Theorem 9.1 we can and will take Y as RDE solution driven by $\mathbf{B}^{\text{Strat}}(\omega)$. For $h \in \mathcal{H}$ and some fixed $\alpha \in (\frac{1}{3},\frac{1}{2})$, we furthermore denote by $S^{(2)}(h) = (h, \int h \otimes dh) \in \mathscr{C}_g^{0,\alpha}$ the canonical lift given by computing the iterated integrals using usual Riemann-Stieltjes integration. It was then shown in

[FLS06][1] that for every $\delta > 0$,

$$\lim_{\varepsilon\to 0}\mathbf{P}\Big(\varrho_{\alpha;[0,T]}\big(\mathbf{B}^{\text{Strat}}, S^{(2)}(h)\big) < \delta \;\Big|\; \|B-h\|_{\infty;[0,T]} < \varepsilon\Big) = 1. \tag{9.3}$$

The conditional statement then follows easily from continuity of the Itô–Lyons map and so yields the "difficult" support inclusion: every y^h is in the support of Y. The easy inclusion, support of Y contained in the closure of $\{y^h\}$, follows from the Wong–Zakai theorem, Theorem 9.3. If one is only interested in the support statement, but without the conditional statement (9.2), there are "softer" proofs; see Exercise 9.6 below. □

The second result to be discussed here, due to Freidlin–Wentzell, concerns the behaviour of diffusion in the singular ($\varepsilon \to 0$) limit when B is replaced by εB. We assume the reader is familar with large deviation theory.

Theorem 9.5 (Freidlin–Wentzell large deviations). *Let f, f_0, ξ be as in Theorem 9.1 above. Let $\alpha < 1/2$, B be a d-dimensional Brownian motion and consider the unique Stratonovich SDE solution $Y = Y^\varepsilon$ on $[0,T]$ to*

$$dY = f_0(Y)dt + \sum_{i=1}^{d} f_i(Y)\circ \varepsilon dB^i \tag{9.4}$$

started at $Y_0 = \xi \in \mathbf{R}^e$. Write Y^h for the ODE solution obtained by replacing $\circ\varepsilon dB$ with dh where $h \in \mathcal{H} = W_0^{1,2}$. Then $(Y_t^\varepsilon : 0 \le t \le T)$ satisfies a large deviation principle (in α-Hölder topology) with good rate function on pathspace given by

$$J(y) = \inf\big\{I(h) : Y^h = y\big\}.$$

Here I is Schilder's rate function for Brownian motion, i.e. $I(h) = \frac{1}{2}\|\dot{h}\|^2_{L^2([0,T],\mathbf{R}^d)}$ for $h \in \mathcal{H}$ and $I(h) = +\infty$ otherwise.

Proof. The key remark is that large deviation principles are robust under continuous maps, a simple fact known as *contraction principle*. The problem is then reduced to establishing a suitable large deviation principle for the Stratonovich lift of εB (which is exacly $\delta_\varepsilon \mathbf{B}^{\text{Strat}}$) in the α-Hölder rough path topology. Readers familiar with general facts of large deviation theory, in particular the *inverse* and *generalized* contraction principles, are invited to complete the proof along Exercise 9.7 below. □

9.4 Exercises

Exercise 9.6 (support of Brownian rough path, see [FV10b]). Fix $\alpha \in (\frac{1}{3}, \frac{1}{2})$ and view the law μ of $\mathbf{B}^{\text{Strat}}$ as probability measure on the Polish space $\mathscr{C}^{0,\alpha}_{g,0}$, the

[1] Strictly speaking, this was shown for $h \in \mathcal{C}^2$; the extension to $h \in \mathcal{H}$ is non-trivial and found in [FV10b].

(closed) subspace of $\mathscr{C}_g^{0,\alpha}$ of rough paths **X** started at $X_0 = 0$. Show that $\mathbf{B}^{\text{Strat}}$ has full support. The "easy" inclusion, supp $\mu \subset \mathscr{C}_g^{0,\alpha}$ is clear from Proposition 3.6. For the other inclusion, recall the translation operator from Exercise 2.19 and follow the steps below.

a) **(Cameron–Martin theorem for Brownian rough path)** Let $h \in [0,T] \in \mathcal{H} = W_0^{1,2}$. Show that $\mathbf{X} \in \operatorname{supp}\mu$ implies $T_h(\mathbf{X}) \in \operatorname{supp}\mu$.
b) Show that the support of μ contains at least one point, say $\hat{\mathbf{X}} \in \mathscr{C}_g^{0,\alpha}$ with the property that there exists a sequence of Lipschitz paths $(h^{(n)})$ so that $T_{h^{(n)}}(\hat{\mathbf{X}}) \to (0,0)$ in α-Hölder rough path metric. **Hint:** Almost every realization of $\mathbf{B}^{\text{Strat}}(\omega)$ will do, with $-h^{(n)} = B^{(n)}$, the dyadic piecewise-linear approximations from Proposition 3.6.
c) Conclude that $(0,0) = \lim_{n\to\infty} T_{h^{(n)}}(\hat{\mathbf{X}}) \in \operatorname{supp}\mu$.
d) As a consequence, any $(h, \int h \otimes dh) = T_h(0,0) \in \operatorname{supp}\mu$, for any $h \in \mathcal{H}$ and taking the closure yields the "difficult" inclusion.
e) Appeal to continuity of the Itô–Lyons map to obtain the "difficult" support inclusion ("every y^h is in the support of Y") in the context of Theorem 9.4.

Exercise 9.7 ("Schilder" large deviations, see [FV10b]). Fix $\alpha \in (\frac{1}{3}, \frac{1}{2})$ and consider

$$\delta_\varepsilon \mathbf{B}^{\text{Strat}} = (\varepsilon B, \varepsilon^2 \mathbb{B}^{\text{Strat}})\ ,$$

the laws of which are viewed as probability measures μ^ε on the Polish space $\mathscr{C}_{g,0}^{0,\alpha}$. Show that $(\mu^\varepsilon) : \varepsilon > 0$ satisfies a large deviation principle in α-Hölder rough path topology with good rate function

$$J(\mathbf{X}) = I(X)\ ,$$

where $\mathbf{X} = (X, \mathbb{X})$ and I is Schilder's rate function for Brownian motion, i.e. $I(h) = \frac{1}{2}\|\dot{h}\|^2_{L^2([0,T],\mathbf{R}^d)}$ for $h \in \mathcal{H} = W_0^{1,2}$ and $I(h) = +\infty$ otherwise.

Hint: Thanks to Fernique estimates for the homogeneous rough paths norm of $\mathbf{B}^{\text{Strat}}$ (which can be obtained by carefully tracking the moment-growth in Theorem 3.1 applied to $\mathbf{B}^{\text{Strat}}$; alternatively see Theorem 11.9 below for an elegant Gaussian argument) it is actually enough to establish a large deviation principle for $(\delta_\varepsilon \mathbf{B}^{\text{Strat}} : \varepsilon > 0)$ in the (much coarser) uniform topology, which is not very hard to do "by hand", cf. [FV10b].

9.5 Comments

Lyons [Lyo98] used the Wong–Zakai theorem in conjunction with his continuity result to deduce the fact that RDE solutions (driven by the Brownian rough path $\mathbf{B}^{\text{Strat}}$) coincide with solution to (Stratonovich) stochastic differential equations. Similar to Friz–Victoir [FV10b], the logic is reversed here: thanks to an a priori identification of $\int f(Y)\, d\mathbf{B}^{\text{Strat}}$ as a Stratonovich stochastic integral, the Wong–Zakai results is obtained. Almost sure rates for Wong–Zakai approximations in Brownian (and then more general Gaussian) situations, were studied by Hu–Nualart [HN09], Deya, Tindel and Neuenkirch [DNT12b] and Friz–Riedel [FR14]; see also Riedel–Xu [RX13]. Let us also note that L^q-rates for the convergence of approximations are not easy to obtain with rough path techniques (in contrast to Itô-calculus which is ideally suited for moment calculations). Nonetheless, such rates can be obtained by Gaussian techniques, as discussed in Section 11.2.3 below; applications include multi-level Monte Carlo for rough differential equations [BFRS13]. The material in Section 9.3 goes back to Ledoux, Qian and Zhang ([LQZ02]; in p-variation). The results in stronger Hölder topolgy are due to Friz and Victoir [Fri05, FV05, FV07, FV10b], the conditional estimate (9.3) is due to Friz, Lyons and Stroock [FLS06].

Chapter 10
Gaussian rough paths

Abstract We investigate when multidimensional stochastic processes can be viewed – in a "canonical" fashion – as random rough paths. Gaussianity only enters through equivalence of moments. A simple criterion is given which applies in particular to fractional Brownian motion with suitable Hurst parameter.

10.1 A simple criterion for Hölder regularity

We now consider a driving signal modelled by a continuous, centred Gaussian process with values in $V = \mathbf{R}^d$. We thus have continuous sample paths

$$X(\omega) : [0,T] \to \mathbf{R}^d$$

and may take the underlying probability space as $\mathcal{C}\big([0,T], \mathbf{R}^d\big)$, equipped with a Gaussian measure μ so that $X_t(\omega) = \omega(t)$. Recall that μ, the law of X, is fully determined by its covariance function

$$\begin{aligned} R : [0,T]^2 &\to \mathbf{R}^{d\times d} \\ (s,t) &\mapsto \mathbf{E}[X_s \otimes X_t] \; . \end{aligned}$$

In this section, a major role will be played by the *rectangular increments* of the covariance, namely

$$R\begin{pmatrix} s\,,t \\ s',t' \end{pmatrix} \stackrel{\text{def}}{=} \mathbf{E}[X_{s,t} \otimes X_{s',t'}] \; .$$

As far as the Hölder regularity of sample paths is concerned, we have the following classical result, which is nothing but a special case of Kolmogorov's continuity criterion:

Proposition 10.1. *Assume there exists positive ϱ and M such that for every $0 \le s \le t \le T$,*

$$\left| R\begin{pmatrix} s,t \\ s,t \end{pmatrix} \right| \le M|t-s|^{1/\varrho}. \tag{10.1}$$

Then, for every $\alpha < 1/(2\varrho)$ there exists $K_\alpha \in L^q$, for all $q < \infty$, such that

$$|X_{s,t}(\omega)| \le K_\alpha(\omega)|t-s|^\alpha.$$

Proof. We may argue componentwise and thus take $d = 1$ without loss of generality. Since

$$|X_{s,t}|_{L^2} = (\mathbf{E}[X_{s,t}X_{s,t}])^{1/2} \le \left| R\begin{pmatrix} s,t \\ s,t \end{pmatrix} \right|^{1/2} \le M^{1/2}|t-s|^{\frac{1}{2\varrho}}$$

and $|X_{s,t}|_{L^q} \le c_q|X_{s,t}|_{L^2}$ by Gaussianity, we conclude immediately with an application of the Kolmogorov criterion. □

Whenever the above proposition applies with $\varrho < 1$, the resulting sample paths can be taken with Hölder exponent $\alpha \in (\frac{1}{2}, \frac{1}{2\varrho})$; differential equations driven by X can then be handled with Young's theory, cf. Section 8.3. Therefore, our focus will be on Gaussian processes which satisfy a suitable modification of condition (10.1) with $\varrho \ge 1$ such that the process X allows for a *probabilistic* construction of a suitable second order process[1]

$$\mathbb{X}(\omega) : [0,T]^2 \to \mathbf{R}^{d\times d} \ ,$$

which is tantamount to making sense of the "formal" stochastic integrals

$$\int_s^t X^i_{s,r} dX^j_r \qquad \text{for} \qquad 0 \le s < t \le T, \quad 1 \le i,j \le d \ , \tag{10.2}$$

such that almost every realisation $\mathbb{X}(\omega)$ satisfies the algebraic and analytical properties of Section 2, notably (2.1) and (2.3) for some $\alpha \in \left(\frac{1}{3}, \frac{1}{2}\right]$. We shall also look for $(X, \mathbb{X})$ as (random) *geometric* rough path; thanks to (2.5), only the case $i < j$ in (10.2) then needs to be considered.

At the risk of being repetitive, the reader should keep in mind the following three points: (i) the sample paths $X(\omega)$ will not have, in general, enough regularity to define (10.2) as Young integrals; (ii) the process X will not be, in general, a semimartingale, so (10.2) cannot be defined using classical stochastic integrals; (iii) a lift of the process X to $(X, \mathbb{X}) \in \mathscr{C}^\alpha_g$ for some $\alpha \in \left(\frac{1}{3}, \frac{1}{2}\right]$, if at all possible, will never be unique (as discussed in Chapter 2, one can always perturb the area, i.e. Anti($\mathbb{X}$) by the increments of a 2α-Hölder path). But there might still be one distinguished *canonical* choice for $\mathbb{X}$, in the same way as $\mathbb{B}^{\text{Strat}}$ is canonically obtained as limit (in probability) of $\int B^n \otimes dB^n$, for many natural approximations B^n of Brownian motion B.

[1] Despite the two parameters (s,t) one should not think of a random field here: as was noted in Exercise 2.7, $(X, \mathbb{X})$ is really a *path*.

10.2 Stochastic integration and variation regularity of the covariance

Our standing assumption from here on is independence of the d components of X, which is tantamount to saying that the covariance takes values in the diagonal matrices. Basic examples to have in mind are d-dimensional standard Brownian motion B with

$$R(s,t) = (s \wedge t) \times I_d \in \mathbf{R}^{d \times d}$$

(here I_d denotes the identity matrix in $\mathbf{R}^{d\times d}$) or fractional Brownian motion B^H, with

$$R_H(s,t) = \frac{1}{2}\Big[s^{2H} + t^{2H} - |t-s|^{2H}\Big] \times I_d \in \mathbf{R}^{d\times d}$$

where $H \in (0,1)$; note the implication $\mathbf{E}\big[\big(B^H_t - B^H_s\big)^2\big] = |t-s|^{2H}$. The reader should observe that Proposition 10.1 above applies with $\varrho = 1/(2H)$; the focus on $\varrho \geq 1$ (to avoid trivial situations covered by Young theory) translates to $H \leq 1/2$.

We return to the task of making sense of (10.2), componentwise for fixed $i < j$, and it will be enough to do so for the unit interval; the interval $[s,t]$ is handled by considering $\big(X_{s+\tau(t-s)} : 0 \leq \tau \leq 1\big)$. Writing $\big(X, \tilde{X}\big)$, rather than $\big(X^i, X^j\big)$, we attempt a definition of the form

$$\int_0^1 X_{0,u}\, d\tilde{X}_u \stackrel{\text{def}}{=} \lim_{|\mathcal{P}|\downarrow 0} \sum_{[s,t]\in\mathcal{P}} X_{0,\xi}\tilde{X}_{s,t} \quad \text{with} \quad \xi \in [s,t]\ , \tag{10.3}$$

where the limit is understood in probability, say. Classical stochastic analysis (e.g. [RY91, p144]) tells us that care is necessary: if $X, \tilde{X}$ are semimartingales, the choice $\xi = s$ ("left-point evaluation") leads to the Itô integral; $\xi = t$ ("right-point evaluation") to the backward Itô - and $\xi = (s+t)/2$ to the Stratonovich integral. On the other hand, all these integrals only differ by a bracket term $\langle X, \tilde{X}\rangle$ which vanishes if $X, \tilde{X}$ are independent. While we do not assume a semi-martingale structure here, we do have the standing assumption of componentwise independence. This suggests a Riemann sum approximation of (10.2) in which we expect the precise point of evaluation to play no rôle; we thus consider left-point evaluation (but mid- or rightpoint evaluation would lead to the same result; cf. Exercise 10.18, (ii) below). Give partitions $\mathcal{P}, \mathcal{P}'$ of $[0,1]$ we set

$$\int_{\mathcal{P}} X_{0,s}\, d\tilde{X}_s := \sum_{[s,t]\in\mathcal{P}} X_{0,s}\tilde{X}_{s,t}\ ,$$

so that under the *assumption that X and $\tilde{X}$ are independent*, we have

$$\mathbf{E}\Big[\int_{\mathcal{P}} X_{0,s}\, d\tilde{X}_s \int_{\mathcal{P}'} X_{0,s}\, d\tilde{X}_s\Big] = \sum_{\substack{[s,t]\in\mathcal{P}\\ [s',t']\in\mathcal{P}'}} R\begin{pmatrix}0, s\\ 0, s'\end{pmatrix}\tilde{R}\begin{pmatrix}s\,, t\\ s', t'\end{pmatrix}. \tag{10.4}$$

On the right-hand side we recognise a 2D Riemann-Stieltjes sum and set

$$\int_{\mathcal{P}\times\mathcal{P}'} R\,d\tilde{R} := \sum_{\substack{[s,t]\in\mathcal{P}\\ [s',t']\in\mathcal{P}'}} R\begin{pmatrix}0,s\\0,s'\end{pmatrix}\tilde{R}\begin{pmatrix}s\,,t\\s',t'\end{pmatrix}.$$

Let us now assume that R has finite ϱ-variation in the sense $\|R\|_{\varrho;[0,1]^2} < \infty$ where the ϱ-variation on a rectangle $I \times I'$ is given by

$$\|R\|_{\varrho;I\times I'} := \left(\sup_{\substack{\mathcal{P}\subset I,\\ \mathcal{P}'\subset I'}} \sum_{\substack{[s,t]\in\mathcal{P}\\ [s',t']\in\mathcal{P}'}} \left|R\begin{pmatrix}s\,,t\\s',t'\end{pmatrix}\right|^{\varrho}\right)^{1/\varrho} < \infty, \tag{10.5}$$

and similarly for $\tilde{R}$, with $\theta = 1/\varrho + 1/\tilde{\varrho} > 1$. A generalisation of Young's maximal inequality due to Towghi [Tow02] states that [2]

$$\sup_{\substack{\mathcal{P}\subset I,\\ \mathcal{P}'\subset I'}} \left|\int_{\mathcal{P}\times\mathcal{P}'} R\,d\tilde{R}\right| \le C(\theta)\big\|R\big\|_{\varrho;I\times I'}\big\|\tilde{R}\big\|_{\tilde{\varrho};I\times I'}.$$

In particular, if the covariance of $\tilde{X}$ has similar variation regularity as X, the condition simplifies to $\varrho < 2$ and we obtain the following *L^2-maximal inequality.*

Lemma 10.2. *Let $X, \tilde{X}$ be independent, continuous, centred Gaussian processes with respective covariances $R, \tilde{R}$ of finite ϱ-variation, some $\varrho < 2$. Then*

$$\sup_{\mathcal{P}\subset[0,1]} \mathbf{E}\left[\left(\int_{\mathcal{P}} X_{0,r}\,d\tilde{X}_r\right)^2\right] \le C\big\|R\big\|_{\varrho;[0,1]^2}\big\|\tilde{R}\big\|_{\varrho;[0,1]^2}\,,$$

where the constant C depends on ϱ.

We can now show existence of (10.3) as L^2-limit.

Proposition 10.3. *Under the assumptions of the previous lemma,*

$$\lim_{\varepsilon\to 0} \sup_{\substack{\mathcal{P},\mathcal{P}'\subset[0,1]:\\ |\mathcal{P}|\vee|\mathcal{P}'|<\varepsilon,}} \left|\int_{\mathcal{P}} X_{0,r}d\tilde{X}_r - \int_{\mathcal{P}'} X_{0,r}d\tilde{X}_r\right|_{L^2} = 0. \tag{10.6}$$

Hence, $\int_0^1 X_{0,r}d\tilde{X}_r$ exists as the L^2-limit of $\int_{\mathcal{P}} X_{0,r}d\tilde{X}_r$ as $|\mathcal{P}| \downarrow 0$ and

$$\mathbf{E}\left[\left(\int_0^1 X_{0,r}d\tilde{X}_r\right)^2\right] \le C\big\|R\big\|_{\varrho;[0,1]^2}\big\|\tilde{R}\big\|_{\varrho;[0,1]^2} \tag{10.7}$$

with a constant $C = C(\varrho)$.

[2] This holds more generally if R is evaluated at $[0,\xi]\times[0,\xi']$ where $\xi\in[s,t],\ \xi'\in[s',t']$.

Proof. At first glance, the situation looks similar to Young's part in the proof of Theorem 4.10 where we deduce (4.12) from Young's maximal inequality. However, the same argument fails if re-run with $\Xi_{s,t} = X_{0,s}\tilde{X}_{s,t}$ and $|\cdot|$ replaced by $|\cdot|_{L^2}$; in effect, the triangle inequality is too crude and does not exploit probabilistic cancellations present here. We now present two arguments for the key estimate (10.6). **First argument**: at the price of adding/subtracting $\mathcal{P} \cap \mathcal{P}'$, we may assume without loss of generality that $\mathcal{P}'$ refines $\mathcal{P}$. This allows to write

$$\int_{\mathcal{P}'} X_{0,r}\, d\tilde{X}_r - \int_{\mathcal{P}} X_{0,r}\, d\tilde{X}_r = \sum_{[u,v]\in\mathcal{P}} \int_{\mathcal{P}'\cap[u,v]} X_{u,r}\, d\tilde{X}_r \stackrel{\text{def}}{=} \mathcal{I} ,$$

and we need to show convergence of $\mathcal{I}$ to zero in L^2 as $|\mathcal{P}| = |\mathcal{P}| \vee |\mathcal{P}'| \to 0$. To see this, we rewrite the square of the expectation of this quantity as

$$\begin{aligned}\mathbf{E}\mathcal{I}^2 &= \sum_{[u,v]\in\mathcal{P}} \sum_{[u',v']\in\mathcal{P}} \mathbf{E}\left(\int_{\mathcal{P}'\cap[u,v]} X_{u,r}\, d\tilde{X}_r \int_{\mathcal{P}'\cap[u',v']} X_{u',r'}\, d\tilde{X}_{r'}\right) \\ &= \sum_{[u,v]\in\mathcal{P}} \sum_{[u',v']\in\mathcal{P}} \int_{\mathcal{P}'\cap[u,v]\times\mathcal{P}'\cap[u',v']} R\, d\tilde{R} .\end{aligned}$$

Thanks to Towghi's maximal inequality, the absolute value of this term is bounded from above by a constant $C = C(\varrho)$ times

$$\begin{aligned}&\sum_{[u,v]\in\mathcal{P}} \sum_{[u',v']\in\mathcal{P}} \|R\|_{\varrho;[u,v]\times[u',v']} \|\tilde{R}\|_{\varrho;[u,v]\times[u',v']} \\ &\qquad\qquad \le \sum_{[u,v]\in\mathcal{P}} \sum_{[u',v']\in\mathcal{P}} \omega([u,v]\times[u',v'])^{\frac{1}{\varrho}} \tilde{\omega}([u,v]\times[u',v'])^{\frac{1}{\varrho}} ,\end{aligned}$$

where $\omega = \omega([s,t]\times[s',t'])$ (and similarly for $\tilde{\omega}$) is a so-called 2D control [FV11]: super-additive, continuous and zero when $s = t$ or $s' = t'$. A possible choice, if finite, is

$$\omega([s,t]\times[s',t']) \stackrel{\text{def}}{=} \sup_{\mathcal{Q}\subset[s,t]\times[s',t']} \sum_{[u,v]\times[u',v']\in\mathcal{Q}} \left| R\begin{pmatrix} u, v \\ u', v' \end{pmatrix} \right|^{\varrho} . \tag{10.8}$$

The difference to (10.5) is that the sup is taken over all (finite) partitions $\mathcal{Q}$ of $[s,t]\times[s',t']$ into rectangles; not just "grid-like" partitions induced by $\mathcal{P}\times\mathcal{P}'$. At this stage it looks like one should the change the assumption "covariance of finite ϱ-variation" to "finite *controlled* ϱ-variation", which by definition means $\omega\big([0,1]^2\big) < \infty$. But in fact there is little difference [FV11]: finite controlled ϱ-variation trivially implies finite ϱ-variation; conversely, finite ϱ-variation implies finite controlled ϱ'-variation, any $\varrho' > \varrho$. Since (10.6) does not depend on ϱ, we may as well (at the price of replacing ϱ by ϱ') assume finite controlled ϱ-variation. The Cauchy–Schwarz inequality for finite sums shows that $\bar{\omega} := \omega^{1/2}\tilde{\omega}^{1/2}$ is again a 2D control; the above estimates can then be continued to

$$
\begin{aligned}
\mathbf{E}\mathcal{I}^2 &\le C \sum_{[u,v]\in\mathcal{P}} \sum_{[u',v']\in\mathcal{P}} \bar{\omega}([u,v]\times[u',v'])^{2/\varrho} \\
&\le C \max_{\substack{[u,v]\in\mathcal{P}\\ [u',v']\in\mathcal{P}}} \bar{\omega}([u,v]\times[u',v'])^{\frac{2-\varrho}{\varrho}} \times \sum_{[u,v]\in\mathcal{P}} \sum_{[u',v']\in\mathcal{P}} \bar{\omega}([u,v]\times[u',v']) \\
&\le \mathrm{o}(1) \times \bar{\omega}([0,1]\times[0,1]) \;,
\end{aligned}
$$

where we used the facts that $|\mathcal{P}| \downarrow 0$, $\varrho < 2$ and super-additivity of $\bar{\omega}$ to obtain the last inequality. This is precisely the required bound. The **second argument** makes use of Riemann-Stieltjes theory, applicable after mollification of $\tilde{X}$, and a uniformity property of ϱ-variation upon mollification. Let thus denote $\tilde{X}^n := \tilde{X} * f_n$ the convolution of $t \mapsto \tilde{X}_t$ with (f_n), a family of smooth, compactly supported probability density functions, weakly convergent to a Dirac at 0. Writing $\tilde{R}^n_{s,t} := \mathbf{E}\big(\tilde{X}^n_s \tilde{X}^n_t\big)$ for the covariance of $\tilde{X}^n$, and also $\tilde{S}^n_{s,t} := \mathbf{E}\big(\tilde{X}_s \tilde{X}^n_t\big)$ for the "mixed" covariance, we leave the fact that

$$
\sup_n \big\|\tilde{R}^n\big\|_{\varrho;[0,1]^2}, \sup_n \big\|\tilde{S}^n\big\|_{\varrho;[0,1]^2} \le \big\|\tilde{R}\big\|_{\varrho;[0,1]^2} \;, \tag{10.9}
$$

as and easy exercise for the reader. (**Hint:** Note $\tilde{R}^n = \tilde{R} * (f_n \otimes f_n)$, $\tilde{S}^n = \tilde{R} * (\delta \otimes f_n)$; estimate then the rectangular increments of $\tilde{R}_n$, respectively $\tilde{S}^n$, to the power ϱ with Jensen's inequality.)

Since $\tilde{X}^n$ has finite variation sample paths, basic Riemann-Stieltjes theory implies

$$
\int_{\mathcal{P}} X_{0,r}\, d\tilde{X}^n_r \to \int X_{0,r}\, d\tilde{X}^n_r \quad \text{as} \quad |\mathcal{P}| \to 0. \tag{10.10}
$$

In fact, this convergence (n fixed) takes also place in L^2 which may be seen as consequence of Lemma 10.2. On the other hand, pick $\varrho' \in (\varrho, 2)$ and apply Lemma 10.2 to obtain[3]

$$
\begin{aligned}
\sup_{\mathcal{P}} \Big| \int_{\mathcal{P}} X_{0,r}\, d\tilde{X}_r - \int_{\mathcal{P}} X_{0,r}\, d\tilde{X}^n_r \Big|^2_{L^2} &\le C \|R_X\|_{\varrho';[0,1]^2} \big\|R_{\tilde{X}-\tilde{X}^n}\big\|_{\varrho';[0,1]^2} \\
&\le C \|R_X\|_{\varrho';[0,1]^2} \big\|R_{\tilde{X}-\tilde{X}^n}\big\|^{\varrho/\varrho'}_{\varrho;[0,1]^2} \big\|R_{\tilde{X}-\tilde{X}^n}\big\|^{1-\varrho/\varrho'}_{\infty;[0,1]^2} \;,
\end{aligned} \tag{10.11}
$$

where $C = C(\varrho)$. Now $\varrho' > \varrho$ implies $\|R_X\|_{\varrho';[0,1]^2} \le \|R_X\|_{\varrho;[0,1]^2}$ (immediate consequence of $|x|_{\varrho'} \le |x|_{\varrho} \equiv (\sum_{i=1}^m |x_i|^{\varrho})^{1/\varrho}$ on $\mathbf{R}^m$) and thanks to (10.9) we also have the (uniform in n) estimate

$$
\big\|R_{\tilde{X}-\tilde{X}^n}\big\|_{\varrho;[0,1]^2} \le C_{\varrho} \Big(\big\|R_{\tilde{X}}\big\|_{\varrho;[0,1]^2} + 2\big\|\tilde{S}^n\big\|_{\varrho;[0,1]^2} + \big\|R_{\tilde{X}^n}\big\|_{\varrho;[0,1]^2} \Big)
$$

[3] Define $|f|_{\infty;[0,1]^2} = \sup \left| f\begin{pmatrix} u,v \\ u',v' \end{pmatrix} \right|$ where the sup is taken over all $[u,v]$, $[u',v'] \subset [0,1]$.

$$\leq 4C_\varrho \|\tilde{R}\|_{\varrho;[0,1]^2}.$$

Since $\tilde{X}^n$ converges to $\tilde{X}$ uniformly and in L^2, it is not hard to see that $R_{\tilde{X}-\tilde{X}^n} \to 0$ uniformly on $[0,1]^2$. We then see that (10.11) tends to zero as $n \to \infty$. It is now an elementary exercise to combine this with (10.10) to conclude the (second) proof of (10.6).

At last, the L^2-estimate is an immediate corollary of the maximal inequality given in Lemma 10.2 and L^2-convergence of the approximating Riemann-Stieltjes sums. □

Note that there was nothing special about the time horizon $[0,1]$ in the above discussion. Indeed, given any time horizon $[s,t]$ of interest, it suffices to apply the same argument to the process $\big(X_{s+\tau(t-s)} : 0 \leq \tau \leq 1\big)$. Since variation norms are conveniently invariant under reparametrisation, (10.7) translates immediately to an estimate of the form

$$\mathbf{E}\left[\left(\int_s^t X_{s,r}\, d\tilde{X}_r\right)^2\right] \leq C\|R\|_{\varrho;[s,t]^2}\|\tilde{R}\|_{\varrho;[s,t]^2}\ , \tag{10.12}$$

first for the approximating Riemann-Stieltjes sums and then for their L^2-limits.

Theorem 10.4. *Let $(X_t : 0 \leq t \leq T)$ be a d-dimensional, continuous, centred Gaussian process with independent components and covariance R such that there exists $\varrho \in [1,2)$ and $M < \infty$ such that for every $i \in \{1,\dots,d\}$ and $0 \leq s \leq t \leq T$,*

$$\|R_{X^i}\|_{\varrho;[s,t]^2} \leq M|t-s|^{1/\varrho}. \tag{10.13}$$

Define, for $1 \leq i < j \leq d$ and $0 \leq s \leq t \leq T$, in L^2-sense (cf. Proposition 10.3),

$$\mathbb{X}^{i,j}_{s,t} := \lim_{|\mathcal{P}|\to 0} \int_{\mathcal{P}} \big(X^i_r - X^i_s\big) dX^j_r\ ,$$

and then also (the algebraic conditions (2.1) and (2.5) leave no other choice!)

$$\mathbb{X}^{i,i}_{s,t} := \frac{1}{2}\big(X^i_{s,t}\big)^2 \quad \text{and} \quad \mathbb{X}^{j,i}_{s,t} := -\mathbb{X}^{i,j}_{s,t} + X^i_{s,t}X^j_{s,t}\ . \tag{10.14}$$

Then, the following properties hold:

a) For every $q \in [1,\infty)$ there exists $C_1 = C_1(q,\varrho,d,T)$ such that for all $0 \leq s \leq t \leq T$,

$$\mathbf{E}\Big(|X_{s,t}|^{2q} + |\mathbb{X}_{s,t}|^q\Big) \leq C_1 M^q |t-s|^{q/\varrho}. \tag{10.15}$$

b) There exists a continuous modification of $\mathbb{X}$, denoted by the same letter from here on. Moreover, for any $\alpha < 1/(2\varrho)$ and $q \in [1,\infty)$ there exists $C_2 = C_2(q,\varrho,d,\alpha)$ such that

$$\mathbf{E}\Big(\|X\|_\alpha^{2q} + \|\mathbb{X}\|_{2\alpha}^q\Big) \leq C_2 M^q. \tag{10.16}$$

c) For any $\alpha < \frac{1}{2\varrho}$, with probability one, the pair $(X, \mathbb{X})$ satisfies conditions (2.1), (2.3) and (2.5). In particular, for $\varrho \in [1, \frac{3}{2})$ and any $\alpha \in (\frac{1}{3}, \frac{1}{2\varrho})$ we have $(X, \mathbb{X}) \in \mathscr{C}_g^\alpha$ almost surely.

Proof. By scaling, we can take $M = 1$ without loss of generality. Regarding the first property, the "first level" estimates are contained in Proposition 10.1. Thus, in view of (10.14), in order to establish (10.15) only $\mathbf{E}\big(\big|\mathbb{X}_{s,t}^{i,j}\big|^q\big)$ for $i < j$ needs to be considered. For $q = 2$ this is an immediate consequence of (10.12) and our assumption (10.13). The case of general q follows from the well known equivalence of L^q- and L^2-norm on the second Wiener–Itô chaos (e.g. [FV10b, Appendix D]).

Regarding the remaining two properties, almost sure validity of the algebraic constraint (2.1) for any fixed pair of times is an easy consequence of algebraic identities for Riemann sums. The construction of a continuous modification of $(s,t) \mapsto \mathbb{X}_{s,t}$ under the assumed bound is then standard (in fact, the proof of Theorem 3.1 shows this for dyadic times and the unique continuous extension is the desired modification). At last, Theorem 3.1 yields $K_\alpha, \mathbb{K}_\alpha$, with moments of all orders, such that

$$|X_{s,t}| \le K_\alpha(\omega)|t-s|^\alpha\;, \qquad |\mathbb{X}_{s,t}| \le \mathbb{K}_\alpha(\omega)|t-s|^{2\alpha}\;.$$

The dependence of the moments of K_α and $\mathbb{K}_\alpha$ on M finally follows by simple rescaling. $\square$

Theorem 10.5. *Let $(X,Y) = \big(X^1, Y^1, \ldots, X^d, Y^d\big)$ be a centred continuous Gaussian process on $[0,T]$ such that $\big(X^i, Y^i\big)$ is independent of $\big(X^j, Y^j\big)$ when $i \ne j$. Assume that there exists $\varrho \in [1,2)$ and $M \in (0,\infty)$ such that the bounds*

$$\begin{aligned} \|R_{X^i}\|_{\varrho;[s,t]^2} &\le M|t-s|^{1/\varrho}\;, \qquad \|R_{Y^i}\|_{\varrho;[s,t]^2} \le M|t-s|^{1/\varrho}\;, \\ \|R_{X^i - Y^i}\|_{\varrho;[s,t]^2} &\le \varepsilon^2 M|t-s|^{1/\varrho}\;, \end{aligned} \tag{10.17}$$

hold for all $i \in \{1,\ldots,d\}$ and all $0 \le s \le t \le T$. Then

a) For every $q \in [1,\infty)$, the bounds

$$\begin{aligned} \mathbf{E}(|Y_{s,t} - X_{s,t}|^q)^{\frac{1}{q}} &\lesssim \varepsilon\sqrt{M}|t-s|^{\frac{1}{2\varrho}}\;, \\ \mathbf{E}(|\mathbb{Y}_{s,t} - \mathbb{X}_{s,t}|^q)^{\frac{1}{q}} &\lesssim \varepsilon\, M|t-s|^{\frac{1}{\varrho}}\;, \end{aligned}$$

hold for all $0 \le s \le t \le T$.

b) For any $\alpha < 1/(2\varrho)$ and $q \in [1,\infty)$, one has

$$\begin{aligned} |\mathbf{E}(\|Y - X\|_\alpha^q)|^{\frac{1}{q}} &\lesssim \varepsilon\sqrt{M}\;, \\ |\mathbf{E}(\|\mathbb{Y} - \mathbb{X}\|_{2\alpha}^q)|^{\frac{1}{q}} &\lesssim \varepsilon M\;. \end{aligned}$$

c) For $\varrho \in [1, \frac{3}{2})$ and any $\alpha \in (\frac{1}{3}, \frac{1}{2\varrho})$, $q < \infty$, one has

$$|\varrho_\alpha(\mathbf{X},\mathbf{Y})|_{L^q} \lesssim \varepsilon \ .$$

(Here, $\varrho_\alpha(\mathbf{X},\mathbf{Y})$ denotes the α-Hölder rough path distance between $\mathbf{X}=(X,\mathbb{X})$ and $\mathbf{Y}=(Y,\mathbb{X})$ in $\mathscr{C}_g^\alpha$.)

Proof. By scaling we may without loss of generality assume $M=1$. As for a) we note (again) that equivalence of L^q- and L^2-norm on Wiener–Itô chaos allow to reduce our discussion to $q=2$. The first level estimate being easy, we focus on the second level estimate; to this end fix $i\neq j$. Since L^2-convergence implies a.s. convergence along a subsequence there exists $(\mathcal{P}_n)$, with mesh tending to zero, we can use Fatou's lemma to estimate

$$\begin{aligned}
\mathbf{E}\big(\big|\mathbb{Y}^{i,j}_{s,t}-\mathbb{X}^{i,j}_{s,t}\big|^2\big) &= \mathbf{E}\Big(\lim_{n\to\infty}\Big|\int_{\mathcal{P}_n} Y^i_{s,r}dY^j_r - X^i_{s,r}dX^j_r\Big|^2\Big)\\
&\le \liminf_n \mathbf{E}\Big(\Big|\int_{\mathcal{P}_n} Y^i_{s,r}dY^j_r - X^i_{s,r}dX^j_r\Big|^2\Big)\\
&\le \sup_{\mathcal{P}} \mathbf{E}\Big(\Big|\int_{\mathcal{P}} Y^i_{s,r}dY^j_r - X^i_{s,r}dX^j_r\Big|^2\Big).
\end{aligned}$$

The result now follows from the bound

$$\Big|\int_{\mathcal{P}} Y^i_{s,r}dY^j_r - X^i_{s,r}dX^j_r\Big| \le \Big|\int_{\mathcal{P}} Y^i_{s,r}d(Y-X)^j_r\Big| + \Big|\int_{\mathcal{P}} (Y-X)^i_{s,r}dX^j_r\Big| \ ,$$

where we estimate the second moment of each term on the right hand side by the respective variation norms of the covariances; e.g.

$$\begin{aligned}
\mathbf{E}\Big(\Big|\int_{\mathcal{P}} Y^i_{s,r}d(Y-X)^j_r\Big|^2\Big) &\le C\|R_{Y^i}\|_{\varrho;[s,t]^2}\|R_{Y^j-X^j}\|_{\varrho;[s,t]^2}\\
&\le C\varepsilon^2|t-s|^{\frac{2}{\varrho}} \ .
\end{aligned}$$

The case $i=j$ is easier: it suffices to note that

$$\begin{aligned}
\mathbf{E}\big(\big|\mathbb{Y}^{i,i}_{s,t}-\mathbb{X}^{i,i}_{s,t}\big|^2\big) &= \frac{1}{4}\mathbf{E}\big(\big(Y^i_{s,t}\big)^2-\big(X^i_{s,t}\big)^2\big)\\
&= \frac{1}{4}\big|\mathbf{E}\big(\big(Y^i_{s,t}-X^i_{s,t}\big)\big(Y^i_{s,t}+X^i_{s,t}\big)\big)\big| \ ,
\end{aligned}$$

then conclude with Cauchy–Schwarz.

Regarding b), given the pointwise L^q-estimates as stated in a), the L^q-estimates for $\|X-Y\|_\alpha$ and $\|\mathbb{Y}-\mathbb{X}\|_{2\alpha}$ are obtained from Theorem 3.3. The last statement is then an immediate consequence of the definition of ϱ_α. □

Corollary 10.6. *As above, let $(X,Y)=\big(X^1,Y^1,\dots,X^d,Y^d\big)$ be a centred continuous Gaussian process such that $\big(X^i,Y^i\big)$ is independent of $\big(X^j,Y^j\big)$ when $i\neq j$. Assume that there exists $\varrho\in[1,\frac{3}{2})$ and $M\in(0,\infty)$ such that*

$$\big\|R_{(X,Y)}\big\|_{\varrho;[s,t]^2} \le M|t-s|^{1/\varrho}\ \forall 0\le s\le t\le T. \tag{10.18}$$

Then, for every $\alpha \in (\frac{1}{3}, \frac{1}{2\varrho})$, every $\theta \in \left(0, \frac{1}{2} - \varrho\alpha\right)$ and $q < \infty$, there exists a constant C such that

$$|\varrho_\alpha(\mathbf{X}, \mathbf{Y})|_{L^q} \leq C \sup_{s,t\in[0,T]} \left[\mathbf{E}|X_{s,t} - Y_{s,t}|^2\right]^\theta. \tag{10.19}$$

Proof. At the price of replacing (X, Y) by the rescaled process $M^{-1/2}(X, Y)$ we may take $M = 1$. (The concluding L^q-estimate on $\varrho_\alpha\big(M^{-1/2}X, M^{-1/2}Y\big)$ is then readily translated into an estimate on $\varrho_\alpha(X, Y)$, given that we allow the final constant to depend on M.) Assumption (10.18) then spells out precisely to

$$\|R_{X^i}\|_{\varrho;[s,t]^2} \leq |t-s|^{1/\varrho}, \|R_{Y^i}\|_{\varrho;[s,t]^2} \leq |t-s|^{1/\varrho}$$

and (not present in the assumptions of the previous theorem!)

$$\left\|R_{(X^i,Y^i)}\right\|_{\varrho;[s,t]^2} \leq |t-s|^{1/\varrho}$$

where $R_{(X^i,Y^i)}(u, v) = \mathbf{E}\big(X^i_u Y^i_v\big)$. Thanks to this assumption we have

$$\begin{aligned}\|R_{X^i-Y^i}\|_{\varrho;[s,t]^2} &\leq C_\varrho\Big(\|R_{X^i}\|_{\varrho;[s,t]^2} + 2\left\|R_{(X^i,Y^i)}\right\|_{\varrho;[s,t]^2} + \|R_{Y^i}\|_{\varrho;[s,t]^2}\Big)\\ &\leq 4C_\varrho|t-s|^{1/\varrho}\ ,\end{aligned}$$

which is handy in the following interpolation argument. Set

$$\eta := \max\{\|R_{X^i-Y^i}\|_{\infty;[0,T]^2} : 1 \leq i \leq d\}$$

and note that, for any $\varrho' > \varrho$,

$$\begin{aligned}\|R_{X^i-Y^i}\|_{\varrho';[s,t]^2} &\leq \|R_{X^i-Y^i}\|^{1-\varrho/\varrho'}_{\infty;[s,t]^2}\|R_{X^i-Y^i}\|^{\varrho/\varrho'}_{\varrho;[s,t]^2}\\ &\leq (4C_\varrho)^{\varrho/\varrho'}\eta^{1-\varrho/\varrho'}|t-s|^{1/\varrho'}.\end{aligned}$$

Also, with $\tilde{M} = 1 \vee T^{1/p-1/p'}$, and then similar for R_{Y^i},

$$\|R_{X^i}\|_{\varrho';[s,t]^2} \leq \|R_{X^i}\|_{\varrho;[s,t]^2} \leq |t-s|^{1/\varrho} \leq \tilde{M}|t-s|^{1/\varrho'}$$

and so, picking $\varrho' = \frac{\varrho}{1-2\theta}$ the previous theorem (with $\varrho' \leftarrow \varrho$ and $\varepsilon^2 \leftarrow \eta^{1-\varrho/\varrho'}$, $M \leftarrow \tilde{M} \vee (4C_\varrho)^{\varrho/\varrho'}$ ...) yields

$$|\varrho_\alpha(X, Y)|_{L^q} \leq C\varepsilon = C\eta^{\frac{1}{2}-\varrho\frac{1}{2\varrho'}} = C\eta^\theta.$$

for any given $\theta \in \big(0, \frac{1}{2} - \varrho\alpha\big)$. At last, take $i_* \in \{1, \ldots, d\}$ as the $\arg\max$ in the definition of η and set $\varDelta = X^{i_*} - Y^{i_*}$. Then, by Cauchy–Schwarz,

$$\eta = \|R_\Delta\|_{\infty;[0,T]^2} = \sup_{\substack{0\le s\le t\le T\\ 0\le s'\le t'\le T}} \mathbf{E}(\Delta_{s,t}\Delta_{s',t'}) \le \sup_{0\le s\le t\le T} \mathbf{E}\Delta_{s,t}^2$$

and the proof is finished. □

Remark 10.7. Corollary 10.6 suggests an alternative route to the construction of a rough path lift $\mathbf{X} = (X, \mathbb{X})$ for some Gaussian process X as in Theorem 10.4. The idea is to establish the crucial estimate (10.19) only for processes with regular sample paths, in which case $\mathbb{X}$ is canonically given by iterated Riemann–Stieltjes integration. Apply this to piecewise linear (or mollifier) approximations X^n, X^m to see that $(X^n, \mathbb{X}^n)$ is Cauchy, in probability and rough path metric in the space $\mathscr{C}_g^{0,\alpha}$. The resulting limiting (random) rough path $\mathbf{X}$ is easily seen to be indistinguishable from the one construct in Theorem 10.4. All estimates are then seen to remain valid in the limit. (This is the approach taken in [FV10b].)

10.3 Fractional Brownian motion and beyond

We remarked in the beginning of Section 10.2 that (d-dimensional) fractional Brownian motion B^H, with Hurst parameter $H \in (0,1)$, determined through its covariance

$$R_H(s,t) = \frac{1}{2}\Big[s^{2H} + t^{2H} - |t-s|^{2H}\Big] \times I_d \in \mathbf{R}^{d\times d}$$

has α-Hölder sample paths for any $\alpha < H$. For $H > 1/2$, there is little need for rough path analysis - after all, Young's theory is applicable. For $H = 1/2$, one deals with d-dimensional standard Brownian motion which, of course, renders the classical martingale based stochastic analysis applicable. For $H < 1/2$, however, all these theories fail but rough path analysis works. In the remainder of this section we detail the construction of a fractional Brownian rough path.

In fact, we shall consider centred, continuous Gaussian processes with independent components $X = \big(X^1, \dots, X^d\big)$ and *stationary increments*. The construction of a (geometric) rough path associated to X then naturally passes through an understanding of the two-dimensional ϱ-variation of $R = R_X$, the covariance of X; cf. Theorem 10.4. To this end, it is enough to focus on one component and we may take X to be scalar until further notice. The law of such a process is fully determined by

$$\sigma^2(u) := \mathbf{E}\big[X_{t,t+u}^2\big] = R\begin{pmatrix} t, t+u \\ t, t+u \end{pmatrix}.$$

Lemma 10.8. *Assume that $\sigma^2(\cdot)$ is concave on $[0,h]$ for some $h > 0$. Then, one has non-positive correlation of non-overlapping increments in the sense that, for $0 \le s \le t \le u \le v \le h$,*

$$\mathbf{E}[X_{s,t}X_{u,v}] = R\begin{pmatrix} s,t \\ u,v \end{pmatrix} \le 0.$$

If in addition $\sigma^2(\cdot)$ restricted to $[0,h]$ is non-decreasing (which is always the case for some possibly smaller h), then for $0 \le s \le u \le v \le t \le h$,

$$0 \le \mathbf{E}[X_{s,t}X_{u,v}] = |\mathbf{E}[X_{s,t}X_{u,v}]| \le \mathbf{E}\big[X_{u,v}^2\big] = \sigma^2(v-u)\ .$$

Proof. Using the identity $2ac = (a+b+c)^2 + b^2 - (b+c)^2 - (a+b)^2$ with $a = X_{s,t}, b = X_{t,u}$ and $c = X_{u,v}$, we see that

$$\begin{aligned} 2\mathbf{E}[X_{s,t}X_{u,v}] &= \mathbf{E}\big[X_{s,v}^2\big] + \mathbf{E}\big[X_{t,u}^2\big] - \mathbf{E}\big[X_{t,v}^2\big] - \mathbf{E}\big[X_{s,u}^2\big] \\ &= \sigma^2(v-s) + \sigma^2(u-t) - \sigma^2(v-t) - \sigma^2(u-s). \end{aligned}$$

The first claim now easily follows from concavity, cf. [MR06, Lemma 7.2.7].

To show the second bound, note that $X_{s,t}X_{u,v} = (a+b+c)b$ where $a = X_{s,u}$, $b = X_{u,v}$, and $c = X_{v,t}$. Applying the algebraic identity

$$2(a+b+c)b = (a+b)^2 - a^2 + (c+b)^2 - c^2$$

and taking expectations yields

$$\begin{aligned} 2\mathbf{E}[X_{s,t}X_{u,v}] &= \mathbf{E}\big[X_{s,v}^2\big] - \mathbf{E}\big[X_{s,u}^2\big] + \mathbf{E}\big[X_{u,t}^2\big] - \mathbf{E}\big[X_{v,t}^2\big] \\ &= \big(\sigma^2(v-s) - \sigma^2(u-s)\big) + \big(\sigma^2(t-u) - \sigma^2(t-v)\big) \ge 0\ , \end{aligned}$$

where we used that $\sigma^2(\cdot)$ is non-decreasing. On the other hand, using $(a+b+c)b = b^2 + ab + cb$ and the non-positive correlation of non-overlapping increments, we have

$$\mathbf{E}[X_{s,t}X_{u,v}] = \mathbf{E}\big[X_{u,v}^2\big] + \mathbf{E}[X_{s,u}X_{u,v}] + \mathbf{E}[X_{v,t}X_{u,v}] \le \mathbf{E}\big[X_{u,v}^2\big]\ ,$$

thus concluding the proof. □

Theorem 10.9. *Let X be a real-valued Gaussian process with stationary increments and $\sigma^2(\cdot)$ concave and non-decreasing on $[0,h]$, some $h > 0$. Assume also, for constants $L, \varrho \ge 1$, and all $\tau \in [0,h]$,*

$$|\sigma^2(\tau)| \le L|\tau|^{1/\varrho}\ .$$

Then the covariance of X has finite ϱ-variation. More precisely

$$\|R_X\|_{\varrho\text{-var};[s,t]^2} \le M|t-s|^{1/\varrho} \tag{10.20}$$

for all intervals $[s,t]$ with length $|t-s| \le h$ and some $M = M(\varrho, L) > 0$.

Proof. Consider some interval $[s,t]$ with length $|t-s| \le h$. The proof relies on separating "diagonal" and "off-diagonal" contributions. Let $\mathcal{D} = \{t_i\}$, $\mathcal{D}' = \{t_j'\}$ be two dissections of $[s,t]$. For fixed i, we have

$$
\begin{aligned}
3^{1-\varrho} \sum_{t'_j \in \mathcal{D}'} \Big| \mathbf{E}\big(X_{t_i,t_{i+1}} X_{t'_j,t'_{j+1}}\big) \Big|^{\varrho} &\le 3^{1-\varrho} \big\| \mathbf{E} X_{t_i,t_{i+1}} X_{\cdot} \big\|^{\varrho}_{\varrho\text{-var};[s,t]} \qquad (10.21) \\
&\le \big\| \mathbf{E} X_{t_i,t_{i+1}} X_{\cdot} \big\|^{\varrho}_{\varrho\text{-var};[s,t_i]} + \big\| \mathbf{E} X_{t_i,t_{i+1}} X_{\cdot} \big\|^{\varrho}_{\varrho\text{-var};[t_i,t_{i+1}]} \\
&\quad + \big\| \mathbf{E} X_{t_i,t_{i+1}} X_{\cdot} \big\|^{\varrho}_{\varrho\text{-var};[t_{i+1},t]} .
\end{aligned}
$$

By Lemma 10.8 above, we have

$$
\begin{aligned}
\big\| \mathbf{E} X_{t_i,t_{i+1}} X_{\cdot} \big\|_{\varrho\text{-var};[s,t_i]} \le |\mathbf{E} X_{t_i,t_{i+1}} X_{s,t_i}| &\le |\mathbf{E} X_{t_i,t_{i+1}} X_{s,t_{i+1}}| + |\mathbf{E} X^2_{t_i,t_{i+1}}| \\
&\le 2\sigma^2 (t_{i+1} - t_i) \ .
\end{aligned}
$$

The third term is bounded analogously. For the middle term in (10.21) we estimate

$$
\begin{aligned}
\big\| \mathbf{E} X_{t_i,t_{i+1}} X_{\cdot} \big\|^{\varrho}_{\varrho\text{-var};[t_i,t_{i+1}]} &= \sup_{\mathcal{D}'} \sum_{t'_j \in \mathcal{D}'} |\mathbf{E} X_{t_i,t_{i+1}} X_{t'_j,t'_{j+1}}|^{\varrho} \\
&\le \sup_{\mathcal{D}'} \sum_{t'_j \in \mathcal{D}'} \big| \sigma^2 (t'_{j+1} - t'_j) \big|^{\varrho} \le L |t_{i+1} - t_i| \ ,
\end{aligned}
$$

where we used the second estimate of Lemma 10.8 for the penultimate bound and the assumption on σ^2 for the last bound. Using these estimates in (10.21) yields

$$
\sum_{t'_j \in \mathcal{D}'} |\mathbf{E} X_{t_i,t_{i+1}} X_{t'_j,t'_{j+1}}|^{\varrho} \le C |t_{i+1} - t_i| \ ,
$$

and (10.20) follows by summing over t_i and taking the supremum over all dissections of $[s,t]$. $\square$

Corollary 10.10. *Let $X = (X^1, ..., X^d)$ be a centred continuous Gaussian process with independent components such that each X^i satisfies the assumption of the previous theorem, with common values of h, L and $\varrho \in [1, 3/2)$. Then X, restricted to any interval $[0,T]$, lifts to $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^{\alpha}_g\big([0,T], \mathbf{R}^d\big)$.*

Proof. Set $I_n = [(n-1)h, nh]$ so that $[0,T] \subset I_1 \cup I_2 \cup \dots \cup I_{[T/h]+1}$. On each interval I_n, we may apply Theorem 10.4 to lift $X_n := X|_{I_n}$ to a (random) rough path $\mathbf{X}_n \in \mathscr{C}^{\alpha}_g\big(I_n, \mathbf{R}^d\big)$. The concatenation of $\mathbf{X}_1, \mathbf{X}_2, \dots$ then yields the desired rough path lift on $[0,T]$. $\square$

Example 10.11 (Fractional Brownian motion). Clearly, d-dimensional fractional Brownian motion B^H with Hurst parameter $H \in (\frac{1}{3}, \frac{1}{2}]$ satisfies the assumptions of the above theorem / corollary for all components with

$$
\sigma(u) = u^{2H} \ ,
$$

obviously non-decreasing and concave for $H \le \frac{1}{2}$ and on any time interval $[0,T]$. This also identifies

$$
\varrho = \frac{1}{2H}
$$

and $\varrho < \frac{3}{2}$ translates to $H > \frac{1}{3}$ in which case we obtain a canonical geometric rough path $\mathbf{B}^H = (B^H, \mathbb{B}^H)$ associated to fBm. In fact, a canonical "level-3" rough path $\mathbf{B}^H$ can be constructed as long as $\varrho < \varrho^* = 2$, corresponding to $H > 1/4$ but this requires level-3 considerations which we do not discuss here (see [FV10b, Ch.15]).

Example 10.12 (Ornstein-Uhlenbeck process). Consider the d-dimensional (stationary) OU process, consisting of i.i.d. copies of a scalar Gaussian process X with covariance

$$\mathbf{E}[X_s X_t] = K(|t-s|) , \qquad K(u) = \exp(-cu) ,$$

where $c > 0$ is fixed. Note that $\sigma^2(u) = \mathbf{E}X_{t,t+u}^2 = \mathbf{E}X_{t+u}^2 + \mathbf{E}X_t^2 - 2\mathbf{E}X_{t,t+u} = 2[K(0) - K(u)] = 1 - \exp(-cu)$, so that $\sigma^2(u)$ is indeed increasing and concave:

$$\partial_u \big[\sigma^2(u)\big] = c \exp(-cu) > 0$$
$$\partial_u^2 \big[\sigma^2(u)\big] = -c^2 \exp(-cu) < 0 .$$

One also has the bound $\sigma^2(u) = 1 - \exp(-cu) \le cu$, which shows that the assumptions of the above corollary are satisfied with $\varrho = 1$, $L = c$ and arbitrary $h > 0$.

10.4 Exercises

Exercise 10.13. Let $X^{\mathcal{D}}$ be a piecewise linear approximation to X. Show that $(\mathbb{X}_{s,t})$ as constructed in Theorem 10.4 is the limit, in probability and uniformly on $\{(s,t) : 0 \le s \le t \le T\}$ say, of $\int_s^t X_{s,u}^{\mathcal{D}} \otimes dX_u^{\mathcal{D}}$ as $|\mathcal{D}| \to 0$. (In particular, any algebraic relations which hold for (piecewise) smooth paths and their iterated integrals then hold true in the limit. This yields an alternative proof that $(X, \mathbb{X})$ satisfies conditions (2.1) and (2.5).)

Exercise 10.14 (Brownian rough path, rate of convergence [HN09, FR11]). Let $X = B$ and $Y = B^n$ be d-dimensional Brownian motion and piecewise linear approximations (with mesh size $1/n$), respectively. Show that the covariance of (B, B^n) has finite 1-variation, uniformly in n. Show also that

$$\sup_{s,t \in [0,T]} \Big[\mathbf{E}\big(\big|B_{s,t} - B_{s,t}^n\big|^2\big)\Big] = \mathrm{O}\Big(\frac{1}{n}\Big) .$$

Conclude that, for any $\theta < 1/2 - \alpha$

$$\Big|\|B - B^n\|_\alpha + \sqrt{\|\mathbb{B} - \mathbb{B}^n\|_{2\alpha}}\Big|_{L^q} = \mathrm{O}\Big(\frac{1}{n^\theta}\Big) .$$

Use a Borel–Cantelli argument to show that, also for any $\theta < 1/2 - \alpha$,

$$\|B - B^n\|_\alpha + \|\mathbb{B} - \mathbb{B}^n\|_{2\alpha} \le C(\omega) \frac{1}{n^\theta} .$$

When $\alpha \in \left(\frac{1}{3}, \frac{1}{2}\right)$ we can conclude convergence in α-Hölder rough path metric, i.e.

$$\varrho_\alpha((B, \mathbb{B}), (B^n, \mathbb{B}^n)) \to 0 ,$$

almost surely with rate $1/2 - \alpha - \varepsilon$ for every $\varepsilon > 0$.

Exercise 10.15. Let $(B, \tilde{B})$ be a 2-dimensional standard Brownian motion. The (Gaussian) process given by

$$X = (B_t, B_t + \tilde{B}_t)$$

fails to have independent components and yet lifts to a Gaussian rough path. Explain how and detail the construction.

Exercise 10.16. Assume $R(s,t) = K(|t-s|)$ for some $\mathcal{C}^2$-function K. (This was exactly the situation in the above Ornstein–Uhlenbeck case, Example 10.12.) Give a direct proof that R has finite 2-dimensional 1-variation, more precisely,

$$\|R\|_{1\text{-var};[s,t]^2} \le C|t-s| , \qquad \forall\, 0 \le s \le t \le T ,$$

for a constant C which depends on T and K.

Solution 10.17. If $(s,t) \mapsto R(s,t) := \mathbf{E}[X_s X_t]$ is smooth, the 2-dimensional 1-variation is given by

$$\|R\|_{1\text{-var};[0,T]^2} = \int_{[0,T]^2} \left|\partial^2_{s,t} R(s,t)\right| ds\, dt$$

This remains true when the mixed derivative is a signed measure, which in turn is the case when $R(s,t) = K(|t-s|)$ for some $\mathcal{C}^2$-function K. Indeed, write H and 2δ for the distributional derivatives of $|\cdot|$. Formal application of the chain-rule gives $\partial_t R = K'(|t-s|)H(t-s)$ and then, using $|H| \le 1$ a.s.,

$$\left|\partial^2_{s,t} R(s,t)\right| \le |K''(|t-s|)| + 2|K'(|t-s|)|\delta(t-s).$$

Integration again over $[s,t]^2 \subset [0,T]^2$ yields

$$\|R\|_{1\text{-var};[s,t]^2} = \int_{[s,t]^2} \left|\partial^2_{u,v} R(u,v)\right| du\, dv \le (T|K''|_\infty + 2|K'(0)|)|t-s|.$$

This is easily made rigorous by replacing $|\cdot|$ (and then $H, 2\delta$) by a mollified version, say $|\cdot|_\varepsilon$ (and $H_\varepsilon, 2\delta_\varepsilon$), noting that variation-norms behave in a lower-semi-continuous fashion under pointwise limits; that is

$$\|R\|_{1\text{-var};[s,t]^2} \le \liminf_{\varepsilon \to 0} \|R_\varepsilon\|_{1\text{-var};[s,t]^2}$$

whenever $R^\varepsilon \to R$ pointwise. To see this, it suffices to take arbitrary dissections $\mathcal{D} = (t_i)$ and $\mathcal{D}' = (t'_j)$ of $[u,v]$ and note that

$$\sum_{i,j}\left|R\binom{t_{i-1},t_i}{t'_{j-1},t'_j}\right| = \lim_{\varepsilon\to 0}\sum_{i,j}\left|R^{\varepsilon}\binom{t_{i-1},t_i}{t'_{j-1},t'_j}\right| \le \liminf_{\varepsilon\to 0}\|R_\varepsilon\|_{1\text{-var};[u,v]^2}\,.$$

Exercise 10.18. Assume $X = \left(X^1,\ldots,X^d\right)$ is a centred, continuous Gaussian process with independent components.

(i) Assume covariance of finite ϱ-variation with $\varrho < 2$. Show that each component $X = X^i$, for $i = 1,\ldots,d$, has almost surely *vanishing compensated quadratic variation* on $[0,T]$ by which we mean

$$\lim_{n\to\infty}\sum_{[s,t]\in\mathcal{P}_n}\left(X^2_{s,t} - \mathbf{E}(X^2_{s,t})\right) = 0\ ,$$

in probability (and L^q, any $q<\infty$) for any sequence of partitions $(\mathcal{P}_n)$ of $[0,T]$ with mesh $|\mathcal{P}_n|\to 0$.

(ii) Under the assumptions of (i), show that there exists $(\mathcal{P}_n)$ with $|\mathcal{P}_n|\to 0$ so that, with probability one, the quadratic (co)variation $\left[X^i,X^j\right]$, in the sense of definition 5.8, vanishes, for any $i\neq j$, with $i,j\in\{1,\ldots,d\}$.
Conclude that, with regard to Theorem 10.4, the off-diagonal elements $\mathbb{X}^{i,j}_{s,t}$, defined as the L^2 limit of left-point Riemann-Stieltjes sums, could have been equivalently defined via mid- or right-point Riemann sums.

(iii) Assume $\varrho = 1$. Show that, for all $i = 1,\ldots,d$, there exists a sequence $(\mathcal{P}_n)$ with mesh $|\mathcal{P}_n|\to 0$ so that, with probability, the quadratic variation $\left[X^i,X^i\right]$, in the sense of definition 5.8, exists and equals

$$\left[X^i\right]_t := \lim_{\varepsilon\to 0}\sup_{|\mathcal{P}|<\varepsilon}\sum_{\substack{[u,v]\in\mathcal{P}\\ u<t}}\mathbf{E}\big(X^i_{u,v}\big)^2.$$

Discuss the possibility of lifting X to a (random) non-geometric rough path, similar to the Itô-lift of Brownian motion.

(iv) Consider the case of a zero-mean, stationary Gaussian process on $[0,2\pi]$ with i.i.d. components, each specified by

$$\mathbf{E}(X^2_{s,t}) = \cosh(-\pi) - \cosh(|t-s|-\pi).$$

Verify that $\varrho = 1$ and compute $[X]$. (This example is related to the stochastic heat equation, where s,t should be thought of as spatial variables; cf Lemma 12.17)

Solution 10.19. (i) Using Wick's formula for the expectation of products of centred Gaussians, namely

$$\mathbf{E}[ABCD] = \mathbf{E}[AB]\mathbf{E}[CD] + \mathbf{E}[AC]\mathbf{E}[BD] + \mathbf{E}[AD]\mathbf{E}[BC]\ ,$$

we obtain the identity

$$\mathbf{E}\Big|\sum_{[s,t]\in\mathcal{P}_n}X^2_{s,t} - \mathbf{E}(X^2_{s,t})\Big|^2$$

$$
\begin{aligned}
&= \sum_{[s,t]\in\mathcal{P}_n} \sum_{[s',t']\in\mathcal{P}_n} \Big(\mathbf{E}\big[X_{s,t}^2 X_{s',t'}^2\big] - \mathbf{E}\big[X_{s,t}^2\big]\mathbf{E}(X_{s',t'}^2) \Big) \\
&= \sum_{[s,t]\in\mathcal{P}_n} \sum_{[s',t']\in\mathcal{P}_n} 2\mathbf{E}[X_{s,t}X_{s',t'}]\mathbf{E}[X_{s,t}X_{s',t'}] \\
&= 2 \sum_{[s,t]\in\mathcal{P}_n} \sum_{[s',t']\in\mathcal{P}_n} \left| R\begin{pmatrix} s\,, t \\ s', t' \end{pmatrix} \right|^2 \\
&\le \sup_{\substack{t-s\le|\mathcal{P}_n| \\ t'-s'\le|\mathcal{P}_n|}} \left| R\begin{pmatrix} s\,, t \\ s', t' \end{pmatrix} \right|^{2-\varrho} \|R\|^{\varrho}_{\varrho\text{-var};[0,T]^2} \;.
\end{aligned}
$$

This term on the other hand converges to 0 as $|\mathcal{P}_n| \to 0$. This gives L^2-convergence and hence convergence in probability. Convergence in L^q for any $q < \infty$ follows from general facts on Wiener–Itô chaos.

(ii) Left to the reader.

(iii) We fix i and drop the index. We easily see that (i) holds uniformly on compacts, say, in the sense that

$$\sup_{t\in[0,T]} \sum_{\substack{[u,v]\in D_n \\ u<t}} \big(X_{u,v}^2 - \mathbf{E}(X_{u,v}^2) \big) \to 0 \text{ as } n \to \infty$$

in probability whenever $|\mathcal{P}_n| \to 0$. On the other hand,

$$\sup_{|\mathcal{P}|<\varepsilon} \sum_{\substack{[u,v]\in\mathcal{P} \\ u<t}} \mathbf{E}(X_{u,v})^2 < \infty$$

thanks to finite 1-variation of the covariance. By monotonicity, the limit as $\varepsilon = 1/n \to 0$ exists, and we call it $[[X]]_t$. Then, along a suitable sequence $\big(\tilde{\mathcal{P}}_n\big)$,

$$[[X]]_t = \lim_n \sum_{\substack{[u,v]\in\tilde{\mathcal{P}}_n \\ u<t}} \mathbf{E}(X_{u,v})^2 \;.$$

On the other hand, at the price of passing to another subsequence also denoted by $\big(\tilde{\mathcal{P}}_n\big)$, we have

$$\sup_{t\in[0,T]} \sum_{\substack{[u,v]\in\tilde{\mathcal{P}}_n \\ u<t}} \big(X_{u,v}^2 - \mathbf{E}(X_{u,v}^2) \big) \to 0 \qquad \text{almost surely,}$$

and so with probability one, and uniformly in $t \in [0,T]$,

$$\sum_{\substack{[u,v]\in\tilde{\mathcal{P}}_n \\ u<t}} X_{u,v}^2 \to [[X]]_t \;.$$

(iv) One has $\mathbf{E}(X_{s,t}^2) = \cosh(-\pi) - \cosh(|t-s| - \pi) = \sinh(\pi)|t-s| + \mathrm{o}(|t-s|)$ and so $[X]_t = t \sinh(\pi)$.

Exercise 10.20. Assume finite 1-variation of the covariance (as e.g. defined in (10.5)) of a zero-mean Gaussian process X and $\mathbf{E}[X_{t,t+h}^2] = f(t)h + \mathrm{o}(h)$ as $h \downarrow 0$, for some $f \in \mathcal{C}([0,T], \mathbf{R})$. Show that, for every smooth test function φ,

$$\int_0^T \varphi(t) \frac{X_{t,t+h}^2}{h} \, dt \;\to\; \int_0^T \varphi(t) f(t) \, dt \qquad \text{as } h \to 0,$$

where the convergence takes places in L^q for any $q < \infty$ (and hence also in probability).

Solution 10.21. Since all types of L^q-convergence are equivalent on the finite Wiener–Itô chaos (here we only need the chaos up to level 2), it suffices to consider $q = 2$. A dissection (t_k) of $[0,T]$ is given by $t_k = kh \wedge T$. We have

$$\begin{aligned}\sum_k \frac{1}{h} \int_{t_k}^{t_{k+1}} \varphi(t) X_{t,t+h}^2 dt &= \int_0^1 d\theta \sum_k \varphi(t_k + \theta h) X_{t_k+\theta h, t_k+\theta h+h}^2 \\ &\equiv \int_0^1 \langle \varphi, \mu_{\theta,h} \rangle d\theta \;,\end{aligned}$$

where the random measure $\mu_{\theta,h} := \sum_k \delta_{t_k+\theta h} X_{t_k+\theta h, t_k+\theta h+h}^2$ acts on test functions by integration. It obviously suffices to establish $\langle \varphi, \mu_{\theta,h} \rangle \to \langle \varphi, f \rangle$ in L^2, uniformly in $\theta \in [0,1]$. Define the (random) distribution function of $\mu_{\theta,h}$

$$F(t) := \mu_{\theta,h}([0,t]) = \sum_{k: t_k+\theta h \le t} X_{t_k+\theta h, t_k+\theta h+h}^2 \;,$$

and also $\bar{F}(t) = \mathbf{E}F(t)$. Note that,

$$\bar{F}(t) = \sum_{k: t_k+\theta h \le t} f(t_k + \theta h) h + \mathrm{o}(h) \sim \int_0^t f(s) ds \quad \text{ as } h \downarrow 0,$$

uniformly in $\theta \in [0,1], t \in [0,T]$. On the other hand, the Gaussian (or Wick) identity $\mathbf{E}\big[A^2 B^2\big] - \mathbf{E}[A^2]\mathbf{E}\big[B^2\big] = 2(\mathbf{E}[AB])^2$, applied with $A = X_{t_k+\theta h, t_k+\theta h+h}$ and $B = X_{t_j+\theta h, t_j+\theta h+h}$, gives

$$\begin{aligned}\mathbf{E}\big(F(t) - \bar{F}(t)\big)^2 &= \mathbf{E}\big(F^2(t)\big) - \bar{F}^2(t) \\ &= 2 \sum_{\substack{k: t_k+\theta h \le t \\ j: t_j+\theta h \le t}} R_X \begin{pmatrix} t_k + \theta h, t_k + \theta h + h \\ t_j + \theta h, t_j + \theta h + h \end{pmatrix}^2 \\ &\lesssim \operatorname{osc}\big(R^{2-\varrho}; h\big) \to 0 \quad \text{ as } h \to 0 \;,\end{aligned}$$

uniformly in $\theta \in [0,1], t \in [0,T]$. It follows that

$$F(t) = \mu_{\theta,h}([0,t]) \to \int_0^t f(s)ds$$

in L^2, again uniformly in t and θ. Now, for fixed smooth φ, one has the bound

$$\begin{aligned}\left|\int \varphi(t)\mu_{\theta,h}(dt) - \int \varphi(t)f(t)dt\right|^2 &= \left|\int \left(\int_0^t f(s)ds - \mu_{\theta,h}([0,t])\right)\dot{\varphi}(t)dt\right|^2 \\ &\lesssim \int_0^1 \left(\int_0^t f(s)ds - \mu_{\theta,h}([0,t])\right)^2 dt\end{aligned}$$

and so

$$\mathbf{E}\left|\int \varphi(t)\mu_{\theta,h}(dt) - \int \varphi(t)f(t)dt\right|^2 \lesssim \int_0^1 \mathbf{E}\left(\int_0^t f(s)ds - \mu_{\theta,h}([0,t])\right)^2 dt\ .$$

This expression converges to 0 as $h \to 0$, uniformly in θ, thus completing the proof.

10.5 Comments

Classes of Gaussian processes which admit (canonical) lifts to random rough paths were first studied by Coutin–Qian [CQ02], with focus on fBm with Hurst parameter $H > 1/4$. Ledoux, Qian and Zhang [LQZ02] used Gaussian techniques to establish large deviation and support for the Brownian rough paths, extensions to fractional Brownian motions were investigated by Millet–Sanz-Solé [MSS06], Feyel and de la Pradelle [FdLP06], Friz–Victoir [FV07, FV06a]. When $H \leq 1/4$, there is no canonical rough path lift: as noted in [CQ02], the L^2-norm of the area associated to piecewise linear approximations to fBm diverges. See however the works of Unterberger and then Nualart–Tindel [Unt10, NT11].

The notion of two-dimensional ϱ-variation of the covariance, as adopted in this chapter, is due to Friz–Victoir, [FV10a], [FV10b, Ch.15], [FV11], and allows for an elegant and general construction of Gaussian rough paths. It also leads naturally to useful Cameron–Martin embeddings, see Section 11.1. If restricted to the "diagonal", ϱ-variation of the covariance relates to a classical criterion of Jain–Monrad [JM83]. The question remains how one checks finite ϱ-variation when faced with a non-trivial (and even non-explicit, e.g. given as Fourier series) covariance function. A general criterion based on a certain covariance measure structure (reminiscent of Kruk, Russo and Tudor [KRT07]) was recently given by Friz, Gess, Gulisashvili and Riedel [FGGR13], a special case of which is the "concavity criterion" of Theorem 10.9.

Chapter 11
Cameron–Martin regularity and applications

Abstract A continuous Gaussian process gives rise to a Gaussian measure on path-space. Thanks to variation regularity properties of Cameron–Martin paths, powerful tools from the analysis on Gaussian spaces become available. A general Fernique type theorem leads us to integrability properties of rough integrals with Gaussian integrator akin to those of classical stochastic integrals. We then discuss Malliavin calculus for differential equations driven by Gaussian rough paths. As application a version of Hörmander's theorem in this non-Markovian setting is established.

11.1 Complementary Young regularity

Although we have chosen to introduce (rough) paths subject to α-Hölder regularity, the arguments are not difficult to adapt to p-variation with $p = 1/\alpha$. In particular, one uses the p-variation semi-norm given by

$$\|X\|^p_{p\text{-var};[0,T]} = \sup_{\mathcal{P}} \sum_{[s,t]\in\mathcal{P}} |X_{s,t}|^p \;, \tag{11.1}$$

where $X \in \mathcal{C}([0,T],\mathbf{R}^d)$, say, and the supremum is taken over all partitions of $[0,T]$. The 1-variation ($p = 1$) of such a path is of course nothing but its length, possibly $+\infty$. Hölder implies variation regularity, one has the immediate estimate

$$\|X\|_{p\text{-var};[0,T]} \le T^\alpha \|X\|_{\alpha;[0,T]} .$$

Conversely, a time-change renders p-variation paths Hölder continuous with exponent $\alpha = 1/p$. Given two paths $X \in \mathcal{C}^{p\text{-var}}([0,T],\mathbf{R}^d)$, $h \in \mathcal{C}^{q\text{-var}}([0,T],\mathbf{R}^d)$ let us say that they enjoy *complementary Young regularity* if Young's condition

$$\frac{1}{p} + \frac{1}{q} > 1 \;, \tag{11.2}$$

is satisfied.

We are now interested in the regularity of Cameron–Martin paths. As in the last section, X is an $\mathbf{R}^d$-valued, continuous and centred Gaussian process on $[0,T]$, realized as $X(\omega) = \omega \in \mathcal{C}([0,T],\mathbf{R}^d)$, a Banach space under the uniform norm, equipped with a Gaussian measure. General principles of Gaussian measures on (separable) Banach spaces thus apply [Led96]. Specializing to the situation at hand, the associated *Cameron–Martin space* $\mathcal{H} \subset \mathcal{C}([0,T],\mathbf{R}^d)$ consists of paths $t \mapsto h_t = \mathbf{E}(ZX_t)$ where $Z \in \mathcal{W}^1$ is an element in the so-called *first Wiener chaos*, the L^2-closure of $\operatorname{span}\{X_t^i : t \in [0,T], 1 \le i \le d\}$, consisting of Gaussian random variables. We recall that if $h' = \mathbf{E}(Z'X_\cdot)$ denotes another element in $\mathcal{H}$, the inner product $\langle h, h'\rangle_{\mathcal{H}} = \mathbf{E}(ZZ')$ makes $\mathcal{H}$ a Hilbert space; $Z \mapsto h$ is an isometry between $\mathcal{W}^1$ and $\mathcal{H}$.

Example 11.1. (Brownian motion). Let B be a d-dimensional Brownian motion, let $g \in L^2([0,T],\mathbf{R}^d)$, and set

$$Z = \sum_{i=1}^d \int_0^T g_s^i dB_s^i \equiv \int_0^T \langle g, dB\rangle \ .$$

By Itô's isometry, $h_t^i := \mathbf{E}(ZB_t^i) = \int_0^t g_s^i ds$ so that $\dot{h} = g$ and $\|h\|_{\mathcal{H}}^2 := \mathbf{E}(Z^2) = \int_0^T |g_s|^2 ds = \|\dot{h}\|_{L^2}^2$ where $|\cdot|$ denotes Euclidean norm on $\mathbf{R}^d$. Clearly, h is of finite 1-variation, and its length is given by $\|\dot{h}\|_{L^1}$. On the other hand, Cauchy–Schwarz shows any $h \in \mathcal{H}$ is $1/2$-Hölder which, in general, "only" implies 2-variation.

The proposition below applies to Brownian motion with $\varrho = 1$, also recalling that $\|R\|_{1;[s,t]^2} = |t-s|$ in the Brownian motion case.

Proposition 11.2. *Assume the covariance $R : (s,t) \mapsto \mathbf{E}(X_s \otimes X_t)$ is of finite ϱ-variation (in 2D sense) for $\varrho \in [1,\infty)$. Then $\mathcal{H}$ is continuously embedded in the space of continuous paths of finite ϱ-variation. More, precisely, for all $h \in \mathcal{H}$ and all $s < t$ in $[0,T]$,*

$$\|h\|_{\varrho\text{-var};[s,t]} \le \|h\|_{\mathcal{H}} \sqrt{\|R\|_{\varrho\text{-var};[s,t]^2}}\,.$$

Proof. We assume X, h to be scalar. The extension to d-dimensional X is straightforward (and even trivial when X has independent components, which will always be the case for us). Let $h = \mathbf{E}(ZX_\cdot)$. By scaling, we may assume without loss of generality, that $\|h\|_{\mathcal{H}}^2 := \mathbf{E}(Z^2) = 1$. Let (t_j) be a dissection of $[s,t]$. Let ϱ' be the Hölder conjugate of ϱ. Using duality for l^ϱ-spaces, we have[1]

$$\begin{aligned}\Big(\sum_j |h_{t_j,t_{j+1}}|^\varrho\Big)^{1/\varrho} &= \sup_{\beta,|\beta|_{l^{\varrho'}}\le 1} \sum_j \langle \beta_j, h_{t_j,t_{j+1}}\rangle \\ &= \sup_{\beta,|\beta|_{l^{\varrho'}}\le 1} \mathbf{E}\Big(Z\sum_j \langle \beta_j, X_{t_j,t_{j+1}}\rangle\Big)\end{aligned}$$

[1] The case $\varrho = 1$ may be seen directly by taking $\beta_j = \operatorname{sgn}(h_{t_j,t_{j+1}})$.

$$\le \sup_{\beta, |\beta|_{l^{\varrho'}} \le 1} \sqrt{\sum_{j,k} \big\langle \beta_j \otimes \beta_k, \mathbf{E}\big(X_{t_j,t_{j+1}} \otimes X_{t_k,t_{k+1}}\big)\big\rangle}$$

$$\le \sup_{\beta, |\beta|_{l^{\varrho'}} \le 1} \sqrt{\Big(\sum_{j,k} |\beta_j|^{\varrho'} |\beta_k|^{\varrho'}\Big)^{\frac{1}{\varrho'}} \Big(\sum_{j,k} \big|\mathbf{E}\big(X_{t_j,t_{j+1}} \otimes X_{t_k,t_{k+1}}\big)\big|^{\varrho}\Big)^{\frac{1}{\varrho}}}$$

$$\le \Big(\sum_{j,k} \big|\mathbf{E}\big(X_{t_j,t_{j+1}} \otimes X_{t_k,t_{k+1}}\big)\big|^{\varrho}\Big)^{1/(2\varrho)} \le \sqrt{\|R\|_{\varrho\text{-var};[s,t]^2}}.$$

The proof is then completed by taking the supremum over all dissections (t_j) of $[0,t]$. □

Remark 11.3. It is typical (e.g. for Brownian or fractional Brownian motion, with $\varrho = 1/(2H) \ge 1$) that

$$\forall s < t \text{ in } [0,T]: \qquad \|R\|_{\varrho\text{-var};[s,t]^2} \le M|t-s|^{1/\varrho} \ .$$

In such a situation, Proposition 11.2 implies that

$$|h_{s,t}| \le \|h\|_{\varrho\text{-var};[s,t]} \le \|h\|_{\mathcal{H}} M^{1/2} |t-s|^{1/(2\varrho)} \ ,$$

which tells us that $\mathcal{H}$ is continuously embedded in the space of $1/(2\varrho)$-Hölder continuous paths (which can also be seen directly from $h_{s,t} = \mathbf{E}(Z X_{s,t})$ and Cauchy–Schwarz). The point is that $1/(2\varrho)$-Hölder only implies 2ϱ-variation regularity, in contrast to the sharper result of Proposition 11.2.

In part i) of the following lemma we allow $\mathbf{X} = (X, \mathbb{X})$ to be a (continuous) rough path of finite p-variation rather than of α-Hölder regularity. More formally, we write $\mathbf{X} \in \mathscr{C}^{p\text{-var}}$ when $p \in [2,3)$ and the analytic Hölder type condition (2.3) in the definition of a rough path is replaced by

$$\begin{aligned} \|X\|_{p\text{-var}} &\stackrel{\text{def}}{=} \Big(\sup_{\mathcal{P}} \sum_{[s,t]\in\mathcal{P}} |X_{s,t}|^p\Big)^{1/p} < \infty \ , \\ \|\mathbb{X}\|_{p/2\text{-var}} &\stackrel{\text{def}}{=} \Big(\sup_{\mathcal{P}} \sum_{[s,t]\in\mathcal{P}} |\mathbb{X}_{s,t}|^{p/2}\Big)^{2/p} < \infty \ . \end{aligned} \tag{11.3}$$

The homogeneous p-variation rough path norm over $[0,T]$ is then given by

$$|||\mathbf{X}|||_{p\text{-var};[0,T]} \stackrel{\text{def}}{=} |||\mathbf{X}|||_{p\text{-var}} \stackrel{\text{def}}{=} \|X\|_{p\text{-var}} + \sqrt{\|\mathbb{X}\|_{p/2\text{-var}}}. \tag{11.4}$$

Of course, a geometric rough path of finite p-variation, $\mathbf{X} \in \mathscr{C}_g^{p\text{-var}}$ is one for which the "first order calculus" condition (2.5) holds.

The following results will prove crucial in Section 11.2 where we will derive, based on the Gaussian isoperimetric inequality, good probabilistic estimates on Gaussian rough path objects. They are equally crucial for developing the Malliavin calculus for (Gaussian) rough differential equations in Section 11.3.

Recall from Exercise 2.19 that the translation of a rough path $\mathbf{X} = (X, \mathbb{X})$ in direction h is given by

$$T_h(\mathbf{X}) \stackrel{\text{def}}{=} \big(X^h, \mathbb{X}^h\big) \tag{11.5}$$

where $X^h := X + h$ and

$$\mathbb{X}^h_{s,t} := \mathbb{X}_{s,t} + \int_s^t h_{s,r} \otimes dX_r + \int_s^t X_{s,r} \otimes dh_r + \int_s^t h_{s,r} \otimes dh_r \;, \tag{11.6}$$

provided that h is sufficiently regular to make the final three integrals above well-defined.

Lemma 11.4. *i) Let $\mathbf{X} \in \mathscr{C}_g^{p\text{-var}}([0,T], \mathbf{R}^d)$, with $p \in [2,3)$ and consider a function $h \in \mathcal{C}^{q\text{-var}}([0,T], \mathbf{R}^d)$ with complementary Young regularity in the sense that*

$$1/p + 1/q > 1 \;.$$

Then the translation of $\mathbf{X}$ in direction h is well-defined in the sense that the integrals appearing in (11.6) are well-defined Young integrals and $T_h : \mathbf{X} \mapsto T_h(\mathbf{X})$ maps $\mathscr{C}_g^{p\text{-var}}\big([0,T], \mathbf{R}^d\big)$ into itself. Moreover, one has the estimate, for some constant $C = C(p,q)$,

$$\|T_h(\mathbf{X})\|_{p\text{-var}} \le C\Big(\|\mathbf{X}\|_{p\text{-var}} + \|h\|_{q\text{-var}}\Big) \;.$$

ii) Similarly, let $\alpha = 1/p \in (\frac{1}{3}, \frac{1}{2}]$, $\mathbf{X} \in \mathscr{C}_g^{\alpha}\big([0,T], \mathbf{R}^d\big)$ and $h : [0,T] \to \mathbf{R}^d$ again of complementary Young regularity, but now "respectful" of α-Hölder regularity in the sense that [2]

$$\|h\|_{q\text{-var};[s,t]} \le K|t-s|^{\alpha}, \tag{11.7}$$

uniformly in $0 \le s < t \le T$. Write $\|h\|_{q,\alpha}$ for the smallest constant K in the bound (11.7). Then again T_h is well-defined and now maps $\mathscr{C}_g^{\alpha}\big([0,T], \mathbf{R}^d\big)$ into itself. Moreover, one has the estimate, again with $C = C(p,q)$,

$$\|T_h(\mathbf{X})\|_{\alpha} \le C(\|\mathbf{X}\|_{\alpha} + \|h\|_{q,\alpha}) \;.$$

Proof. This is essentially a consequence of Young's inequality which gives

$$\left| \int_s^t h_{s,r} \otimes dX_r \right| \le C \|h\|_{q\text{-var};[s,t]} \|X\|_{p\text{-var};[s,t]} \;,$$

and then similar estimates for the other (Young) integrals appearing in the definition of $\mathbb{X}^h$. One then uses elementary estimates of the form $\sqrt{ab} \le a + b$ (for non-negative reals a, b), in view of the definition of *homogeneous* norm (which involves $\mathbb{X}^h$ with a square root). Details are left to the reader. □

[2] From Remark 11.3, $\|h\|_{\varrho,\alpha} \lesssim \|h\|_{\mathcal{H}}$ for all $\alpha \le \frac{1}{2\varrho}$.

By combining the Cameron–Martin regularity established in Proposition 11.2, see also Remark 11.3, with the previous lemma we obtain the following result.

Theorem 11.5. *Assume $(X_t : 0 \le t \le T)$ is a continuous d-dimensional, centred Gaussian process with independent components and covariance R such that there exists $\varrho \in [1, \frac{3}{2})$ and $M < \infty$ such that for every $i \in \{1, \dots, d\}$ and $0 \le s \le t \le T$,*

$$\|R_{X^i}\|_{\varrho\text{-var};[s,t]^2} \le M|t-s|^{1/\varrho}.$$

Let $\alpha \in (\frac{1}{3}, \frac{1}{2\varrho}]$ and $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha\big([0,T], \mathbf{R}^d\big)$ a.s. be the random Gaussian rough path constructed in Theorem 10.4. Then there exists a null set N such that for every $\omega \in N^c$ and every $h \in \mathcal{H}$,

$$T_h(\mathbf{X}(\omega)) = \mathbf{X}(\omega + h) \ .$$

Proof. Note that complementary Young regularity holds, with $p = \frac{1}{\alpha} < 3$ and $q = \varrho < \frac{3}{2}$, as is seen from $\frac{1}{p} + \frac{1}{q} > \frac{1}{3} + \frac{2}{3} > 1$. As a consequence of Lemma 11.4, the translation $T_h(\mathbf{X}(\omega))$ is well-defined whenever $\mathbf{X}(\omega) \in \mathscr{C}^\alpha$. The proof requires a close look at the precise construction of $\mathbf{X}(\omega) = (X(\omega), \mathbb{X}(\omega))$ in Theorem 10.4, using Kolmogorov's criterion to build a suitable (continuous, and then Hölder) modification from $\mathbf{X}$ restricted to dyadic times. We recall that $X(\omega) = \omega \in \mathcal{C}([0,T], \mathbf{R}^d)$. Let N_1 be the null set of ω where $X(\omega)$ fails to be of α-Hölder (or p-variation) regularity. Note that $\omega \in N_1^c$ implies $\omega + h \in N_1^c$ for all $h \in \mathcal{H}$. By the very construction of $\mathbb{X}_{s,t}$ as an L^2-limit, for fixed s, t there exists a sequence of partitions $(\mathcal{P}^m)$ of $[s,t]$ such that $\mathbb{X}_{s,t}(\omega) = \lim_m \int_{\mathcal{P}^m} X \otimes dX$ exists for a.e. ω, and we write $N_{2;s,t}$ for the null set on which this fails. The intersections of all these, for dyadic times s, t, is again a null set, denoted by N_2. Now take $\omega \in N_1^c \cap N_2^c$. For fixed dyadic s, t, consider the aforementioned partitions $(\mathcal{P}^m)$ and note

$$\begin{aligned}\int_{\mathcal{P}^m} &X(\omega + h) \otimes dX(\omega + h) \\ &= \int_{\mathcal{P}^m} X(\omega) \otimes dX(\omega) + \int_{\mathcal{P}^m} h \otimes dX + \int_{\mathcal{P}^m} X \otimes dh + \int_{\mathcal{P}^m} h \otimes dh \ .\end{aligned}$$

Thanks to $\omega \in N_1^c$ and Proposition 11.2, $X(\omega)$ and h have complementary Young regularities, which guarantees convergence of the last three integrals to their respective Young integrals. On the other hand, $\omega \in N_2^c$ guarantees that $\int_{\mathcal{P}^m} X(\omega) \otimes dX(\omega) \to \mathbb{X}_{s,t}(\omega)$. This shows that the left hand side converges, the limit being by definition $\mathbb{X}(\omega + h)$. In other words, for all $\omega \in N_1^c \cap N_2^c$, $h \in \mathcal{H}$ and dyadic times s, t,

$$T_h(\mathbf{X}(\omega))_{s,t} = \mathbf{X}(\omega + h)_{s,t} \ .$$

The construction of $\mathbf{X}_{s,t}$ for non-dyadic times was obtained by continuity (see Theorem 10.4) and the above almost-sure identity remains valid. $\square$

11.2 Concentration of measure

11.2.1 Borell's inequality

Let us first recall a remarkable isoperimetric inequality for Gaussian measures. Following [Led96], we state it in the form due to C. Borell [Bor75], but an essentially equivalent result was obtained independently by Sudakov and Tsirelson [SC74]. In order to state things in their natural generality, we consider in this section an abstract Wiener-space $(E, \mathcal{H}, \mu)$. The reader may have in mind the Banach space $E = \mathcal{C}\big([0,T], \mathbf{R}^d\big)$, equipped with norm $\|x\|_E := \sup_{0 \le t \le T} |x_t|$ and a Gaussian measure μ, the law of a d-dimensional, continuous centred Gaussian process X. In this example, the Cameron–Martin space is given by $\mathcal{H} = \big\{\mathbf{E}(X.Z) : Z \in \mathcal{W}^1\big\}$ with $\|h\|_{\mathcal{H}} = \mathbf{E}\big(Z^2\big)^{1/2}$ for $h = \mathbf{E}(X.Z)$. Let us write

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-x^2/2} dx$$

for the cumulative distribution function of a standard Gaussian, noting the elementary tail estimate

$$\bar{\Phi}(y) := 1 - \Phi(y) \le \exp\left(-y^2/2\right), \quad y \ge 0.$$

Theorem 11.6 (Borell's inequality). *Let $(E, \mathcal{H}, \mu)$ be an abstract Wiener space and $A \subset E$ a measurable Borel set with $\mu(A) > 0$ so that*

$$\hat{a} := \Phi^{-1}(\mu(A)) \in (-\infty, \infty]$$

Then, if $\mathcal{K}$ denotes the unit ball in $\mathcal{H}$, for every $r \ge 0$,

$$\mu((A + r\mathcal{K})^c) \le \bar{\Phi}(\hat{a} + r).$$

where $A + r\mathcal{K} = \{x + rh : x \in A,\ h \in \mathcal{K}\}$ is the so-called Minkowski sum.[3]

Theorem 11.7 (Generalized Fernique Theorem). *Let $a, \sigma \in (0, \infty)$ and consider measurable maps $f, g : E \to [0, \infty]$ such that*

$$A_a = \{x : g(x) \le a\}$$

has (strictly) positive μ measure[4] *and set*

$$\hat{a} := \Phi^{-1}(\mu(A_a)) \in (-\infty, \infty].$$

Assume furthermore that there exists a null-set N such that for all $x \in N^c, h \in \mathcal{H}$:

$$f(x) \le g(x - h) + \sigma \|h\|_{\mathcal{H}}. \tag{11.8}$$

[3] Measurability is a delicate matter but circumventable by reading μ as outer measure; [Led96].

[4] Unless $g = +\infty$ almost surely, this holds true for sufficiently large a.

Then f has a Gaussian tail. More precisely, for all $r > a$ and with $\bar{a} := \hat{a} - a/\sigma$,

$$\mu(\{x : f(x) > r\}) \leq \bar{\Phi}(\bar{a} + r/\sigma).$$

Proof. Note that $\mu(A_a) > 0$ implies $\hat{a} = \Phi^{-1}(\mu(A_a)) > -\infty$. We have for all $x \notin N$ and arbitrary $r, M > 0$ and $h \in r\mathcal{K}$,

$$\begin{aligned}\{x : f(x) \leq M\} &\supset \{x : g(x-h) + \sigma\|h\|_{\mathcal{H}} \leq M\} \\ &\supset \{x : g(x-h) + \sigma r \leq M\} \\ &= \{x + h : g(x) \leq M - \sigma r\}.\end{aligned}$$

Since $h \in r\mathcal{K}$ was arbitrary, this immediately implies the inclusion

$$\begin{aligned}\{x : f(x) \leq M\} &\supset \bigcup_{h \in r\mathcal{K}} \{x + h : g(x) \leq M - \sigma r\} \\ &= \{x : g(x) \leq M - \sigma r\} + r\mathcal{K}\ ,\end{aligned}$$

and we see that

$$\mu(f(x) \leq M) \geq \mu(\{x : g(x) \leq M - \sigma r\} + r\mathcal{K})\ .$$

Setting $M = \sigma r + a$ and $A := \{x : g(x) \leq a\}$, it then follows from Borell's inequality that

$$\mu(f(x) > \sigma r + a) \leq \mu((A + r\mathcal{K})^c) \leq \bar{\Phi}(\hat{a} + r)\ .$$

It then suffices to rewrite the estimate in terms of $\tilde{r} := \sigma r + a > a$, noting that $\hat{a} + r = \bar{a} + \tilde{r}/\sigma$. $\square$

Example 11.8 (Classical Fernique estimate). Take $f(x) = g(x) = \|x\|_E$. Then the assumptions of the generalized Fernique Theorem are satisfied with σ equal to the operator norm of the continuous embedding $\mathcal{H} \hookrightarrow E$. This applies in particular to Wiener measure on $\mathcal{C}\big([0,T], \mathbf{R}^d\big)$.

11.2.2 Fernique theorem for Gaussian rough paths

Theorem 11.9. *Let $(X_t : 0 \leq t \leq T)$ be a d-dimensional, centred Gaussian process with independent components and covariance R such that there exists $\varrho \in [1, \frac{3}{2})$ and $M < \infty$ such that for every $i \in \{1, \ldots, d\}$ and $0 \leq s \leq t \leq T$,*

$$\|R_{X^i}\|_{\varrho\text{-var};[s,t]^2} \leq M|t-s|^{1/\varrho}.$$

Then, for any $\alpha \in (\frac{1}{3}, \frac{1}{2\varrho})$, the associated rough path $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}_g^\alpha$ built in Theorem 10.4 is such that there exists $\eta = \eta(M, T, \alpha, \varrho)$ with

$$\mathbf{E}\exp\Big(\eta\|\mathbf{X}\|_\alpha^2\Big) < \infty\ . \tag{11.9}$$

Remark 11.10. Recall that the homogeneous "norm" $\|\mathbf{X}\|_\alpha$ was defined in (2.4) as the sum of $\|X\|_\alpha$ and $\sqrt{\|\mathbb{X}\|_{2\alpha}}$. Since $\mathbb{X}$ is "quadratic" in X (more precisely: in the second Wiener–Itô chaos), the square root is crucial for the Gaussian estimate (11.9) to hold.

Proof. Combining Theorem 11.5 with Lemma 11.4 and Proposition 11.2 shows that for a.e. ω and all $h \in \mathcal{H}$

$$\|\mathbf{X}(\omega)\|_\alpha \le C\Big(\|(\mathbf{X}(\omega - h))\|_\alpha + M^{1/2}\|h\|_{\mathcal{H}}\Big)\ .$$

We can thus apply the generalized Fernique Theorem with $f(\omega) = \|\mathbf{X}\|_\alpha(\omega)$ and $g(\omega) = Cf(\omega)$, noting that $\|\mathbf{X}\|_\alpha(\omega) < \infty$ almost surely implies that

$$A_a \stackrel{\text{def}}{=} \{x : g(x) \le a\}$$

has positive probability for a large enough (and in fact, any $a > 0$ thanks to a support theorem for Gaussian rough paths, [FV10b]). Gaussian integrability of the homogeneous rough path norm, for a fixed Gaussian rough path $\mathbf{X}$ is thus established. The claimed uniformity, $\eta = \eta(M, T, \alpha, \varrho)$ and not depending on the particular $\mathbf{X}$ under consideration requires an additional argument. We need to make sure that $\mu(A_a)$ is uniformly positive over all $\mathbf{X}$ with given bounds on the parameters (in particular M, ϱ, a, d); but this is easy, using (10.16),

$$\mu(\|\mathbf{X}\|_\alpha \le a) \ge 1 - \frac{1}{a^2}\mathbf{E}\|\mathbf{X}\|_\alpha^2 \ge 1 - \frac{1}{a^2}C\ ,$$

where $C = C(M, \varrho, \alpha, d)$ and so, say, $a = \sqrt{2C}$ would do. □

11.2.3 Integrability of rough integrals and related topics

The price of a pathwise integration / SDE theory is that all estimates (have to) deal with the worst possible scenario. To wit, given $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}_g^\alpha$ and a nice 1-form, $F \in \mathcal{C}_b^2$ say, we had the estimate

$$\left|\int_0^T F(X)d\mathbf{X}\right| \le C\Big(\|\mathbf{X}\|_{\alpha;[0,T]} \vee \|\mathbf{X}\|_{\alpha;[0,T]}^{1/\alpha}\Big)\ ,$$

where C may depend on F, T and $\alpha \in \big(\frac{1}{3}, \frac{1}{2}\big]$. In terms of p-variation, $p = 1/\alpha$, one can show similarly, with $\|\mathbf{X}\|_{p\text{-var};[0,T]}$ as introduced earlier, cf. (11.4),

$$\left|\int_0^T F(X)d\mathbf{X}\right| \le C\Big(\|\mathbf{X}\|_{p\text{-var};[0,T]} \vee \|\mathbf{X}\|_{p\text{-var};[0,T]}^p\Big)\ , \tag{11.10}$$

where C depends on F and $\alpha \in \left(\frac{1}{3}, \frac{1}{2}\right]$ but not on T, thanks to invariance under reparametrisation. For the same reason, the integration domain $[0,T]$ in (11.10) may be replaced by any other interval.

Example 11.11. The estimate (11.10) is sharp, at least when $p = 1/\alpha = 2$, in the following sense. Consider the ("pure-area") rough path given by

$$t \mapsto (0, At)\;, \quad A = \begin{pmatrix} 0 & c \\ -c & 0 \end{pmatrix}\;,$$

for some $c > 0$. The homogeneous (p-variation, or α-Hölder) rough path norm here scales with $c^{1/2}$. Hence, the right-hand side of (11.10) scales like c (for c large), as does the left-hand side which in fact is given by $T|DF(0)A|$.

The "trouble", in Brownian ($\varrho = 1$) or worse ($\varrho > 1$) regimes of Gaussian rough paths is that, despite Gaussian tails of the random variable $\|\mathbf{X}(\omega)\|_\alpha$, established in Theorem 11.9, the above estimate (11.10) fails to deliver Gaussian, or even exponential, integrability of the "random" rough integral

$$Z(\omega) \stackrel{\text{def}}{=} \int_0^T F(X(\omega))d\mathbf{X}(\omega)\;,$$

something which is rather straightforward in the context of (Itô or Stratonovich) stochastic integration against Brownian motion.

As we shall now see, Borell's inequality, in the manifestation of our generalized Fernique estimate, allows to fully close this "gap" between integrability properties. The key idea, due to Cass–Litterer–Lyons [CLL13] is to define, for a fixed rough path $\mathbf{X}$ of finite homogeneous p-variation in the sense of (11.4), a tailor-made partition[5] of $[0,T]$, say

$$\mathcal{P} = \{[\tau_i, \tau_{i+1}] : i = 0, \ldots, N\}$$

with the property that for all $i < N$

$$\|\mathbf{X}\|_{p\text{-var};[\tau_i,\tau_{i+1}]} = 1,$$

i.e. for all but the very last interval for which one has $\|\mathbf{X}\|_{p\text{-var};[\tau_N,\tau_{N+1}]} \le 1$. One can then exploit rough path estimates such as (11.10) on (small) intervals $[\tau_i, \tau_{i+1}]$ on which estimates are linear in $\|\mathbf{X}\|_{p\text{-var}} \sim 1$. The problem of estimating rough integrals is thus reduced to estimating $N = N(\mathbf{X})$ and it was a key technical result in [CLL13] to use Borell's inequality to establish good (probabilistic) estimates on N when $\mathbf{X} = \mathbf{X}(\omega)$ is a Gaussian rough path. (Our proof below is different from [CLL13] and makes good use of the generalized Fernique estimate.)

To formalize this construction, we fixed a (1D) control function $w = w(s,t)$, i.e. a continuous map on $\{0 \le s \le t \le T\}$, super-additive, continuous and zero on the

[5] The construction is purely deterministic. Of course, when $\mathbf{X} = \mathbf{X}(\omega)$ is random, then so is the partition.

diagonal.[6] The canonical example of a control in this context is[7]

$$w_{\mathbf{X}}(s,t) = \|\mathbf{X}\|^p_{p\text{-var};[s,t]}.$$

Thanks to continuity of $w = w_{\mathbf{X}}$ we can then define a partition tailor-made for $\mathbf{X}$ based on eating up unit ($\beta = 1$ below) pieces of p-variation as follows. Set

$$\tau_0 = 0\,, \quad \tau_{i+1} = \inf\{t \,:\, w(\tau_i, t) \geq \beta,\ \tau_i < t \leq T\} \wedge T\,, \tag{11.11}$$

so that $w(\tau_i, \tau_{i+1}) = \beta$ for all $i < N$, while $w(\tau_N, \tau_{N+1}) \leq \beta$, where N is given by

$$N(w) \equiv N_\beta(w;[0,T]) := \sup\{i \geq 0 :\ \tau_i < T\}.$$

As immediate consequence of super-additivity of controls,

$$\beta N_\beta(w;[0,T]) = \sum_{i=0}^{N-1} w(\tau_i, \tau_{i+1}) \leq w(0, \tau_N) \leq w(0, \tau_{N+1}) = w(0,T).$$

Note also that N is monotone in w, i.e. $w \leq \tilde{w}$ implies $N(w) \leq N(\tilde{w})$. At last, let us set $N(\mathbf{X}) = N(w_{\mathbf{X}})$. The following (purely deterministic) lemma is most naturally stated in variation regularity.

Lemma 11.12. *Assume $\mathbf{X} \in \mathscr{C}^{p\text{-var}}_g$, $p \in [2,3)$, and $h \in \mathcal{C}^{q\text{-var}}$, $q \geq 1$, of complementary Young regularity in the sense that $\frac{1}{p} + \frac{1}{q} > 1$. Then there exists $C = C(p,q)$ so that*

$$N_1(\mathbf{X};[0,T])^{\frac{1}{q}} \leq C\Big(\|T_{-h}(\mathbf{X})\|^{\frac{p}{q}}_{p\text{-var};[0,T]} + \|h\|_{q\text{-var};[0,T]}\Big). \tag{11.12}$$

Proof. (Riedel) It is easy to see that all $N_\beta, N_{\beta'}$, with $\beta, \beta' > 0$ are comparable, it is therefore enough to prove the lemma for some fixed $\beta > 0$.

Given $h \in \mathcal{C}^{q\text{-var}}$, $w_h(s,t) = \|h\|^q_{q\text{-var};[s,t]}$ is a control and so is w_h^θ whenever $\theta \geq 1$. (Noting $1 \leq q \leq p$, we shall use this fact with $\theta = p/q$.) From Lemma 11.4 we have, for any interval I

$$\|T_h \mathbf{X}\|_{p\text{-var};I} \lesssim \|\mathbf{X}\|_{p\text{-var};I} + \|h\|_{q\text{-var};I}\ .$$

Raise everything to the pth power to see that

$$(s,t) \mapsto \|T_h \mathbf{X}\|^p_{p\text{-var};[s,t]} \leq C\Big(\|\mathbf{X}\|^p_{p\text{-var};[s,t]} + \|h\|^p_{q\text{-var};[s,t]}\Big) =: C\tilde{w}(s,t)\ .$$

where $C = C(p,q)$ and $\tilde{w}$ is a control. Choose $\beta = C$. By monotonicity of N_β in the control,

[6] Do not confuse a control w with "randomness" ω.

[7] Super-additivity, i.e. $\omega(s,t) + \omega(t,u) \leq \omega(s,u)$ whenever $s \leq t \leq u$ is immediate, but continuity is non-trivial see e.g. [FV10b, Prop. 5.8])

$$N_\beta(T_h\mathbf{X};[0,T]) \le N_\beta(C\tilde{w};[0,T]) = N_1(\tilde{\omega};[0,T]).$$

By definition, $\tilde{N} := N_1(\tilde{\omega};[0,T])$ is the number of consecutive intervals $[\tau_i,\tau_{i+1}]$ for which

$$1 = \tilde{\omega}(\tau_i,\tau_{i+1}) = \|\mathbf{X}\|^p_{p\text{-var};[\tau_i,\tau_{i+1}]} + \|h\|^p_{q\text{-var};[\tau_i,\tau_{i+1}]}.$$

Using the manifest estimate $\|h\|^p_{q\text{-var};[\tau_i,\tau_{i+1}]} \le 1$ and $q/p \le 1$ we have

$$1 \le \|\mathbf{X}\|^p_{p\text{-var};[\tau_i,\tau_{i+1}]} + \|h\|^q_{q\text{-var};[\tau_i,\tau_{i+1}]} = w_{\mathbf{X}}(\tau_i,\tau_{i+1}) + w_h(\tau_i,\tau_{i+1})$$

for $0 \le i < \tilde{N}$. Summation over i yields

$$\tilde{N} \le w_{\mathbf{X}}(0,\tau_{\tilde{N}}) + w_h(0,\tau_{\tilde{N}}) \le \|\mathbf{X}\|^p_{p\text{-var};[0,T]} + \|h\|^q_{q\text{-var};[0,T]}.$$

Combination of these estimate hence shows that

$$N_\beta(T_h\mathbf{X};[0,T]) \le \|\mathbf{X}\|^p_{p\text{-var};[0,T]} + \|h\|^q_{q\text{-var};[0,T]}.$$

Replace $\mathbf{X} = T_h T_{-h}\mathbf{X}$ by $T_{-h}\mathbf{X}$ and then use elementary estimates of the type $(a+b)^{1/q} \le (a^{1/q} + b^{1/q})$ for non-negative reals a, b, to obtain the claimed estimate (11.12). □

The previous lemma, combined with variation regularity of Cameron–Martin paths (Proposition 11.2) and the generalized Fernique Theorem 11.7 then gives immediately

Theorem 11.13 (Cass–Litterer–Lyons). *Let $\mathbf{X} = (X,\mathbb{X}) \in \mathscr{C}^\alpha_g$ a.s. be a Gaussian rough path, as in Theorem 11.9. (In particular, the covariance is assumed to have finite 2D ϱ-variation.) Then the integer-valued random variable*

$$N(\omega) := N_1(\mathbf{X}(\omega);[0,T])$$

has a Weibull tail with shape parameter $2/\varrho$ (by which we mean that $N^{1/\varrho}$ has a Gaussian tail).

Let us quickly illustrate how to use the above estimate.

Corollary 11.14. *Let $\mathbf{X}$ be as in the previous theorem and assume $F \in \mathcal{C}^2_b$. Then the random rough integral*

$$Z(\omega) \stackrel{\text{def}}{=} \int_0^T F(X(\omega))d\mathbf{X}(\omega)$$

has a Weibull tail with shape parameter $2/\varrho$ by which we mean that $|Z|^{1/\varrho}$ has a Gaussian tail.

Proof. Let (τ_i) be the (random) partition associated to the p-variation of $\mathbf{X}(\omega)$ as defined in (11.11), with $\beta = 1$ and $w = w_{\mathbf{X}}$. Thanks to (11.10) we may estimate

$$
\begin{aligned}
\left|\int_0^T F(X(\omega))d\mathbf{X}(\omega)\right| &\leq \sum_{[\tau_i,\tau_{i+1}]\in\mathcal{P}} \left|\int_{\tau_i}^{\tau_{i+1}} F(X(\omega))d\mathbf{X}(\omega)\right| \\
&\lesssim (N(\omega)+1)\sup_i \left(\|\mathbf{X}\|_{p\text{-var};[\tau_i,\tau_{i+1}]} \vee \|\mathbf{X}\|^p_{p\text{-var};[\tau_i,\tau_{i+1}]}\right) \\
&= (N(\omega)+1)\ ,
\end{aligned}
$$

where the proportionality constant may depend on F, T and $\alpha \in \left(\frac{1}{3}, \frac{1}{2\varrho}\right]$. $\square$

11.3 Malliavin calculus for rough differential equations

In this section, we assume that the reader is already familiar with the basics of Malliavin calculus as exposed for example in the monographs [Mal97, Nua06].

11.3.1 Bouleau–Hirsch criterion and Hörmander's theorem

Consider some abstract Wiener space $(W,\mathcal{H},\mu)$ and a Wiener functional of the form $F: W \to \mathbf{R}^e$. In the context of stochastic - or rough differential equations (driven by Gaussian signals), the Banach space W is of the form $\mathcal{C}\big([0,T],\mathbf{R}^d\big)$ where μ describes the statistics of the driving noise. If F denotes the solution to a stochastic differential equation at some time $t\in(0,T]$, then, in general, F is not a continuous, let alone Fréchet regular, function of the driving path. However, as we will see in this section, it can be the case that for μ-almost every ω, the map $\mathcal{H}\ni h\mapsto F(\omega+h)$, i.e. $F(\omega+\cdot)$ restricted to the Cameron-Martin space $(\mathcal{H},\langle\cdot,\cdot\rangle)$ is Fréchet differentiable. (This implies $\mathbb{D}^{1,p}_{\text{loc}}$-regularity, based on the commonly used Shigekawa Sobolev space $\mathbb{D}^{1,p}$; our notation here follows [Mal97] or [Nua06, Sec. 1.2, 1.3.4].) More precisely, we introduce the following notion, see for example [Nua06, Sec. 4.1.3]:

Definition 11.15. Given an abstract Wiener space $(W,\mathcal{H},\mu)$, a random variable $F\colon W\to\mathbf{R}$ is said to be continuously $\mathcal{H}$-differentiable, in symbols $F\in\mathcal{C}^1_{\mathcal{H}}$, if for μ-almost every ω, the map

$$
\mathcal{H}\ni h\mapsto F(\omega+h)
$$

is continuously Fréchet differentiable. A vector-valued random variable is said to be in $\mathcal{C}^1_{\mathcal{H}}$ if this is the case for each of its components. In particular, μ-almost surely, $DF(\omega)=\big(DF^1(\omega),\dots,DF^e(\omega)\big)$ is a linear bounded map from $\mathcal{H}$ to $\mathbf{R}^e$.

Given an $\mathbf{R}^e$-valued random variable F in $\mathcal{C}^1_{\mathcal{H}}$, we define the *Malliavin covariance matrix*

$$
\mathcal{M}_{ij}(\omega)\stackrel{\text{def}}{=}\big\langle DF^i(\omega),DF^j(\omega)\big\rangle\ . \tag{11.13}
$$

The following well known criterion of Bouleau–Hirsch, see [BH91, Thm 5.2.2] and [Nua06, Sec. 1.2, 1.3.4] then provides a condition under which the law of F has a density with respect to Lebesgue measure:

Theorem 11.16. *Let $(W, \mathcal{H}, \mu)$ be an abstract Wiener space and let F be an $\mathbf{R}^e$-valued random variable F in $\mathcal{C}^1_{\mathcal{H}}$. If the associated Malliavin matrix $\mathcal{M}$ is invertible μ-almost surely, then the law of F is has a density with respect to Lebesgue measure on $\mathbf{R}^e$.*

Remark 11.17. Higher-order differentiability, together with control of inverse moments of $\mathcal{M}$ allow to strengthen this result to obtain smoothness of this density.

As beautifully explained in his own book [Mal97], Malliavin realised that the strong solution to the stochastic differential equation

$$dY_t = \sum_{i=1}^{d} V_i(Y_t) \circ dB_t^i \;, \tag{11.14}$$

started at $Y_0 = y_0 \in \mathbf{R}^e$ and driven along $\mathcal{C}^\infty$-bounded vector fields V_i on $\mathbf{R}^e$, gives rise to a non-degenerate Wiener functional $F = Y_T$, admitting a density with respect to Lebesgue measure, provided that the vector fields satisfy Hörmander's famous "bracket condition" at the starting point y_0:

$$\text{Lie}\,\{V_1, \ldots, V_d\}\Big|_{y_0} = \mathbf{R}^e \;. \tag{H}$$

(Here, Lie $\mathcal{V}$ denotes the Lie algebra generated by a collection $\mathcal{V}$ of smooth vector fields.) There are many variations on this theme, one can include a drift vector field (which gives rise to a modified Hörmander condition) and under the same assumptions one can show that Y_T admits a smooth density. This result can also (and was originally, see [Hör67, Koh78]) be obtained by using purely functional analytic techniques, exploiting the fact that the density solves Kolmogorov's forward equation. On the other hand, Malliavin's approach is purely stochastic and allows to go beyond the Markovian / PDE setting. In particular, we will see that it is possible to replace B by a somewhat generic sufficiently non-degenerate Gaussian process, with the interpretation of (11.14) as a random RDE driven by some Gaussian rough path $\mathbf{X}$ rather than Brownian motion.

11.3.2 Calculus of variations for ODEs and RDEs

Throughout, we assume that $V = (V_1, \ldots, V_d)$ is a given set of smooth vector fields, bounded and with bounded derivatives of all orders. In particular, there is a unique solution flow to the RDE

$$dY = V(Y)\, d\mathbf{X} \;, \tag{11.15}$$

for any α-Hölder geometric driving rough path $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}_g^{0,\alpha}$, which may be obtained as limit of smooth, or piecewise smooth, paths in α-Hölder rough path metric. Set $p = 1/\alpha$. Recall that, thanks to continuity of the Itô–Lyons maps, RDE solutions are limits of the corresponding ODE solutions.

The unique RDE solution (11.15) passing through $Y_{t_0} = y_0$ gives rise to the solution flow $y_0 \mapsto U^{\mathbf{X}}_{t\leftarrow t_0}(y_0) = Y_t$. We call the derivative of the flow with respect to the starting point the *Jacobian* and denote it by $J^{\mathbf{X}}_{t\leftarrow t_0}$, so that

$$J^{\mathbf{X}}_{t\leftarrow t_0} a = \frac{d}{d\varepsilon} U^{\mathbf{X}}_{t\leftarrow t_0}(y_0 + \varepsilon a)\Big|_{\varepsilon=0}.$$

We also consider the directional derivative

$$D_h U^{\mathbf{X}}_{t\leftarrow 0} = \frac{d}{d\varepsilon} U^{T_{\varepsilon h}\mathbf{X}}_{t\leftarrow 0}\Big|_{\varepsilon=0} ,$$

for any sufficiently smooth path $h \colon \mathbf{R}_+ \to \mathbf{R}^e$. Recall that the translation operator T_h was defined in (11.5). In particular, we have seen in Lemma 11.4 that, if $\mathbf{X}$ arises from a smooth path X together with its iterated integrals, then the translated rough path $T_h\mathbf{X}$ is nothing but $X + h$ together with its iterated integrals. In the general case, given $h \in \mathcal{C}^{q\text{-var}}$ of complementary Young regularity, i.e. with $1/p + 1/q > 1$, the translation $T_h\mathbf{X}$ can be written in terms of $\mathbf{X}$ and cross-integrals between X and h.

Suppose for a moment that the rough path $\mathbf{X}$ is the canonical lift of a smooth $\mathbf{R}^d$-valued path X. Then, it is classical to prove that $J^{\mathbf{X}}_{t\leftarrow t_0}$ solves the linear ODE

$$dJ^X_{t\leftarrow t_0} = \sum_{i=1}^d DV_i(Y_t) J^X_{t\leftarrow t_0}\, dX^i_t \,, \tag{11.16}$$

and satisfies $J^X_{t_2\leftarrow t_0} = J^X_{t_2\leftarrow t_1} \cdot J^X_{t_1\leftarrow t_0}$. Furthermore, the variation of constants formula leads to

$$D_h U^X_{t\leftarrow 0} = \int_0^t \sum_{i=1}^d J^X_{t\leftarrow s}\, V_i(Y_s)\, dh^i_s \,. \tag{11.17}$$

Similarly, given any smooth vector field W, a straightforward application of the chain rule yields

$$d\big(J^X_{0\leftarrow t}\, W(Y_t)\big) = \sum_{i=1}^d J^X_{0\leftarrow t}\, [V_i, W](Y_t)\, dX^i_t \,, \tag{11.18}$$

where $[V, W]$ denotes the Lie bracket between the vector fields V and W. All this extends to the rough path limit without difficulties. For instance, (11.16) can be interpreted as a linear equation driven by the rough path $\mathbf{X}$, using the fact that $DV(Y)$ is controlled by X to give meaning to the equation. It is then still the case that $J^{\mathbf{X}}_{t\leftarrow t_0}$ is the derivative of the flow associated to (11.15) with respect to its initial condition.

Proposition 11.18. *Let* $\mathbf{X} \in \mathscr{C}_g^{0,\alpha}([0,T],\mathbf{R}^d)$ *with* $\alpha \in (\frac{1}{3},\frac{1}{2}]$ *and* $h \in \mathcal{C}^{q\text{-var}}\big([0,T],\mathbf{R}^d\big)$ *with complementary Young regularity in the sense that* $\alpha + \frac{1}{q} > 1$. *Then*

$$D_h U^{\mathbf{X}}_{t\leftarrow 0}(y_0) = \int_0^t \sum_{i=1}^d J^{\mathbf{X}}_{t\leftarrow s}\big(V_i\big(U^{\mathbf{X}}_{s\leftarrow 0}\big)\big)dh^i_s \tag{11.19}$$

where the right hand side is well-defined as Young integral.

Proof. Both $J^{\mathbf{X}}_{t\leftarrow 0}$ and $D_h U^{\mathbf{X}}_{t\leftarrow 0}$ satisfy (jointly with $U^{\mathbf{X}}_{t\leftarrow 0}$) a RDE driven by $\mathbf{X}$. This is well known in the ODE case, i.e. when both X, h are smooth, (Duhamel's principle, variation of constant formula, ...) and remains valid in the geometric rough path limit by appealing to continuity of the Itô–Lyons and continuity properties of the Young integral. A little care is needed since the resulting vector fields are not bounded anymore. It suffices to rule out explosion so that the problem can be localized. The required remark is that that $J^{\mathbf{X}}_{t\leftarrow 0}$ also satisfies a linear RDE of form

$$dJ^{\mathbf{X}}_{t\leftarrow 0} = d\mathbf{M}^{\mathbf{X}} \cdot J^{\mathbf{X}}_{t\leftarrow 0}(y_0)$$

and linear RDEs do not explode; cf. Exercise 8.12. □

Consider now an RDE driven by a Gaussian rough path $\mathbf{X} = \mathbf{X}(\omega)$. We now show that the $\mathbf{R}^e$-valued random variable obtained from solving this random RDE enjoys $\mathcal{C}^1_{\mathcal{H}}$-regularity.

Proposition 11.19. *With* $\varrho \in [1,\frac{3}{2})$ *and* $\alpha \in (\frac{1}{3},\frac{1}{2\varrho})$, *let* $\mathbf{X} = (X,\mathbb{X}) \in \mathscr{C}_g^{\alpha}$ *be a Gaussian rough path as constructed in Theorem 10.4. For fixed* $t \geq 0$, *the* $\mathbf{R}^e$*-valued random variable*

$$\omega \mapsto U^{\mathbf{X}(\omega)}_{t\leftarrow 0}(y_0)$$

is continuously $\mathcal{H}$*-differentiable.*

Proof. Recall $h \in \mathcal{H} \subset \mathcal{C}^{\varrho\text{-var}}$ so that a.e. $\mathbf{X}(\omega)$ and h enjoy complementary Young regularity. As a consequence, we saw that the event

$$\{\omega : \mathbf{X}(\omega + h) \equiv T_h\mathbf{X}(\omega) \text{ for all } h \in \mathcal{H}\} \tag{11.20}$$

has full measure. We show that $h \in \mathcal{H} \mapsto U^{\mathbf{X}(\omega+h)}_{t\leftarrow 0}(y_0)$ is continuously Fréchet differentiable for every ω in the above set of full measure. By basic facts of Fréchet theory, we must show (a) Gateaux differentiability and (b) continuity of the Gateaux differential.
Ad (a): Using $\mathbf{X}(\omega + g + h) \equiv T_g T_h \mathbf{X}(\omega)$ for $g, h \in \mathcal{H}$ it suffices to show Gateaux differentiability of $U^{\mathbf{X}(\omega+\cdot)}_{t\leftarrow 0}(y_0)$ at $0 \in \mathcal{H}$. For fixed t, define

$$Z_{i,s} \equiv J^{\mathbf{X}}_{t\leftarrow s}\big(V_i\big(U^{\mathbf{X}}_{s\leftarrow 0}\big)\big).$$

Note that $s \mapsto Z_{i,s}$ is of finite p-variation, with $p = 1/\alpha$. We have, with implicit summation over i,

$$
\begin{aligned}
\left|D_h U^{\mathbf{X}}_{t\leftarrow 0}(y_0)\right| &= \left|\int_0^t J^{\mathbf{X}}_{t\leftarrow s}\big(V_i\big(U^{\mathbf{X}}_{s\leftarrow 0}\big)\big)dh^i_s\right| \\
&= \left|\int_0^t Z_i dh^i\right| \\
&\lesssim (\|Z\|_{p\text{-var}} + |Z(0)|) \times \|h\|_{\varrho\text{-var}} \\
&\lesssim (\|Z\|_{p\text{-var}} + |Z(0)|) \times \|h\|_{\mathcal{H}}.
\end{aligned}
$$

Hence, the linear map $DU^{\mathbf{X}}_{t\leftarrow 0}(y_0) : h \mapsto D_h U^{\mathbf{X}}_{t\leftarrow 0}(y_0) \in \mathbf{R}^e$ is bounded and each component is an element of $\mathcal{H}^*$. We just showed that

$$
h \mapsto \left\{\frac{d}{d\varepsilon} U^{T_{\varepsilon h}\mathbf{X}(\omega)}_{t\leftarrow 0}(y_0)\right\}_{\varepsilon=0} = \left\langle DU^{\mathbf{X}(\omega)}_{t\leftarrow 0}(y_0), h\right\rangle_{\mathcal{H}}
$$

and hence

$$
h \mapsto \left\{\frac{d}{d\varepsilon} U^{\mathbf{X}(\omega+\varepsilon h)}_{t\leftarrow 0}(y_0)\right\}_{\varepsilon=0} = \left\langle DU^{\mathbf{X}(\omega)}_{t\leftarrow 0}(y_0), h\right\rangle_{\mathcal{H}}
$$

emphasizing again that $\mathbf{X}(\omega + h) \equiv T_h\mathbf{X}(\omega)$ almost surely for all $h \in \mathcal{H}$ simultaneously. Repeating the argument with $T_g\mathbf{X}(\omega) = \mathbf{X}(\omega + g)$ shows that the Gateaux differential of $U^{\mathbf{X}(\omega+\cdot)}_{t\leftarrow 0}$ at $g \in \mathcal{H}$ is given by

$$
DU^{\mathbf{X}(\omega+g)}_{t\leftarrow 0} = DU^{T_g\mathbf{X}(\omega)}_{t\leftarrow 0}.
$$

(b) It remains to be seen that $g \in \mathcal{H} \mapsto DU^{T_g\mathbf{X}(\omega)}_{t\leftarrow 0} \in \mathcal{L}(\mathcal{H}, \mathbf{R}^e)$, the space of linear bounded maps equipped with operator norm, is continuous. We leave this as exercise to the reader, cf. Exercise 11.23 below. □

11.3.3 Hörmander's theorem for Gaussian RDEs

Recall that $\varrho \in [1, \frac{3}{2}), \alpha \in (\frac{1}{3}, \frac{1}{2\varrho})$ and $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^{\alpha}_g$ a.s. is the Gaussian rough path constructed in Theorem 10.4. Any $h \in \mathcal{H} \subset \mathcal{C}^{\varrho\text{-var}}$ and a.e. $\mathbf{X}(\omega)$ enjoy complementary Young regularity. We now present the remaining conditions on X, followed by some commentary on each of the conditions, explaining their significance in the context of the problem and verifying them for some explicit examples of Gaussian processes.

Condition 1 *Fix $T > 0$. For every $t \in (0, T]$ we assume non-degeneracy of the law of X on $[0, t]$ in the following sense. Given $f \in \mathcal{C}^{\alpha}([0, t], \mathbf{R}^d)$, if $\sum_{j=1}^d \int_0^t f_j dh^j = 0$ for all $h \in \mathcal{H}$, then one has $f = 0$.*

Note that, thanks to complementary Young regularity, the integral $\int_0^t f_j dh^j$ makes sense as a Young integral. Some assumption along the lines of Condition 1 is certainly necessary: just consider the trivial rough differential equation $dY = dX$, starting at $Y_0 = 0$, with driving process $X = X(\omega)$ given by a Brownian bridge which returns to the origin at time T (i.e. $X_t = B_t - \frac{t}{T} B_T$ in terms of a standard Brownian motion B). Clearly $Y_T = X_T = 0$ and so Y_T does not admit a density, despite the equation $dY = dX$ being even "elliptic". However, it is straightforward to verify that in this example $\int_0^T dh = 0$ for every h belonging to the Cameron-Martin space of the Brownian bridge, so that Condition 1 is violated by taking for f a non-zero constant function.

Condition 2 *With probability one, sample paths of X are truly rough, at least in a right-neighbourhood of* 0.

These conditions obviously hold for d-dimensional Brownian motion: the first condition is satisfied because 0 is the only (continuous) function orthogonal to all of $L^2([0,T], \mathbf{R}^d)$; the second condition was already verified in Section 6.3. More interestingly, these conditions are very robust and also hold for the Ornstein-Uhlenbeck process, a Brownian bridge which returns to the origin at a time strictly greater than T, and some non-semimartingale examples such as fractional Brownian motion, including the rough regime of Hurst parameter less than $1/2$. We now show that under these conditions the process admits a density at strictly positive times. Note that the aforementioned situations are not at all covered by the "usual" Hörmander theorem.

Theorem 11.20. *With $\varrho \in [1, \frac{3}{2})$ and $\alpha \in (\frac{1}{3}, \frac{1}{2\varrho})$, let $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}_g^\alpha$ be a Gaussian rough path as constructed in Theorem 10.4. Assume that the Gaussian process X satisfies Conditions 1 and 2. Let $V = (V_1, \ldots, V_d)$ be a collection of $\mathcal{C}^\infty$-bounded vector fields on $\mathbf{R}^e$, which satisfies Hörmander's condition (H) at some point $y_0 \in \mathbf{R}^e$. Then the law of the RDE solution*

$$dY_t = V(Y_t)\, d\mathbf{X}_t\,, \qquad Y(0) = y_0\,,$$

admits a density with respect to Lebesgue measure on $\mathbf{R}^e$ for all $t \in (0, T]$.

Proof. Thanks to Proposition 11.19 and in view of the Bouleau–Hirsch criterion, Theorem 11.16 we only need to show almost sure invertibility of the Malliavin matrix associated to the solution map. As a consequence of (11.13) and (11.19), we have for every $z \in \mathbf{R}^e$ the identity

$$z^\mathsf{T} \mathcal{M}_t z = \sum_{j=1}^d \left\| z^\mathsf{T} J^{\mathbf{X}}_{t \leftarrow \cdot} V_j(Y_\cdot) \right\|_t^2\,,$$

where we wrote $\|\cdot\|_t$ for the norm given by

$$\|f\|_t = \sup_{h \in \mathcal{H}\,:\, \|h\|=1} \int_0^t f(s)\, dh(s)\,.$$

Before we proceed we note that, by the multiplicative property of $J^{\mathbf{X}}_{t \leftarrow s}$, see the remark following (11.16), one has

$$\mathcal{M}_t = J^{\mathbf{X}}_{t\leftarrow 0}\tilde{\mathcal{M}}_t\big(J^{\mathbf{X}}_{t\leftarrow 0}\big)^{\mathsf{T}}\;,$$

where $\tilde{\mathcal{M}}_t$ is given by

$$z^{\mathsf{T}}\tilde{\mathcal{M}}_t z = \sum_{j=1}^{d}\big\|z^{\mathsf{T}}J^{\mathbf{X}}_{0\leftarrow\cdot}V_j(Y_\cdot)\big\|_t^2\;.$$

Since we know that the Jacobian is invertible, invertibility of $\mathcal{M}_t$ is equivalent to that of $\tilde{\mathcal{M}}_t$, and it is the invertibility of the latter that we are going to show.

Assume now by contradiction that $\tilde{\mathcal{M}}_t$ is not almost surely invertible. This implies that there exists a random unit vector $z \in \mathbf{R}^e$ such that $z^{\mathsf{T}}\tilde{\mathcal{M}}_t z = 0$ with non-zero probability. It follows immediately from Condition 1 that, with non-zero probability, the functions $s \mapsto z^{\mathsf{T}}J^{\mathbf{X}(\omega)}_{0\leftarrow s}V_j(Y_s)$ vanish identically on $[0,t]$ for every $j \in \{1,\dots,d\}$. By (11.18), this is equivalent to

$$\sum_{i=1}^{d}\int_0^{\cdot} z^{\mathsf{T}}J^{\mathbf{X}}_{0\leftarrow s}\,[V_i,V_j](Y_s)\,d\mathbf{X}^i(s) \equiv 0$$

on $[0,t]$. Thanks to Condition 2, true roughness of X, we can apply Theorem 6.5 to conclude that one has

$$z^{\mathsf{T}}J^{\mathbf{X}}_{0\leftarrow\cdot}\,[V_i,V_j](Y_\cdot) \equiv 0\;,$$

for every $i,j \in \{1,\dots,d\}$. Iterating this argument shows that, with non-zero probability, the processes $s \mapsto z^{\mathsf{T}}J^{\mathbf{X}}_{0\leftarrow s}\,W(Y_s)$ vanish identically for every vector field W obtained as a Lie bracket of the vector fields V_i. In particular, this is the case for $s = 0$, which implies that with positive probability, z is orthogonal to $W(z_0)$ for all such vector fields. Since Hörmander's condition (H) asserts precisely that these vector fields span the tangent space at the starting point y_0, we conclude that $z = 0$ with positive probability, which is in contradiction with the fact that z is a random unit vector and thus concludes the proof. □

11.4 Exercises

Exercise 11.21 (Improved Cameron–Martin regularity, [FGGR13]). A combination of Theorem 10.9 with the Cameron–Martin embedding, Proposition 11.2, shows that every Cameron–Martin path associated to a Gaussian process enjoys finite q-variation regularity with $q = \varrho$. Show that, under the assumptions of Theorem 10.9, this can be improved to

$$q = \frac{1}{\frac{1}{2}+\frac{1}{2\varrho}}\;. \tag{11.21}$$

As a consequence, "complementary Young regularity", now holds for all $\varrho < 2$. In the fBm setting, this covers every Hurst parameter $H > 1/4$. (To exploit this in the newly covered regime $H \in (1/4, 1/3]$, one would need to work in a "level-3" rough path setting.)

Exercise 11.22. Formulate a quantitative version of Theorem 11.14. Show in particular that the Gaussian tail of $|Z|^{1/\varrho}$ is uniform over rough integrals against Gaussian rough paths, provided that $\|F\|_{\mathcal{C}_b^2}$ and the ϱ-variation of the covariance, say in the form of the constant M in Theorem 11.9, are uniformly bounded.

Exercise 11.23. Finish the proof of part (b) of Proposition 11.19.

Solution 11.24. In the notation of the (proof of) this Proposition, we have to show that $g \in \mathcal{H} \mapsto DU_{t\leftarrow 0}^{T_g \mathbf{X}(\omega)} \in \mathcal{L}(\mathcal{H}, \mathbf{R}^e)$ is continuous. To this end, assume $g_n \to g$ in $\mathcal{H}$ (and hence in $\mathcal{C}^{\varrho\text{-var}}$). Continuity properties of the Young integral imply continuity of the translation operator viewed as map $h \in \mathcal{C}^{\varrho\text{-var}} \mapsto T_h \mathbf{X}(\omega)$ and so

$$T_{g_n} \mathbf{X}(\omega) \to T_g \mathbf{X}(\omega)$$

in p-variation rough path metric. To point here is that

$$\mathbf{x} \mapsto J_{t\leftarrow \cdot}^{\mathbf{x}} \text{ and } J_{t\leftarrow \cdot}^{\mathbf{x}}(V_i(U_{\cdot\leftarrow 0}^{\mathbf{x}})) \in \mathcal{C}^{p\text{-var}}$$

depends continuously on $\mathbf{x}$ with respect to p-variation rough path metric: using the fact that $J_{t\leftarrow \cdot}^{\mathbf{x}}$ and $U_{\cdot\leftarrow 0}^{\mathbf{x}}$ both satisfy rough differential equations driven by $\mathbf{x}$ this is just a consequence of Lyons' limit theorem (the *universal limit theorem* of rough path theory). We apply this with $\mathbf{x} = \mathbf{X}(\omega)$ where ω remains a fixed element in (11.20). It follows that

$$\left\| DU_{t\leftarrow 0}^{T_{g_n}\mathbf{X}(\omega)} - DU_{t\leftarrow 0}^{T_g\mathbf{X}(\omega)} \right\|_{op} = \sup_{h:\|h\|_{\mathcal{H}}=1} \left| D_h U_{t\leftarrow 0}^{T_{g_n}\mathbf{X}(\omega)} - D_h U_{t\leftarrow 0}^{T_g\mathbf{X}(\omega)} \right|$$

and defining $Z_i^g(s) \equiv J_{t\leftarrow s}^{T_g\mathbf{X}(\omega)}\Big(V_i\Big(U_{s\leftarrow 0}^{T_g\mathbf{X}(\omega)}\Big)\Big)$, and similarly $Z_i^{g_n}(s)$, the same reasoning as in part (a) leads to the estimate

$$\left\| DU_{t\leftarrow 0}^{T_{g_n}\mathbf{X}(\omega)} - DU_{t\leftarrow 0}^{T_g\mathbf{X}(\omega)} \right\|_{op} \le c\Big(|Z^{g_n} - Z^g|_{p\text{-var}} + |Z^{g_n}(0) - Z^g(0)| \Big).$$

From the explanations just given this tends to zero as $n \to \infty$ which establishes continuity of the Gateaux differential, as required, and the proof is finished.

Exercise 11.25. Prove Theorem 11.20 in presence of a drift vector field V_0. In particular, show that in this case condition (H) can be weakened to

$$\mathrm{Lie}\left\{V_1, \dots, V_d, [V_0, V_1], \dots, [V_0, V_d]\right\}\Big|_{y_0} = \mathbf{R}^e \ . \tag{11.22}$$

11.5 Comments

Section 11.1: Regularity of Cameron–Martin paths (q-variation, with $q = \varrho$) under the assumption of finite ϱ-variation of the covariance was established in Friz–Victoir, [FV10a], see also [FV10b, Ch.15]. In the context of Gaussian rough paths, this leads to complementary Young regularity (CYR) whenever $\varrho < \frac{3}{2}$ which covers general "level-2" Gaussian rough paths as discussed in Chapter 10. On the other hand, "level-3" Gaussian rough paths can be constructed for any $\varrho < 2$ which includes fBm with $H = \frac{1}{2\varrho} > \frac{1}{4}$). A sharper Cameron regularity result specific to fBm follows from a Besov–variation embedding theorem [FV06b], thereby leading to CYR for all $H > \frac{1}{4}$. The general case was understood in [FGGR13]: one can take q as in (11.21), provided one makes the slightly stronger assumption of finite "mixed" $(1, \varrho)$-variation of the covariance. The conclusion concerning ϱ-variation of Theorem 10.9 can in fact be strengthened to finite mixed $(1, \varrho)$-variation at no extra prize and indeed this theorem is only a special case of a general criterion given in [FGGR13].

Section 11.2: Theorem 11.9 was originally obtained by careful tracking of constants via the Garsia–Rodemich–Rumsey Lemma, see [FV10b]. The generalized Fernique estimate is taken from Friz–Oberhauser and then Diehl, Oberhauser and Riedel [FO10, DOR13]. It yields an elegant proof of Theorem 11.13 with which Cass, Litterer, and Lyons [CLL13] have overcome the longstanding problem of obtaining moment bounds for the Jacobian of the flow of a rough differential equation driven by Gaussian rough paths, thereby paving the way for the proof of the Hörmander-type results, see below. As was illustrated, this above methodology can be adapted to many other situations of interest, a number of which are discussed in [FR13].

Section 11.3: Baudoin–Hairer [BH07] proved a Hörmander theorem for differential equations driven by fBm in the regular regime of Hurst parameter $H > 1/2$ in a framework of Young differential equations. The Brownian case $H = 1/2$ of course classical, see the monographs [Nua06, Mal97] or the original articles [Mal78, KS84, KS85, KS87, Bis81b, Bis81a, Nor86], a short self-contained proof can be found in [Hai11a]. In the case of rough differential equations driven by less regular Gaussian processes (including fBm with $H > 1/4$), the relevance of complementary Young regularity of Cameron–Martin paths to Malliavin regularity or (Gaussian) RDE solutions was first recognised by Cass, Friz and Victoir [CFV09]. Existence of a density under Hörmander's condition for such RDEs was obtained by Cass–Friz [CF10], see also [FV10b, Ch.20], but with a Stroock-Varadhan support type argument instead of true roughness (already commented on at the end of Chapter 6.) Smoothness of densities was subsequently established by Hairer–Pillai [HP13] in the case case of fBm and then Cass, Hairer, Litterer and Tindel [CHLT12] in the general Gaussian setting of Chapter 10, making crucial use of the integrability estimates discussed in Section 11.2. Indeed, combined with known estimates for the Jacobian of RDE flows (Friz–Victoir, [FV10b, Thm 10.16]) one readily obtains finite moments of the Jacobian of the inverse flow, a key ingredient in the smoothness proof via Malliavin calculus. See also Inahama [Ina13] for a discussion about higher-order Malliavin differentiability of Gaussian RDE solutions.

Chapter 12
Stochastic partial differential equations

Abstract Second order stochastic partial differential equations are discussed from a rough path point of view. In the linear and finite-dimensional noise case we follow a Feynman–Kac approach which makes good use of concentration of measure results, as those obtained in Section 11.2. Alternatively, one can proceed by flow decomposition and this approach also works in a number of non-linear situations. Secondly, now motivated by some semi-linear SPDEs of Burgers' type with infinite-dimension noise, we study the stochastic heat equation (in space dimension 1) as evolution in Gaussian rough path space *relative to the spatial variable*, in the sense of Chapter 10.

12.1 Rough partial differential equations

12.1.1 Linear theory: Feynman–Kac

The second order stochastic partial differential equations we will be concerned with here take the form of a terminal value problem,

$$-du = L[u]dt + \sum_{i=1}^{d} \Gamma_i[u] \circ dW_t^i(\omega) , \qquad u(T,\cdot) = g , \tag{12.1}$$

for $u = u(\omega) : [0,T] \times \mathbf{R}^n \to \mathbf{R}$, with differential operators L and Γ_i given by

$$\begin{aligned} L[u] &\overset{\text{def}}{=} \frac{1}{2}\mathrm{Tr}\big(\sigma(x)\sigma^T(x)D^2u\big) + \langle b(x), Du\rangle + c(x)u , \\ \Gamma_i[u] &\overset{\text{def}}{=} \langle \beta_i(x), Du\rangle + \gamma_i(x)u . \end{aligned} \tag{12.2}$$

The coefficients $\sigma = (\sigma_1, \ldots, \sigma_m)$, b and $\beta = (\beta_1, \ldots, \beta_d)$ are viewed as vector fields on $\mathbf{R}^n$, while $c, \gamma_1, \ldots, \gamma_d$ are scalar functions. For simplicity only, all coefficients are assumed to be bounded with bounded derivatives of all orders (but see Remark 12.3). We assume $g \in \mathcal{BC}(\mathbf{R}^n)$, that is bounded and continuous.[1] The reader may think of $\circ dW$ in (12.1) as Stratonovich differential of a d-dimensional Brownian motion. But of course, we are interested in replacing W by a genuine (geometric) rough path $\mathbf{W}$, such as to give meaning to the *rough partial differential equation* (RPDE)

$$-du = L[u]dt + \Gamma[u]d\mathbf{W} , \qquad u(T, \cdot) = g . \tag{12.3}$$

To this end, since geometric rough paths are limits of smooth paths, we first consider the case $W \in \mathcal{C}^1\big([0,T], \mathbf{R}^d\big)$. It is a basic exercise in Itô-calculus, that any bounded $\mathcal{C}^{1,2}$ solution to

$$-\partial_t u = L[u] + \sum_{i=1}^{d} \Gamma_i[u]\dot{W}_t^i , \qquad u(T, \cdot) = g , \tag{12.4}$$

is given by the classical Feynman–Kac formula (and hence also unique),

$$\begin{aligned} u(s,x) \ &= \ \mathbf{E}^{s,x}\left[g(X_T)\exp\left(\int_s^T c(X_t)dt + \int_s^T \gamma(X_t)\dot{W}_t dt\right)\right] && (12.5)\\ &=: \mathcal{S}[W; g](s,x), && (12.6) \end{aligned}$$

where X is the (unique) strong solution to

$$dX_t = \sigma(X_t)dB(\omega) + b(X_t)dt + \beta(X_t)\dot{W}_t dt, \tag{12.7}$$

where B is a m-dimensional standard Brownian motion.

Remark 12.1. The natural form of the Feynman–Kac formula is the reason for considering terminal value problems here, rather than Cauchy problems of the form $\partial_t u = L[u] + \Gamma[u]\dot{W}$ with given initial data $u(0,\cdot)$. Of course, a change of the time variable $t \mapsto T - t$ allows to switch between these problems.

Clearly, there are situations when solutions cannot be expected to be $\mathcal{C}^{1,2}$, notably when $g \notin \mathcal{C}^2$ and L fails to provide smoothing as is the case, for example, in "transport" equations where L is of first order. In such a case, formula (12.5) is a perfectly good way to define a generalized solution to (12.4). Such a solution need not be $\mathcal{C}^{1,2}$ although it is bounded and continuous on $[0,T] \times \mathbf{R}^n$, as one can see directly from (12.5). As a matter of fact, (12.5) yields a (analytically) weak PDE solution (cf. Exercise 12.22). It is also a stochastic representation of the unique (bounded) viscosity solution [CIL92, FS06] to (12.4) although this will play no rôle for us in the present section.

Theorem 12.2. *Let* $\alpha \in (\frac{1}{3}, \frac{1}{2}]$. *Given a geometric rough path* $\mathbf{W} = (W, \mathbb{W}) \in \mathscr{C}_g^{0,\alpha}([0,T], \mathbf{R}^d)$*, pick* $W^\varepsilon \in \mathcal{C}^1\big([0,T], \mathbf{R}^d\big)$ *so that*

[1] In contrast to the space $\mathcal{C}_b$ we shall equip $\mathcal{BC}$ with the topology of locally uniform convergence.

$$(W^\varepsilon, \mathbb{W}^\varepsilon) := \left(W^\varepsilon, \int_0^\cdot W^\varepsilon_{0,t} \otimes dW^\varepsilon_t \right) \to \mathbf{W}$$

in α-Hölder rough path metric. Then there exists $u = u(t,x) \in \mathcal{BC}([0,T] \times \mathbf{R}^n)$, not dependent on the approximating (W^ε) but only on $\mathbf{W} \in \mathscr{C}_g^{0,\alpha}([0,T], \mathbf{R}^d)$, so that, for $g \in \mathcal{BC}(\mathbf{R}^n)$,

$$u^\varepsilon = \mathcal{S}[W^\varepsilon; g] \to u =: \mathcal{S}[\mathbf{W}; g]$$

as $\varepsilon \to 0$ in the sense of locally uniform convergence. Moreover, the resulting solution map

$$\mathcal{S} : \mathscr{C}_g^{0,\alpha}([0,T], \mathbf{R}^d) \times \mathcal{BC}(\mathbf{R}^n) \to \mathcal{BC}([0,T] \times \mathbf{R}^n)$$

is continuous. We say that u satisfies the RPDE (12.3).

Proof. **Step 1:** Write $X = X^W$ for the solution to (12.7) whenever $W \in \mathcal{C}^1$. The first step is to make sense of the *hybrid Itô-rough differential equation*

$$dX_t = \sigma(X_t) dB_t + b(X_t) dt + \beta(X_t) d\mathbf{W}_t. \tag{12.8}$$

This is clearly not an equation that can be solved by Itô theory alone. But is also not immediately well-posed as rough differential equation since for this we would need to understand B and $\mathbf{W} = (W, \mathbb{W})$ jointly as a rough path. In view of the Itô-differential dB in (12.8), we take $\left(B, \mathbb{B}^{\text{Itô}}\right)$, as constructed in Section 3.2), and are basically short of the cross-integrals between B and W. (For simplicity of notation only, pretend over the next few lines W, B to be scalar.) We can define $\int W dB(\omega)$ as Wiener integral (Itô with deterministic integrand), and then $\int B dW = WB - \int W dB$ by imposing integration by parts. We then easily get the estimate

$$E\left(\int_s^t W_{s,r} dB_r \right)^2 \lesssim \|W\|_\alpha^2 |t-s|^{2\alpha+1} ,$$

also when switching the roles of W, B, thanks to the integration by parts formula. It follows from Kolmogorov's criterion that $\mathbf{Z}^W(\omega) := \mathbf{Z} = (Z, \mathbb{Z}) \in \mathscr{C}^{\alpha'}$ a.s. for any $\alpha' \in (1/3, \alpha)$ where

$$Z_t = \begin{pmatrix} B_t(\omega) \\ W_t \end{pmatrix}, \qquad \mathbb{Z}_{s,t} = \begin{pmatrix} \mathbb{B}^{\text{Itô}}_{s,t}(\omega) & \int_s^t W_{s,r} \otimes dB_r(\omega) \\ \int_s^t B_{s,r} \otimes dW_r(\omega) & \mathbb{W}_{s,t} \end{pmatrix}$$

where we reverted to tensor notation reflecting the multidimensional nature of B, W. It is easy to deduce from Theorem 3.3 that, for any $q < \infty$,

$$\left| \varrho_{\alpha'}\left(\mathbf{Z}^{\mathbf{W}}, \mathbf{Z}^{\tilde{\mathbf{W}}} \right) \right|_{L^q} \lesssim \varrho_\alpha\left(\mathbf{W}, \tilde{\mathbf{W}} \right). \tag{12.9}$$

We are hence able to say that a solution $X = X(\omega)$ of (12.8) is, by definition, a solution to the genuine (random) rough differential equation

$$dX = (\sigma, \beta)(X) d\mathbf{Z}^{\mathbf{W}}(\omega) + b(X) dt \tag{12.10}$$

driven by the random rough path $\mathbf{Z} = \mathbf{Z}^{\mathbf{W}}(\omega)$. Moreover, as an immediate consequence of (12.9) and continuity of the Itô–Lyons map, we see that X is really the limit, e.g. in probability and uniformly on $[0, T]$, of classical Itô SDE solutions X^{ε}, obtained by replacing $d\mathbf{W}_t$ by the $\dot{W}^{\varepsilon}_t dt$ in (12.8).

Step 2: Given (s, x) we have a solution $(X_t : s \leq t \leq T)$ to the hybrid equation (12.8), started at $X_s = x$. In fact $(X, X') \in \mathscr{D}^{2\alpha'}_Z$ with $X' = (\sigma, \beta)(X)$. In particular, the rough integral

$$\int \gamma(X) d\mathbf{W} := \int (0, \gamma(X)) d\mathbf{Z}$$

is well-defined, as is - with regard to the Feynman–Kac formula (12.5) - the random variable

$$g(X_T) \exp\left(\int_s^T c(X_t) dt + \int_s^T \gamma(X_t) d\mathbf{W}_t\right)(\omega). \tag{12.11}$$

One can see, similar to (11.10), but now also relying on RDE growth estimates as established in Proposition 8.3), with $p = 1/\alpha'$,

$$\left| \int_s^t \gamma(X) d\mathbf{W} \right| \lesssim \|\mathbf{Z}\|_{p\text{-var};[s,t]}$$

whenever $\|\mathbf{Z}\|_{p\text{-var};[s,t]}$ is of order one. An application of the generalized Fernique Theorem 11.7, similar to the proof of Theorem 11.13 but with $\varrho = 1$ in the present context, then shows that the number of consecutive intervals on which $\mathbf{Z}$ accumulates unit p-variation has Gaussian tails; in fact, uniformly in $\varepsilon \in (0, 1]$, if $\mathbf{W}$ is replaced by W^{ε} with limit $\mathbf{W}$.) This implies that (12.11) is integrable (and uniformly integrable with respect to ε when $\mathbf{W}$ is replaced by W^{ε}). It follows that

$$u(s, x) := \mathbf{E}^{s,x}\left[g(X_T) \exp\left(\int_s^T c(X_t) dt + \int_s^T \gamma(X_t) d\mathbf{W}_t\right)\right] \tag{12.12}$$

is indeed well-defined and the pointwise limit of u^{ε} (defined in the same way, with $\mathbf{W}$ replaced by W^{ε}). By an Arzela–Ascoli argument, the limit is locally uniform. At last, the claimed continuity of the solution map follows from the same arguments, essentially by replacing W^{ϵ} by $\mathbf{W}^{\epsilon}$ everywhere in the above argument, and of course using (12.12) with $g, \mathbf{W}$ replaced by $g^{\varepsilon}, \mathbf{W}^{\varepsilon}$, respectively. □

Remark 12.3. The proof actually shows that our solution $u = u(s, x; \mathbf{W})$ to the linear RDPE (12.3) enjoys a Feynman–Kac type representation, namely (12.12), in terms of the process constructed as solution to the hybrid Itô-rough differential equation (12.8). Assume now W is a Brownian motion, independent of B, and $\mathbf{W}(\omega) = \mathbf{W}^{\text{Strat}} = (W, \mathbb{W}^{\text{Strat}}) \in \mathscr{C}^{0,\alpha}_g$ a.s. It is not difficult to show that $u = u(., ., \mathbf{W}^{\text{Strat}}(\omega))$ coincides with the Feynman–Kac SPDE solution derived by Pardoux [Par79] or Kunita [Kun82], via conditional expectations given $\sigma(\{W_{u,v} : s \leq u \leq v \leq T\}$, and so provides an identification with classical SPDE theory. In conjunction with continuity of the solution map $\mathcal{S} = \mathcal{S}[\mathbf{W}; g]$ one obtains, along the lines of Sections 9.2, SPDE limit theorems of Wong–Zakai type, Stroock–Varadhan type support statements and Freidlin–Wentzell type small noise large deviations.

Remark 12.4. It is easy to quantify the required regularity of the coefficients. The argument essentially relies on solving (12.10) as bona fide rough differential equation. It is then clear that we need to impose $\mathcal{C}_b^3$-regularity for the vector fields σ and β. The drift vector field b may be taken to be Lipschitz and $c \in \mathcal{C}_b$.

Remark 12.5. We have not given meaning to the actual equation (12.3),

$$-du = L[u]dt + \Gamma[u]d\mathbf{W} , \qquad u(T, \cdot) = g .$$

Indeed, in the absence of ellipticity or Hörmander type conditions on L, the solution may not be any more regular than $g \in \mathcal{BC}$, so that in general the action of the first order differential operator $\Gamma = (\Gamma_1, \ldots, \Gamma_d)$ on u has no pointwise meaning, let alone its rough integral against $\mathbf{W}$. On the other hand, we can (at least formally) test the equation against spatial Schwartz functions $\varphi \in \mathcal{D}$ and so arrive the following "analytically weak" formulation of (12.3),

$$\langle u_s, \varphi \rangle = \langle g, \varphi \rangle - \int_s^T \langle u_t, L^* \varphi \rangle dt - \int_s^T \langle u_t, \Gamma^* \varphi \rangle d\mathbf{W}_t. \tag{12.13}$$

In Exercise 12.22 the reader is invited to check that this formulation is indeed meaningful. In particular, the integral term $\int \langle u_t, \Gamma^* \varphi \rangle d\mathbf{W}_t$ is a bona-fide rough integral of the controlled rough path $(Y, Y') \in \mathscr{D}_W^{2\alpha}$ against $\mathbf{W}$, where

$$Y_t = \langle u_t, \Gamma^* \varphi \rangle , \qquad Y'_t = -\langle u_t, \Gamma^* \Gamma^* \varphi \rangle . \tag{12.14}$$

Assume now that W is a Brownian motion and take $\mathbf{W}(\omega) = \mathbf{W}^{\text{Strat}}$ as above. Then, thanks to Theorem 5.12, one can see that $u = u(.,., \mathbf{W}^{\text{Strat}}(\omega))$ is an analytically weak SPDE solution in the sense that

$$\langle u_s, \varphi \rangle = \langle g, \varphi \rangle - \int_s^T \langle u_t, L^* \varphi \rangle dt - \int_s^T \langle u_t, M^* \varphi \rangle \circ \overleftarrow{dW}_t,$$

where the final integral is a backward Stratonovich integral.

12.1.2 Nonlinear theory: flow transformation method

We now turn our attention to initial value problems of the form

$$du = F[u]dt + \sum_{i=1}^{d} H_i[u] \circ dW_t^i(\omega) , \qquad u(0, \cdot) = g , \tag{12.15}$$

with (possibly non-linear) differential operators,

$$F[u] = F(x, u, Du, D^2u), \quad H_i[u] = H_i(x, u, Du), \quad i = 1, \ldots, d,$$

given (in abusive notation) in terms of continuous functions F, H. As in the previous section we aim to replace $\circ dW$ by a "rough" differential $d\mathbf{W}$, for some geometric rough path $\mathbf{W} \in C_g^{0,\alpha}([0,T], \mathbf{R}^d)$, and show that an RPDE solution arises as the unique limit under approximations $(W^\varepsilon, \mathbb{W}^\varepsilon) \to \mathbf{W}$. Of course, there is little one can say at this level of generality and we have not even clarified in which sense we mean to solve (12.15) when $W \in \mathcal{C}^1$! Let us postpone this discussion and assume momentarily that F and H are sufficiently "nice" so that, for every $W \in \mathcal{C}^1$ and $g \in \mathcal{BC}$, say, there is a classical solution $u = u(t, x)$ for $t > 0$. We shall focus on three types of noise.

a) **Transport noise.** For sufficienly nice vector fields β_i on $\mathbf{R}^n$,

$$H_i[u] = \langle \beta_i(x), Du \rangle;$$

b) **Semilinear[2] noise.** For a sufficienly nice function H_i on $\mathbf{R}^n \times \mathbf{R}$,

$$H_i[u] = H_i(x, u);$$

c) **Linear noise.** With β_i as above and sufficiently nice functions γ_i on $\mathbf{R}^n$

$$H_i[u] = \Gamma_i[u] := \langle \beta_i(x), Du \rangle + \gamma_i(x)u.$$

We now develop the "calculus" for the transformations associated to each of the above cases. All proofs consist of elementary computations and are left to the reader.

Proposition 12.6 (Case a). *Assume that $\psi = \psi^W$ is a $\mathcal{C}^3$ solution flow of diffeomorphisms associated to the ODE $\dot{Y} = -\beta(Y)\dot{W}$, where $W \in \mathcal{C}^1$. (This is the case if $\beta \in \mathcal{C}_b^3$.) Then u is a classical solution to*

$$\partial_t u = F\big(x, u, Du, D^2u\big) + \langle \beta(x), Du \rangle \dot{W}$$

if and only if $v(t, x) = u(t, \psi_t(x))$ is a classical solution to

$$\partial_t v - F^\psi\big(t, x, v, Dv, D^2v\big) = 0$$

where F^ψ is determined from

$$\begin{aligned} &F^\psi(t, \psi_t(x), r, p, X) \\ &\quad \stackrel{\text{def}}{=} F\big(x, r, \big\langle p, D\psi_t^{-1}\big\rangle, \big\langle X, D\psi_t^{-1} \otimes D\psi_t^{-1}\big\rangle + \big\langle p, D^2\psi_t^{-1}\big\rangle\big) \ . \end{aligned}$$

Proposition 12.7 (Case b). *For any fixed $x \in \mathbf{R}^n$, assume that the one-dimensional ODE*

[2] The terminology here follows [LS00a].

$$\dot{\varphi} = H(x, \varphi)\dot{W} , \qquad \varphi(0; x) = r ,$$

has a unique solution flow $\varphi = \varphi^W = \varphi(t, r; x)$ *which is of class* $\mathcal{C}^2$ *as a function of both* r *and* x*. Then* u *is a classical solution to*

$$\partial_t u = F\big(x, u, Du, D^2 u\big) + H(x, u)\dot{W}$$

if and only if $v(t, x) = \varphi^{-1}(t, u(t, x), x)$*, or equivalently* $\varphi(t, v(t, x), x) = u(t, x)$ *, is a solution of*

$$\partial_t v - {}^{\varphi}F\big(t, x, r, Dv, D^2 v\big) = 0 ,$$

with

$$\begin{aligned} {}^{\varphi}F(t, x, r, p, X) \overset{\text{def}}{=} \frac{1}{\varphi'} F(t, x, \varphi, D\varphi + \varphi' p, \\ \varphi'' p \otimes p + D\varphi' \otimes p + p \otimes D\varphi' + D^2\varphi + \varphi' X\big) , \end{aligned} \tag{12.16}$$

where φ' *denotes the derivative of* $\varphi = \varphi(t, r, x)$ *with respect to* r*.*

Remark 12.8. It is worth noting that the "quadratic gradient" term $\varphi'' p \otimes p$ disappears in (12.16) whenever $\varphi'' = 0$. This happens when $H(x, u)$ is linear in u, i.e. when

$$H_i[u] = \gamma_i(x)u , \qquad i = 1, \ldots, d .$$

in which case we have

$$\varphi(t, r, x) = r \exp\left(\int_0^t \gamma(x) dW_s\right) = r \exp\left(\sum_{i=1}^{d} \gamma_i(x) W_{0,t}^i\right) . \tag{12.17}$$

Remark 12.9. Note that all dependence on $\dot{W}$ has disappeared in (12.17), and consequently (12.16). In the SPDE / filtering context this is known as *robustification*: the transformed PDE $(\partial_t - {}^{\varphi}F)v = 0$ can be solved for any $W \in \mathcal{C}([0, T], \mathbf{R}^d)$. This provides a way to solve SPDEs of the form $du = F[u]dt + \sum_{i=1}^d \gamma_i(x)u \circ dW_t$ pathwise, so that u depends continuously on W in uniform topology.

We now turn our attention to case c). The point here is that the "inner" and "outer" transformation seen above, namely

$$v(t, x) = u(t, \psi_t(x)) , \qquad v(t, x) = \varphi^{-1}(t, u(t, x), x) ,$$

respectively, can be combined to handle noise coefficients obtained by adding those from cases a) and b), i.e. noise coefficients of the type $\langle \beta_i(x), Du\rangle + H_i(x, u)$. We content ourselves with the linear case

$$H_i[u] = \langle \beta_i(x), Du\rangle + \gamma_i(x)u .$$

Proposition 12.10 (Case c). *Let* $\psi = \psi^W$ *be as in case a) and set* $\varphi(t, r, x) = r \exp\left(\int_0^t \gamma(\psi_s(x)) dW_s\right)$*. Then* u *is a (classical) solution to*

$$\partial_t u = F\big(x, u, Du, D^2u\big) + \big(\langle \beta(x), Du\rangle + \gamma(x)u\big)\dot{W}\ ,$$

if and only if $v(t,x) = u(t, \psi_t(\cdot)) \exp\Big(-\int_0^t \gamma(\psi_s(x))dW_s\Big)$ *is a (classical) solution to*

$$\partial_t v - {}^{\varphi}(F^{\psi})\big(t, x, v, Dv, D^2v\big) = 0.$$

Remark 12.11. It is worth noting that the outer transformation $F \to F^{\psi}$ preserves the class of linear operators. That is, if $F[u] = L[u]$ as given in (12.2), then F^{ψ} is again a linear operator. Because of the appearance of quadratic terms in Du, this is not true for the inner transformation $F \to {}^{\varphi}F$ *unless* $\varphi'' = 0$. Fortunately, this happens in the linear case and it follows that the transformation $F \to {}^{\varphi}(F^{\psi})$ used in case c) above does preserve the class of linear operators.

Let us reflect for a moment on what has been achieved. We started with a PDE that involves $\dot{W}$ and in all cases we managed to transform the original problem to a PDE where all dependence on $\dot{W}$ has been isolated in some auxiliary ODEs. In the stochastic context ($\circ dW$ instead of $dW = \dot{W}dt$) this is nothing but the reduction, via stochastic flows, from a *stochastic* PDE to a *random* PDE, to be solved ω-wise. In the same spirit, the rough case is now handled with the aid of flows for RDEs and their stability properties.

Given $\mathbf{W} \in \mathscr{C}_g^{0,\alpha}$, we pick an approximating sequence (W^{ε}), and transform

$$\partial_t u^{\varepsilon} = F[u^{\varepsilon}] + H[u^{\varepsilon}]\dot{W}^{\varepsilon} \tag{12.18}$$

to a PDE of the form

$$\partial_t v^{\varepsilon} = F^{\varepsilon}[v^{\varepsilon}], \tag{12.19}$$

e.g. with $F^{\varepsilon} = F^{\psi}$ and $\psi = \psi^{W^{\epsilon}}$ in case a) and accordingly in the other cases. Then

$$F^{\varepsilon}[w] = F^{\varepsilon}[t, x, w, Dw, D^2w]$$

(in abusive notation) and the function F^{ε} which appears on the right-hand side above converges (e.g. locally uniformly) as $\varepsilon \to 0$, due to stability properties of flows associated to RDEs as discussed in Section 8.9.

All one now needs is a (deterministic) PDE framework with a number of good properties, along the following "wish list".

1. All approximate problems, i.e. with $W^{\varepsilon} \in \mathcal{C}^1([0,T], \mathbf{R}^d)$

$$\partial_t u^{\varepsilon} = F[u^{\varepsilon}] + \sum_{i=1}^{d} H_i[u^{\varepsilon}]\dot{W}_t^{\varepsilon,i}\ , \qquad u^{\varepsilon}(0,\cdot) = g^{\varepsilon},$$

   should admit a unique solution, in a suitable class $\mathcal{U}$ of functions on $[0,T] \times \mathbf{R}^n$, for a suitable class of initial conditions in some space $\mathcal{G}$.
2. The change of variable calculus (Propositions 12.6 - 12.10) should remain valid, so that $u^{\varepsilon} \in \mathcal{U}$ is a solution to (12.18) if and only if its transformation $v^{\varepsilon} \in \mathcal{U}$ is a solution to (12.19).

3. There should be a good stability theory, so that $g^\varepsilon \to g^0$ in $\mathcal{G}$ and $F^\varepsilon \to F^0$ (in a suitable sense) allows to obtain convergence in $\mathcal{U}$ of solutions v^ε to (12.19) with intitial data g^ε to the (unique) solution of the limiting problem $\partial_t v^0 = F^0[v^0]$ with initial data g^0.
4. At last, the topology of $\mathcal{U}$ should be weak enough to make sure that $v^\epsilon \to v^0$ implies that the "back-transformed" u^ϵ converges in $\mathcal{U}$, with limit u^0 being v^0 back-transformed.[3]

The final point suggests to *define* a solution to

$$du = F[u]dt + H[u]d\mathbf{W}\ , \qquad u(0,\cdot) = g\ , \tag{12.20}$$

as an element in $\mathcal{U}$ which, under the correct flow transformation associated to $\mathbf{W}$ and H, solves the transformed equation $\partial_t v = F^0[v]$, $v(0,\cdot) = g$. To make this more concrete, consider the transport case a). As before, $\psi = \psi^{\mathbf{W}}$ is the flow associated to the RDE $dY = -\beta(Y)d\mathbf{W}$ and u solves the above RPDE (with $H[u] = \langle \beta(x), Du\rangle$) if, by definition, $v(t,x) := u(t,\psi_t(x))$ solves $\partial_t v = F^\psi[v]$, with $v(0,\cdot) = g$. The same logic applies to cases b) and c).

We then have the following (meta-) theorem, subject to a PDE framework with the above properties.

Theorem 12.12. *Let $\alpha \in (\frac{1}{3}, \frac{1}{2}]$. Given a geometric rough path $\mathbf{W} = (W, \mathbb{W}) \in \mathscr{C}_g^{0,\alpha}([0,T], \mathbf{R}^d)$, pick $W^\varepsilon \in \mathcal{C}^1\big([0,T], \mathbf{R}^d\big)$ so that*

$$(W^\varepsilon, \mathbb{W}^\varepsilon) := \left(W^\varepsilon, \int_0^\cdot W^\varepsilon_{0,t} \otimes dW^\varepsilon_t \right) \to \mathbf{W}$$

in α-Hölder rough path metric. Consider unique solutions $u^\epsilon \in \mathcal{U}$ to the PDEs

$$\begin{cases} \partial_t u^\epsilon = F[u^\epsilon] + H[u^\epsilon]\dot{W}^\epsilon \\ u^\epsilon(0,\cdot) = g \in \mathcal{G}. \end{cases} \tag{12.21}$$

Then there exists $u = u(t,x) \in \mathcal{U}$, not dependent on the approximating (W^ε) but only on $\mathbf{W} \in \mathscr{C}_g^{0,\alpha}([0,T], \mathbf{R}^d)$, so that

$$u^\varepsilon = \mathcal{S}[W^\varepsilon; g] \to u =: \mathcal{S}[\mathbf{W}; g]$$

as $\varepsilon \to 0$ in $\mathcal{U}$. This u is the unique solution to the RPDE (12.20) in the sense of the above definition. Moreover, the resulting solution map,

$$\mathcal{S} : \mathscr{C}_g^{0,\alpha}([0,T], \mathbf{R}^d) \times \mathcal{G} \to \mathcal{U}$$

is continuous.

It remains to identify suitable PDE frameworks, depending on the non-linearity F. When $\partial_t u = F[u]$ is a scalar conservation law, entropy solutions actually provide

[3] Given the roughness in t of our transformations, typically α-Hölder, it would not be wise to incorporate temporal $\mathcal{C}^1$-regularity in the definition of the space $\mathcal{U}$.

a suitable framework to handle additional rough noise, at least of (linear) type c), [FG14]. On the other hand, when $F = F[u]$ is a fully non-linear second order operator, say of Hamilton-Jacobi-Bellman (HJB) or Isaacs type, the natural framework is viscosity theory [CIL92, FS06] and the problem of handling additional "rough" noise, in the sense of $W \notin \mathcal{C}^1$, also with non-linear $H = H(Du)$, was first raised by Lions–Sougandis [LS98a, LS98b, LS00a, LS00b].

12.1.3 Rough viscosity solutions

Consider a real-valued function $u = u(x)$ with $x \in \mathbf{R}^m$ and assume $u \in \mathcal{C}^2$ is a classical supersolution,

$$-G\big(x, u, Du, D^2u\big) \geq 0,$$

where G is continuous and *degenerate elliptic* in the sense that $G(x, u, p, A) \leq G(x, u, p, A + B)$ whenever $B \geq 0$ in the sense of symmetric matrices. The idea is to consider a (smooth) test function φ which touches u from below at some interior point $\bar{x}$. Basic calculus implies that $Du(\bar{x}) = D\varphi(\bar{x}),\ D^2u(\bar{x}) \geq D^2\varphi(\bar{x})$ and, from degenerate ellipticity,

$$-G\big(\bar{x}, \varphi, D\varphi, D^2\varphi\big) \geq 0. \tag{12.22}$$

This motivates the definition of a *viscosity supersolution* (at the point $\bar{x}$) to $-G = 0$ as a (lower semi-)continuous function u with the property that (12.22) holds for any test function which touches u from below at $\bar{x}$. Similarly, *viscosity subsolutions* are (upper semi-)continuous functions defined via test functions touching u from above and by reversing inequality in (12.22); *viscosity solutions* are both super- and subsolutions. Observe that this definition covers (completely degenerate) first order equations as well as parabolic equations, e.g. by considering $\partial_t - F = 0$ on $[0, T] \times \mathbf{R}^n$ where F is degenerate elliptic. Let us mention a few key results of viscosity theory, with special regard to our "wish list".

1. One has existence and uniqueness results in the class of $\mathcal{BC}$ solutions to the initial value problem $(\partial_t - F)u = 0,\ u(0, \cdot) = g \in \mathcal{BUC}(\mathbf{R}^n)$, provided $F = F(t, x, u, Du, D^2u)$ is continuous, degenerate elliptic, there exists $\gamma \in \mathbf{R}$ such that, uniformly in t, x, p, X,

$$\gamma(s - r) \leq F(t, x, r, p, X) - F(t, x, s, p, X) \text{ whenever } r \leq s, \tag{12.23}$$

   and some technical conditions hold. [4] Without going into technical details, the conditions are met for $F = L$ as in (12.2) and are robust under taking inf and sup (provided the regularity of the coefficients holds uniformly). As a consequence, HJB and Isaacs type non-linearities, where F takes the form $\inf_a L_a, \inf_a sup_{a'} L_{a,a'}$, are also covered.

[4] ... the most important of which is [CIL92, (3.14)]. Additional assumptions on F are necessary, however, in particular due to the unboundedness of the domain $\mathbf{R}^n$, and these are not easily found in the literature; see [DFO14]. One can also obtain existence and uniqueness result in $\mathcal{BUC}$.

2. The change-of-variable "calculus" Propositions 12.6 - 12.10 remain valid for (continuous) viscosity solutions. Indeed, this can be checked directly from the definition of a viscosity solution.
3. In fact, the technical conditions mentioned in 1. imply a particularly strong form of uniqueness, known as *comparison*: assume u (resp. v) is a subsolution (resp. supersolution) and $u_0 \leq v_0$; then $u \leq v$ on $[0,T] \times \mathbf{R}^n$. A key feature of viscosity theory is what workers in the field simply call *stability*, a powerful incarnation of which is known as *Barles and Perthame procedure* [FS06, Section VII.3] and relies on comparison for (semi-continuous) sub- and super-solutions. In the for us relevant form, one assumes comparison for $\partial_t - F^0$ and considers viscosity solutions to $(\partial_t - F^\varepsilon)v^\varepsilon = 0$, with $v^\varepsilon(0,\cdot) = g^\varepsilon$, assuming locally uniform boundedness of v^ε and $g^\varepsilon \to g^0$ locally uniformly. Then $v^\varepsilon \to v^0$ locally uniformly where v^0 is the (unique) solution to the limiting problem $\big(\partial_t - F^0\big)v^0 = 0$, with $v^0(0,\cdot) = g^0$.

In the context of RPDEs above, again with focus on the transport case a) for the sake of argument, $F^0 = F^\psi$ where $\psi = \psi^{\mathbf{W}}$, where ψ is a flow of $\mathcal{C}^3$-diffeomorphisms (associated to the RDE $dY = -\beta(Y)d\mathbf{W}$ thereby leading to the assumption $\beta \in \mathcal{C}^5_b$). As a structural condition on F, we may simply assume "ψ-invariant comparison" meaning that comparison holds for $\partial_t - F^\psi$, for any $\mathcal{C}^3$-diffeomorphism with bounded derivatives. Checking this condition turns out to be easy. First, when $F = L$ is linear, we have $F^\psi = L^\psi$ also linear, with similar bounds on the coefficients as L due to the stringent assumptions on the derivatives of ψ. From the above discussion, and in particular from what was said in 1., it is then clear that L satisfies ψ-invariant comparison. In fact, stability of the condition in 1. under taking inf and sup, also implies that HJB and Isaacs type non-linearities satisfy ψ-invariant comparison.

It is now possible to implement the arguments of the previous (meta-) Theorem 12.12 in the viscosity framework [CFO11]. We tacitly assume that all approximate problems of the form (12.24) below have a viscosity solution, for all $W^\varepsilon \in C^1$ and $g \in \mathcal{BUC}$, but see Remark 12.14.

Theorem 12.13. *Let $\alpha \in (\frac{1}{3}, \frac{1}{2}]$. Given a geometric rough path $\mathbf{W} = (W, \mathbb{W}) \in \mathscr{C}^{0,\alpha}_g([0,T], \mathbf{R}^d)$, pick $W^\varepsilon \in \mathcal{C}^1\big([0,T], \mathbf{R}^d\big)$ so that $(W^\varepsilon, \mathbb{W}^\varepsilon) \to \mathbf{W}$ in α-Hölder rough path metric. Consider unique $\mathcal{BC}$ viscosity solutions u^ϵ to*

$$\begin{cases} \partial_t u^\epsilon = F[u^\epsilon] + \langle \beta(x), Du \rangle \dot{W}^\epsilon \\ u^\epsilon(0,\cdot) = g \in \mathcal{BUC}(\mathbf{R}^n) \end{cases} \tag{12.24}$$

where F satisfies ψ-invariant comparison. Then there exists $u = u(t,x) \in \mathcal{BC}$, not dependent on the approximating (W^ε) but only on $\mathbf{W} \in \mathscr{C}^{0,\alpha}_g([0,T], \mathbf{R}^d)$, so that

$$u^\varepsilon = \mathcal{S}[W^\varepsilon; g] \to u =: \mathcal{S}[\mathbf{W}; g]$$

as $\varepsilon \to 0$ in local uniform sense. This u is the unique solution to the RPDE (12.20) with transport noise $H[u] = \langle \beta(x), Du \rangle$ in the sense of the definition given previous to Theorem 12.12. Moreover, we have continuity of the solution map,

$$\mathcal{S} : \mathscr{C}_g^{0,\alpha}([0,T],\mathbf{R}^d) \times \mathcal{BUC}(\mathbf{R}^n) \to \mathcal{BC}([0,T] \times \mathbf{R}^n) \ .$$

Remark 12.14. In the above theorem, existence of RPDE solutions actually relies on existence of approximate solutions u^ε, which one of course expects from standard viscosity theory. Mild structural conditions on F, satisfied by HJB and Isaacs examples, which imply this existence are reviewed in [DFO14]. One can also establish a modulus of continuity for RPDE solutions, so that $u \in \mathcal{BUC}$ after all.

Remark 12.15. The RPDE solution to $du = F[u]dt + \langle \beta(x), Du \rangle d\mathbf{W}$ as constructed above, when $F = \inf_a L_a$ is of HJB form, arises in pathwise stochastic control [LS98b, BM07, DFG13].

Unfortunately, in the "semi-linear" noise case b), it turns out the structural assumptions one has to impose on F in order to have the necessary comparison for $\partial_t - F^0 = 0$ is rather restrictive, although semilinear situations are certainly covered. Even in this case, due to the appearance of a quadratic non-linearity in Du, the argument is involved and requires a careful analysis on consecutive small time intervals, rather than $[0,T]$; see [LS00a, DF12]. A non-linear Feynman–Kac representation, in terms of *rough backward stochastic differential equations* is given in [DF12].

At last, we return to the fully linear case of Section 12.1.1. That is, we consider the (linear noise) case c) with linear $F = L$. With some care [FO14], the double transformation leading to the transformed equation $\partial_t - {}^{\varphi}(F^{\psi}) = 0$ can be implemented with the aid of coupled flows of rough differential equations. We can then recover Theorem 12.2, but with somewhat different needs concerning the regularity of the coefficients. (For instance, in the aforementioned theorem we really needed $\sigma, \beta \in \mathcal{C}_b^3$ whereas now, using flow decomposition, we need $\beta \in \mathcal{C}_b^5$ but only $\sigma \in \mathcal{C}_b^1$.

Remark 12.16. By either approach, case c) with linear $F = L$ or Theorem 12.2, we obtain a robust view on classes SPDEs which contain the Zakai equation from filtering theory, provided the initial law admits a $\mathcal{BUC}$-density. Robustness is an important issue in filtering theory, see also Exercise 12.24.

12.2 Stochastic heat equation as a rough path

Nonlinear stochastic partial differential equations driven by very singular noise, say space-time white noise, may suffer from the fact that their nonlinearities are ill-posed. For instance, even in space dimension one, there is no obvious way of giving "weak" meaning to Burgers-like stochastic PDEs of the type

$$\partial_t u^i = \partial_x^2 u^i + f(u) + \sum_{j=1}^n g_j^i(u) \partial_x u^j + \xi^i \ , \qquad i = 1, \dots, n \ , \tag{12.25}$$

where $\xi = (\xi^i)$ denotes space-time white noise (strictly speaking, n independent copies of scalar space-time white noise). Recall that, at least formally, space-time white noise is a Gaussian generalized stochastic process such that

$$\mathbf{E}\xi^i(t,x)\xi^j(s,y) = \delta_{ij}\delta(t-s)\delta(x-y) \; .$$

As a consequence of the lack of regularity of ξ, it turns out that the solution to the stochastic heat equation (i.e. the case $f = g = 0$ in (12.25) above) is only α-Hölder continuous in the spatial variable x for any $\alpha < 1/2$. In other words, one would not expect any solution u to (12.25) to exhibit spatial regularity better than that of a Brownian motion.

As a consequence, even when aiming for a weak solution theory, it is not clear how to define the integral of a spatial test function φ against the nonlinearity. Indeed, this would require us to make sense of expressions of the type

$$\int \varphi(x) g_j^i(u)\partial_x u^j(t,x)\, dx \; ,$$

for fixed t. When g happens to be a gradient, such an integral can be defined by postulating that the chain rule holds and integrating by parts. For a general g, as arising in applications from path sampling [HSV07], this approach fails. This suggests to seek an understanding of $u(t,\cdot)$ as a spatial rough path. Indeed, this would solve the problem just explained by allowing us to define the nonlinearity in a weak sense as

$$\int \varphi(x) g_j^i(u)\, d\mathbf{u}^{\mathbf{j}}(t,x) \; ,$$

where $\mathbf{u}$ is the rough path associated to u.

In the particular case of (12.25), it is actually sufficient to be able to associate a rough path to the solution ψ to the stochastic heat equation

$$\partial_t \psi = \partial_x^2 \psi + \xi \; .$$

Indeed, writing $u = \psi + v$ and proceeding formally for the moment, we then see that v should solve

$$\partial_t v^i = \partial_x^2 v^i + f(v+\psi) + \sum_{j=1}^n g_j^i(v+\psi)\big(\partial_x \psi^j + \partial_x v^j\big) \; .$$

If we were able to make sense of the term appearing in the right hand side of this equation, one would expect it to have the same regularity as $\partial_x \psi$ so that, since $\psi(t,\cdot)$ turns out to belong to $\mathcal{C}^\alpha$ for every $\alpha < 1/2$, one would expect $v(t,\cdot)$ to be of regularity $\mathcal{C}^{\alpha+1}$ for every $\alpha < 1/2$. In particular, we would not expect the term involving $\partial_x v^j$ to cause any trouble, so that it only remains to provide a meaning for the term $g_j^i(v+\psi)\partial_x \psi^j$. If we know that $v \in \mathcal{C}^1$ and we have an interpretation of $\psi(t,\cdot)$ as a rough path $\boldsymbol{\psi}$ (in space), then this can be interpreted as the distribution whose action, when tested against a test function φ, is given by

$$\int \varphi(x) g^i_j(\psi + v))\, d\boldsymbol{\psi}^{\mathbf{j}}(t,x)\ .$$

This reasoning can actually be made precise, see the original article [Hai11b]. In this section we limit ourselves to providing the construction of $\boldsymbol{\psi}$ and giving some of its basic properties.

12.2.1 The linear stochastic heat equation

We now study the model problem in this context - the construction of a spatial rough path associated, in essence, to the above SPDE in the case $f = g = 0$. More precisely, we are considering stationary (in time) solution to the stochastic heat equation[5],

$$d\psi_t = -A\psi_t dt + \sigma dW_t, \tag{12.26}$$

where, for fixed $\lambda > 0$

$$Au = -\partial_x^2 u + \lambda u;$$

and W is a cylindrical Wiener process over $L^2(\mathbb{T})$, the L^2-space over the one-dimensional torus. Let $(e_k : k \in \mathbf{Z})$ denote the standard Fourier-basis of $L^2(\mathbb{T})$

$$e_k(x) = \begin{cases} \frac{1}{\sqrt{\pi}} \sin(kx) & \text{for } k > 0 \\ \frac{1}{\sqrt{2\pi}} & \text{for } k = 0 \\ \frac{1}{\sqrt{\pi}} \cos(kx) & \text{for } k < 0 \end{cases}$$

which diagonalises the operator A in the sense that

$$Ae_k = \mu_k e_k\ , \qquad mu_k = k^2 + \lambda\ , \qquad k \in \mathbf{Z}\ .$$

Thanks to the fact that we chose $\lambda > 0$, the stochastic heat equation (12.26) has indeed a stationary solution which, by taking Fourier transforms, may be decomposed as $\psi(x,t;\omega) = \sum_k Y^k_t(\omega) e_k(x)$. The components Y^k_t are then a family of independent stationary one-dimensional Ornstein-Uhlenbeck processes given by

$$dY^k_t = -\mu_k Y^k_t dt + \sigma dB^k_t\ ,$$

where $(B^k : k \in \mathbf{Z})$ is a family of i.i.d. standard Brownian motions. An explicit calculation yields

$$\mathbf{E}\big(Y^k_s Y^k_t\big) = \frac{\sigma^2}{2\mu_k} \exp\left(-\mu_k |t-s|\right)\ ,$$

so that in particular

---

[5] With $\lambda = 0$, the 0^{th} mode of ψ behaves like a Brownian motion and ψ cannot be stationary in time, unless one identifies functions that only differ by a constant.

$$\mathbf{E}\big(Y_t^k\big)^2 = \frac{\sigma^2}{2\mu_k} \ ,$$

for any fixed time t.

Lemma 12.17. *For each fixed t, the spatial covariance of ψ is given by*

$$\mathbf{E}(\psi(x,t)\psi(y,t)) = K(|x-y|)$$

where K is given by

$$K(u) := \frac{1}{4\pi}\sigma^2 \sum_{k\in\mathbf{Z}} \frac{\cos(ku)}{\mu_k} = \frac{\sigma^2}{4\sqrt{\lambda}\sinh\big(\sqrt{\lambda}\pi\big)} \cosh\Big(\sqrt{\lambda}(u-\pi)\Big).$$

Here, the second equality holds for u restricted to $[0,2\pi]$. In fact, the cosine series is the periodic continuation of the r.h.s. restricted to $[0,2\pi]$.

Proof. From the basic identity $\cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$,

$$e_{-k}(x)e_{-k}(y) + e_k(x)e_k(y) = \frac{1}{\pi}\cos(k(x-y)), \ k\in\mathbf{Z} \ .$$

Inserting the respective expansion in $R(x,y) := \mathbf{E}(\psi(x,t)\psi(y,t))$, and using the independence of the $\big(Y^k : k\in\mathbf{Z}\big)$, gives

$$\begin{aligned} R(x,y) &= \sum_{k\in\mathbf{Z}} e_k(x)e_k(y)\mathbf{E}\big(Y_t^k\big)^2 = \frac{1}{2\pi}\mathbf{E}\big(Y_t^0\big)^2 + \frac{1}{\pi}\sum_{k=1}^{\infty}\cos(k(x-y))\mathbf{E}\big(Y_t^k\big)^2 \\ &= \frac{\sigma^2}{4\pi}\sum_{k\in\mathbf{Z}}\frac{\cos(k(x-y))}{\lambda+k^2} \ , \end{aligned}$$

and then $R(x,y) = K(|x-y|)$ where

$$K(x) = \frac{\sigma^2}{4\pi}\sum_{k\in\mathbf{Z}}\frac{\cos(kx)}{\lambda+k^2} \ .$$

At last, expand the (even) function $\cosh\big(\sqrt{\lambda}(|\cdot|-\pi)\big)$ in its (cosine) Fourier-series to get the claimed equality. $\square$

Proposition 12.18. *Fix $t\geq 0$. Then $\psi_t(x;\omega) = \psi(t,x;\omega)$, indexed by $x\in[0,2\pi]$, is a centred Gaussian process with covariance of finite 1-variation. More precisely,*

$$\big\|R_{\psi(t,\cdot)}\big\|_{1;[x,y]^2} \leq 2\pi\|K\|_{\mathcal{C}^2;[0,2\pi]}|x-y| \ ,$$

and so (cf. Theorem 10.4), for each fixed $t\geq 0$, the $\mathbf{R}^d$-valued process

$$[0,2\pi] \ni x \mapsto \big(\psi_t^1(x),\ldots,\psi_t^d(x)\big),$$

consisting of d i.i.d. copies of ψ_t, lifts canonically to a Gaussian rough path $\boldsymbol{\psi}_t(\cdot) \in \mathscr{C}_g^{0,\alpha}\big([0,2\pi],\mathbf{R}^d\big)$.

Proof. This follows immediately from Exercise 10.16. □

Remark 12.19. There are ad-hoc ways to construct a (spatial) rough path lift associated to the stochastic heat-equation, for instance be writing $\psi(t,\cdot)$ as Brownian bridge plus a random smooth function. In this way, however, one ignores the large body of results available for general Gaussian rough paths: for instance, rough path convergence of hyper-viscosity or Galerkin approximation, extensions to fractional stochastic heat equations, concentration of measure can all be deduced from general principles.

We now show that solutions to the stochastic heat equation induces a continuous stochastic evolution in rough path space.

Theorem 12.20. *There exists a continuous modification of the map $t \mapsto \boldsymbol{\psi}_t$ with values in $\mathscr{C}_g^{\alpha}\big([0,2\pi],\mathbf{R}^d\big)$.*

Proof. Fix s and t. The proof then proceeds in two steps. First, we will verify the assumptions of Corollary 10.6, namely we will show that

$$|\varrho_\alpha(\psi_s,\psi_t)|_{L^q} \le C \sup_{x,y\in[0,2\pi]} \Big[\mathbf{E}(|\psi_s(x,y)-\psi_t(x,y)|^2)\Big]^\theta ,$$

for some constant C that is independent of s and t. In the second step, we will show that (here we may assume $d=1$), with $\psi_s(x,y) := \psi_s(y)-\psi_s(x)$, one has the bound

$$\sup_{x,y\in[0,2\pi]} \mathbf{E}\Big[|\psi_s(x,y)-\psi_t(x,y)|^2\Big] = \mathrm{O}\big(|t-s|^{1/2}\big) .$$

The existence of a continuous (and even Hölder) modification is then a consequence of the classical Kolmogorov criterion.

For the first step, we write $X = \big(\psi_s^1(\cdot),\ldots,\psi_s^d(\cdot)\big)$ and $Y = \big(\psi_t^1(\cdot),\ldots,\psi_t^d(\cdot)\big)$. Note that one has independence of $\big(X^i,Y^i\big)$ with $\big(X^j,Y^j\big)$ for $i\neq j$. We have to verify finite 1-variation (in the 2D sense) of the covariance of (X,Y). In view of Proposition 12.18, it remains to establish finite 1-variation of

$$\begin{aligned}(x,y) \mapsto R_{(X^1,Y^1)}(x,y) = \mathbf{E}\big[\psi_s^1(x)\psi_t^1(y)\big] &= \sum_{k\in\mathbf{Z}} e_k(x)e_k(y)\mathbf{E}\big(Y_s^kY_t^k\big)\\ &= \frac{\sigma^2}{4\pi}\sum_{k\in\mathbf{Z}} \frac{\cos\big(k(x-y)\big)}{\lambda+k^2} e^{-\left(\lambda+k^2\right)|t-s|\cdot} =: R_\tau(x,y).\end{aligned}$$

For every $\tau>0$, exponential decay of the Fourier-modes implies smoothness of R_τ. We claim

$$\|R_\tau\|_{1\text{-var};[u,v]^2} \le C|v-u| < \infty,$$

uniformly in $\tau\in(0,1]$ and u,v. To see this, write

$$
\begin{aligned}
\|R_\tau\|_{1\text{-var};[u,v]^2} &= \int_u^v \int_u^v |\partial_{xy} R_\tau| dx\, dy \\
&\sim \int_u^v \int_u^v \left| \sum k^2 \frac{e^{ik(x-y)}}{\lambda + k^2} e^{-(\lambda+k^2)\tau} \right| dx\, dy \\
&\sim \int_u^v \int_u^v \left| \sum e^{ik(x-y)} e^{-k^2\tau} \right| dx\, dy \\
&= \int_u^v \int_u^v p_\tau(x-y) dy\, dx \le |v-u| \,,
\end{aligned}
$$

where we used the trivial estimate $\int_u^v p_\tau(x-y)dy \le \int_0^{2\pi} p_\tau(x-y)dy = 1$. In this expression, p denotes the (positive) transition kernel of the heat semigroup on the torus. The step above, between second and third line, where we effectively set $\lambda = 0$ is harmless. The factor $e^{-\lambda\tau}$ may simply be taken out, and

$$
\left| \sum_k \left(1 - \frac{k^2}{\lambda + k^2}\right) e^{ik(x-y)} e^{-k^2\tau} \right| \le \sum_k \left| 1 - \frac{k^2}{\lambda + k^2} \right| = \sum_k \frac{\lambda}{\lambda + k^2} < \infty \,.
$$

After integrating over $[u,v]^2$, we see that the error made above is actually of order $\mathrm{O}\big(|v-u|^2\big)$. This is more than enough to conclude that

$$
\big\| R_{(X^1,Y^1)} \big\|_{1\text{-var};[u,v]^2} \le C|v-u| < \infty \,,
$$

uniformly in $\tau \in (0,1]$ and u, v.

We now turn to the second step of our proof. We claim that $\mathbf{E}|\psi_s^1(x,y) - \psi_t^1(x,y)|^2 = \mathrm{O}\big(|t-s|^{1/2}\big)$, uniformly in $x, y \in [0,2\pi]$. Since

$$
\big|\psi_s^1(x,y) - \psi_t^1(x,y)\big| \le \big|\psi_s^1(x) - \psi_t^1(x)\big| + \big|\psi_s^1(y) - \psi_t^1(y)\big| \,,
$$

the question reduces to a similar bound on $\mathbf{E}|\psi_s^1(x) - \psi_t^1(x)|^2$, uniform in $x \in [0,2\pi]$. This quantity is equal to

$$
\begin{aligned}
\mathbf{E}\big[\psi_s^1(x)\psi_s^1(x)\big] - 2\mathbf{E}\big[\psi_s^1(x)\psi_t^1(x)\big] + \mathbf{E}\big[\psi_t^1(x)\psi_t^1(x)\big] \\
&= \frac{\sigma^2}{4\pi} \sum_{k\in\mathbf{Z}} \frac{2\big(1 - e^{-(\lambda+k^2)|t-s|}\big)}{\lambda + k^2} . \\
&\le \frac{\sigma^2}{4\pi} \sum_{|k|<N} 2|t-s| + 2\frac{\sigma^2}{4\pi} \sum_{k\ge N} \frac{2\big(1 - e^{-(\lambda+k^2)|t-s|}\big)}{\lambda + k^2} \,,
\end{aligned}
$$

where we used that $1 - e^{-cx} \le cx$ for $c, x > 0$ in the first sum. We then take $N \sim |t-s|^{-1/2}$, so that the first sum is of order $\mathrm{O}\big(|t-s|^{1/2}\big)$. For the second sum, we use the trivial bound $1 - e^{-(\lambda+k^2)|t-s|} \le 1$. It then suffices to note that

$$\sum_{k \geq N} \frac{1}{\lambda + k^2} \leq \sum_{k \geq N} \frac{1}{k^2} = \mathrm{O}(1/N) = \mathrm{O}\big(|t-s|^{1/2}\big) ,$$

which completes the proof. □

Remark 12.21. The final estimate in the above proof, namely

$$\mathbf{E}\big|\psi_s^1(x) - \psi_t^1(x)\big|^2 = \mathrm{O}\big(|t-s|^{1/2}\big) ,$$

also implies "almost $\frac{1}{4}$-Hölder" temporal regularity of the stochastic heat equation.

12.3 Exercises

Exercise 12.22 (From [DFS14]).

a) Assume $W \in \mathcal{C}^1$. Show that the Feynman–Kac solution (equivalently: viscosity) solution to (12.4) is an analytically weak solution in the sense of (12.13) with $d\mathbf{W}$ replaced by $\dot{W} dt$.
b) Assume now $\mathbf{W} = (W, \mathbb{W}) \in \mathscr{C}_g^{0,\alpha}$. Show that $(Y, Y') \in \mathscr{D}_W^{2\alpha}$.
c) Show that the Feynmann–Kac solution constructed in Theorem 12.2 is an analytically weak solution in the sense of (12.13).

Exercise 12.23 (From [CDFO13]). A crucial rôle in the proof of Theorem 12.2 was played by a *hybrid Itô-rough differential equation* of the form

$$dX_t = \sigma(X_t) dB + \beta(X_t) d\mathbf{W}, \tag{12.27}$$

ultimately solved as (random) rough differential equation, subject to $\sigma, \beta \in \mathcal{C}_b^3$. Give an alternative construction to the hybrid equation based on flow decomposition. That is, use the flow associated to the RDE $dY = \beta(Y) d\mathbf{W}$ and transform (12.27) into a bona-fide Itô differential equations. **Hint:** When $\mathbf{W}$ is replaced by a $\mathcal{C}^1$ path W^ε this a straight-forward computation. Use stability of RDE flows, combined with stability results for Itô SDEs to conclude. Specify the regularity requirements on σ, β.

Exercise 12.24 (Robust filtering, [CDFO13]). Consider a pair of processes (X, Y) with dynamics

$$dX_t = V_0(X_t, Y_t) dt + \sum_k Z_k(X_t, Y_t) dW_t^k + \sum_j V_j(X_t, Y_t) dB_t^j, \tag{12.28}$$

$$dY_t = h(X_t, Y_t) dt + dW_t, \tag{12.29}$$

with $X_0 \in L^\infty$ and $Y_0 = 0$. For simplicity, assume coefficients $V_0, V_1, \ldots, V_{d_B} : \mathbf{R}^{d_X + d_Y} \to \mathbf{R}^{d_X}$, $Z_1, \ldots, Z_{d_Y} : \mathbf{R}^{d_X + d_Y} \to \mathbf{R}^{d_X}$ and $h = (h^1, \ldots, h^{d_Y}) : \mathbf{R}^{d_X + d_Y} \to \mathbf{R}^{d_Y}$ to be bounded with bounded derivatives of all orders; W and B are independent Brownian motions of the correct dimension. We now interpret X

as a signal and Y as noisy and incomplete observation. The *filtering problem* consists in computing the conditional distribution of the unobserved component X, given the observation Y. Equivalently, one is interested in computing

$$\pi_t(g) = \mathbf{E}[g(X_t, Y_t)|\mathcal{Y}_t] \ ,$$

where $\mathcal{Y}_t$ is the observation filtration and g is a suitably chosen test function. Measure theory tells us that there exists a Borel-measurable map $\theta_t^g : \mathcal{C}([0,t], \mathbf{R}^{d_Y}) \to \mathbf{R}$, such that a.s. $\pi_t(g) = \theta_t^g(Y)$ where we consider $Y = Y(\omega)$ as a $\mathcal{C}([0,t], \mathbf{R}^{d_Y})$-valued random variable. Note that θ_t^g is not uniquely determined (after all, modifications on null sets are always possible). On the other hand, there is obvious interest to have a *robust filter*, in the sense of having a *continuous* version of θ_t^g, so that close observations lead to nearby conclusions about the signal.

a) Give an example to show that, in general, θ_t^g does not admit a continuous version.
b) Let $\alpha \in (1/2, 1/3)$. Show that there exists a continuous map on rough path space

$$\Theta_t^g : \mathscr{C}_g^{0,\alpha}([0,t], \mathbf{R}^{d_Y}) \to \mathbf{R} \ ,$$

such that a.s.

$$\pi_t(g) = \Theta_t^g(\mathbf{Y}) \ , \tag{12.30}$$

where $\mathbf{Y}$ is the random geometric rough path obtained from Y by iterated Stratonovich integration.

Hint: You may use the “Kallianpur–Striebel formula”, a standard result in filtering theory which asserts that

$$\pi_t(g) = \frac{p_t(g)}{p_t(1)}, \quad p_t(g) := \mathbf{E}_0[g(X_t, Y_t)v_t|\mathcal{Y}_t]$$

where

$$\left.\frac{\mathrm{d}\mathbf{P}_0}{\mathrm{d}\mathbf{P}}\right|_{\mathcal{F}_t} = \exp\left(-\sum_i \int_0^t h^i(X_s, Y_s)dW_s^i - \frac{1}{2}\int_0^t ||h(X_s, Y_s)||^2 ds\right)$$

and $v = \{v_t, t > 0\}$ is defined as the right-hand side above with $-W$ replaced by Y.

Exercise 12.25. Show almost sure “$(\frac{1}{4} - \varepsilon)$-Hölder” temporal regularity of $\psi = \psi_t(x;\omega)$, solution to the stochastic heat equation. Show that, for fixed x, $\psi_t(x;\omega)$ is not a semi-martingale.

Exercise 12.26 (Spatial Itô–Stratonovich correction; from [HM12]). Writing $\mathbb{T}$ for $[0, 2\pi]$ with periodic boundary, let us say that

$$u = u(t, x; \omega) : [0, T] \times \mathbb{T} \times \Omega \to \mathbf{R}$$

is a (analytically) weak solution to

$$\partial_t u = \partial_{xx} u - u + \partial_x \left(\frac{1}{2} u^2 \right) + \xi \, , \tag{$\star$}$$

if and only if $u = v + \psi$ where ψ is the stationary solution to $\partial_t \psi = \partial_{xx}\psi - \psi + \xi$ and, for all test functions $\varphi \in \mathcal{C}^\infty(\mathbb{T})$,

$$\partial_t \langle v, \varphi \rangle = \langle v, \partial_{xx}\varphi \rangle - \langle v, \varphi \rangle - \left\langle \frac{1}{2} u^2, \partial_x \varphi \right\rangle .$$

a) Replace $\partial_x(\frac{1}{2}u^2)$ in ($\star$) by a (spatially right) finite-difference approximation,

$$\frac{1}{2}\frac{u(.+\varepsilon)^2 - u^2}{\varepsilon};$$

write u^ε for a solution to the resulting equation. Assume $u^\varepsilon \to u$ locally uniformly in probability. Show that u is a solution to ($\star$).

b) At least formally, $\partial_x\left(\frac{1}{2}u^2\right) = u\partial_x u$ in ($\star$), which suggests an alternative finite difference approximation, namely,

$$u\frac{(u(.+\varepsilon) - u)}{\varepsilon};$$

Assume $v^\varepsilon = u^\varepsilon - \psi \to v := u - \psi$ and its first (spatial) derivatives converge locally uniformly in probability. Show that u is an analytically weak solution to the perturbed equation

$$\partial_t u = \partial_{xx} u + \partial_x \left(\frac{1}{2} u^2 \right) + C + \xi$$

with $C \neq 0$. Determine the value of C. **Hint:** Use Exercise 10.20.

Solution 12.27. a) By switching to suitable subsequences, we may assume $u^\varepsilon \to u$ locally uniformly with probability one. Write $D_{\varepsilon,l}, D_{\varepsilon,r}$ for a discrete (left, right) finite difference approximation. Note

$$\left\langle D_{\varepsilon,r}\left(\frac{1}{2}u^2\right), \varphi \right\rangle = -\left\langle \frac{1}{2}u^2, D_{\varepsilon,l}\varphi \right\rangle \to -\left\langle \frac{1}{2}u^2, \partial_x\varphi \right\rangle.$$

Given that $v^\varepsilon = u^\varepsilon - \psi \to v := u - \psi$ locally uniform it then suffices to pass to the limit in the (integral formulation) of

$$\partial_t \langle v^\varepsilon, \varphi \rangle = \langle v^\varepsilon, \partial_{xx}\varphi \rangle - \langle v^\varepsilon, \varphi \rangle + \left\langle \frac{1}{2}u^2, D_{\varepsilon,l}\varphi \right\rangle.$$

b) We note

$$D_{\varepsilon,r}\left(\frac{1}{2}u^2\right) = \frac{1}{2}\frac{u(.+\varepsilon)^2 - u^2}{\varepsilon} = \frac{(u(.+\varepsilon) + u)}{2}\frac{(u(.+\varepsilon) - u)}{\varepsilon}$$

$$= u\frac{(u(.+\varepsilon)-u)}{\varepsilon} + \frac{1}{2\varepsilon}(u(.+\varepsilon)-u)^2 \ .$$

It follows that

$$\begin{aligned}
\partial_t\langle v^\varepsilon,\varphi\rangle &= \langle v^\varepsilon,\partial_{xx}\varphi\rangle - \langle v^\varepsilon,\varphi\rangle + \left\langle u^\varepsilon\frac{(u^\varepsilon(.+\varepsilon)-u^\varepsilon)}{\varepsilon},\varphi\right\rangle. \\
&= \langle v^\varepsilon,\partial_{xx}\varphi\rangle - \langle v^\varepsilon,\varphi\rangle \\
&\quad - \left\langle \frac{1}{2}(u^\varepsilon)^2, D_{\varepsilon,l}\varphi\right\rangle - \left\langle \frac{1}{2\varepsilon}(u^\varepsilon(.+\varepsilon)-u^\varepsilon)^2,\varphi\right\rangle \ .
\end{aligned}$$

In order to pass to the $\varepsilon \to 0$ limit, we must understand the final "quadratic variation" term. By assumption v^ε are of class $\mathcal{C}^1$, uniformly in ε. Hence

$$\begin{aligned}
[u^\varepsilon(.+\varepsilon)-u^\varepsilon] &= \psi(.+\varepsilon)-\psi+v^\varepsilon(.+\varepsilon)-v^\varepsilon \\
&= \psi(.+\varepsilon)-\psi+\mathrm{O}(\varepsilon)
\end{aligned}$$

and so, with $\operatorname{osc}(\psi;\varepsilon)\mathrm{O}(1)+\mathrm{O}(\varepsilon)=\mathrm{o}(1)$ as $\varepsilon\to 0$,

$$\frac{1}{2\varepsilon}(u^\varepsilon(.+\varepsilon)-u^\varepsilon)^2 = \frac{1}{2\varepsilon}(\psi(.+\varepsilon)-\psi)^2 + \mathrm{o}(1)$$

we have

$$\left\langle \frac{1}{2\varepsilon}(u^\varepsilon(.+\varepsilon)-u^\varepsilon)^2,\varphi\right\rangle = \left\langle \frac{1}{2\varepsilon}(\psi(.+\varepsilon)-\psi)^2,\varphi\right\rangle + \mathrm{o}(1) \ .$$

From Lemma 12.17 we know that

$$\mathbf{E}[\psi^2_{x,x+\varepsilon}] = 2(K(0)-K(\varepsilon)) = -2K'(0)\varepsilon + \mathrm{o}(\varepsilon) = C\varepsilon + \mathrm{o}(\varepsilon) \ .$$

Since $K(u) = \frac{\cosh(u-\pi)}{4\sinh(\pi)}$, we have $C = -2K'(0) = \frac{1}{2}$, and it follows from Exercise 10.20 that

$$\begin{aligned}
\left\langle \frac{1}{2\varepsilon}(\psi(.+\varepsilon)-\psi)^2,\varphi\right\rangle &= \frac{1}{2}\int \varphi(x)\frac{\psi^{\,2}_{x,x+\varepsilon}}{\varepsilon}dx \\
&\to \frac{1}{2}\int \varphi(x)C\,dx = \left\langle \frac{1}{4},\varphi\right\rangle ,
\end{aligned}$$

where the convergence takes place in probability. It follows that u is a solution (in the above analytically weak sense) of

$$\partial_t u = \partial_{xx}u - u + \partial_x\left(\frac{1}{2}u^2\right) + \frac{1}{4} + \xi.$$

12.4 Comments

Section 12.1: Linear stochastic partial differential equations go back at least to Krylov–Rozovskii [KR77]. A Feynmann-Kac representation appears in Pardoux [Par79] and Kunita [Kun82]. Kunita also has flow decompositions of SPDE solutions. Caruana–Friz [CF09] implement this in the rough path setting in a framework of classical PDE solutions. In the context of Crandall–Ishii-Lions viscosity setting, non-linear SPDE problems ("stochastic viscosity solution") where introduced by Lions–Souganidis [LS98a, LS98b, LS00a, LS00b]. Caruana, Friz and Oberhauser [CFO11] introduce "rough viscosity solutions", for classes of non-linear SPDEs with transport noise. Extensions to different noise situations are due to Diehl–Friz, [DF12] and then [FO14]. Non-linear noise, quadratic in Du is considered by Friz–Gassiat [FG13]. See [LPS13, FG14] for similar investigations in the context of stochastic conservation laws. A non-linear Feynman–Kac representation (with relations to "rough BSDEs") is given in [DF12]. In a filtering context, a (rough path) robustified Kalianpur–Striebel formula (cf. Exercise 12.24) was given by Crisan, Diehl, Friz and Oberhauser [CDFO13], which is also the first source of hybrid differential equations. The construction of hybrid stochastic/rough differential equations as encountered in the proof of Theorem 12.2 is taken from [DOR13]; see also [DFS14]. At last, we refer to Gubinelli–Tindel, Deya et al. and Teichmann [GT10, DGT12, Tei11] for some other rough path approaches to SPDEs.

Section 12.2: The construction of a spatial rough path associated to the stochastic heat equation is due to Hairer [Hai11b] and allows to deal with otherwise ill-posed SPDEs of stochastic Burgers type, see also Hairer–Weber [HW13] and Friz, Gess, Gulisashvili, Riedel [FGGR13] for various extensions (including multiplicative noise, and fractional Laplacian / non-periodic boundary respectively). This construction is also important in giving meaning to the KPZ equation, Hairer [Hai13] and Chapter 15. Exercise 12.26, in the spirit of Föllmer – rather than rough path – integration, is taken from Hairer–Maas [HM12]. Similar results are available for rough SPDEs of type (12.25), see Hairer, Maas and Weber [HMW14], but this is beyond the scope of these notes.

Chapter 13
Introduction to regularity structures

Abstract We give a short introduction to the main concepts of the general theory of regularity structures. This theory unifies the theory of (controlled) rough paths with the usual theory of Taylor expansions and allows to treat situations where the underlying space is multidimensional.

13.1 Introduction

While a full exposition of the theory of regularity structures is well beyond the scope of this book, we aim to give a concise overview to most of its concepts and to show how the theory of controlled rough paths fits into it. In most cases, we will only state results in a rather informal way and give some ideas as to how the proofs work, focusing on conceptual rather than technical issues. The only exception is the "reconstruction theorem", Theorem 13.12 below, which is one of the linchpins of the whole theory. Since its proof (or rather a slightly simplified version of it) is relatively concise, we provide a fully self-contained version. For precise statements and complete proofs of most of the results exposed here, we refer to the original article [Hai14c]. See also the review articles [Hai14a, Hai14b] for shorter expositions that complement the one given here.

It should be clear by now that a controlled rough path $(Y, Y') \in \mathscr{D}_W^{2\alpha}$ bears a strong resemblance to a differentiable function, with the Gubinelli derivative Y' describing the coefficient in front of a "first-order Taylor expansion" of the type

$$Y_t = Y_s + Y'_s W_{s,t} + \mathrm{O}(|t-s|^{2\alpha}) \,. \tag{13.1}$$

Compare this to the fact that a function $f \colon \mathbf{R} \to \mathbf{R}$ is of class $\mathcal{C}^\gamma$ with $\gamma \in (k, k+1)$ if for every $s \in \mathbf{R}$ there exist coefficients $f_s^{(1)}, \ldots, f_s^{(k)}$ such that

$$f_t = f_s + \sum_{\ell=1}^{k} f_s^{(\ell)} (t-s)^\ell + \mathrm{O}(|t-s|^\gamma) \,. \tag{13.2}$$

Of course, $f_s^{(\ell)}$ is nothing but the ℓth derivative of f at the point s, divided by $\ell!$. In this sense, one should really think of a controlled rough path $(Y, Y') \in \mathscr{D}_W^{2\alpha}$ as a 2α-Hölder continuous function, but with respect to a "model" given by W, rather than the usual Taylor polynomials. This formal analogy between controlled rough paths and Taylor expansions suggests that it might be fruitful to systematically investigate what are the "right" objects that could possibly take the place of Taylor polynomials, while still retaining many of their nice properties.

13.2 Definition of a regularity structure and first examples

The first step in such an endeavour is to set up an algebraic structure reflecting the properties of Taylor expansions. First of all, such a structure should contain a vector space T that will contain the coefficients of our expansion. It is natural to assume that T has a graded structure: $T = \bigoplus_{\alpha \in A} T_\alpha$, for some set A of possible "homogeneities". For example, in the case of the usual Taylor expansion (13.2), it is natural to take for A the set of natural numbers and to have T_ℓ contain the coefficients corresponding to the derivatives of order ℓ. In the case of controlled rough paths however, it is natural to take $A = \{0, \alpha\}$, to have again T_0 contain the value of the function Y at any time s, and to have T_α contain the Gubinelli derivative Y_s'. This reflects the fact that the "monomial" $t \mapsto X_{s,t}$ only vanishes at order α near $t = s$, while the usual monomials $t \mapsto (t-s)^\ell$ vanish at integer order ℓ.

This however isn't the full algebraic structure describing Taylor-like expansions. Indeed, one of the characteristics of Taylor expansions is that an expansion around some point x_0 can be re-expanded around any other point x_1 by writing

$$(x - x_0)^m = \sum_{k+\ell=m} \frac{m!}{k!\ell!} (x_1 - x_0)^k \cdot (x - x_1)^\ell \; . \tag{13.3}$$

(In the case when $x \in \mathbf{R}^d$, k, ℓ and m denote multi-indices and $k! = k_1! \ldots k_d!$.) Somewhat similarly, in the case of controlled rough paths, we have the (rather trivial) identity

$$W_{s_0,t} = W_{s_0,s_1} \cdot 1 + 1 \cdot W_{s_1,t} \; . \tag{13.4}$$

What is a natural abstraction of this fact? In terms of the coefficients of a "Taylor expansion", the operation of reexpanding around a different point is ultimately just a linear operation from $\Gamma \colon T \to T$, where the precise value of the map Γ depends on the starting point x_0, the endpoint x_1, and possibly also on the details of the particular "model" that we are considering. In view of the above examples, it is natural to impose furthermore that Γ has the property that if $\tau \in T_\alpha$, then $\Gamma\tau - \tau \in \bigoplus_{\beta<\alpha} T_\beta$. In other words, when reexpanding a homogeneous monomial around a different point, the leading order coefficient remains the same, but lower order monomials may appear.

These heuristic considerations can be summarised in the following definition of an abstract object we call a *regularity structure*:

Definition 13.1. A *regularity structure* $\mathscr{T} = (A, T, G)$ consists of the following elements:

- An index set $A \subset \mathbf{R}$ such that $0 \in A$, A is bounded from below, and A is locally finite.
- A *model space* T, which is a graded vector space $T = \bigoplus_{\alpha \in A} T_\alpha$, with each T_α a Banach space; elements in T_α are said to have *homogeneity* (or *degree*) α. Furthermore $T_0 = \langle \mathbf{1} \rangle \cong \mathbf{R}$. Given $\tau \in T$, we will write $\|\tau\|_\alpha$ for the norm of its component in T_α.
- A *structure group* G of (continuous[1]) linear operators acting on T such that, for every $\Gamma \in G$, every $\alpha \in A$, and every $\tau_\alpha \in T_\alpha$, one has

$$\Gamma \tau_\alpha - \tau_\alpha \in T_{<\alpha} \stackrel{\text{def}}{=} \bigoplus_{\beta < \alpha} T_\beta \;. \tag{13.5}$$

  Furthermore, $\Gamma \mathbf{1} = \mathbf{1}$ for every $\Gamma \in G$.

Remark 13.2. The assumption $T_0 \cong \mathbf{R}$ is not really crucial, but it is convenient in some cases and all of the natural examples we know of do satisfy it.

Remark 13.3. In principle, the index set A can be infinite. By analogy with the polynomials, it is then natural to consider T as the set of all formal series of the form $\sum_{\alpha \in A} \tau_\alpha$, where only finitely many of the τ_α's are non-zero. This also dovetails nicely with the particular form of elements in G. In practice however we will only ever work with finite subsets of A so that the precise topology on T does not matter as long as each of the T_α is finite-dimensional which is the case in all of the examples we will consider here.

The model space should be thought of as consisting of "abstract" Taylor expansions (or "jets") , where each element of T_α would correspond to a "monomial of degree (homogeneity) α" (this will be made meaningful with the definition of a model below). To avoid confusion between "abstract" elements of T and "concrete" associated functions (or Schwartz distributions), we will use color to denote elements of T, e.g. τ. Typically, T will be generated (as a free vector space) by a set of "basis symbols", so that T consists of all formal (finite) linear combination obtained from regarding these symbols as basis vectors. Given basis symbols/vectors $\tau_1, \tau_2, \ldots$ we indicate this by

$$T = \langle \tau_1, \tau_2, \ldots \rangle. \tag{13.6}$$

Important convention: basis symbols will always by listed in order of increasing homogeneities. That is, $\tau_i \in T_{\alpha_i}$ with $\alpha_1 \leq \alpha_2 \leq \ldots$ in (13.6). We now turn to some first examples of regularity structures.

---

[1] This only matters if $\dim T_{<\alpha} = +\infty$ for some $\alpha \in A$.

13.2.1 The canonical polynomial structure

We start with two simple special cases followed by the general polynomial structure. Fix $\gamma \in (0,1)$ and consider a real-valued function belonging to the Hölder space of exponent γ, say $f \in \mathcal{C}^\gamma$. In other words, $f : \mathbf{R} \to \mathbf{R}$, and $|f_x - f_y| \lesssim |y-x|^\gamma$ uniformly for x, y on compacts. The trivial regularity structure

$$A = \{0\} , \qquad T = T_0 = \langle \mathbf{1} \rangle \cong \mathbf{R} , \qquad G = \{I\} ,$$

allows us to interpret the function f as a T-valued map

$$x \mapsto f(x) := f_x \mathbf{1}.$$

Consider next a real-valued function $f : \mathbf{R} \to \mathbf{R}$ of class $\mathcal{C}^{2+\gamma}$, with $\gamma \in (0,1)$. By this we mean that continuous derivatives Df and $D^2 f$ exist, with $D^2 f$ locally γ-Hölder continuous. The minimal regularity structure allowing to capture the fact that $f \in \mathcal{C}^{2+\gamma}$ is

$$A = \{0, 1, 2\} , \qquad T = T_0 \oplus T_1 \oplus T_2 = \langle \mathbf{1}, X, X^2 \rangle \cong \mathbf{R}^3 ,$$

with structure group $G = \{\Gamma_h \in \mathcal{L}(T,T) : h \in (\mathbf{R}, +)\}$ where Γ_h is given, with respect to the ordered basis $\mathbf{1}, X, X^2$, by the matrix

$$\begin{pmatrix} 1 & h & h^2 \\ 0 & 1 & 2h \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that $\Gamma_g \circ \Gamma_h = \Gamma_{g+h}$, so that G inherits its group structure from $(\mathbf{R}, +)$. Moreover, the triangular form, with ones on the diagonal, expresses exactly the requirement (13.5), i.e. that the action of Γ_h on any element in T amounts to add terms of lower homogeneity. This structure allows to represent the function f and its first two derivatives as a truncated Taylor series, namely as the T-valued map

$$x \mapsto f(x) := f_x \mathbf{1} + Df_x X + \frac{1}{2} D^2 f_x X^2.$$

It is now an easy matter to generalize the above considerations to general Hölder maps of several variables, say $f : \mathbf{R}^d \to \mathbf{R}$ in the Hölder space $\mathcal{C}^{n+\gamma}$, which is defined by the obvious generalisation of (13.2) to functions on $\mathbf{R}^d$. In this case, we would take $A = \{0, 1, \dots, n\}$ and T is the space of abstract polynomials of degree at most n, in d commuting indeterminants $X_1, \dots, X_d$. This motivates the following definition.

Definition 13.4. The *canonical polynomial regularity structure* on $\mathbf{R}^d$ is given by

- $A = \mathbf{N} = \{0, 1, 2, ...\}$ is the set of nonnegative integers.
- $T = \mathbf{R}[X_1, \dots, X_d]$ is the space of polynomials in d commuting indeterminants with real coefficients and T_α is spanned by the monomials of degree $\alpha \in \mathbf{N}$.

- $G \sim (\mathbf{R}^d, +)$ acts on T via

$$\Gamma_h P(X) = P(X + h\mathbf{1}) , \qquad h \in \mathbf{R}^d ,$$

for any polynomial P.

Given an arbitrary multi-index $k = (k_1, \ldots, k_d)$, we write X^k as a shorthand for $X_1^{k_1} \cdots X_d^{k_d}$, and we write $|k| = k_1 + \cdots + k_d$. With this notation, for any $\alpha \in A = \mathbf{N}$,

$$T_\alpha = \langle X^k : |k| = \alpha \rangle. \tag{13.7}$$

13.2.2 The rough path structure

We start again from simple examples. What structure would be appropriate for Young integration? Fix $\alpha \in (0,1)$ and consider the problem of integrating a (continuous) path Y against a scalar $W \in \mathcal{C}^\alpha$. In the case of smooth W, the indefinite integral $Z = \int Y dW$ exists in Riemann–Stieltjes sense (and then $\dot{Z} = Y\dot{W}$). Otherwise, $\dot{W}$ only exists as a Schwartz distribution (more precisely, $\dot{W}$ is an element of the negative Hölder space $\mathcal{C}^{\alpha-1}$). The corresponding regularity structure is given by

$$A = \{\alpha - 1, 0\} , \qquad T = T_{\alpha-1} \oplus T_0 = \langle \dot{W}, \mathbf{1} \rangle \cong \mathbf{R}^2 , \qquad G = \{\mathrm{Id}|_T\} . \tag{13.8}$$

The potentially ill-defined product $\dot{Z} = Y\dot{W}$ can now be replaced by the perfectly well-defined (abstract) T-valued map

$$s \mapsto \dot{Z}(s) := Y_s \dot{W}.$$

We shall see later how $\dot{Z}$ gives rise to the $\dot{Z}$, the distributional derivative of the indefinite Young integral $\int Y dW$, provided of course that Y has the correct regularity such as $Y \in \mathcal{C}^\beta$ with $\alpha + \beta > 1$.

Let us next consider the "task" of representing a controlled rough path in a suitable regularity structure. More precisely, consider $\alpha \in (1/3, 1/2]$, a path $W \in \mathcal{C}^\alpha$ with values in $\mathbf{R}$, say, and $(Y, Y') \in \mathscr{D}_W^{2\alpha}$ so that

$$Y_t \approx Y_s + Y'_s W_{s,t} . \tag{13.9}$$

The right-hand side above is (some sort of) Taylor expansion, based on $W \in \mathcal{C}^\alpha$, which describes Y well near the (time) point s. We want to formalize this by attaching to each time s the "jet"

$$Y(s) := Y_s \mathbf{1} + Y'_s W .$$

Performing the substitution $\mathbf{1} \mapsto 1$, $W \mapsto (y \mapsto W_{s,t})$ gets us back to the right hand side of (13.9). This suggests to define the following regularity structure

$$A = \{0, \alpha\} , \qquad T = T_0 \oplus T_\alpha = \langle \mathbf{1}, W \rangle \cong \mathbf{R}^2 ,$$

with structure group $G = \{\Gamma_h \in \mathcal{L}(T,T) : h \in (\mathbf{R},+)\}$ where Γ_h is given, with respect to the ordered basis $\mathbf{1}, W$ by the matrix

$$\begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix}.$$

The regularity structure relevant for rough integration is essentially a combination of the two previous one. Let $\mathbf{W} = (W, \mathbb{W}) \in \mathscr{C}^\alpha$ and $(Y, Y') \in \mathscr{D}_W^{2\alpha}$ and consider the rough integral $Z := \int Y \, d\mathbf{W}$. Since, for $s \approx t$, we have

$$Z_{s,t} = \int_s^t Y \, d\mathbf{W} \approx Y_s W_{s,t} + Y'_s \mathbb{W}_{s,t} \; ,$$

this suggests (rather informally at this stage), that in the vicinity of any fixed time s, the distributional derivative of Z should have an expansion of the type

$$\dot{Z} \approx Y_s \dot{W} + Y'_s \dot{\mathbb{W}}_s \; , \tag{13.10}$$

where $\dot{W} := \partial_t W_t$ and $\dot{\mathbb{W}}_s := \partial_t \mathbb{W}_{s,t}$ are distributional derivatives. This suggests to attach the following "jet" at each point s,

$$\dot{Z}(s) := Y_s \dot{W} + Y'_s \dot{\mathbb{W}} \; , \tag{13.11}$$

which can be done with the aid of the following regularity structure.

$$\begin{aligned} &A = \{\alpha - 1, 2\alpha - 1, 0, \alpha\} \; , \\ &T = T_{\alpha-1} \oplus T_{2\alpha-1} \oplus T_0 \oplus T_\alpha = \langle \dot{W}, \dot{\mathbb{W}}, \mathbf{1}, W \rangle \cong \mathbf{R}^4 \; , \end{aligned}$$

with structure group $G = \{\Gamma_h \in \mathcal{L}(T,T) : h \in (\mathbf{R},+)\}$ where Γ_h is given, with respect to the ordered basis $\dot{W}, \dot{\mathbb{W}}, \mathbf{1}, W$, by the matrix

$$\begin{pmatrix} 1 & h & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & h \\ 0 & 0 & 0 & 1 \end{pmatrix} .$$

Equivalently,

$$\Gamma_h \mathbf{1} = \mathbf{1} \; , \quad \Gamma_h \dot{W} = \dot{W} \; , \quad \Gamma_h W = W + h\mathbf{1} \; , \quad \Gamma_h(\dot{\mathbb{W}}) = \dot{\mathbb{W}} + h\dot{W} \; .$$

It will be seen later that in this framework the function $\dot{Z}$ defined in (13.11) does indeed give rise to $\dot{Z}$, the distributional derivative of the indefinite rough integral $\int Y \, d\mathbf{W}$. The extension to multi-component rough paths, $\mathbf{W} \in \mathscr{C}([0,T], \mathbf{R}^e)$ with $e > 1$, is essentially trivial. We just need more basis vectors $\dot{W}^i, \dot{\mathbb{W}}^{j,k}, W^l$ (with $1 \le i,j,k,l \le e$):

Definition 13.5. Let $\alpha \in (1/3, 1/2]$. The *regularity structure for α-Hölder rough paths (over $\mathbf{R}^e$)* is given by

- The set of possible homogeneities is given by $A = \{\alpha - 1, 2\alpha - 1, 0, \alpha\}$.
- The model space T is given by $T = T_{\alpha-1} \oplus T_{2\alpha-1} \oplus T_0 \oplus T_\alpha \cong \mathbf{R}^{e+e^2+1+e}$ with

$$T_0 = \langle \mathbf{1} \rangle\,, \qquad T_\alpha = \langle W^1, \dots, W^e \rangle\,,$$
$$T_{\alpha-1} = \langle \dot W^1, \dots, \dot W^e \rangle\,, \qquad T_{2\alpha-1} = \langle \dot{\mathbb{W}}^{ij} : 1 \le i,j \le e \rangle\,.$$

- The group $G \sim (\mathbf{R}^e, +)$ acts on T via

$$\begin{aligned} \Gamma_h \mathbf{1} &= \mathbf{1}\,, & \Gamma_h W^i &= W^i + h^i \mathbf{1}\,,\\ \Gamma_h \dot W^i &= \dot W^i\,, & \Gamma_h \dot{\mathbb{W}}^{ij} &= \dot{\mathbb{W}}^{ij} + h^i \dot W^j\,. \end{aligned} \tag{13.12}$$

In a Brownian (rough path) context, one has Hölder regularity with exponent $\alpha = 1/2 - \kappa$, for arbitrarily small $\kappa > 0$. The above index set A, relevant for a "regularity structure view" on stochastic integration, then becomes

$$A = \Big\{ -\frac{1}{2} - \kappa, -2\kappa, 0, \frac{1}{2} - \kappa \Big\}\,,$$

which, in abusive but convenient notation, we write as

$$A = \Big\{ -\frac{1}{2}^-, 0^-, 0, \frac{1}{2}^- \Big\}\,.$$

Index sets of this form ("half-integers$^-$") will also be typical in later SPDE situations driven by spatial or space-time white noise.

13.3 Definition of a model and first examples

At this stage, a regularity structure is a completely abstract object. It only becomes useful when endowed with a *model*, which is a concrete way of associating to any $\tau \in T$ and $x_0 \in \mathbf{R}^d$, the actual "Taylor polynomial based at x_0" represented by τ. Furthermore, we want elements $\tau \in T_\alpha$ to represent functions (or possibly distributions!) that "vanish at order α" around the given point x_0 (thereby justifying our calling α homogeneity).

Since we would like to allow A to contain negative values and therefore allow elements in T to represent actual distributions, we need a suitable notion of "vanishing at order α". We achieve this by considering the size of our distributions, when tested against test functions that are localised around the given point x_0. Given a test function φ on $\mathbf{R}^d$, we write φ_x^λ as a shorthand for

$$\varphi_x^\lambda(y) = \lambda^{-d} \varphi\big(\lambda^{-1}(y - x)\big)\,.$$

Given an integer $r > 0$, we also denote by $\mathcal{B}_r$ the set of all functions $\varphi\colon \mathbf{R}^d \to \mathbf{R}$ such that $\varphi \in \mathcal{C}_b^r$ with $\|\varphi\|_{\mathcal{C}_b^r} \le 1$ that are furthermore supported in the unit ball around the origin. We also write $\mathcal{D}'(\mathbf{R}^d)$ for the space of Schwartz distributions on $\mathbf{R}^d$. With these notations, our definition of a model for a given regularity structure $\mathscr{T}$ is as follows.

Definition 13.6. Given a regularity structure $\mathscr{T}$ and an integer $d \geq 1$, a *model* $\mathrm{M} = (\Pi, \Gamma)$ for $\mathscr{T}$ on $\mathbf{R}^d$ consists of maps

$$\begin{aligned} \Pi \colon \mathbf{R}^d &\to \mathcal{L}\big(T, \mathcal{D}'(\mathbf{R}^d)\big) & \qquad \Gamma \colon \mathbf{R}^d \times \mathbf{R}^d &\to G \\ x &\mapsto \Pi_x & (x,y) &\mapsto \Gamma_{xy} \end{aligned}$$

such that $\Gamma_{xy}\Gamma_{yz} = \Gamma_{xz}$ and $\Pi_x \Gamma_{xy} = \Pi_y$. We then say that Π_x *realizes* an element of T as a Schwartz distribution.

Furthermore, write r for the smallest integer such that $r > |\min A| \geq 0$. We then impose that for every compact set $\mathfrak{K} \subset \mathbf{R}^d$ and every $\gamma > 0$, there exists a constant $C = C(\mathfrak{K}, \gamma)$ such that the bounds

$$\big|\big(\Pi_x \tau\big)(\varphi_x^\lambda)\big| \leq C\lambda^\alpha \|\tau\|_\alpha\;, \qquad \|\Gamma_{xy}\tau\|_\beta \leq C|x-y|^{\alpha-\beta}\|\tau\|_\alpha\;, \tag{13.13}$$

hold uniformly over $\varphi \in \mathcal{B}_r$, $(x,y) \in \mathfrak{K}$, $\lambda \in (0,1]$, $\tau \in T_\alpha$ with $\alpha \leq \gamma$ and $\beta < \alpha$.

One very important remark is that the space $\mathscr{M}$ of all models for a given regularity structure is *not* a linear space. However, it can be viewed as a closed subset (determined by the nonlinear constraints $\Gamma_{xy} \in G$, $\Gamma_{xy}\Gamma_{yz} = \Gamma_{xz}$, and $\Pi_y = \Pi_x\Gamma_{xy}$) of the linear space with seminorms (indexed by the compact set $\mathfrak{K}$) given by the smallest constant C in (13.13). Also, there is a natural distance between models (Π, Γ) and $(\bar\Pi, \bar\Gamma)$ given by the smallest constant C in (13.13), when replacing Π_x by $\Pi_x - \bar\Pi_x$ and Γ_{xy} by $\Gamma_{xy} - \bar\Gamma_{xy}$.

Remark 13.7. In principle, test functions appearing in (13.13) should be smooth. It turns out that if these bounds hold for smooth elements of $\mathcal{B}_r$, then $\Pi_x\tau$ can be extended canonically to allow any $\mathcal{C}_b^r$ test function with compact support.

Remark 13.8. The identity $\Pi_x\Gamma_{xy} = \Pi_y$ reflects the fact that Γ_{xy} is the linear map that takes an expansion around y and turns it into an expansion around x. The first bound in (13.13) states what we mean precisely when we say that $\tau \in T_\alpha$ represents a term that vanishes at order α. (See Exercise 13.31; note that α can be negative, so that this may actually not vanish at all!) The second bound in (13.13) is very natural in view of both (13.3) and (13.4). It states that when expanding a monomial of order α around a new point at distance h from the old one, the coefficient appearing in front of lower-order monomials of order β is of order at most $h^{\alpha-\beta}$.

Remark 13.9. In many cases of interest, it is natural to scale the different directions of $\mathbf{R}^d$ in a different way. This is the case for example when using the theory of regularity structures to build solution theories for parabolic stochastic PDEs, in which case the time direction "counts as" two space directions. This "parabolic scaling" can be formalized by the integer vector $(2,1,\ldots,1)$. More generally, one can introduce a scaling $\mathfrak{s}$ of $\mathbf{R}^d$, which is just a collection of d mutually prime strictly positive integers and to define φ_x^λ in such a way that the ith direction is scaled by $\lambda^{\mathfrak{s}_i}$. The polynomial structure introduced earlier, in particular (13.7), should be changed accordingly by postulating that the homogeneity of X^k is given by $|k|_\mathfrak{s} = \sum_{i=1}^d \mathfrak{s}_i k_i$. In this case, the Euclidean distance between two points should be replaced everywhere by the corresponding scaled distance $|x|_\mathfrak{s} = \sum_i |x_i|^{1/\mathfrak{s}_i}$. See [Hai14c] for more details.

With these definitions at hand, it is then natural to define an equivalent in this context of the space of γ-Hölder continuous functions in the following way.

Definition 13.10. Given a regularity structure $\mathscr{T}$ equipped with a model $\mathrm{M} = (\Pi, \Gamma)$ over $\mathbf{R}^d$, the space $\mathscr{D}_\mathrm{M}^\gamma = \mathscr{D}_\mathrm{M}^\gamma(\mathscr{T})$ is given by the set of functions $f\colon \mathbf{R}^d \to T_{<\gamma}$ such that, for every compact set $\mathfrak{K}$ and every $\alpha < \gamma$, there exists a constant C with

$$\|f(x) - \Gamma_{xy} f(y)\|_\alpha \le C|x-y|^{\gamma-\alpha} \tag{13.14}$$

uniformly over $x, y \in \mathfrak{K}$. Such functions f are called *modelled distributions*. For fixed $\mathfrak{K}$, a semi-norm

$$\|f\|_{\mathrm{M},\gamma;\mathfrak{K}}$$

is defined as the smallest constant C in the bound (13.14). The space $\mathscr{D}_\mathrm{M}^\gamma$ endowed with this family of seminorms is then a Fréchet space.

It is furthermore convenient to be able to compare two modelled distributions defined over two different models. In this case, a natural way of comparing them is to take as a "metric" the smallest constant C in the bound

$$\|f(x) - \Gamma_{xy} f(y) - \bar f(x) + \bar\Gamma_{xy} \bar f(y)\|_\alpha \le C|x-y|^{\gamma-\alpha}\;.$$

Remark 13.11. (Compare with Remark 4.8 in the rough path context.) It is important to note that while the space of models $\mathscr{M}$ is not a linear space, the space $\mathscr{D}_\mathrm{M}^\gamma$ is a linear space (with Banach, or at least Fréchet structure), given a model $\mathrm{M} \in \mathscr{M}$. The twist of course is that the space in question depends in a crucial way on the choice of M. The total space then is

$$\mathscr{M} \ltimes \mathscr{D}^\gamma \stackrel{\text{def}}{=} \bigcup_{\mathrm{M}\in\mathscr{M}} \{\mathrm{M}\} \times \mathscr{D}_\mathrm{M}^{2\alpha},$$

with base space $\mathscr{M}$ and "fibres" $\mathscr{D}_\mathrm{M}^\gamma$.

The most fundamental result in the theory of regularity structures then states that given $f \in \mathscr{D}^\gamma$ with $\gamma > 0$, there exists a *unique* Schwartz distribution $\mathcal{R}f$ on $\mathbf{R}^d$ such that, for every $x \in \mathbf{R}^d$, $\mathcal{R}f$ "looks like $\Pi_x f(x)$ near x". More precisely, one has

Theorem 13.12 (Reconstruction). *Let* $\mathrm{M} = (\Pi, \Gamma)$ *be a model for a regularity structure* $\mathscr{T}$ *on* $\mathbf{R}^d$. *Assume* $f \in \mathscr{D}_\mathrm{M}^\gamma$ *with* $\gamma > 0$. *Then, there exists a unique linear map*

$$\mathcal{R} = \mathcal{R}_\mathrm{M}\colon \mathscr{D}_\mathrm{M}^\gamma \to \mathcal{D}'(\mathbf{R}^d)$$

such that

$$\big|\big(\mathcal{R}f - \Pi_x f(x)\big)(\varphi_x^\lambda)\big| \lesssim \lambda^\gamma\;, \tag{13.15}$$

uniformly over $\varphi \in \mathcal{B}_r$ and λ as before, and locally uniformly in x. Without the positivity assumption on γ, everything remains valid but uniqueness of $\mathcal{R}$.

Remark 13.13. Actually, $\mathcal{R}$ should really be viewed as a (nonlinear!) map from the total space $\mathscr{M} \ltimes \mathscr{D}^\gamma$ into $\mathcal{D}'(\mathbf{R}^d)$. It is then also continuous with respect to the natural topology on this space, which is essential when using it to prove convergence results. We will however not prove this stronger continuity statement here.

In the particular case where $\Pi_x\tau$ happens to be a continuous function for every $\tau \in T$ (and every $x \in \mathbf{R}^d$), we will see that $\mathcal{R}f$ is also a continuous function and $\mathcal{R}$ is given by the somewhat trivial explicit formula

$$\big(\mathcal{R}f\big)(x) = \big(\Pi_x f(x)\big)(x)\;.$$

We postpone the proof of the reconstruction theorem, as well as the above remark, and turn instead to our previous list of regularity structures, now adding the relevant models and indicate the interest of the reconstruction map.

13.3.1 The canonical polynomial model

Recall the canonical polynomial regularity structure in d variables. In this context, the canonical polynomial model P is given by

$$\big(\Pi_x X^k\big) = (y \mapsto (y-x)^k)\;, \qquad \Gamma_{xy} = \Gamma_h\big|_{h=x-y}\;.$$

We leave it as an exercise to the reader to verify that this does indeed satisfy the bounds and relations of Definition 13.6.

In the sense of the following proposition, modelled distributions in the context of the polynomial model are nothing than classical Hölder functions.

Proposition 13.14. *Let $\beta = n + \gamma$ with $n \in \mathbf{N}$ and $\gamma \in (0,1)$. Then f is an element in the Hölder space $\mathcal{C}^\beta$ if and only if there exists a function $\hat{f} \in \mathscr{D}^\beta_{\mathrm{P}}$ with $\langle \hat{f}, \mathbf{1}\rangle = f$.*

The proof is not difficult. Given $f \in \mathcal{C}^{n+\gamma}$, write $f(x)$ for the Taylor expansion up to order n with all monomials $(y-x)^k$ replaced by X^k. It is immediate to check that $\hat{f} = f$ will do. The converse is obvious when $n = 0$, the general case can be seen by induction. The proposition remains valid for integer values of β with the usual caveat that in this context $\mathcal{C}^\beta$ means $\beta - 1$ times continuously differentiable with the highest order derivatives locally Lipschitz continuous.

Validity of such a proposition for negative exponents requires a suitable notion "negative" Hölder spaces. In fact, the considerations above (see also Exercise 13.31) suggest that a very natural space of distributions is obtained in the following way.

Given $\alpha > 0$, we denote by $\mathcal{C}^{-\alpha}$ the space of all Schwartz distributions η such that η belongs to the dual of $\mathcal{C}_c^r$ (elements in $\mathcal{C}_b^r$ with compact support) with r the smallest integer such that $r > \alpha$, and such that

$$\big|\eta(\varphi_x^\lambda)\big| \lesssim \lambda^{-\alpha} ,$$

uniformly over all $\varphi \in \mathcal{B}_r$ and $\lambda \in (0,1]$, and locally uniformly in x. Given any compact set $\mathfrak{K}$, the best possible constant such that the above bound holds uniformly over $x \in \mathfrak{K}$ yields a seminorm. The collection of these seminorms endows $\mathcal{C}^{-\alpha}$ with a Fréchet space structure.

Remark 13.15. In terms of the scale of classical Besov spaces, the space $\mathcal{C}^{-\alpha}$ is a local version of $\mathcal{B}^{-\alpha}_{\infty,\infty}$. It is in some sense the largest space of distributions that is invariant under the scaling $\varphi(\cdot) \mapsto \lambda^{-\alpha}\varphi(\lambda^{-1}\cdot)$, see for example [BP08].

Let us now give a very simple application of the reconstruction theorem. It is a classical result in the "folklore" of harmonic analysis (see for example [BCD11, Thm 2.52] for a very similar statement) that the product extends naturally to $\mathcal{C}^{-\alpha} \times \mathcal{C}^{\beta}$ into $\mathcal{D}'(\mathbf{R}^d)$ if and only if $\beta > \alpha$. Let us illustrate how to use the reconstruction theorem in order to obtain a straightforward proof of the "if" part of this result:

Theorem 13.16. *For $\beta > \alpha > 0$, there is a continuous bilinear map*

$$B \colon \mathcal{C}^{\beta} \times \mathcal{C}^{-\alpha} \to \mathcal{D}'(\mathbf{R}^d)$$

such that $B(f,g) = fg$ for any two continuous functions f and g.

Proof. Assume from now on that $g = \xi \in \mathcal{C}^{-\alpha}$ for some $\alpha > 0$ and that $f \in \mathcal{C}^{\beta}$ for some $\beta > \alpha$. We then build a regularity structure $\mathscr{T}$ in the following way. For the set A, we take $A = \mathbf{N} \cup (\mathbf{N} - \alpha)$ and for T, we set $T = V \oplus W$, where each one of the spaces V and W is a copy of the canonical polynomial model (in d commuting variables). We also choose Γ as in the canonical polynomial model above, acting simultaneously on each of the two instances.

As before, we denote by X^k the canonical basis vectors in V. We also use the suggestive notation "ΞX^k" for the corresponding basis vector in W, but we postulate that $\Xi X^k \in T_{|k|-\alpha}$ rather than $\Xi X^k \in T_{|k|}$. Given any distribution $\xi \in \mathcal{C}^{-\alpha}$, we then define a model (Π^ξ, Γ), where Γ is as in the canonical model, while Π^ξ acts as

$$\big(\Pi_x^\xi X^k\big)(y) = (y-x)^k , \qquad \big(\Pi_x^\xi \Xi X^k\big)(y) = (y-x)^k \xi(y) ,$$

with the obvious abuse of notation in the second expression. It is then straightforward to verify that $\Pi_y = \Pi_x \circ \Gamma_{xy}$ and that the relevant analytical bounds are satisfied, so that this is indeed a model.

Denote now by $\mathcal{R}^\xi$ the reconstruction map associated to the model (Π^ξ, Γ) and, for $f \in \mathcal{C}^\beta$, denote by f the element in $\mathscr{D}^\beta$ given by the local Taylor expansion of f of order β at each point. Note that even though the space $\mathscr{D}^\beta$ does in principle depend on the choice of model, in our situation $f \in \mathscr{D}^\beta$ for any choice of ξ. It follows immediately from the definitions that the map $x \mapsto \Xi f(x)$ belongs to $\mathscr{D}^{\beta-\alpha}$ so that, provided that $\beta > \alpha$, one can apply the reconstruction operator to it. This suggests that the multiplication operator we are looking for can be defined as

$$B(f,\xi) = \mathcal{R}^{\xi}\big(\Xi f\big) .$$

By Theorem 13.12, this is a jointly continuous map from $\mathcal{C}^{\beta} \times \mathcal{C}^{-\alpha}$ into $\mathcal{D}'(\mathbf{R}^d)$, provided that $\beta > \alpha$. If ξ happens to be a smooth function, then it follows immediately from the remark after Theorem 13.12 that $B(f,\xi) = f(x)\xi(x)$, so that B is indeed the requested continuous extension of the usual product. □

Remark 13.17. In the context of this theorem, one can actually show that $B(f,g) \in \mathcal{C}^{-\alpha}$. More generally, denoting by $-\alpha$ the smallest homogeneity arising in a given regularity structure $\mathscr{T}$, i.e. $\alpha = -\min A$, it is possible to show that the reconstruction operator $\mathcal{R}$ takes values in $\mathcal{C}^{-\alpha}$.

The reader may notice that one can also work with a finite-dimensional regularity structure, based on index set $\tilde{N} \cup (\tilde{N} - \alpha)$, with $\tilde{N} = \{0, 1, \ldots, n\}$ and $\beta = n + \gamma$. In particular, if $n = 0$, the regularity structure used here is exacty the one already encountered in (13.8).

13.3.2 The rough path model

Let us see now how some of the results of Section 4 can be reinterpreted in the light of this theory. Fix $\alpha \in (1/3, 1/2]$ and let $\mathscr{T}$ be the rough path regularity structure put forward in Definition 13.5. Recall that this means $A = \{\alpha - 1, 2\alpha - 1, 0, \alpha\}$. We have for T_0 a copy of $\mathbf{R}$ with unit vector $\mathbf{1}$, for T_α and $T_{\alpha-1}$ a copy of $\mathbf{R}^e$ with respective unit vectors W^j and $\dot{W}^j$, and for $T_{2\alpha-1}$ a copy of $\mathbf{R}^{e\times e}$ with unit vectors $\dot{\mathbb{W}}^{ij}$. The structure group G is isomorphic to $\mathbf{R}^e$ and, for $h \in \mathbf{R}^e$, acts on T via

$$\Gamma_h \mathbf{1} = \mathbf{1} \,, \quad \Gamma_h \dot{W}^i = \dot{W}^i \,, \quad \Gamma_h W^i = W^i + h^i \mathbf{1} \,, \quad \Gamma_h \dot{\mathbb{W}}^{ij} = \dot{\mathbb{W}}^{ij} + h^i \dot{W}^j \,. \tag{13.16}$$

Let now $\mathbf{W} = (W, \mathbb{W})$ be an α-Hölder continuous rough path over $\mathbf{R}^e$. It turns out that this defines a model for $\mathscr{T}$ in the following way:

Lemma 13.18. *Given an α-Hölder continuous rough path $\mathbf{W}$, one can define a model $\mathrm{M} = \mathrm{M}_{\mathbf{W}}$ for $\mathscr{T}$ on $\mathbf{R}$ by setting $\Gamma_{t,s} = \Gamma_{W_{s,t}}$ and*

$$\begin{aligned} \big(\Pi_s \mathbf{1}\big)(t) &= 1 \,, & \big(\Pi_s W^j\big)(t) &= W^j_{s,t} \\ \big(\Pi_s \dot{W}^j\big)(\psi) &= \int \psi(t)\, dW^j_t \,, & \big(\Pi_s \dot{\mathbb{W}}^{ij}\big)(\psi) &= \int \psi(t)\, d\mathbb{W}^{ij}_{s,t} \,. \end{aligned}$$

Here, both integrals are perfectly well-defined Riemann integrals, with the differential in the second case taken with respect to the variable t. Given a controlled rough path $(Y, Y') \in \mathscr{D}^{2\alpha}_W$, this then defines an element $Y \in \mathscr{D}^{2\alpha}_{\mathrm{M}}$ by

$$\mathbf{Y}(s) = Y(s)\,\mathbf{1} + Y_i'(s)\,W^i\;,$$

with summation over i implied.

Proof. We first check that the algebraic properties of Definition 13.6 are satisfied. It is clear that $\Gamma_{s,u}\Gamma_{u,t} = \Gamma_{s,t}$ and that $\Pi_s \Gamma_{s,u}\tau = \Pi_u \tau$ for $\tau \in \{\mathbf{1}, W^j, \dot W^j\}$. Regarding $\dot{\mathbb{W}}^{ij}$, we differentiate Chen's relations (2.1) which yields the identity

$$d\mathbb{W}^{i,j}_{s,t} = d\mathbb{W}^{i,j}_{u,t} + W^i_{s,u}\,dW^j_t\;.$$

The last missing algebraic relation then follows at once. The required analytic bounds follow immediately (exercise!) from the definition of the rough path space $\mathscr{C}^\alpha$.

Regarding the function $\mathbf{Y}$ defined in the statement, we have

$$\begin{aligned}
\|\mathbf{Y}(s) - \Gamma_{s,u}\mathbf{Y}(u)\|_0 &= |Y(s) - Y(u) + Y_i'(u)W^i_{s,u}|\;,\\
\|\mathbf{Y}(s) - \Gamma_{s,u}\mathbf{Y}(u)\|_\alpha &= |Y'(s) - Y'(u)|\;,
\end{aligned}$$

so that the condition (13.14) with $\gamma = 2\alpha$ does indeed coincide with the definition of a controlled rough path. □

Theorems 4.4 and 4.10 can then be recovered as a particular case of the reconstruction theorem in the following way.

Proposition 13.19. *In the same context as above, let $\alpha \in (\frac{1}{3}, \frac{1}{2}]$, and consider the modelled distribution $\mathbf{Y} \in \mathscr{D}^{2\alpha}_{\mathrm{M_W}}$ built as above from a controlled rough path $(Y, Y') \in \mathscr{D}^{2\alpha}_W$. Then, the map $\mathbf{Y}\dot W^i$ given by*

$$\big(\mathbf{Y}\dot W^j\big)(s) := Y(s)\,\dot W^j + Y_i'(s)\,\dot{\mathbb{W}}^{ij}$$

belongs to $\mathscr{D}^{3\alpha-1}$. Furthermore, there exists an essentially unique function Z such that

$$\big(\mathcal{R}\mathbf{Y}\dot W^j\big)(\psi) = \int \psi(t)\,dZ(t)\;,$$

and such that $Z_{s,t} = Y(s)\,W^j_{s,t} + Y_i'(s)\,\mathbb{W}^{i,j}_{s,t} + \mathrm{O}(|t-s|^{3\alpha})$.

Remark 13.20. The function Z is unique up to addition of constants.

Proof. The fact that $\mathbf{Y}\dot W^j \in \mathscr{D}^{3\alpha-1}$ is an immediate consequence of the definitions. Since $\alpha > \frac{1}{3}$ by assumption, we can apply the reconstruction theorem to it, from which it follows that there exists a unique distribution η such that, if ψ is a smooth compactly supported test function, one has

$$\eta(\psi_s^\lambda) = \int \psi_s^\lambda(t) Y(s)\,dW^j_t + \int \psi_s^\lambda(t) Y_i'(s)\,d\mathbb{W}^{i,j}_{s,t} + \mathrm{O}(\lambda^{3\alpha-1})\;.$$

By a simple approximation argument, it turns out that one can take for ψ the indicator function of the interval $[0,1]$, so that

$$\eta(\mathbf{1}_{[s,t]}) = Y(s)\, W^j_{s,t} + Y'_i(s)\, \mathbb{W}^{i,j}_{s,t} + \mathrm{O}(|t-s|^{3\alpha}) \;.$$

Here, the reason why one obtains an exponent 3α rather than $3\alpha - 1$ is that it is really $|t-s|^{-1}\mathbf{1}_{[s,t]}$ that scales like an approximate δ-distribution as $t \to s$. $\square$

Remark 13.21. Using the formula (13.25), it is straightforward to verify that if W happens to be a smooth function and $\mathbb{W}$ is defined from W via (2.2), but this time viewing it as a definition for the right hand side, with the left hand side given by a usual Riemann integral, then the function Z constructed in Proposition 13.19 coincides with the usual Riemann integral of Y against W^j.

Remark 13.22. The theory of (controlled) rough paths of lower regularity already hinted at in Section 2.4 can be recovered from the reconstruction operator and a suitable choice of regularity structure (essentially two copies of the truncated tensor algebra) in virtually the same way.

Let us give another application to rough path theory. Given an arbitrary path $W \in \mathcal{C}^\alpha$ with values in $\mathbf{R}^e$, does there exist a (since $\alpha \leq 1/2$: non-unique) rough path lift? In dimension $e = 1$, the answer is trivially yes, it suffices set $\mathbb{W}_{s,t} = \frac{1}{2} W^2_{s,t}$ but the case of $e > 1$ is non-trivial. The following can be obtained as easy application of the reconstruction theorem in the case $\gamma \leq 0$.

Proposition 13.23 (Lyons–Victoir extension; [LV07]). *For any $W \in \mathcal{C}^\alpha$ with values in $\mathbf{R}^e$ for $e > 1$, there exist a rough path lift, i.e. $\mathbb{W}$ so that*

$$\mathbf{W} = (W, \mathbb{W}) \in \mathscr{C}^\alpha([0,T], \mathbf{R}^e) \;.$$

Furthermore, this can be done is such a way that the map $W \mapsto \mathbf{W}$ is continuous.

Remark 13.24. The reader may wonder how this dovetails with Proposition 1.1. The point is that if we define $W \mapsto \mathbf{W}$ by an application of the reconstruction theorem with $\gamma < 0$, this map restricted to smooth paths does in general *not* coincide with the Riemann–Stieltjes integral of W against itself.

13.4 Wavelets and the reconstruction theorem

We trust the reader is familiar with the Haar (wavelet) basis. The analysis seen earlier in the rough path context (e.g. the proof of the sewing lemma, based on dyadic refinements) can be viewed as based on this wavelet basis. The Haar basis, however, suffers from lack of regularity. Fortunately, the following result due to Daubechies [Dau88] provides us with much more regular functions that enjoy analogous properties:

Theorem 13.25. *Given any integer $0 < r < \infty$, there exists a function $\varphi \colon \mathbf{R}^d \to \mathbf{R}$ with the following properties:*

1. *The function φ is of class $\mathcal{C}^r_b$ and has compact support.*

2. *For every polynomial P of degree r, there exists a polynomial $\hat{P}$ of degree r such that, for every $x \in \mathbf{R}^d$, one has $\sum_{y\in\mathbf{Z}^d} \hat{P}(y)\varphi(x-y) = P(x)$.*
3. *One has $\int \varphi(x)\varphi(x-y)\,dx = \delta_{y,0}$ for every $y \in \mathbf{Z}^d$.*
4. *There exist coefficients $\{a_k\}_{k\in\mathbf{Z}^d}$ such that $2^{-d/2}\varphi(x/2) = \sum_{k\in\mathbf{Z}^d} a_k\varphi(x-k)$.*

The existence of such a function φ is highly non-trivial and actually equivalent to the existence of a wavelet basis consisting of $\mathcal{C}_b^r$ functions with compact support. Let us restate the reconstruction theorem for the reader's convenience. (We only consider the case $\gamma > 0$ here.)

Theorem 13.26. *Let $\mathscr{T}$ be a regularity structure as above and let (Π, Γ) a model for $\mathscr{T}$ on $\mathbf{R}^d$. Then, there exists a unique linear map $\mathcal{R}\colon \mathscr{D}^\gamma \to \mathcal{D}'(\mathbf{R}^d)$ such that*

$$\big|\big(\mathcal{R}f - \Pi_x f(x)\big)(\varphi_x^\lambda)\big| \lesssim \lambda^\gamma\;,$$

uniformly over $\varphi \in \mathcal{B}_r$ and $\lambda \in (0,1]$, and locally uniformly in x.

Proof. We pick φ with properties (1–4), as provided by Theorem 13.25, for some $r > |\inf A|$. We also set $\Lambda^n = 2^{-n}\mathbf{Z}^d$ and, for $y \in \Lambda^n$, we set $\varphi_y^n(x) = 2^{nd/2}\varphi(2^n(x-y))$. Here, the normalisation is chosen in such a way that the set $\{\varphi_y^n\}_{y\in\Lambda^n}$ is again orthonormal in L^2. We then denote by $V_n \subset \mathcal{C}^r$ the linear span of $\{\varphi_y^n\}_{y\in\Lambda^n}$, so that, by the property (4) above, one has $V_0 \subset V_1 \subset V_2 \subset \ldots$. We furthermore denote by $\hat{V}_n$ the L^2-orthogonal complement of V_{n-1} in V_n, so that $V_n = V_0 \oplus \hat{V}_1 \oplus \ldots \oplus \hat{V}_n$. In order to keep notations compact, it will also be convenient to define the coefficients a_k^n with $k \in \Lambda^n$ by $a_k^n = a_{2^n k}$.

With these notations at hand, we then define a sequence of linear operators $\mathcal{R}^n\colon \mathscr{D}^\gamma \to \mathcal{C}^r$ by

$$\big(\mathcal{R}^n f\big)(y) = \sum_{x\in\Lambda^n} \big(\Pi_x f(x)\big)(\varphi_x^n)\,\varphi_x^n(y)\;.$$

We claim that there then exists a Schwartz distribution $\mathcal{R}f$ such that, for every compactly supported test function ψ of class $\mathcal{C}^r$, one has $\langle \mathcal{R}^n f, \psi\rangle \to \big(\mathcal{R}f\big)(\psi)$, and that $\mathcal{R}f$ furthermore satisfies the properties stated in the theorem.

Let us first consider the size of the components of $\mathcal{R}^{n+1}f - \mathcal{R}^n f$ in V_n. Given $x \in \Lambda^n$, we make use of properties (3-4), so that

$$\begin{aligned}
\langle \mathcal{R}^{n+1}f - \mathcal{R}^n f, \varphi_x^n\rangle &= \sum_{k\in\Lambda^{n+1}} a_k^n \langle \mathcal{R}^{n+1}f, \varphi_{x+k}^{n+1}\rangle - \big(\Pi_x f(x)\big)(\varphi_x^n)\\
&= \sum_{k\in\Lambda^{n+1}} a_k^n \big(\Pi_{x+k} f(x+k)\big)(\varphi_{x+k}^{n+1}) - \big(\Pi_x f(x)\big)(\varphi_x^n)\\
&= \sum_{k\in\Lambda^{n+1}} a_k^n \Big(\big(\Pi_{x+k} f(x+k)\big)(\varphi_{x+k}^{n+1}) - \big(\Pi_x f(x)\big)(\varphi_{x+k}^{n+1})\Big)\\
&= \sum_{k\in\Lambda^{n+1}} a_k^n \Big(\Pi_{x+k}\big(f(x+k) - \Gamma_{x+k,x} f(x)\big)\Big)(\varphi_{x+k}^{n+1})\;,
\end{aligned}$$

where we used the algebraic relations between Π_x and Γ_{xy} to obtain the last identity. Since only finitely many of the coefficients a_k are non-zero, it follows from the definition of $\mathscr{D}^\gamma$ that for the non-vanishing terms in this sum we have the bound

$$\|f(x+k) - \Gamma_{x+k,x} f(x)\|_\alpha \lesssim 2^{-n(\gamma-\alpha)} ,$$

uniformly over $n \geq 0$ and x in any compact set. Furthermore, for any $\tau \in T_\alpha$, it follows from the definition of a model that one has the bound

$$\big|\big(\Pi_x \tau\big)(\varphi_x^n)\big| \lesssim 2^{-\alpha n - \frac{nd}{2}} ,$$

again uniformly over $n \geq 0$ and x in any compact set. Here, the additional factor $2^{-\frac{nd}{2}}$ comes from the fact that the functions φ_x^n are normalised in L^2 rather than L^1. Combining these two bounds, we immediately obtain that

$$\big|\langle \mathcal{R}^{n+1} f - \mathcal{R}^n f, \varphi_x^n\rangle\big| \lesssim 2^{-\gamma n - \frac{nd}{2}} , \tag{13.17}$$

uniformly over $n \geq 0$ and x in compact sets. Take now a test function $\psi \in \mathcal{C}_b^r$ with compact support and let us try to estimate $\langle \mathcal{R}^{n+1} f - \mathcal{R}^n f, \psi\rangle$. Since $\mathcal{R}^{n+1} f - \mathcal{R}^n f \in V_{n+1}$, we can decompose it into a part $\delta\mathcal{R}^n f \in V_n$ and a part $\hat\delta\mathcal{R}^n f \in \hat V_{n+1}$ and estimate both parts separately. Regarding the part in V_n, we have

$$\big|\langle \delta\mathcal{R}^n f, \psi\rangle\big| = \Big|\sum_{x\in\Lambda^{n+1}} \langle \delta\mathcal{R}^n f, \varphi_x^n\rangle\langle\varphi_x^n,\psi\rangle\Big| \lesssim 2^{-\gamma n - \frac{nd}{2}} \sum_{x\in\Lambda^{n+1}} \big|\langle\varphi_x^n,\psi\rangle\big| , \tag{13.18}$$

where we made use of the bound (13.17). At this stage we use the fact that, due to the boundedness of ψ, we have $\big|\langle\varphi_x^n,\psi\rangle\big| \lesssim 2^{-nd/2}$. Furthermore, thanks to the boundedness of the support of ψ, the number of non-vanishing terms appearing in this sum is bounded by 2^{nd}, so that we eventually obtain the bound

$$\big|\langle \delta\mathcal{R}^n f, \psi\rangle\big| \lesssim 2^{-\gamma n} . \tag{13.19}$$

Regarding the second term, we use the standard fact coming from wavelet analysis [Mey92] that a basis of $\hat V_{n+1}$ can be obtained in the same way as the basis of V_n, but replacing the function φ by functions $\hat\varphi$ from some finite set Φ. In other words, $\hat V_{n+1}$ is the linear span of $\{\hat\varphi_x^n\}_{x\in\Lambda^n;\hat\varphi\in\Phi}$. Furthermore, as a consequence of property (2), the functions $\hat\varphi \in \Phi$ all have the property that

$$\int \hat\varphi(x)\, P(x)\, dx = 0 , \tag{13.20}$$

for any polynomial P of degree less or equal to r. In particular, this shows that one has the bound

$$|\langle\hat\varphi_x^n,\psi\rangle| \lesssim 2^{-\frac{nd}{2} - nr} .$$

As a consequence, one has

$$\big|\langle\hat\delta\mathcal{R}^n f,\psi\rangle\big| = \Big|\sum_{\substack{x\in\Lambda^n\\ \hat\varphi\in\Phi}}\langle\mathcal{R}^{n+1}f,\hat\varphi_x^n\rangle\langle\hat\varphi_x^n,\psi\rangle\Big| \lesssim 2^{-\frac{nd}{2}-nr}\Big|\sum_{\substack{x\in\Lambda^n\\ \hat\varphi\in\Phi}}\langle\mathcal{R}^{n+1}f,\hat\varphi_x^n\rangle\Big| .$$

At this stage, we note that, thanks to the definition of $\mathcal{R}^{n+1}$ and the bounds on the model (Π,Γ), we have $|\langle\mathcal{R}^{n+1}f,\hat\varphi_x^n\rangle| \lesssim 2^{-\frac{nd}{2}-\alpha_0 n}$, where $\alpha_0 = \inf A$, so that $\big|\langle\hat\delta\mathcal{R}^n f,\psi\rangle\big| \lesssim 2^{-nr-\alpha_0 n}$. Combining this with (13.19), we see that one has indeed $\mathcal{R}^n f \to \mathcal{R}f$ for some Schwartz distribution $\mathcal{R}f$.

It remains to show that the bound (13.15) holds. For this, given a distribution $\eta\in\mathcal{C}^\alpha$ for some $\alpha > -r$, we first introduce the notation

$$\mathcal{P}_n\eta = \sum_{x\in\Lambda^n}\eta(\varphi_x^n)\,\varphi_x^n , \qquad \hat{\mathcal{P}}_n\eta = \sum_{\hat\varphi\in\Phi}\sum_{x\in\Lambda^n}\eta(\hat\varphi_x^n)\,\hat\varphi_x^n .$$

We also choose an integer value $n\ge 0$ such that $2^{-n}\sim\lambda$ and we write

$$\begin{aligned}\mathcal{R}f - \Pi_x f(x) &= \mathcal{R}^n f - \mathcal{P}_n\Pi_x f(x) + \sum_{m\ge n}\big(\mathcal{R}^{m+1}f - \mathcal{R}^m f - \hat{\mathcal{P}}_m\Pi_x f(x)\big)\\ &= \mathcal{R}^n f - \mathcal{P}_n\Pi_x f(x) + \sum_{m\ge n}\big(\hat\delta\mathcal{R}^m f - \hat{\mathcal{P}}_m\Pi_x f(x)\big)\\ &\quad + \sum_{m\ge n}\delta\mathcal{R}^m f .\end{aligned}\tag{13.21}$$

We then test these terms against ψ_x^λ and we estimate the resulting terms separately. For the first term, we have the identity

$$\big(\mathcal{R}^n f - \mathcal{P}_n\Pi_x f(x)\big)(\psi_x^\lambda) = \sum_{y\in\Lambda^n}\big(\Pi_y f(y) - \Pi_x f(x)\big)(\varphi_y^n)\,\langle\varphi_y^n,\psi_x^\lambda\rangle .\tag{13.22}$$

We have the bound $|\langle\varphi_y^n,\psi_x^\lambda\rangle| \lesssim \lambda^{-d}2^{-dn/2}\sim 2^{dn/2}$. Since one furthermore has $|y-x|\lesssim\lambda$ for all non-vanishing terms in the sum, one also has similarly to before

$$|\big(\Pi_y f(y) - \Pi_x f(x)\big)(\varphi_y^n)| \lesssim \sum_{\alpha<\gamma}\lambda^{\gamma-\alpha}2^{-\frac{dn}{2}-\alpha n}\sim 2^{-\frac{dn}{2}-\gamma n} .\tag{13.23}$$

Since only finitely many (independently of n) terms contribute to the sum in (13.22), it is indeed bounded by a constant proportional to $2^{-\gamma n}\sim\lambda^\gamma$ as required.

We now turn to the second term in (13.21), where we consider some fixed value $m\ge n$. We rewrite this term very similarly to before as

$$\begin{aligned}&\big(\hat\delta\mathcal{R}^m f - \hat{\mathcal{P}}_m\Pi_x f(x)\big)(\psi_x^\lambda)\\ &\qquad = \sum_{\hat\varphi\in\Phi}\sum_{y,z}\big(\Pi_y f(y) - \Pi_x f(x)\big)(\varphi_y^{m+1})\,\langle\varphi_y^{m+1},\hat\varphi_z^m\rangle\,\langle\hat\varphi_z^m,\psi_x^\lambda\rangle ,\end{aligned}$$

where the sum runs over $y\in\Lambda^{m+1}$ and $z\in\Lambda^m$. This time, we use the fact that by the property (13.20) of the wavelets $\hat\varphi$, one has the bound

$$|\langle \hat\varphi_z^m, \psi_x^\lambda\rangle| \lesssim \lambda^{-d-r} 2^{-rm-\frac{md}{2}}\;, \tag{13.24}$$

and the L^2-scaling implies that $|\langle \varphi_y^{m+1}, \hat\varphi_z^m\rangle| \lesssim 1$. Furthermore, for each $z \in \Lambda^m$, only finitely many elements $y \in \Lambda^{m+1}$ contribute to the sum, and these elements all satisfy $|y-z| \lesssim 2^{-m}$. Bounding the first factor as in (13.23) and using the fact that there are of the order of $\lambda^d 2^{md}$ terms contributing for every fixed m, we thus see that the contribution of the second term in (13.21) is bounded by

$$\sum_{m\ge n} \lambda^d 2^{md} \sum_{\alpha<\gamma} \lambda^{\gamma-\alpha-d-r} 2^{-dm-\alpha m - rm} \sim \sum_{\alpha<\gamma} \lambda^{\gamma-\alpha-r} \sum_{m\ge n} 2^{-\alpha m - rm} \sim \lambda^\gamma\;.$$

For the last term in (13.21), we combine (13.18) with the bound $|\langle \varphi_y^m, \psi_x^\lambda\rangle| \lesssim \lambda^{-d} 2^{-dm/2}$ and the fact that there are of the order of $\lambda^d 2^{-md}$ terms appearing in the sum (13.18) to conclude that the mth summand is bounded by a constant proportional to $2^{-\gamma m}$. Summing over m yields again the desired bound and concludes the proof. □

Remark 13.27. There are obvious analogies between the construction of the reconstruction operator $\mathcal{R}$ and that of the "rough integral" in Section 4, see also Exercise 13.33. As a matter of fact, there exists a slightly more abstract formulation of the reconstruction theorem which can be interpreted as a multidimensional analogue to the sewing lemma, Lemma 4.2.

Remark 13.28. With a look to remark 13.11, and $\mathrm{M} = (\Pi, \Gamma) \in \mathscr{M}$, one should really view $\mathcal{R} = \mathcal{R}_{\mathrm{M}} f$ as a map from $\mathscr{M} \ltimes \mathscr{D}^\gamma$ into $\mathcal{D}'$. Since the space $\mathscr{M} \ltimes \mathscr{D}^\gamma$ is *not* a linear space, this shows that the map $\mathcal{R}$ isn't actually linear, despite appearances. However, the map $(\Pi, \Gamma, f) \mapsto \mathcal{R} f$ turns out to be locally Lipschitz continuous provided that the distance between (Π, Γ, f) and $(\bar\Pi, \bar\Gamma, \bar f)$ is given by the smallest constant ϱ such that

$$\begin{aligned} \|f(x) - \bar f(x) - \Gamma_{xy} f(y) + \bar\Gamma_{xy} \bar f(y)\|_\alpha &\le \varrho |x-y|^{\gamma-\alpha}\;, \\ \big|\big(\Pi_x \tau - \bar\Pi_x \tau\big)(\varphi_x^\lambda)\big| &\le \varrho \lambda^\alpha \|\tau\|\;, \\ \|\Gamma_{xy}\tau - \bar\Gamma_{xy}\tau\|_\beta &\le \varrho |x-y|^{\alpha-\beta} \|\tau\|\;. \end{aligned}$$

Here, in order to obtain bounds on $\big(\mathcal{R} f - \bar{\mathcal{R}} \bar f\big)(\psi)$ for some smooth compactly supported test function ψ, the above bounds should hold uniformly for x and y in a neighbourhood of the support of ψ. The proof that this stronger continuity property also holds is actually crucial when showing that sequences of solutions to mollified equations all converge to the same limiting object. However, its proof is somewhat more involved which is why we chose not to give it here.

Remark 13.29. In the particular case where $\Pi_x \tau$ happens to be a continuous function for every $\tau \in T$ (and every $x \in \mathbf{R}^d$), $\mathcal{R} f$ is also a continuous function and one has the identity

$$\big(\mathcal{R} f\big)(x) = \big(\Pi_x f(x)\big)(x)\;. \tag{13.25}$$

This can be seen from the fact that

$$\big(\mathcal{R}f\big)(y) = \lim_{n\to\infty} \big(\mathcal{R}^n f\big)(y) = \lim_{n\to\infty} \sum_{x\in\Lambda^n} \big(\Pi_x f(x)\big)(\varphi_x^n)\, \varphi_x^n(y)\;.$$

Indeed, our assumptions imply that the function $(x,z) \mapsto \big(\Pi_x f(x)\big)(z)$ is jointly continuous and since the non-vanishing terms in the above sum satisfy $|x-y| \lesssim 2^{-n}$, one has $2^{dn/2}\big(\Pi_x f(x)\big)(\varphi_x^n) \approx \big(\Pi_y f(y)\big)(y)$ for large n. Since furthermore $\sum_{x\in\Lambda^n} \varphi_x^n(y) = 2^{dn/2}$, the claim follows.

13.5 Exercises

Exercise 13.30. Use wavelets to construct an example demonstrating the "only if" part of Theorem 13.16.

Exercise 13.31 (Hölder spaces).

a) For $k \in \mathbf{N}$ and $\alpha \in (0,1)$, it is customary to define $\mathcal{C}^{k+\alpha}$ as the space of k times continuously differentiable functions $f : \mathbf{R}^d \to \mathbf{R}$ such that their derivatives of order k are α-Hölder continuous. Show that this agrees with the obvious extension to $\mathbf{R}^d$ of the definition given earlier in (13.2).
b) Fix $\alpha > 0$. Show that $f \in \mathcal{C}^\alpha$ if and only if, for each x, there exists a polynomial P_x such that
$$\big\langle f - P_x, \psi_x^\lambda\big\rangle \lesssim \lambda^\alpha\;,$$
locally uniformly in x, uniformly over $\lambda \in (0,1]$ and uniformly over smooth functions $\psi \in \mathcal{D}$ with support in $B_1(0)$ such that $\|\psi\|_{\infty;B_1(0)} \le 1$.
c) Define $\mathcal{C}^{-\alpha}$ as the space of all Schwartz distributions η belonging to the dual of $\mathcal{C}^r$ with $r > \alpha$ some integer and such that
$$\big|\eta(\varphi_x^\lambda)\big| \lesssim \lambda^{-\alpha}\;,$$
uniformly over all $\varphi \in \mathcal{B}_r$ and $\lambda \in (0,1]$, and locally uniformly in x. Show that the space $\mathcal{C}^{-\alpha}$ is independent of the choice of r in the definition given above, which justifies the notation. Take now $d = 1$ and $\alpha \in (0,1)$ for simplicity. Show that any $f \in \mathcal{C}^{-\alpha}$ is the distributional derivative of some Hölder continuous function $F \in \mathcal{C}^{1-\alpha}$.

Exercise 13.32. Show that in general, the function Z defined by (15.2) coincides, up to an additive constant, with the integral $\int^t Y(s)\, dX_s^j$, interpreted in the sense of (4.19).

Exercise 13.33. Retrace the proof of Theorem 13.12 in the context of Proposition 13.19 with the Haar basis as the choice of wavelet basis (i.e. set $\varphi(x) = \mathbf{1}_{[0,1]}(x)$). Convince yourself that this is equivalent to the proof of Lemma 4.2.

Exercise 13.34. Let (Π, Γ) be a model for the "rough path" regularity structure given in Definition 13.5 with the additional property that $\Pi_s \dot W^i$ is the distributional derivative of $\Pi_s W^i$ for every s. Show that it is then necessarily of the form $\mathrm{M}_{\mathbf{W}}$ for some α-Hölder rough path $\mathbf{W}$ as in Lemma 13.18.

Exercise 13.35. Give a detailed proof of Proposition 13.23.

13.6 Comments

An alternative theory to the theory of regularity structures [Hai14c] has been introduced more or less simultaneously in Gubinelli–Imkeller–Perkowski [GIP12]. Instead of the reconstruction theorem, that theory builds instead on properties of Bony's paraproduct [Bon81, BMN10, BCD11]. It is also in principle able to deal with stochastic PDEs like the KPZ equation or the dynamical Φ^4_3 equation, see Catellier–Chouk [CC13], but its scope is not as wide as that of the theory of regularity structures. (For example, it cannot deal with classical one-dimensional parabolic SPDEs driven by space-time white noise with a diffusion coefficient depending on the solution.)

One advantage of the paraproduct-based theory is that one generally deals with globally defined objects rather than the "jets" used in the theory of regularity structures. It also uses some already well-studied objects, so that it can rely on a substantial body of existing literature. However, besides being less systematic than the theory of regularity structures, it achieves a less clean break between the analytical and the algebraic aspects of a given problem.

Chapter 14
Operations on modelled distributions

Abstract The original motivation for the development of the theory of regularity structures was to provide robust solution theories for singular stochastic PDEs like the KPZ equation or the dynamical Φ^4_3 model. The idea is to reformulate them as fixed point problems in some space $\mathscr{D}^\gamma$ (or rather a slightly modified version that takes into account possible singular behaviour near time 0) based on a suitable random model in a regularity structure purpose-built for the problem at hand. In order to achieve this this chapter provides a systematic way of formulating the standard operations arising in the construction of the corresponding fixed point problem (differentiation, multiplication, composition by a regular function, convolution with the heat kernel) as operations on the spaces $\mathscr{D}^\gamma$.

14.1 Differentiation

Being a local operation, differentiating a modelled distribution is straightforward, provided that the model one works with is sufficiently rich. Denote by $\mathcal{L}$ some (formal) differential operator with constant coefficients that is homogeneous of degree m, i.e. it is of the form

$$\mathcal{L} = \sum_{|k|=m} a_k D^k \;,$$

where k is a d-dimensional multi-index, $a_k \in \mathbf{R}$, and D^k denotes the kth mixed derivative in the distributional sense.

Given a regularity structure (A, T, G), it is convenient to define "abstract" differentiation only on certain subspaces of T. More precisely, we say that a subspace $V \subset T$ is a *sector* if it is invariant under the action of the structure group G and if it can furthermore be written as $V = \bigoplus_{\alpha\in A} V_\alpha$ with $V_\alpha \subset T_\alpha$. We then have the following

Definition 14.1. Let V be a sector of T. A linear operator $\partial\colon V \to T$ is said to *realise* $\mathcal{L}$ (of degree m) for the model (Π, Γ) if

- one has $\partial\tau \in T_{\alpha-m}$ for every $\tau \in V_\alpha$,
- one has $\Gamma\partial\tau = \partial\Gamma\tau$ for every $\tau \in V$.
- one has $\Pi_x\partial\tau = \mathcal{L}\Pi_x\tau$ for every $\tau \in V$ and every $x \in \mathbf{R}^d$.

Writing $\mathscr{D}^\gamma(V)$ for those elements in $\mathscr{D}^\gamma$ taking values in the sector V, it then turns out that one has the following fact:

Proposition 14.2. *Let ∂ be a map that realises $\mathcal{L}$ for the model (Π, Γ) and let $f \in \mathscr{D}^\gamma(V)$ for some $\gamma > m$. Then, $\partial f \in \mathscr{D}^{\gamma-m}$ and the identity $\mathcal{R}\partial f = \mathcal{L}\mathcal{R}f$ holds.*

Proof. The fact that $\partial f \in \mathscr{D}^{\gamma-m}$ is an immediate consequence of the definitions, so we only need to show that $\mathcal{R}\partial f = \mathcal{L}\mathcal{R}f$.

By the "uniqueness" part of the reconstruction theorem, this on the other hand follows immediately if we can show that, for every fixed test function ψ and every $x \in \mathbf{R}^d$, one has

$$\big(\Pi_x\partial f(x) - \mathcal{L}\mathcal{R}f\big)(\psi_x^\lambda) \lesssim \lambda^\delta ,$$

for some $\delta > 0$. Here, we defined ψ_x^λ as before. By the assumption on the model Π, we have the identity

$$\big(\Pi_x\partial f(x) - \mathcal{L}\mathcal{R}f\big)(\psi_x^\lambda) = \big(\partial\Pi_x f(x) - \mathcal{L}\mathcal{R}f\big)(\psi_x^\lambda) = -\big(\Pi_x f(x) - \mathcal{R}f\big)(\mathcal{L}^*\psi_x^\lambda) ,$$

where $\mathcal{L}^*$ is the formal adjoint of $\mathcal{L}$. Since, as a consequence of the homogeneity of $\mathcal{L}$, one has the identity $\mathcal{L}^*\psi_x^\lambda = \lambda^{-m}\big(\mathcal{L}^*\psi\big)_x^\lambda$, it then follows immediately from the reconstruction theorem that the right hand side of this expression is of order $\lambda^{\gamma-m}$, as required. □

14.2 Products and composition by regular functions

One of the main purposes of the theory presented here is to give a robust way to multiply distributions (or functions with distributions) that goes beyond the barrier illustrated by Theorem 13.16. Provided that our functions / distributions are represented as elements in $\mathscr{D}^\gamma$ for some model and regularity structure, we can multiply their "Taylor expansions" pointwise, provided that we give ourselves a table of multiplication on T.

It is natural to consider products with the following properties.

Definition 14.3. Given a regularity structure (T, A, G) and two sectors $V, \bar{V} \subset T$, a *product* on $(V, \bar{V})$ is a bilinear map $\star\colon V \times \bar{V} \to T$ such that, for any $\tau \in V_\alpha$ and $\bar{\tau} \in \bar{V}_\beta$, one has $\tau \star \bar{\tau} \in T_{\alpha+\beta}$ and such that, for any element $\Gamma \in G$, one has $\Gamma(\tau \star \bar{\tau}) = \Gamma\tau \star \Gamma\bar{\tau}$.

Remark 14.4. The condition that homogeneities add up under multiplication is very natural, bearing in mind the case of the polynomial regularity structure. The second condition is also very natural since it merely states that if one reexpands the product of two "polynomials" around a different point, one should obtain the same result as if one reexpands each factor first and then multiplies them together.

Given such a product, we can ask ourselves when the pointwise product of an element $\mathscr{D}^{\gamma_1}$ with an element in $\mathscr{D}^{\gamma_2}$ again belongs to some $\mathscr{D}^{\gamma}$. In order to answer this question, we introduce the notation $\mathscr{D}^{\gamma}_{\alpha}$ to denote those elements $f \in \mathscr{D}^{\gamma}$ such that furthermore

$$f(x) \in T_{\geq\alpha} = \bigoplus_{\beta\geq\alpha} T_\beta \;,$$

for every x. With this notation at hand, it is not hard to show:

Theorem 14.5. *Let $f_1 \in \mathscr{D}^{\gamma_1}_{\alpha_1}(V)$, $f_2 \in \mathscr{D}^{\gamma_2}_{\alpha_2}(\bar{V})$, and let $\star$ be a product on $(V, \bar{V})$. Then, the function f given by $f(x) = f_1(x) \star f_2(x)$ belongs to $\mathscr{D}^{\gamma}_{\alpha}$ with*

$$\alpha = \alpha_1 + \alpha_2 \;, \qquad \gamma = (\gamma_1 + \alpha_2) \wedge (\gamma_2 + \alpha_1) \;. \tag{14.1}$$

Proof. It is clear that $f(x) \in T_{\geq\alpha}$, so it remains to show that it belongs to $\mathscr{D}^{\gamma}$. Furthermore, since we are only interested in showing that $f_1 \star f_2 \in \mathscr{D}^{\gamma}$, we discard all of the components in T_β for $\beta \geq \gamma$.

By the properties of the product $\star$, it remains to obtain a bound of the type

$$\|\Gamma_{xy} f_1(y) \star \Gamma_{xy} f_2(y) - f_1(x) \star f_2(x)\|_\beta \lesssim |x-y|^{\gamma-\beta} \;.$$

By adding and subtracting suitable terms, we obtain

$$\begin{aligned}\|\Gamma_{xy} f(y) - f(x)\|_\beta \leq{}& \|\big(\Gamma_{xy} f_1(y) - f_1(x)\big) \star \big(\Gamma_{xy} f_2(y) - f_2(x)\big)\|_\beta \\ &+ \|\big(\Gamma_{xy} f_1(y) - f_1(x)\big) \star f_2(x)\|_\beta \\ &+ \|f_1(x) \star \big(\Gamma_{xy} f_2(y) - f_2(x)\big)\|_\beta \;.\end{aligned} \tag{14.2}$$

It follows from the properties of the product $\star$ that the first term in (14.2) is bounded by a constant times

$$\begin{aligned}&\sum_{\beta_1+\beta_2=\beta} \|\Gamma_{xy} f_1(y) - f_1(x)\|_{\beta_1} \|\Gamma_{xy} f_2(y) - f_2(x)\|_{\beta_2} \\ &\qquad \lesssim \sum_{\beta_1+\beta_2=\beta} \|x-y\|^{\gamma_1-\beta_1} \|x-y\|^{\gamma_2-\beta_2} \lesssim \|x-y\|^{\gamma_1+\gamma_2-\beta} \;.\end{aligned}$$

Since $\gamma_1 + \gamma_2 \geq \gamma$, this bound is as required. The second term is bounded by a constant times

$$\begin{aligned}\sum_{\beta_1+\beta_2=\beta} \|\Gamma_{xy} f_1(y) - f_1(x)\|_{\beta_1} \|f_2(x)\|_{\beta_2} &\lesssim \sum_{\beta_1+\beta_2=\beta} \|x-y\|^{\gamma_1-\beta_1} \mathbf{1}_{\beta_2\geq\alpha_2} \\ &\lesssim \|x-y\|^{\gamma_1+\alpha_2-\beta} \;,\end{aligned}$$

where the second inequality uses the identity $\beta_1 + \beta_2 = \beta$. Since $\gamma_1 + \alpha_2 \geq \gamma$, this bound is again of the required type. The last term is bounded similarly by reversing the roles played by f_1 and f_2. $\square$

Remark 14.6. It is clear that the formula (14.1) for γ is optimal in general as can be seen from the following two "reality checks". First, consider the case of the polynomial model and take $f_i \in \mathcal{C}^{\gamma_i}$. In this case, the (abstract) truncated Taylor series f_i for f_i belong to $\mathscr{D}_0^{\gamma_i}$. It is clear that in this case, the product cannot be expected to have better regularity than $\gamma_1 \wedge \gamma_2$ in general, which is indeed what (14.1) states. The second reality check comes from (the proof of) Theorem 13.16. In this case, with $\beta > \alpha \geq 0$, one has $f \in \mathscr{D}_0^{\beta}$, while the constant function $x \mapsto \varXi$ belongs to $\mathscr{D}_{-\alpha}^{\infty}$ so that, according to (14.1), one expects their product to belong to $\mathscr{D}_{-\alpha}^{\beta-\alpha}$, which is indeed the case.

It turns out that if we have a product on a regularity structure, then in many cases this also naturally yields a notion of composition with regular functions. Of course, one could in general not expect to be able to compose a regular function with a distribution of negative order. As a matter of fact, we will only define the composition of regular functions with elements in some $\mathscr{D}^{\gamma}$ for which it is guaranteed that the reconstruction operator yields a continuous function. One might think at this case that this would yield a triviality, since we know of course how to compose arbitrary continuous function. The subtlety is that we would like to design our composition operator in such a way that the result is again an element of $\mathscr{D}^{\gamma}$.

For this purpose, we say that a given sector $V \subset T$ is *function-like* if $\alpha < 0 \Rightarrow V_\alpha = 0$ and if V_0 is one-dimensional. (Denote the unit vector of V_0 by $\mathbf{1}$.) We will furthermore always assume that our models are *normal* in the sense that $\big(\Pi_x \mathbf{1}\big)(y) = 1$. In this case, it turns out that if $f \in \mathscr{D}^{\gamma}(V)$, then $\mathcal{R} f$ is a continuous function and one has the identity $\big(\mathcal{R} f\big)(x) = \langle \mathbf{1}, f(x)\rangle$, where we denote by $\langle \mathbf{1}, \cdot\rangle$ the element in the dual of V which picks out the prefactor of $\mathbf{1}$.

Assume now that we are given a regularity structure with a function-like sector V and a product $\star\colon V \times V \to V$. For any smooth function $G\colon \mathbf{R} \to \mathbf{R}$ and any $f \in \mathscr{D}^{\gamma}(V)$ with $\gamma > 0$, we can then *define* $G(f)$ to be the V-valued function given by

$$\big(G \circ f\big)(x) = \sum_{k \geq 0} \frac{G^{(k)}(\bar f(x))}{k!} \tilde f(x)^{\star k}\;,$$

where we have set

$$\bar f(x) = \langle \mathbf{1}, f(x)\rangle\;, \qquad \tilde f(x) = f(x) - \bar f(x)\mathbf{1}\;.$$

Here, $G^{(k)}$ denotes the kth derivative of G and $\tau^{\star k}$ denotes the k-fold product $\tau \star \cdots \star \tau$. We also used the usual conventions $G^{(0)} = G$ and $\tau^{\star 0} = \mathbf{1}$.

Note that as long as G is $\mathcal{C}^{\infty}$, this expression is well-defined. Indeed, by assumption, there exists some $\alpha_0 > 0$ such that $\tilde f(x) \in T_{\geq \alpha_0}$. By the properties of the product, this implies that one has $\tilde f(x)^{\star k} \in T_{\geq k\alpha_0}$. As a consequence, when considering the component of $G \circ f$ in T_β for $\beta < \gamma$, the only terms that give a contribution are those with $k < \gamma/\alpha_0$. Since we cannot possibly hope in general that $G \circ f \in \mathscr{D}^{\gamma'}$ for some $\gamma' > \gamma$, this is all we really need.

It turns out that if G is sufficiently regular, then the map $f \mapsto G \circ f$ enjoys similarly nice continuity properties to what we are used to from classical Hölder spaces. The following result is the analogue in this context to Lemma 7.3:

Proposition 14.7. *In the same setting as above, provided that G is of class $\mathcal{C}^k$ with $k > \gamma/\alpha_0$, the map $f \mapsto G \circ f$ is continuous from $\mathscr{D}^\gamma(V)$ into itself. If $k > \gamma/\alpha_0 + 1$, then it is locally Lipschitz continuous.*

The proof of this result can be found in [Hai14c]. It is somewhat lengthy, but ultimately rather straightforward.

14.3 Schauder estimates and admissible models

One of the reasons why the theory of regularity structures is very successful at providing detailed descriptions of the small-scale features of solutions to semilinear (S)PDEs is that it comes with very sharp Schauder estimates. Recall that the classical Schauder estimates state that if $K \colon \mathbf{R}^d \to \mathbf{R}$ is a kernel that is smooth everywhere, except for a singularity at the origin that is (approximately) homogeneous of degree $\beta - d$ for some $\beta > 0$, then the operator $f \mapsto K * f$ maps $\mathcal{C}^\alpha$ into $\mathcal{C}^{\alpha+\beta}$ for every $\alpha \in \mathbf{R}$, except for those values for which $\alpha + \beta \in \mathbf{N}$. (See for example [Sim97].)

It turns out that similar Schauder estimates hold in the context of general regularity structures in the sense that it is in general possible to build an operator $\mathcal{K} \colon \mathscr{D}^\gamma \to \mathscr{D}^{\gamma+\beta}$ with the property that $\mathcal{R}\mathcal{K}f = K * \mathcal{R}f$. Of course, such a statement can only be true if our regularity structure contains not only the objects necessary to describe $\mathcal{R}f$ up to order γ, but also those required to describe $K * \mathcal{R}f$ up to order $\gamma + \beta$. What are these objects? At this stage, it might be useful to reflect on the effect of the convolution of a singular function (or distribution) with K.

Let us assume for a moment that a given real-valued function f is smooth everywhere, except at some point x_0. It is then straightforward to convince ourselves that $K * f$ is also smooth everywhere, except at x_0. Indeed, for any $\delta > 0$, we can write $K = K_\delta + K_\delta^c$, where K_δ is supported in a ball of radius δ around 0 and K_δ^c is a smooth function. Similarly, we can decompose f as $f = f_\delta + f_\delta^c$, where f_δ is supported in a δ-ball around x_0 and f_δ^c is smooth. Since the convolution of a smooth function with an arbitrary distribution is smooth, it follows that the only non-smooth component of $K * f$ is given by $K_\delta * f_\delta$, which is supported in a ball of radius 2δ around x_0. Since δ was arbitrary, the statement follows. By linearity, this strongly suggests that the local structure of the singularities of $K * f$ can be described completely by only using knowledge on the local structure of the singularities of f. It also suggests that the "singular part" of the operator $\mathcal{K}$ should be local, with the non-local parts of $\mathcal{K}$ only contributing to the "regular part".

This discussion suggests that we certainly need the following ingredients to build an operator $\mathcal{K}$ with the desired properties:

- The canonical polynomial structure should be part of our regularity structure in order to be able to describe the "regular parts".
- We should be given an "abstract integration operator" $\mathcal{I}$ on T which describes how the "singular parts" of $\mathcal{R}f$ transform under convolution by K.
- We should restrict ourselves to models which are "compatible" with the action of $\mathcal{I}$ in the sense that the behaviour of $\Pi_x \mathcal{I}\tau$ should relate in a suitable way to the behaviour of $K * \Pi_x \tau$ near x.

One way to implement these ingredients is to assume first that our model space T contains abstract polynomials in the following sense.

Assumption 14.8 *There exists a sector $\bar{T} \subset T$ isomorphic to the space of abstract polynomials in d commuting variables. In other words, $\bar{T}_\alpha \neq 0$ if and only if $\alpha \in \mathbf{N}$, and one can find basis vectors X^k of $T_{|k|}$ such that every element $\Gamma \in G$ acts on $\bar{T}$ by $\Gamma X^k = (X + h\mathbf{1})^k$ for some $h \in \mathbf{R}^d$.*

Furthermore, we assume that there exists an abstract integration operator $\mathcal{I}$ with the following properties.

Assumption 14.9 *There exists a linear map $\mathcal{I}\colon T \to T$ such that $\mathcal{I}T_\alpha \subset T_{\alpha+\beta}$, such that $\mathcal{I}\bar{T} = 0$, and such that, for every $\Gamma \in G$ and $\tau \in T$, one has*

$$\Gamma \mathcal{I}\tau - \mathcal{I}\Gamma\tau \in \bar{T} \; . \tag{14.3}$$

Finally, we want to consider models that are compatible with this structure for a given kernel K. For this, we first make precise what we mean exactly when we said that K is approximately homogeneous of degree $\beta - d$.

Assumption 14.10 *One can write $K = \sum_{n\geq 0} K_n$ where each of the kernels $K_n\colon \mathbf{R}^d \to \mathbf{R}$ is smooth and compactly supported in a ball of radius 2^{-n} around the origin. Furthermore, we assume that for every multi-index k, one has a constant C such that the bound*

$$\sup_x |D^k K_n(x)| \leq C 2^{n(d-\beta+|k|)} \; , \tag{14.4}$$

holds uniformly in n. Finally, we assume that $\int K_n(x) P(x)\,dx = 0$ for every polynomial P of degree at most N, for some sufficiently large value of N.

Remark 14.11. It turns out that in order to define the operator $\mathcal{K}$ on $\mathscr{D}^\gamma$, we will need K to annihilate polynomials of degree N for some $N \geq \gamma + \beta$.

Remark 14.12. The last assumption may appear to be extremely stringent at first sight. In practice, this turns out not to be a problem at all. Say for example that we want to define an operator that represents convolution with $\mathcal{G}$, the Green's function of the Laplacian. Then, $\mathcal{G}$ can be decomposed into a sum of terms satisfying the bound (14.4) with $\beta = 2$, but it does of course not annihilate generic polynomials and it is not supported in the ball of radius 1.

However, for any fixed value of $N > 0$, it is straightforward to decompose $\mathcal{G}$ as $\mathcal{G} = K + R$, where the kernel K is compactly supported and satisfies all of the properties mentioned above, and the kernel R is smooth. Lifting the convolution with R to an operator from $\mathscr{D}^\gamma \to \mathscr{D}^{\gamma+\beta}$ (actually to $\mathscr{D}^{\bar\gamma}$ for any $\bar\gamma > 0$) is straightforward, so that we have reduced our problem to that of constructing an operator describing the convolution by K.

Given such a kernel K, we can now make precise what we meant earlier when we said that the models under consideration should be compatible with the kernel K.

Definition 14.13. Given a kernel K as in Assumption 14.10 and a regularity structure $\mathscr{T}$ satisfying Assumptions 14.8 and 14.9, we say that a model (Π, Γ) is *admissible* if the identities

$$\big(\Pi_x X^k\big)(y) = (y-x)^k\;, \qquad \Pi_x \mathcal{I}\tau = K * \Pi_x \tau - \Pi_x \mathcal{J}(x)\tau\;, \tag{14.5}$$

holds for every $\tau \in T$ with $|\tau| \le N$. Here, $\mathcal{J}(x)\colon T \to \bar T$ is the linear map given on homogeneous elements by

$$\mathcal{J}(x)\tau = \sum_{|k|<|\tau|+\beta} \frac{X^k}{k!} \int D^{(k)} K(x-y)\,\big(\Pi_x\tau\big)(dy)\;. \tag{14.6}$$

Remark 14.14. Note first that if $\tau \in \bar T$, then the definition given above is coherent as long as $|\tau| < N$. Indeed, since $\mathcal{I}\tau = 0$, one necessarily has $\Pi_x \mathcal{I}\tau = 0$. On the other hand, the properties of K ensure that in this case one also has $K * \Pi_x\tau = 0$, as well as $\mathcal{J}(x)\tau = 0$.

Remark 14.15. While $K * \xi$ is well-defined for any distribution ξ, it is not so clear *a priori* whether the operator $\mathcal{J}(x)$ given in (14.6) is also well-defined. It turns out that the axioms of a model do ensure that this is the case. The correct way of interpreting (14.6) is by

$$\mathcal{J}(x)\tau = \sum_{|k|<|\tau|+\beta} \sum_{n\ge 0} \frac{X^k}{k!} \big(\Pi_x\tau\big)\big(D^{(k)} K_n(x-\cdot)\big)\;.$$

Note now that the scaling properties of the K_n ensure that $2^{(\beta-|k|)n} D^{(k)} K_n(x-\cdot)$ is a test function that is localised around x at scale 2^{-n}. As a consequence, one has

$$\big|\big(\Pi_x\tau\big)\big(D^{(k)} K_n(x-\cdot)\big)\big| \lesssim 2^{(|k|-\beta-|\tau|)n}\;,$$

so that this expression is indeed summable as long as $|k| < |\tau| + \beta$.

Remark 14.16. As a matter of fact, it turns out that the above definition of an admissible model dovetails very nicely with our axioms defining a general model. Indeed, starting from *any* regularity structure $\mathscr{T}$, *any* model (Π, Γ) for $\mathscr{T}$, and a kernel K satisfying Assumption 14.10, it is usually possible to build a larger regularity structure $\hat{\mathscr{T}}$ containing $\mathscr{T}$ (in the "obvious" sense that $T \subset \hat T$ and the action of $\hat G$ on T is compatible with that of G) and endowed with an abstract integration map $\mathcal{I}$, as well as an admissible model $(\hat{\Pi}, \hat{\Gamma})$ on $\hat{\mathscr{T}}$ which reduces to (Π, Γ) when restricted to T. See [Hai14c] for more details.

The only exception to this rule arises when the original structure T contains some homogeneous element τ which does not represent a polynomial and which is such that $|\tau| + \beta \in \mathbf{N}$. Since the bounds appearing both in the definition of a model and in Assumption 14.10 are only upper bounds, it is in practice easy to exclude such a situation by slightly tweaking the definition of either the exponent β or of the original regularity structure $\mathscr{T}$.

With all of these definitions in place, we can finally build the operator $\mathcal{K} \colon \mathscr{D}^\gamma \to \mathscr{D}^{\gamma+\beta}$ announced at the beginning of this section. Recalling the definition of $\mathcal{J}$ from (14.6), we set

$$\big(\mathcal{K} f\big)(x) = \mathcal{I} f(x) + \mathcal{J}(x) f(x) + \big(\mathcal{N} f\big)(x) \;, \tag{14.7}$$

where the operator $\mathcal{N}$ is given by

$$\big(\mathcal{N} f\big)(x) = \sum_{|k|<\gamma+\beta} \frac{X^k}{k!} \int D^{(k)} K(x-y) \, \big(\mathcal{R} f - \Pi_x f(x)\big)(dy) \;. \tag{14.8}$$

Note first that thanks to the reconstruction theorem, it is possible to verify that the right hand side of (14.8) does indeed make sense for every $f \in \mathscr{D}^\gamma$ in virtually the same way as in Remark 14.15. One has:

Theorem 14.17. *Let K be a kernel satisfying Assumption 14.10, let $\mathscr{T} = (A, T, G)$ be a regularity structure satisfying Assumptions 14.8 and 14.9, and let (Π, Γ) be an admissible model for $\mathscr{T}$. Then, for every $f \in \mathscr{D}^\gamma$ with $\gamma \in (0, N-\beta)$ and $\gamma+\beta \notin \mathbf{N}$, the function $\mathcal{K} f$ defined in (14.7) belongs to $\mathscr{D}^{\gamma+\beta}$ and satisfies $\mathcal{R}\mathcal{K} f = K * \mathcal{R} f$.*

Proof. The complete proof of this result can be found in [Hai14c] and will not be given here. Let us simply show that one has indeed $\mathcal{R}\mathcal{K} f = K * \mathcal{R} f$ in the particular case when our model consists of continuous functions so that Remark 13.29 applies. In this case, one has

$$\big(\mathcal{R}\mathcal{K} f\big)(x) = \big(\Pi_x(\mathcal{I} f(x) + \mathcal{J}(x) f(x))\big)(x) + \big(\Pi_x \big(\mathcal{N} f\big)(x)\big)(x) \;.$$

As a consequence of (14.5), the first term appearing in the right hand side of this expression is given by

$$\big(\Pi_x(\mathcal{I} f(x) + \mathcal{J}(x) f(x))\big)(x) = \big(K * \Pi_x f(x)\big)(x) \;.$$

On the other hand, the only term contributing to the second term is the one with $k = 0$ (which is always present since $\gamma > 0$ by assumption) which then yields

$$\big(\Pi_x \big(\mathcal{N} f\big)(x)\big)(x) = \int K(x-y) \, \big(\mathcal{R} f - \Pi_x f(x)\big)(dy) \;.$$

Adding both of these terms, we see that the expression $\big(K * \Pi_x f(x)\big)(x)$ cancels, leaving us with the desired result. $\square$

We are now in principle in possession of all of the ingredients required to formulate a large number of semilinear stochastic PDEs: multiplication, composition by regular functions, differentiation, and integration against the Green's function of the linearised equation.

In the next chapter we show how this can be leveraged in practice in order to build a robust solution theory for the KPZ equation.

14.4 Exercises

Exercise 14.18. a) Construct an example of a regularity structure with trivial group G in which both $\mathcal{R}f_1$ and $\mathcal{R}f_2$ are continuous functions but where the identity

$$\mathcal{R}(f_1 \star f_2)(x) = (\mathcal{R}f_1)(x)\,(\mathcal{R}f_2)(x)$$

fails.

b) Transfer Exercise 2.17 to the present context.

Solution 14.19. (We only address the first part.) Consider for instance the regularity structure given by $A = (-2\kappa, -\kappa, 0)$ for fixed $\kappa > 0$ with each T_α being a copy of **R** given by $T_{-n\kappa} = \langle \Xi^n \rangle$. We furthermore take for G the trivial group. This regularity structure comes with an obvious product by setting $\Xi^m \star \Xi^n = \Xi^{m+n}$ provided that $m + n \le 2$.

Then, we could for example take as a model for $\mathscr{T} = (T, A, G)$:

$$\big(\Pi_x \Xi^0\big)(y) = 1\;, \quad \big(\Pi_x \Xi\big)(y) = 0\;, \quad \big(\Pi_x \Xi^2\big)(y) = c\;, \tag{14.9}$$

where c is an arbitrary constant. Let furthermore

$$f_1(x) = f_1(x)\Xi^0 + \tilde f_1(x)\Xi\;, \qquad f_2(x) = f_2(x)\Xi^0 + \tilde f_2(x)\Xi\;.$$

Since our group G is trivial, one has $f_i \in \mathscr{D}^\gamma$ provided that each of the f_i belongs to $\mathcal{C}^\gamma$ and each of the $\tilde f_i$ belongs to $\mathcal{C}^{\gamma+\kappa}$. (And one has $\gamma + \kappa < 1$.) One furthermore has the identity $\big(\mathcal{R}f_i\big)(x) = f_i(x)$.

However, the pointwise product is given by

$$\big(f_1 \star f_2\big)(x) = f_1(x)f_2(x)\Xi^0 + \big(\tilde f_1(x)f_2(x) + \tilde f_2(x)f_1(x)\big)\Xi + \tilde f_1(x)\tilde f_2(x)\Xi^2\;,$$

which by Theorem 14.5 belongs to $\mathscr{D}^{\gamma-\kappa}$. Provided that $\gamma > \kappa$, one can then apply the reconstruction operator to this product and one obtains

$$\mathcal{R}\big(f_1 \star f_2\big)(x) = f_1(x)f_2(x) + c\tilde f_1(x)\tilde f_2(x)\;,$$

which is obviously quite different from the pointwise product $(\mathcal{R}f_1)(x) \cdot (\mathcal{R}f_2)(x)$.

How should this be interpreted? For $n > 0$, we could have defined a model $\Pi^{(n)}$ by

$$\big(\Pi_x \Xi^0\big)(y) = 1\;, \quad \big(\Pi_x \Xi\big)(y) = \sqrt{2c}\sin(ny)\;, \quad \big(\Pi_x \Xi^2\big)(y) = 2c\sin^2(ny)\;.$$

Denoting by $\mathcal{R}^{(n)}$ the corresponding reconstruction operator, we have the identity

$$\big(\mathcal{R}^{(n)} f_i\big)(x) = f_i(x) + \sqrt{2c}\tilde{f}_i(x)\sin(nx)\;,$$

as well as $\mathcal{R}^{(n)}(f_1 \star f_2) = \mathcal{R}^{(n)} f_1 \cdot \mathcal{R}^{(n)} f_2$. As a model, the model $\Pi^{(n)}$ actually converges to the limiting model Π defined in (14.9). As a consequence of the continuity of the reconstruction operator, this implies that

$$\mathcal{R}^{(n)} f_1 \cdot \mathcal{R}^{(n)} f_2 = \mathcal{R}^{(n)}(f_1 \star f_2) \to \mathcal{R}(f_1 \star f_2) \neq \mathcal{R} f_1 \cdot \mathcal{R} f_2\;,$$

which is of course also easy to see "by hand". This shows that in some cases, the "non-canonical" models as in (14.9) can be interpreted as limits of "canonical" models for which the usual rules of calculus hold. Even this is however not always the case (think of the Itô Brownian rough path).

Exercise 14.20. Consider space-time $\mathbf{R}^d$ with one temporal and $(d-1)$ spatial dimensions, under the parabolic scaling $(2, 1, \ldots, 1)$, as introduced in Remark 13.9. Denote by $\mathcal{G}$ the heat kernel (i.e. the Green's function of the operator $\partial_t - \partial_x^2$). Show that one has the decomposition

$$\mathcal{G} = K + \hat{K}\;,$$

where the kernel K satisfies all of the assumptions of Section 14.3 (with $\beta = 2$) and the remainder $\hat{K}$ is smooth and bounded.

Chapter 15
Application to the KPZ equation

Abstract We show how the theory of regularity structures can be used to build a robust solution theory for the KPZ equation. We also give a very short survey of the original approach to the same problem using controlled rough paths and we discuss how the two approaches are linked.

15.1 Formulation of the main result

Let us now briefly explain how the theory of regularity structures can be used to make sense of solutions to very singular semilinear stochastic PDEs. We will keep the discussion in this section at a very informal level without attempting to make mathematically precise statements. The interested reader may find more details in [Hai13, Hai14c].

For definiteness, we focus on the case of the KPZ equation [KPZ86], which is formally given by

$$\partial_t h = \partial_x^2 h + (\partial_x h)^2 + \xi - C \;, \tag{15.1}$$

where ξ denotes space-time white noise, the spatial variable takes values in the circle (i.e. in the interval $[0, 2\pi]$ endowed with periodic boundary conditions), and C is a fixed constant. The problem with such an equation is that even the solution to the linear part of the equation, namely

$$\partial_t \Psi = \partial_x^2 \Psi + \xi \;,$$

is not differentiable as a function of the spatial variable. As a matter of fact, as already noted in Section 12.2, for any fixed time t, Ψ has the regularity of Brownian motion as a function of the spatial variable x. As a consequence, it turns out that the only way of giving meaning to (15.1) is to "renormalise" the equation by subtracting to its right hand side an "infinite constant", which counteracts the divergence of the term $(\partial_x h)^2$.

This has usually been interpreted in the following way. Assuming for a moment that ξ is a smooth function, a simple consequence of the change of variables formula shows that if we define $h = \log Z$, then Z satisfies the PDE

$$\partial_t Z = \partial_x^2 Z + Z\,\xi\;.$$

The only ill-posed product appearing in this equation now is the product of the solution Z with white noise ξ. As long as Z takes values in L^2, this product can be given a meaning as a classical Itô integral, so that the equation for Z can be interpreted as the Itô equation

$$dZ = \partial_x^2 Z\,dt + Z\,dW\;, \tag{15.2}$$

were W is an L^2-cylindrical Wiener process. It is well known [DPZ92] that this equation has a unique (mild) solution and we can then go backwards and *define* the solution to the KPZ equation as $h = \log Z$. The expert reader will have noticed that this argument is flawed: since (15.2) is interpreted as an Itô equation, we should really use Itô's formula to find out what equation h satisfies. If one does this a bit more carefully, one notices that the Itô correction term appearing in this way is indeed an infinite constant! This is the case in the following sense. If W_ε is a Wiener process with a covariance given by $x \mapsto \varepsilon^{-1}\varrho(\varepsilon^{-1}x)$ for some smooth compactly supported function ϱ integrating to 1 and Z_ε solves (15.2) with W replaced by W_ε, then $h_\varepsilon = \log Z_\varepsilon$ solves

$$dh = \partial_x^2 h\,dt + (\partial_x h)^2\,dt + dW_\varepsilon - \varepsilon^{-1}C_\varrho\,dt\;, \tag{15.3}$$

for some constant C_ϱ depending on ϱ. Since Z_ε converges to a strictly positive limit Z, this shows that the sequence of functions h_ε solving (15.3) converges to a limit h. This limit is called the Hopf–Cole solution to the KPZ equation [Hop50, Col51, BG97].

This notion of solution is of course not very satisfactory since it relies on a nonlinear transformation and provides no direct interpretation of the term $(\partial_x h)^2$ appearing in the right hand side of (15.1). Furthermore, many natural growth models lead to equations that structurally "look like" (15.1), rather than (15.2). Since perturbations are usually rather badly behaved under exponentiation and since there is no really good approximation theory for (15.2) either (for example it has been an open problem whether space-time regularisations of the noise lead to the same notion of solution), one would like to have a robust solution theory for (15.1) directly.

Such a robust solution theory is precisely what the theory of regularity structures provides. More precisely, it provides spaces $\mathscr{M}$ (a suitable space of "admissible models") and $\mathscr{D}^\gamma$, maps $\mathcal{S}_a$ (an abstract "solution map"), $\mathcal{R}$ (the reconstruction operator) and Ψ (a "canonical lift map"), as well as a finite-dimensional group $\mathfrak{R}$ acting both on $\mathbf{R}$ and $\mathscr{M}$ such that the following diagram commutes:

$$
\begin{tikzcd}
C \in \mathcal{F} \arrow[d, leftrightarrow] & \times & \mathscr{M} \arrow[loop above, "\mathfrak{R}"] & \times & \mathcal{C}^\alpha \arrow[r, "\mathcal{S}_a"] \arrow[d, leftrightarrow] \arrow[dr, phantom, "."] & \mathscr{D}^\gamma \arrow[d, "\mathcal{R}"] \\
\mathcal{F} \arrow[loop below, "\mathfrak{R}"] & \times & \mathcal{C} \arrow[u, "\Psi"] & \times & \mathcal{C}^\alpha \arrow[r, "\mathcal{S}_c"] & \mathcal{C}([0,T],\mathcal{C}^\alpha) \\
& & \rotatebox[origin=c]{90}{$\in$} & & \rotatebox[origin=c]{90}{$\in$} & \rotatebox[origin=c]{90}{$\in$} \\
& & \xi & & h_0 & h
\end{tikzcd}
\tag{15.4}
$$

Here, $\mathcal{S}_c$ denotes the classical solution map $\mathcal{S}_c(C,\xi,h_0)$ which provides the solution (up to some fixed final time T) to the equation

$$\partial_t h = \partial_x^2 h + (\partial_x h)^2 + \xi - C\;, \quad h(0,x) = h_0(x)\;, \tag{15.5}$$

for regular instances of the noise ξ. The space $\mathcal{F}$ of "formal right hand sides" is in this case just a copy of $\mathbf{R}$ which holds the value of the constant C appearing in (15.5). The diagram commutes in the sense that if $M \in \mathfrak{R}$, then

$$\mathcal{S}_c(M(C),\xi,h_0) = \mathcal{R}\mathcal{S}_a(C, M(\Psi(\xi)), h_0)\;,$$

where we identify M with its respective actions on $\mathbf{R}$ and $\mathscr{M}$. The important additional features are the following:

- If ξ_ε denotes a "natural" regularisation of space-time white noise, then there exists a sequence M_ε of elements in $\mathfrak{R}$ such that $M_\varepsilon\Psi(\xi_\varepsilon)$ converges to a limiting random element $(\Pi,\Gamma) \in \mathscr{M}$. This element can also be characterised directly without resorting to specific approximation procedures and $\mathcal{R}\mathcal{S}_a(0,(\Pi,\Gamma),h_0)$ coincides almost surely with the Hopf–Cole solution to the KPZ equation.
- The maps $\mathcal{S}_a$ and $\mathcal{R}$ are both continuous, unlike the map $\mathcal{S}_c$ which is discontinuous in its second argument for any topology for which ξ_ε converges to ξ.
- As an abstract group, the "renormalisation group" $\mathfrak{R}$ is simply equal to $(\mathbf{R}^3,+)$. However, it is possible to extend the picture to deal with much larger classes of approximations, which has the effect of increasing both $\mathfrak{R}$ and the space $\mathcal{F}$ of possible right hand sides. See for example [HQ14] for a proof of convergence to KPZ for a much larger class of interface growth models.

An example of statement that can be proved from these considerations (see [Hai13, Hai14c, HQ14]) is the following.

Theorem 15.1. *Consider the sequence of equations*

$$\partial_t h_\varepsilon = \partial_x^2 h_\varepsilon + (\partial_x h_\varepsilon)^2 + \xi_\varepsilon - C_\varepsilon\;, \tag{15.6}$$

where $\xi_\varepsilon = \delta_\varepsilon * \xi$ with $\delta_\varepsilon(t,x) = \varepsilon^{-3}\varrho(\varepsilon^{-2}t, \varepsilon^{-1}x)$, for some smooth and compactly supported function ϱ, and ξ denotes space-time white noise. Then, there exists a *(diverging) choice of constants C_ε such that the sequence h_ε converges in probability to a limiting process h.*

Furthermore, one can ensure that the limiting process h does not depend on the choice of mollifier ϱ and that it coincides with the Hopf–Cole solution to the KPZ equation.

Remark 15.2. It is important to note that although the limiting process is independent of the choice of mollifier ϱ, the constant C_ε *does* very much depend on this choice, as we already alluded to earlier.

Remark 15.3. Regarding the initial condition, one can take $h_0 \in \mathcal{C}^\beta$ for any fixed $\beta > 0$. Unfortunately, this result does not cover the case of "infinite wedge" initial conditions, see for example [Cor12].

The aim of this section is to sketch how the theory of regularity structures can be used to obtain this kind of convergence results and how (15.4) is constructed. First of all, we note that while our solution h will be a Hölder continuous space-time function (or rather an element of $\mathscr{D}^\gamma$ for some regularity structure with a model over $\mathbf{R}^2$), the "time" direction has a different scaling behaviour from the three "space" directions. As a consequence, it turns out to be effective to slightly change our definition of "localised test functions" by setting

$$\varphi_{(s,x)}^\lambda(t,y) = \lambda^{-3}\varphi\big(\lambda^{-2}(t-s), \lambda^{-1}(y-x)\big)\;.$$

Accordingly, the "effective dimension" of our space-time is actually 3, rather than 2. The theory presented in Section 13 extends *mutatis mutandis* to this setting. (Note however that when considering the homogeneity of a regular monomial, powers of the time variable should now be counted double.) Note also that with this way of measuring regularity, space-time white noise belongs to $\mathcal{C}^{-\alpha}$ for every $\alpha > \frac{3}{2}$. This is because of the bound

$$\big(\mathbf{E}\langle \xi, \varphi_x^\lambda\rangle^2\big)^{1/2} = \|\varphi_x^\lambda\|_{L^2} \approx \lambda^{-\frac{3}{2}}\;,$$

combined with an argument somewhat similar to the proof of Kolmogorov's continuity lemma.

15.2 Construction of the associated regularity structure

Our first step is to build a regularity structure that is sufficiently large to allow to reformulate (15.1) as a fixed point in $\mathscr{D}^\gamma$ for some $\gamma > 0$. Denoting by $\mathcal{G}$ the heat kernel (i.e. the Green's function of the operator $\partial_t - \partial_x^2$), we can rewrite the solution to (15.1) with initial condition h_0 as

$$h = \mathcal{G} * \big((\partial_x h)^2 + \xi\big) + \mathcal{G}h_0\;,$$

where $*$ denotes space-time convolution and where we denote by $\mathcal{G}h_0$ the harmonic extension of h_0. (That is the solution to the heat equation with initial condition h_0.) In order to have a chance of fitting this into the framework described above, we first decompose the heat kernel $\mathcal{G}$ as in Exercise 14.20 as

$$\mathcal{G} = K + \hat{K} \;,$$

where the kernel K satisfies all of the assumptions of Section 14.3 (with $\beta = 2$) and the remainder $\hat{K}$ is smooth. If we consider any regularity structure containing the usual Taylor polynomials and equipped with an admissible model, is straightforward to associate to $\hat{K}$ an operator $\hat{\mathcal{K}}\colon \mathscr{D}^\gamma \to \mathscr{D}^\infty$ via

$$\bigl(\hat{\mathcal{K}} f\bigr)(z) = \sum_k \frac{X^k}{k!} \bigl(D^{(k)} \hat{K} * \mathcal{R} f\bigr)(z) \;,$$

where z denotes a space-time point and k runs over all possible 2-dimensional multiindices. Similarly, the harmonic extension of h_0 can be lifted to an element in $\mathscr{D}^\infty$ which we denote again by $\mathcal{G}h_0$ by considering its Taylor expansion around every space-time point. At this stage, we note that we actually cheated a little: while $\mathcal{G}h_0$ is smooth in $\{(t,x) \,:\, t > 0, x \in S^1\}$ and vanishes when $t < 0$, it is of course singular on the time-0 hyperplane $\{(0,x) \,:\, x \in S^1\}$. This problem can be cured by introducing weighted versions of the spaces $\mathscr{D}^\gamma$ allowing for singularities on a given hyperplane. A precise definition of these spaces and their behaviour under multiplication and the action of the integral operator $\mathcal{K}$ can be found in [Hai14c]. For the purpose of the informal discussion given here, we will simply ignore this problem.

This suggests that the "abstract" formulation of (15.1) should be given by

$$H = \mathcal{K}\bigl((\partial H)^2 + \Xi\bigr) + \hat{\mathcal{K}}\bigl((\partial H)^2 + \Xi\bigr) + \mathcal{G}h_0 \;, \tag{15.7}$$

where it still remains to be seen how to define an "abstract differentiation operator" ∂ realising the spatial derivative ∂_x as in Section 14.1. In view of (14.7), this equation is of the type

$$H = \mathcal{I}\bigl((\partial H)^2 + \Xi\bigr) + (\ldots) \;, \tag{15.8}$$

where the terms $(\ldots)$ consist of functions that take values in the subspace $\bar{T}$ of T spanned by regular Taylor polynomials in the time variable X_0 and the space variable X_1. (As previously, X denotes the collection of both.) In order to build a regularity structure in which (15.8) can be formulated, it is then natural to start with the structure $\bar{T}$ given by these abstract polynomials (again with the parabolic scaling which causes the abstract "time" variable to have homogeneity 2 rather than 1), and to then add a symbol Ξ to it which we postulate to have homogeneity $-\frac{3}{2}^-$, where we denote by α^- an exponent strictly smaller than, but arbitrarily close to, the value α. As a consequence of our definitions, it will also turn out that the symbol ∂ is always immediately followed by the symbol $\mathcal{I}$, so that it makes sense to introduce the shorthand $\mathcal{I}' = \partial\mathcal{I}$. This is also suggestive of the fact that $\mathcal{I}'$ can itself be considered an abstract integration map, associated to the kernel $K' = \partial_x K$.

We then simply add to T all of the formal expressions that an application of the right hand side of (15.8) can generate for the description of H, ∂H, and $(\partial H)^2$. The homogeneity of a given expression is furthermore completely determined by the rules $|\mathcal{I}\tau| = |\tau| + 2$, $|\partial\tau| = |\tau| - 1$ and $|\tau\bar\tau| = |\tau| + |\bar\tau|$. For example, it follows from (15.8) that the symbol $\mathcal{I}(\Xi)$ is required for the description of H, so that $\mathcal{I}'(\Xi)$ is required for the description of ∂H. This then implies that $\mathcal{I}'(\Xi)^2$ is required for the description of the right hand side of (15.8), which in turn implies that $\mathcal{I}(\mathcal{I}'(\Xi)^2)$ is also required for the description of H, etc.

Remark 15.4. Here we made a distinction between $\mathcal{I}(\Xi)$, interpreted as the linear map $\mathcal{I}$ applied to the symbol Ξ, and the symbol $\mathcal{I}(\Xi)$. Since the map $\mathcal{I}$ is then defined by $\mathcal{I}(\Xi) = \mathcal{I}(\Xi)$, this distinction is somewhat moot and will be blurred in the sequel.

More formally, denote by $\mathcal{U}$ the collection of those formal expressions that are required to describe H. This is then defined as the smallest collection containing $\mathbf{1}$, X, and $\mathcal{I}(\Xi)$, and such that

$$\tau_1, \tau_2 \in \mathcal{U} \quad \Rightarrow \quad \mathcal{I}(\partial\tau_1\partial\tau_2) \in \mathcal{U}\;,$$

where it is understood that $\mathcal{I}(X^k) = 0$ for every multi-index k. We then set

$$\mathcal{W} = \mathcal{U} \cup \{\Xi\} \cup \{\partial\tau_1\partial\tau_2 \,:\, \tau_i \in \mathcal{U}\}\;, \tag{15.9}$$

and we define our space T as the set of all linear combinations of elements in $\mathcal{W}$. Naturally, T_α consists of those linear combinations that only involve elements in $\mathcal{W}$ that are of homogeneity α. It is not too difficult to convince oneself that, for every $\alpha \in \mathbf{R}$, $\mathcal{W}$ contains only finitely many elements of homogeneity less than α, so that each T_α is finite-dimensional.

In order to simplify expressions later, we will use the following shorthand graphical notation for elements of $\mathcal{W}$. For Ξ, we draw a small circle. The integration map $\mathcal{I}$ is then represented by a downfacing wavy line and $\mathcal{I}'$ is represented by a downfacing plain line. The multiplication of symbols is obtained by joining them at the root. For example, we have

$$\mathcal{I}'(\Xi)^2 = \;, \qquad (\mathcal{I}'(\mathcal{I}'(\Xi)^2))^2 = \;, \qquad \mathcal{I}(\mathcal{I}'(\Xi)^2) = \;.$$

Symbols containing factors of X have no particular graphical representation, so we will for example write $X_i\mathcal{I}'(\Xi)^2 = X_i$. With this notation, the space T (up to homogeneity $\frac{3}{2}$) is given by

$$T = \langle \Xi, \ldots, \mathbf{1}, \ldots, X_1, \ldots \rangle\;, \tag{15.10}$$

where we ordered symbols in increasing order of homogeneity and used $\langle\cdot\rangle$ to denote the linear span.

Exercise 15.5. Compute the homogeneities of the symbols appearing in the list (15.10).

15.3 The structure group

Recall that the purpose of the group G is to provide a class of linear maps $\Gamma\colon T \to T$ arising as possible candidates for the action of "reexpanding" a "Taylor series" around a different point. In our case, in view of (14.5), the coefficients of these reexpansions will naturally be some polynomials in x and in the expressions appearing in (14.6). This suggests that we should define a space T^+ whose basis vectors consist of formal expressions of the type

$$X^k \prod_{i=1}^{N} \mathcal{J}_{\ell_i}(\tau_i) \;, \tag{15.11}$$

where N is an arbitrary but finite number, the τ_i are canonical basis elements in $\mathcal{W}$ defined in (15.9), and the ℓ_i are d-dimensional multiindices satisfying $|\ell_i| < |\tau_i| + 2$. (The last bound is a reflection of the restriction of the summands in (14.6) with $\beta = 2$.) The space T^+ is endowed with a natural commutative product. (In fact, T^+ is nothing but the free commutative algebra over the symbols $\{X_i, \mathcal{J}_\ell(\tau)\}$ with $i \in \{1, \ldots, d\}$ and $\tau \in \mathcal{W}$ with $|\tau| < |\ell|$.)

Remark 15.6. While the canonical basis of T^+ is related to that of T, it should be viewed as a completely disjoint space. We emphasise this by not colouring the basis vectors of T^+.

The space T^+ also has a natural graded structure $T^+ = \bigoplus T^+_\alpha$ similarly to before by setting

$$|\mathcal{J}_\ell(\tau)| = |\tau| + 2 - |\ell| \;, \qquad |X^k| = |k| \;,$$

and by postulating that the degree of a product is the sum of the degrees of its factors. Unlike in the case of T however, elements of T^+ all have strictly positive homogeneity, except for the empty product $\mathbf{1}$ which we postulate to have homogeneity 0.

To any given admissible model (Π, Γ), it is then natural to associate linear maps $f_x\colon T^+ \to \mathbf{R}$ by setting $f_x(X^k) = (-x)^k$, $f_x(\sigma\bar\sigma) = f_x(\sigma) f_x(\bar\sigma)$, and

$$f_x(\mathcal{J}_{\ell_i}(\tau_i)) = \int D^{(\ell_i)} K(x-y) \big(\Pi_x \tau_i\big)(dy) \;. \tag{15.12}$$

It turns out that with this definition, the coefficients of the linear maps Γ_{xy} can be expressed as polynomials of the numbers $f_x(\mathcal{J}_{\ell_i}(\tau_i))$ and $f_y(\mathcal{J}_{\ell_i}(\tau_i))$ for suitable expressions τ_i and multiindices ℓ_i. In order to formalize this, we consider the following construction. We define a linear map $\Delta\colon T \to T \otimes T^+$ in the following way. For the basic elements Ξ, $\mathbf{1}$ and X_i ($i \in \{0, 1\}$), we set

$$\Delta \mathbf{1} = \mathbf{1} \otimes \mathbf{1} , \quad \Delta \Xi = \Xi \otimes \mathbf{1} , \quad \Delta X_i = X_i \otimes \mathbf{1} + \mathbf{1} \otimes X_i .$$

We then extend this recursively to all of T by imposing the following identities

$$\begin{aligned} \Delta(\tau\bar\tau) &= \Delta\tau \cdot \Delta\bar\tau , \\ \Delta\mathcal{I}(\tau) &= (\mathcal{I} \otimes I)\Delta\tau + \sum_{\ell,m} \frac{X^\ell}{\ell!} \otimes \frac{X^m}{m!} \mathcal{J}_{\ell+m}(\tau) , \\ \Delta\mathcal{I}'(\tau) &= (\mathcal{I}' \otimes I)\Delta\tau + \sum_{\ell,m} \frac{X^\ell}{\ell!} \otimes \frac{X^m}{m!} \mathcal{J}_{\ell+m+(0,1)}(\tau) . \end{aligned}$$

Here, we extend $\tau \mapsto \mathcal{J}_k(\tau) \stackrel{\text{def}}{=} \mathcal{J}_k(\tau)$ to a linear map $\mathcal{J}_k \colon T \to T^+$ by setting $\mathcal{J}_k(\tau) = 0$ for those basis vectors $\tau \in \mathcal{W}$ for which $|\tau| < |k| - 2$.

Let now G_+ denote the set of all linear maps $g \colon T^+ \to \mathbf{R}$ with the property that $g(\sigma\bar\sigma) = g(\sigma)g(\bar\sigma)$ for any two elements σ and $\bar\sigma$ in T^+. Then, to any such map, we can associate a linear map $\Gamma_g \colon T \to T$ by

$$\Gamma_g \tau = (I \otimes g)\Delta\tau . \tag{15.13}$$

In principle, this definition makes sense for every $g \in (T^+)^*$. However, it turns out that the set of such maps with $g \in G_+$ forms a group, which is our structure group G.

Furthermore, there exists a linear map $\Delta^+ \colon T^+ \to T^+ \otimes T^+$ such that

$$(\Delta \otimes I)\Delta = (I \otimes \Delta^+)\Delta , \qquad \Delta^+(\sigma\bar\sigma) = \Delta^+\sigma \cdot \Delta^+\bar\sigma . \tag{15.14}$$

With this map at hand, we can define a product $\circ$ on the dual of T^+ by

$$(f \circ g)(\sigma) = (f \otimes g)\Delta^+\sigma .$$

This has the property that $\Gamma_{f\circ g} = \Gamma_f \circ \Gamma_g$, with the symbol $\circ$ on the right denoting the composition of linear maps as usual. The second identity of (15.14) furthermore ensures that if f and g belong to G_+, then $f \circ g \in G_+$. It also turns out that every $f \in G_+$ admits a unique inverse f^{-1} such that $f^{-1} \circ f = f \circ f^{-1} = e$, where $e \colon T^+ \to \mathbf{R}$ maps every basis vector of the form (15.11) to zero, except for $e(\mathbf{1}) = 1$. The element e is neutral in the sense that Γ_e is the identity operator.

It is a highly non-trivial fact [Hai14c, Sec. 8] that if Π comes from an admissible model as in Definition 14.13 and we define $F_x \colon T \to T$ by

$$F_x = \Gamma_{f_x} ,$$

with f_x given by (15.12), then $\Pi_z F_z^{-1}$ is independent of the space-time point z. In particular, for any admissible model, it turns out that Γ is determined by Π through the identity

$$\Gamma_{xy} = F_x^{-1} F_y = \Gamma_{\gamma_{xy}} , \qquad \gamma_{xy} = f_x^{-1} \circ f_y .$$

While this is a very nice coherent algebraic framework, it begs the question whether in general there do even exist any non-trivial admissible models. This is a valid question since the analytic bounds and algebraic identities that any admissible model should satisfy are extremely stringent. The next section shows that fortunately there exists a very rich class of admissible models.

15.4 Canonical lifts of regular functions

Given any sufficiently regular function ξ (say a continuous space-time function), there is then a canonical way of lifting ξ to a model $\iota\xi = (\Pi, \Gamma)$ for T by setting

$$\big(\Pi_z \Xi\big)(\bar z) = \xi(\bar z)\;, \qquad \big(\Pi_z X^k\big)(\bar z) = (\bar z - z)^k\;,$$

and then recursively by

$$\big(\Pi_z \tau\bar\tau\big)(\bar z) = \big(\Pi_z \tau\big)(\bar z)\cdot\big(\Pi_z \bar\tau\big)(\bar z)\;, \tag{15.15}$$

as well as (14.5). Here we used z and $\bar z$ as notations for generic space-time points in order not to overload the notations. The maps Γ_{xy} are then determined from Π by the discussion in the previous subsection.

With such a model $\iota\xi$ at hand, it follows from (15.15), (13.25), and the admissibility of $\iota\xi$ that the associated reconstruction operator satisfies the properties

$$\mathcal{R}\mathcal{K} f = K * \mathcal{R} f\;, \qquad \mathcal{R}(fg) = \mathcal{R} f\cdot\mathcal{R} g\;,$$

as long as all the functions to which $\mathcal{R}$ is applied belong to $\mathscr{D}^\gamma$ for some $\gamma > 0$. As a consequence, applying the reconstruction operator $\mathcal{R}$ to both sides of (15.7), we see that if H solves (15.7) then, provided that the model $(\Pi, \Gamma) = \iota\xi$ was built as above starting from any *continuous* realisation ξ of the driving noise, the function $h = \mathcal{R}H$ solves the equation (15.1).

At this stage, the situation is as follows. For any *continuous* realisation ξ of the driving noise, we have factorized the solution map $(h_0, \xi) \mapsto h$ associated to (15.1) into maps

$$(h_0, \xi) \mapsto (h_0, \iota\xi) \mapsto H \mapsto h = \mathcal{R}H\;,$$

where the middle arrow corresponds to the solution to (15.7) in some weighted $\mathscr{D}^\gamma$-space. The advantage of such a factorisation is that the last two arrows yield *continuous* maps, even in topologies sufficiently weak to be able to describe driving noise having the lack of regularity of space-time white noise. The only arrow that isn't continuous in such a weak topology is the first one. At this stage, it should be believable that a similar construction can be performed for a very large class of semilinear stochastic PDEs, provided that certain scaling properties are satisfied. This is indeed the case and large parts of this programme have been carried out in [Hai14c].

Given this construction, one is lead naturally to the following question: given a sequence ξ_ε of "natural" regularisations of space-time white noise, for example as in (15.6), do the lifts $\iota\xi_\varepsilon$ converge in probably in a suitable space of admissible models? Unfortunately, unlike in the theory of rough paths where this is very often the case (see Section 10), the answer to this question in the context of SPDEs is often an emphatic **no**. Indeed, if it were the case for the KPZ equation, then one could have been able to choose the constant C_ε to be independent of ε in (15.6), which is certainly not the case.

15.5 Renormalisation of the KPZ equation

One way of circumventing the fact that $\iota\xi_\varepsilon$ does not converge to a limiting model as $\varepsilon \to 0$ is to consider instead a sequence of *renormalised* models. The main idea is to exploit the fact that our abstract definitions of a model do not impose the identity (15.15), even in situations where ξ itself happens to be a continuous function. One question that then imposes itself is: what are the natural ways of "deforming" the usual product which still lead to lifts to an admissible model? It turns out that the regularity structure whose construction was sketched above comes equipped with a natural *finite-dimensional* group of continuous transformations $\mathfrak{R}$ on its space of admissible models (henceforth called the "renormalisation group"), which essentially amounts to the space of all natural deformations of the product. It then turns out that even though $\iota\xi_\varepsilon$ does not converge, it is possible to find a sequence M_ε of elements in $\mathfrak{R}$ such that the sequence $M_\varepsilon\iota\xi_\varepsilon$ converges to a limiting model $(\hat{\Pi}, \hat{\Gamma})$. Unfortunately, the elements M_ε no *not* preserve the image of ι in the space of admissible models. As a consequence, when solving the fixed point map (15.7) with respect to the model $M_\varepsilon\iota\xi_\varepsilon$ and inserting the solution into the reconstruction operator, it is not clear *a priori* that the resulting function (or distribution) can again be interpreted as the solution to some modified PDE. It turns out that in our case, at least for a suitable subgroup of $\mathfrak{R}$, this is again the case and the modified equation is precisely given by (15.6), where C_ε is some linear combination of the constants appearing in the description of M_ε.

There are now three questions that remain to be answered:

1. How does one construct the renormalisation group $\mathfrak{R}$?
2. How does one derive the new equation obtained when renormalising a model?
3. What is the right choice of M_ε ensuring that the renormalised models converge?

15.5.1 The renormalisation group

How does all this help with the identification of a natural class of deformations for the usual product? First, it turns out that for every continuous function ξ, if we denote again by (Π, Γ) the model $\iota\xi$, then the linear map $\boldsymbol{\Pi}\colon T \to \mathcal{C}$ given by

$$\boldsymbol{\Pi} = \Pi_y F_y^{-1} , \tag{15.16}$$

which is independent of the choice of y by the above discussion, is given by

$$\big(\boldsymbol{\Pi}\Xi\big)(x) = \xi(x) , \qquad \big(\boldsymbol{\Pi}X^k\big)(x) = x^k , \tag{15.17}$$

and then recursively by

$$\boldsymbol{\Pi}\tau\bar\tau = \boldsymbol{\Pi}\tau \cdot \boldsymbol{\Pi}\bar\tau , \qquad \boldsymbol{\Pi}\mathcal{I}\tau = K * \boldsymbol{\Pi}\tau . \tag{15.18}$$

Note that this is very similar to the definition of $\iota\xi$, with the notable exception that (14.5) is replaced by the more "natural" identity $\boldsymbol{\Pi}\mathcal{I}\tau = K * \boldsymbol{\Pi}\tau$. It turns out that the knowledge of $\boldsymbol{\Pi}$ and the knowledge of (Π, Γ) are equivalent since one has $\Pi_x = \boldsymbol{\Pi} F_x$ and the map F_x can be recovered from Π_x by (15.12). (This argument appears circular but it is possible to put a suitable recursive structure on T and T^+ ensuring that this actually works.) Furthermore, the translation $(\Pi, \Gamma) \leftrightarrow \boldsymbol{\Pi}$ actually works for *any* admissible model and does not at all rely on the fact that it was built by lifting a continuous function. However, in the general case, the first identity in (15.17) does of course not make any sense anymore and might fail even if the coordinates of $\boldsymbol{\Pi}$ consist of continuous functions.

At this stage we note that if ξ happens to be a stationary stochastic process and $\boldsymbol{\Pi}$ is built from ξ by following the above procedure, then $\boldsymbol{\Pi}\tau$ is a stationary stochastic process for every $\tau \in T$. In order to define $\mathfrak{R}$, it is natural to consider only transformations of the space of admissible models that preserve this property. Since we are not in general allowed to multiply components of $\boldsymbol{\Pi}$, the only remaining operation is to form linear combinations. It is therefore natural to describe elements of $\mathfrak{R}$ by linear maps $M\colon T \to T$ and to postulate their action on admissible models by $\boldsymbol{\Pi} \mapsto \boldsymbol{\Pi}^M$ with

$$\boldsymbol{\Pi}^M\tau = \boldsymbol{\Pi} M\tau . \tag{15.19}$$

It is not clear *a priori* whether given such a map M and an admissible model (Π, Γ) there is a coherent way of building a new model (Π^M, Γ^M) such that $\boldsymbol{\Pi}^M$ is the map associated to (Π^M, Γ^M) as above. It turns out that one has the following statement:

Proposition 15.7. *In the above context, for every linear map $M\colon T \to T$ commuting with $\mathcal{I}$ and multiplication by X^k, there exist unique linear maps $\Delta^M\colon T \to T \otimes T^+$ and $\hat\Delta^M\colon T^+ \to T^+ \otimes T^+$ such that if we set*

$$\Pi_x^M\tau = \big(\Pi_x \otimes f_x\big)\Delta^M\tau , \qquad \gamma_{xy}^M(\sigma) = (\gamma_{xy} \otimes f_x)\hat\Delta^M\sigma ,$$

then Π_x^M satisfies again (14.5) and the identity $\Pi_x^M\Gamma_{xy}^M = \Pi_y^M$.

At this stage it may look like *any* linear map $M\colon T \to T$ commuting with $\mathcal{I}$ and multiplication by X^k yields a transformation on the space of admissible models by Proposition 15.7. This however is not true since we have completely disregarded the *analytical* bounds that every model has to satisfy. It is clear from Definition 13.6 that in the absence of any additional knowledge these are satisfied if and only if $\Pi_x^M\tau$ is a linear combination of the $\Pi_x\bar\tau$ for some symbols $\bar\tau$ with $|\bar\tau| \ge |\tau|$. This suggests the following definition.

Definition 15.8. The renormalisation group $\mathfrak{R}$ consists of the set of linear maps $M\colon T \to T$ commuting with $\mathcal{I}$, $\mathcal{I}'$, and with multiplication by X^k, such that for $\tau \in T_\alpha$ one has

$$\Delta^M \tau - \tau \otimes \mathbf{1} \in T_{>\alpha} \otimes T^+ \;. \tag{15.20}$$

Its action on the space of admissible models is given by Proposition 15.7.

Remark 15.9. In principle, one should of course also impose that

$$\hat\Delta^M \sigma - \sigma \otimes \mathbf{1} \in T^+_{>\alpha} \otimes T^+ \;.$$

However, it turns out that this is always the case, provided that Δ^M satisfies (15.20). The reason for this is that it is possible to verify that one always has the identity $\hat\Delta^M \mathcal{J}_k(\tau) = (\mathcal{J}_k \otimes I)\Delta^M \tau$.

15.5.2 The renormalised equations

In the case of the KPZ equation, it turns out that we need a three-parameter subgroup of $\mathfrak{R}$ to renormalise the equations, but in order to explain the procedure we will consider a larger 4-dimensional subgroup of $\mathfrak{R}$. More precisely, we consider elements $M \in \mathfrak{R}$ of the form $M = \exp(-\sum_{i=0}^3 C_i L_i)$, where the generators L_i are determined by the following contraction rules:

$$L_0\colon \;\mapsto \mathbf{1}\;, \qquad L_1\colon \;\mapsto \mathbf{1}\;, \qquad L_2\colon \;\mapsto \mathbf{1} \qquad L_3\colon \;\mapsto \mathbf{1}\;. \tag{15.21}$$

This should be understood in the sense that if τ is an arbitrary formal expression, then $L_0\tau$ is the sum of all formal expressions obtained from τ by performing a substitution of the type $\;\mapsto \mathbf{1}$. For example, one has

$$L_0 \; = 2 \;, \qquad L_0 \; = 2 \; + \;,$$

etc. The extension of the other operators L_i to all of T proceeds in principle along the same lines. However, as a consequence of the fact that $\mathcal{I}(\mathbf{1}) = \mathcal{I}'(\mathbf{1}) = 0$ by construction, it actually turns out that $L_i\tau = 0$ for $i \neq 0$ and every τ for which L_i wasn't defined in (15.21). It is possible to verify that one has the following result.

Proposition 15.10. *The linear maps M of the type just described belong to $\mathfrak{R}$. Furthermore, if (Π, Γ) is an admissible model such that $\Pi_x\tau$ is a continuous function for every $\tau \in T$, then one has the identity*

$$\big(\Pi^M_x \tau\big)(x) = \big(\Pi_x M\tau\big)(x)\;. \tag{15.22}$$

Remark 15.11. Note that it is the same value x that appears twice on each side of (15.22). It is in fact *not* the case that one has $\Pi_x^M \tau = \Pi_x M \tau$ in general! However, the identity (15.22) is all we need to derive the renormalised equation.

It is now rather straightforward to show the following:

Proposition 15.12. *Let $M = \exp(-\sum_{i=0}^{3} C_i L_i)$ as above and let $(\Pi^M, \Gamma^M) = M\iota\xi$ for some smooth function ξ. Let furthermore H be the solution to (15.7) with respect to the model (Π^M, Γ^M). Then, writing $\mathcal{R}^M$ for the reconstruction operator associated to this renormalised model, the function $h(t,x) = \big(\mathcal{R}^M H\big)(t,x)$ solves the equation*

$$\partial_t h = \partial_x^2 h + (\partial_x h)^2 - 4C_0\, \partial_x h + \xi - (C_1 + C_2 + 4C_3)\;.$$

Proof. By Theorem 14.5, it turns out that (15.7) can be solved in $\mathscr{D}^\gamma$ as soon as γ is a little bit greater than $3/2$. Therefore, we only need to keep track of its solution H up to terms of homogeneity $3/2$. By repeatedly applying the identity (15.8), we see that the solution $H \in \mathscr{D}^\gamma$ for γ close enough to $3/2$ is necessarily of the form

$$H = h\,\mathbf{1} + \quad + \quad + h'\,X_1 + 2 \quad + 2h' \quad,$$

for some real-valued functions h and h'. (Note that h' is treated as an independent function here, we certainly do not suggest that the function h is differentiable! Our notation is only by analogy with the classical Taylor expansion...) As an immediate consequence, ∂H is given by

$$\partial H = \quad + \quad + h'\,\mathbf{1} + 2 \quad + 2h' \quad, \tag{15.23}$$

as an element of $\mathscr{D}^\gamma$ for γ sufficiently close to $1/2$. Similarly, the right hand side of the equation is given up to order 0 by

$$(\partial H)^2 + \Xi = \Xi + \quad + 2 \quad + 2h' \quad + \quad + 4 \quad + 2h' \quad + 4h' \quad + (h')^2\,\mathbf{1}\;. \tag{15.24}$$

It follows from the definition of M that one then has the identity

$$M\partial H = \partial H - 4C_0 \quad,$$

so that, as an element of $\mathscr{D}^\gamma$ with very small (but positive) γ, one has the identity

$$(M\partial H)^2 = (\partial H)^2 - 8C_0 \quad.$$

As a consequence, after neglecting all terms of strictly positive order, one has the identity (writing c instead of $c\mathbf{1}$ for real constants c)

$$\begin{aligned} M\big((\partial H)^2 + \Xi\big) &= (\partial H)^2 + \Xi - C_0\big(4 \quad + 4 \quad + 8 \quad + 4h'\,\mathbf{1}\big) - C_1 - C_2 - 4C_3 \\ &= (M\partial H)^2 + \Xi - 4C_0\, M\partial H - (C_1 + C_2 + 4C_3)\;. \end{aligned}$$

Combining this with (15.22), the claim now follows at once. $\square$

Remark 15.13. It turns out that, thanks to the symmetry $x \mapsto -x$ enjoyed by our problem, the corresponding model can be renormalised by a map M as above, but with $C_0 = 0$. The reason why we considered the general case here is twofold. First, it shows that it is possible to obtain renormalised equations that differ from the original equation in a more complicated way than just by the addition of a large constant. Second, it is plausible that if one tries to approximate the KPZ equation by a microscopic model which is not symmetric under space inversion, then the constant C_0 could play a non-trivial role.

15.5.3 Convergence of the renormalised models

It remains to argue why one expects to be able to find constants C_i^ε such that the sequence of renormalised models $M^\varepsilon \iota \xi_\varepsilon$ with $M^\varepsilon = \exp(\sum_{i=1}^3 C_i^\varepsilon L_i)$ converges to a limiting model. Instead of considering the actual sequence of models, we only consider the sequence of stationary processes $\hat{\boldsymbol{\Pi}}^\varepsilon \tau := \boldsymbol{\Pi}^\varepsilon M^\varepsilon \tau$, where $\boldsymbol{\Pi}^\varepsilon$ is associated to $(\Pi^\varepsilon, \Gamma^\varepsilon) = \iota \xi_\varepsilon$ as in Section 15.5.1.

Remark 15.14. It is important to note that we do *not* attempt here to give a full proof that the renormalised model converges to a limit in the correct topology for the space of admissible models. We only aim to argue that it is *plausible* that $\hat{\boldsymbol{\Pi}}^\varepsilon$ converges to a limit in *some* topology. A full proof of convergence (but in a slightly different setting) can be found in [Hai13], see also [Hai14c, Section 10].

Since there are general arguments available to deal with all the expressions τ of positive homogeneity as well as expressions of the type $\mathcal{I}'(\tau)$ and Ξ itself, we restrict ourselves to those that remain. Inspecting (15.10), we see that they are given by

For this part, some elementary notions from the theory of Wiener chaos expansions are required, but we'll try to hide this as much as possible. At a formal level, one has the identity

$$\boldsymbol{\Pi}^\varepsilon \,\cdot\, = K' * \xi_\varepsilon = K'_\varepsilon * \xi \;,$$

where the kernel K'_ε is given by $K'_\varepsilon = K' * \delta_\varepsilon$. This shows that, at least formally, one has

$$\big(\boldsymbol{\Pi}^\varepsilon \,\cdot\,\big)(z) = \big(K' * \xi_\varepsilon\big)(z)^2 = \int\!\!\int K'_\varepsilon(z - z_1) K'_\varepsilon(z - z_2)\, \xi(z_1)\xi(z_2)\, dz_1\, dz_2 \;.$$

Similar but more complicated expressions can be found for any formal expression τ. This naturally leads to the study of random variables of the type

$$I_k(f) = \int \cdots \int f(z_1, \ldots, z_k)\, \xi(z_1) \cdots \xi(z_k)\, dz_1 \cdots dz_k \;. \tag{15.25}$$

Ideally, one would hope to have an Itô isometry of the type $\mathbf{E} I_k(f) I_k(g) = \langle f^{\mathrm{sym}}, g^{\mathrm{sym}} \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the L^2-scalar product and f^{sym} denotes the symmetrisation of f. This is unfortunately *not* the case. Instead, one should replace the products in (15.25) by *Wick products*, which are formally generated by all possible *contractions* of the type

$$\xi(z_i)\xi(z_j) \mapsto \xi(z_i) \diamond \xi(z_j) + \delta(z_i - z_j) \;.$$

If we then set

$$\hat I_k(f) = \int \cdots \int f(z_1, \ldots, z_k)\, \xi(z_1) \diamond \cdots \diamond \xi(z_k)\, dz_1 \cdots dz_k \;,$$

One has indeed

$$\mathbf{E} \hat I_k(f) \hat I_k(g) = \langle f^{\mathrm{sym}}, g^{\mathrm{sym}} \rangle \;.$$

Furthermore, one has equivalence of moments in the sense that, for every $k > 0$ and $p > 0$ there exists a constant $C_{k,p}$ such that

$$\mathbf{E} |\hat I_k(f)|^p \le C_{k,p} \|f^{\mathrm{sym}}\|^p \;.$$

Finally, one has $\mathbf{E} \hat I_k(f) \hat I_\ell(g) = 0$ if $k \neq \ell$. Random variables of the form $\hat I_k(f)$ for some $k \ge 0$ and some square integrable function f are said to belong to the *kth homogeneous Wiener chaos*.

Returning to our problem, we first argue that it should be possible to choose M^ε in such a way that $\hat{\boldsymbol{\Pi}}^\varepsilon \vee$ converges to a limit as $\varepsilon \to 0$. The above considerations suggest that one should rewrite $\boldsymbol{\Pi}^\varepsilon \vee$ as

$$\begin{aligned} \big(\boldsymbol{\Pi}^\varepsilon \vee\big)(z) &= \big(K' * \xi_\varepsilon\big)(z)^2 \\ &= \iint K'_\varepsilon(z - z_1) K'_\varepsilon(z - z_2)\, \xi(z_1) \diamond \xi(z_2)\, dz_1\, dz_2 + C_\varepsilon^{(1)} \;, \end{aligned} \tag{15.26}$$

where the constant $C_\varepsilon^{(1)}$ is given by the contraction

$$C_\varepsilon^{(1)} = \overset{\frown}{\vee} \stackrel{\mathrm{def}}{=} \int \big(K'_\varepsilon(z)\big)^2\, dz \;.$$

Note now that K'_ε is an ε-approximation of the kernel K' which has the same singular behaviour as the derivative of the heat kernel. In terms of the parabolic distance, the singularity of the derivative of the heat kernel scales like $K(z) \sim |z|^{-2}$ for $z \to 0$. (Recall that we consider the parabolic distance $|(t,x)| = \sqrt{|t|} + |x|$, so that this is consistent with the fact that the derivative of the heat kernel is bounded by t^{-1}.) This suggests that one has $\big(K'_\varepsilon(z)\big)^2 \sim |z|^{-4}$ for $|z| \gg \varepsilon$. Since parabolic space-time has scaling dimension 3 (time counts double!), this is a non-integrable singularity. As a matter of fact, there is a whole power of z missing to make it borderline integrable, which suggests that one has

$$C_\varepsilon^{(1)} \sim \frac{1}{\varepsilon}\;.$$

This already shows that one should not expect $\boldsymbol{\Pi}^\varepsilon \vee$ to converge to a limit as $\varepsilon \to 0$. However, it turns out that the first term in (15.26) converges to a distribution-valued stationary space-time process, so that one would like to somehow get rid of this diverging constant $C_\varepsilon^{(1)}$. This is exactly where the renormalisation map M^ε (in particular the factor $\exp(-C_1 L_1)$) enters into play. Following the above definitions, we see that one has

$$\big(\hat{\boldsymbol{\Pi}}^\varepsilon \vee\big)(z) = \big(\boldsymbol{\Pi}^\varepsilon M \vee\big)(z) = \big(\boldsymbol{\Pi}^\varepsilon \vee\big)(z) - C_1\;.$$

This suggests that if we make the choice $C_1 = C_\varepsilon^{(1)}$, then $\hat{\boldsymbol{\Pi}}^\varepsilon \vee$ does indeed converge to a non-trivial limit as $\varepsilon \to 0$. This limit is a distribution given, at least formally, by

$$\big(\boldsymbol{\Pi}^\varepsilon \vee\big)(\psi) = \int\!\!\int \psi(z) K'(z - z_1) K'(z - z_2)\, dz\, \xi(z_1) \diamond \xi(z_2)\, dz_1\, dz_2\;.$$

Using again the scaling properties of the kernel K', it is not too difficult to show that this yields indeed a random variable belonging to the second homogeneous Wiener chaos for every choice of smooth test function ψ.

The case $\tau = \Diamond$ is treated in a somewhat similar way. This time one has

$$\begin{aligned}\big(\boldsymbol{\Pi}^\varepsilon \Diamond\big)(z) &= \big(K' * \xi_\varepsilon\big)(z)\big(K' * K' * \xi_\varepsilon\big)(z)\\ &= \int\!\!\int K'_\varepsilon(z - z_1)(K * K'_\varepsilon)(z - z_2)\, \xi(z_1) \diamond \xi(z_2)\, dz_1\, dz_2 + C_\varepsilon^{(0)}\;,\end{aligned}$$

where the constant $C_\varepsilon^{(0)}$ is given by the contraction

$$C_\varepsilon^{(0)} = \Diamond \stackrel{\text{def}}{=} \int K'_\varepsilon(z)\big(K' * K'_\varepsilon\big)(z)\, dz\;.$$

This time however K'_ε is an odd function (in the spatial variable) and $K' * K'_\varepsilon$ is an even function, so that $C_\varepsilon^{(0)}$ vanishes for every $\varepsilon > 0$. This is why we can set $C_0 = 0$ and no renormalisation is required for $\Diamond$.

Turning to our list of terms of negative homogeneity, it remains to consider $\Diamond\!\vee$, $\vee\!\vee$, and $\vee\!\Diamond$. It turns out that the latter two are the more difficult ones, so we only discuss these. Let us first argue why we expect to be able to choose the constant C_2 in such a way that $\hat{\boldsymbol{\Pi}}^\varepsilon \vee\!\vee$ converges to a limit. In this case, the "bad" term comes from the part of $\big(\boldsymbol{\Pi}^\varepsilon \vee\!\vee\big)(z)$ belonging to the homogeneous chaos of order 0. This is simply a constant, which is given by

$$C_\varepsilon^{(2)} = 2\,\vee\!\vee \stackrel{\text{def}}{=} 2\int K'(z)K'(\bar z)Q_\varepsilon^2(z - \bar z)\, dz\, d\bar z\;, \tag{15.27}$$

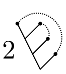
where the kernel Q_ε is given by

$$Q_\varepsilon(z) = \int K'_\varepsilon(\bar z) K'_\varepsilon(\bar z - z)\, d\bar z\;.$$

Remark 15.15. The factor 2 comes from the fact that the contraction (15.27) appears twice, since it is equal to the contraction . In principle, one would think that the contraction also contributes to $C_\varepsilon^{(2)}$. This term however vanishes due to the fact that the integral of K'_ε vanishes.

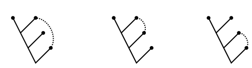
Since K'_ε is an ε-mollification of a kernel with a singularity of order -2 and the scaling dimension of the underlying space is 3, we see that Q_ε behaves like an ε-mollification of a kernel with a singularity of order $-2-2+3=-1$ at the origin. As a consequence, the singularity of the integrand in (15.27) is of order -6, which gives rise to a logarithmic divergence as $\varepsilon \to 0$. This suggests that one should choose $C_2 = C_\varepsilon^{(2)}$ in order to cancel out this diverging term and obtain a non-trivial limit for $\hat{\boldsymbol{\Pi}}^\varepsilon$ as $\varepsilon \to 0$. This is indeed the case.

We finally turn to the case $\tau =$. In this case, there are "bad" terms appearing in the Wiener chaos decomposition of $\boldsymbol{\Pi}^\varepsilon$ both in the second and the zeroth Wiener chaos. This time, the constant appearing in the zeroth Wiener chaos is given by

$$C_\varepsilon^{(3)} = 2$$  $$\stackrel{\text{def}}{=} 2 \int K'(z)K'(\bar z)Q_\varepsilon(\bar z)Q_\varepsilon(z+\bar z)\, dz\, d\bar z\;,$$

which diverges logarithmically for exactly the same reason as $C_\varepsilon^{(2)}$. Setting $C_2 = C_\varepsilon^{(2)}$, this diverging constant can again be cancelled out. The combinatorial factor 2 arises in essentially the same way as for and the contribution of the term where the two top nodes are contracted vanishes for the same reason as previously.

It remains to consider the contribution of $\boldsymbol{\Pi}^\varepsilon$ to the second Wiener chaos. This contribution consists of three terms, which correspond to the contractions



It turns out that the first one of these terms does not give raise to any singularity. The last two terms can be treated in essentially the same way, so we focus on the last one, which we denote by η^ε. For fixed ε, the distribution (actually smooth function) η^ε is given by

$$\begin{aligned}\eta^\varepsilon(\psi) = \int & \psi(z_0)K'(z_0-z_1)Q_\varepsilon(z_0-z_1)K'(z_2-z_1)\\ &\times K'_\varepsilon(z_3-z_2)K'_\varepsilon(z_4-z_2)\,\xi(z_3)\diamond\xi(z_4)\, dz\;.\end{aligned}$$

The problem with this is that as $\varepsilon \to 0$, the product $\hat Q_\varepsilon := K' Q_\varepsilon$ converges to a kernel $\hat Q = K'Q$, which has a non-integrable singularity at the origin. In particular, it is not clear *a priori* whether the action of integrating a test function against $\hat Q_\varepsilon$ converges to a limiting distribution as $\varepsilon \to 0$. Our saving grace here is that since Q_ε is even and K' is odd, the kernel $\hat Q_\varepsilon$ integrates to 0 for every fixed ε.

This is akin to the problem of making sense of the "Cauchy principal value" distribution, which formally corresponds to the integration against $1/x$. For the sake of the argument, let us consider a function $W : \mathbf{R} \to \mathbf{R}$ which is compactly supported and smooth everywhere except at the origin, where it diverges like $|W(x)| \sim 1/|x|$. It is then natural to associate to W a "renormalised" distribution $\mathscr{R}W$ given by

$$\big(\mathscr{R}W\big)(\varphi) = \int W(x)\big(\varphi(x) - \varphi(0)\big)\,dx\;.$$

Note that $\mathscr{R}W$ has the property that if $\varphi(0) = 0$, then it simply corresponds to integration against W, which is the standard way of associating a distribution to a function. Furthermore, the above expression is always well-defined, since φ is smooth and therefore the factor $(\varphi(x) - \varphi(0))$ cancels out the singularity of W at the origin. It is also straightforward to verify that if W_ε is a sequence of smooth approximations to W (say one has $W_\varepsilon(x) = W(x)$ for $|x| > \varepsilon$ and $|W_\varepsilon| \lesssim 1/\varepsilon$ otherwise) which has the property that each W_ε integrates to 0, then $W^\varepsilon \to \mathscr{R}W$ in a distributional sense.

In the same way, one can show that $\hat{Q}_\varepsilon$ converges as $\varepsilon \to 0$ to a limiting distribution $\mathscr{R}\hat{Q}$. As a consequence, one can show that η^ε converges to a limiting (random) distribution η given by

$$\eta(\psi) = \int \psi(z_0)\,\mathscr{R}\hat{Q}(z_0-z_1)K'(z_2-z_1)K'(z_3-z_2)K'(z_4-z_2)\,\xi(z_3)\diamond\xi(z_4)\,dz\;.$$

It should be clear from this whole discussion that while the precise values of the constants C_i depend on the details of the mollifier δ_ε, the limiting (random) model $(\hat{\Pi}, \hat{\Gamma})$ obtained in this way is independent of it. Combining this with the continuity of the solution to the fixed point map (15.7) and of the reconstruction operator $\mathcal{R}$ with respect to the underlying model, we see that the statement of Theorem 15.1 follows almost immediately.

15.6 The KPZ equation and rough paths

In the particular case of the KPZ equation, it turns out that is possible to give a robust solution theory by only using "classical" controlled rough path theory, as exposed in the earlier part of this book. This is actually how it was originally treated in [Hai13]. To see how this can be the case, we make the following crucial remarks:

1. First, looking at the expression (15.23) for ∂H, we see that most symbols come with constant coefficients. The only non-constant coefficients that appear are in front of the term $\mathbf{1}$, which is some kind of renormalised value for ∂H, and in front of the term . This suggests that the problem of finding a solution h to the KPZ equation (or equivalently a solution h' to the corresponding Burgers equation) can be simplified considerably by considering instead the function v given by

$$v = \partial_x h - \boldsymbol{\Pi}\big(\upharpoonleft + \curlyvee + 2\Psi\big)\;, \tag{15.28}$$

where $\boldsymbol{\Pi}$ is the operator given in (15.16).

2. The only symbol τ appearing in ∂H such that $|\tau| + |\lessdot| < 0$ is the symbol $\upharpoonleft$. Furthermore, one has

$$\begin{aligned} \Delta\mathbf{1} &= \mathbf{1}\otimes\mathbf{1}\;, \qquad & \Delta\lessdot &= \lessdot\otimes\mathbf{1} + \mathbf{1}\otimes\mathcal{J}'(\upharpoonleft)\;,\\ \Delta\upharpoonleft &= \upharpoonleft\otimes\mathbf{1}\;, \qquad & \Delta\diamond &= \diamond\otimes\mathbf{1} + \upharpoonleft\otimes\mathcal{J}'(\upharpoonleft)\;. \end{aligned}$$

It then follows from this and the definition (15.13) of the structure group G that the space $\langle \upharpoonleft, \diamond, \mathbf{1}, \lessdot\rangle \subset T$ is invariant under the action of G. Furthermore, its action on this subspace is completely described by one real number corresponding to $\mathcal{J}'(\upharpoonleft)$. Finally, viewing this subspace as a regularity structure in its own right, we see that it is nothing but the regularity structure of Section 13.3.2, provided that we make the identifications $\upharpoonleft \sim \dot W$, $\lessdot \sim W$, and $\diamond \sim \dot{\mathbb{W}}$.

3. One has the identities

$$\Delta\Psi = \Psi\otimes\mathbf{1} + \upharpoonleft\otimes\mathcal{J}'(\Psi)\;, \qquad \Delta\Phi = \Phi\otimes\mathbf{1} + \upharpoonleft\otimes\mathcal{J}'(\Phi)\;,$$

so that the pair of symbols $\{\mathcal{J}'(\Psi), \Phi\}$ could also have played the role of $\{W, \dot{\mathbb{W}}\}$ in the previous remark.

Let now ξ be a *smooth* function and let h be given by the solution to the unrenormalised KPZ equation (15.1). Defining $\boldsymbol{\Pi}$ by $\boldsymbol{\Pi}\Xi = \xi$ and then recursively as in (15.18), and defining v by (15.28), we then obtain for v the equation

$$\partial_t v = \partial_x^2 v + \partial_x\big(v\,\boldsymbol{\Pi}\upharpoonleft + 4\,\boldsymbol{\Pi}\Phi\big) + R\;, \tag{15.29}$$

where the "remainder" R belongs to $\mathcal{C}^\alpha$ for every $\alpha < -1$. Similarly to before, it also turns out that if we replace $\boldsymbol{\Pi}$ bi $\hat{\boldsymbol{\Pi}} = \boldsymbol{\Pi}^M$ defined as in (15.19) (with $C_0 = 0$) and h as the solution to the renormalised KPZ equation (15.6) with $C_\varepsilon = C_1 + C_2 + 4C_3$, then v also satisfies (15.29), but with $\boldsymbol{\Pi}$ replaced by the renormalised model $\hat{\boldsymbol{\Pi}}$.

We are now in the following situation. As a consequence of (15.23) we can *guess* that for any fixed time t, the solution v should be controlled by the function $\hat{\boldsymbol{\Pi}}\lessdot$, which we can interpret as one component (say W^1) of some rough path $(W, \mathbb{W})$. Note that here the spatial variable plays the role of time! The time variables merely plays the role of a parameter, so we really have a family of rough paths indexed by time. Furthermore, $\hat{\boldsymbol{\Pi}}\upharpoonleft$ can be interpreted as the distributional derivative of another component (say W^0) of the rough path W. Finally, the function $\hat{\boldsymbol{\Pi}}\Psi$ can be interpreted as a third component W^2 of W.

As a consequence of the second and third remarks above, the two distributions $\hat{\boldsymbol{\Pi}}\diamond$ and $\hat{\boldsymbol{\Pi}}\Phi$ can then be interpreted as the distributional derivatives of the "iterated integrals" $\mathbb{W}^{1,0}$ and $\mathbb{W}^{2,1}$. It follows automatically from these algebraic relations combined with the analytic bounds (13.13) that $\mathbb{W}^{1,0}$ and $\mathbb{W}^{2,1}$ then satisfy the required estimates (2.3). Our model does not provide any values for $\mathbb{W}^{1,2}$, but these turn out not to be required. Assuming that v is indeed controlled by $X_1 = \hat{\boldsymbol{\Pi}}\lessdot$, it is then possible to give meaning to the term $v\,\boldsymbol{\Pi}$ appearing in (15.29) by using "classical" rough integration.

As a consequence, we then see that the right hand side of (15.29) is of the form $\partial_x^2 Y$, for some function Y controlled by W^0. One of the main technical results of [Hai13] guarantees that if Z solves

$$\partial_t Z = \partial_x^2 Z + \partial_x^2 Y \;,$$

and Y is controlled by W^0, then Z is necessarily controlled by $W^1 = \hat{\Pi}$. This "closes the loop" and allows to set up a fixed point equation for v that is stable as a function of the underlying model $\hat{\boldsymbol{\Pi}}$ and therefore also allows to deal with the limiting case of the KPZ equation driven by space-time white noise.

References

[Alm66] F. J. ALMGREN, JR. *Plateau's problem: An invitation to varifold geometry*. W. A. Benjamin, Inc., New York-Amsterdam, 1966.

[AR91] S. ALBEVERIO and M. RÖCKNER. Stochastic differential equations in infinite dimensions: solutions via Dirichlet forms. *Probab. Theory Related Fields* **89**, no. 3, (1991), 347–386. doi:10.1007/BF01198791.

[Bal00] E. J. BALDER. Lectures on Young measure theory and its applications in economics. *Rend. Istit. Mat. Univ. Trieste* **31**, no. suppl. 1, (2000), 1–69. Workshop on Measure Theory and Real Analysis (Italian) (Grado, 1997).

[Bau04] F. BAUDOIN. *An introduction to the geometry of stochastic flows*. Imperial College Press, London, 2004.

[BCD11] H. BAHOURI, J.-Y. CHEMIN, and R. DANCHIN. *Fourier analysis and nonlinear partial differential equations*, vol. 343 of *Grundlehren der Mathematischen Wissenschaften*. Springer, Heidelberg, 2011.

[BF13] C. BAYER and P. K. FRIZ. Cubature on Wiener space: pathwise convergence. *Appl. Math. Optim.* **67**, no. 2, (2013), 261–278. doi:10.1007/s00245-012-9187-8.

[BFH09] E. BREUILLARD, P. FRIZ, and M. HUESMANN. From random walks to rough paths. *Proc. Amer. Math. Soc.* **137**, no. 10, (2009), 3487–3496. doi:10.1090/S0002-9939-09-09930-4.

[BFRS13] C. BAYER, P. K. FRIZ, S. RIEDEL, and J. SCHOENMAKERS. From rough path estimates to multilevel Monte Carlo. *ArXiv e-prints* (2013). arXiv:1305.5779.

[BG97] L. BERTINI and G. GIACOMIN. Stochastic Burgers and KPZ equations from particle systems. *Comm. Math. Phys.* **183**, no. 3, (1997), 571–607.

[BH91] N. BOULEAU and F. HIRSCH. *Dirichlet forms and analysis on Wiener space*, vol. 14 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1991.

[BH07] F. BAUDOIN and M. HAIRER. A version of Hörmander's theorem for the fractional Brownian motion. *Probab. Theory Related Fields* **139**, no. 3-4, (2007), 373–395. doi:10.1007/s00440-006-0035-0.

[Bis81a] J.-M. BISMUT. Martingales, the Malliavin calculus and Hörmander's theorem. In *Stochastic integrals (Proc. Sympos., Univ. Durham, Durham, 1980)*, vol. 851 of *Lecture Notes in Math.*, 85–109. Springer, Berlin, 1981.

[Bis81b] J.-M. BISMUT. Martingales, the Malliavin calculus and hypoellipticity under general Hörmander's conditions. *Z. Wahrsch. Verw. Gebiete* **56**, no. 4, (1981), 469–505.

[BM07] R. BUCKDAHN and J. MA. Pathwise stochastic control problems and stochastic HJB equations. *SIAM Journal on Control and Optimization* **45**, no. 6, (2007), 2224–2256. doi:10.1137/S036301290444335X.

[BMN10] Á. BÉNYI, D. MALDONADO, and V. NAIBO. What is . . . a paraproduct? *Notices Amer. Math. Soc.* **57**, no. 7, (2010), 858–860.

[BNQ13] H. BOEDIHARDJO, H. NI, and Z. QIAN. Uniqueness of signature for simple curves. *ArXiv e-prints* (2013). arXiv:1304.0755.

[Bog98] V. I. BOGACHEV. *Gaussian measures*, vol. 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998.

[Bon81] J.-M. BONY. Calcul symbolique et propagation des singularités pour les équations aux dérivées partielles non linéaires. *Ann. Sci. École Norm. Sup. (4)* **14**, no. 2, (1981), 209–246.

[Bor75] C. BORELL. The Brunn-Minkowski inequality in Gauss space. *Invent. Math.* **30**, no. 2, (1975), 207–216.

[BP08] J. BOURGAIN and N. PAVLOVIĆ. Ill-posedness of the Navier-Stokes equations in a critical space in 3D. *J. Funct. Anal.* **255**, no. 9, (2008), 2233–2247. doi:10.1016/j.jfa.2008.07.008.

[CC13] R. CATELLIER and K. CHOUK. Paracontrolled Distributions and the 3-dimensional Stochastic Quantization Equation. *ArXiv e-prints* (2013). arXiv:1310.6869.

[CDFO13] D. CRISAN, J. DIEHL, P. K. FRIZ, and H. OBERHAUSER. Robust filtering: correlated noise and multidimensional observation. *Ann. Appl. Probab.* **23**, no. 5, (2013), 2139–2160. doi:10.1214/12-AAP896.

[CF09] M. CARUANA and P. FRIZ. Partial differential equations driven by rough paths. *J. Differential Equations* **247**, no. 1, (2009), 140–173. doi:10.1016/j.jde.2009.01.026.

[CF10] T. CASS and P. FRIZ. Densities for rough differential equations under Hörmander's condition. *Ann. of Math. (2)* **171**, no. 3, (2010), 2115–2141. doi:10.4007/annals.2010.171.2115.

[CFO11] M. CARUANA, P. K. FRIZ, and H. OBERHAUSER. A (rough) pathwise approach to a class of non-linear stochastic partial differential equations. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **28**, no. 1, (2011), 27–46. doi:10.1016/j.anihpc.2010.11.002.

[CFV07] L. COUTIN, P. FRIZ, and N. VICTOIR. Good rough path sequences and applications to anticipating stochastic calculus. *Ann. Probab.* **35**, no. 3, (2007), 1172–1193. doi:10.1214/009117906000000827.

[CFV09] T. CASS, P. FRIZ, and N. VICTOIR. Non-degeneracy of Wiener functionals arising from rough differential equations. *Trans. Amer. Math. Soc.* **361**, no. 6, (2009), 3359–3371. doi:10.1090/S0002-9947-09-04677-7.

[Che54] K.-T. CHEN. Iterated integrals and exponential homomorphisms. *Proc. London Math. Soc. (3)* **4**, (1954), 502–512.

[Che13] I. CHEVYREV. Unitary representations of geometric rough paths. *ArXiv e-prints* (2013). arXiv:1307.3580.

[CHLT12] T. CASS, M. HAIRER, C. LITTERER, and S. TINDEL. Smoothness of the density for solutions to Gaussian Rough Differential Equations. *ArXiv e-prints* (2012). arXiv:1209.3100. Ann. Probab., to appear.

[CIL92] M. G. CRANDALL, H. ISHII, and P.-L. LIONS. User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.)* **27**, no. 1, (1992), 1–67.

[CLL13] T. CASS, C. LITTERER, and T. LYONS. Integrability and tail estimates for Gaussian rough differential equations. *Ann. Probab.* **41**, no. 4, (2013), 3026–3050. doi:doi:10.1214/12-AOP821.

[Col51] J. D. COLE. On a quasi-linear parabolic equation occurring in aerodynamics. *Quart. Appl. Math.* **9**, (1951), 225–236.

[Cor12] I. CORWIN. The Kardar-Parisi-Zhang equation and universality class. *Random Matrices Theory Appl.* **1**, no. 1, (2012), 1130001, 76. doi:10.1142/S2010326311300014.

[CQ02] L. COUTIN and Z. QIAN. Stochastic analysis, rough path analysis and fractional Brownian motions. *Probab. Theory Related Fields* **122**, no. 1, (2002), 108–140.

[Dau88] I. DAUBECHIES. Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* **41**, no. 7, (1988), 909–996. doi:10.1002/cpa.3160410705.

[Dav08] A. M. DAVIE. Differential equations driven by rough paths: an approach via discrete approximation. *Appl. Math. Res. Express. AMRX* **2008**, no. 2, (2008), 1–40. `doi:10.1093/amrx/abm009`.

[Der10] S. DEREICH. Rough paths analysis of general Banach space-valued Wiener processes. *J. Funct. Anal.* **258**, no. 9, (2010), 2910–2936. `doi:10.1016/j.jfa.2010.01.018`.

[DF12] J. DIEHL and P. FRIZ. Backward stochastic differential equations with rough drivers. *Ann. Probab.* **40**, no. 4, (2012), 1715–1758. `doi:10.1214/11-AOP660`.

[DFG13] J. DIEHL, P. FRIZ, and P. GASSIAT. Stochastic control with rough paths. *ArXiv e-prints* (2013). `arXiv:1303.7160`.

[DFM13] J. DIEHL, P. FRIZ, and H. MAI. Pathwise stability of likelihood estimators for diffusions via rough paths. *ArXiv e-prints* (2013). `arXiv:1311.1061`.

[DFO14] J. DIEHL, P. K. FRIZ, and H. OBERHAUSER. Regularity theory for rough partial differential equations and parabolic comparison revisited. *Preprint; earlier version available on arXiv* (2014).

[DFS14] J. DIEHL, P. FRIZ, and W. STANNAT. Stochastic partial differential equations: a rough path view, 2014. Preprint.

[DGT12] A. DEYA, M. GUBINELLI, and S. TINDEL. Non-linear rough heat equations. *Probab. Theory Related Fields* **153**, no. 1-2, (2012), 97–147. `doi:10.1007/s00440-011-0341-z`.

[DNT12a] A. DEYA, A. NEUENKIRCH, and S. TINDEL. A Milstein-type scheme without Lévy area terms for SDEs driven by fractional Brownian motion. *Ann. Inst. Henri Poincaré Probab. Stat.* **48**, no. 2, (2012), 518–550. `doi:10.1214/10-AIHP392`.

[DNT12b] A. DEYA, A. NEUENKIRCH, and S. TINDEL. A Milstein-type scheme without Lévy area terms for SDEs driven by fractional Brownian motion. *Ann. Inst. Henri Poincaré Probab. Stat.* **48**, no. 2, (2012), 518–550. `doi:10.1214/10-AIHP392`.

[DOR13] J. DIEHL, H. OBERHAUSER, and S. RIEDEL. A Levy-area between Brownian motion and rough paths with applications to robust non-linear filtering and RPDEs. *ArXiv e-prints* (2013). `arXiv:1301.3799`.

[DPD03] G. DA PRATO and A. DEBUSSCHE. Strong solutions to the stochastic quantization equations. *Ann. Probab.* **31**, no. 4, (2003), 1900–1916. `doi:10.1214/aop/1068646370`.

[DPZ92] G. DA PRATO and J. ZABCZYK. *Stochastic equations in infinite dimensions*, vol. 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.

[Faw04] T. FAWCETT. *Non-commutative harmonic analysis.* Ph.D. thesis, University of Oxford, 2004.

[FdLP06] D. FEYEL and A. DE LA PRADELLE. Curvilinear integrals along enriched paths. *Electronic Journal of Probability* **11**, (2006), 860–892. `doi:10.1214/EJP.v11-356`.

[FG13] P. FRIZ and P. GASSIAT. Eikonal equations and pathwise solutions to fully nonlinear SPDEs, 2013. Preprint.

[FG14] P. K. FRIZ and B. GESS. Stochastic scalar conservation laws driven by rough paths. *ArXiv e-prints* (2014). `arXiv:1403.6785`.

[FGGR13] P. K. FRIZ, B. GESS, A. GULISASHVILI, and S. RIEDEL. Jain-Monrad criterion for rough paths. *ArXiv e-prints* (2013). `arXiv:1307.3460`.

[FGL13] P. FRIZ, P. GASSIAT, and T. LYONS. Physcial Brownian motion in magnetic field as rough path. *ArXiv e-prints* (2013). `arXiv:1302.2531`.

[FLS06] P. FRIZ, T. LYONS, and D. STROOCK. Lévy's area under conditioning. *Ann. Inst. H. Poincaré Probab. Statist.* **42**, no. 1, (2006), 89–101. `doi:10.1016/j.anihpb.2005.02.003`.

[FO10] P. FRIZ and H. OBERHAUSER. A generalized Fernique theorem and applications. *Proc. Amer. Math. Soc.* **138**, (2010), 3679–3688. `doi:10.1090/S0002-9939-2010-10528-2`.

[FO14] P. FRIZ and H. OBERHAUSER. Rough path stability of (semi-)linear SPDEs. *Probab. Theory Related Fields* **158**, no. 1-2, (2014), 401–434. `doi:10.1007/s00440-013-0483-2`.

[Föl81] H. FÖLLMER. Calcul d'Itô sans probabilités. In *Seminar on Probability, XV (Univ. Strasbourg, Strasbourg, 1979/1980) (French)*, vol. 850 of *Lecture Notes in Math.*, 143–150. Springer, Berlin, 1981.

[FR11] P. FRIZ and S. RIEDEL. Convergence rates for the full Brownian rough paths with applications to limit theorems for stochastic flows. *Bull. Sci. Math.* **135**, no. 6-7, (2011), 613–628. `doi:10.1016/j.bulsci.2011.07.006`.

[FR13] P. FRIZ and S. RIEDEL. Integrability of (non-)linear rough differential equations and integrals. *Stochastic Analysis and Applications* **31**, no. 2, (2013), 336–358. `doi:10.1080/07362994.2013.759758`.

[FR14] P. FRIZ and S. RIEDEL. Convergence rates for the full Gaussian rough paths. *Ann. Inst. Henri Poincaré Probab. Stat.* **50**, no. 1, (2014), 154–194. `doi:10.1214/12-AIHP507`.

[Fri05] P. K. FRIZ. Continuity of the Itô-map for Hölder rough paths with applications to the support theorem in Hölder norm. In *Probability and partial differential equations in modern applied mathematics*, vol. 140 of *IMA Vol. Math. Appl.*, 117–135. Springer, New York, 2005.

[FS06] W. H. FLEMING and H. M. SONER. *Controlled Markov processes and viscosity solutions*, vol. 25 of *Stochastic Modelling and Applied Probability*. Springer, New York, second ed., 2006.

[FS12a] P. FRIZ and A. SHEKHAR. Doob–Meyer for rough paths. *Special Varadhan issue of Bulletin of Institute of Mathematics Academia Sinica New Series, to appear* (2012).

[FS12b] P. FRIZ and A. SHEKHAR. The Lévy-Kintchine formula for rough paths. *ArXiv e-prints* (2012). `arXiv:1212.5888`.

[FV05] P. FRIZ and N. VICTOIR. Approximations of the Brownian rough path with applications to stochastic analysis. *Ann. Inst. H. Poincaré Probab. Statist.* **41**, no. 4, (2005), 703–724. `doi:10.1016/j.anihpb.2004.05.003`.

[FV06a] P. FRIZ and N. VICTOIR. A note on the notion of geometric rough paths. *Probab. Theory Related Fields* **136**, no. 3, (2006), 395–416. `doi:10.1007/s00440-005-0487-7`.

[FV06b] P. FRIZ and N. VICTOIR. A variation embedding theorem and applications. *J. Funct. Anal.* **239**, no. 2, (2006), 631–637. `doi:10.1016/j.jfa.2005.12.021`.

[FV07] P. FRIZ and N. VICTOIR. Large deviation principle for enhanced Gaussian processes. *Ann. Inst. H. Poincaré Probab. Statist.* **43**, no. 6, (2007), 775 – 785. `doi:10.1016/j.anihpb.2006.11.002`.

[FV10a] P. FRIZ and N. VICTOIR. Differential equations driven by Gaussian signals. *Ann. Inst. H. Poincaré Probab. Statist.* **46**, no. 2, (2010), 369–413. `doi:10.1214/09-AIHP202`.

[FV10b] P. FRIZ and N. VICTOIR. *Multidimensional Stochastic Processes as Rough Paths*, vol. 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[FV11] P. FRIZ and N. VICTOIR. A note on higher dimensional p-variation. *Electronic Journal of Probability* **16**, (2011), 1880–1899. `doi:10.1214/EJP.v16-951`.

[GIP12] M. GUBINELLI, P. IMKELLER, and N. PERKOWSKI. Paracontrolled distributions and singular PDEs. *ArXiv e-prints* (2012). `arXiv:1210.2684`.

[GL09] M. GUBINELLI and J. LÖRINCZI. Gibbs measures on Brownian currents. *Comm. Pure Appl. Math.* **62**, no. 1, (2009), 1–56. `doi:10.1002/cpa.20260`.

[GLP99] G. GIACOMIN, J. L. LEBOWITZ, and E. PRESUTTI. Deterministic and stochastic hydrodynamic equations arising from simple microscopic model systems. In *Stochastic partial differential equations: six perspectives*, vol. 64 of *Math. Surveys Monogr.*, 107–152. Amer. Math. Soc., Providence, RI, 1999.

[GT10] M. GUBINELLI and S. TINDEL. Rough evolution equations. *Ann. Probab.* **38**, no. 1, (2010), 1–75. `doi:10.1214/08-AOP437`.

[Gub04] M. GUBINELLI. Controlling rough paths. *J. Funct. Anal.* **216**, no. 1, (2004), 86–140. `doi:10.1016/j.jfa.2004.01.002`.

[Gub10] M. GUBINELLI. Ramification of rough paths. *J. Differential Equations* **248**, no. 4, (2010), 693–721. `doi:10.1016/j.jde.2009.11.015`.

[Gub12] M. GUBINELLI. Rough solutions for the periodic Korteweg–de Vries equation. *Commun. Pure Appl. Anal.* **11**, no. 2, (2012), 709–733. doi:10.3934/cpaa.2012.11.709.

[Hai11a] M. HAIRER. On Malliavin's proof of Hörmander's theorem. *Bull. Sci. Math.* **135**, no. 6-7, (2011), 650–666. doi:10.1016/j.bulsci.2011.07.007.

[Hai11b] M. HAIRER. Rough stochastic PDEs. *Comm. Pure Appl. Math.* **64**, no. 11, (2011), 1547–1585. doi:10.1002/cpa.20383.

[Hai13] M. HAIRER. Solving the KPZ equation. *Ann. of Math. (2)* **178**, no. 2, (2013), 559–664. doi:10.4007/annals.2013.178.2.4.

[Hai14a] M. HAIRER. Introduction to regularity structures. *ArXiv e-prints* (2014). arXiv:1401.3014.

[Hai14b] M. HAIRER. Singular stochastic PDEs. *ArXiv e-prints* (2014). arXiv:1403.6353.

[Hai14c] M. HAIRER. A theory of regularity structures. *Invent. Math.* (2014). doi:10.1007/s00222-014-0505-4.

[HK12] M. HAIRER and D. KELLY. Geometric versus non-geometric rough paths. *ArXiv e-prints* (2012). arXiv:1210.6294. Ann. IHP (B), to appear.

[HM11] M. HAIRER and J. C. MATTINGLY. A theory of hypoellipticity and unique ergodicity for semilinear stochastic PDEs. *Electron. J. Probab.* **16**, (2011), no. 23, 658–738. doi:10.1214/EJP.v16-875.

[HM12] M. HAIRER and J. MAAS. A spatial version of the Itô-Stratonovich correction. *Ann. Probab.* **40**, no. 4, (2012), 1675–1714. doi:10.1214/11-AOP662.

[HMW14] M. HAIRER, J. MAAS, and H. WEBER. Approximating rough stochastic PDEs. *Comm. Pure Appl. Math.* **67**, no. 5, (2014), 776–870. doi:10.1002/cpa.21495.

[HN09] Y. HU and D. NUALART. Rough path analysis via fractional calculus. *Trans. Amer. Math. Soc.* **361**, no. 5, (2009), 2689–2718. doi:10.1090/S0002-9947-08-04631-X.

[Hop50] E. HOPF. The partial differential equation $u_t + uu_x = \mu u_{xx}$. *Comm. Pure Appl. Math.* **3**, (1950), 201–230.

[Hör67] L. HÖRMANDER. Hypoelliptic second order differential equations. *Acta Math.* **119**, (1967), 147–171.

[HP13] M. HAIRER and N. S. PILLAI. Regularity of laws and ergodicity of hypoelliptic SDEs driven by rough paths. *Ann. Probab.* **41**, no. 4, (2013), 2544–2598. doi:10.1214/12-AOP777.

[HQ14] M. HAIRER and J. QUASTEL. Continuous interface growth models rescale to KPZ, 2014. Preprint.

[HSV07] M. HAIRER, A. M. STUART, and J. VOSS. Analysis of SPDEs arising in path sampling. II. The nonlinear case. *Ann. Appl. Probab.* **17**, no. 5-6, (2007), 1657–1706. doi:10.1214/07-AAP441.

[HT13] Y. HU and S. TINDEL. Smooth density for some nilpotent rough differential equations. *J. Theoret. Probab.* **26**, no. 3, (2013), 722–749. doi:10.1007/s10959-011-0388-x.

[HW13] M. HAIRER and H. WEBER. Rough Burgers-like equations with multiplicative noise. *Probab. Theory Related Fields* **155**, no. 1-2, (2013), 71–126. doi:10.1007/s00440-011-0392-1.

[IK06] Y. INAHAMA and H. KAWABI. Large deviations for heat kernel measures on loop spaces via rough paths. *J. London Math. Soc. (2)* **73**, no. 3, (2006), 797–816. doi:10.1112/S0024610706022654.

[Ina10] Y. INAHAMA. A stochastic Taylor-like expansion in the rough path theory. *J. Theor. Probab.* **23**, (2010), 671–714.

[Ina13] Y. INAHAMA. Malliavin differentiability of solutions of rough differential equations. *ArXiv e-prints* (2013). arXiv:1312.7621.

[INY78] N. IKEDA, S. NAKAO, and Y. YAMATO. A class of approximations of Brownian motion. *Publ. Res. Inst. Math. Sci.* **13**, no. 1, (1977/78), 285–300.

[IW89] N. IKEDA and S. WATANABE. *Stochastic differential equations and diffusion processes*. North-Holland Publishing Co., Amsterdam, second ed., 1989.

[JLM85] G. JONA-LASINIO and P. K. MITTER. On the stochastic quantization of field theory. *Comm. Math. Phys.* **101**, no. 3, (1985), 409–436.

[JM83] N. C. JAIN and D. MONRAD. Gaussian measures in B_p. *Ann. Probab.* **11**, no. 1, (1983), 46–57.

[Kal02] O. KALLENBERG. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second ed., 2002.

[KM14] D. KELLY and I. MELBOURNE. Smooth approximation of stochastic differential equations. *ArXiv e-prints* (2014). arXiv:1403.7281.

[Koh78] J. J. KOHN. Lectures on degenerate elliptic problems. In *Pseudodifferential operator with applications (Bressanone, 1977)*, 89–151. Liguori, Naples, 1978.

[KPZ86] M. KARDAR, G. PARISI, and Y.-C. ZHANG. Dynamic scaling of growing interfaces. *Phys. Rev. Lett.* **56**, no. 9, (1986), 889–892.

[KR77] N. V. KRYLOV and B. L. ROZOVSKII. The Cauchy problem for linear stochastic partial differential equations. *Izv. Akad. Nauk SSSR Ser. Mat.* **41**, no. 6, (1977), 1329–1347, 1448.

[KRT07] I. KRUK, F. RUSSO, and C. A. TUDOR. Wiener integrals, Malliavin calculus and covariance measure structure. *J. Funct. Anal.* **249**, no. 1, (2007), 92–142. doi:10.1016/j.jfa.2007.03.031.

[KS84] S. KUSUOKA and D. STROOCK. Applications of the Malliavin calculus. I. In *Stochastic analysis (Katata/Kyoto, 1982)*, vol. 32 of *North-Holland Math. Library*, 271–306. North-Holland, Amsterdam, 1984.

[KS85] S. KUSUOKA and D. STROOCK. Applications of the Malliavin calculus. II. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **32**, no. 1, (1985), 1–76.

[KS87] S. KUSUOKA and D. STROOCK. Applications of the Malliavin calculus. III. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **34**, no. 2, (1987), 391–442.

[Kun82] H. KUNITA. Stochastic partial differential equations connected with nonlinear filtering. In *Nonlinear filtering and stochastic control (Cortona, 1981)*, vol. 972 of *Lecture Notes in Math.*, 100–169. Springer, Berlin, 1982.

[Kus01] S. KUSUOKA. Approximation of expectation of diffusion process and mathematical finance. In *Taniguchi Conference on Mathematics Nara '98*, vol. 31 of *Adv. Stud. Pure Math.*, 147–165. Math. Soc. Japan, Tokyo, 2001.

[LCL07] T. J. LYONS, M. CARUANA, and T. LÉVY. *Differential equations driven by rough paths*, vol. 1908 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6–24, 2004, With an introduction concerning the Summer School by Jean Picard.

[Led96] M. LEDOUX. Isoperimetry and Gaussian analysis. In *Lectures on probability theory and statistics (Saint-Flour, 1994)*, vol. 1648 of *Lecture Notes in Math.*, 165–294. Springer, Berlin, 1996.

[LLQ02] M. LEDOUX, T. LYONS, and Z. QIAN. Lévy area of Wiener processes in Banach spaces. *Ann. Probab.* **30**, no. 2, (2002), 546–578. doi:10.1214/aop/1023481002.

[LN11] T. LYONS and H. NI. Expected signature of Brownian Motion up to the first exit time from a bounded domain. *ArXiv e-prints* (2011). arXiv:1101.5902.

[LPS13] P.-L. LIONS, B. PERTHAME, and P. E. SOUGANIDIS. Scalar conservation laws with rough (stochastic) fluxes. *ArXiv e-prints* (2013). arXiv:1309.1931.

[LQ02] T. LYONS and Z. QIAN. *System control and rough paths*. Oxford Mathematical Monographs. Oxford University Press, Oxford, 2002. Oxford Science Publications.

[LQZ02] M. LEDOUX, Z. QIAN, and T. ZHANG. Large deviations and support theorem for diffusion processes via rough paths. *Stochastic Process. Appl.* **102**, no. 2, (2002), 265–283. doi:10.1016/S0304-4149(02)00176-X.

[LS98a] P.-L. LIONS and P. E. SOUGANIDIS. Fully nonlinear stochastic partial differential equations. *C. R. Acad. Sci. Paris Sér. I Math.* **326**, no. 9, (1998), 1085–1092. doi:10.1016/S0764-4442(98)80067-0.

[LS98b] P.-L. LIONS and P. E. SOUGANIDIS. Fully nonlinear stochastic partial differential equations: non-smooth equations and applications. *C. R. Acad. Sci. Paris Sér. I Math.* **327**, no. 8, (1998), 735–741. doi:10.1016/S0764-4442(98)80161-4.

[LS00a] P.-L. LIONS and P. E. SOUGANIDIS. Fully nonlinear stochastic PDE with semilinear stochastic dependence. *C. R. Acad. Sci. Paris Sér. I Math.* **331**, no. 8, (2000), 617–624. doi:10.1016/S0764-4442(00)00583-8.

[LS00b] P.-L. LIONS and P. E. SOUGANIDIS. Uniqueness of weak solutions of fully nonlinear stochastic partial differential equations. *C. R. Acad. Sci. Paris Sér. I Math.* **331**, no. 10, (2000), 783–790. doi:10.1016/S0764-4442(00)01597-4.

[LS01] W. V. LI and Q.-M. SHAO. Gaussian processes: inequalities, small ball probabilities and applications. In *Stochastic processes: theory and methods*, vol. 19 of *Handbook of Statist.*, 533–597. North-Holland, Amsterdam, 2001.

[LV04] T. LYONS and N. VICTOIR. Cubature on Wiener space. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **460**, no. 2041, (2004), 169–198. doi:10.1098/rspa.2003.1239. Stochastic analysis with applications to mathematical finance.

[LV07] T. LYONS and N. VICTOIR. An extension theorem to rough paths. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **24**, no. 5, (2007), 835–847. doi:10.1016/j.anihpc.2006.07.004.

[LY02] F. LIN and X. YANG. *Geometric measure theory—an introduction*, vol. 1 of *Advanced Mathematics (Beijing/Boston)*. Science Press Beijing, Beijing, 2002.

[Lyo91] T. LYONS. On the nonexistence of path integrals. *Proc. Roy. Soc. London Ser. A* **432**, no. 1885, (1991), 281–290. doi:10.1098/rspa.1991.0017.

[Lyo94] T. LYONS. Differential equations driven by rough signals. I. An extension of an inequality of L. C. Young. *Math. Res. Lett.* **1**, no. 4, (1994), 451–464. doi:10.4310/MRL.1994.v1.n4.a5.

[Lyo98] T. J. LYONS. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana* **14**, no. 2, (1998), 215–310. doi:10.4171/RMI/240.

[Mal78] P. MALLIAVIN. Stochastic calculus of variations and hypoelliptic operators. *Proc. Intern. Symp. SDE* 195–263.

[Mal97] P. MALLIAVIN. *Stochastic analysis*, vol. 313 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1997.

[McK69] H. P. MCKEAN, JR. *Stochastic integrals*. Probability and Mathematical Statistics, No. 5. Academic Press, New York-London, 1969.

[McS72] E. J. MCSHANE. Stochastic differential equations and models of random processes. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. III: Probability theory*, 263–294. Univ. California Press, Berkeley, Calif., 1972.

[Mey92] Y. MEYER. *Wavelets and operators*, vol. 37 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1992. Translated from the 1990 French original by D. H. Salinger.

[MR06] M. B. MARCUS and J. ROSEN. *Markov processes, Gaussian processes, and local times*, vol. 100 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

[MSS06] A. MILLET and M. SANZ-SOLÉ. Large deviations for rough paths of the fractional Brownian motion. *Ann. Inst. H. Poincaré Probab. Statist.* **42**, no. 2, (2006), 245–271.

[Nor86] J. NORRIS. Simplified Malliavin calculus. In *Séminaire de Probabilités, XX, 1984/85*, vol. 1204 of *Lecture Notes in Math.*, 101–130. Springer, Berlin, 1986.

[NP88] D. NUALART and É. PARDOUX. Stochastic calculus with anticipating integrands. *Probab. Theory Related Fields* **78**, no. 4, (1988), 535–581. doi:10.1007/BF00353876.

[NT11] D. NUALART and S. TINDEL. A construction of the rough path above fractional Brownian motion using Volterra's representation. *Ann. Probab.* **39**, no. 3, (2011), 1061–1096. doi:10.1214/10-AOP578.

[Nua06] D. NUALART. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, Berlin, second ed., 2006.

[Par79] E. PARDOUX. Stochastic partial differential equations and filtering of diffusion processes. *Stochastics* **3**, no. 2, (1979), 127–167. doi:10.1080/17442507908833142.

[PS08] G. A. PAVLIOTIS and A. M. STUART. *Multiscale methods*, vol. 53 of *Texts in Applied Mathematics*. Springer, New York, 2008. Averaging and homogenization.

[PW81] G. PARISI and Y. S. WU. Perturbation theory without gauge fixing. *Sci. Sinica* **24**, no. 4, (1981), 483–496.

[Qua11] J. QUASTEL. Introduction to KPZ. *Current Developments in Mathematics* **2011**, (2011), 125–194.

[RX13] S. RIEDEL and W. XU. A simple proof of distance bounds for gaussian rough paths. *Electron. J. Probab.* **18**, (2013), no. 108, 1–18. doi:10.1214/EJP.v18-2387.

[RY91] D. REVUZ and M. YOR. *Continuous martingales and Brownian motion*, vol. 293 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1991.

[Rya02] R. A. RYAN. *Introduction to tensor products of Banach spaces*. Springer Monographs in Mathematics. Springer-Verlag London Ltd., London, 2002.

[SC74] V. N. SUDAKOV and B. S. CIREL′SON. Extremal properties of half-spaces for spherically invariant measures. *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **41**, (1974), 14–24, 165. Problems in the theory of probability distributions, II.

[Sim97] L. SIMON. Schauder estimates by scaling. *Calc. Var. Partial Differential Equations* **5**, no. 5, (1997), 391–407. doi:10.1007/s005260050072.

[Sip93] E.-M. SIPILÄINEN. *A pathwise view of solutions of stochastic differential equations*. Ph.D. thesis, University of Edinburgh, 1993.

[Str11] D. W. STROOCK. *Probability theory*. Cambridge University Press, Cambridge, second ed., 2011. An analytic view.

[SV73] D. W. STROOCK and S. R. S. VARADHAN. Limit theorems for random walks on Lie groups. *Sankhyā Ser. A* **35**, no. 3, (1973), 277–294.

[Tei11] J. TEICHMANN. Another approach to some rough and stochastic partial differential equations. *Stoch. Dyn.* **11**, no. 2-3, (2011), 535–550. doi:10.1142/S0219493711003437.

[Tow02] N. TOWGHI. Multidimensional extension of L. C. Young's inequality. *JIPAM. J. Inequal. Pure Appl. Math.* **3**, no. 2, (2002), Article 22, 13 pp. (electronic).

[Unt10] J. UNTERBERGER. A rough path over multidimensional fractional Brownian motion with arbitrary Hurst index by Fourier normal ordering. *Stochastic Process. Appl.* **120**, no. 8, (2010), 1444–1472. doi:10.1016/j.spa.2010.04.001.

[Wer12] B. M. WERNESS. Regularity of Schramm-Loewner evolutions, annular crossings, and rough path theory. *Electron. J. Probab.* **17**, (2012), no. 81, 21. doi:10.1214/EJP.v17-2331.

[You36] L. C. YOUNG. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Math.* **67**, no. 1, (1936), 251–282.

Index