Rough Path Theory

Lecture Notes

Andrew L. Allan

Abstract

These notes are based on a lecture course I gave at ETH Zürich in Spring semester 2021. They are intended to provide a gentle but rigorous introduction to the theory of rough paths, with a particular focus on their integration theory and associated rough differential equations, and how the theory relates to and enhances the field of stochastic calculus.

The first motivation is to understand the limitations of classical notions of integration to handle paths of very low regularity, and to see how the rough integral succeeds where other notions fail. We then construct rough integrals and establish solutions of differential equations driven by rough paths, as well as the continuity of these objects with respect to the paths involved, and their consistency with stochastic integration and SDEs. Various applications and extensions of the theory are then discussed.

Contents

1 Introduction

Numerous real world phenomena require us to model and analyze systems of controlled[1] differential equations of the form

$$\mathrm{d}Y_t = f(Y_t)\,\mathrm{d}X_t. \tag{1.1}$$

Here f is some (typically nonlinear) function (also known as a "vector field"), $X\colon [0,T] \to \mathbb{R}^d$ is an input signal, and $Y\colon [0,T] \to \mathbb{R}^m$ is the output/solution.

Of course, we haven't yet given a precise meaning to the right-hand side of (1.1). If the signal X is "nice", let's say at least absolutely continuous, then we have the natural interpretation $\mathrm{d}X_t = \frac{\mathrm{d}X_t}{\mathrm{d}t}\,\mathrm{d}t = \dot{X}_t\,\mathrm{d}t$, and we obtain the very classical ODE

$$\dot{Y}_t = f(Y_t)\dot{X}_t.$$

In many applications, particularly in the context of systems affected by random noise, as is commonplace in e.g. engineering and financial applications, the signal X is not so nice, and in particular does not admit a classical derivative $\frac{\mathrm{d}X_t}{\mathrm{d}t}$ (at least not as a function in the usual sense). We therefore need to be a bit cleverer in how we interpret the differential $\mathrm{d}X_t$. A common first step is to rewrite the differential equation (1.1) as the integral equation

$$Y_t = Y_0 + \int_0^t f(Y_s)\,\mathrm{d}X_s, \tag{1.2}$$

and our central question becomes how we should define the integral $\int_0^t f(Y_s)\,\mathrm{d}X_s$.

One way of getting around having to deal directly with complicated objects is to first deal with simpler ones and then take the limit in a suitable topology.

In the context of integration, one might suggest that for general continuous paths X, Y, one could simply take an approximating sequence of smooth paths with $X^n \to X$ and $Y^n \to Y$, and then define $\int_0^t f(Y_s)\,\mathrm{d}X_s$ by the limit $\lim_{n\to\infty} \int_0^t f(Y_s^n)\,\mathrm{d}X_s^n$, since each of the approximations $\int_0^t f(Y_s^n)\,\mathrm{d}X_s^n$ is already well understood.

Although this is in principle possible, the problem is knowing in which topology to take the limit. The obvious choice for continuous functions is to use the topology corresponding to uniform convergence; that is, to take limits with respect to the supremum norm, so that $X^n \to X$ is interpreted as $\sup_{s\in[0,T]} |X_s^n - X_s| \to 0$ as $n \to \infty$. However, this is not sufficient, as the following example demonstrates.

Example 1.1. For each $n \geq 1$, define the functions $X^n, Y^n\colon [0, 2\pi] \to \mathbb{R}$ by

$$X_t^n = -n^{-\frac{1}{3}}\cos(nt), \qquad Y_t^n = n^{-\frac{1}{3}}\sin(nt).$$

Then

$$\int_0^{2\pi} Y_t^n\,\mathrm{d}X_t^n = \int_0^{2\pi} Y_t^n \dot{X}_t^n\,\mathrm{d}t = n^{\frac{1}{3}} \int_0^{2\pi} \sin^2(nt)\,\mathrm{d}t$$
$$= n^{\frac{1}{3}} \int_0^{2\pi} \frac{1}{2}\big(1 - \cos(2nt)\big)\,\mathrm{d}t = n^{\frac{1}{3}}\pi.$$

[1]The description "controlled" differential equation is not particularly enlightening, but is quite common in the literature. It simply refers to the fact that the equation is driven by a signal X, which then determines the behaviour of the solution.

Thus, we have that $X^n \to 0$ and $Y^n \to 0$ uniformly, but $\int_0^{2\pi} Y_t^n \, \mathrm{d}X_t^n \to \infty$ as $n \to \infty$. In particular, the map $(X, Y) \mapsto \int_0^T Y_t \, \mathrm{d}X_t$ is *not* continuous with respect to the supremum norm.

Given a smooth path X and some initial value $y \in \mathbb{R}^d$, let Y be the solution of the equation

$$Y_t = y + \int_0^t f(Y_s) \, \mathrm{d}X_s, \qquad t \in [0, T].$$

Given the above example, it should not be surprising that the map $X \mapsto Y$ is also not continuous in the supremum norm. Clearly, if this strategy is to work we would need a topology considerably stronger than that of uniform convergence.

1.1 Riemann–Stieltjes integration

A sensible first attempt to define the integral in (1.2) is via Riemann–Stieltjes integration. Here and throughout, we shall denote a partition of the time interval $[0, T]$ by $\pi = \{0 = t_0 < t_1 < \cdots < t_N = T\}$. We shall denote the "mesh size" of a partition π by $|\pi| = \max\{|t_{i+1} - t_i| : i = 0, 1, \ldots, N-1\}$.

Let $\pi^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \geq 1$, be a sequence of partitions with vanishing mesh size, i.e. such that $|\pi^n| \to 0$ as $n \to \infty$. For each $n \geq 1$ and $i = 0, 1, \ldots, N-1$, let u_i^n be an arbitrary point in the interval $[t_i^n, t_{i+1}^n]$. The Riemann–Stieltjes integral of Y against X, when it exists, is defined as

$$\int_0^T Y_s \, \mathrm{d}X_s := \lim_{n \to \infty} \sum_{i=0}^{N_n - 1} Y_{u_i^n} (X_{t_{i+1}^n} - X_{t_i^n}),$$

where the limit does not depend on the choice of the sequence of partitions $(\pi^n)_{n \geq 1}$, or on the choice of the intermediate times $u_i^n \in [t_i^n, t_{i+1}^n]$.

If X and Y are two continuous paths, then the Riemann–Stieltjes integral of Y against X exists, for example, whenever at least one of X or Y is Lipschitz continuous or, more generally, of bounded variation over the interval $[0, T]$. One could say: the integral exists provided that at least one of X or Y is "nice". The interested reader may see [Str11] for a thorough explanation of Riemann–Stieltjes integration.

It is worth highlighting the fact that on each interval $[t_i^n, t_{i+1}^n]$ the intermediate time $u_i^n \in [t_i^n, t_{i+1}^n]$ may be chosen arbitrarily. It is quite straightforward to see why this should be the case. Suppose for instance that X were Lipschitz continuous with Lipschitz constant C, and let $u_i^n, v_i^n \in [t_i^n, t_{i+1}^n]$. Then

$$\begin{aligned}
\bigg| \sum_{i=0}^{N_n - 1} (Y_{v_i^n} - Y_{u_i^n})(X_{t_{i+1}^n} - X_{t_i^n}) \bigg| &\leq C \sum_{i=0}^{N_n - 1} |Y_{v_i^n} - Y_{u_i^n}| |t_{i+1}^n - t_i^n| \\
&\leq C \bigg(\sum_{i=0}^{N_n - 1} |t_{i+1}^n - t_i^n| \bigg) \max_{0 \leq i < N_n} |Y_{v_i^n} - Y_{u_i^n}| \\
&= CT \max_{0 \leq i < N_n} |Y_{v_i^n} - Y_{u_i^n}| \longrightarrow 0 \qquad \text{as} \quad n \to \infty,
\end{aligned}$$

where the convergence holds since the mesh size $|\pi^n| \to 0$ and Y is uniformly continuous on the compact interval $[0, T]$. Thus,

$$\begin{aligned}
\lim_{n\to\infty} \sum_{i=0}^{N_n-1} Y_{v_i^n}(X_{t_{i+1}^n} - X_{t_i^n}) &= \lim_{n\to\infty} \sum_{i=0}^{N_n-1} Y_{u_i^n}(X_{t_{i+1}^n} - X_{t_i^n}) + \sum_{i=0}^{N_n-1} (Y_{v_i^n} - Y_{u_i^n})(X_{t_{i+1}^n} - X_{t_i^n}) \\
&= \lim_{n\to\infty} \sum_{i=0}^{N_n-1} Y_{u_i^n}(X_{t_{i+1}^n} - X_{t_i^n}),
\end{aligned}$$

as desired.

For example, we have that

$$\int_0^T Y_s \, \mathrm{d}X_s = \lim_{n\to\infty} \sum_{i=0}^{N_n-1} Y_{t_i^n}(X_{t_{i+1}^n} - X_{t_i^n}) = \lim_{n\to\infty} \sum_{i=0}^{N_n-1} Y_{t_{i+1}^n}(X_{t_{i+1}^n} - X_{t_i^n}),$$

corresponding to choosing the left endpoint $u_i^n = t_i^n$ and right endpoint $u_i^n = t_{i+1}^n$ respectively. Intuitively, the path Y does not vary enough over the interval $[t_i^n, t_{i+1}^n]$ for the value of the integral $\int_0^T Y_s \, \mathrm{d}X_s$ to be affected by whether we take the left endpoint, right endpoint, or any other point in between. As we will see, this property is a luxury that we cannot expect to hold in general for less regular paths X, Y, for which the choice of the intermediate time u_i^n will become crucial.

1.2 Young integration

In general we wish to be able to handle situations where neither X nor Y is particularly nice. When considering paths of low regularity, it's helpful to have a quantitative measure of *how* irregular a given path is. For this purpose, we recall the notion of *Hölder continuity.* For $\alpha \in (0, 1]$, we say that a path $X \colon [0, T] \to \mathbb{R}^d$ is *α-Hölder continuous* if there exists a constant C such that

$$|X_t - X_s| \le C|t - s|^\alpha$$

for all $s, t \in [0, T]$ with $s < t$. Clearly, any Hölder continuous path is continuous, and saying that a path is 1-Hölder continuous is the same as saying that it is Lipschitz continuous.

Theorem 1.2. *Let $\alpha, \beta \in (0, 1]$ such that*

$$\alpha + \beta > 1. \tag{1.3}$$

Let X be α-Hölder continuous, and let Y be β-Hölder continuous. Let $\pi^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \ge 1$, be a sequence of partitions with vanishing mesh size, i.e. such that $|\pi^n| \to 0$ as $n \to \infty$. For each $n \ge 1$ and $i = 0, 1, \ldots, N - 1$, let u_i^n be an arbitrary point in the interval $[t_i^n, t_{i+1}^n]$. Then the limit

$$\int_0^T Y_s \, \mathrm{d}X_s := \lim_{n\to\infty} \sum_{i=0}^{N_n-1} Y_{u_i^n}(X_{t_{i+1}^n} - X_{t_i^n}) \tag{1.4}$$

exists, and does not depend on the choice of the sequence of partitions $(\pi^n)_{n\ge1}$, or on the choice of the intermediate times $u_i^n \in [t_i^n, t_{i+1}^n]$.

A proof of Theorem 1.2 will be given in Section 5.

The limit in (1.4) is called the *Young integral* of Y against X. The key part of the hypothesis of the above theorem is the inequality in (1.3). As we already observed above, if either $\alpha = 1$ or $\beta = 1$ then the limit in (1.4) exists as a Riemann–Stieltjes integral. The point here then is that we can trade off the regularity of the paths X and Y to allow both $\alpha < 1$ and $\beta < 1$, provided that $\alpha + \beta > 1$.

If $\alpha = \beta$ (as is often the case in practice), then the inequality (1.3) becomes

$$\alpha > \frac{1}{2}. \tag{1.5}$$

Thus, Young integration is suitable when the underlying paths are α-Hölder continuous for some α strictly greater than $\frac{1}{2}$.

We note again that here the choice of the intermediate time u_i^n does not affect the value of the integral. However, this relies on the fact that the paths X, Y are continuous. In general this property fails to hold when we allow the paths X, Y to have jumps, i.e. when we drop the continuity assumption, but we will restrict ourselves to continuous paths in this course.

1.3 Stochastic integration

It is expected that the reader is already familiar with the fundamentals of stochastic calculus, so we will not spend much time here to recall the relevant details. However, as the course progresses we will recall the relevant concepts as and when they are needed, so a thorough knowledge of the subject should not be essential.

One of the main motivations for considering paths of low regularity is the study of systems under the influence of stochastic noise. In the most standard setting one supposes that the system noise is generated by a *Brownian motion* W defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \le t \le T}, \mathbb{P})$. Recall that a (standard one-dimensional) Brownian motion is an $\mathbb{R}$-valued adapted stochastic process W with continuous trajectories and stationary independent Gaussian increments, such that

$$W_t - W_s \sim N(0, t - s) \tag{1.6}$$

and $W_t - W_s$ is independent of $\mathcal{F}_s$ for every $s, t \in [0, T]$ with $s < t$.

It follows easily from (1.6) and Kolmogorov's continuity criterion that the trajectories of Brownian motion are almost surely α-Hölder continuous for every $\alpha < \frac{1}{2}$. On the other hand, it can be shown that they are almost surely *not* α-Hölder continuous for any $\alpha \ge \frac{1}{2}$. Recalling the condition (1.5) above, we see that Young integration is not capable of providing an integration theory for Brownian motion.

A very satisfying resolution was provided by the introduction of Itô calculus, which has become a cornerstone of stochastic analysis. Very briefly, given an L^2-bounded continuous martingale M, and a progressively measurable process Y which has sufficient integrability (specifically such that $\mathbb{E}[\int_0^T |Y_s|^2 \,\mathrm{d}\langle M \rangle_s] < \infty$, where $\langle M \rangle$ denotes the quadratic variation of M), one can define the Itô integral $\int_0^T Y_s \,\mathrm{d}M_s$ as a limit in $L^2(\mathbb{P})$ of integrals of simple adapted processes against M. Extensions to more general integrators and integrands then follow by localization arguments. For details, see one of the many available textbooks on stochastic calculus.

Let X be a continuous semimartingale, Y be a left-continuous locally bounded adapted process, and $\pi^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \geq 1$, be a sequence of partitions with vanishing mesh size. Then the Itô integral of Y against X can be expressed as the limit in probability:

$$\sum_{i=0}^{N_n-1} Y_{t_i^n}(X_{t_{i+1}^n} - X_{t_i^n}) \xrightarrow{\mathbb{P}} \int_0^T Y_s \,\mathrm{d}X_s \qquad \text{as} \quad n \to \infty. \tag{1.7}$$

The main point here for us is the fact that the Itô integral is constructed *using probability*. In this sense, it is not a purely analytical theory. Moreover, the limit in (1.7) is only a limit in probability, and does not in general hold almost surely. In other words stochastic integration really is *stochastic*—it is not a “pathwise” theory.

Another important point here is the necessity of taking the left endpoint $Y_{t_i^n}$ of Y in (1.7). Taking a different choice of endpoint here will in general change the value of the integral. Another common choice is to take the average of the left and right endpoints; that is, to replace $Y_{t_i^n}$ in (1.7) by $\frac{1}{2}(Y_{t_i^n} + Y_{t_{i+1}^n})$. This gives an alternative definition of stochastic integral, known as the *Stratonovich integral*:

$$\int_0^T Y_s \circ \mathrm{d}X_s = \lim_{n\to\infty} \sum_{i=0}^{N_n-1} \frac{1}{2}(Y_{t_i^n} + Y_{t_{i+1}^n})(X_{t_{i+1}^n} - X_{t_i^n}), \tag{1.8}$$

which exists as a limit in probability.

Note that

$$\begin{aligned}
\int_0^T Y_s \circ \mathrm{d}X_s &= \lim_{n\to\infty} \sum_{i=0}^{N_n-1} \frac{1}{2}(Y_{t_i^n} + Y_{t_{i+1}^n})(X_{t_{i+1}^n} - X_{t_i^n}) \\
&= \lim_{n\to\infty} \sum_{i=0}^{N_n-1} Y_{t_i^n}(X_{t_{i+1}^n} - X_{t_i^n}) + \frac{1}{2} \sum_{i=0}^{N_n-1} (Y_{t_{i+1}^n} - Y_{t_i^n})(X_{t_{i+1}^n} - X_{t_i^n}) \\
&= \int_0^T Y_s \,\mathrm{d}X_s + \frac{1}{2}\langle Y, X\rangle_T,
\end{aligned}$$

where $\langle Y, X\rangle$ is the quadratic covariation of Y and X. For example, if $X = Y = W$ for a Brownian motion W, then we have

$$\int_0^T W_s \circ \mathrm{d}W_s = \int_0^T W_s \,\mathrm{d}W_s + \frac{T}{2}.$$

Thus, the answer to the question “What is the value of the integral of Y against X?” depends crucially on which notion of integral one chooses. One should also recognise that both these types of integral are perfectly valid; neither of them is the “correct” choice in general, and the integral one chooses typically depends on the application one has in mind. In financial and biological applications it is typically better from a modelling perspective to choose the Itô integral, which also often proves useful due to the fact that it preserves the martingale property. On the other hand, the Stratonovich integral is arguably more natural from an abstract calculus perspective, as it satisfies the classical integration by parts and chain rules, or "first order calculus", which is not true of the Itô integral. We shall revisit these ideas and explore them in more precise detail later in this course.

1.4 Rough integration

All of the integrals we have discussed above essentially start from the basic notion of constructing integrals as limits of "Riemann sums". That is, we try to define the integral of a path Y against another path X via

$$\lim_{|\pi|\to 0} \sum_{[s,t]\in\pi} Y_u(X_t - X_s), \tag{1.9}$$

where $u \in [s,t]$, and the limit is taken over any sequence of partitions $(\pi^n)_{n\geq 1}$ with vanishing mesh size. (Here we abuse notation slightly by writing $[s,t] \in \pi$, but the meaning should be clear.) However, for general paths X, Y which do not satisfy the Young condition (1.3), this limit may not exist, or the limit may depend on the choice of sequence of partitions, in which case it is unclear that any particular limit is actually meaningful.

Moreover, we saw in the context of stochastic integration that even when the limit does exist, it may depend crucially on how we select the intermediate points $u \in [s,t]$. Intuitively, the paths X and Y vary so rapidly during the small time interval $[s,t]$ that a simple Riemann sum, as in (1.9), is not enough to capture these rapid variations. As we will see, there is, in a certain sense, a lack of information. To resolve this, we will now look more closely at what happens over a small time interval.

Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a smooth function, and let $X = (X^1, \dots, X^d)\colon [0,T] \to \mathbb{R}^d$ be an α-Hölder continuous path for some $\alpha \in (0,1]$. Suppose that we wish to integrate the path $f(X)$ against X itself. That is, we wish to give a meaning to $\int_0^T f(X_r)\,\mathrm{d}X_r$. Let $[s,t] \subset [0,T]$ be a "small" time interval, and let $r \in [s,t]$. By Taylor expansion, we have that

$$f(X_r) = f(X_s) + Df(X_s)(X_r - X_s) + \dots$$

where Df denotes the gradient of f. Integrating with respect to X, we obtain

$$\int_s^t f(X_r)\,\mathrm{d}X_r = f(X_s)(X_t - X_s) + Df(X_s)\int_s^t (X_r - X_s) \otimes \mathrm{d}X_r + \dots$$

Note that, given two vectors $x, y \in \mathbb{R}^d$, the notation $x \otimes y$, known as the *tensor product* of x and y, is used to denote the $d \times d$-matrix with (i,j)-entry given by $[x \otimes y]^{ij} = x^i y^j$. For clarity, we rewrite the above in component form: for $j = 1, \dots, d$, we have

$$\int_s^t f(X_r)\,\mathrm{d}X_r^j = f(X_s)(X_t^j - X_s^j) + \sum_{i=1}^d \partial_i f(X_s)\int_s^t (X_r^i - X_s^i)\,\mathrm{d}X_r^j + \dots$$

It turns out that, provided $\alpha > \frac{1}{3}$, the higher order terms we have omitted in the above expansion vanish upon applying $\lim_{|\pi|\to 0}\sum_{[s,t]\in\pi}$. In fact, if $\alpha > \frac{1}{2}$ (recall the Young condition (1.5)), then one can show that

$$\lim_{|\pi|\to 0} \sum_{[s,t]\in\pi} \int_s^t (X_r - X_s) \otimes \mathrm{d}X_r = 0. \tag{1.10}$$

In this case we simply obtain

$$\int_0^T f(X_r)\,\mathrm{d}X_r = \lim_{|\pi|\to 0} \sum_{[s,t]\in\pi} \int_s^t f(X_r)\,\mathrm{d}X_r = \lim_{|\pi|\to 0} \sum_{[s,t]\in\pi} f(X_s)(X_t - X_s),$$

which we recognise as the definition of the Young integral of $f(X)$ against X.

However, when $\alpha \leq \frac{1}{2}$ the convergence in (1.10) does not necessarily hold, and this "second order" term remains:

$$\begin{aligned}\int_0^T f(X_r)\,\mathrm{d}X_r &= \lim_{|\pi|\to 0} \sum_{[s,t]\in\pi} \int_s^t f(X_r)\,\mathrm{d}X_r \\ &= \lim_{|\pi|\to 0} \sum_{[s,t]\in\pi} \left(f(X_s)(X_t - X_s) + Df(X_s) \int_s^t (X_r - X_s) \otimes \mathrm{d}X_r \right)\end{aligned}$$

This suggests that, for $\alpha \in (\frac{1}{3}, \frac{1}{2}]$, in order to compute the integral of $f(X)$ against X, we need as inputs both the path increments $X_t - X_s$ as well as the integrals $\int_s^t (X_r - X_s) \otimes \mathrm{d}X_r$ for each pair of times $s < t$. We therefore make the definition:

$$\mathbb{X}_{s,t} =: \int_s^t (X_r - X_s) \otimes \mathrm{d}X_r. \tag{1.11}$$

Of course, for $\alpha \leq \frac{1}{2}$, the integral on the right-hand side is not generally understood (at least without probability). However, as we will see, rough path theory will tell us precisely the features of the integral which are actually necessary, and these features will be expressed as conditions which must be satisfied by $\mathbb{X}$. We should therefore think of the object $\mathbb{X}$, sometimes referred to as the "enhancement" or the "lift" of X, as providing a "candidate" for the value of the integral. At first glance one may have presumed that the symbol =: in (1.11) was a typo, and that the left-hand side should be defined by the right-hand side, but this is not the case.

By a *rough path*, we mean the pair $(X, \mathbb{X})$. But to be clear, when we come later to the proper definition of a rough path we will abandon the equality in (1.11), and instead provide analytical and algebraic conditions which must be satisfied by $\mathbb{X}$. One should therefore just think of (1.11) as motivation for the "information" encoded by $\mathbb{X}$.

We observed earlier that if we define Y as the solution of the equation

$$Y_t = y + \int_0^t f(Y_s)\,\mathrm{d}X_s, \qquad t \in [0,T], \tag{1.12}$$

for a smooth path X, then the solution map $X \mapsto Y$ is not continuous. It turns out that if we interpret the integral in (1.12) as a "rough integral" against the pair $(X, \mathbb{X})$ (which we will define properly later), then the solution map $(X, \mathbb{X}) \mapsto Y$ is continuous with respect to a suitable rough path topology.

Solving such a "rough differential equation" thus essentially involves finding the two mappings:

$$X \longmapsto (X, \mathbb{X}) \longmapsto Y.$$

The first of these maps involves adding new information, and hence depends on the particular problem one is trying to study. There is sometimes some work involved in constructing a suitable rough path lift $\mathbb{X}$, but, as we will see, there are many situations where this lift can be obtained very naturally.

Given this lift, the second map—known as the *Itô–Lyons map*—is then continuous, and in standard situations is even locally Lipschitz continuous.

2 Hölder spaces

2.1 Basic properties

For brevity, given a path $X\colon [0,T] \to \mathbb{R}^d$ and a pair of times $s, t \in [0,T]$, we write

$$X_{s,t} = X_t - X_s$$

for the increment of X from time s to time t.

We write $\mathcal{C} = \mathcal{C}([0,T];\mathbb{R}^d)$ for the space of continuous paths $X\colon [0,T] \to \mathbb{R}^d$. We will sometimes write $\|X\|_\infty = \sup_{t\in[0,T]} |X_t|$ for the supremum norm.

Definition 2.1. For $\alpha \in (0,1]$ we define the α-Hölder seminorm of a path $X\colon [0,T] \to \mathbb{R}^d$ by

$$\|X\|_\alpha = \sup_{0\le s<t\le T} \frac{|X_{s,t}|}{|t-s|^\alpha}.$$

We define the space of α-Hölder continuous paths as the family of paths X such that $\|X\|_\alpha < \infty$. We denote this space by $\mathcal{C}^\alpha = \mathcal{C}^\alpha([0,T];\mathbb{R}^d)$.

Note that $\|\cdot\|_\alpha$ is only a seminorm, as it does not distinguish between additive constants. We can obtain a genuine norm via the map

$$X \mapsto |X_0| + \|X\|_\alpha.$$

Equipped with this norm, $\mathcal{C}^\alpha$ becomes a Banach space (i.e. a normed vector space such that the norm is complete, meaning that every Cauchy sequence converges to a limit within the space).

It is easy to see that if $0 < \alpha < \beta \le 1$ then $\mathcal{C}^\beta \subset \mathcal{C}^\alpha$, and that this inclusion is strict. It turns out that the Hölder spaces $\mathcal{C}^\alpha$ are not separable.

Lemma 2.2 (Lower semi-continuity)**.** *Let $\alpha \in (0,1]$ and $X\colon [0,T] \to \mathbb{R}^d$. Let $(X^n)_{n\ge1} \subset \mathcal{C}^\alpha$ be a sequence of α-Hölder continuous paths and assume that $X^n \to X$ pointwise. Then*

$$\|X\|_\alpha \le \liminf_{n\to\infty} \|X^n\|_\alpha.$$

Proof. Simply note that for all $0 \le s < t \le T$, we have

$$\frac{|X_{s,t}|}{|t-s|^\alpha} = \liminf_{n\to\infty} \frac{|X^n_{s,t}|}{|t-s|^\alpha} \le \liminf_{n\to\infty} \|X^n\|_\alpha,$$

and take the supremum over $0 \le s < t \le T$. ∎

Lemma 2.3 (Interpolation)**.** *Let $0 < \alpha < \beta \le 1$ and let $X\colon [0,T] \to \mathbb{R}^d$. Then*

$$\|X\|_\alpha \le \|X\|_\beta^{\frac{\alpha}{\beta}} \Big(\sup_{0 \le s < t \le T} |X_{s,t}| \Big)^{1-\frac{\alpha}{\beta}}.$$

Proof. Note that

$$\frac{|X_{s,t}|}{|t-s|^\alpha} = \bigg(\frac{|X_{s,t}|}{|t-s|^\beta} \bigg)^{\frac{\alpha}{\beta}} |X_{s,t}|^{1-\frac{\alpha}{\beta}}.$$

Taking the supremum over $0 \le s < t \le T$, we obtain the desired inequality. □

Lemma 2.4. *Let $0 < \alpha < \beta \le 1$ and let $X \in \mathcal{C}$ be a continuous path. Let $(X^n)_{n \ge 1} \subset \mathcal{C}^\beta$ be a sequence of β-Hölder continuous paths and assume that $\sup_{n \ge 1} \|X^n\|_\beta < \infty$. If $X^n \to X$ uniformly, then $X \in \mathcal{C}^\beta$, and $\|X^n - X\|_\alpha \to 0$ as $n \to \infty$.*

Proof. By Lemma 2.3, we have that

$$\|X^n - X\|_\alpha \le \|X^n - X\|_\beta^{\frac{\alpha}{\beta}} \Big(\sup_{0 \le s < t \le T} |X^n_{s,t} - X_{s,t}| \Big)^{1-\frac{\alpha}{\beta}}. \tag{2.1}$$

By Lemma 2.2, we know that

$$\|X\|_\beta \le \liminf_{n \to \infty} \|X^n\|_\beta \le \sup_{n \ge 1} \|X^n\|_\beta < \infty,$$

and hence that $\sup_{n \ge 1} \|X^n - X\|_\beta < \infty$. Since $X^n \to X$ uniformly, it follows that the right-hand side of (2.1) tends to zero as $n \to \infty$. □

Lemma 2.5 (Compactness)**.** *Let $0 < \alpha < \beta \le 1$. Let $(X^n)_{n \ge 1} \subset \mathcal{C}^\beta$ be a sequence of β-Hölder continuous paths and assume that*

$$\sup_{n \ge 1} \big(|X^n_0| + \|X^n\|_\beta \big) < \infty. \tag{2.2}$$

Then there exists a path $X \in \mathcal{C}^\beta$ and a subsequence $(n_k)_{k \ge 1}$ such that $\|X^{n_k} - X\|_\alpha \to 0$ as $k \to \infty$.

Proof. It follows from (2.2) that the sequence of paths $(X^n)_{n \ge 1}$ is uniformly bounded and uniformly equicontinuous. It therefore follows from Arzelà–Ascoli (see e.g. [FV10, Theorem 1.4]) that there exists a continuous path X and a subsequence $(n_k)_{k \ge 1}$ such that $X^{n_k} \to X$ uniformly. The result then follows from Lemma 2.4. □

The result of Lemma 2.5 shows that $\mathcal{C}^\beta$ is compactly embedded in $\mathcal{C}^\alpha$ whenever $0 < \alpha < \beta \le 1$. That is, any bounded subset of $\mathcal{C}^\beta$ is relatively compact in $\mathcal{C}^\alpha$.

2.2 The closure of smooth paths

Definition 2.6. For $\alpha \in (0,1]$, we define $\mathcal{C}^{0,\alpha} = \mathcal{C}^{0,\alpha}([0,T];\mathbb{R}^d)$ to be the closure of the space of smooth paths from $[0,T] \to \mathbb{R}^d$ with respect to the α-Hölder seminorm.

It is clear that $\mathcal{C}^{0,\alpha}$ is a closed linear subspace of $\mathcal{C}^\alpha$, and thus is itself a Banach space.

Proposition 2.7. *Let $\alpha \in (0,1)$ and let $X\colon [0,T] \to \mathbb{R}^d$ be a path. Then $X \in \mathcal{C}^{0,\alpha}$ if and only if*

$$\lim_{\delta \to 0} \sup_{|t-s|<\delta} \frac{|X_{s,t}|}{|t-s|^\alpha} = 0. \tag{2.3}$$

Proof. ($\Longrightarrow$) Suppose first that $X \in \mathcal{C}^{0,\alpha}$. Let $\varepsilon > 0$. By the definition of $\mathcal{C}^{0,\alpha}$, there exists a smooth path Y such that $\|X - Y\|_\alpha < \frac{\varepsilon}{2}$.

Since Y is smooth, it is Lipschitz continuous. Let $L \geq 0$ be the Lipschitz constant of Y. Let $\delta > 0$ be sufficiently small such that $L\delta^{1-\alpha} < \frac{\varepsilon}{2}$. (Note that this would not work for $\alpha = 1$.)

Then, for any $s < t$ with $|t-s| < \delta$, we have

$$\begin{aligned}
\frac{|X_{s,t}|}{|t-s|^\alpha} &\leq \frac{|X_{s,t} - Y_{s,t}|}{|t-s|^\alpha} + \frac{|Y_{s,t}|}{|t-s|^\alpha} \\
&\leq \|X - Y\|_\alpha + L|t-s|^{1-\alpha} \\
&\leq \|X - Y\|_\alpha + L\delta^{1-\alpha} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

Thus,

$$\sup_{|t-s|<\delta} \frac{|X_{s,t}|}{|t-s|^\alpha} \leq \varepsilon,$$

and we deduce that (2.3) holds.

($\Longleftarrow$) Now suppose instead that (2.3) holds. Let $\varepsilon > 0$. By assumption, there exists a $\delta > 0$ such that

$$\sup_{|t-s|<\delta} \frac{|X_{s,t}|}{|t-s|^\alpha} < \varepsilon. \tag{2.4}$$

We can assume without loss of generality that $\delta \leq 1$.

Let $\pi = \{0 = u_0 < u_1 < \cdots < u_N = T\}$ be a partition of the interval $[0,T]$ with equidistant points, so that $u_i = iT/N$ for each $i = 0, 1, \ldots, N$, and such that the mesh size $|\pi| = |u_{i+1} - u_i| < \delta$.

Consider the piecewise linear approximation of X, which is equal to X at each of the points u_i in the partition π, and linear on the interval $[u_i, u_{i+1}]$ for each $i = 0, 1, \ldots, N-1$. Note that this approximation is Lipschitz continuous with Lipschitz constant equal to

$$\max_{0 \leq i < N} \frac{|X_{u_i,u_{i+1}}|}{|u_{i+1} - u_i|} = \frac{1}{|\pi|} \max_{0 \leq i < N} |X_{u_i,u_{i+1}}|.$$

Note that we can smooth out this path in a small neighbourhood of each of the points u_i in π, whilst only increasing the Lipschitz constant by an arbitrarily small amount. Thus, there exists a smooth path Y such that

- $Y_{u_i} = X_{u_i}$ for every $i = 0, 1, \ldots, N$,
- and the Lipschitz constant L of Y satisfies

$$L \leq \left(\frac{1}{|\pi|} \max_{0 \leq i < N} |X_{u_i,u_{i+1}}| \right) + \varepsilon.$$

Since $|u_{i+1} - u_i| = |\pi| < \delta$ for each i, it follows from (2.4) that

$$\max_{0 \le i < N} \frac{|X_{u_i,u_{i+1}}|}{|u_{i+1} - u_i|^\alpha} < \varepsilon,$$

and hence that

$$L \le \left(\frac{1}{|\pi|^{1-\alpha}} \max_{0 \le i < N} \frac{|X_{u_i,u_{i+1}}|}{|u_{i+1} - u_i|^\alpha}\right) + \varepsilon < \varepsilon\left(\frac{1}{|\pi|^{1-\alpha}} + 1\right). \tag{2.5}$$

Let $0 \le s < t \le T$. Let j, k be such that $s \in [u_j, u_{j+1})$ and $t \in (u_k, u_{k+1}]$. In particular, we must have that $j \le k$. We will deal with the cases $j = k$ and $j < k$ separately.

If $j = k$, then $|t - s| \le |u_{j+1} - u_j| = |\pi| < \delta$, so it follows immediately from (2.4) that

$$\frac{|X_{s,t}|}{|t - s|^\alpha} < \varepsilon.$$

Moreover, using (2.5), we have

$$\frac{|Y_{s,t}|}{|t - s|^\alpha} \le L|t - s|^{1-\alpha} \le \varepsilon\left(\frac{1}{|\pi|^{1-\alpha}} + 1\right)|t - s|^{1-\alpha} \le \varepsilon\big(1 + |t - s|^{1-\alpha}\big) \le 2\varepsilon,$$

where we used the fact that $|t - s|^{1-\alpha} \le |\pi|^{1-\alpha} < \delta^{1-\alpha} \le 1$. Hence, in this case we have

$$\frac{|X_{s,t} - Y_{s,t}|}{|t - s|^\alpha} \le \frac{|X_{s,t}|}{|t - s|^\alpha} + \frac{|Y_{s,t}|}{|t - s|^\alpha} < \varepsilon + 2\varepsilon = 3\varepsilon. \tag{2.6}$$

If $j < k$, then we have

$$\begin{aligned} X_{s,t} - Y_{s,t} &= (X_{s,u_{j+1}} + X_{u_{j+1},u_k} + X_{u_k,t}) - (Y_{s,u_{j+1}} + Y_{u_{j+1},u_k} + Y_{u_k,t}) \\ &= X_{s,u_{j+1}} - Y_{s,u_{j+1}} + X_{u_k,t} - Y_{u_k,t}. \end{aligned} \tag{2.7}$$

Since $u_j \le s < u_{j+1} \le u_k < t$, we have that $|u_{j+1} - s| < |t - s|$, and $|u_{j+1} - s| \le |u_{j+1} - u_j| = |\pi| < \delta$. Thus, using (2.4), we have

$$\frac{|X_{s,u_{j+1}}|}{|t - s|^\alpha} \le \frac{|X_{s,u_{j+1}}|}{|u_{j+1} - s|^\alpha} < \varepsilon.$$

Using (2.5), we also have

$$\frac{|Y_{s,u_{j+1}}|}{|t - s|^\alpha} \le L|u_{j+1} - s|^{1-\alpha} < \varepsilon\left(\frac{1}{|\pi|^{1-\alpha}} + 1\right)|\pi|^{1-\alpha} = \varepsilon\big(1 + |\pi|^{1-\alpha}\big) \le 2\varepsilon.$$

Dealing with the terms $X_{u_k,t}$ and $Y_{u_k,t}$ similarly, we deduce from (2.7) that

$$\frac{|X_{s,t} - Y_{s,t}|}{|t - s|^\alpha} \le \frac{|X_{s,u_{j+1}}|}{|t - s|^\alpha} + \frac{|Y_{s,u_{j+1}}|}{|t - s|^\alpha} + \frac{|X_{u_k,t}|}{|t - s|^\alpha} + \frac{|Y_{u_k,t}|}{|t - s|^\alpha} < 6\varepsilon. \tag{2.8}$$

It follows from (2.6) and (2.8) that $\|X - Y\|_\alpha \le 6\varepsilon$. Since Y is smooth and $\varepsilon > 0$ was arbitrary, we have that $X \in \mathcal{C}^{0,\alpha}$. □

Example 2.8. Let $\alpha \in (0, 1)$ and let $X \colon [0, T] \to \mathbb{R}$ be the path given by $X_t = t^\alpha$. Then $X \in \mathcal{C}^\alpha$, but $X \notin \mathcal{C}^{0,\alpha}$.

Lemma 2.9. *Let $0 < \alpha < \beta \le 1$. Then $\mathcal{C}^\beta \subset \mathcal{C}^{0,\alpha}$.*

Proof. Let $X \in \mathcal{C}^\beta$. Let $\delta > 0$, and let $s < t$ such that $|t-s| < \delta$. Then

$$\frac{|X_{s,t}|}{|t-s|^\alpha} = \frac{|X_{s,t}|}{|t-s|^\beta}|t-s|^{\beta-\alpha} \le \|X\|_\beta \delta^{\beta-\alpha},$$

and we see that

$$\lim_{\delta\to 0} \sup_{|t-s|<\delta} \frac{|X_{s,t}|}{|t-s|^\alpha} = 0.$$

By Proposition 2.7, we have that $X \in \mathcal{C}^{0,\alpha}$. □

We conclude that, for $0 < \alpha < \beta \le 1$, we have

$$\mathcal{C}^\beta \subset \mathcal{C}^{0,\alpha} \subset \mathcal{C}^\alpha \tag{2.9}$$

and that each of these inclusions is strict.

Remark 2.10. It turns out that the closure of smooth paths in $\mathcal{C}^1$ is equal to the space of continuously differentiable paths. The key to seeing this is to show that $\|X\|_1 = \sup_{t\in[0,T]} |\dot{X}_t|$ for any continuously differentiable path X. It then follows that a sequence of smooth paths is Cauchy with respect to the 1-Hölder norm $X \mapsto |X_0| + \|X\|_1$ if and only if it is Cauchy with respect to the norm $X \mapsto |X_0| + \sup_{t\in[0,T]} |\dot{X}_t|$, which we note is a norm on the space of continuously differentiable paths.

2.3 Two-parameter functions

As well as paths defined on the interval $[0,T]$, we will also consider two-parameter functions defined on

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

We will denote by $\mathcal{C}_2 = \mathcal{C}_2(\Delta_{[0,T]}; E)$ the space of continuous functions from $\Delta_{[0,T]} \to E$, where E will typically be either $\mathbb{R}^m$ or $\mathbb{R}^{d\times d}$.

The notion of Hölder continuity is also valid for such two-parameter functions. For $A\colon \Delta_{[0,T]} \to E$, we similarly define

$$\|A\|_\alpha = \sup_{0\le s<t\le T} \frac{|A_{s,t}|}{|t-s|^\alpha}.$$

We will denote the space of α-Hölder continuous functions on $\Delta_{[0,T]}$ by $\mathcal{C}_2^\alpha$.

To avoid confusion, we stress that if X is a path then $X_{s,t}$ means the increment $X_t - X_s$, but if A is a two-parameter function defined on $\Delta_{[0,T]}$ then $A_{s,t}$ just means A evaluated at the pair of times $(s,t) \in \Delta_{[0,T]}$.

Although a path which is α-Hölder continuous for some $\alpha > 1$ is necessarily equal to a constant, it is perfectly possible to have non-trivial functions $A \in \mathcal{C}_2^\alpha$ for $\alpha > 1$. Note, however, that in this case we have

$$\sum_{[s,t]\in\pi} |A_{s,t}| \le \|A\|_\alpha \sum_{[s,t]\in\pi} |t-s|^\alpha \le \|A\|_\alpha \Bigg(\sum_{[s,t]\in\pi} |t-s|\Bigg)|\pi|^{\alpha-1} = \|A\|_\alpha T|\pi|^{\alpha-1},$$

which vanishes upon letting the mesh size $|\pi| \to 0$.

3 The space of rough paths

3.1 Basic definitions

Definition 3.1. For $\alpha \in (\frac{1}{3}, \frac{1}{2}]$, an *$\alpha$-Hölder rough path* (over $\mathbb{R}^d$) is a pair $\mathbf{X} = (X, \mathbb{X})$, where $X \colon [0,T] \to \mathbb{R}^d$ is an α-Hölder continuous path, $\mathbb{X} \colon \Delta_{[0,T]} \to \mathbb{R}^{d\times d}$ is 2α-Hölder continuous, and such that *Chen's relation:*

$$\mathbb{X}_{s,t} = \mathbb{X}_{s,u} + \mathbb{X}_{u,t} + X_{s,u} \otimes X_{u,t} \tag{3.1}$$

holds for all $0 \leq s \leq u \leq t \leq T$. We shall denote the space of α-Hölder rough paths by $\mathscr{C}^\alpha = \mathscr{C}^\alpha([0,T]; \mathbb{R}^d)$.

Note that in component form, Chen's relation states that

$$\mathbb{X}^{ij}_{s,t} = \mathbb{X}^{ij}_{s,u} + \mathbb{X}^{ij}_{u,t} + X^i_{s,u} X^j_{u,t} \tag{3.2}$$

for each $1 \leq i,j \leq d$.

Thus, a rough path is an element $\mathbf{X} = (X, \mathbb{X}) \in \mathcal{C}^\alpha \times \mathcal{C}^{2\alpha}_2$ such that the algebraic condition (3.1) holds. We will refer to $\mathbb{X}$ as the "lift" or "enhancement" of X, and think of a rough path $\mathbf{X}$ as a path X which has been "lifted" or "enhanced" by the addition of $\mathbb{X}$.

Let us briefly discuss this definition. First, we recall from the introduction that Young integration provides an adequate integration theory for α-Hölder continuous paths when $\alpha > \frac{1}{2}$. We are therefore interested in cases when $\alpha \leq \frac{1}{2}$. As also indicated earlier, it will turn out that the framework we focus on here is not sufficient to deal with the case when $\alpha \leq \frac{1}{3}$, so we will restrict ourselves to $\alpha > \frac{1}{3}$. This is not a worry for us, as many very interesting and important situations fit nicely into the regime $\alpha \in (\frac{1}{3}, \frac{1}{2}]$. We will briefly discuss extensions to $\alpha \leq \frac{1}{3}$ later in Section 12.

As discussed in the introduction, one should think of $\mathbb{X}_{s,t}$ as *postulating* the value of the integral

$$\int_s^t X_{s,r} \otimes \mathrm{d}X_r.$$

That is, the (i,j)-component $\mathbb{X}^{ij}_{s,t}$ corresponds to the integral $\int_s^t X^i_{s,r}\,\mathrm{d}X^j_r$. Note that $\mathbb{X}$ is more regular than X, assumed to be 2α-Hölder continuous. We will see examples later to justify this condition, but for now it can be taken on trust that this is a sensible assumption.

A simple but important special case is when the path X is smooth. In this case we can simply define $\mathbb{X}_{s,t} := \int_s^t X_{s,r} \otimes \mathrm{d}X_r$, with the integral being defined in the Riemann–Stieltjes sense. It is then easy to verify that X and $\mathbb{X}$ do indeed satisfy Chen's relation (3.1), and thus that $\mathbf{X} := (X, \mathbb{X})$ is a rough path for any $\alpha \in (\frac{1}{3}, \frac{1}{2}]$.

Definition 3.2. Given rough paths $\mathbf{X} = (X, \mathbb{X}), \tilde{\mathbf{X}} = (\tilde{X}, \tilde{\mathbb{X}}) \in \mathscr{C}^\alpha$, we define the α-Hölder rough path distance by

$$\begin{aligned}\|\mathbf{X}; \tilde{\mathbf{X}}\|_\alpha &= \|X - \tilde{X}\|_\alpha + \|\mathbb{X} - \tilde{\mathbb{X}}\|_{2\alpha}\\ &= \sup_{0\leq s<t\leq T} \frac{|X_{s,t} - \tilde{X}_{s,t}|}{|t-s|^\alpha} + \sup_{0\leq s<t\leq T} \frac{|\mathbb{X}_{s,t} - \tilde{\mathbb{X}}_{s,t}|}{|t-s|^{2\alpha}}.\end{aligned}$$

Note that $(\mathbf{X}, \tilde{\mathbf{X}}) \mapsto \|\mathbf{X}; \tilde{\mathbf{X}}\|_\alpha$ is a pseudometric; that is, it satisfies the usual conditions of being a metric, except that $\|\mathbf{X}; \tilde{\mathbf{X}}\|_\alpha = 0$ does not necessarily imply that $\mathbf{X} = \tilde{\mathbf{X}}$. However, the map

$$(\mathbf{X}, \tilde{\mathbf{X}}) \mapsto |X_0 - \tilde{X}_0| + \|\mathbf{X}; \tilde{\mathbf{X}}\|_\alpha$$

does define a genuine metric.

It is not hard to see that $\mathcal{C}^\alpha \times \mathcal{C}_2^{2\alpha}$ is a Banach space with norm $(X, \mathbb{X}) \mapsto |X_0| + |||\mathbf{X}|||_\alpha$, where

$$|||\mathbf{X}|||_\alpha := \|X\|_\alpha + \|\mathbb{X}\|_{2\alpha}.$$

Note however that since Chen's relation (3.1) is nonlinear, the space of rough paths $\mathscr{C}^\alpha$ is not even a vector space. Nevertheless, it is a closed subset of $\mathcal{C}^\alpha \times \mathcal{C}_2^{2\alpha}$. The space $\mathscr{C}^\alpha$ equipped with the metric $(\mathbf{X}, \tilde{\mathbf{X}}) \mapsto |X_0 - \tilde{X}_0| + \|\mathbf{X}; \tilde{\mathbf{X}}\|_\alpha$ is therefore a complete metric space.

Let $(X, \mathbb{X})$ be a rough path, and let $F \in \mathcal{C}^{2\alpha}$. Note that if we let $\tilde{\mathbb{X}}_{s,t} = \mathbb{X}_{s,t} + F_{s,t}$ for all $(s,t) \in \Delta_{[0,T]}$, then the pair $(X, \tilde{\mathbb{X}})$ still satisfies Chen's relation, and is thus also a rough path. We infer that, given a rough path $(X, \mathbb{X})$, the enhancement $\mathbb{X}$ is never unique.

On the other hand, suppose now that both $(X, \mathbb{X})$ and $(X, \tilde{\mathbb{X}})$ are rough paths with the same underlying path X, and let $G = \mathbb{X} - \tilde{\mathbb{X}}$. It then follows from Chen's relation that

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

for $s \le u \le t$, so that in particular $G_{s,t} = G_{0,t} - G_{0,s}$ for every $s \le t$. That is, $G_{s,t}$ is actually just the increment of the path given by $t \mapsto G_{0,t}$.

It follows that for any rough path $(X, \mathbb{X})$, the enhancement $\mathbb{X}$ is determined up to the addition of the increments of some path $F \in \mathcal{C}^{2\alpha}$. The choice of F does matter, and there is in general no obvious canonical choice. However, as we will see, there are important examples where such a canonical choice does exist.

The fact that the enhancement $\mathbb{X}$ is not unique should not be too surprising, given the discussion in the introduction. We think of $\mathbb{X}_{s,t}$ as postulating the value of the integral $\int_s^t X_{s,r} \otimes \mathrm{d}X_r$. But recall that, particularly in the setting of stochastic integration, the value of such an integral depends on the choice of the intermediate point in the definition of the integral, and that therefore there are in general multiple different ways of defining such an integral. For example, we saw that the Itô and Stratonovich integrals give two different but equally valid interpretations of a stochastic integral. The choice of the enhancement $\mathbb{X}$ corresponds, in a meaningful sense, to the choice of Itô, Stratonovich, or any other choice of integral.

Note that, given knowledge of just the path $t \mapsto (X_{0,t}, \mathbb{X}_{0,t})$, we can reconstruct the entire enhancement $\mathbb{X}$ via Chen's relation: $\mathbb{X}_{s,t} = \mathbb{X}_{0,t} - \mathbb{X}_{0,s} - X_{0,s} \otimes X_{s,t}$. In this sense, the "rough path" $(X, \mathbb{X})$ is indeed a genuine *path*, rather than just some two-parameter function.

3.2 Geometric rough paths

Chen's relation (3.1) captures the basic additive structure that one would expect any reasonable notion of integral to respect, but it doesn't encode any form of integration by parts or chain rule. We now seek an additional condition which will allow us to recover such classical rules of calculus.

Let $X = (X^1, \dots, X^d) \colon [0,T] \to \mathbb{R}^d$ be a smooth path, and let $\mathbb{X}_{s,t} := \int_s^t X_{s,r} \otimes \mathrm{d}X_r$ be its canonical lift (with the integral being defined in the Riemann–Stieltjes sense), so that $\mathbf{X} = (X, \mathbb{X})$ is a rough path. Applying integration by parts, we have

$$\mathbb{X}^{ij}_{s,t} + \mathbb{X}^{ji}_{s,t} = \int_s^t X^i_{s,r}\,\mathrm{d}X^j_r + \int_s^t X^j_{s,r}\,\mathrm{d}X^i_r = X^i_{s,t}X^j_{s,t}.$$

That is,

$$\mathrm{Sym}(\mathbb{X}_{s,t}) = \frac{1}{2} X_{s,t} \otimes X_{s,t}. \tag{3.3}$$

The condition (3.3) is thus a consequence of classical "first order calculus". This motivates the following definition.

Definition 3.3. We define the space of *weakly geometric α-Hölder rough paths* $\mathscr{C}^\alpha_g$ as the set of elements $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ such that (3.3) holds for all $(s,t) \in \Delta_{[0,T]}$.

Note that $\mathscr{C}^\alpha_g$ is a closed subset of $\mathscr{C}^\alpha$, and hence is itself a complete metric space. We also make the following alternative definition.

Definition 3.4. We define the space of *geometric α-Hölder rough paths* $\mathscr{C}^{0,\alpha}_g$ as the closure of canonical lifts of smooth paths with respect to the α-Hölder rough path distance.

To spell out this definition, $\mathbf{X} = (X, \mathbb{X})$ is a geometric rough path if and only if there exists a sequence of smooth paths $(X^n)_{n \ge 1}$ such that $\|\mathbf{X}^n; \mathbf{X}\|_\alpha \to 0$ as $n \to \infty$, where $\mathbf{X}^n = (X^n, \mathbb{X}^n)$, and $\mathbb{X}^n_{s,t} = \int_s^t X^n_{s,r} \otimes \mathrm{d}X^n_r$ for all $(s,t) \in \Delta_{[0,T]}$.

It is clear that $\mathscr{C}^{0,\alpha}_g \subset \mathscr{C}^\alpha_g$, and it turns out that this inclusion is strict. It can also be shown that $\mathscr{C}^\beta_g \subset \mathscr{C}^{0,\alpha}_g$ whenever $\frac{1}{3} < \alpha < \beta \le \frac{1}{2}$. Recall the embeddings of Hölder spaces in (2.9). In the rough path setting we have the analogous inclusions

$$\mathscr{C}^\beta_g \subset \mathscr{C}^{0,\alpha}_g \subset \mathscr{C}^\alpha_g \subset \mathscr{C}^\alpha, \tag{3.4}$$

each of which is strict.

4 Brownian motion as a rough path

In this section will we exhibit an important example of a (random) rough path, and see in particular how stochastic processes can be lifted to rough paths.

4.1 Kolmogorov criterion for rough paths

In the following we will write L^q for the standard Lebesgue space on the underlying probability space, so that $\|X_{s,t}\|_{L^q} = \mathbb{E}[|X_{s,t}|^q]^{1/q}$. Recall that we say $\tilde{X}$ is a *modification* of X if, for every $t \in [0,T]$, we have that $\tilde{X}_t = X_t$ almost surely.

Theorem 4.1. *Let $(X, \mathbb{X}) \colon \Omega \times [0,T] \to \mathbb{R}^d \times \mathbb{R}^{d \times d}$ be a measurable stochastic process which almost surely satisfies Chen's relation. Let $q \ge 2$ and $\beta > \frac{1}{q}$. Suppose that there exists a constant $C > 0$ such that, for all $(s,t) \in \Delta_{[0,T]}$,*

$$\|X_{s,t}\|_{L^q} \le C|t-s|^\beta, \qquad \|\mathbb{X}_{s,t}\|_{L^{q/2}} \le C|t-s|^{2\beta}. \tag{4.1}$$

Then, for all $\alpha \in [0, \beta - \frac{1}{q})$, there exists a modification $(\tilde{X}, \tilde{\mathbb{X}})$ of $(X, \mathbb{X})$ and random variables $K_\alpha \in L^q$, $\mathbb{K}_\alpha \in L^{q/2}$ such that, for all $(s,t) \in \Delta_{[0,T]}$,

$$|\tilde{X}_{s,t}| \le K_\alpha |t-s|^\alpha, \qquad |\tilde{\mathbb{X}}_{s,t}| \le \mathbb{K}_\alpha |t-s|^{2\alpha}. \tag{4.2}$$

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

Proof. Without loss of generality we may take $T = 1$. For each $n \ge 0$, let $D_n = \{\frac{k}{2^n} : k = 0, 1, \dots, 2^n - 1\}$ denote the dyadic partition of the interval $[0,1)$ with mesh size 2^{-n}. Let

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

It follows from (4.1) that

$$\mathbb{E}[K_n^q] \le \mathbb{E}\bigg[\sum_{t \in D_n} |X_{t,t+2^{-n}}|^q\bigg] \le \sum_{t \in D_n} C^q 2^{-n\beta q} = C^q 2^{-n(\beta q - 1)},$$
$$\mathbb{E}[\mathbb{K}_n^{q/2}] \le \mathbb{E}\bigg[\sum_{t \in D_n} |\mathbb{X}_{t,t+2^{-n}}|^{q/2}\bigg] \le \sum_{t \in D_n} C^{q/2} 2^{-n\beta q} = C^{q/2} 2^{-n(\beta q - 1)}.$$

Fix $s < t$ in $\cup_{n \ge 0} D_n$. Choose $m \ge 0$ such that $2^{-(m+1)} < t - s \le 2^{-m}$. The interval $[s,t]$ can be expressed as the finite union of intervals of the form $[u,v] \in D_n$ with $n \ge m+1$ and where no three intervals have the same length. In other words, we have a partition of $[s,t]$ of the form

$$s = u_0 < u_1 < \dots < u_N = t,$$

where $[u_i, u_{i+1}] \in D_n$ for some $n \ge m+1$, and for each fixed $n \ge m+1$ there are at most two such intervals taken from D_n. It follows that

$$|X_{s,t}| \le \max_{0 \le i < N} |X_{s,u_{i+1}}| \le \sum_{i=0}^{N-1} |X_{u_i,u_{i+1}}| \le 2 \sum_{n=m+1}^{\infty} K_n,$$

and similarly,

$$\begin{aligned}
|\mathbb{X}_{s,t}| &= \bigg|\sum_{i=0}^{N-1} \big(\mathbb{X}_{u_i,u_{i+1}} + X_{s,u_i} \otimes X_{u_i,u_{i+1}}\big)\bigg| \le \sum_{i=0}^{N-1} \big(|\mathbb{X}_{u_i,u_{i+1}}| + |X_{s,u_i}||X_{u_i,u_{i+1}}|\big) \\
&\le \sum_{i=0}^{N-1} |\mathbb{X}_{u_i,u_{i+1}}| + \Big(\max_{0 \le i < N} |X_{s,u_{i+1}}|\Big)\bigg(\sum_{i=0}^{N-1} |X_{u_i,u_{i+1}}|\bigg) \\
&\le 2 \sum_{n=m+1}^{\infty} \mathbb{K}_n + \bigg(2 \sum_{n=m+1}^{\infty} K_n\bigg)^2.
\end{aligned}$$

We thus obtain

$$\frac{|X_{s,t}|}{|t-s|^\alpha} \le 2 \sum_{n=m+1}^{\infty} \frac{K_n}{2^{-(m+1)\alpha}} \le 2 \sum_{n=m+1}^{\infty} \frac{K_n}{2^{-n\alpha}} \le 2 \sum_{n=0}^{\infty} \frac{K_n}{2^{-n\alpha}} =: K_\alpha.$$

Since

$$\|K_\alpha\|_{L^q} \le 2 \sum_{n=0}^{\infty} \frac{\mathbb{E}[K_n^q]^{1/q}}{2^{-n\alpha}} \le 2 \sum_{n=0}^{\infty} \frac{C 2^{-n(\beta - \frac{1}{q})}}{2^{-n\alpha}} = 2C \sum_{n=0}^{\infty} 2^{-n(\beta - \frac{1}{q} - \alpha)} < \infty,$$

we have that $K_\alpha \in L^q$. Similarly,

$$\frac{|\mathbb{X}_{s,t}|}{|t-s|^{2\alpha}} \leq 2\sum_{n=m+1}^{\infty} \frac{\mathbb{K}_n}{2^{-2(m+1)\alpha}} + \left(2\sum_{n=m+1}^{\infty} \frac{K_n}{2^{-(m+1)\alpha}}\right)^2 \leq 2\sum_{n=0}^{\infty} \frac{\mathbb{K}_n}{2^{-2n\alpha}} + K_\alpha^2 =: \mathbb{K}_\alpha.$$

Since

$$\begin{aligned}
\|\mathbb{K}_\alpha\|_{L^{q/2}} &\leq 2\sum_{n=0}^{\infty} \frac{\mathbb{E}[\mathbb{K}_n^{q/2}]^{2/q}}{2^{-2n\alpha}} + \mathbb{E}[K_\alpha^q]^{2/q} \leq 2\sum_{n=0}^{\infty} \frac{C2^{-2n(\beta-\frac{1}{q})}}{2^{-2n\alpha}} + \|K_\alpha\|_{L^q}^2 \\
&= 2C\sum_{n=0}^{\infty} 2^{-2n(\beta-\frac{1}{q}-\alpha)} + \|K_\alpha\|_{L^q}^2 < \infty,
\end{aligned}$$

we have that $\mathbb{K}_\alpha \in L^{q/2}$.

So far we have shown that (4.2) holds for $(X, \mathbb{X})$ at every pair of times $s < t$ in $\cup_{n\geq 0} D_n$. We now need to extend this to all times in between, which is where a modification is required.

For each $t \in [0,1]$, let $(t_k)_{k\geq 1} \subset \cup_{n\geq 0} D_n$ be a sequence of times with $t_k \to t$ as $k \to \infty$. It follows from the above that X is Hölder continuous on $\cup_{n\geq 0} D_n$, and hence that the limit $\tilde{X}_t := \lim_{k\to\infty} X_{t_k}$ exists. By Fatou's lemma, we have

$$\|\tilde{X}_t - X_t\|_{L^q} \leq \liminf_{k\to\infty} \|X_{t_k} - X_t\|_{L^q} \leq \liminf_{k\to\infty} C|t-t_k|^\beta = 0,$$

so that $\tilde{X}_t = X_t$ almost surely. Thus, $\tilde{X}$ is indeed a modification of X. Moreover, assuming $s_k \to s$ and $t_k \to t$ with $s_k, t_k \in \cup_{n\geq 0} D_n$, we have

$$|\tilde{X}_{s,t}| = \lim_{k\to\infty} |X_{s_k,t_k}| \leq \lim_{k\to\infty} K_\alpha |t_k - s_k|^\alpha = K_\alpha |t-s|^\alpha. \tag{4.3}$$

(This argument also implies that $\tilde{X}_t$ almost surely does not depend on the choice of the sequence of times $(t_k)_{k\geq 1}$.)

For each pair $0 < s \leq t < 1$ (the cases with $s = 0$ or $t = 1$ may be dealt with similarly), let $(s_k)_{k\geq 1} \subset \cup_{n\geq 0} D_n$ and $(t_k)_{k\geq 1} \subset \cup_{n\geq 0} D_n$ be sequences such that $s_k \nearrow s$ and $t_k \searrow t$ as $k \to \infty$, and define $\tilde{\mathbb{X}}_{s,t} = \lim_{k\to\infty} \mathbb{X}_{s_k,t_k}$. Since $s_k \leq s \leq t \leq t_k$, Chen's relation implies that

$$\mathbb{X}_{s_k,t_k} - \mathbb{X}_{s,t} = \mathbb{X}_{s_k,s} + \mathbb{X}_{t,t_k} + X_{s_k,s} \otimes X_{s,t_k} + X_{s,t} \otimes X_{t,t_k}.$$

We then have

$$\begin{aligned}
\|\mathbb{X}_{s_k,t_k} - \mathbb{X}_{s,t}\|_{L^{q/2}} &= \|\mathbb{X}_{s_k,s} + \mathbb{X}_{t,t_k} + X_{s_k,s} \otimes X_{s,t_k} + X_{s,t} \otimes X_{t,t_k}\|_{L^{q/2}} \\
&\leq \|\mathbb{X}_{s_k,s}\|_{L^{q/2}} + \|\mathbb{X}_{t,t_k}\|_{L^{q/2}} + \|X_{s_k,s}\|_{L^q}\|X_{s,t_k}\|_{L^q} + \|X_{s,t}\|_{L^q}\|X_{t,t_k}\|_{L^q} \\
&\leq C|s-s_k|^{2\beta} + C|t_k - t|^{2\beta} + C^2|s-s_k|^\beta |t_k - s|^\beta + C^2|t-s|^\beta|t_k-t|^\beta,
\end{aligned}$$

and, similarly to above, it follows from Fatou's lemma that $\tilde{\mathbb{X}}_{s,t} = \mathbb{X}_{s,t}$ almost surely for every (s,t), so that $\tilde{\mathbb{X}}$ is a modification of $\mathbb{X}$. Finally, we have that

$$|\tilde{\mathbb{X}}_{s,t}| = \lim_{k\to\infty} |\mathbb{X}_{s_k,t_k}| \leq \lim_{k\to\infty} \mathbb{K}_\alpha |t_k - s_k|^{2\alpha} = \mathbb{K}_\alpha |t-s|^{2\alpha},$$

which, combined with (4.3), implies that (4.2) holds for $(\tilde{X}, \tilde{\mathbb{X}})$ for all $s < t$. $\square$

Although in general $(\tilde{X}, \tilde{\mathbb{X}})$ is a modification of $(X, \mathbb{X})$, in practice one usually assumes that, whenever it exists, such a modification has always been adopted. It is therefore usual to keep the same notation $(X, \mathbb{X})$, rather than introducing $(\tilde{X}, \tilde{\mathbb{X}})$.

4.2 Itô Brownian motion

Consider a d-dimensional standard Brownian motion B. We can enhance B by defining

$$\mathbb{B}^{\text{Itô}}_{s,t} := \int_s^t B_{s,r} \otimes \mathrm{d}B_r, \qquad (s,t) \in \Delta_{[0,T]}, \tag{4.4}$$

where the stochastic integral is understood in the sense of Itô. By the additivity of the integral, it is easy to check that the pair $(B, \mathbb{B}^{\text{Itô}})$ satisfies Chen's relation. It remains to use the Kolmogorov criterion for rough paths to check that $(B, \mathbb{B}^{\text{Itô}})$ has the required Hölder regularity.

Proposition 4.2. *For any $\alpha \in (\frac{1}{3}, \frac{1}{2})$, we have that, almost surely,*

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

Proof. Let $q \geq 2$ and $(s,t) \in \Delta_{[0,T]}$. We have

$$\|B_{s,t}\|_{L^q} = \|(t-s)^{\frac{1}{2}} B_1\|_{L^q} = \|B_1\|_{L^q} |t-s|^{\frac{1}{2}}.$$

Applying the Burkholder–Davis–Gundy inequality twice, we also have

$$\begin{aligned}\mathbb{E}\big[|\mathbb{B}^{\text{Itô}}_{s,t}|^{\frac{q}{2}}\big] = \mathbb{E}\bigg[\bigg|\int_s^t B_{s,r} \otimes \mathrm{d}B_r\bigg|^{\frac{q}{2}}\bigg] &\leq C_q \mathbb{E}\bigg[\bigg|\int_s^t |B_{s,r}|^2 \,\mathrm{d}r\bigg|^{\frac{q}{4}}\bigg] \\ &\leq C_q \mathbb{E}\Big[\sup_{r \in [s,t]} |B_{s,r}|^{\frac{q}{2}}\Big] |t-s|^{\frac{q}{4}} \leq C_q^2 |t-s|^{\frac{q}{2}},\end{aligned}$$

so that

$$\|\mathbb{B}^{\text{Itô}}_{s,t}\|_{L^{q/2}} \leq C_q^{\frac{4}{q}} |t-s|.$$

We can therefore apply Theorem 4.1 with $\beta = \frac{1}{2}$, to deduce (possibly after taking a suitable modification) that $(B, \mathbb{B}^{\text{Itô}}) \in \mathcal{C}^\alpha \times \mathcal{C}_2^{2\alpha}$ for any $\alpha \in (0, \frac{1}{2} - \frac{1}{q})$. By taking $q \to \infty$, it follows that $(B, \mathbb{B}^{\text{Itô}}) \in \mathscr{C}^\alpha$ for all $\alpha \in (\frac{1}{3}, \frac{1}{2})$. $\square$

To be explicit, this means that for almost every $\omega \in \Omega$, we have that

$$\mathbf{B}(\omega) = (B(\omega), \mathbb{B}^{\text{Itô}}(\omega)) \in \mathscr{C}^\alpha.$$

In other words, $\mathbf{B}$ is a random rough path, or a rough path-valued random variable. We refer to $\mathbf{B} = (B, \mathbb{B}^{\text{Itô}})$ as (Itô enhanced) Brownian rough path.

An obvious question is whether $\mathbf{B}$ so defined is geometric. The answer, sadly, is no. Indeed, Itô's formula tells us that, for each $1 \leq i, j \leq d$,

$$B^i_{s,t} B^j_{s,t} = \int_s^t B^i_{s,r} \mathrm{d}B^j_r + \int_s^t B^j_{s,r} \mathrm{d}B^i_r + \langle B^i, B^j \rangle_{s,t},$$

where, for a Brownian motion B, we have $\langle B^i, B^j \rangle_{s,t} = \delta_{ij}(t-s)$ (where δ_{ij} is the Kronecker delta). Thus,

$$\operatorname{Sym}(\mathbb{B}^{\text{Itô}}_{s,t}) = \frac{1}{2}\big(B_{s,t} \otimes B_{s,t} - (t-s)I\big), \tag{4.5}$$

where I denotes the $d \times d$-identity matrix, and we see that (3.3) does *not* hold.

4.3 Stratonovich Brownian motion

For one-dimensional continuous semimartingales, X, Y, the *Stratonovich integral* of Y against X is defined by

$$\int_0^t Y_s \circ \mathrm{d}X_s = \int_0^t Y_s \,\mathrm{d}X_s + \frac{1}{2}\langle Y, X\rangle_t. \tag{4.6}$$

Recall also the limit in (1.8). An advantage of Stratonovich integration is that it obeys "first order calculus". By this, we mean that it satisfies the classical integration by parts formula

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s \circ \mathrm{d}Y_s + \int_0^t Y_s \circ \mathrm{d}X_s,$$

and chain rule/fundamental theorem of calculus

$$f(X_t) = f(X_0) + \int_0^t Df(X_s) \circ \mathrm{d}X_s.$$

As we will see in the next proposition, as a consequence it turns out that Stratonovich-enhanced Brownian motion gives a rough path which is also geometric.

As above, let B be a d-dimensional standard Brownian motion. Instead of using Itô integration, we can alternatively enhance B via

$$\mathbb{B}^{\mathrm{Strat}}_{s,t} := \int_s^t B_{s,r} \otimes \circ \,\mathrm{d}B_r, \qquad (s,t) \in \Delta_{[0,T]}.$$

It is again easy to see that the pair $(B, \mathbb{B}^{\mathrm{Strat}})$ satisfies Chen's relation. We also see from (4.6) that

$$\mathbb{B}^{\mathrm{Strat}}_{s,t} = \mathbb{B}^{\mathrm{Itô}}_{s,t} + \frac{1}{2}(t-s)I, \tag{4.7}$$

which means in particular that

$$\mathrm{Sym}(\mathbb{B}^{\mathrm{Strat}}_{s,t}) = \mathrm{Sym}(\mathbb{B}^{\mathrm{Itô}}_{s,t}) + \frac{1}{2}(t-s)I = \frac{1}{2} B_{s,t} \otimes B_{s,t}. \tag{4.8}$$

Proposition 4.3. *For any $\alpha \in (\frac{1}{3}, \frac{1}{2})$, we have that, almost surely,*

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

Proof. Since the function $(s,t) \mapsto \frac{1}{2}(t-s)I$ is 1-Hölder continuous, it follows immediately from (4.7) that the Hölder regularity of $\mathbb{B}^{\mathrm{Itô}}$ is inherited by $\mathbb{B}^{\mathrm{Strat}}$. We therefore have that $\mathbf{B} = (B, \mathbb{B}^{\mathrm{Strat}}) \in \mathscr{C}^\alpha$.

Let $\beta \in (\alpha, \frac{1}{2})$. It follows from (4.8) that $\mathbf{B} \in \mathscr{C}^\beta_g$. Recalling the inclusions in (3.4), we conclude that $\mathbf{B} \in \mathscr{C}^{0,\alpha}_g$. □

We refer to $\mathbf{B} = (B, \mathbb{B}^{\mathrm{Strat}})$ as (Stratonovich enhanced) Brownian rough path.

Recalling (4.5) and (4.8), it is worth noting that for both Itô and Stratonovich enhanced Brownian rough paths, given the path $t \mapsto B_t$, the symmetric part of the enhancement is immediately known. This is a general feature, whereby the "new information" encoded by the rough path lift $\mathbb{X}$ is actually in its *antisymmetric part*, given, for $1 \leq i, j \leq d$, by

$$\mathrm{Anti}(\mathbb{X}_{s,t})^{ij} = \frac{1}{2}\bigg(\int_s^t X^i_{s,r}\,\mathrm{d}X^j_r - \int_s^t X^j_{s,r}\,\mathrm{d}X^i_r\bigg),$$

which we recognise as the *Lévy area* of the two-dimensional path $t \mapsto (X^i_t, X^j_t)$. In this sense, a geometric rough path may be equivalently defined as a path X along with its Lévy area $\mathrm{Anti}(\mathbb{X})$.

Since a 1×1-matrix is already symmetric, any one-dimensional path $X \in \mathcal{C}^\alpha$ may be readily lifted to a weakly geometric rough path by simply setting

$$\mathbb{X}_{s,t} := \frac{1}{2}(X_{s,t})^2.$$

5 Integration

5.1 The sewing lemma

The following result may look at first glance like abstract nonsense, but it will actually turn out to be a very useful tool for constructing integrals. Although this result is commonly known as the sewing *lemma*, we will give it the recognition it deserves by calling it a theorem.

Theorem 5.1 (Sewing lemma)**.** *Let $(E, \|\cdot\|)$ be a Banach space, and let $A\colon \Delta_{[0,T]} \to E$ be a continuous function. For each triplet $0 \leq s \leq u \leq t \leq T$, write $\delta A_{s,u,t} := A_{s,t} - A_{s,u} - A_{u,t}$. Suppose that there exist constants $\lambda \geq 0$ and $\varepsilon > 0$ such that*

$$\|\delta A_{s,u,t}\| \leq \lambda |t-s|^{1+\varepsilon}$$

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

Then there exists a continuous path $\gamma\colon [0,T] \to E$, with $\gamma_0 = 0$, such that

$$\|\gamma_t - \gamma_s - A_{s,t}\| \leq C\lambda |t-s|^{1+\varepsilon} \tag{5.1}$$

for all $(s,t) \in \Delta_{[0,T]}$, where the constant C depends only on ε. Moreover, for all $(s,t) \in \Delta_{[0,T]}$, we have that

$$\lim_{|\pi|\to 0} \sum_{[u,v]\in\pi} A_{u,v} = \gamma_t - \gamma_s,$$

where the limit is taken over any sequence of partitions π of the interval $[s,t]$ with mesh size $|\pi| \to 0$.

Proof. Let $(s,t) \in \Delta_{[0,T]}$. For each integer $n \geq 0$, let $\{s = t^n_0 < t^n_1 < \cdots < t^n_{2^n} = t\}$ be the dyadic partition of $[s,t]$, so that $t^n_i = s + \frac{i}{2^n}(t-s)$, which has mesh size $|\pi^n| = |t^n_{i+1} - t^n_i| = 2^{-n}|t-s|$. Let

$$A^n_{s,t} = \sum_{i=0}^{2^n-1} A_{t^n_i, t^n_{i+1}}.$$

For each n and each i, let u_i^n be the midpoint of the interval $[t_i^n, t_{i+1}^n]$. We have that

$$A_{s,t}^n - A_{s,t}^{n+1} = \sum_{i=0}^{2^n-1} \delta A_{t_i^n, u_i^n, t_{i+1}^n},$$

and hence that

$$\begin{aligned}
\|A_{s,t}^n - A_{s,t}^{n+1}\| &\leq \sum_{i=0}^{2^n-1} \|\delta A_{t_i^n, u_i^n, t_{i+1}^n}\| \leq \sum_{i=0}^{2^n-1} \lambda |t_{i+1}^n - t_i^n|^{1+\varepsilon} \\
&= \lambda |t-s|^{1+\varepsilon} \sum_{i=0}^{2^n-1} 2^{-n(1+\varepsilon)} = \lambda |t-s|^{1+\varepsilon} 2^{-n\varepsilon}.
\end{aligned}$$

We then have that

$$\sum_{n=k}^{\infty} \|A_{s,t}^n - A_{s,t}^{n+1}\| \leq \lambda |t-s|^{1+\varepsilon} \sum_{n=k}^{\infty} 2^{-n\varepsilon} = \lambda |t-s|^{1+\varepsilon} \frac{2^{-k\varepsilon}}{1-2^{-\varepsilon}} \longrightarrow 0 \qquad \text{as} \quad k \to \infty,$$

from which it follows that $(A_{s,t}^n)_{n \geq 0}$ is a Cauchy sequence. Since A^n takes values in a Banach space, we have that the limit

$$\Gamma_{s,t} := \lim_{n \to \infty} A_{s,t}^n$$

exists. Since

$$\|\Gamma_{s,t} - A_{s,t}^k\| = \left\| \sum_{n=k}^{\infty} (A_{s,t}^n - A_{s,t}^{n+1}) \right\| \leq \lambda |t-s|^{1+\varepsilon} \frac{2^{-k\varepsilon}}{1-2^{-\varepsilon}},$$

we see that the convergence $A_{s,t}^n \to \Gamma_{s,t}$ holds uniformly in $(s,t) \in \Delta_{[0,T]}$, and moreover that

$$\|\Gamma_{s,t} - A_{s,t}\| \leq \frac{\lambda |t-s|^{1+\varepsilon}}{1-2^{-\varepsilon}}. \tag{5.2}$$

Since A^n is continuous, it follows from the uniform convergence that Γ is also continuous.

It follows from the above construction that

$$\Gamma_{s,u} + \Gamma_{u,t} = \Gamma_{s,t} \tag{5.3}$$

for all dyadic times $s \leq u \leq t$, and it then follows by continuity that (5.3) holds for *all* times $s \leq u \leq t$. We thus infer that Γ is really just the increments of a continuous path. That is, if we define $\gamma_t = \Gamma_{0,t}$ for $t \in [0,T]$, then we have that

$$\gamma_t - \gamma_s = \Gamma_{s,t} \qquad \text{for all} \quad (s,t) \in \Delta_{[0,T]},$$

and in particular that $\gamma_0 = 0$. The inequality (5.2) then reads

$$\|\gamma_t - \gamma_s - A_{s,t}\| \leq \frac{\lambda |t-s|^{1+\varepsilon}}{1-2^{-\varepsilon}},$$

which implies (5.1). For any $(s,t) \in \Delta_{[0,T]}$ and any (not necessarily dyadic) partition $\pi = \{s = t_0 < t_1 < \cdots < t_N = t\}$ of $[s,t]$, we then have

$$\begin{aligned}
\left\| \gamma_t - \gamma_s - \sum_{i=0}^{N-1} A_{t_i, t_{i+1}} \right\| &= \left\| \sum_{i=0}^{N-1} (\gamma_{t_{i+1}} - \gamma_{t_i} - A_{t_i, t_{i+1}}) \right\| \leq \frac{\lambda}{1-2^{-\varepsilon}} \sum_{i=0}^{N-1} |t_{i+1} - t_i|^{1+\varepsilon} \\
&\leq \frac{\lambda}{1-2^{-\varepsilon}} |t-s| |\pi|^{\varepsilon},
\end{aligned}$$

and we deduce that $\sum_{i=0}^{N-1} A_{t_i, t_{i+1}} \to \gamma_t - \gamma_s$ as $|\pi| \to 0$. ∎

5.2 Young integration

Proposition 5.2. *Let $X \in \mathcal{C}^\alpha$ and $Y \in \mathcal{C}^\beta$ for some $\alpha, \beta \in (0,1]$ with $\alpha + \beta > 1$. Then the limit*

$$\int_0^t Y_u \,\mathrm{d}X_u := \lim_{|\pi| \to 0} \sum_{[u,v] \in \pi} Y_u X_{u,v}$$

exists for every $t \in [0,T]$, where the limit is taken over any sequence of partitions π of the interval $[0,t]$ with mesh size $|\pi| \to 0$. This limit is called the Young integral *of Y against X, which moreover comes with the estimate*

$$\left| \int_s^t Y_u \,\mathrm{d}X_u - Y_s X_{s,t} \right| \leq C \|Y\|_\beta \|X\|_\alpha |t-s|^{\alpha+\beta} \tag{5.4}$$

for all $(s,t) \in \Delta_{[0,T]}$, where the constant C depends only on $\alpha + \beta$.

Proof. Let $A_{s,t} = Y_s X_{s,t}$, and let $\delta A_{s,u,t} = A_{s,t} - A_{s,u} - A_{u,t}$ for $s \leq u \leq t$. We have

$$\begin{aligned} \delta A_{s,u,t} &= Y_s X_{s,t} - Y_s X_{s,u} - Y_u X_{u,t} \\ &= Y_s X_{u,t} - Y_u X_{u,t} \\ &= -Y_{s,u} X_{u,t}, \end{aligned}$$

and hence

$$|\delta A_{s,u,t}| = |Y_{s,u} X_{u,t}| \leq \|Y\|_\beta \|X\|_\alpha |t-s|^{\alpha+\beta}.$$

By the sewing lemma (Theorem 5.1), there exists a continuous path $\gamma =: \int_0^\cdot Y_u \,\mathrm{d}X_u$ with the desired properties. □

The Young integral as defined in the previous proposition is given as a limit of left endpoint Riemann sums. In this setting (in particular with continuous paths), the left endpoint may be replaced by any other intermediate point without changing the value of the integral, as shown in the next lemma.

Lemma 5.3. *Let $X \in \mathcal{C}^\alpha$ and $Y \in \mathcal{C}^\beta$ for some $\alpha, \beta \in (0,1]$ with $\alpha + \beta > 1$. For a partition π and interval $[u,v] \in \pi$, let $r \in [u,v]$ denote an arbitrary point in the interval $[u,v]$. The Young integral of Y against X is equal to the limit*

$$\int_0^t Y_u \,\mathrm{d}X_u = \lim_{|\pi| \to 0} \sum_{[u,v] \in \pi} Y_r X_{u,v}$$

for any $t \in [0,T]$.

Proof. We have

$$\begin{aligned} \left| \sum_{[u,v] \in \pi} Y_{u,r} X_{u,v} \right| &\leq \|Y\|_\beta \|X\|_\alpha \sum_{[u,v] \in \pi} |v-u|^{\alpha+\beta} \\ &\leq \|Y\|_\beta \|X\|_\alpha T |\pi|^{\alpha+\beta-1} \longrightarrow 0 \end{aligned}$$

as $|\pi| \to 0$. Thus

$$\lim_{|\pi| \to 0} \sum_{[u,v] \in \pi} Y_r X_{u,v} = \lim_{|\pi| \to 0} \left(\sum_{[u,v] \in \pi} Y_u X_{u,v} + \sum_{[u,v] \in \pi} Y_{u,r} X_{u,v} \right) = \int_0^t Y_u \,\mathrm{d}X_u.$$

□

Combining the results of Proposition 5.2 and Lemma 5.3, we have proven Theorem 1.2.

Proposition 5.4. *Let $X, \tilde{X} \in \mathcal{C}^\alpha$ and $Y, \tilde{Y} \in \mathcal{C}^\beta$ for some $\alpha, \beta \in (0,1]$ with $\alpha+\beta > 1$. Then there exists a constant C, depending only on α, β and T, such that*

$$\begin{aligned}&\left\| \int_0^{\cdot} Y_u \,\mathrm{d}X_u - \int_0^{\cdot} \tilde{Y}_u \,\mathrm{d}\tilde{X}_u \right\|_\alpha \\ &\quad\leq C\Big(\big(|Y_0 - \tilde{Y}_0| + \|Y - \tilde{Y}\|_\beta\big)\|X\|_\alpha + \big(|\tilde{Y}_0| + \|\tilde{Y}\|_\beta\big)\|X - \tilde{X}\|_\alpha \Big).\end{aligned}$$

Proof. Let $A_{s,t} = Y_s X_{s,t}$, $\tilde{A}_{s,t} = \tilde{Y}_s \tilde{X}_{s,t}$, $\Delta = A - \tilde{A}$, and $\delta\Delta_{s,u,t} = \Delta_{s,t} - \Delta_{s,u} - \Delta_{u,t}$. Then

$$\delta\Delta_{s,u,t} = \delta A_{s,u,t} - \delta\tilde{A}_{s,u,t} = -\big(Y_{s,u}X_{u,t} - \tilde{Y}_{s,u}\tilde{X}_{u,t}\big),$$

so that

$$\begin{aligned}|\delta\Delta_{s,u,t}| &\leq \big|Y_{s,u}X_{u,t} - \tilde{Y}_{s,u}\tilde{X}_{u,t}\big| \leq |Y_{s,u} - \tilde{Y}_{s,u}||X_{u,t}| + |\tilde{Y}_{s,u}||X_{u,t} - \tilde{X}_{u,t}| \\ &\leq \big(\|Y - \tilde{Y}\|_\beta \|X\|_\alpha + \|\tilde{Y}\|_\beta \|X - \tilde{X}\|_\alpha\big)|t-s|^{\alpha+\beta}.\end{aligned}$$

By the sewing lemma (Theorem 5.1), there exists a path γ and a constant C such that

$$|\gamma_t - \gamma_s - \Delta_{s,t}| \leq C\big(\|Y - \tilde{Y}\|_\beta \|X\|_\alpha + \|\tilde{Y}\|_\beta \|X - \tilde{X}\|_\alpha\big)|t-s|^{\alpha+\beta},$$

and, letting π denote a partition of the interval $[s,t]$,

$$\gamma_t - \gamma_s = \lim_{|\pi|\to 0} \sum_{[u,v]\in\pi} \Delta_{u,v} = \lim_{|\pi|\to 0} \left(\sum_{[u,v]\in\pi} A_{u,v} - \sum_{[u,v]\in\pi} \tilde{A}_{u,v} \right) = \int_s^t Y_u \,\mathrm{d}X_u - \int_s^t \tilde{Y}_u \,\mathrm{d}\tilde{X}_u.$$

Combining the above, we have that

$$\begin{aligned}&\left| \int_s^t Y_u \,\mathrm{d}X_u - \int_s^t \tilde{Y}_u \,\mathrm{d}\tilde{X}_u - \big(Y_s X_{s,t} - \tilde{Y}_s \tilde{X}_{s,t}\big) \right| \\ &\qquad\leq C\big(\|Y - \tilde{Y}\|_\beta \|X\|_\alpha + \|\tilde{Y}\|_\beta \|X - \tilde{X}\|_\alpha\big)|t-s|^{\alpha+\beta}.\end{aligned} \tag{5.5}$$

We also have

$$\begin{aligned}\big|Y_s X_{s,t} - \tilde{Y}_s \tilde{X}_{s,t}\big| &\leq |Y_s - \tilde{Y}_s||X_{s,t}| + |\tilde{Y}_s||X_{s,t} - \tilde{X}_{s,t}| \\ &\leq \big(\|Y - \tilde{Y}\|_\infty \|X\|_\alpha + \|\tilde{Y}\|_\infty \|X - \tilde{X}\|_\alpha\big)|t-s|^{\alpha}.\end{aligned} \tag{5.6}$$

Combining (5.5) and (5.6) and noting the simple bounds $\|\tilde{Y}\|_\infty \leq |\tilde{Y}_0| + T^\beta \|\tilde{Y}\|_\beta$ and $\|Y - \tilde{Y}\|_\infty \leq |Y_0 - \tilde{Y}_0| + T^\beta \|Y - \tilde{Y}\|_\beta$, it follows that

$$\begin{aligned}&\left| \int_s^t Y_u \,\mathrm{d}X_u - \int_s^t \tilde{Y}_u \,\mathrm{d}\tilde{X}_u \right| \\ &\leq (1+C)(1+T^\beta)\Big(\big(|Y_0 - \tilde{Y}_0| + \|Y - \tilde{Y}\|_\beta\big)\|X\|_\alpha + \big(|\tilde{Y}_0| + \|\tilde{Y}\|_\beta\big)\|X - \tilde{X}\|_\alpha \Big)|t-s|^\alpha,\end{aligned}$$

and thus

$$\begin{aligned}&\left\| \int_0^{\cdot} Y_u \,\mathrm{d}X_u - \int_0^{\cdot} \tilde{Y}_u \,\mathrm{d}\tilde{X}_u \right\|_\alpha \\ &\quad\leq (1+C)(1+T^\beta)\Big(\big(|Y_0 - \tilde{Y}_0| + \|Y - \tilde{Y}\|_\beta\big)\|X\|_\alpha + \big(|\tilde{Y}_0| + \|\tilde{Y}\|_\beta\big)\|X - \tilde{X}\|_\alpha \Big).\end{aligned}$$

□

As the next lemma shows, Young integration satisfies the classical integration by parts formula.

Lemma 5.5. *Let $X \in \mathcal{C}^\alpha$ and $Y \in \mathcal{C}^\beta$ for some $\alpha, \beta \in (0,1]$ with $\alpha + \beta > 1$. Then*

$$X_T Y_T = X_0 Y_0 + \int_0^T X_u \,\mathrm{d}Y_u + \int_0^T Y_u \,\mathrm{d}X_u.$$

Proof. Let π be a partition of the interval $[0,T]$. We have

$$\begin{aligned} \Big| \sum_{[s,t]\in\pi} X_{s,t} Y_{s,t} \Big| &\le \|X\|_\alpha \|Y\|_\beta \sum_{[s,t]\in\pi} |t-s|^{\alpha+\beta} \\ &\le \|X\|_\alpha \|Y\|_\beta T |\pi|^{\alpha+\beta-1} \longrightarrow 0 \end{aligned}$$

as $|\pi| \to 0$. Then

$$\begin{aligned} X_T Y_T - X_0 Y_0 &= \lim_{|\pi|\to 0} \sum_{[s,t]\in\pi} (X_t Y_t - X_s Y_s) \\ &= \lim_{|\pi|\to 0} \sum_{[s,t]\in\pi} \big(X_s Y_{s,t} + Y_s X_{s,t} + X_{s,t} Y_{s,t} \big) \\ &= \int_0^T X_u \,\mathrm{d}Y_u + \int_0^T Y_u \,\mathrm{d}X_u. \end{aligned}$$

□

We write $C(\mathbb{R}^d;\mathbb{R})$ for the space of continuous functions $f\colon \mathbb{R}^d \to \mathbb{R}$. Given a function $f\colon \mathbb{R}^d \to \mathbb{R}$ and $k \in \mathbb{N}$, we write $D^k f$ for the k^{th} order derivative of f. For $\gamma \in (0,1]$, we say that f is locally γ-Hölder continuous if, for every bounded subset $K \subset \mathbb{R}^d$, there exists a constant C such that $|f(x) - f(y)| \le C|x-y|^\gamma$ for all $x, y \in K$.

For $k \in \mathbb{N}$ and $\gamma \in (0,1]$, we write $f \in C^{k+\gamma} = C^{k+\gamma}(\mathbb{R}^d;\mathbb{R})$ whenever a function f is k times continuously differentiable, and the k^{th} order derivative $D^k f$ is locally γ-Hölder continuous.

Lemma 5.6. *Let $X \in \mathcal{C}^\alpha([0,T];\mathbb{R}^d)$ and $f \in C^{1+\gamma}(\mathbb{R}^d;\mathbb{R})$ for some $\alpha, \gamma \in (0,1]$, such that $\alpha(1+\gamma) > 1$. Then $\int_0^T Df(X_u)\,\mathrm{d}X_u$ is a well-defined Young integral, and*

$$f(X_T) = f(X_0) + \int_0^T Df(X_u)\,\mathrm{d}X_u.$$

The proof of Lemma 5.6 is left as an exercise. We conclude our discussion of Young integration with the following lemma.

Lemma 5.7. *Let $X \in \mathcal{C}^\alpha$ and $Y, K \in \mathcal{C}^\beta$ for some $\alpha, \beta \in (0,1]$ with $\alpha + \beta > 1$. Let $Z = \int_0^\cdot K_u \,\mathrm{d}X_u$. Then $Z \in \mathcal{C}^\alpha$, and*

$$\int_0^T Y_u \,\mathrm{d}Z_u = \int_0^T Y_u K_u \,\mathrm{d}X_u.$$

Proof. It follows from (5.4) that

$$|Z_{s,t}| = \left| \int_s^t K_u \, \mathrm{d}X_u - K_s X_{s,t} + K_s X_{s,t} \right| \leq C \|K\|_\beta \|X\|_\alpha |t-s|^{\alpha+\beta} + \|K\|_\infty \|X\|_\alpha |t-s|^\alpha,$$

and hence that $\|Z\|_\alpha \leq C\|K\|_\beta \|X\|_\alpha T^\beta + \|K\|_\infty \|X\|_\alpha < \infty$, so we indeed have that $Z \in \mathcal{C}^\alpha$. Since $Y \in \mathcal{C}^\beta$, the Young integral $\int_0^T Y_u \, \mathrm{d}Z_u$ is then well-defined.

By repeated use of the estimate in (5.4), we have

$$\begin{aligned} \int_s^t Y_u \, \mathrm{d}Z_u &= Y_s Z_{s,t} + O(|t-s|^{\alpha+\beta}) \\ &= Y_s K_s X_{s,t} + O(|t-s|^{\alpha+\beta}) \\ &= \int_s^t Y_u K_u \, \mathrm{d}X_u + O(|t-s|^{\alpha+\beta}). \end{aligned}$$

Taking $\lim_{|\pi| \to 0} \sum_{[s,t] \in \pi}$ on both sides, we deduce the result. $\square$

Using $\bullet$ as formal notation for integration, the previous lemma states that

$$Y \bullet (K \bullet X) = (YK) \bullet X.$$

This property is therefore known as the *associativity* of Young integration.

5.3 Controlled paths

Let us recall some of the motivation discussed in the introduction. Given a sufficiently smooth function f and a path X, a Taylor expansion tells us that

$$f(X_r) \simeq f(X_s) + Df(X_s) X_{s,r},$$

and integrating with respect to X then gives

$$\int_s^t f(X_r) \, \mathrm{d}X_r \simeq f(X_s) X_{s,t} + Df(X_s) \int_s^t X_{s,r} \otimes \mathrm{d}X_r.$$

This suggests that we might expect to be able to establish a limit of the form

$$\int_0^T f(X_r) \, \mathrm{d}X_r \overset{?}{=} \lim_{|\pi| \to 0} \sum_{[s,t] \in \pi} f(X_s) X_{s,t} + Df(X_s) \mathbb{X}_{s,t}. \tag{5.7}$$

Indeed, we will soon prove that this limit does indeed exist, giving us a notion of "rough integration".

Notice however that the integrand here is not an arbitrary path, but is assumed to be a given function of the path X. This is, unfortunately, a drawback of the theory; the space of valid integrands is actually very restrictive. Happily, as we will see, this is rarely a problem in practice. It is quite typical that the integrand one wishes to consider is of the required form to allow for rough integration, including when considering solutions of differential equations driven by rough paths.

Having said that, we can do a bit better than explicit functions f of X. We now introduce our space of valid integrands, namely the space of *controlled paths.*

In the following we will write e.g. $\mathcal{L}(\mathbb{R}^d;\mathbb{R}^m)$ for the space of linear functions from $\mathbb{R}^d \to \mathbb{R}^m$. Naturally, one is welcome to identify this space with the space of $m \times d$-matrices, but we do not wish to start worrying about such trivial issues as the order in which we write products of variables.

Definition 5.8. Let $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ and $X \in \mathcal{C}^\alpha([0,T];\mathbb{R}^d)$. We say that a pair (Y, Y') is a *controlled path* (with respect to X), if $Y \in \mathcal{C}^\alpha([0,T];\mathbb{R}^m)$, $Y' \in \mathcal{C}^\alpha([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^m))$ and $R^Y \in \mathcal{C}_2^{2\alpha}([0,T];\mathbb{R}^m)$, where $R^Y \colon \Delta_{[0,T]} \to \mathbb{R}^m$ is defined implicitly by

$$Y_{s,t} = Y'_s X_{s,t} + R^Y_{s,t}, \qquad (s,t) \in \Delta_{[0,T]}. \tag{5.8}$$

We write $\mathscr{D}_X^{2\alpha} = \mathscr{D}_X^{2\alpha}([0,T];\mathbb{R}^m)$ for the space of controlled paths (with respect to X).

This definition essentially says that the path Y "looks like" X on very small time scales. We call Y' the *Gubinelli derivative* of Y (with respect to X), and we call R^Y the *remainder.*

It is easy to see that, for a fixed X, the space $\mathscr{D}_X^{2\alpha}$ of controlled paths is a vector space. In fact, it is a Banach space when equipped with the norm

$$\|Y, Y'\|_{\mathscr{D}_X^{2\alpha}} = |Y_0| + |Y'_0| + \|Y'\|_\alpha + \|R^Y\|_{2\alpha}.$$

Note however that the space $\mathscr{D}_X^{2\alpha}$ depends crucially on the choice of the path X.

Note that, given paths X and Y, the Gubinelli derivative Y', when it exists, is not unique in general. For instance, if it happens that $X \in \mathcal{C}^{2\alpha}$ and $Y \in \mathcal{C}^{2\alpha}$, then *any* continuous path Y' would satisfy (5.8) with $\|R^Y\|_{2\alpha} < \infty$. On the other hand, as shown in [FH20, Chapter 6], if X is far from smooth, i.e. genuinely rough in all directions, then Y' is uniquely determined by Y.

Recall that we write $f \in C^k$ whenever a function f is k times continuously differentiable. We will also write $f \in C_b^k$ when additionally f and all its derivatives up to order k are uniformly bounded. Writing $D^k f$ for the k^{th} order derivative of f, we write $\|\cdot\|_{C_b^k}$ for the norm given by

$$\|f\|_{C_b^k} = \|f\|_\infty + \|Df\|_\infty + \cdots + \|D^k f\|_\infty.$$

Example 5.9. Let $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ and $X \in \mathcal{C}^\alpha$. Let $f \in C_b^2$. Then the pair $(f(X), Df(X))$ is a controlled path with respect to X. Indeed, it is clear that $f(X) \in \mathcal{C}^\alpha$ and $Df(X) \in \mathcal{C}^\alpha$, and we have

$$\begin{aligned}\big|f(X_t) - f(X_s) - Df(X_s)X_{s,t}\big| &= \left|\int_0^1 \big(Df(X_s + rX_{s,t}) - Df(X_s)\big) X_{s,t}\,\mathrm{d}r\right| \\ &\le \|f\|_{C_b^2} |X_{s,t}|^2 \le \|f\|_{C_b^2} \|X\|_\alpha^2 |t-s|^{2\alpha},\end{aligned}$$

which implies that $\|R^{f(X)}\|_{2\alpha} \le \|f\|_{C_b^2}\|X\|_\alpha^2 < \infty$, where $R^{f(X)}_{s,t} := f(X_t) - f(X_s) - Df(X_s)X_{s,t}$.

In fact, it is enough to take $f \in C^2$, since the path X is bounded and, since f and its derivatives are continuous, they are locally bounded, and hence bounded on the image of X.

Example 5.10. Let $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ and $X \in \mathcal{C}^\alpha$. Let $(Y, Y'), (Z, Z') \in \mathscr{D}_X^{2\alpha}$ be two controlled paths. Then the product YZ is a controlled path with Gubinelli derivative $(YZ)' = YZ' + Y'Z$. Indeed, we have

$$\begin{aligned} R^{YZ}_{s,t} &:= (YZ)_{s,t} - (YZ)'_s X_{s,t} \\ &= Y_t Z_t - Y_s Z_s - (Y_s Z'_s + Y'_s Z_s) X_{s,t} \\ &= Y_s Z_{s,t} + Y_{s,t} Z_s + Y_{s,t} Z_{s,t} - (Y_s Z'_s + Y'_s Z_s) X_{s,t} \\ &= Y_s R^Z_{s,t} + R^Y_{s,t} Z_s + Y_{s,t} Z_{s,t}, \end{aligned}$$

and hence

$$\|R^{YZ}\|_{2\alpha} \leq \|Y\|_\infty \|R^Z\|_{2\alpha} + \|R^Y\|_{2\alpha} \|Z\|_\infty + \|Y\|_\alpha \|Z\|_\alpha < \infty.$$

5.4 Rough integration

Recall the proof of Proposition 5.2, in which we let $A_{s,t} = Y_s X_{s,t}$, and saw that then $\delta A_{s,u,t} = -Y_{s,u} X_{u,t}$. This then meant that $|\delta A_{s,u,t}| \leq \|Y\|_\beta \|X\|_\alpha |t-s|^{\alpha+\beta}$, and since $\alpha + \beta > 1$, we could then apply the sewing lemma. Clearly, if $\alpha + \beta \leq 1$ then this no longer works. However, once we have lifted a path X to a rough path $\mathbf{X} = (X, \mathbb{X})$, we can get around this by using the additional information encoded in the lift $\mathbb{X}$.

Proposition 5.11. *Let $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ and let $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha([0,T]; \mathbb{R}^d)$ be a rough path. Let $(Y, Y') \in \mathscr{D}_X^{2\alpha}([0,T]; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^m))$ be a controlled path and let R^Y be the corresponding remainder term. Then the limit*

$$\int_0^t Y_u \,\mathrm{d}\mathbf{X}_u := \lim_{|\pi| \to 0} \sum_{[u,v] \in \pi} Y_u X_{u,v} + Y'_u \mathbb{X}_{u,v}$$

exists for every $t \in [0,T]$, where the limit is taken over any sequence of partitions π of the interval $[0,t]$ with mesh size $|\pi| \to 0$. This limit is called the rough integral *of (Y, Y') against $\mathbf{X}$, which moreover comes with the estimate*

$$\left| \int_s^t Y_u \,\mathrm{d}\mathbf{X}_u - Y_s X_{s,t} - Y'_s \mathbb{X}_{s,t} \right| \leq C \big(\|R^Y\|_{2\alpha} \|X\|_\alpha + \|Y'\|_\alpha \|\mathbb{X}\|_{2\alpha} \big) |t-s|^{3\alpha} \tag{5.9}$$

for all $(s,t) \in \Delta_{[0,T]}$, where the constant C depends only on α.

Proof. Let $A_{s,t} = Y_s X_{s,t} + Y'_s \mathbb{X}_{s,t}$, and let $\delta A_{s,u,t} = A_{s,t} - A_{s,u} - A_{u,t}$ for $s \leq u \leq t$. We have

$$\begin{aligned} \delta A_{s,u,t} &= A_{s,t} - A_{s,u} - A_{u,t} \\ &= Y_s X_{s,t} - Y_s X_{s,u} - Y_u X_{u,t} + Y'_s \mathbb{X}_{s,t} - Y'_s \mathbb{X}_{s,u} - Y'_u \mathbb{X}_{u,t} \\ &= Y_s X_{u,t} - Y_u X_{u,t} + Y'_s (\mathbb{X}_{s,t} - \mathbb{X}_{s,u}) - Y'_u \mathbb{X}_{u,t} \\ &= -Y_{s,u} X_{u,t} + Y'_s (\mathbb{X}_{u,t} + X_{s,u} \otimes X_{u,t}) - Y'_u \mathbb{X}_{u,t} \\ &= (-Y_{s,u} + Y'_s X_{s,u}) X_{u,t} - Y'_{s,u} \mathbb{X}_{u,t} \\ &= -R^Y_{s,u} X_{u,t} - Y'_{s,u} \mathbb{X}_{u,t}, \end{aligned}$$

and hence

$$|\delta A_{s,u,t}| = |R^Y_{s,u} X_{u,t} + Y'_{s,u} \mathbb{X}_{u,t}| \leq \big(\|R^Y\|_{2\alpha} \|X\|_\alpha + \|Y'\|_\alpha \|\mathbb{X}\|_{2\alpha} \big) |t-s|^{3\alpha}.$$

Since $3\alpha > 1$, it follows from the sewing lemma (Theorem 5.1) that there exists a continuous path $\gamma =: \int_0^\cdot Y_u \,\mathrm{d}\mathbf{X}_u$ with the desired properties. □

Remark 5.12. Note that the rough integral $\int Y \,\mathrm{d}\mathbf{X}$ is really the integral of the *pair* (Y, Y') against the rough path $\mathbf{X} = (X, \mathbb{X})$. However, the standard convention is to hide the dependence on the Gubinelli derivative Y' in the notation. In practice Y' is generally always clear from the context, so there is rarely any ambiguity in doing this. Nevertheless, one should remember that the choice of Y' does generally matter.

Remark 5.13. In Proposition 5.11, the path Y is prescribed in take values in $\mathcal{L}(\mathbb{R}^d; \mathbb{R}^m)$. Recalling Definition 5.8, we then have that the Gubinelli derivative Y' takes values in $\mathcal{L}(\mathbb{R}^d; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^m))$. In particular, this means that the product $Y'X$ takes values in $\mathcal{L}(\mathbb{R}^d; \mathbb{R}^m)$, consistent with Y, so that the relation $Y_{s,t} = Y'_s X_{s,t} + R^Y_{s,t}$ makes sense. However, here we also identify the space $\mathcal{L}(\mathbb{R}^d; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^m))$ with $\mathcal{L}(\mathbb{R}^{d\times d}; \mathbb{R}^m)$. This allows us to also make sense of the product $Y'\mathbb{X}$, which then takes values in $\mathbb{R}^m$.

We saw in Example 5.9 that for any function $f \in C^2$, the pair $(f(X), Df(X))$ is a controlled path. Hence, the rough integral $\int_0^T f(X_r) \,\mathrm{d}\mathbf{X}_r$ exists, and is given by the limit in (5.7).

For any $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ and $(Y, Y') \in \mathscr{D}^{2\alpha}_X$, the pair $(Z, Z') := (\int_0^\cdot Y_u \,\mathrm{d}\mathbf{X}_u, Y)$ is itself another controlled path with respect to X. Indeed, with $R^Z_{s,t} := Z_{s,t} - Z'_s X_{s,t}$, we see from (5.9) that

$$
\begin{aligned}
|R^Z_{s,t}| &= \left| \int_s^t Y_u \,\mathrm{d}\mathbf{X}_u - Y_s X_{s,t} \right| \\
&\leq |Y'_s \mathbb{X}_{s,t}| + C\big(\|R^Y\|_{2\alpha} \|X\|_\alpha + \|Y'\|_\alpha \|\mathbb{X}\|_{2\alpha}\big)|t-s|^{3\alpha} \\
&\leq \|Y'\|_\infty \|\mathbb{X}\|_{2\alpha} |t-s|^{2\alpha} + C\big(\|R^Y\|_{2\alpha} \|X\|_\alpha + \|Y'\|_\alpha \|\mathbb{X}\|_{2\alpha}\big)|t-s|^{3\alpha},
\end{aligned}
$$

and hence that $\|R^Z\|_{2\alpha} < \infty$.

In future we will denote R^Z by $R^{\int_0^\cdot Y_u \,\mathrm{d}\mathbf{X}_u}$.

Lemma 5.14. *For some $\alpha \in (\frac{1}{3}, \frac{1}{2}]$, let $F \in \mathcal{C}^{2\alpha}$ be a 2α-Hölder continuous path, and let $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ and $\tilde{\mathbf{X}} = (\tilde{X}, \tilde{\mathbb{X}}) \in \mathscr{C}^\alpha$ be two rough paths such that*

$$
\tilde{X}_t = X_t, \qquad\qquad \tilde{\mathbb{X}}_{s,t} = \mathbb{X}_{s,t} + F_{s,t}.
$$

Let $(Y, Y') \in \mathscr{D}^{2\alpha}_X = \mathscr{D}^{2\alpha}_{\tilde{X}}$. Then

$$
\int_0^T Y_u \,\mathrm{d}\tilde{\mathbf{X}}_u = \int_0^T Y_u \,\mathrm{d}\mathbf{X}_u + \int_0^T Y'_u \,\mathrm{d}F_u.
$$

The proof of Lemma 5.14 is left as an exercise.

Theorem 5.15 (Stability of rough integration)**.** *Let $\alpha \in (\frac{1}{3}, \frac{1}{2}]$, and let $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ and $\tilde{\mathbf{X}} = (\tilde{X}, \tilde{\mathbb{X}}) \in \mathscr{C}^\alpha$ be rough paths. Let $(Y, Y') \in \mathscr{D}^{2\alpha}_X$ and $(\tilde{Y}, \tilde{Y}') \in \mathscr{D}^{2\alpha}_{\tilde{X}}$ be controlled paths, and let R^Y and $R^{\tilde{Y}}$ be the corresponding remainder terms. There exists a constant C, depending only on α and T, such that*

$$
\begin{aligned}
\|Y - \tilde{Y}\|_\alpha \leq C\Big(&\big(|Y'_0 - \tilde{Y}'_0| + \|Y' - \tilde{Y}'\|_\alpha\big)\|X\|_\alpha \\
&+ \big(|\tilde{Y}'_0| + \|\tilde{Y}'\|_\alpha\big)\|X - \tilde{X}\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha} T^\alpha \Big) \qquad (5.10)
\end{aligned}
$$

and

$$\left\|R^{\int_0^\cdot Y_u \,\mathrm{d}\mathbf{X}_u} - R^{\int_0^\cdot \tilde{Y}_u \,\mathrm{d}\tilde{\mathbf{X}}_u}\right\|_{2\alpha} \leq C\Big(\big(|Y_0' - \tilde{Y}_0'| + \|Y' - \tilde{Y}'\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha}\big)\|\mathbf{X}\|_\alpha \\ + \big(|\tilde{Y}_0'| + \|\tilde{Y}'\|_\alpha + \|R^{\tilde{Y}}\|_{2\alpha}\big)\|\mathbf{X};\tilde{\mathbf{X}}\|_\alpha\Big). \qquad (5.11)$$

Proof. We have

$$\begin{aligned}|Y_{s,t} - \tilde{Y}_{s,t}| &= \big|Y_s' X_{s,t} + R^Y_{s,t} - \tilde{Y}_s'\tilde{X}_{s,t} - R^{\tilde{Y}}_{s,t}\big| \\ &\leq |Y_s' - \tilde{Y}_s'||X_{s,t}| + |\tilde{Y}_s'||X_{s,t} - \tilde{X}_{s,t}| + |R^Y_{s,t} - R^{\tilde{Y}}_{s,t}| \\ &\leq \big(\|Y' - \tilde{Y}'\|_\infty \|X\|_\alpha + \|\tilde{Y}'\|_\infty \|X - \tilde{X}\|_\alpha\big)|t-s|^\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha}|t-s|^{2\alpha},\end{aligned}$$

so that

$$\|Y - \tilde{Y}\|_\alpha \leq \|Y' - \tilde{Y}'\|_\infty \|X\|_\alpha + \|\tilde{Y}'\|_\infty \|X - \tilde{X}\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha} T^\alpha,$$

which gives the estimate in (5.10).

Let $A_{s,t} = Y_s X_{s,t} + Y_s' \mathbb{X}_{s,t}$, $\tilde{A}_{s,t} = \tilde{Y}_s \tilde{X}_{s,t} + \tilde{Y}_s' \tilde{\mathbb{X}}_{s,t}$, $\Delta = A - \tilde{A}$, and $\delta\Delta_{s,u,t} = \Delta_{s,t} - \Delta_{s,u} - \Delta_{u,t}$. Then

$$\delta\Delta_{s,u,t} = \delta A_{s,u,t} - \delta\tilde{A}_{s,u,t} = -\big(R^Y_{s,u}X_{u,t} + Y'_{s,u}\mathbb{X}_{u,t} - R^{\tilde{Y}}_{s,u}\tilde{X}_{u,t} - \tilde{Y}'_{s,u}\tilde{\mathbb{X}}_{u,t}\big),$$

so that

$$\begin{aligned}|\delta\Delta_{s,u,t}| &= \big|R^Y_{s,u}X_{u,t} + Y'_{s,u}\mathbb{X}_{u,t} - R^{\tilde{Y}}_{s,u}\tilde{X}_{u,t} - \tilde{Y}'_{s,u}\tilde{\mathbb{X}}_{u,t}\big| \\ &\leq |R^Y_{s,u} - R^{\tilde{Y}}_{s,u}||X_{u,t}| + |R^{\tilde{Y}}_{s,u}||X_{u,t} - \tilde{X}_{u,t}| + |Y'_{s,u} - \tilde{Y}'_{s,u}||\mathbb{X}_{u,t}| + |\tilde{Y}'_{s,u}||\mathbb{X}_{u,t} - \tilde{\mathbb{X}}_{u,t}| \\ &\leq \big(\|R^Y - R^{\tilde{Y}}\|_{2\alpha}\|X\|_\alpha + \|R^{\tilde{Y}}\|_{2\alpha}\|X - \tilde{X}\|_\alpha \\ &\qquad + \|Y' - \tilde{Y}'\|_\alpha \|\mathbb{X}\|_{2\alpha} + \|\tilde{Y}'\|_\alpha \|\mathbb{X} - \tilde{\mathbb{X}}\|_{2\alpha}\big)|t-s|^{3\alpha}.\end{aligned}$$

By the sewing lemma (Theorem 5.1), there exists a path γ and a constant C such that

$$|\gamma_t - \gamma_s - \Delta_{s,t}| \leq C\big(\|R^Y - R^{\tilde{Y}}\|_{2\alpha}\|X\|_\alpha + \|R^{\tilde{Y}}\|_{2\alpha}\|X - \tilde{X}\|_\alpha \\ + \|Y' - \tilde{Y}'\|_\alpha \|\mathbb{X}\|_{2\alpha} + \|\tilde{Y}'\|_\alpha \|\mathbb{X} - \tilde{\mathbb{X}}\|_{2\alpha}\big)|t-s|^{3\alpha}$$

and, letting π denote a partition of the interval $[s,t]$,

$$\gamma_t - \gamma_s = \lim_{|\pi|\to 0} \sum_{[u,v]\in\pi} \Delta_{u,v} = \lim_{|\pi|\to 0}\Bigg(\sum_{[u,v]\in\pi} A_{u,v} - \sum_{[u,v]\in\pi} \tilde{A}_{u,v}\Bigg) = \int_s^t Y_u \,\mathrm{d}\mathbf{X}_u - \int_s^t \tilde{Y}_u \,\mathrm{d}\tilde{\mathbf{X}}_u.$$

Combining the above, we have that

$$\begin{aligned}&\left|\int_s^t Y_u \,\mathrm{d}\mathbf{X}_u - \int_s^t \tilde{Y}_u \,\mathrm{d}\tilde{\mathbf{X}}_u - \big(Y_s X_{s,t} + Y_s'\mathbb{X}_{s,t} - \tilde{Y}_s\tilde{X}_{s,t} - \tilde{Y}_s'\tilde{\mathbb{X}}_{s,t}\big)\right| \\ &\leq C\big(\|R^Y - R^{\tilde{Y}}\|_{2\alpha}\|X\|_\alpha + \|R^{\tilde{Y}}\|_{2\alpha}\|X - \tilde{X}\|_\alpha \\ &\qquad + \|Y' - \tilde{Y}'\|_\alpha \|\mathbb{X}\|_{2\alpha} + \|\tilde{Y}'\|_\alpha \|\mathbb{X} - \tilde{\mathbb{X}}\|_{2\alpha}\big)|t-s|^{3\alpha}. \qquad (5.12)\end{aligned}$$

We also have

$$\begin{aligned}|Y'_s\mathbb{X}_{s,t} - \tilde{Y}'_s\tilde{\mathbb{X}}_{s,t}| &\leq |Y'_s - \tilde{Y}'_s||\mathbb{X}_{s,t}| + |\tilde{Y}'_s||\mathbb{X}_{s,t} - \tilde{\mathbb{X}}_{s,t}| \\ &\leq \big(\|Y' - \tilde{Y}'\|_\infty\|\mathbb{X}\|_{2\alpha} + \|\tilde{Y}'\|_\infty\|\mathbb{X} - \tilde{\mathbb{X}}\|_{2\alpha}\big)|t-s|^{2\alpha}.\end{aligned} \tag{5.13}$$

Combining (5.12) and (5.13), we have

$$\begin{aligned}\big|R^{\int_0^\cdot Y_u\,\mathrm{d}\mathbf{X}_u}_{s,t} - R^{\int_0^\cdot \tilde{Y}_u\,\mathrm{d}\tilde{\mathbf{X}}_u}_{s,t}\big| &= \left|\int_s^t Y_u\,\mathrm{d}\mathbf{X}_u - Y_sX_{s,t} - \int_s^t \tilde{Y}_u\,\mathrm{d}\tilde{\mathbf{X}}_u + \tilde{Y}_s\tilde{X}_{s,t}\right| \\ &\leq \big(\|Y' - \tilde{Y}'\|_\infty\|\mathbb{X}\|_{2\alpha} + \|\tilde{Y}'\|_\infty\|\mathbb{X} - \tilde{\mathbb{X}}\|_{2\alpha}\big)|t-s|^{2\alpha} \\ &\quad + C\Big(\|R^Y - R^{\tilde{Y}}\|_{2\alpha}\|X\|_\alpha + \|R^{\tilde{Y}}\|_{2\alpha}\|X - \tilde{X}\|_\alpha \\ &\qquad + \|Y' - \tilde{Y}'\|_\alpha\|\mathbb{X}\|_{2\alpha} + \|\tilde{Y}'\|_\alpha\|\mathbb{X} - \tilde{\mathbb{X}}\|_{2\alpha}\Big)|t-s|^{3\alpha},\end{aligned}$$

and hence that

$$\begin{aligned}\big\|R^{\int_0^\cdot Y_u\,\mathrm{d}\mathbf{X}_u} - R^{\int_0^\cdot \tilde{Y}_u\,\mathrm{d}\tilde{\mathbf{X}}_u}\big\|_{2\alpha} &\leq \big(\|Y' - \tilde{Y}'\|_\infty\|\mathbb{X}\|_{2\alpha} + \|\tilde{Y}'\|_\infty\|\mathbb{X} - \tilde{\mathbb{X}}\|_{2\alpha}\big) \\ &\quad + C\Big(\|R^Y - R^{\tilde{Y}}\|_{2\alpha}\|X\|_\alpha + \|R^{\tilde{Y}}\|_{2\alpha}\|X - \tilde{X}\|_\alpha \\ &\qquad + \|Y' - \tilde{Y}'\|_\alpha\|\mathbb{X}\|_{2\alpha} + \|\tilde{Y}'\|_\alpha\|\mathbb{X} - \tilde{\mathbb{X}}\|_{2\alpha}\Big)T^\alpha,\end{aligned}$$

which gives the estimate in (5.11) for a new constant C. □

Corollary 5.16. *Let $\alpha \in (\frac{1}{3}, \frac{1}{2}]$, and let $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ and $\tilde{\mathbf{X}} = (\tilde{X}, \tilde{\mathbb{X}}) \in \mathscr{C}^\alpha$ be rough paths. Let $(Y, Y') \in \mathscr{D}^{2\alpha}_X$ and $(\tilde{Y}, \tilde{Y}') \in \mathscr{D}^{2\alpha}_{\tilde{X}}$ be controlled paths, and let R^Y and $R^{\tilde{Y}}$ be the corresponding remainder terms. There exists a constant C, depending only on α and T, such that*

$$\begin{aligned}&\left\|\int_0^\cdot Y_u\,\mathrm{d}\mathbf{X}_u - \int_0^\cdot \tilde{Y}_u\,\mathrm{d}\tilde{\mathbf{X}}_u\right\|_\alpha \\ &\leq C\big(1 + \|X\|_\alpha + \|\tilde{X}\|_\alpha\big)\Big(\|\tilde{Y}, \tilde{Y}'\|_{\mathscr{D}^{2\alpha}_{\tilde{X}}}\|\mathbf{X}; \tilde{\mathbf{X}}\|_\alpha \\ &\qquad + \big(|Y_0 - \tilde{Y}_0| + |Y'_0 - \tilde{Y}'_0| + \|Y' - \tilde{Y}'\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha}\big)\|\!|\mathbf{X}|\!\|_\alpha\Big).\end{aligned}$$

Proof. Since $(\int_0^\cdot Y_u\,\mathrm{d}\mathbf{X}_u, Y) \in \mathscr{D}^{2\alpha}_X$ and $(\int_0^\cdot \tilde{Y}_u\,\mathrm{d}\tilde{\mathbf{X}}_u, \tilde{Y}) \in \mathscr{D}^{2\alpha}_{\tilde{X}}$ are controlled paths, we have

$$\begin{aligned}&\left|\int_s^t Y_u\,\mathrm{d}\mathbf{X}_u - \int_s^t \tilde{Y}_u\,\mathrm{d}\tilde{\mathbf{X}}_u\right| = \big|Y_sX_{s,t} + R^{\int_0^\cdot Y_u\,\mathrm{d}\mathbf{X}_u}_{s,t} - \tilde{Y}_s\tilde{X}_{s,t} - R^{\int_0^\cdot \tilde{Y}_u\,\mathrm{d}\tilde{\mathbf{X}}_u}_{s,t}\big| \\ &\leq |Y_s - \tilde{Y}_s||X_{s,t}| + |\tilde{Y}_s||X_{s,t} - \tilde{X}_{s,t}| + \big|R^{\int_0^\cdot Y_u\,\mathrm{d}\mathbf{X}_u}_{s,t} - R^{\int_0^\cdot \tilde{Y}_u\,\mathrm{d}\tilde{\mathbf{X}}_u}_{s,t}\big| \\ &\leq \big(\|Y - \tilde{Y}\|_\infty\|X\|_\alpha + \|\tilde{Y}\|_\infty\|X - \tilde{X}\|_\alpha\big)|t-s|^\alpha + \big\|R^{\int_0^\cdot Y_u\,\mathrm{d}\mathbf{X}_u} - R^{\int_0^\cdot \tilde{Y}_u\,\mathrm{d}\tilde{\mathbf{X}}_u}\big\|_{2\alpha}|t-s|^{2\alpha},\end{aligned}$$

so that

$$\begin{aligned}&\left\|\int_0^\cdot Y_u\,\mathrm{d}\mathbf{X}_u - \int_0^\cdot \tilde{Y}_u\,\mathrm{d}\tilde{\mathbf{X}}_u\right\|_\alpha \\ &\leq \|Y - \tilde{Y}\|_\infty\|X\|_\alpha + \|\tilde{Y}\|_\infty\|X - \tilde{X}\|_\alpha + \big\|R^{\int_0^\cdot Y_u\,\mathrm{d}\mathbf{X}_u} - R^{\int_0^\cdot \tilde{Y}_u\,\mathrm{d}\tilde{\mathbf{X}}_u}\big\|_{2\alpha}T^\alpha.\end{aligned} \tag{5.14}$$

Since $\tilde{Y}_t = \tilde{Y}_0 + \tilde{Y}_{0,t}$, we have $|\tilde{Y}_t| \le |\tilde{Y}_0| + |\tilde{Y}_{0,t}| \le |\tilde{Y}_0| + \|\tilde{Y}\|_\alpha T^\alpha$, so that

$$\|\tilde{Y}\|_\infty \le |\tilde{Y}_0| + \|\tilde{Y}\|_\alpha T^\alpha,$$

and similarly

$$\|Y - \tilde{Y}\|_\infty \le |Y_0 - \tilde{Y}_0| + \|Y - \tilde{Y}\|_\alpha T^\alpha.$$

It is also easy to see from $\tilde{Y}_{s,t} = \tilde{Y}'_s \tilde{X}_{s,t} + R^{\tilde{Y}}_{s,t}$ that

$$\|\tilde{Y}\|_\alpha \le (|\tilde{Y}'_0| + \|\tilde{Y}'\|_\alpha T^\alpha)\|\tilde{X}\|_\alpha + \|R^{\tilde{Y}}\|_{2\alpha} T^\alpha.$$

Substituting these estimates along with (5.10) and (5.11) into (5.14), we deduce the desired inequality. □

6 Further topics in rough integration

In this section we will see how rough path theory allows various results from stochastic calculus to be recovered in a pathwise sense, without the use of probability.

6.1 Associativity

We have seen how one can define the rough integral of a controlled path (Y, Y') against a rough path $\mathbf{X}$. Recalling the definition of controlled paths, this amounts to saying that we know how to integrate a path that "locally looks like X" against X itself. One may then point out that, since any two controlled paths (with respect to X) both locally look like X, they actually look locally like *each other*, which suggests that it should be possible to integrate controlled paths against each other. This is indeed the case, as the next proposition shows.

Proposition 6.1. *Let $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ be a rough path, and let $(Y, Y'), (Z, Z') \in \mathscr{D}^{2\alpha}_X$ be two controlled paths with remainders R^Y and R^Z respectively. Then the limit*

$$\int_0^t Y_u \,\mathrm{d}Z_u := \lim_{|\pi| \to 0} \sum_{[u,v] \in \pi} Y_u Z_{u,v} + Y'_u Z'_u \mathbb{X}_{u,v}$$

exists for every $t \in [0, T]$, where the limit is taken over any sequence of partitions π of the interval $[0, t]$ with mesh size $|\pi| \to 0$. Moreover, we have the estimate

$$\begin{aligned}\left| \int_s^t Y_u \,\mathrm{d}Z_u - Y_s Z_{s,t} - Y'_s Z'_s \mathbb{X}_{s,t} \right| \le C\big(&\|Y'\|_\infty \|Z'\|_\alpha \|X\|_\alpha^2 + \|Y\|_\alpha \|R^Z\|_{2\alpha} \\ &+ \|R^Y\|_{2\alpha} \|Z'\|_\infty \|X\|_\alpha + \|Y'Z'\|_\alpha \|\mathbb{X}\|_{2\alpha}\big)|t - s|^{3\alpha}\end{aligned} \tag{6.1}$$

for all $(s, t) \in \Delta_{[0,T]}$, where the constant C depends only on α.

Proof. Let $A_{s,t} = Y_s Z_{s,t} + Y'_s Z'_s \mathbb{X}_{s,t}$ and define $\delta A_{s,u,t} = A_{s,t} - A_{s,u} - A_{u,t}$. Using Chen's relation (3.1), one can show that

$$\delta A_{s,u,t} = -Y'_s Z'_{s,u} X_{s,u} \otimes X_{u,t} - Y_{s,u} R^Z_{u,t} - R^Y_{s,u} Z'_u X_{u,t} - (Y'Z')_{s,u} \mathbb{X}_{u,t},$$

so that

$$|\delta A_{s,u,t}| \le \big(\|Y'\|_\infty \|Z'\|_\alpha \|X\|_\alpha^2 + \|Y\|_\alpha \|R^Z\|_{2\alpha} \\ + \|R^Y\|_{2\alpha}\|Z'\|_\infty \|X\|_\alpha + \|Y'Z'\|_\alpha \|\mathbb{X}\|_{2\alpha}\big)|t-s|^{3\alpha}.$$

By the sewing lemma (Theorem 5.1), there exists a continuous path $\gamma =: \int_0^{\cdot} Y_u \,\mathrm{d}Z_u$ with the desired properties. $\square$

We exhibited in Lemma 5.7 the associativity of Young integration. We can now give a corresponding result for rough integration.

Proposition 6.2. *Let $\mathbf{X} = (X, \mathbb{X})$ be a rough path and let $(Y, Y'), (K, K') \in \mathscr{D}_X^{2\alpha}$ be two controlled paths, so that in particular the rough integral $\int_0^{\cdot} K_u \,\mathrm{d}\mathbf{X}_u$ exists by Proposition 5.11, and the pair $(Z, Z') := (\int_0^{\cdot} K_u \,\mathrm{d}\mathbf{X}_u, K) \in \mathscr{D}_X^{2\alpha}$ is also a controlled path. Then*

$$\int_0^{\cdot} Y_u \,\mathrm{d}Z_u = \int_0^{\cdot} Y_u K_u \,\mathrm{d}\mathbf{X}_u,$$

where on the left-hand side we have the integral of (Y, Y') against (Z, Z') as defined in Proposition 6.1, and on the right-hand side we have the rough integral of $(YK, (YK)')$ against $\mathbf{X}$.

Proof. Recall from Example 5.10 that the product YK is itself a controlled path with Gubinelli derivative $(YK)' = YK' + Y'K$. It follows from (5.9) that

$$Z_{s,t} = \int_s^t K_u \,\mathrm{d}\mathbf{X}_u = K_s X_{s,t} + K'_s \mathbb{X}_{s,t} + O(|t-s|^{3\alpha})$$

and

$$\int_s^t Y_u K_u \,\mathrm{d}\mathbf{X}_u = Y_s K_s X_{s,t} + (YK)'_s \mathbb{X}_{s,t} + O(|t-s|^{3\alpha}).$$

Similarly, by (6.1), we have

$$\int_s^t Y_u \,\mathrm{d}Z_u = Y_s Z_{s,t} + Y'_s Z'_s \mathbb{X}_{s,t} + O(|t-s|^{3\alpha}).$$

We then calculate

$$\begin{aligned}
\int_s^t Y_u \,\mathrm{d}Z_u &= Y_s Z_{s,t} + Y'_s Z'_s \mathbb{X}_{s,t} + O(|t-s|^{3\alpha}) \\
&= Y_s\big(K_s X_{s,t} + K'_s \mathbb{X}_{s,t}\big) + Y'_s K_s \mathbb{X}_{s,t} + O(|t-s|^{3\alpha}) \\
&= Y_s K_s X_{s,t} + \big(Y_s K'_s + Y'_s K_s\big)\mathbb{X}_{s,t} + O(|t-s|^{3\alpha}) \\
&= Y_s K_s X_{s,t} + (YK)'_s \mathbb{X}_{s,t} + O(|t-s|^{3\alpha}) \\
&= \int_s^t Y_u K_u \,\mathrm{d}\mathbf{X}_u + O(|t-s|^{3\alpha}).
\end{aligned}$$

Taking $\lim_{|\pi| \to 0} \sum_{[s,t] \in \pi}$ on both sides, we obtain $\int_0^T Y_u \,\mathrm{d}Z_u = \int_0^T Y_u K_u \,\mathrm{d}\mathbf{X}_u$. $\square$

Any controlled path with respect to a rough path can itself be lifted in a canonical way to a rough path. Indeed, let $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ be a rough path, and let $(Z, Z') \in \mathscr{D}_X^{2\alpha}$ be a controlled path. Define

$$\mathbb{Z}_{s,t} := \int_s^t Z_{s,u} \, \mathrm{d}Z_u, \qquad (s,t) \in \Delta_{[0,T]}, \tag{6.2}$$

where the integral is defined in the sense of Proposition 6.1. It is easy to see that the pair $(Z, \mathbb{Z})$ satisfies Chen's relation, and it follows from the estimate in (6.1) that $\mathbb{Z}$ is 2α-Hölder continuous. Thus, the pair $\mathbf{Z} = (Z, \mathbb{Z})$ is another rough path.

Notice that, if $\mathbf{Z}$ is the rough path defined above, then we can integrate a controlled path Y either with respect to Z as in Proposition 6.1, or against $\mathbf{Z}$ as in Proposition 5.11. The next lemma shows that these two different notions of rough integral coincide.

Lemma 6.3. *Let $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ be a rough path. Let $(Z, Z') \in \mathscr{D}_X^{2\alpha}$ be a controlled path, and let $\mathbf{Z} = (Z, \mathbb{Z})$ be the canonical rough path lift of Z, as defined in (6.2). Let $(Y, Y') \in \mathscr{D}_Z^{2\alpha}$ be a controlled path with respect to Z. Then $(Y, Y'Z') \in \mathscr{D}_X^{2\alpha}$ is a controlled path with respect to X, and, moreover, we have that*

$$\int_0^\cdot Y_u \, \mathrm{d}\mathbf{Z}_u = \int_0^\cdot Y_u \, \mathrm{d}Z_u,$$

where on the left-hand side we have the rough integral of (Y, Y') against $\mathbf{Z}$, and on the right-hand side we have the integral of $(Y, Y'Z')$ against (Z, Z') as defined in Proposition 6.1.

Proof. We have that $Z_{s,t} = Z'_s X_{s,t} + R^Z_{s,t}$ and $Y_{s,t} = Y'_s Z_{s,t} + R^Y_{s,t}$. Then

$$Y_{s,t} - Y'_s Z'_s X_{s,t} = Y_{s,t} - Y'_s (Z_{s,t} - R^Z_{s,t}) = R^Y_{s,t} + Y'_s R^Z_{s,t}.$$

Since the right-hand side is 2α-Hölder continuous, we see that $(Y, Y'Z') \in \mathscr{D}_X^{2\alpha}$.

Similarly to the proof of Proposition 6.2, using the estimates in (5.9) and (6.1), we have

$$\begin{aligned} \int_s^t Y_u \, \mathrm{d}\mathbf{Z}_u &= Y_s Z_{s,t} + Y'_s \mathbb{Z}_{s,t} + O(|t-s|^{3\alpha}) \\ &= Y_s Z_{s,t} + Y'_s Z'_s Z'_s \mathbb{X}_{s,t} + O(|t-s|^{3\alpha}) \\ &= \int_s^t Y_u \, \mathrm{d}Z_u + O(|t-s|^{3\alpha}). \end{aligned}$$

Taking $\lim_{|\pi| \to 0} \sum_{[s,t] \in \pi}$ on both sides, we obtain $\int_0^T Y_u \, \mathrm{d}\mathbf{Z}_u = \int_0^T Y_u \, \mathrm{d}Z_u$. $\square$

6.2 The bracket of a rough path

An important object in stochastic calculus is the *quadratic variation* of a stochastic process. Rough paths do not generally admit quadratic variation (although there are cases when they do). However, there is a corresponding object in rough path theory which plays the same role that the quadratic variation does in stochastic calculus. This object is called the *bracket* of a rough path.

Definition 6.4. Let $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ be a rough path, and let $\mathrm{Sym}(\mathbb{X})$ denote the symmetric part of $\mathbb{X}$. The bracket of $\mathbf{X}$ is defined as the path $[\mathbf{X}]\colon [0,T] \to \mathbb{R}^{d\times d}$ given by

$$[\mathbf{X}]_t := X_{0,t} \otimes X_{0,t} - 2\,\mathrm{Sym}(\mathbb{X}_{0,t}).$$

Lemma 6.5. *Let* $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$*. Then*

$$[\mathbf{X}]_{s,t} = X_{s,t} \otimes X_{s,t} - 2\,\mathrm{Sym}(\mathbb{X}_{s,t})$$

for all $(s,t) \in \Delta_{[0,T]}$*. In particular, we have that* $[\mathbf{X}] \in \mathcal{C}^{2\alpha}$*.*

The proof of Lemma 6.5 is left as an exercise.

Example 6.6. Recall that a rough path $\mathbf{X} = (X, \mathbb{X})$ is said to be weakly geometric if it satisfies the equality

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

We therefore see that a rough path $\mathbf{X}$ is weakly geometric if and only if $[\mathbf{X}]_t = 0$ for all $t \in [0,T]$.

Example 6.7. Let B be a Brownian motion, and let $\mathbf{B} = (B, \mathbb{B}^{\text{Itô}})$ be the Itô lift of B, as defined in (4.4). Recall from (4.5) that the symmetric part of the Itô enhancement is given by

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

Thus,

$$[\mathbf{B}]_{s,t} = (t-s)I.$$

At least in the case of Itô Brownian rough path, we see that the bracket of Brownian motion does in fact coincide with its quadratic variation.

Lemma 6.8. *Let* $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ *be a rough path and let* $(K, K') \in \mathscr{D}^{2\alpha}_X$*. Recall that* $(Z, Z') := (\int_0^\cdot K_u \,\mathrm{d}\mathbf{X}_u, K) \in \mathscr{D}^{2\alpha}_X$*. Let* $\mathbf{Z} = (Z, \mathbb{Z})$ *be the canonical rough path lift of* Z*, as defined in* (6.2)*, so that in particular the bracket* $[\mathbf{Z}]$ *of* $\mathbf{Z}$ *exists. Then*

$$[\mathbf{Z}] = \int_0^\cdot (K_u \otimes K_u)\,\mathrm{d}[\mathbf{X}]_u,$$

where the integral on the right-hand side is a Young integral.

Proof. Since $[\mathbf{X}]$ is 2α-Hölder continuous,

$$\int_0^T (K_u \otimes K_u)\,\mathrm{d}[\mathbf{X}]_u = \lim_{|\pi|\to 0} \sum_{[s,t]\in\pi} (K_s \otimes K_s)[\mathbf{X}]_{s,t}$$

exists as a Young integral. We have

$$\begin{aligned}
[\mathbf{Z}]_{s,t} &= Z_{s,t} \otimes Z_{s,t} - 2\,\mathrm{Sym}(\mathbb{Z}_{s,t}) \\
&= (K_s X_{s,t} + K'_s \mathbb{X}_{s,t}) \otimes (K_s X_{s,t} + K'_s \mathbb{X}_{s,t}) - 2(Z'_s \otimes Z'_s)\,\mathrm{Sym}(\mathbb{X}_{s,t}) + O(|t-s|^{3\alpha}) \\
&= (K_s X_{s,t}) \otimes (K_s X_{s,t}) - 2(K_s \otimes K_s)\,\mathrm{Sym}(\mathbb{X}_{s,t}) + O(|t-s|^{3\alpha}) \\
&= (K_s \otimes K_s)[\mathbf{X}]_{s,t} + O(|t-s|^{3\alpha}).
\end{aligned}$$

Taking $\lim_{|\pi|\to 0} \sum_{[s,t]\in\pi}$ on both sides, we obtain $[\mathbf{Z}]_T = \int_0^T (K_u \otimes K_u)\,\mathrm{d}[\mathbf{X}]_u$. $\square$

Abusing notation slightly, one could rewrite the result of Lemma 6.8 as

$$\left[\int_0^{\cdot} K_u \,\mathrm{d}\mathbf{X}_u\right]_t = \int_0^t K_u^2 \,\mathrm{d}[\mathbf{X}]_u,$$

giving us an analogous result to the well-known formula for the quadratic variation of Itô integrals.

6.3 The Itô formula for rough paths

One of the most useful results in stochastic calculus is Itô's formula, which plays the role of the chain rule/fundamental theorem of calculus. In the setting of rough paths, we have the following analogous result.

Proposition 6.9. *Let $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ be a rough path, and let $f \in C^3$. Then*

$$f(X_T) = f(X_0) + \int_0^T Df(X_u)\,\mathrm{d}\mathbf{X}_u + \frac{1}{2}\int_0^T D^2 f(X_u)\,\mathrm{d}[\mathbf{X}]_u,$$

where the first integral on the right-hand side is the rough integral of $(Df(X), D^2 f(X))$ against $\mathbf{X}$, and the second integral is the Young integral of $D^2 f(X)$ against the bracket $[\mathbf{X}]$.

Proof. Since X is bounded, we may assume without loss of generality that $f \in C_b^3$. Since $Df \in C_b^2$, we have that the pair $(Df(X), D^2 f(X))$ is indeed a controlled path with respect to X. We have

$$\begin{aligned} f(X_t) - f(X_s) &= Df(X_s)X_{s,t} + \frac{1}{2}D^2 f(X_s)(X_{s,t}\otimes X_{s,t}) + R_{s,t} \\ &= Df(X_s)X_{s,t} + D^2 f(X_s)\mathbb{X}_{s,t} + \frac{1}{2}D^2 f(X_s)(X_{s,t}\otimes X_{s,t}) \\ &\quad - D^2 f(X_s)\mathbb{X}_{s,t} + R_{s,t}, \end{aligned}$$

where

$$R_{s,t} := \int_0^1\int_0^1 \big(D^2 f(X_s + r_1 r_2 X_{s,t}) - D^2 f(X_s)\big)(X_{s,t}\otimes X_{s,t}) r_1 \,\mathrm{d}r_2\,\mathrm{d}r_1.$$

Note that

$$|R_{s,t}| \le \|f\|_{C_b^3}|X_{s,t}|^3 \le \|f\|_{C_b^3}\|X\|_\alpha^3 |t-s|^{3\alpha},$$

so that $\lim_{|\pi|\to 0}\sum_{[s,t]\in\pi}|R_{s,t}| = 0$.

Recall that the contraction of a symmetric matrix with an antisymmetric matrix is zero. That is, if A is symmetric and B is antisymmetric, then $\sum_{i,j} A^{ij}B^{ij} = -\sum_{i,j} A^{ji}B^{ji} = -\sum_{i,j} A^{ij}B^{ij}$, which implies that $\sum_{i,j} A^{ij}B^{ij} = 0$.

Since the Hessian matrix $D^2 f(X_s)$ is symmetric, it therefore kills the antisymmetric part of $\mathbb{X}$. Thus,

$$D^2 f(X_s)\mathbb{X}_{s,t} = D^2 f(X_s)\,\mathrm{Sym}(\mathbb{X}_{s,t}).$$

By Lemma 6.5, we then have that

$$f(X_t) - f(X_s) = \big(Df(X_s)X_{s,t} + D^2 f(X_s)\mathbb{X}_{s,t}\big) + \frac{1}{2}D^2 f(X_s)[\mathbf{X}]_{s,t} + R_{s,t}.$$

Taking $\lim_{|\pi|\to 0}\sum_{[s,t]\in\pi}$ on both sides, we deduce the result. □

Remark 6.10. In Proposition 6.9, and also in Proposition 6.11 below, it is actually enough to take $f \in C^{\frac{1}{\alpha}+\varepsilon}$ for any $\varepsilon > 0$. This can be shown by using the more general notion of (β, γ)-controlled paths introduced in the second exercise sheet.

Proposition 6.11. *Let $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ be a rough path, and suppose that $(Y, Y') \in \mathscr{D}_X^{2\alpha}$ and $(Y', Y'') \in \mathscr{D}_X^{2\alpha}$ are controlled paths. Suppose further that*

$$Y_t = Y_0 + \int_0^t Y'_s \, \mathrm{d}\mathbf{X}_s + \Gamma_t$$

for all $t \in [0, T]$, for some path $\Gamma \in \mathcal{C}^{2\alpha}$. Let $f \in C^3$. Then

$$f(Y_T) = f(Y_0) + \int_0^T Df(Y_u) Y'_u \, \mathrm{d}\mathbf{X}_u + \int_0^T Df(Y_u) \, \mathrm{d}\Gamma_u + \frac{1}{2} \int_0^T D^2 f(Y_u)(Y'_u \otimes Y'_u) \, \mathrm{d}[\mathbf{X}]_u.$$

The proof of Proposition 6.11 is left as an exercise.

6.4 The rough exponential

In this section we shall see our first example of a rough differential equation (RDE).

In the following, we will write $\|X\|_{\alpha,[s,t]}$ for the α-Hölder seminorm over the interval $[s, t]$, i.e. $\sup_{s \le u < v \le t} |X_{u,v}|/|v - u|^\alpha$. We also define $\|\mathbb{X}\|_{2\alpha,[s,t]}$ and $|||\mathbf{X}|||_{\alpha,[s,t]}$ similarly.

Proposition 6.12. *Let $\beta \in (\frac{1}{3}, \frac{1}{2}]$, and let $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\beta$ be rough path over $\mathbb{R}$ (so that in particular X is real-valued) such that $X_0 = 0$. Let*

$$V_t = \exp\Big(X_t - \frac{1}{2}[\mathbf{X}]_t\Big), \qquad t \in [0, T]. \tag{6.3}$$

Then V is the unique solution to the linear rough differential equation

$$V_t = 1 + \int_0^t V_u \, \mathrm{d}\mathbf{X}_u. \tag{6.4}$$

By a solution to (6.4) we mean a path $V \in \mathcal{C}^\beta([0, T]; \mathbb{R})$ such that $(V, V) \in \mathscr{D}_X^{2\beta}$, and such that the equation holds with the integral defined as the rough integral of (V, V) against $\mathbf{X}$.

Proof. Applying the Itô formula of Proposition 6.11 with $Y = X - \frac{1}{2}[\mathbf{X}]$, $Y' = 1$ and $f = \exp$, we obtain

$$\begin{aligned} V_t &= 1 + \int_0^t V_u \, \mathrm{d}\mathbf{X}_u - \frac{1}{2} \int_0^t V_u \, \mathrm{d}[\mathbf{X}]_u + \frac{1}{2} \int_0^t V_u \, \mathrm{d}[\mathbf{X}]_u \\ &= 1 + \int_0^t V_u \, \mathrm{d}\mathbf{X}_u \end{aligned}$$

so that V does indeed satisfy (6.4).

We now turn to proving uniqueness. Suppose that $(\tilde{V}, \tilde{V}) \in \mathscr{D}_X^{2\beta}$ were another solution of (6.4). Let $\alpha \in (\frac{1}{3}, \beta)$. Since $\alpha < \beta$, it is clear that $\mathbf{X} \in \mathscr{C}^\alpha$, $(V, V) \in \mathscr{D}_X^{2\alpha}$ and $(\tilde{V}, \tilde{V}) \in \mathscr{D}_X^{2\alpha}$.

In the following we will let $\lesssim$ denote inequality up to a multiplicative constant which may depend on α, T and $\|X\|_\alpha$.

By Corollary 5.16, we have

$$\|V - \tilde{V}\|_\alpha = \left\| \int_0^\cdot V_u \,\mathrm{d}\mathbf{X}_u - \int_0^\cdot \tilde{V}_u \,\mathrm{d}\mathbf{X}_u \right\|_\alpha \lesssim \big(\|V - \tilde{V}\|_\alpha + \|R^V - R^{\tilde{V}}\|_{2\alpha}\big) \|\!|\mathbf{X}|\!\|_\alpha.$$

By the estimate in (5.11), we have

$$\|R^V - R^{\tilde{V}}\|_{2\alpha} = \big\|R^{\int_0^\cdot V_u \,\mathrm{d}\mathbf{X}_u} - R^{\int_0^\cdot \tilde{V}_u \,\mathrm{d}\mathbf{X}_u}\big\|_{2\alpha} \lesssim \big(\|V - \tilde{V}\|_\alpha + \|R^V - R^{\tilde{V}}\|_{2\alpha}\big) \|\!|\mathbf{X}|\!\|_\alpha.$$

Combining these inequalities, we have that

$$\|V - \tilde{V}\|_\alpha + \|R^V - R^{\tilde{V}}\|_{2\alpha} \leq C\big(\|V - \tilde{V}\|_\alpha + \|R^V - R^{\tilde{V}}\|_{2\alpha}\big) \|\!|\mathbf{X}|\!\|_\alpha \tag{6.5}$$

for some constant C. Note that

$$\|\!|\mathbf{X}|\!\|_{\alpha,[0,t]} = \|X\|_{\alpha,[0,t]} + \|\mathbb{X}\|_{2\alpha,[0,t]} \leq \|X\|_{\beta,[0,t]} t^{\beta-\alpha} + \|\mathbb{X}\|_{2\beta,[0,t]} t^{2(\beta-\alpha)}.$$

Therefore, by taking the terminal time $t = t_0 > 0$ sufficiently small, we can ensure that $C\|\!|\mathbf{X}|\!\|_{\alpha,[0,t_0]} < 1$. It then follows from (6.5) that $V = \tilde{V}$ on the interval $[0, t_0]$.

Since the constant C in (6.5) does not depend on the initial condition $V_0 = \tilde{V}_0$, we can simply infer from the same argument that uniqueness also holds over the next interval $[t_0, 2t_0]$, and so on, to deduce uniqueness over the entire original interval $[0, T]$. □

We call the path V defined in (6.3) the *rough exponential* of $\mathbf{X}$.

Corollary 6.13. *Suppose that $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ and $(K, K') \in \mathscr{D}_X^{2\alpha}$ are such that the rough integral $\int_0^\cdot K_u \,\mathrm{d}\mathbf{X}_u$ takes values in $\mathbb{R}$. Let V be the path given by*

$$V_t = \exp\bigg(\int_0^t K_u \,\mathrm{d}\mathbf{X}_u - \frac{1}{2}\int_0^t (K_u \otimes K_u) \,\mathrm{d}[\mathbf{X}]_u\bigg), \qquad t \in [0, T].$$

Then V is the unique solution of the rough differential equation

$$V_t = 1 + \int_0^t V_u K_u \,\mathrm{d}\mathbf{X}_u, \qquad t \in [0, T].$$

The proof Corollary 6.13 is left as an exercise.

7 Differential equations

In the previous section we saw our first example of a rough differential equation. In this section we shall study such equations in a more general setting. It is useful to first consider the case where the driving path is regular enough to allow for Young integration, before moving on to study equations driven by rough paths.

Throughout this section we will use the symbol $\lesssim$ to mean an inequality up to a multiplicative constant which may depend on α, T and $\|f\|_{C_b^k}$ (for $k = 2$ or 3).

7.1 Young differential equations

Lemma 7.1. *Let $\alpha \in (0,1]$ and $f \in C_b^2$. Then there exists a constant C, depending only on α, T and $\|f\|_{C_b^2}$, such that*

$$\|f(Y) - f(\tilde{Y})\|_\alpha \le C\big(1 + \|Y\|_\alpha + \|\tilde{Y}\|_\alpha\big)\big(|Y_0 - \tilde{Y}_0| + \|Y - \tilde{Y}\|_\alpha\big)$$

holds for all $Y, \tilde{Y} \in \mathcal{C}^\alpha$.

Proof.

$$\begin{aligned}
&\big|f(Y)_{s,t} - f(\tilde{Y})_{s,t}\big| \\
&= \left| \int_0^1 Df(\tilde{Y}_t + r(Y_t - \tilde{Y}_t))(Y_t - \tilde{Y}_t)\,\mathrm{d}r - \int_0^1 Df(\tilde{Y}_s + r(Y_s - \tilde{Y}_s))(Y_s - \tilde{Y}_s)\,\mathrm{d}r \right| \\
&\lesssim |(Y - \tilde{Y})_{s,t}| + (|Y_{s,t}| + |\tilde{Y}_{s,t}|)\|Y - \tilde{Y}\|_\infty,
\end{aligned}$$

so that

$$\begin{aligned}
\|f(Y) - f(\tilde{Y})\|_\alpha &\lesssim \|Y - \tilde{Y}\|_\alpha + (\|Y\|_\alpha + \|\tilde{Y}\|_\alpha)\|Y - \tilde{Y}\|_\infty \\
&\lesssim \big(1 + \|Y\|_\alpha + \|\tilde{Y}\|_\alpha\big)\big(|Y_0 - \tilde{Y}_0| + \|Y - \tilde{Y}\|_\alpha\big).
\end{aligned}$$

□

Theorem 7.2. *Let $\beta \in (\frac{1}{2}, 1]$ and let $X \in \mathcal{C}^\beta([0,T];\mathbb{R}^d)$. Let $f \in C_b^2(\mathbb{R}^m;\mathcal{L}(\mathbb{R}^d;\mathbb{R}^m))$, and let $y \in \mathbb{R}^m$. There exists a unique path $Y \in \mathcal{C}^\beta([0,T];\mathbb{R}^m)$ which satisfies*

$$Y_t = y + \int_0^t f(Y_s)\,\mathrm{d}X_s \tag{7.1}$$

for all $t \in [0,T]$.

Proof. Let $\alpha \in (\frac{1}{2}, \beta)$. Define the map $\mathcal{M}_t \colon \mathcal{C}^\alpha([0,t];\mathbb{R}^m) \to \mathcal{C}^\alpha([0,t];\mathbb{R}^m)$ by

$$\mathcal{M}_t(Y) = y + \int_0^\cdot f(Y_s)\,\mathrm{d}X_s.$$

Let

$$\mathcal{B}_t = \{Y \in \mathcal{C}^\alpha([0,t];\mathbb{R}^m) : Y_0 = y,\ \|Y\|_{\alpha,[0,t]} \le 1\}$$

which, being a closed subset of the Banach space $\mathcal{C}^\alpha$, is a complete metric space.

Invariance: Let $Y \in \mathcal{B}_t$. Using the estimate in (5.4), we have

$$\begin{aligned}
\|\mathcal{M}_t(Y)\|_\alpha = \left\| \int_0^\cdot f(Y_s)\,\mathrm{d}X_s \right\|_\alpha &\lesssim \big(\|f(Y)\|_\infty \|X\|_\alpha + \|f(Y)\|_\alpha \|X\|_\alpha\big) \\
&\lesssim (1 + \|Y\|_\alpha)\|X\|_\alpha \le 2\|X\|_\alpha.
\end{aligned}$$

Thus,

$$\|\mathcal{M}_t(Y)\|_{\alpha,[0,t]} \le C_1 \|X\|_{\alpha,[0,t]} \le C_1 \|X\|_{\beta,[0,t]} t^{\beta-\alpha}$$

for some constant C_1. Choosing $t = t_1 > 0$ sufficiently small so that $C_1\|X\|_{\beta,[0,T]} t_1^{\beta-\alpha} \le 1$, we then have that $\|\mathcal{M}_{t_1}(Y)\|_{\alpha,[0,t_1]} \le 1$. Since $\mathcal{M}_{t_1}(Y)_0 = y$, it follows that the set $\mathcal{B}_{t_1}$ is invariant under the map $\mathcal{M}_{t_1}$. That is, $\mathcal{M}_{t_1} \colon \mathcal{B}_{t_1} \to \mathcal{B}_{t_1}$.

Contraction: Let $Y, \tilde{Y} \in \mathcal{B}_t$. Using Proposition 5.4 and Lemma 7.1, we have

$$\begin{aligned}\|\mathcal{M}_t(Y) - \mathcal{M}_t(\tilde{Y})\|_\alpha &= \left\| \int_0^\cdot f(Y_s)\,\mathrm{d}X_s - \int_0^\cdot f(\tilde{Y}_s)\,\mathrm{d}X_s \right\|_\alpha \\ &\lesssim \|f(Y) - f(\tilde{Y})\|_\alpha \|X\|_\alpha \lesssim \|Y - \tilde{Y}\|_\alpha \|X\|_\alpha.\end{aligned}$$

This gives

$$\|\mathcal{M}_t(Y) - \mathcal{M}_t(\tilde{Y})\|_{\alpha,[0,t]} \leq C_2 \|Y - \tilde{Y}\|_{\alpha,[0,t]} \|X\|_{\alpha,[0,t]} \leq C_2 \|Y - \tilde{Y}\|_{\alpha,[0,t]} \|X\|_{\beta,[0,t]} t^{\beta-\alpha},$$

for some constant C_2. Taking $t = t_2 \in (0, t_1]$ sufficiently small so that $C_2 \|X\|_{\beta,[0,T]} t_2^{\beta-\alpha} \leq \frac{1}{2}$, we obtain

$$\|\mathcal{M}_{t_2}(Y) - \mathcal{M}_{t_2}(\tilde{Y})\|_{\alpha,[0,t_2]} \leq \frac{1}{2} \|Y - \tilde{Y}\|_{\alpha,[0,t_2]}.$$

Thus, the map $\mathcal{M}_{t_2}$ is a contraction on $\mathcal{B}_{t_2}$. By the Banach fixed point theorem, there exists a unique fixed point. That is, there exists a unique $Y \in \mathcal{C}^\alpha$ which satisfies (7.1) over the time interval $[0, t_2]$.

Since the constants C_1, C_2 above did not depend on the initial condition, we can then simply apply this argument over the next interval $[t_2, 2t_2]$, and so on. By pasting these solutions together, we deduce the existence of a unique solution Y over the entire interval $[0, T]$.

So far we only have that $Y \in \mathcal{C}^\alpha$. However, since $X \in \mathcal{C}^\beta$, it follows from (7.1) and (5.4) that $Y \in \mathcal{C}^\beta$. Since any solution in $\mathcal{C}^\beta$ is automatically also in $\mathcal{C}^\alpha$, our solution Y is also unique in $\mathcal{C}^\beta$. □

Proposition 7.3. *Let $\beta \in (\frac{1}{2}, 1]$ and $f \in C_b^2$. Let $X, \tilde{X} \in \mathcal{C}^\beta$ and $y, \tilde{y} \in \mathbb{R}^m$, and let Y and $\tilde{Y}$ be the (unique) solutions of* (7.1) *with the data (y, X) and $(\tilde{y}, \tilde{X})$ respectively. Let $M > 0$ be a constant such that $\|X\|_\beta, \|\tilde{X}\|_\beta \leq M$. Then, for any $\alpha \in (\frac{1}{2}, \beta)$, there exists a constant $C_M > 0$, depending on $\alpha, T, \|f\|_{C_b^2}$ and M, such that*

$$\|Y - \tilde{Y}\|_\alpha \leq C_M \big(|y - \tilde{y}| + \|X - \tilde{X}\|_\alpha\big).$$

Proof. Recall from the proof of Theorem 7.2 that the local solution Y over the time interval $[0, t_2]$ is an element of the set $\mathcal{B}_{t_2}$, which means in particular that $\|Y\|_{\alpha,[0,t_2]} \leq 1$. Similarly, we have that $\|\tilde{Y}\|_{\alpha,[0,\tilde{t}_2]} \leq 1$ for some $\tilde{t}_2 > 0$.

By Proposition 5.4, for any $t \in (0, t_2 \wedge \tilde{t}_2]$, we then have

$$\begin{aligned}\|Y - \tilde{Y}\|_\alpha &= \left\| \int_0^\cdot f(Y_s)\,\mathrm{d}X_s - \int_0^\cdot f(\tilde{Y}_s)\,\mathrm{d}\tilde{X}_s \right\|_\alpha \\ &\lesssim \big(|f(Y_0) - f(\tilde{Y}_0)| + \|f(Y) - f(\tilde{Y})\|_\alpha\big)\|X\|_\alpha + \big(|f(\tilde{y}) + \|f(\tilde{Y})\|_\alpha\big)\|X - \tilde{X}\|_\alpha \\ &\lesssim \big(|Y_0 - \tilde{Y}_0| + \|Y - \tilde{Y}\|_\alpha\big)\|X\|_\alpha + \|X - \tilde{X}\|_\alpha.\end{aligned}$$

This means that

$$\|Y - \tilde{Y}\|_{\alpha,[0,t]} \leq C_3 \Big(\big(|Y_0 - \tilde{Y}_0| + \|Y - \tilde{Y}\|_{\alpha,[0,t]}\big)\|X\|_{\beta,[0,t]} t^{\beta-\alpha} + \|X - \tilde{X}\|_{\alpha,[0,t]}\Big)$$

for some constant C_3.

Choosing $t = t_3 \in (0, t_2 \wedge \tilde{t}_2]$ sufficiently small so that $C_3\|X\|_{\beta,[0,T]}t_3^{\beta-\alpha} \leq \frac{1}{2}$ and rearranging, we deduce that

$$\|Y - \tilde{Y}\|_{\alpha,[0,t_3]} \lesssim |Y_0 - \tilde{Y}_0| + \|X - \tilde{X}\|_{\alpha,[0,t_3]}.$$

It follows that there exists a $\delta > 0$, depending on $\alpha, T, \|f\|_{C_b^2}$ and M, such that, for any interval $[s,t] \subset [0,T]$ with $|t-s| \leq \delta$, we have

$$\|Y - \tilde{Y}\|_{\alpha,[s,t]} \lesssim |Y_s - \tilde{Y}_s| + \|X - \tilde{X}\|_{\alpha,[s,t]}. \tag{7.2}$$

Take a partition π of the interval $[0,T]$ with mesh size $|\pi| \leq \delta$. The estimate in (7.2) then holds over every interval $[s,t] \in \pi$, and by combining these estimates one can deduce that the same estimate holds over the entire interval $[0,T]$. $\square$

7.2 Functions of controlled paths

Lemma 7.4. *Let $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ and $X \in \mathcal{C}^\alpha$. Let $f \in C_b^2$. For any $(Y, Y') \in \mathscr{D}_X^{2\alpha}$, the pair*

$$\big(f(Y), Df(Y)Y'\big) \in \mathscr{D}_X^{2\alpha}$$

is a controlled path. Moreover, we have the estimates

$$\begin{aligned} \|Df(Y)Y'\|_\alpha &\leq C\big(1 + |Y_0'| + \|Y'\|_\alpha + \|R^Y\|_{2\alpha}\big)^2\big(1 + \|X\|_\alpha\big), \\ \|R^{f(Y)}\|_{2\alpha} &\leq C\big(1 + |Y_0'| + \|Y'\|_\alpha + \|R^Y\|_{2\alpha}\big)^2\big(1 + \|X\|_\alpha\big)^2, \end{aligned}$$

where the constant C depends on α, T and $\|f\|_{C_b^2}$.

Proof. We have

$$\begin{aligned} \big|Df(Y_t)Y_t' - Df(Y_s)Y_s'\big| &\leq |Df(Y_t)||Y_{s,t}'| + |Df(Y_t) - Df(Y_s)||Y_s'| \\ &\lesssim |Y_{s,t}'| + |Y_{s,t}||Y_s'| \end{aligned}$$

so that

$$\begin{aligned} \|Df(Y)Y'\|_\alpha &\lesssim \|Y'\|_\alpha + \|Y\|_\alpha\|Y'\|_\infty \\ &\lesssim \|Y'\|_\alpha + \big(\|Y'\|_\infty\|X\|_\alpha + \|R^Y\|_{2\alpha}\big)\|Y'\|_\infty \\ &\lesssim \|Y'\|_\alpha + \big((|Y_0'| + \|Y'\|_\alpha)\|X\|_\alpha + \|R^Y\|_{2\alpha}\big)\big(|Y_0'| + \|Y'\|_\alpha\big) \\ &\lesssim \big(1 + |Y_0'| + \|Y'\|_\alpha + \|R^Y\|_{2\alpha}\big)^2\big(1 + \|X\|_\alpha\big), \end{aligned}$$

which gives the first estimate.

We also have

$$\begin{aligned} R_{s,t}^{f(Y)} &:= f(Y_t) - f(Y_s) - Df(Y_s)Y_s'X_{s,t} \\ &= f(Y_t) - f(Y_s) - Df(Y_s)(Y_{s,t} - R_{s,t}^Y) \\ &= \int_0^1 \int_0^1 D^2 f(Y_s + r_1 r_2 Y_{s,t})Y_{s,t}^{\otimes 2}\, r_1 \,\mathrm{d}r_2 \,\mathrm{d}r_1 + Df(Y_s)R_{s,t}^Y, \end{aligned}$$

so that

$$\big|R_{s,t}^{f(Y)}\big| \lesssim |Y_{s,t}|^2 + |R_{s,t}^Y|.$$

Then

$$
\begin{aligned}
\|R^{f(Y)}\|_{2\alpha} &\lesssim \|Y\|_{\alpha}^2 + \|R^Y\|_{2\alpha} \\
&\lesssim \big(\|Y'\|_\infty \|X\|_\alpha + \|R^Y\|_{2\alpha}\big)^2 + \|R^Y\|_{2\alpha} \\
&\lesssim \big(1 + |Y_0'| + \|Y'\|_\alpha + \|R^Y\|_{2\alpha}\big)^2 \big(1 + \|X\|_\alpha\big)^2,
\end{aligned}
$$

which gives the second estimate. □

Lemma 7.5. *Let $\alpha \in (\frac{1}{3}, \frac{1}{2}]$, $X, \tilde{X} \in \mathcal{C}^\alpha$, $(Y, Y') \in \mathscr{D}_X^{2\alpha}$, $(\tilde{Y}, \tilde{Y}') \in \mathscr{D}_{\tilde{X}}^{2\alpha}$ and $f \in C_b^3$. Let $M > 0$ be a constant such that $\|X\|_\alpha \leq M$, $\|\tilde{X}\|_\alpha \leq M$, $|Y_0'| + \|Y'\|_\alpha + \|R^Y\|_{2\alpha} \leq M$ and $|\tilde{Y}_0'| + \|\tilde{Y}'\|_\alpha + \|R^{\tilde{Y}}\|_{2\alpha} \leq M$. We have that*

$$
\begin{aligned}
&\big\|Df(Y)Y' - Df(\tilde{Y})\tilde{Y}'\big\|_\alpha \\
&\quad \leq C\big(|Y_0 - \tilde{Y}_0| + |Y_0' - \tilde{Y}_0'| + \|Y' - \tilde{Y}'\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha} + \|X - \tilde{X}\|_\alpha\big),
\end{aligned}
$$

and

$$
\begin{aligned}
&\big\|R^{f(Y)} - R^{f(\tilde{Y})}\big\|_{2\alpha} \\
&\quad \leq C\big(|Y_0 - \tilde{Y}_0| + |Y_0' - \tilde{Y}_0'| + \|Y' - \tilde{Y}'\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha} + \|X - \tilde{X}\|_\alpha\big),
\end{aligned}
$$

where the constant C depends on $\alpha, T, \|f\|_{C_b^3}$ and M.

Proof. In the following we shall allow the multiplicative constant indicated by the symbol $\lesssim$ to also depend on M.

We have

$$
\begin{aligned}
&\big|\big(Df(Y)Y' - Df(\tilde{Y})\tilde{Y}'\big)_{s,t}\big| \\
&\leq \big|\big(Df(Y)(Y' - \tilde{Y}')\big)_{s,t}\big| + \big|\big((Df(Y) - Df(\tilde{Y}))\tilde{Y}'\big)_{s,t}\big| \\
&\leq |Df(Y_t)(Y' - \tilde{Y}')_{s,t}| + |Df(Y)_{s,t}(Y_s' - \tilde{Y}_s')| \\
&\qquad + |(Df(Y_t) - Df(\tilde{Y}_t))\tilde{Y}_{s,t}'| + |(Df(Y) - Df(\tilde{Y}))_{s,t}\tilde{Y}_s'|
\end{aligned}
$$

and hence

$$
\begin{aligned}
&\big\|Df(Y)Y' - Df(\tilde{Y})\tilde{Y}'\big\|_\alpha \\
&\leq \|Df(Y)\|_\infty \|Y' - \tilde{Y}'\|_\alpha + \|Df(Y)\|_\alpha \|Y' - \tilde{Y}'\|_\infty \\
&\qquad + \|Df(Y) - Df(\tilde{Y})\|_\infty \|\tilde{Y}'\|_\alpha + \|Df(Y) - Df(\tilde{Y})\|_\alpha \|\tilde{Y}'\|_\infty \\
&\lesssim \|Y' - \tilde{Y}'\|_\alpha + \|Y' - \tilde{Y}'\|_\infty + \|Y - \tilde{Y}\|_\infty + \|Df(Y) - Df(\tilde{Y})\|_\alpha.
\end{aligned}
$$

By Lemma 7.1, we have

$$
\|Df(Y) - Df(\tilde{Y})\|_\alpha \lesssim |Y_0 - \tilde{Y}_0| + \|Y - \tilde{Y}\|_\alpha,
$$

and by (5.10) we have

$$
\|Y - \tilde{Y}\|_\alpha \lesssim |Y_0' - \tilde{Y}_0'| + \|Y' - \tilde{Y}'\|_\alpha + \|X - \tilde{X}\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha}. \tag{7.3}
$$

Putting this together, we obtain the first estimate.

We also have

$$
\begin{aligned}
&\big|R^{f(Y)}_{s,t} - R^{f(\tilde{Y})}_{s,t}\big| \\
&= \big|f(Y)_{s,t} - Df(Y_s)Y'_sX_{s,t} - f(\tilde{Y})_{s,t} + Df(\tilde{Y}_s)\tilde{Y}'_s\tilde{X}_{s,t}\big| \\
&= \big|f(Y)_{s,t} - Df(Y_s)(Y_{s,t} - R^Y_{s,t}) - f(\tilde{Y})_{s,t} + Df(\tilde{Y}_s)(\tilde{Y}_{s,t} - R^{\tilde{Y}}_{s,t})\big| \\
&\le \big|f(Y)_{s,t} - Df(Y_s)Y_{s,t} - f(\tilde{Y})_{s,t} + Df(\tilde{Y}_s)\tilde{Y}_{s,t}\big| + \big|Df(Y_s)R^Y_{s,t} - Df(\tilde{Y}_s)R^{\tilde{Y}}_{s,t}\big| \\
&= \bigg|\int_0^1\int_0^1 D^2f(Y_s + r_1r_2Y_{s,t})Y^{\otimes 2}_{s,t}\, r_1\,\mathrm{d}r_2\,\mathrm{d}r_1 - \int_0^1\int_0^1 D^2f(\tilde{Y}_s + r_1r_2\tilde{Y}_{s,t})\tilde{Y}^{\otimes 2}_{s,t}\, r_1\,\mathrm{d}r_2\,\mathrm{d}r_1\bigg| \\
&\qquad + \big|Df(Y_s)R^Y_{s,t} - Df(\tilde{Y}_s)R^{\tilde{Y}}_{s,t}\big| \\
&\lesssim \|Y - \tilde{Y}\|_\infty |Y_{s,t}|^2 + |Y^{\otimes 2}_{s,t} - \tilde{Y}^{\otimes 2}_{s,t}| + \big|Df(Y_s)R^Y_{s,t} - Df(\tilde{Y}_s)R^{\tilde{Y}}_{s,t}\big| \\
&\le \|Y - \tilde{Y}\|_\infty |Y_{s,t}|^2 + |Y_{s,t}||Y_{s,t} - \tilde{Y}_{s,t}| + |Y_{s,t} - \tilde{Y}_{s,t}||\tilde{Y}_{s,t}| \\
&\qquad + |Df(Y_s) - Df(\tilde{Y}_s)||R^Y_{s,t}| + |Df(\tilde{Y}_s)||R^Y_{s,t} - R^{\tilde{Y}}_{s,t}|,
\end{aligned}
$$

so that

$$
\begin{aligned}
\big\|R^{f(Y)} - R^{f(\tilde{Y})}\big\|_{2\alpha} &\lesssim \|Y - \tilde{Y}\|_\infty \|Y\|_\alpha^2 + (\|Y\|_\alpha + \|\tilde{Y}\|_\alpha)\|Y - \tilde{Y}\|_\alpha \\
&\qquad + \|Y - \tilde{Y}\|_\infty \|R^Y\|_{2\alpha} + \|R^Y - R^{\tilde{Y}}\|_{2\alpha} \\
&\lesssim |Y_0 - \tilde{Y}_0| + \|Y - \tilde{Y}\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha}.
\end{aligned}
$$

Using (7.3) again, we obtain the second estimate. □

Lemma 7.6. *Let $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ and $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$. Let $f \in C^2_b$. For any $(Y, Y') \in \mathscr{D}^{2\alpha}_X$, the pair*

$$
\left(\int_0^\cdot f(Y_u)\,\mathrm{d}\mathbf{X}_u, f(Y)\right) \in \mathscr{D}^{2\alpha}_X
$$

is a controlled path. Moreover, we have the estimates

$$
\begin{aligned}
\|f(Y)\|_\alpha &\le C\big((|Y'_0| + \|Y'\|_\alpha)\|X\|_\alpha + \|R^Y\|_{2\alpha}T^\alpha\big), \\
\big\|R^{\int_0^\cdot f(Y_u)\,\mathrm{d}\mathbf{X}_u}\big\|_{2\alpha} &\le C\big(1 + |Y'_0| + \|Y'\|_\alpha + \|R^Y\|_{2\alpha}\big)^2\big(1 + \|X\|_\alpha\big)^2 |||\mathbf{X}|||_\alpha,
\end{aligned}
$$

where the constant C depends on α, T and $\|f\|_{C^2_b}$.

Proof. The first estimate follows easily from the Lipschitz continuity of f and the relation $Y_{s,t} = Y'_sX_{s,t} + R^Y_{s,t}$.

By Lemma 7.4, we know that the pair $(f(Y), Df(Y)Y')$ is a controlled path, and hence by Proposition 5.11 that the rough integral $\int_0^\cdot f(Y_u)\,\mathrm{d}\mathbf{X}_u$ exists. Moreover, it follows from the estimate in (5.9) that

$$
\begin{aligned}
\big|R^{\int_0^\cdot f(Y_u)\,\mathrm{d}\mathbf{X}_u}_{s,t}\big| &= \left|\int_s^t f(Y_u)\,\mathrm{d}\mathbf{X}_u - f(Y_s)X_{s,t}\right| \\
&\lesssim |Df(Y_s)Y'_s\mathbb{X}_{s,t}| + \big(\|R^{f(Y)}\|_{2\alpha}\|X\|_\alpha + \|Df(Y)Y'\|_\alpha\|\mathbb{X}\|_{2\alpha}\big)|t-s|^{3\alpha},
\end{aligned}
$$

and hence, using the estimates in Lemma 7.4,

$$
\begin{aligned}
\big\|R^{\int_0^\cdot f(Y_u)\,\mathrm{d}\mathbf{X}_u}\big\|_{2\alpha} &\lesssim \|Y'\|_\infty\|\mathbb{X}\|_{2\alpha} + \|R^{f(Y)}\|_{2\alpha}\|X\|_\alpha + \|Df(Y)Y'\|_\alpha\|\mathbb{X}\|_{2\alpha} \\
&\lesssim \big(1 + |Y'_0| + \|Y'\|_\alpha + \|R^Y\|_{2\alpha}\big)^2\big(1 + \|X\|_\alpha\big)^2\big(\|X\|_\alpha + \|\mathbb{X}\|_{2\alpha}\big),
\end{aligned}
$$

which gives the second estimate. □

Lemma 7.7. *Let $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ and $\mathbf{X} = (X, \mathbb{X}), \tilde{\mathbf{X}} = (\tilde{X}, \tilde{\mathbb{X}}) \in \mathscr{C}^\alpha$. Let $(Y, Y') \in \mathscr{D}_X^{2\alpha}$ and $(\tilde{Y}, \tilde{Y}') \in \mathscr{D}_{\tilde{X}}^{2\alpha}$, and $f \in C_b^3$. Let $M > 0$ be a constant such that $\|X\|_\alpha \leq M$, $\|\tilde{X}\|_\alpha \leq M$, $|Y_0'| + \|Y'\|_\alpha + \|R^Y\|_{2\alpha} \leq M$ and $|\tilde{Y}_0'| + \|\tilde{Y}'\|_\alpha + \|R^{\tilde{Y}}\|_{2\alpha} \leq M$. Then*

$$\begin{aligned}
&\|f(Y) - f(\tilde{Y})\|_\alpha \\
&\leq C\Big(|Y_0 - \tilde{Y}_0| + \big(|Y_0' - \tilde{Y}_0'| + \|Y' - \tilde{Y}'\|_\alpha\big)\|X\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha} T^\alpha + \|X - \tilde{X}\|_\alpha\Big),
\end{aligned}$$

and

$$\begin{aligned}
&\big\|R^{\int_0^\cdot f(Y_u)\,\mathrm{d}\mathbf{X}_u} - R^{\int_0^\cdot f(\tilde{Y}_u)\,\mathrm{d}\tilde{\mathbf{X}}_u}\big\|_{2\alpha} \\
&\leq C\Big(\big(|Y_0 - \tilde{Y}_0| + |Y_0' - \tilde{Y}_0'| + \|Y' - \tilde{Y}'\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha} + \|X - \tilde{X}\|_\alpha\big)\|\mathbf{X}\|_\alpha + \|\mathbf{X}; \tilde{\mathbf{X}}\|_\alpha\Big)
\end{aligned}$$

where the constant C depends on $\alpha, T, \|f\|_{C_b^3}$ and M.

Proof. In the following we shall allow the multiplicative constant indicated by the symbol $\lesssim$ to also depend on M.

By Lemma 7.1, we have that

$$\begin{aligned}
\|f(Y) - f(\tilde{Y})\|_\alpha &\lesssim \big(1 + \|Y\|_\alpha + \|\tilde{Y}\|_\alpha\big)\big(|Y_0 - \tilde{Y}_0| + \|Y - \tilde{Y}\|_\alpha\big) \\
&\lesssim |Y_0 - \tilde{Y}_0| + \|Y - \tilde{Y}\|_\alpha.
\end{aligned}$$

By (5.10), we have

$$\|Y - \tilde{Y}\|_\alpha \lesssim \big(|Y_0' - \tilde{Y}_0'| + \|Y' - \tilde{Y}'\|_\alpha\big)\|X\|_\alpha + \|X - \tilde{X}\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha} T^\alpha.$$

Combining these two inequalities, we obtain the first estimate.

By (5.11), we have

$$\begin{aligned}
&\big\|R^{\int_0^\cdot f(Y_u)\,\mathrm{d}\mathbf{X}_u} - R^{\int_0^\cdot f(\tilde{Y}_u)\,\mathrm{d}\tilde{\mathbf{X}}_u}\big\|_{2\alpha} \\
&\lesssim \big(|Df(Y_0)Y_0' - Df(\tilde{Y}_0)\tilde{Y}_0'| + \|Df(Y)Y' - Df(\tilde{Y})\tilde{Y}'\|_\alpha + \|R^{f(Y)} - R^{f(\tilde{Y})}\|_{2\alpha}\big)\|\mathbf{X}\|_\alpha \\
&\qquad + \big(|Df(\tilde{Y}_0)\tilde{Y}_0'| + \|Df(\tilde{Y})\tilde{Y}'\|_\alpha + \|R^{f(\tilde{Y})}\|_{2\alpha}\big)\|\mathbf{X}; \tilde{\mathbf{X}}\|_\alpha.
\end{aligned}$$

We know from Lemma 7.4 that the norms $\|Df(\tilde{Y})\tilde{Y}'\|_\alpha$ and $\|R^{f(\tilde{Y})}\|_{2\alpha}$ are both bounded by

$$\big(1 + |\tilde{Y}_0'| + \|\tilde{Y}'\|_\alpha + \|R^{\tilde{Y}}\|_{2\alpha}\big)^2\big(1 + \|\tilde{X}\|_\alpha\big)^2 \lesssim 1,$$

and we know from Lemma 7.5 that the norms $\big\|Df(Y)Y' - Df(\tilde{Y})\tilde{Y}'\big\|_\alpha$ and $\big\|R^{f(Y)} - R^{f(\tilde{Y})}\big\|_{2\alpha}$ can both be estimated by

$$|Y_0 - \tilde{Y}_0| + |Y_0' - \tilde{Y}_0'| + \|Y' - \tilde{Y}'\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha} + \|X - \tilde{X}\|_\alpha.$$

Substituting these estimates into the above, we obtain the desired inequality. □

7.3 Rough differential equations

We are now ready to establish existence and uniqueness of solutions to RDEs of the form

$$\mathrm{d}Y_t = f(Y_t)\,\mathrm{d}\mathbf{X}_t$$

for sufficiently regular vector fields f.

Theorem 7.8. *Let $\beta \in (\frac{1}{3}, \frac{1}{2}]$ and let $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\beta([0,T];\mathbb{R}^d)$ be a rough path. Let $f \in C_b^3(\mathbb{R}^m; \mathcal{L}(\mathbb{R}^d;\mathbb{R}^m))$, and let $y \in \mathbb{R}^m$. There exists a unique controlled path $(Y, Y') \in \mathscr{D}_X^{2\beta}$ such that $Y' = f(Y)$, and such that*

$$Y_t = y + \int_0^t f(Y_s)\,\mathrm{d}\mathbf{X}_s \tag{7.4}$$

for all $t \in [0,T]$.

Proof. Let $\alpha \in (\frac{1}{3}, \beta)$. Define the map $\mathcal{M}_t \colon \mathscr{D}_X^{2\alpha}([0,t];\mathbb{R}^m) \to \mathscr{D}_X^{2\alpha}([0,t];\mathbb{R}^m)$ by

$$\mathcal{M}_t(Y,Y') = \left(y + \int_0^\cdot f(Y_s)\,\mathrm{d}\mathbf{X}_s, f(Y)\right),$$

which we know defines a controlled path by Lemma 7.6. Let

$$\mathcal{B}_t = \big\{(Y,Y') \in \mathscr{D}_X^{2\alpha}([0,t];\mathbb{R}^m) : Y_0 = y,\ Y_0' = f(y),\ \|Y,Y'\|_{X,\alpha,[0,t]} \leq 1\big\},$$

where

$$\|Y,Y'\|_{X,\alpha} = \|Y'\|_\alpha + \|R^Y\|_{2\alpha}.$$

Since $\mathcal{B}_t$ a closed subset of the Banach space $\mathscr{D}_X^{2\alpha}$, it is a complete metric space with the metric induced by the norm $\|\cdot,\cdot\|_{X,\alpha}$. Note that the path $s \mapsto (y + f(y)X_{0,s}, f(y))$ is an element of $\mathcal{B}_t$, so the set $\mathcal{B}_t$ is nonempty.

Let $M = 1 + \|f\|_{C_b^3} + \|X\|_{\alpha,[0,T]}$. Note that, if $(Y,Y') \in \mathcal{B}_t$, then $|Y_0'| + \|Y'\|_\alpha + \|R^Y\|_{2\alpha} = |f(y)| + \|Y,Y'\|_{X,\alpha} \leq \|f\|_{C_b^3} + 1 \leq M$, so that the hypotheses of Lemma 7.7 are satisfied. In the following we shall allow the multiplicative constant indicated by the symbol $\lesssim$ to also depend on M.

Invariance: Let $(Y,Y') \in \mathcal{B}_t$. By Lemma 7.6, we have

$$\begin{aligned}
\|\mathcal{M}_t(Y,Y')\|_{X,\alpha} &= \|f(Y)\|_\alpha + \big\|R^{\int_0^\cdot f(Y_s)\,\mathrm{d}\mathbf{X}_s}\big\|_{2\alpha} \\
&\lesssim (|Y_0'| + \|Y'\|_\alpha)\|X\|_\alpha + \|R^Y\|_{2\alpha} t^\alpha \\
&\quad + \big(1 + |Y_0'| + \|Y'\|_\alpha + \|R^Y\|_{2\alpha}\big)^2 \big(1 + \|X\|_\alpha\big)^2 \|\!|\mathbf{X}|\!\|_\alpha \\
&\lesssim \|\!|\mathbf{X}|\!\|_\alpha + t^\alpha,
\end{aligned}$$

so that

$$\|\mathcal{M}_t(Y,Y')\|_{X,\alpha,[0,t]} \leq C_1\big(\|\!|\mathbf{X}|\!\|_{\alpha,[0,t]} + t^\alpha\big)$$

for some constant C_1. Then

$$\begin{aligned}
\|\mathcal{M}_t(Y,Y')\|_{X,\alpha,[0,t]} &\leq C_1\big(\|X\|_{\alpha,[0,t]} + \|\mathbb{X}\|_{2\alpha,[0,t]} + t^\alpha\big) \\
&\leq C_1\big(\|X\|_{\beta,[0,t]} t^{\beta-\alpha} + \|\mathbb{X}\|_{2\beta,[0,t]} t^{2(\beta-\alpha)} + t^\alpha\big).
\end{aligned}$$

Choosing $t = t_1 > 0$ sufficiently small, we can ensure that $\|\mathcal{M}_{t_1}(Y,Y')\|_{X,\alpha,[0,t_1]} \leq 1$ for all $(Y,Y') \in \mathcal{B}_{t_1}$. Thus, the set $\mathcal{B}_{t_1}$ is invariant under the map $\mathcal{M}_{t_1}$.

Contraction: Let $(Y,Y'),(\tilde{Y},\tilde{Y}') \in \mathcal{B}_t$ for some $t \in (0,t_1]$. By Lemma 7.7, we have

$$\begin{aligned}\big\|\mathcal{M}_t(Y,Y') - \mathcal{M}_t(\tilde{Y},\tilde{Y}')\big\|_{X,\alpha} &= \|f(Y) - f(\tilde{Y})\|_\alpha + \big\|R^{\int_0^\cdot f(Y_s)\,\mathrm{d}\mathbf{X}_s} - R^{\int_0^\cdot f(\tilde{Y}_s)\,\mathrm{d}\tilde{\mathbf{X}}_s}\big\|_{2\alpha}\\ &\lesssim \big(\|Y' - \tilde{Y}'\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha}\big)\big(\|\mathbf{X}\|_\alpha + t^\alpha\big),\end{aligned}$$

so that

$$\begin{aligned}&\big\|\mathcal{M}_t(Y,Y') - \mathcal{M}_t(\tilde{Y},\tilde{Y}')\big\|_{X,\alpha,[0,t]}\\ &\quad\leq C_2\big(\|Y' - \tilde{Y}'\|_{\alpha,[0,t]} + \|R^Y - R^{\tilde{Y}}\|_{2\alpha,[0,t]}\big)\big(\|\mathbf{X}\|_{\alpha,[0,t]} + t^\alpha\big)\end{aligned}$$

for some constant C_2. Then

$$\begin{aligned}&\big\|\mathcal{M}_t(Y,Y') - \mathcal{M}_t(\tilde{Y},\tilde{Y}')\big\|_{X,\alpha,[0,t]}\\ &\quad\leq C_2\big(\|Y' - \tilde{Y}'\|_{\alpha,[0,t]} + \|R^Y - R^{\tilde{Y}}\|_{2\alpha,[0,t]}\big)\big(\|X\|_{\beta,[0,t]}t^{\beta-\alpha} + \|\mathbb{X}\|_{2\beta,[0,t]}t^{2(\beta-\alpha)} + t^\alpha\big).\end{aligned}$$

Choosing $t = t_2 \in (0,t_1]$ sufficiently small, we can then ensure that

$$\big\|\mathcal{M}_{t_2}(Y,Y') - \mathcal{M}_{t_2}(\tilde{Y},\tilde{Y}')\big\|_{X,\alpha,[0,t_2]} \leq \frac{1}{2}\big\|(Y,Y') - (\tilde{Y},\tilde{Y}')\big\|_{X,\alpha,[0,t_2]}.$$

The map $\mathcal{M}_{t_2}$ is therefore a contraction on $\mathcal{B}_{t_2}$. The unique fixed point of this map is then the unique element $(Y,Y') \in \mathscr{D}_X^{2\alpha}$ of the RDE over the interval $[0,t_2]$ satisfying $Y' = f(Y)$.

Since the constants C_1, C_2 above did not depend on the initial condition, we can then simply apply this argument over the next interval $[t_2, 2t_2]$, and so on. By pasting these solutions together, we deduce the existence of a unique solution $(Y,Y') \in \mathscr{D}_X^{2\alpha}$ over the entire interval $[0,T]$.

So far we only have that $(Y,Y') \in \mathscr{D}_X^{2\alpha}$. However, since $\mathbf{X} \in \mathscr{C}^\beta$, we actually have that $(Y,Y') \in \mathscr{D}_X^{2\beta}$. Indeed, since $Y_{s,t} = Y_s' X_{s,t} + R_{s,t}^Y$ and $X \in \mathcal{C}^\beta$, we see that $Y \in \mathcal{C}^\beta$, and since $Y' = f(Y)$ and f is Lipschitz, we then have that $Y' \in \mathcal{C}^\beta$. Moreover, by (5.9) we have that

$$\begin{aligned}|R_{s,t}^Y| = |Y_{s,t} - Y_s' X_{s,t}| &= \bigg|\int_s^t f(Y_u)\,\mathrm{d}\mathbf{X}_u - f(Y_s)X_{s,t}\bigg|\\ &\lesssim \|Df(Y)Y'\|_\infty |\mathbb{X}_{s,t}| + O(|t-s|^{3\alpha}),\end{aligned}$$

and since $\mathbb{X} \in \mathcal{C}_2^{2\beta}$, we see that $R^Y \in \mathcal{C}_2^{2\beta}$. Since $\mathscr{D}_X^{2\beta} \subset \mathscr{D}_X^{2\alpha}$, we have that the solution (Y,Y') is also the unique solution in $\mathscr{D}_X^{2\beta}$ satisfying $Y' = f(Y)$. □

Suppose that we have fixed a vector field f and an initial condition y. The previous theorem shows that, given a rough path $\mathbf{X}$, we can assign a unique (in a suitable sense) solution (Y,Y') to the corresponding rough differential equation. The solution map

$$\mathbf{X} \longmapsto (Y,Y')$$

in the context of rough paths is known as the Itô–Lyons map.

We now come to a version of what is arguably the most important result in the theory: the continuity of the Itô–Lyons map.

Theorem 7.9. *Let $\beta \in (\frac{1}{3}, \frac{1}{2}]$ and $f \in C_b^3$. Let $\mathbf{X} = (X, \mathbb{X}), \tilde{\mathbf{X}} = (\tilde{X}, \tilde{\mathbb{X}}) \in \mathscr{C}^\beta$ and $y, \tilde{y} \in \mathbb{R}^m$, and let $(Y, Y') \in \mathscr{D}_X^{2\beta}$ and $(\tilde{Y}, \tilde{Y}') \in \mathscr{D}_{\tilde{X}}^{2\beta}$ be the solutions of the RDE (7.4) given by Theorem 7.8 with the data $(y, \mathbf{X})$ and $(\tilde{y}, \tilde{\mathbf{X}})$ respectively. Let $M > 0$ be a constant such that $\|\mathbf{X}\|_\beta, \|\tilde{\mathbf{X}}\|_\beta \leq M$. Then, for any $\alpha \in (\frac{1}{3}, \beta)$, there exists a constant $C_M > 0$, depending on $\alpha, T, \|f\|_{C_b^3}$ and M, such that*

$$\|Y - \tilde{Y}\|_\alpha + \|Y' - \tilde{Y}'\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha} \leq C_M\big(|y - \tilde{y}| + \|\mathbf{X}; \tilde{\mathbf{X}}\|_\alpha\big).$$

Proof. Recall from the proof of Theorem 7.8 that the local solution (Y, Y') of the RDE over the time interval $[0, t_2]$ is an element of $\mathcal{B}_{t_2}$, which means in particular that $\|Y, Y'\|_{X,\alpha,[0,t_2]} \leq 1$. Similarly, we have that $\|\tilde{Y}, \tilde{Y}'\|_{X,\alpha,[0,t_2]} \leq 1$ for some $\tilde{t}_2 > 0$. In the following we shall allow the multiplicative constant indicated by the symbol $\lesssim$ to also depend on M.

By Lemma 7.7, for any $t \in (t_2 \wedge \tilde{t}_2]$, we have that

$$\begin{aligned}
&\|Y' - \tilde{Y}'\|_\alpha = \|f(Y) - f(\tilde{Y})\|_\alpha \\
&\lesssim |Y_0 - \tilde{Y}_0| + \big(|Y_0' - \tilde{Y}_0'| + \|Y' - \tilde{Y}'\|_\alpha\big)\|X\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha} t^\alpha + \|X - \tilde{X}\|_\alpha
\end{aligned}$$

and

$$\begin{aligned}
&\|R^Y - R^{\tilde{Y}}\|_{2\alpha} = \big\|R^{\int_0^\cdot f(Y_u)\,\mathrm{d}\mathbf{X}_u} - R^{\int_0^\cdot f(\tilde{Y}_u)\,\mathrm{d}\tilde{\mathbf{X}}_u}\big\|_{2\alpha} \\
&\lesssim \big(|Y_0 - \tilde{Y}_0| + |Y_0' - \tilde{Y}_0'| + \|Y' - \tilde{Y}'\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha} + \|X - \tilde{X}\|_\alpha\big)\|\mathbf{X}\|_\alpha + \|\mathbf{X}; \tilde{\mathbf{X}}\|_\alpha.
\end{aligned}$$

Noting that $|Y_0' - \tilde{Y}_0'| = |f(Y_0) - f(\tilde{Y}_0)| \lesssim |Y_0 - \tilde{Y}_0|$, we then have

$$\begin{aligned}
&\|Y' - \tilde{Y}'\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha} \\
&\leq C_3\Big(|Y_0 - \tilde{Y}_0| + \big(\|Y' - \tilde{Y}'\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha}\big)\big(\|\mathbf{X}\|_\alpha + t^\alpha\big) + \|\mathbf{X}; \tilde{\mathbf{X}}\|_\alpha\Big)
\end{aligned}$$

for some constant C_3. We have that

$$\|\mathbf{X}\|_{\alpha,[0,t]} + t^\alpha \leq \|X\|_{\beta,[0,t]} t^{\beta-\alpha} + \|\mathbb{X}\|_{2\beta,[0,t]} t^{2(\beta-\alpha)} + t^\alpha.$$

Choosing $t = t_3 \in (t_2 \wedge \tilde{t}_2]$ sufficiently small such that

$$C_3\big(\|X\|_{\beta,[0,t_3]} t_3^{\beta-\alpha} + \|\mathbb{X}\|_{2\beta,[0,t_3]} t_3^{2(\beta-\alpha)} + t_3^\alpha\big) \leq \frac{1}{2}$$

and rearranging, we obtain

$$\|Y' - \tilde{Y}'\|_{\alpha,[0,t_3]} + \|R^Y - R^{\tilde{Y}}\|_{2\alpha,[0,t_3]} \lesssim |Y_0 - \tilde{Y}_0| + \|\mathbf{X}; \tilde{\mathbf{X}}\|_{\alpha,[0,t_3]}.$$

It then follows from the estimate in (5.10) that

$$\|Y - \tilde{Y}\|_{\alpha,[0,t_3]} \lesssim |Y_0 - \tilde{Y}_0| + \|\mathbf{X}; \tilde{\mathbf{X}}\|_{\alpha,[0,t_3]}.$$

It follows that there exists a $\delta > 0$, depending on $\alpha, T, \|f\|_{C_b^3}$ and M, such that, for any interval $[s, t] \subset [0, T]$ with $|t - s| \leq \delta$, we have

$$\|Y' - \tilde{Y}'\|_{\alpha,[s,t]} + \|R^Y - R^{\tilde{Y}}\|_{2\alpha,[s,t]} \lesssim |Y_0 - \tilde{Y}_0| + \|\mathbf{X}; \tilde{\mathbf{X}}\|_{\alpha,[s,t]}, \tag{7.5}$$

$$\|Y - \tilde{Y}\|_{\alpha,[s,t]} \lesssim |Y_0 - \tilde{Y}_0| + \|\mathbf{X}; \tilde{\mathbf{X}}\|_{\alpha,[s,t]}. \tag{7.6}$$

Take a partition π of the interval $[0, T]$ with mesh size $|\pi| \leq \delta$. The estimates in (7.5) and (7.6) then hold over every interval $[s, t] \in \pi$, and by combining the estimates on different intervals one can deduce that the same estimates hold over the entire interval $[0, T]$. $\square$

8 Consistency with stochastic calculus

8.1 Stochastic integration

Recall from Section 4 that a d-dimensional Brownian motion B can be lifted to a random rough path $\mathbf{B} = (B, \mathbb{B})$. As always, the enhancement $\mathbb{B}$ is not unique, but we have two important examples, namely the Itô and Stratonovich enhancements. Let us first consider the Itô enhancement:

$$\mathbb{B}^{\text{Itô}}_{s,t} = \int_s^t B_{s,r} \otimes \mathrm{d}B_r, \qquad (s,t) \in \Delta_{[0,T]}.$$

We recall from Proposition 4.2 that $\mathbf{B}(\omega) = (B(\omega), \mathbb{B}^{\text{Itô}}(\omega)) \in \mathscr{C}^\alpha$ for almost every $\omega \in \Omega$. The following result shows that rough and stochastic integrals against Itô Brownian motion coincide whenever both are well-defined.

Proposition 8.1. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,T]}, \mathbb{P})$ be a filtered probability space. Let $\alpha \in (\frac{1}{3}, \frac{1}{2})$ and let $\mathbf{B} = \mathbf{B}^{\text{Itô}} = (B, \mathbb{B}) = (B, \mathbb{B}^{\text{Itô}})$ be an $\mathcal{F}_t$-adapted Itô enhanced Brownian rough path, so that $\mathbf{B} \in \mathscr{C}^\alpha$ almost surely. Let (Y, Y') be an adapted stochastic process such that $(Y(\omega), Y'(\omega)) \in \mathscr{D}^{2\alpha}_{B(\omega)}$ for almost every $\omega \in \Omega$. Then*

$$\int_0^T Y_u \,\mathrm{d}\mathbf{B}_u = \int_0^T Y_u \,\mathrm{d}B_u \tag{8.1}$$

almost surely.

It is helpful to make the dependence on ω explicit here. The equality in (8.1) means that, for almost every $\omega \in \Omega$, we have that

$$\int_0^T Y_u(\omega) \,\mathrm{d}\mathbf{B}_u(\omega) = \Big(\int_0^T Y_u \,\mathrm{d}B_u \Big)(\omega),$$

where, by Proposition 5.11, we have that

$$\int_0^T Y_u(\omega) \,\mathrm{d}\mathbf{B}_u(\omega) = \lim_{|\pi| \to 0} \sum_{[s,t] \in \pi} Y_s(\omega) B_{s,t}(\omega) + Y'_s(\omega) \mathbb{B}_{s,t}(\omega). \tag{8.2}$$

Proof of Proposition 8.1. Let $(\pi^n)_{n \geq 1}$ be a sequence of partitions with $|\pi^n| \to 0$ as $n \to \infty$. Recall that the Itô integral against Brownian motion can be written as the limit in probability

$$\sum_{[s,t] \in \pi^n} Y_s B_{s,t} \xrightarrow{\mathbb{P}} \int_0^T Y_u \,\mathrm{d}B_u \qquad \text{as} \quad n \to \infty.$$

There then exists a subsequence $(n_k)_{k \geq 1}$ such that

$$\sum_{[s,t] \in \pi^{n_k}} Y_s B_{s,t} \to \int_0^T Y_u \,\mathrm{d}B_u \qquad \text{as} \quad k \to \infty$$

almost surely. Combining this with (8.2), we have that

$$\sum_{[s,t] \in \pi^{n_k}} Y'_s \mathbb{B}_{s,t} \to \int_0^T Y_u \,\mathrm{d}\mathbf{B}_u - \int_0^T Y_u \,\mathrm{d}B_u \qquad \text{as} \quad k \to \infty \tag{8.3}$$

almost surely.

Since the Itô integral $\int_0^\cdot B_u \otimes \mathrm{d}B_u$ is a martingale, we can use the orthogonality of martingale increments. That is, for $[u,v],[s,t] \in \pi$ with $v \le s$, we have that $\mathbb{E}[Y'_u \mathbb{B}_{u,v} Y'_s \mathbb{B}_{s,t}] = \mathbb{E}[\mathbb{E}[Y'_u \mathbb{B}_{u,v} Y'_s \mathbb{B}_{s,t} \,|\, \mathcal{F}_s]] = \mathbb{E}[Y'_u \mathbb{B}_{u,v} Y'_s \mathbb{E}[\mathbb{B}_{s,t} \,|\, \mathcal{F}_s]] = 0$. For any partition π, we then have that

$$\mathbb{E}\bigg[\bigg|\sum_{[s,t]\in\pi} Y'_s \mathbb{B}_{s,t}\bigg|^2\bigg] = \sum_{[s,t]\in\pi} \mathbb{E}\big[|Y'_s \mathbb{B}_{s,t}|^2\big].$$

Let us assume for the moment that $\|Y'\|_{L^\infty(\Omega\times[0,T])} \le M$ for some constant $M > 0$. Recall from the proof of Proposition 4.2 that $\mathbb{E}[|\mathbb{B}_{s,t}|^2] \le C|t-s|^2$ for a constant C. Then

$$\mathbb{E}\bigg[\bigg|\sum_{[s,t]\in\pi} Y'_s \mathbb{B}_{s,t}\bigg|^2\bigg] = \sum_{[s,t]\in\pi} \mathbb{E}\big[|Y'_s \mathbb{B}_{s,t}|^2\big] \le CM^2 \sum_{[s,t]\in\pi} |t-s|^2 \le CM^2 T|\pi|. \tag{8.4}$$

Applying this with $\pi = \pi^{n_k}$, we have that

$$\sum_{[s,t]\in\pi^{n_k}} Y'_s \mathbb{B}_{s,t} \xrightarrow{L^2(\mathbb{P})} 0 \qquad \text{as} \quad k \to \infty.$$

We also know from (8.3) that this sequence of random variables converges almost surely. It follows that these limits must be equal almost surely.

If $\|Y'\|_{L^\infty(\Omega\times[0,T])}$ is not finite then we can use a localization argument: Let $M > 0$ and define the stopping time $\tau_M = T \wedge \inf\{t \in [0,T] : |Y'_t| \ge M\}$. Applying the argument above with the stopped processes $(Y')^{\tau_M}$ and $\mathbf{B}^{\tau_M} = (B^{\tau_M}, \mathbb{B}^{\tau_M})$, we deduce that $\int_0^{\tau_M} Y_u \,\mathrm{d}\mathbf{B}_u = \int_0^{\tau_M} Y_u \,\mathrm{d}B_u$ almost surely. Since $\tau_M \to T$ as $M \to \infty$ almost surely, the result then follows upon letting $M \to \infty$. $\square$

We now turn our attention to Stratonovich enhanced Brownian motion $\mathbf{B} = (B, \mathbb{B}^{\mathrm{Strat}})$, where

$$\mathbb{B}^{\mathrm{Strat}}_{s,t} = \int_s^t B_{s,r} \otimes \circ\,\mathrm{d}B_r, \qquad (s,t) \in \Delta_{[0,T]}.$$

Recall that Stratonovich integration is related to Itô integration by the relation

$$\int_0^T Y_s \circ \mathrm{d}X_s = \int_0^T Y_s \,\mathrm{d}X_s + \frac{1}{2}\langle Y, X\rangle_T \tag{8.5}$$

for semimartingales X, Y, where the limit in probability

$$\sum_{[s,t]\in\pi} Y_{s,t} X_{s,t} \xrightarrow{\mathbb{P}} \langle Y, X\rangle_T \qquad \text{as} \quad |\pi| \to 0,$$

is the quadratic covariation of Y and X.

By Proposition 4.3, we know that, almost surely, $\mathbf{B} = (B, \mathbb{B}^{\mathrm{Strat}}) \in \mathscr{C}_g^{0,\alpha}$ is a geometric rough path. Let us also recall the following useful equalities:

$$\mathbb{B}^{\mathrm{Strat}}_{s,t} = \mathbb{B}^{\mathrm{Itô}}_{s,t} + \frac{1}{2}(t-s)I, \tag{8.6}$$

and

$$\mathrm{Sym}(\mathbb{B}^{\mathrm{Strat}}_{s,t}) = \mathrm{Sym}(\mathbb{B}^{\mathrm{Itô}}_{s,t}) + \frac{1}{2}(t-s)I = \frac{1}{2}B_{s,t} \otimes B_{s,t}. \tag{8.7}$$

Similarly to the Itô case above, we now show that rough and stochastic integrals against Stratonovich Brownian motion coincide whenever both integrals are well-defined.

Proposition 8.2. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space. Let $\alpha \in (\frac{1}{3}, \frac{1}{2})$ and let $\mathbf{B} = \mathbf{B}^{\mathrm{Strat}} = (B, \mathbb{B}) = (B, \mathbb{B}^{\mathrm{Strat}})$ be an $\mathcal{F}_t$-adapted Stratonovich enhanced Brownian rough path, so that $\mathbf{B} \in \mathscr{C}_g^{0,\alpha}$ almost surely. Let (Y, Y') be an adapted stochastic process such that $(Y(\omega), Y'(\omega)) \in \mathscr{D}^{2\alpha}_{B(\omega)}$ for almost every $\omega \in \Omega$. Then*

$$\int_0^T Y_u \,\mathrm{d}\mathbf{B}_u = \int_0^T Y_u \circ \mathrm{d}B_u$$

almost surely.

Proof. The main step is to identify the quadratic covariation of Y and B. Recalling that $Y_{s,t} = Y'_s B_{s,t} + R^Y_{s,t}$, and using (8.7), we have that

$$\begin{aligned} Y_{s,t}B_{s,t} &= Y'_s(B_{s,t} \otimes B_{s,t}) + R^Y_{s,t}B_{s,t} \\ &= 2Y'_s \operatorname{Sym}(\mathbb{B}^{\mathrm{Itô}}_{s,t}) + Y'_s(t-s) + R^Y_{s,t}B_{s,t}. \end{aligned}$$

By the proof of Proposition 8.1, specifically (8.4), we saw that $\sum_{[s,t]\in\pi} Y'_s \mathbb{B}^{\mathrm{Itô}}_{s,t} \xrightarrow{\mathbb{P}} 0$ as $|\pi| \to 0$. It is easy to see that the same argument applies to the symmetric part $\operatorname{Sym}(\mathbb{B}^{\mathrm{Itô}})$, so that

$$\sum_{[s,t]\in\pi} Y'_s \operatorname{Sym}(\mathbb{B}^{\mathrm{Itô}}_{s,t}) \xrightarrow{\mathbb{P}} 0 \qquad \text{as} \quad |\pi| \to 0.$$

We have that

$$\sum_{[s,t]\in\pi} Y'_s(t-s) \to \int_0^T Y'_u \,\mathrm{d}u \qquad \text{as} \quad |\pi| \to 0,$$

almost surely, and that

$$\Bigg|\sum_{[s,t]\in\pi} R^Y_{s,t}B_{s,t}\Bigg| \leq \|R^Y\|_{2\alpha}\|B\|_\alpha \sum_{[s,t]\in\pi} |t-s|^{3\alpha} \leq \|R^Y\|_{2\alpha}\|B\|_\alpha T|\pi|^{3\alpha-1} \to 0 \qquad \text{as} \quad |\pi| \to 0$$

almost surely. Putting this all together, we have that

$$\sum_{[s,t]\in\pi} Y_{s,t}B_{s,t} \xrightarrow{\mathbb{P}} \int_0^T Y'_u \,\mathrm{d}u \qquad \text{as} \quad |\pi| \to 0,$$

so that

$$\langle Y, B\rangle_T = \int_0^T Y'_u \,\mathrm{d}u.$$

Applying Lemma 5.14 with $F_t = \frac{1}{2}tI$, in view of (8.6), we have that

$$\int_0^T Y_u \,\mathrm{d}\mathbf{B}^{\mathrm{Strat}}_u = \int_0^T Y_u \,\mathrm{d}\mathbf{B}^{\mathrm{Itô}}_u + \frac{1}{2}\int_0^T Y'_u \,\mathrm{d}u.$$

Thus, by Proposition 8.1 and (8.5), we then have

$$\int_0^T Y_u \,\mathrm{d}\mathbf{B}^{\mathrm{Strat}}_u = \int_0^T Y_u \,\mathrm{d}B_u + \frac{1}{2}\langle Y, B\rangle_T = \int_0^T Y_u \circ \mathrm{d}B_u.$$

$\square$

8.2 Stochastic differential equations

Given the consistency of rough and stochastic integrals shown above, it is straightforward to deduce consistency of rough and stochastic differential equations.

Proposition 8.3. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space. Let $f \in C_b^3$, $\alpha \in (\frac{1}{3}, \frac{1}{2})$, and $y \in L^2(\mathbb{P})$. Let B be a d-dimensional Brownian motion.*

(i) If $\mathbf{B}^{\text{Itô}} = (B, \mathbb{B}^{\text{Itô}}) \in \mathscr{C}^\alpha$ is Itô enhanced Brownian rough path, and (Y, Y') is the solution of the RDE

$$\mathrm{d}Y_t = f(Y_t)\,\mathrm{d}\mathbf{B}_t^{\text{Itô}}, \qquad Y_0 = y,$$

as given in Theorem 7.8, then Y is the unique strong solution of the Itô SDE

$$\mathrm{d}Y_t = f(Y_t)\,\mathrm{d}B_t, \qquad Y_0 = y.$$

(ii) Similarly, if $\mathbf{B}^{\text{Strat}} = (B, \mathbb{B}^{\text{Strat}}) \in \mathscr{C}_g^{0,\alpha}$ is Stratonovich enhanced Brownian rough path, and (Y, Y') is the solution of the RDE

$$\mathrm{d}Y_t = f(Y_t)\,\mathrm{d}\mathbf{B}_t^{\text{Strat}}, \qquad Y_0 = y,$$

as given in Theorem 7.8, then Y is the unique strong solution of the Stratonovich SDE

$$\mathrm{d}Y_t = f(Y_t) \circ \mathrm{d}B_t, \qquad Y_0 = y.$$

Proof. Since the Itô integral $\int_0^\cdot B_u \otimes \mathrm{d}B_u$ is adapted to the natural filtration generated by B, we have in particular that $\mathbb{B}_{s,r}^{\text{Itô}}$ is $\sigma(B_u : 0 \leq u \leq r)$-measurable. It follows that

$$\sigma(B_s, \mathbb{B}_{s,r}^{\text{Itô}} : 0 \leq s \leq r \leq t) = \sigma(B_u : 0 \leq u \leq t).$$

The continuity of the Itô–Lyons map (Theorem 7.9) tells us that the map

$$(B, \mathbb{B}^{\text{Itô}}) \longmapsto (Y, Y')$$

is continuous (with respect to suitable Hölder norms). It follows that the solution (Y, Y') is also adapted to the natural filtration generated by B.

By Proposition 8.1, we then have that

$$Y_t = y + \int_0^t f(Y_s)\,\mathrm{d}\mathbf{B}_s^{\text{Itô}} = y + \int_0^t f(Y_s)\,\mathrm{d}B_s,$$

so that Y does indeed solve the Itô SDE.

The proof in the Stratonovich case is the same, using Proposition 8.2 to conclude. □

9 Pathwise stability of likelihood estimators

In this section we shall see one of the earliest applications of rough path theory to statistics. We will see how rough paths can be used to obtain pathwise stability of maximum likelihood estimators for diffusion processes.

Let $\mathbb{V}$ be a finite-dimensional vector space, let $\Sigma \in \mathbb{R}^{d\times d}$, and let $h\colon \mathbb{R}^d \to \mathcal{L}(\mathbb{V};\mathbb{R}^d)$ be a Lipschitz continuous map. Consider the stochastic dynamics:

$$\mathrm{d}X_t = h(X_t)A\,\mathrm{d}t + \Sigma\,\mathrm{d}W_t \tag{9.1}$$

with $X_0 = x_0$. Here, $A \in \mathbb{V}$ is a parameter, W is a d-dimensional Brownian motion, and $x_0 \in \mathbb{R}^d$ is a (deterministic) initial value. We assume that Σ is nondegenerate, so that

$$C = \Sigma\Sigma^\top$$

is a positive definite symmetric matrix.

We are interested in estimating the parameter A by observing the process X up to some time T. We therefore consider the Maximum Likelihood Estimator (MLE) $\hat{A}_T$, which we can think of as a function on pathspace:

$$\hat{A}_T\colon C([0,T];\mathbb{R}^d) \to \mathbb{V}.$$

That is, given any observed path $X(\omega) = (X_t(\omega))_{t\in[0,T]}$, we have a corresponding estimate $\hat{A}_T(X(\omega)) \in \mathbb{V}$.

9.1 The classical MLE

Let's derive the MLE. Let $\mathbb{P}^0$ be a probability measure under which W is a d-dimensional standard Brownian motion, and define the process X by

$$X_t = x_0 + \Sigma W_t$$

for $t \in [0,T]$. For each $A \in \mathbb{V}$ and $t \in [0,T]$, let

$$f_t^A = \Sigma^{-1}h(X_t)A$$

and

$$W_t^A = W_t - \int_0^t f_s^A\,\mathrm{d}s.$$

By Girsanov's theorem (see e.g. [CE15, Chapter 15]), we have that W^A is a Brownian motion under the measure $\mathbb{P}^A$ defined via the Radon–Nikodym derivative:

$$\frac{\mathrm{d}\mathbb{P}^A}{\mathrm{d}\mathbb{P}^0} = \exp\left(\int_0^T (f_s^A)^\top\,\mathrm{d}W_s - \frac{1}{2}\int_0^T |f_s^A|^2\,\mathrm{d}s\right). \tag{9.2}$$

In particular, we then have that

$$\begin{aligned}\mathrm{d}X_t &= \Sigma\,\mathrm{d}W_t\\ &= \Sigma(f_t^A\,\mathrm{d}t + \mathrm{d}W_t^A)\\ &= h(X_t)A\,\mathrm{d}t + \Sigma\,\mathrm{d}W_t^A,\end{aligned}$$

so that X has the desired dynamics under $\mathbb{P}^A$.

The Radon–Nikodym derivative in (9.2) is precisely the likelihood function that we wish to maximize. We have

$$\begin{aligned}\log \frac{\mathrm{d}\mathbb{P}^A}{\mathrm{d}\mathbb{P}^0} &= \int_0^T (h(X_s)A)^\top (\Sigma^{-1})^\top \,\mathrm{d}W_s - \frac{1}{2}\int_0^T |\Sigma^{-1}h(X_s)A|^2 \,\mathrm{d}s \\ &= \int_0^T (h(X_s)A)^\top C^{-1} \,\mathrm{d}X_s - \frac{1}{2}\int_0^T (h(X_s)A)^\top C^{-1} h(X_s)A \,\mathrm{d}s.\end{aligned}$$

We wish to find the value of A which maximizes this expression. This is essentially just a case of finding the stationary point of a quadratic equation, and it is straightforward to see that the MLE $\hat{A}_T$ is characterized by the relation:

$$I_T \hat{A}_T = S_T,$$

where

$$S_T = \int_0^T h(X_s)^\top C^{-1} \,\mathrm{d}X_s \in \mathbb{V}^* \tag{9.3}$$

and

$$I_T = \int_0^T h(X_s)^\top C^{-1} h(X_s) \,\mathrm{d}s \in \mathcal{L}(\mathbb{V}; \mathbb{V}^*).$$

Here we write $\mathbb{V}^*$ for the dual space of $\mathbb{V}$, and the integral in (9.3) should be interpreted as an Itô integral. Thus, provided that I_T is invertible, the MLE is then given by

$$\hat{A}_T = I_T^{-1} S_T \in \mathbb{V}. \tag{9.4}$$

Example 9.1. Suppose that W and X are 1-dimensional, $\mathbb{V} = \mathbb{R}$, $\Sigma = \sigma > 0$, and h is just the identity map on $\mathbb{R}$, so that the underlying dynamics are given by

$$\mathrm{d}X_t = AX_t \,\mathrm{d}t + \sigma \,\mathrm{d}W_t$$

with $X_0 = x_0 \in \mathbb{R}$. In this case, we have that

$$I_T = \sigma^{-2} \int_0^T X_s^2 \,\mathrm{d}s,$$

and

$$S_T = \sigma^{-2} \int_0^T X_s \,\mathrm{d}X_s = \frac{\sigma^{-2}}{2}(X_T^2 - x_0^2 - \langle X \rangle_T) = \frac{\sigma^{-2}}{2}(X_T^2 - x_0^2 - \sigma^2 T), \tag{9.5}$$

so that, by (9.4), the MLE is given by

$$\hat{A}_T = \frac{X_T^2 - x_0^2 - \sigma^2 T}{2\int_0^T X_s^2 \,\mathrm{d}s}. \tag{9.6}$$

Note that this expression is well-defined provided that the path of X is not identically zero (which is actually impossible if $x_0 \neq 0$ by continuity).

In the simple case of Example 9.1 we have the explicit expression (9.6) for the MLE. Moreover, note that in this case the MLE is a continuous function on pathspace with respect to the supremum norm. This means that if two observation paths X and $\tilde{X}$ are close, in the sense that the distance $\sup_{t \in [0,T]} |X_t - \tilde{X}_t|$ is small, then the difference between the corresponding MLEs $|\hat{A}_T(X) - \hat{A}_T(\tilde{X})|$ will also be small. This stability property is very desirable. In particular, it means that in practice any small errors in our observations of X will not result in a large error in our estimation of the parameter A.

Unfortunately, pathwise stability with respect to the supremum norm does not hold in general. The fact that it holds in the example above is essentially because things tend to be nice when working in 1-dimension. We will see below an explicit example where this stability fails. An interesting question then arises: can we recover pathwise stability of the MLE with respect to a different topology? As we will see, it turns out that this stability can indeed be recovered if we consider the observation path X as a rough path.

To make our discussion rigorous, we should start by checking that the MLE as derived above is well-defined.

Lemma 9.2. *Define*

$$R_h = \big\{X \in C([0,T];\mathbb{R}^d) : \forall A \in \mathbb{V} \text{ with } A \neq 0,\, \exists t \in [0,T] \ \text{ such that } \ h(X_t)A \neq 0\big\}.$$

We claim that $I_T = I_T(X)$ is invertible for any $X \in R_h$.

In particular, if $\mathbb{P}^0(X \in R_h) = 1$, then the MLE $\hat{A}_T = \hat{A}_T(X) = I_T^{-1}(X)S_T(X)$ (as given in (9.4)*) is almost surely well-defined.*

Proof. Let $A \in \mathbb{V}$ with $A \neq 0$. We have that

$$I_T(A,A) = \int_0^T (h(X_s)A)^\top C^{-1}(h(X_s)A)\,\mathrm{d}s \geq 0.$$

Since C is positive definite by assumption, this expression is equal to zero if and only if $h(X_s)A = 0$ for all $s \in [0,T]$. Hence, for any $X \in R_h$, I_T is non-degenerate (i.e. has trivial kernel), and is therefore also invertible by standard results on bilinear forms over finite-dimensional spaces. □

9.2 Lack of continuity for the classical MLE

Example 9.3. Suppose that W and $X = (X^1, X^2)^\top$ are 2-dimensional, $\mathbb{V} = \mathbb{R}^{2\times 2}$, Σ is the 2×2-identity matrix, and h is the identity map on $\mathbb{R}^2$, so that the underlying dynamics are given by

$$\mathrm{d}X_t = AX_t\,\mathrm{d}t + \mathrm{d}W_t$$

with $X_0 = x_0 \in \mathbb{R}^2$. In this case, the MLE $\hat{A}_T$ satisfies the relation

$$I_T(\hat{A}_T, \cdot) = S_T.$$

That is, $\hat{A}_T \in \mathbb{R}^{2\times 2}$ satisfies

$$I_T(\hat{A}_T, H) = S_T(H) \qquad \text{for all} \quad H \in \mathbb{R}^{2\times 2},$$

where the functionals I_T and S_T are given, for any $A, H \in \mathbb{R}^{2\times 2}$, by

$$I_T(A,H) = \int_0^T X_s^\top H^\top A X_s\,\mathrm{d}s$$

and

$$S_T(H) = \int_0^T X_s^\top H^\top \, \mathrm{d}X_s.$$

Whenever the path X is such that I_T is invertible, the MLE can thus be computed by inverting I_T. In particular (after a slightly tedious calculation), the upper-left component of $\hat{A}_T$ is then given by

$$\hat{A}_T^{1,1} = \frac{1}{U_T}\Big(\int_0^T (X_s^2)^2 \, \mathrm{d}s \int_0^T X_s^1 \, \mathrm{d}X_s^1 - \int_0^T X_s^1 X_s^2 \, \mathrm{d}s \int_0^T X_s^2 \, \mathrm{d}X_s^1 \Big),$$

where

$$U_T = \int_0^T (X_s^1)^2 \, \mathrm{d}s \int_0^T (X_s^2)^2 \, \mathrm{d}s - \Big(\int_0^T X_s^1 X_s^2 \, \mathrm{d}s \Big)^2.$$

Since the variable U_T is defined in terms of simple Lebesgue integrals, it is clearly continuous in the path X with respect to the supremum norm. One can see that the integral $\int_0^T X_s^1 \, \mathrm{d}X_s^1$ is a continuous function of X^1 using Itô's formula, similarly to the calculation in (9.5). Thus, all the integrals in the expression for $\hat{A}_T^{1,1}$ are continuous functions of X, with the exception of the last integral, namely $\int_0^T X_s^2 \, \mathrm{d}X_s^1$. We will now exhibit a sequence of paths $(X^n)_{n \geq 1}$ which converge uniformly to a limiting path X for which I_T is invertible, such that the integral $\int_0^T X_s^{n,2} \, \mathrm{d}X_s^{n,1}$ diverges as $n \to \infty$.

Let $X \colon [0,1] \to \mathbb{R}^2$ be the path which starts at the origin, and moves, at constant speed, anticlockwise along the edges of the square with corners $(0,0)$, $(1,0)$, $(1,1)$ and $(0,1)$, finishing back at the origin. Note that $X \in R_h$, so that $I_T(X)$ is invertible and the corresponding MLE is well-defined by Lemma 9.2.

We now attach a fast spinning loop at the end of this path as follows. For each integer $n \geq 2$, we let

$$\begin{aligned} X_t^n &= X_{\frac{n}{n-1}t}, && t \in [0, (n-1)/n], \\ X_t^n &= \frac{1}{n}\begin{pmatrix} \cos(2\pi n^4 (t - \frac{n-1}{n})) - 1 \\ \sin(2\pi n^4 (t - \frac{n-1}{n})) \end{pmatrix}, && t \in [(n-1)/n, 1]. \end{aligned}$$

Note in particular that $X_0^n = X_{\frac{n-1}{n}}^n = X_1^n = (0,0)^\top$. We can therefore split up the integral as

$$\int_0^1 X_s^{n,2} \, \mathrm{d}X_s^{n,1} = \int_0^{\frac{n-1}{n}} X_{0,s}^{n,2} \, \mathrm{d}X_s^{n,1} + \int_{\frac{n-1}{n}}^1 X_{\frac{n-1}{n},s}^{n,2} \, \mathrm{d}X_s^{n,1}.$$

In particular, recalling the discussion in Section 4.3, we notice that the two integrals on the right-hand side above are equal to (minus) the Lévy area traced out by the path X^n over the corresponding time intervals. The first integral is hence simply (minus) the area of the unit square, i.e.

$$\int_0^{\frac{n-1}{n}} X_{0,s}^{n,2} \, \mathrm{d}X_s^{n,1} = -1.$$

Similarly, the second integral is simply (minus) the area of a circle with radius $1/n$, multiplied by n^3, which is the number of times that this circle is traced out by the path X^n. That is,

$$\int_{\frac{n-1}{n}}^1 X_{\frac{n-1}{n},s}^{n,2} \, \mathrm{d}X_s^{n,1} = -\pi n,$$

so that

$$\int_0^1 X_s^{n,2}\,\mathrm{d}X_s^{n,1} = -1 - \pi n.$$

Thus, we have that $X^n \to X$ uniformly as $n \to \infty$, but

$$|\hat{A}_T^{1,1}(X^n) - \hat{A}_T^{1,1}(X)| \longrightarrow \infty \qquad \text{as} \quad n \longrightarrow \infty.$$

9.3 Stability via rough paths

Clearly then, if we want to restore pathwise stability then we need a stronger topology. We will now see how rough path theory comes to the rescue. To gain access to our theory, we need to assume a bit more regularity on the underlying dynamics. Specifically, we will henceforth assume that $h \in C_b^2(\mathbb{R}^d; \mathcal{L}(\mathbb{V}; \mathbb{R}^d))$.

Let X be the (unique) solution of the SDE (9.1). As we saw in Section 9.1, X can simply be defined by $X = x_0 + \Sigma W$, where W is a Brownian motion under the reference measure $\mathbb{P}^0$, and then for any $A \in \mathbb{V}$, X has the desired dynamics under the corresponding measure $\mathbb{P}^A$.

We now define a rough path lift for X via Itô integration. That is, we let

$$\mathbb{X}_{s,t} = \int_s^t X_{s,r} \otimes \mathrm{d}X_r \tag{9.7}$$

for all $(s,t) \in \Delta_{[0,T]}$, where the integral is defined in the sense of Itô under the measure $\mathbb{P}^0$. We then know that the pair $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ is almost surely an α-Hölder rough path for any $\alpha \in (\frac{1}{3}, \frac{1}{2})$.

Note that since, for any parameter $A \in \mathbb{V}$, the measure $\mathbb{P}^A$ is equivalent to $\mathbb{P}^0$, the stochastic integral in (9.7) is almost surely equal to the same integral defined under $\mathbb{P}^A$.

Let

$$\mathbb{D} = \big\{(X, \mathbb{X}) \in \mathscr{C}^\alpha : X \in R_h\big\}.$$

It is then clear that, if $\mathbb{P}^0(X \in R_h) = 1$, then $\mathbb{P}^0(\mathbf{X} \in \mathbb{D}) = 1$.

Recall the bilinear form I_T, which we may consider as a map from $R_h \to \mathcal{L}(\mathbb{V}; \mathbb{V}^*)$, given by

$$I_T(X) = \int_0^T h(X_s)^\top C^{-1} h(X_s)\,\mathrm{d}s.$$

We now define the map $\mathbf{S}_T : \mathbb{D} \to \mathbb{V}^*$, given by

$$\mathbf{S}_T(X, \mathbb{X}) = \int_0^T h(X_s)^\top C^{-1}\,\mathrm{d}\mathbf{X}_s,$$

where the integral is defined as the (deterministic) rough integral against $\mathbf{X} = (X, \mathbb{X})$. Note that, since $h \in C_b^2$, it is clear that the integrand is a controlled path with respect to X, so that this integral is well-defined.

We can now define a "robust MLE" as the map $\hat{\mathbf{A}}_T : \mathbb{D} \to \mathbb{V}$ given by

$$\hat{\mathbf{A}}_T(X, \mathbb{X}) = I_T^{-1}(X)\mathbf{S}_T(X, \mathbb{X}).$$

Proposition 9.4. *The map* $\hat{\mathbf{A}}_T$ *defined above is a continuous map from* $\mathbb{D} \to \mathbb{V}$ *with respect to the* α*-Hölder rough path distance.*

Moreover, writing $\mathbf{X} = (X, \mathbb{X})$ *for the Itô rough path lift of* X *as in* (9.7)*, we have that*

$$\mathbb{P}^0(\hat{\mathbf{A}}_T(\mathbf{X}) = \hat{A}_T(X)) = 1.$$

Proof. The map I_T is clearly continuous with respect to the (weaker) supremum norm, and hence so is its inverse (which is defined for all $\mathbf{X} \in \mathbb{D}$). The map $\mathbf{S}_T$ is continuous with respect to the rough path distance by the stability of rough integration given in Corollary 5.16. The composition of these two maps is then also continuous.

The second statement follows from the consistency of rough and stochastic integrals, as given in Proposition 8.1. (Strictly speaking this proposition was only stated for standard Brownian motion, but here X is essentially just a Brownian motion under a linear map with a drift, so this extension is trivial.) □

For more details on this topic, including further applications, see the article [DFM16].

10 Parameter uncertainty in stochastic filtering

In this section we shall discuss an application to stochastic filtering. No prior knowledge of filtering theory is assumed, and we shall in any case restrict ourselves to a relatively simple setting in order to avoid virtually all technical difficulties.

10.1 Stochastic filtering

In many applications, from finance and biology to engineering, defence and aerospace, one is interested in the behaviour of a process evolving in time which cannot be observed directly, and must therefore rely on partial observations in the presence of noise. The problem of estimating the current state of such a hidden process from noisy observations is known as *stochastic filtering*.

Suppose for instance that we are interested in the value of a stochastic process S, referred to as the *signal*, but that we must work in the filtration generated by another process Y, referred to as the *observation process*. We also suppose that the dynamics of Y are dependent on the value of S, such as through a relationship of the form

$$\mathrm{d}Y_t = h(S_t)\,\mathrm{d}t + \mathrm{d}W_t,$$

for some observation function h and measurement noise W. It is clear from this relationship that by observing Y over a given period of time one can infer information about S. In short, the filtering problem is concerned with, at each time t, determining the best estimate (typically in the sense of best mean square) for S_t given $\mathcal{Y}_t := \sigma(Y_s : 0 \leq s \leq t)$; that is, finding the best estimate for the current value of S, given our past observations of Y.

An important special case is when the signal and observation are both diffusion processes with linear dynamics. Let's take an underlying filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0})$. We suppose that an $\mathbb{R}^m$-valued signal process S and an $\mathbb{R}^d$-valued observation process Y satisfy the following pair of linear equations

$$\mathrm{d}S_t = (\gamma_t + \alpha_t S_t)\,\mathrm{d}t + \sigma_t\,\mathrm{d}B_t, \tag{10.1}$$

$$\mathrm{d}Y_t = c_t S_t\,\mathrm{d}t + \mathrm{d}W_t, \tag{10.2}$$

with the initial conditions $Y_0 = 0$ and $S_0 \sim N(\mu_0, \Sigma_0)$ for some $\mu_0 \in \mathbb{R}^m$ and $\Sigma_0 \in \mathcal{S}_+^m$, where $\mathcal{S}_+^m$ denotes the set of symmetric, positive definite $m \times m$-matrices. Here B (resp. W) is a standard $\mathbb{R}^l$ (resp. $\mathbb{R}^d$)-valued Brownian motion, and $\gamma\colon [0,T] \to \mathbb{R}^m$, $\alpha\colon [0,T] \to \mathbb{R}^{m\times m}$, $\sigma\colon [0,T] \to \mathbb{R}^{m\times l}$ and $c\colon [0,T] \to \mathbb{R}^{d\times m}$ are parameters. Here we include the case when the signal noise and observation noise are correlated; we suppose that their quadratic covariation is given by

$$\mathrm{d}\langle B, W\rangle_t = \rho_t \,\mathrm{d}t, \tag{10.3}$$

for some correlation matrix $\rho\colon [0,T] \to \mathbb{R}^{l\times d}$. In the scalar case, the correlation should naturally satisfy $\rho^2 \leq 1$. The analogous assumption here is that the matrix $I - \rho\rho^\top$ be positive semi-definite, where I denotes the $l \times l$ identity matrix.

We shall denote by $(\mathcal{Y}_t)_{t\geq 0}$ the (completed) natural filtration generated by the observation process Y. In this setting the signal and observation are both Gaussian processes, and in fact it can be shown that the posterior distribution of the signal given our observations is also Gaussian; see e.g. [BC09, Section 6.2]. That is, at each time $t \geq 0$, we have that

$$S_t \,|\, \mathcal{Y}_t \sim N(q_t, R_t)$$

for some (random) mean vector

$$q_t = \mathbb{E}[S_t \,|\, \mathcal{Y}_t]$$

and covariance matrix

$$R_t = \mathbb{E}[(S_t - q_t)(S_t - q_t)^\top \,|\, \mathcal{Y}_t].$$

Moreover, the conditional mean and covariance satisfy the dynamics:

$$\mathrm{d}q_t = (\gamma_t + \alpha_t q_t)\,\mathrm{d}t + (R_t c_t^\top + \sigma_t\rho_t)(\mathrm{d}Y_t - c_t q_t\,\mathrm{d}t), \tag{10.4}$$

$$\frac{\mathrm{d}R_t}{\mathrm{d}t} = \sigma_t\sigma_t^\top + \alpha_t R_t + R_t\alpha_t^\top - (R_t c_t^\top + \sigma_t\rho_t)(c_t R_t + \rho_t^\top\sigma_t^\top) \tag{10.5}$$

with $q_0 = \mu_0$ and $R_0 = \Sigma_0$. That is, the covariance R satisfies a matrix Riccati equation and, given R, the mean q satisfies a linear SDE driven by the observation process Y.

Given a stream of data corresponding to the observation process Y, one may thus simply solve the equations (10.4)–(10.5) to compute the posterior distribution of the signal S. This procedure (with linear Gaussian underlying dynamics) is known as the Kalman–Bucy filter, and its impact on engineering and aerospace over the last 60 years cannot be overstated.

10.2 Parameter uncertainty

The filtering equations (10.4)–(10.5) involve various parameters so, naturally, in order to run the filter one must first obtain the values of these parameters. In standard treatments of stochastic filtering one often simply runs the filter using an estimate of the parameters. However, this does not take into account the statistical uncertainty introduced by adopting this estimate. Particularly when there is limited available data, resulting in a lack of precision in the estimate, this should cast doubt as to the accuracy of the filter. We therefore now turn our attention to *parameter uncertainty.*

We will focus on the simplest version of the problem, where the drift parameter γ is unknown. For simplicity, we shall assume that the other parameters, namely α, σ, c and ρ are known and constant. We shall also suppose that the prior mean μ_0 is unknown.

Note that the equation for the posterior covariance (10.5) does not depend on the unknown parameter γ. Since the other parameters are constant in time, the solution of this equation will converge as $t \to \infty$ to some (known) stationary value R_∞. It is therefore convenient to assume for simplicity that $\Sigma_0 = R_\infty$, so that the posterior covariance is constant and equal to its stationary value $R = R_\infty$. The remaining filtering equation is given by

$$\mathrm{d}q_t = (\gamma_t + \alpha q_t)\,\mathrm{d}t + (Rc^\top + \sigma\rho)(\mathrm{d}Y_t - cq_t\,\mathrm{d}t),$$

where we suppose that $\gamma\colon [0,T] \to \mathbb{R}^m$ and $q_0 = \mu_0 \in \mathbb{R}^m$ are unknown.

To incorporate parameter uncertainty we adopt the following setup. Let S and Y be $\mathcal{F}_t$-adapted processes. For each parameter choice (γ, μ_0), we let $\mathbb{P}^{\gamma,\mu_0}$ be a probability measure under which S and Y satisfy the dynamics (10.1)–(10.3) with the parameters (γ, μ_0). (In particular, the processes B and W are not fixed, but depend on the choice of the parameters, so that they have the law of a Brownian motion under each measure $\mathbb{P}^{\gamma,\mu_0}$.)

Let γ^* and μ_0^* be some reference parameters. Then, for any choice of parameters γ, μ_0, we can consider the likelihood ratio

$$\left(\frac{\mathrm{d}\mathbb{P}^{\gamma,\mu_0}}{\mathrm{d}\mathbb{P}^{\gamma^*,\mu_0^*}}\right)_{\mathcal{Y}_t}$$

with respect to the observation filtration $(\mathcal{Y}_t)_{t\in[0,T]}$. It is a classical result in filtering theory (see e.g. [BC09, Chapter 2]) that the so-called *innovation* process V, given in this setting by

$$\mathrm{d}V_t = \mathrm{d}Y_t - cq_t\,\mathrm{d}t,$$

is a $\mathcal{Y}_t$-adapted Brownian motion under $\mathbb{P}^{\gamma,\mu_0}$. Writing q^* (resp. V^*) for the posterior mean (resp. innovation process) under the reference measure $\mathbb{P}^{\gamma^*,\mu_0^*}$, we have that

$$\mathrm{d}V_t = \mathrm{d}V_t^* - c(q_t - q_t^*)\,\mathrm{d}t.$$

Hence, by Girsanov's theorem (see e.g. [CE15, Chapter 15]), we can represent the likelihood as a stochastic exponential, namely

$$\left(\frac{\mathrm{d}\mathbb{P}^{\gamma,\mu_0}}{\mathrm{d}\mathbb{P}^{\gamma^*,\mu_0^*}}\right)_{\mathcal{Y}_t} = \exp\left(\int_0^t c(q_s - q_s^*)\cdot \mathrm{d}V_s^* - \frac{1}{2}\int_0^t |c(q_s - q_s^*)|^2\,\mathrm{d}s\right).$$

Substituting $\mathrm{d}V_s^* = \mathrm{d}Y_s - cq_s^*\,\mathrm{d}s$, a short calculation yields that the negative log-likelihood is then given by

$$-\log\left(\frac{\mathrm{d}\mathbb{P}^{\gamma,\mu_0}}{\mathrm{d}\mathbb{P}^{\gamma^*,\mu_0^*}}\right)_{\mathcal{Y}_t} = -\int_0^t c(q_s - q_s^*)\cdot \mathrm{d}Y_s + \frac{1}{2}\int_0^t \left(|cq_s|^2 - |cq_s^*|^2\right)\mathrm{d}s.$$

Since the reference parameters are taken to be fixed, they simply amount to an additive constant in the above expression. That is,

$$-\log\left(\frac{\mathrm{d}\mathbb{P}^{\gamma,\mu_0}}{\mathrm{d}\mathbb{P}^{\gamma^*,\mu_0^*}}\right)_{\mathcal{Y}_t} = -\int_0^t cq_s\cdot \mathrm{d}Y_s + \frac{1}{2}\int_0^t |cq_s|^2\,\mathrm{d}s + \text{const.} \tag{10.6}$$

Since this constant does not depend on the choice of parameters γ, μ_0, it will not affect any of our subsequent analysis, and for simplicity it will therefore be omitted.

It will be useful later to interpret the stochastic integral appearing in (10.6) in the sense of Stratonovich, rather than that of Itô. We therefore make the transformation

$$
\begin{aligned}
-\int_0^t cq_s \cdot \mathrm{d}Y_s &= -\int_0^t cq_s \circ \mathrm{d}Y_s + \frac{1}{2}\langle cq, Y\rangle_t \\
&= -\int_0^t cq_s \circ \mathrm{d}Y_s + \frac{1}{2}\int_0^t \operatorname{tr}\left(c(Rc^\top + \sigma\rho)\right)\mathrm{d}s, \\
&= -\int_0^t cq_s \circ \mathrm{d}Y_s + \text{const.}
\end{aligned}
$$

where $\operatorname{tr}(\cdot)$ denotes the trace. In this simple setting, the additive constant here does not depend in any way on the uncertain parameters γ, μ_0, so we can also simply omit this constant henceforth. We have thus derived the representation:

$$
-\log\left(\frac{\mathrm{d}\mathbb{P}^{\gamma,\mu_0}}{\mathrm{d}\mathbb{P}^{\gamma^*,\mu_0^*}}\right)_{\mathcal{Y}_t} = -\int_0^t cq_s \circ \mathrm{d}Y_s + \frac{1}{2}\int_0^t |cq_s|^2\,\mathrm{d}s. \tag{10.7}
$$

In the setting of parameter uncertainty, we do not know which parameter γ is the correct one. However, it seems sensible to suggest that the most "reasonable" parameter is the one which minimizes the negative log-likelihood. We thus wish to formulate an optimization procedure, in which we try to minimize the expression in (10.7) over the set of all possible parameters γ. We will actually modify this expression somewhat to also incorporate our prior beliefs about which parameter values we believe are most plausible. We shall seek to minimize a general expression of the form

$$
\int_0^t f(q_s, \gamma_s, \dot{\gamma}_s)\,\mathrm{d}s + \int_0^t \psi(q_s) \circ \mathrm{d}Y_s + g(q_0, \gamma_0), \tag{10.8}
$$

over the set of all Lipschitz continuous paths γ, where we recall that the posterior mean q satisfies

$$
\mathrm{d}q_t = (\gamma_t + \alpha q_t)\,\mathrm{d}t + (Rc^\top + \sigma\rho)(\mathrm{d}Y_t - cq_t\,\mathrm{d}t). \tag{10.9}
$$

Here $\psi(q) = -cq$, and we have absorbed the term $\frac{1}{2}|cq|^2$ into the function f, which we also allow to depend on the derivative $\dot{\gamma}$ of γ (which we recall is defined almost everywhere since γ is Lipschitz). This means in particular that we can encode, not only which values of γ we believe are reasonable, but also how quickly we believe they should be able to vary. For example, if we believe that the parameter γ should remain fairly constant in time, then we can include a large penalty on the magnitude of $\dot{\gamma}$.

10.3 A pathwise optimal control problem

Note that, since our inference will naturally depend on our observations, the optimization procedure should be carried out in a pathwise manner, i.e. for each individual realization of the observation process Y. We therefore need to be able to handle the stochastic integral appearing in (10.8) in a pathwise sense. To this end, we lift the observation process Y to a rough path using Stratonovich integration. Under the reference measure $\mathbb{P}^{\gamma^*,\mu_0^*}$, we let

$$
\mathbb{Y}_{s,t} = \int_s^t Y_{s,r} \otimes \circ\,\mathrm{d}Y_r,
$$

for all $(s,t) \in \Delta_{[0,T]}$, so that $\mathbf{Y} = (Y, \mathbb{Y}) \in \mathscr{C}_g^{0,\alpha}$ is almost surely a geometric rough path for any $\alpha \in (\frac{1}{3}, \frac{1}{2})$. In this setting the measures $\mathbb{P}^{\gamma,\mu_0}$ are all equivalent, so we immediately have that $\mathbb{Y}$ coincides almost surely with the same integral defined under any other choice of measure $\mathbb{P}^{\gamma,\mu_0}$.

It follows from (10.9) that q is (almost surely) a controlled path with respect to Y, with Gubinelli derivative given by $q' = Rc^\top + \sigma\rho$. We then have that $\psi(q) = -cq$ is also a controlled path, with derivative $\psi(q)' = -c(Rc^\top + \sigma\rho)$. By Proposition 8.2, we have that

$$\int_0^t \psi(q_s) \circ \mathrm{d}Y_s = \int_0^t \psi(q_s)\, \mathrm{d}\mathbf{Y}_s$$

almost surely.

For notational simplicity, let us write $b(q,\gamma) := \gamma + \alpha q - (Rc^\top + \sigma\rho)cq$ and $\phi = Rc^\top + \sigma\rho$, so that equation (10.9) may be rewritten as

$$\mathrm{d}q_t = b(q_t, \gamma_t)\, \mathrm{d}t + \phi\, \mathrm{d}Y_t. \tag{10.10}$$

We can now formulate our procedure as the following (pathwise) optimal control problem. Let $\mathcal{U}$ be the space of all bounded measurable paths $u\colon [0,T] \to \mathbb{R}^m$. We first fix an (enhanced) observation path $\mathbf{Y} = \mathbf{Y}(\omega) \in \mathscr{C}_g^{0,\alpha}$. In particular, since ϕ is just a constant, we can interpret (10.10), not as an SDE, but simply as an ODE driven by the deterministic path $Y = Y(\omega)$.

We can then define the *value function* $v\colon [0,T] \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ by

$$v(t,x,a) = \inf_{u \in \mathcal{U}} \left\{ \int_0^t f(q_s, \gamma_s, u_s)\, \mathrm{d}s + \int_0^t \psi(q_s)\, \mathrm{d}\mathbf{Y}_s + g(q_0, \gamma_0) \right\}, \tag{10.11}$$

where we interpret the path u as a *control*, and where the state variables q and γ satisfy the controlled dynamics

$$\begin{aligned} \mathrm{d}q_s &= b(q_s, \gamma_s)\, \mathrm{d}s + \phi\, \mathrm{d}Y_s, & q_t &= x, \\ \mathrm{d}\gamma_s &= u_s\, \mathrm{d}s, & \gamma_t &= a. \end{aligned} \tag{10.12}$$

For each posterior value $x \in \mathbb{R}^m$ and parameter value $a \in \mathbb{R}^m$, the control problem above seeks the minimum 'cost' associated with the trajectory of a parameter γ and filter q, which would be consistent with the observed path of Y, where the cost is derived from the negative log-likelihood function. The value function v thus gives, at each time $t \in [0,T]$, a measure of the 'unreasonability' of different posterior values x and parameter values a.

In practice it would be impossible to directly compute the infimum in (10.11). As is a standard technique in control theory, we therefore instead consider the PDE satisfied by the value function, which we can then try to solve.

The first step is to take a smooth approximation of the observation path Y, which is close to $\mathbf{Y}$ in rough path topology.

Let $\eta\colon [0,T] \to \mathbb{R}^d$ be a smooth path. Recall that we can lift η in a canonical way to a rough path $\boldsymbol{\eta} = (\eta, \eta^{(2)})$ by defining

$$\eta^{(2)}_{s,t} := \int_s^t \eta_{s,r} \otimes \mathrm{d}\eta_r \tag{10.13}$$

for $(s,t) \in \Delta_{[0,T]}$, where the integral exists in the Riemann–Stieltjes sense.

By replacing $\mathbf{Y}$ in our control problem by the smooth path η, we obtain the approximate control problem, in which we have the approximate value function

$$v^\eta(t,x,a) = \inf_{u\in\mathcal{U}} \left\{ \int_0^t f(q_s,\gamma_s,u_s)\,\mathrm{d}s + \int_0^t \psi(q_s)\,\mathrm{d}\eta_s + g(q_0,\gamma_0) \right\}, \tag{10.14}$$

where the state variables q and γ satisfy the controlled dynamics

$$\begin{aligned} \mathrm{d}q_s &= b(q_s,\gamma_s)\,\mathrm{d}s + \phi\,\mathrm{d}\eta_s, & q_t &= x,\\ \mathrm{d}\gamma_s &= u_s\,\mathrm{d}s, & \gamma_t &= a. \end{aligned}$$

The PDE associated with this control problem is the Hamilton–Jacobi (HJ) equation

$$\frac{\partial v^\eta}{\partial t} + b\cdot\nabla_x v^\eta + \sup_{u\in\mathbb{R}^m}\{u\cdot\nabla_a v^\eta - f\} + (\phi\cdot\nabla_x v^\eta - \psi)\dot{\eta} = 0 \tag{10.15}$$

with the initial condition

$$v^\eta(0,\cdot,\cdot) = g. \tag{10.16}$$

Here $\dot{\eta}$ denotes the time derivative of η, and ∇_x (resp. ∇_a) denotes the gradient with respect to the x (resp. a) variable.

Theorem 10.1. *Under natural conditions on the functions f,g (local Lipschitz continuity, with superlinear growth of f and asymptotic explosion of g), the approximate value function v^η (as defined in* (10.14)*) is the unique solution of the HJ equation* (10.15)–(10.16).

This result is requires considerable work and careful analysis, and is beyond the scope of this course. We will simply take it for granted.

Replacing the smooth path η by the rough path $\mathbf{Y}$, we formally obtain the rough Hamilton–Jacobi (rough HJ) equation

$$\mathrm{d}v + b\cdot\nabla_x v\,\mathrm{d}t + \sup_{u\in\mathbb{R}^m}\{u\cdot\nabla_a v - f\}\,\mathrm{d}t + (\phi\cdot\nabla_x v - \psi)\,\mathrm{d}\mathbf{Y} = 0 \tag{10.17}$$

with the initial condition

$$v(0,\cdot,\cdot) = g. \tag{10.18}$$

We still need to say what we actually mean by a solution of such a rough PDE, which we do with the following definition.

Definition 10.2. Given a smooth path η, we write $\boldsymbol{\eta} = (\eta,\eta^{(2)})$ for its canonical rough path lift, with $\eta^{(2)}$ defined as in (10.13). We write v^η for the unique solution of (10.15)–(10.16), which, by Theorem 10.1, is precisely the approximate value function, as defined in (10.14). We say that a continuous function v solves the rough HJ equation (10.17)–(10.18) if

$$v^{\eta^n} \longrightarrow v \qquad \text{as} \quad n \longrightarrow \infty$$

locally uniformly, whenever $(\eta^n)_{n\geq 1}$ is a sequence of smooth paths such that $\boldsymbol{\eta}^n \to \mathbf{Y}$ with respect to the α-Hölder rough path distance, i.e. $\|\boldsymbol{\eta}^n;\mathbf{Y}\|_\alpha \to 0$ as $n\to\infty$.

By the uniqueness of limits, if such a solution of (10.17)–(10.18) exists, then it is unique. Moreover, note that, since the rough path $\mathbf{Y}$ is geometric, there certainly exists such a sequence of smooth paths $(\eta^n)_{n\geq 1}$.

Theorem 10.3. *Under suitable conditions on the functions f, g, the value function v solves the rough HJ equation* (10.17)–(10.18) *in the sense of Definition 10.2.*

Moreover, writing $v = v^{\mathbf{Y}}$, the map from $\mathscr{C}_g^{0,\alpha} \to \mathbb{R}$ given by $\mathbf{Y} \mapsto v^{\mathbf{Y}}(t,x,a)$ is locally uniformly continuous with respect to the α-Hölder rough path distance, locally uniformly in (t,x,a).

Sketch of proof. Let $\mathbf{Z} = (Z, \mathbb{Z}) \in \mathscr{C}^\alpha$ be another rough path. Let us write q^Y (resp. q^Z) for the solution of (10.12) driven by Y (resp. Z), and write $v^{\mathbf{Y}}$ (resp. $v^{\mathbf{Z}}$) for the corresponding value function, as defined in (10.11).

For any (t,x,a), we have that

$$\begin{aligned}
&|v^{\mathbf{Y}}(t,x,a) - v^{\mathbf{Z}}(t,x,a)| \\
&\leq \sup_{u\in\mathcal{U}} \bigg| \int_0^t \big(f(q_s^Y, \gamma_s, u_s) - f(q_s^Z, \gamma_s, u_s)\big)\,\mathrm{d}s \\
&\qquad + \int_0^t \psi(q_s^Y)\,\mathrm{d}\mathbf{Y}_s - \int_0^t \psi(q_s^Z)\,\mathrm{d}\mathbf{Z}_s + g(q_0^Y, \gamma_0) - g(q_0^Z, \gamma_0)\bigg| \\
&\lesssim \sup_{u\in\mathcal{U}} \bigg\{ \int_0^t |q_s^Y - q_s^Z|\,\mathrm{d}s + \bigg| \int_0^t \psi(q_s^Y)\,\mathrm{d}\mathbf{Y}_s - \int_0^t \psi(q_s^Z)\,\mathrm{d}\mathbf{Z}_s \bigg| + |q_0^Y - q_0^Z| \bigg\}.
\end{aligned}$$

By the stability of rough integration and differential equations, we deduce that

$$|v^{\mathbf{Y}}(t,x,a) - v^{\mathbf{Z}}(t,x,a)| \lesssim \|\mathbf{Y}; \mathbf{Z}\|_\alpha.$$

Taking a sequence of smooth paths $(\eta^n)_{n\geq 1}$ such that $\|\boldsymbol{\eta}^n; \mathbf{Y}\|_\alpha \to 0$ as $n \to \infty$, the required convergence follows by taking $\mathbf{Z} = \boldsymbol{\eta}^n$ in the above. The stated continuity of the value function with respect to the driving rough path is also immediate from the above. □

In this section we have demonstrated how one can give meaning to a rough PDE, and seen an example of how such an equation arises in an application to statistics. For further details on this topic in a more general setting, see Section 4 of the article [AC20].

11 The stochastic sewing lemma

In Section 5.1 we introduced a basic version of the sewing lemma. Due to its usefulness, this result has been generalized in several ways (e.g. sewing lemmas for discontinuous paths, and nonlinear sewing lemmas). In this section we will study a recent and powerful version of the result, known as the stochastic sewing lemma.

11.1 Proof of the stochastic sewing lemma

A key ingredient in the stochastic sewing lemma is the following observation. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $Z_1, \dots, Z_n$ be a finite sequence of $\mathbb{R}^d$-valued random variables in $L^m = L^m(\Omega, \mathcal{F}, \mathbb{P})$ for some $m \in [2, \infty)$. Suppose that we are given a filtration $(\mathcal{F}_i)_{i=1,\dots,n}$

such that, for each i, the variables $Z_1, \dots, Z_{i-1}$ are all $\mathcal{F}_i$-measurable. Suppose that we wish to estimate the sum

$$S = \sum_{i=1}^{n} Z_i.$$

Rather than estimating this directly, let us first write

$$S = S_1 + S_2$$

where

$$S_1 = \sum_{i=1}^{n} \mathbb{E}^{\mathcal{F}_i} Z_i, \qquad \text{and} \qquad S_2 = \sum_{i=1}^{n} (Z_i - \mathbb{E}^{\mathcal{F}_i} Z_i).$$

We simply estimate S_1 as

$$\|S_1\|_{L^m} \leq \sum_{i=1}^{n} \|\mathbb{E}^{\mathcal{F}_i} Z_i\|_{L^m}. \tag{11.1}$$

However, the martingale structure of S_2 allows us to do better. We have

$$\begin{aligned}
\|S_2\|_{L^m} &= \mathbb{E}\bigg[\bigg|\sum_{i=1}^{n}(Z_i - \mathbb{E}^{\mathcal{F}_i} Z_i)\bigg|^m\bigg]^{\frac{1}{m}} \leq C_m \mathbb{E}\bigg[\bigg(\sum_{i=1}^{n} |Z_i - \mathbb{E}^{\mathcal{F}_i} Z_i|^2\bigg)^{\frac{m}{2}}\bigg]^{\frac{1}{m}} \\
&= C_m \bigg\|\sum_{i=1}^{n} |Z_i - \mathbb{E}^{\mathcal{F}_i} Z_i|^2\bigg\|_{L^{m/2}}^{\frac{1}{2}} \leq C_m \bigg(\sum_{i=1}^{n} \|Z_i - \mathbb{E}^{\mathcal{F}_i} Z_i\|_{L^m}^2\bigg)^{\frac{1}{2}} \\
&\leq C_m \bigg(\sum_{i=1}^{n} \big(\|Z_i\|_{L^m} + \|\mathbb{E}^{\mathcal{F}_i} Z_i\|_{L^m}\big)^2\bigg)^{\frac{1}{2}} \leq 2C_m \bigg(\sum_{i=1}^{n} \|Z_i\|_{L^m}^2\bigg)^{\frac{1}{2}},
\end{aligned} \tag{11.2}$$

where we have used the Burkholder–Davis–Gundy (BDG) inequality, Minkowski's inequality[2] (i.e. the triangle inequality for L^m spaces), and the contraction property of conditional expectations with respect to the L^m norm.

We thus have the estimate

$$\|S\|_{L^m} \leq \sum_{i=1}^{n} \|\mathbb{E}^{\mathcal{F}_i} Z_i\|_{L^m} + 2C_m \bigg(\sum_{i=1}^{n} \|Z_i\|_{L^m}^2\bigg)^{\frac{1}{2}}. \tag{11.3}$$

This estimate will be used multiple times in the proof of the stochastic sewing lemma.

Theorem 11.1 (Stochastic sewing lemma)**.** *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space, and let $m \geq 2$. Let $A\colon \Delta_{[0,T]} \to L^m$ be a continuous map, such that, for all $s \leq t$, $A_{s,s} = 0$ and $A_{s,t}$ is $\mathcal{F}_t$-adapted. Suppose that, for some constants $\lambda_1, \lambda_2 \geq 0$ and $\varepsilon_1, \varepsilon_2 > 0$, we have that*

$$\|\mathbb{E}_s \delta A_{s,u,t}\|_{L^m} \leq \lambda_1 |t-s|^{1+\varepsilon_1}, \tag{11.4}$$

$$\|\delta A_{s,u,t}\|_{L^m} \leq \lambda_2 |t-s|^{\frac{1}{2}+\varepsilon_2}, \tag{11.5}$$

for all $0 \leq s \leq u \leq t \leq T$, where $\mathbb{E}_s$ denotes the conditional expectation at time s, and as usual $\delta A_{s,u,t} := A_{s,t} - A_{s,u} - A_{u,t}$.

Then there exists a unique (up to modifications) stochastic process $\gamma = (\gamma_t)_{t\in[0,T]}$ such that

[2] This requires $\frac{m}{2} \geq 1$, which is why we assumed that $m \geq 2$.

(i) $\gamma_0 = 0$*, and* γ *is* $\mathcal{F}_t$*-adapted and* L^m*-integrable,*

(ii) and there exist constants C_1, C_2 *such that*

$$\|\gamma_t - \gamma_s - A_{s,t}\|_{L^m} \le C_1|t-s|^{1+\varepsilon_1} + C_2|t-s|^{\frac{1}{2}+\varepsilon_2}, \tag{11.6}$$

$$\|\mathbb{E}_s(\gamma_t - \gamma_s - A_{s,t})\|_{L^m} \le C_1|t-s|^{1+\varepsilon_1} \tag{11.7}$$

for all $(s,t) \in \Delta_{[0,T]}$*.*

The constants C_1, C_2 *may be taken to be* $\lambda_1(1-2^{-\varepsilon_1})^{-1}$ *and* $2C_m\lambda_2(1-2^{-\varepsilon_2})^{-1}$ *respectively.*

Moreover, for every $(s,t) \in \Delta_{[0,T]}$*, we have that*

$$\lim_{|\pi|\to 0} \sum_{[u,v]\in\pi} A_{u,v} = \gamma_t - \gamma_s,$$

where the limit exists in L^m*, and is taken over any sequence of partitions* π *of the interval* $[s,t]$ *with mesh size* $|\pi| \to 0$*.*

Proof. Step 1. Let us fix an interval $[s,t] \subseteq [0,T]$. For each $n \ge 1$, let $\pi^n = \{s = t_0^n < t_1^n < \ldots < t_{2^n}^n = t\}$ be the dyadic partition of the interval $[s,t]$, which we note has mesh size $|\pi^n| = 2^{-n}(t-s)$.

For each $n \ge 1$, we define

$$A^n_{s,t} = \sum_{i=0}^{2^n-1} A_{t_i^n, t_{i+1}^n}.$$

For each n and each i, let u_i^n be the midpoint of $[t_i^n, t_{i+1}^n]$. We then have

$$\begin{aligned} A^n_{s,t} - A^{n+1}_{s,t} &= \sum_{i=0}^{2^n-1} \delta A_{t_i^n, u_i^n, t_{i+1}^n} \\ &= \sum_{i=0}^{2^n-1} \mathbb{E}_{t_i^n} \delta A_{t_i^n, u_i^n, t_{i+1}^n} + \sum_{i=0}^{2^n-1} (\delta A_{t_i^n, u_i^n, t_{i+1}^n} - \mathbb{E}_{t_i^n} \delta A_{t_i^n, u_i^n, t_{i+1}^n}) =: I_1^n + I_2^n, \end{aligned} \tag{11.8}$$

where, by an application of the estimates (11.1) and (11.2), we have

$$\|I_1^n\|_{L^m} \le \sum_{i=0}^{2^n-1} \|\mathbb{E}_{t_i^n} \delta A_{t_i^n, u_i^n, t_{i+1}^n}\|_{L^m},$$

$$\|I_2^n\|_{L^m} \le 2C_m \bigg(\sum_{i=0}^{2^n-1} \|\delta A_{t_i^n, u_i^n, t_{i+1}^n}\|^2_{L^m} \bigg)^{\frac{1}{2}}.$$

We can then use the bounds in (11.4) and (11.5) to obtain

$$\begin{aligned} \|I_1^n\|_{L^m} &\le \lambda_1 \sum_{i=0}^{2^n-1} |t_{i+1}^n - t_i^n|^{1+\varepsilon_1} \\ &= \lambda_1 |t-s|^{1+\varepsilon_1} \sum_{i=0}^{2^n-1} 2^{-n(1+\varepsilon_1)} = \lambda_1 |t-s|^{1+\varepsilon_1} 2^{-n\varepsilon_1}, \end{aligned} \tag{11.9}$$

and

$$\begin{aligned}
\|I_2^n\|_{L^m} &\le 2C_m\lambda_2\bigg(\sum_{i=0}^{2^n-1}|t_{i+1}^n - t_i^n|^{1+2\varepsilon_2}\bigg)^{\frac{1}{2}} \\
&= 2C_m\lambda_2|t-s|^{\frac{1}{2}+\varepsilon_2}\bigg(\sum_{i=0}^{2^n-1}2^{-n(1+2\varepsilon_2)}\bigg)^{\frac{1}{2}} = 2C_m\lambda_2|t-s|^{\frac{1}{2}+\varepsilon_2}2^{-n\varepsilon_2}.
\end{aligned}$$

We thus have that

$$\|A_{s,t}^n - A_{s,t}^{n+1}\|_{L^m} \le \lambda_1|t-s|^{1+\varepsilon_1}2^{-n\varepsilon_1} + 2C_m\lambda_2|t-s|^{\frac{1}{2}+\varepsilon_2}2^{-n\varepsilon_2}.$$

It follows that $(A_{s,t}^n)_{n\ge 1}$ is a Cauchy sequence, so that the limit

$$\Gamma_{s,t} := \lim_{n\to\infty} A_{s,t}^n$$

exists in L^m, and satisfies

$$\begin{aligned}
\|\Gamma_{s,t} - A_{s,t}\|_{L^m} &\le \sum_{n=0}^{\infty}\|A_{s,t}^n - A_{s,t}^{n+1}\|_{L^m} \\
&\le \frac{\lambda_1}{1-2^{-\varepsilon_1}}|t-s|^{1+\varepsilon_1} + \frac{2C_m\lambda_2}{1-2^{-\varepsilon_2}}|t-s|^{\frac{1}{2}+\varepsilon_2}. \qquad (11.10)
\end{aligned}$$

It is clear that $\Gamma_{s,t}$ is $\mathcal{F}_t$-measurable. Moreover, since $\mathbb{E}_s I_2^n = 0$, we see from (11.8) that

$$\mathbb{E}_s(A_{s,t}^n - A_{s,t}^{n+1}) = \mathbb{E}_s I_1^n,$$

and it then follows from (11.9) that

$$\begin{aligned}
\|\mathbb{E}_s(\Gamma_{s,t} - A_{s,t})\|_{L^m} &\le \sum_{n=0}^{\infty}\|\mathbb{E}_s(A_{s,t}^n - A_{s,t}^{n+1})\|_{L^m} = \sum_{n=0}^{\infty}\|\mathbb{E}_s I_1^n\|_{L^m} \\
&\le \sum_{n=0}^{\infty}\|I_1^n\|_{L^m} \le \frac{\lambda_1}{1-2^{-\varepsilon_1}}|t-s|^{1+\varepsilon_1}. \qquad (11.11)
\end{aligned}$$

Step 2. Existence: It is straightforward to see that A^n is a continuous function from $\Delta_{[0,T]} \to L^m$, and that the convergence $A_{s,t}^n \to \Gamma_{s,t}$ (in L^m) is uniform on $\Delta_{[0,T]}$. It therefore follows that Γ is itself a continuous function from $\Delta_{[0,T]} \to L^m$.

It follows from the above construction that

$$\Gamma_{s,u} + \Gamma_{u,t} = \Gamma_{s,t} \qquad (11.12)$$

for all dyadic times $s \le u \le t$, and it then follows by continuity that (11.12) holds for *all* times $s \le u \le t$. We thus infer that Γ is really just the increments of a continuous path. That is, if we define $\gamma_t = \Gamma_{0,t}$ for all $t \in [0,T]$, then we have that

$$\gamma_t - \gamma_s = \Gamma_{s,t} \qquad \text{for all} \quad (s,t) \in \Delta_{[0,T]},$$

and in particular that $\gamma_0 = 0$. It is also clear that γ_t is $\mathcal{F}_t$-measurable, and that $\gamma\colon [0,T] \to L^m$ is continuous, so that γ satisfies condition (i). The inequalities in (11.10) and (11.11) now read

$$\|\gamma_t - \gamma_s - A_{s,t}\|_{L^m} \leq \frac{\lambda_1}{1-2^{-\varepsilon_1}}|t-s|^{1+\varepsilon_1} + \frac{2C_m\lambda_2}{1-2^{-\varepsilon_2}}|t-s|^{\frac{1}{2}+\varepsilon_2}, \tag{11.13}$$

$$\|\mathbb{E}_s(\gamma_t - \gamma_s - A_{s,t})\|_{L^m} \leq \frac{\lambda_1}{1-2^{-\varepsilon_1}}|t-s|^{1+\varepsilon_1}, \tag{11.14}$$

which imply the bounds (11.6)–(11.7), so that γ satisfies condition (ii).

Step 3. Uniqueness: Let $\bar{\gamma}$ be another $\mathcal{F}_t$-adapted process satisfying the conditions (i) and (ii), and let $\zeta = \gamma - \bar{\gamma}$. Then ζ satisfies

$$\|\zeta_t - \zeta_s\|_{L^m} \leq \tilde{C}|t-s|^{\frac{1}{2}+\tilde{\varepsilon}}, \tag{11.15}$$

$$\|\mathbb{E}_s(\zeta_t - \zeta_s)\|_{L^m} \leq \tilde{C}|t-s|^{1+\tilde{\varepsilon}} \tag{11.16}$$

for all $(s,t) \in \Delta_{[0,T]}$, for some constants $\tilde{C} \geq 0$ and $\tilde{\varepsilon} > 0$.

Fix a $t \in [0,T]$. For each integer $n \geq 1$, let $\{0 = t_0^n < t_1^n < \ldots < t_{2^n}^n = t\}$ be the dyadic partition of the interval $[0,t]$. Then

$$\zeta_t = \sum_{i=0}^{2^n-1}(\zeta_{t_{i+1}^n} - \zeta_{t_i^n}).$$

Applying the estimate in (11.3), we obtain

$$\|\zeta_t\|_{L^m} \leq \sum_{i=0}^{2^n-1}\|\mathbb{E}_{t_i^n}(\zeta_{t_{i+1}^n} - \zeta_{t_i^n})\|_{L^m} + 2C_m\bigg(\sum_{i=0}^{2^n-1}\|\zeta_{t_{i+1}^n} - \zeta_{t_i^n}\|_{L^m}^2\bigg)^{\frac{1}{2}}.$$

Using the bounds in (11.15)–(11.16), we then have that

$$\|\zeta_t\|_{L^m} \lesssim \sum_{i=0}^{2^n-1} 2^{-n(1+\tilde{\varepsilon})} + \bigg(\sum_{i=0}^{2^n-1} 2^{-n(1+2\tilde{\varepsilon})}\bigg)^{\frac{1}{2}} \lesssim 2^{-n\tilde{\varepsilon}}.$$

Letting $n \to \infty$, we have that $\|\zeta_t\|_{L^m} = 0$, so that $\gamma_t = \bar{\gamma}_t$ almost surely.

Step 4. Convergence of Riemann sums: Let $(s,t) \in \Delta_{[0,T]}$ and let $\pi = \{s = t_0 < t_1 < \cdots < t_N = t\}$ be any (not necessarily dyadic) partition of $[s,t]$. Writing

$$\gamma_t - \gamma_s - \sum_{i=0}^{N-1} A_{t_i,t_{i+1}} = \sum_{i=0}^{N-1}(\gamma_{t_{i+1}} - \gamma_{t_i} - A_{t_i,t_{i+1}})$$

and applying the estimate in (11.3), followed by those in (11.13)–(11.14), we obtain

$$\begin{aligned}
&\bigg\|\gamma_t - \gamma_s - \sum_{i=0}^{N-1} A_{t_i,t_{i+1}}\bigg\|_{L^m} \\
&\leq \sum_{i=0}^{2^n-1}\|\mathbb{E}_{t_i}(\gamma_{t_{i+1}} - \gamma_{t_i} - A_{t_i,t_{i+1}})\|_{L^m} + 2C_m\bigg(\sum_{i=0}^{2^n-1}\|\gamma_{t_{i+1}} - \gamma_{t_i} - A_{t_i,t_{i+1}}\|_{L^m}^2\bigg)^{\frac{1}{2}} \\
&\lesssim \sum_{i=0}^{N-1}|t_{i+1}-t_i|^{1+\varepsilon_1} + \bigg(\sum_{i=0}^{N-1}|t_{i+1}-t_i|^{2+2\varepsilon_1} + |t_{i+1}-t_i|^{1+2\varepsilon_2}\bigg)^{\frac{1}{2}} \\
&\lesssim |\pi|^{\varepsilon_1} + |\pi|^{\frac{1}{2}+\varepsilon_1} + |\pi|^{\varepsilon_2},
\end{aligned}$$

and we deduce that $\sum_{i=0}^{N-1} A_{t_i,t_{i+1}} \to \gamma_t - \gamma_s$ in L^m as $|\pi| \to 0$. $\square$

11.2 Relation to Itô calculus

Example 11.2. Let B be a standard Brownian motion in $\mathbb{R}^d$ with respect to a filtration $(\mathcal{F}_t)_{t\in[0,T]}$. Suppose that $f\colon \mathbb{R}^d \to \mathcal{L}(\mathbb{R}^d;\mathbb{R})$ is β-Hölder continuous for some $\beta \in (0,1]$.

Letting $A_{s,t} = f(B_s)B_{s,t}$, we have that

$$\delta A_{s,u,t} = -f(B)_{s,u}B_{u,t}.$$

Using the independence of Brownian increments, we see that, for every $m \geq 2$,

$$\begin{aligned}
\|\delta A_{s,u,t}\|_{L^m} &= \|f(B)_{s,u}\|_{L^m}\|B_{u,t}\|_{L^m} \\
&\leq \|f\|_{C_b^\beta}\|B_{s,u}\|_{L^{\beta m}}^\beta\|B_{u,t}\|_{L^m} \\
&= \|f\|_{C_b^\beta}\|B_1\|_{L^{\beta m}}^\beta|u-s|^{\frac{\beta}{2}}\|B_1\|_{L^m}|t-u|^{\frac{1}{2}} \\
&\lesssim |t-s|^{\frac{1}{2}+\frac{\beta}{2}}.
\end{aligned}$$

Moreover, it is clear that

$$\mathbb{E}_s\delta A_{s,u,t} = 0.$$

We therefore have that A satisfies (11.4) and (11.5), and we have in particular that $\lambda_1 = 0$ and $\varepsilon_2 = \frac{\beta}{2}$. Thus, by the stochastic sewing lemma (Theorem 11.1), the integral

$$\int_0^t f(B_r)\,\mathrm{d}B_r := \lim_{|\pi|\to 0}\sum_{[u,v]\in\pi} f(B_u)B_{u,v}$$

exists in L^m along any (deterministic) sequence of partitions of $[0,t]$ with vanishing mesh size. Moreover, we have that

$$\left\|\int_s^t f(B_r)\,\mathrm{d}B_r - f(B_s)B_{s,t}\right\|_{L^m} \lesssim |t-s|^{\frac{1}{2}+\frac{\beta}{2}}$$

and

$$\mathbb{E}_s\int_s^t f(B_r)\,\mathrm{d}B_r = 0.$$

In particular, the last line above shows that $\int_0^\cdot f(B_r)\,\mathrm{d}B_r$ is a martingale.

Finally, we have the additional insight that, for any given $\varepsilon > 0$, $\int_0^\cdot f(B_r)\,\mathrm{d}B_r$ is actually the unique $\mathcal{F}_t$-adapted L^m-integrable process γ, with $\gamma_0 = 0$, such that

$$\begin{aligned}
\|\gamma_t - \gamma_s - f(B_s)B_{s,t}\|_{L^m} &\lesssim |t-s|^{1+\varepsilon} + |t-s|^{\frac{1}{2}+\frac{\beta}{2}}, \\
\|\mathbb{E}_s(\gamma_t-\gamma_s)\|_{L^m} &\lesssim |t-s|^{1+\varepsilon}
\end{aligned}$$

for all $(s,t) \in \Delta_{[0,T]}$.

Example 11.3. Let M be an L^4-integrable $\mathcal{F}_t$-adapted martingale in $\mathbb{R}^d$, and assume that

$$\|M_{s,u}\otimes M_{u,t}\|_{L^2} \leq C|t-s|^{\frac{1}{2}+\varepsilon} \tag{11.17}$$

for all $s \leq u \leq t$, for some constants $C, \varepsilon > 0$. (For Brownian motion, this condition is satisfied with $\varepsilon = \frac{1}{2}$.)

Let $A_{s,t} = M_{s,t} \otimes M_{s,t}$. Then

$$\begin{aligned}\delta A_{s,u,t} &= M_{s,t} \otimes M_{s,t} - M_{s,u} \otimes M_{s,u} - M_{u,t} \otimes M_{u,t} \\ &= (M_{s,u} + M_{u,t}) \otimes (M_{s,u} + M_{u,t}) - M_{s,u} \otimes M_{s,u} - M_{u,t} \otimes M_{u,t} \\ &= M_{s,u} \otimes M_{u,t} + M_{u,t} \otimes M_{s,u}.\end{aligned}$$

We then have immediately from (11.17) that

$$\|\delta A_{s,u,t}\|_{L^2} \leq 2C|t-s|^{\frac{1}{2}+\varepsilon},$$

and we also see that

$$\mathbb{E}_s \delta A_{s,u,t} = 0.$$

Thus, by the stochastic sewing lemma (Theorem 11.1), there exists an $\mathcal{F}_t$-adapted process, which we denote by $\langle M \rangle$, such that $\langle M \rangle_0 = 0$, and such that

$$\|\langle M \rangle_{s,t} - M_{s,t} \otimes M_{s,t}\|_{L^2} \lesssim |t-s|^{\frac{1}{2}+\varepsilon}$$

and

$$\mathbb{E}_s \langle M \rangle_{s,t} = \mathbb{E}_s(M_{s,t} \otimes M_{s,t}) \tag{11.18}$$

for all $(s,t) \in \Delta_{[0,T]}$. Moreover, we have that the limit

$$\langle M \rangle_t = \lim_{|\pi| \to 0} \sum_{[u,v] \in \pi} M_{u,v} \otimes M_{u,v}$$

exists in L^2 along any (deterministic) sequence of partitions of $[0,t]$ with vanishing mesh size, from which we deduce, as the notation suggests, that $\langle M \rangle$ is indeed the quadratic variation of M.

Since $\mathbb{E}_s(M_{s,t} \otimes M_{s,t}) = \mathbb{E}_s(M_t \otimes M_t - M_s \otimes M_s)$, it follows from (11.18) that the process $M \otimes M - \langle M \rangle$ is a martingale.

11.3 A rough stochastic integral

Suppose we are given a Brownian motion B and a deterministic (rough) path X, and that we wish to define the integral of $f(B+X)$ against X. If the function f is nonlinear, then neither rough nor stochastic integration alone is able to handle such an integral. We will now see how the stochastic sewing lemma can bring both worlds together to define a suitable notion of integral.

Proposition 11.4. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,T]}, \mathbb{P})$ be a filtered probability space and let $\alpha \in (\frac{1}{3}, \frac{1}{2}]$. Let $\mathbf{X} = (X, \mathbb{X}) \in \mathscr{C}^\alpha$ be a rough path, and let B be an $\mathcal{F}_t$-adapted Brownian motion. Let $m \geq 2$ and $f \in C_b^2$. Then the limit*

$$\int_0^t f(B_r + X_r) \, \mathrm{d}\mathbf{X}_r := \lim_{|\pi| \to 0} \sum_{[u,v] \in \pi} f(B_u + X_u) X_{u,v} + Df(B_u + X_u) \mathbb{X}_{u,v}$$

exists in L^m for every $t \in [0,T]$. Moreover, $\int_0^\cdot f(B_r + X_r)\,\mathrm{d}\mathbf{X}_r$ is the unique $\mathcal{F}_t$-adapted L^m-integrable process, started at zero, such that the estimates

$$\left\| \int_s^t f(B_r + X_r)\,\mathrm{d}\mathbf{X}_r - f(B_s + X_s)X_{s,t} - Df(B_s + X_s)\mathbb{X}_{s,t} \right\|_{L^m} \lesssim |t-s|^{2\alpha}$$

and

$$\left\| \mathbb{E}_s\Big(\int_s^t f(B_r + X_r)\,\mathrm{d}\mathbf{X}_r - f(B_s + X_s)X_{s,t} - Df(B_s + X_s)\mathbb{X}_{s,t} \Big) \right\|_{L^m} \lesssim |t-s|^{3\alpha}$$

hold for all $(s,t) \in \Delta_{[0,T]}$.

Proof. Let

$$A_{s,t} = f(B_s + X_s)X_{s,t} + Df(B_s + X_s)\mathbb{X}_{s,t}.$$

Then

$$\begin{aligned}
\delta A_{s,u,t} &= A_{s,t} - A_{s,u} - A_{u,t} \\
&= f(B_s + X_s)X_{s,t} + Df(B_s + X_s)\mathbb{X}_{s,t} \\
&\quad - f(B_s + X_s)X_{s,u} - Df(B_s + X_s)\mathbb{X}_{s,u} \\
&\quad - f(B_u + X_u)X_{u,t} - Df(B_u + X_u)\mathbb{X}_{u,t} \\
&= -f(B+X)_{s,u}X_{u,t} + Df(B_s + X_s)(\mathbb{X}_{s,t} - \mathbb{X}_{s,u}) - Df(B_u + X_u)\mathbb{X}_{u,t} \\
&= -f(B+X)_{s,u}X_{u,t} + Df(B_s + X_s)(\mathbb{X}_{u,t} + X_{s,u} \otimes X_{u,t}) - Df(B_u + X_u)\mathbb{X}_{u,t} \\
&= -\big(f(B+X)_{s,u} - Df(B_s + X_s)X_{s,u}\big)X_{u,t} - Df(B+X)_{s,u}\mathbb{X}_{u,t}.
\end{aligned}$$

We have that

$$\begin{aligned}
\|f(B+X)_{s,u}\|_{L^m} &\le \|f\|_{C_b^1}\|B_{s,u} + X_{s,u}\|_{L^m} \le \|f\|_{C_b^1}\big(\|B_{s,u}\|_{L^m} + |X_{s,u}|\big) \\
&\le \|f\|_{C_b^1}\big(\|B_1\|_{L^m}|u-s|^{\frac{1}{2}} + \|X\|_\alpha |u-s|^\alpha\big) \\
&\lesssim |t-s|^\alpha
\end{aligned}$$

It then follows from the above that

$$\|\delta A_{s,u,t}\|_{L^m} \lesssim |t-s|^{2\alpha}.$$

Since $2\alpha > \frac{1}{2}$, we have that (11.5) holds.

To obtain (11.4) we need to be a bit sneakier. We note that

$$\begin{aligned}
&f(B+X)_{s,u} - Df(B_s + X_s)X_{s,u} \\
&= \int_0^1 \big(Df(B_s + X_s + r(B_{s,u} + X_{s,u})) - Df(B_s + X_s)\big)(B_{s,u} + X_{s,u})\,\mathrm{d}r \\
&\qquad + Df(B_s + X_s)B_{s,u}.
\end{aligned}$$

Noticing that the final term above vanishes upon taking an $\mathbb{E}_s$ conditional expectation, we then have

$$\begin{aligned}
\big\|\mathbb{E}_s\big(f(B+X)_{s,u} - Df(B_s + X_s)X_{s,u}\big)\big\|_{L^m} &\le \|f\|_{C_b^2}\big\||B_{s,u} + X_{s,u}|^2\big\|_{L^m} \\
&\lesssim \|B_{s,u}\|_{L^{2m}}^2 + |X_{s,u}|^2 \\
&\le \|B_1\|_{L^{2m}}^2|u-s| + \|X\|_\alpha |u-s|^{2\alpha} \\
&\lesssim |t-s|^{2\alpha}.
\end{aligned}$$

It follows that

$$\|\mathbb{E}_s \delta A_{s,u,t}\|_{L^m} \lesssim |t-s|^{3\alpha}.$$

Since $3\alpha > 1$, we have that (11.4) also holds.

By the stochastic sewing lemma (Theorem 11.1), there exists a process with the desired properties. □

For a thorough discussion and further applications of the stochastic sewing lemma, see the article [Lê20].

12 Rough paths of lower regularity

In this course we have focused on the case where the Hölder exponent α of the paths we consider are strictly greater than $\frac{1}{3}$. In particular, we saw how Young integration is adequate to deal with the case when $\alpha \in (\frac{1}{2}, 1]$, and developed a theory of rough paths capable of handling paths with Hölder exponents $\alpha \in (\frac{1}{3}, \frac{1}{2}]$. We saw in particular that this is enough to include important stochastic processes such as Brownian motion, and that we can even derive pathwise results analogous to classical results from stochastic calculus, e.g. the bracket process, Itô formula and the rough exponential.

However, the full theory of rough paths is able to handle paths of arbitrarily small Hölder exponent, i.e. any $\alpha \in (0, 1]$. In this section we shall briefly discuss rough paths of lower regularity. We will not go into much detail, but we will just try to see how the notion of rough paths we have considered fits into the more general theory.

12.1 The signature and Chen's relation

Let $X = (X^1, \dots, X^d) \colon [0, T] \to \mathbb{R}^d$ be a smooth path. Consider a *word*

$$w = w_1 w_2 \dots w_n$$

with letters in the *alphabet* $\{1, 2, \dots, d\}$, i.e. such that $w_i \in \{1, 2, \dots, d\}$ for each $i = 1, \dots, n$. We write $|w| = n \geq 0$ for the length of the word w.

In this course we considered the rough path enhancement as the integral of X against itself, i.e. a matrix consisting of elements of the form

$$\int_s^{r_3} X^{w_1}_{s,r_2} \, \mathrm{d}X^{w_2}_{r_2} = \int_s^{r_3} \int_s^{r_2} \mathrm{d}X^{w_1}_{r_1} \, \mathrm{d}X^{w_2}_{r_2}.$$

Integrating with respect to X^{w_3}, we then obtain $\int_s^{r_4} \int_s^{r_3} \int_s^{r_2} \mathrm{d}X^{w_1}_{r_1} \, \mathrm{d}X^{w_2}_{r_2} \, \mathrm{d}X^{w_3}_{r_3}$. After n such integrals, we arrive at the n-fold iterated integral

$$X^w_{s,t} := \int_s^t \int_s^{r_n} \cdots \int_s^{r_3} \int_s^{r_2} \mathrm{d}X^{w_1}_{r_1} \, \mathrm{d}X^{w_2}_{r_2} \cdots \mathrm{d}X^{w_{n-1}}_{r_{n-1}} \, \mathrm{d}X^{w_n}_{r_n}.$$

As a shorthand, let's write

$$\Delta^n_{[s,t]} = \{s < r_1 < r_2 < \cdots < r_n < t\}$$

for the n-dimensional simplex over the interval $[s,t]$. We can then rewrite the above as

$$X^w_{s,t} = \int_{\Delta^n_{[s,t]}} \mathrm{d}X^{w_1}_{r_1} \cdots \mathrm{d}X^{w_n}_{r_n}. \tag{12.1}$$

The family of all such iterated integrals corresponding to words of length n is given by

$$\int_{\Delta^n_{[s,t]}} \mathrm{d}X_{r_1} \otimes \cdots \otimes \mathrm{d}X_{r_n} \in (\mathbb{R}^d)^{\otimes n}.$$

The collection of all iterated integrals up to arbitrary length, which we can write as

$$S(X)_{s,t} := \left(1, \int_s^t \mathrm{d}X, \int_{\Delta^2_{[s,t]}} \mathrm{d}X \otimes \mathrm{d}X, \dots, \int_{\Delta^n_{[s,t]}} \mathrm{d}X \otimes \cdots \otimes \mathrm{d}X, \dots \right),$$

is called the *signature* of X over the interval $[s,t]$, and takes values in the tensor algebra $T(\mathbb{R}^d) = \bigoplus_{n=0}^{\infty} (\mathbb{R}^d)^{\otimes n}$.

Let us now fix $s \le u \le t$. We observe that, ignoring sets of zero n-dimensional Lebesgue measure, we have that

$$\Delta^n_{[s,t]} = A^n_0 \cup A^n_1 \cup \cdots \cup A^n_n$$

where, for $0 \le j \le n$,

$$\begin{aligned} A^n_j &:= \{s < r_1 < \cdots < r_j < u < r_{j+1} < \cdots < r_n < t\} \\ &= \Delta^j_{[s,u]} \times \Delta^{n-j}_{[u,t]} \end{aligned}$$

That is, the domain of integration may be decomposed, depending on how many variables of integration are to the left of u.

Given a word $w = w_1 w_2 \dots w_n$ of length $|w| = n \ge 1$, we have

$$\begin{aligned} X^w_{s,t} &= \int_{\Delta^n_{[s,t]}} \mathrm{d}X^{w_1}_{r_1} \cdots \mathrm{d}X^{w_n}_{r_n} \\ &= \sum_{j=0}^{n} \int_{A^n_j} \mathrm{d}X^{w_1}_{r_1} \cdots \mathrm{d}X^{w_n}_{r_n} \\ &= \sum_{j=0}^{n} \int_{\Delta^j_{[s,u]}} \mathrm{d}X^{w_1}_{r_1} \cdots \mathrm{d}X^{w_j}_{r_j} \int_{\Delta^{n-j}_{[u,t]}} \mathrm{d}X^{w_{j+1}}_{r_{j+1}} \cdots \mathrm{d}X^{w_n}_{r_n} \\ &= \sum_{j=0}^{n} X^{w_1 \dots w_j}_{s,u} X^{w_{j+1} \dots w_n}_{u,t}. \end{aligned}$$

This motivates a definition. Let us define $X^{\emptyset}_{s,t} = 1$ for the empty word $w = \emptyset$ (of length 0). Then, given two collections of elements $\mathbf{A} = \{A^w : |w| \ge 0\}$ and $\mathbf{B} = \{B^w : |w| \ge 0\}$ indexed by words from the alphabet $\{1, 2, \dots, d\}$, we define a product $\star$ such that, for each word $w = w_1 \dots w_{|w|}$,

$$(\mathbf{A} \star \mathbf{B})^w = \sum_{j=0}^{|w|} A^{w_1 \dots w_j} B^{w_{j+1} \dots w_{|w|}}. \tag{12.2}$$

It then follows from the above that, writing $\mathbf{X}_{s,t} = \{X^w_{s,t} : |w| \geq 0\}$, we have

$$\mathbf{X}_{s,t} = \mathbf{X}_{s,u} \star \mathbf{X}_{u,t} \tag{12.3}$$

for all $s < u < t$. The equality in (12.3) is a general version of Chen's relation.

For example, let's take a word $w = ij$ with length $|w| = 2$. Then (12.2) and (12.3) imply that

$$X^{ij}_{s,t} = X^{ij}_{s,u} + X^{ij}_{u,t} + X^i_{s,u} X^j_{u,t},$$

which is precisely the equality in (3.2), i.e. the version of Chen's relation we have been using throughout this course.

It follows from (12.3) that the signature of X also satisfies its own version of Chen's relation:

$$S(X)_{s,u} \otimes S(X)_{u,t} = S(X)_{s,t}$$

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

12.2 General rough paths

For $N \in \mathbb{N}$, we consider maps of the form

$$\begin{aligned} \Delta_{[0,T]} \ni (s,t) \longmapsto \mathbf{X}_{s,t} &= (X^w_{s,t} : 0 \leq |w| \leq N) \\ &= (1, X^{(1)}_{s,t}, X^{(2)}_{s,t}, \dots, X^{(N)}_{s,t}), \end{aligned}$$

which takes values in the truncated tensor algebra $T^N(\mathbb{R}^d) = \bigoplus_{n=0}^N (\mathbb{R}^d)^{\otimes n}$.

Here, $X^{(n)}_{s,t} = (X^w_{s,t} : |w| = n) \in (\mathbb{R}^d)^{\otimes n}$ is the collection of all elements $X^w_{s,t}$ corresponding to words $w = w_1 w_2 \dots w_n$ of length $|w| = n$.

We can define a seminorm on the space of such maps by

$$|||\mathbf{X}|||_{(\alpha,N)} = \max_{j=1,\dots,N} \|X^{(j)}\|^{\frac{1}{j}}_{j\alpha},$$

where $\|\cdot\|_\alpha$ denotes the usual α-Hölder seminorm.

We can also define an associated distance similarly:

$$\|\mathbf{X}; \tilde{\mathbf{X}}\|_{(\alpha,N)} = \max_{j=1,\dots,N} \|X^{(j)} - \tilde{X}^{(j)}\|^{\frac{1}{j}}_{j\alpha}. \tag{12.4}$$

Definition 12.1. Let $\alpha \in (0,1]$, and let $N = \lfloor \frac{1}{\alpha} \rfloor$, i.e. the integer $N \in \mathbb{N}$ such that $\alpha N \leq 1 < \alpha(N+1)$. We say that a map $(s,t) \mapsto \mathbf{X}_{s,t}$ as above is an *α-Hölder rough path* over $\mathbb{R}^d$, if

- $|||\mathbf{X}|||_{(\alpha,N)} < \infty$,
- and Chen's relation $\mathbf{X}_{s,t} = \mathbf{X}_{s,u} \star \mathbf{X}_{u,t}$ holds for all $0 \leq s \leq u \leq t \leq T$.

Note that the condition that $|||\mathbf{X}|||_{(\alpha,N)} < \infty$ is equivalent to requiring that there exists a constant $C > 0$ such that

$$|X^w_{s,t}| \leq C|t-s|^{\alpha|w|}$$

for all $(s,t) \in \Delta_{[0,T]}$ and all words w with length $|w| \leq N$.

Thus, a "level 1" rough path (with $N = 1$) is simply an α-Hölder continuous path X for some $\alpha \in (\frac{1}{2}, 1]$, which we know fits into the Young regime. Moreover, a "level 2" rough path (with $N = 2$) is precisely the space of rough paths we have considered throughout this course, and is appropriate for paths with regularity $\alpha \in (\frac{1}{3}, \frac{1}{2}]$.

Definition 12.2. We define the space of *geometric α-Hölder rough paths* as the closure of level N canonical lifts of smooth paths with respect to the rough path distance in (12.4). By the level N canonical lift of a smooth path $X\colon [0,T] \to \mathbb{R}^d$, we mean the rough path $\mathbf{X} = (X^w : |w| \leq N)$, where X^w is defined as in (12.1).

The *shuffle product* $\sqcup\!\sqcup$ is a product on pairs of words which, heuristically, gives the sum of all ways of interlacing them. It can be defined inductively by the relations:

$$w \sqcup\!\sqcup \emptyset = \emptyset \sqcup\!\sqcup w = w$$

and

$$vi \sqcup\!\sqcup wj = (v \sqcup\!\sqcup wj)i + (vi \sqcup\!\sqcup w)j,$$

where $\emptyset$ is the empty word, i and j are single letters, and v and w are arbitrary words.

The name "shuffle product" refers to the fact that the product can be thought of as the sum over all the different ways of riffle shuffling two words (or, indeed, the two halves of a deck of playing cards) together. For example, for letters i, j, k, we have that

$$ki \sqcup\!\sqcup j = kji + jki + kij,$$

which is the sum of the ways of arranging the three letters such that the letter k always appears before the letter i.

Definition 12.3. We define a *weakly geometric α-Hölder rough path* as a rough path which additionally satisfies the shuffle identity:

$$X^v_{s,t} X^w_{s,t} = X^{v \sqcup\!\sqcup w}_{s,t} \tag{12.5}$$

for all times $(s,t) \in \Delta_{[0,T]}$ and all words v and w with combined length $|v| + |w| \leq N$.

The purpose of the shuffle identity (12.5) is to impose the chain rule. As an example, suppose that $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ and $N = 2$. For letters i, j, we have that $i \sqcup\!\sqcup j = ij + ji$, so that the shuffle identity gives

$$X^i_{s,t} X^j_{s,t} = X^{i \sqcup\!\sqcup j}_{s,t} = X^{ij}_{s,t} + X^{ji}_{s,t}.$$

Equivalently,

$$\frac{1}{2} X^{(1)}_{s,t} \otimes X^{(1)}_{s,t} = \mathrm{Sym}(X^{(2)}_{s,t}),$$

which is precisely the condition (3.3) for a level 2 rough path to be weakly geometric.

References

[AC20] Andrew L. Allan and Samuel N. Cohen. Pathwise stochastic control with applications to robust filtering. *Ann. Appl. Probab.*, 30(5):2274–2310, 2020.

[BC09] Alan Bain and Dan Crisan. *Fundamentals of Stochastic Filtering.* Springer, New York, 2009.

[CE15] Samuel N. Cohen and Robert J. Elliott. *Stochastic Calculus and Applications.* Springer, New York, 2nd edition, 2015.

[DFM16] Joscha Diehl, Peter K. Friz, and Hilmar Mai. Pathwise stability of likelihood estimators for diffusions via rough paths. *Ann. Appl. Probab.*, 26(4):2169–2192, 2016.

[FH20] Peter K. Friz and Martin Hairer. *A Course on Rough Paths, With an Introduction to Regularity Structures.* Springer, 2nd edition, 2020.

[FV10] Peter K. Friz and Nicolas B. Victoir. *Multidimensional Stochastic Processes as Rough Paths.* Cambridge University Press, 2010.

[Lê20] Khoa Lê. A stochastic sewing lemma and applications. *Electron. J. Probab.*, 25(38):1–55, 2020.

[Str11] D. W. Stroock. *Essentials of Integration Theory for Analysis.* Springer, New York, 2011.