A panorama of rough analysis*

by Nicolas Perkowski

Freie Universität Berlin

July 16, 2024

Abstract

These are lecture notes for a mini course on rough analysis taught at the 9th Regional Summer School on Applied Mathematics in Sinaia in 2024. The material is too much to cover during the school, and the concrete focus will depend on the students' interests. The notes are work in progress and not in polished, final form.

Table of contents

*. This article has been written using GNU TeXmacs [H+98].

Introduction

Let us discuss a few model problems that will guide us through the lecture:

Example 1. (SDEs driven by fractional Brownian motion) Fractional Brownian motion is a class of self-similar Gaussian processes introduced by Mandelbrot and van Ness [MvN68] to model natural time series. Apart from the self-similarity, an important feature are the long-range correlations.

A continuous and centered Gaussian process $(B_t)_{t \geqslant 0}$ with $B_0 = 0$ is called a fractional Brownian motion with Hurst parameter $H \in (0, 1)$ if it has the covariance

$$\mathbb{E}[B_s B_t] = \Gamma(s, t) := \frac{1}{2}(s^{2H} + t^{2H} - |t - s|^{2H}).$$

It is not entirely trivial to see that Γ is indeed a covariance function (i.e. positive definite), but one can show that

$$\Gamma(s, t) = \int_{\mathbb{R}} \Phi(s, r) \Phi(t, r) dr,$$

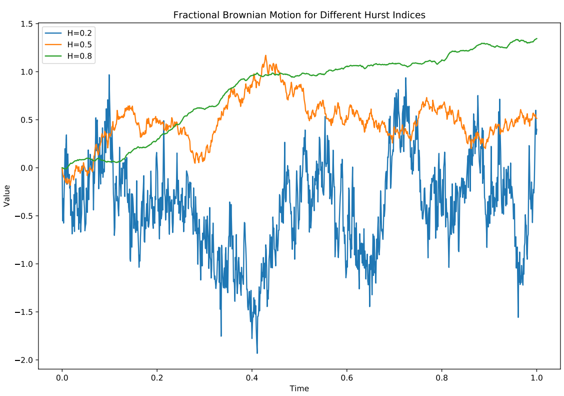
for

$$\Phi(s, r) = \frac{1}{\gamma(H + 1/2)} \big((s - r)_+^{H - 1/2} - (-r)_+^{H - 1/2} \big),$$

and from that representation we easily obtain that Γ is indeed positive definite. Fractional Brownian motion is neither a semimartingale nor a Markov process unless $H = \frac{1}{2}$, for which the covariance becomes $\Gamma(s, t) = s \wedge t$ and thus B is a Brownian motion.

Here is a simulation of the fractional Brownian motion, provided by ChatGPT:

```
>>> import numpy as np
    import matplotlib.pyplot as plt

    def generate_fbm(n, H):
        #Generate fractional Brownian motion using the Cholesky method
        t = np.linspace(0, 1, n+1)
        R = np.zeros((n+1, n+1))
        for i in range(n+1):
            for j in range(n+1):
                R[i, j] = 0.5 * (t[i]**(2*H) + t[j]**(2*H) - abs(t[i] -
    t[j])**(2*H))
        try:
            L = np.linalg.cholesky(R)
        except np.linalg.LinAlgError:
            # Add a small value to the diagonal for numerical stability
            R += np.eye(n+1) * 1e-10
            L = np.linalg.cholesky(R)
        W = np.random.normal(size=n+1)
        fBm = np.dot(L, W)
        return t, fBm

    # Set parameters
    n = 1000  # Number of time steps
    hurst_indices = [0.2, 0.5, 0.8]

    plt.figure(figsize=(12, 8))
    # Generate and plot fractional Brownian motion for each Hurst index
    for H in hurst_indices:
        t, fBm_values = generate_fbm(n, H)
        plt.plot(t, fBm_values, label=f'H={H}')
    plt.title('Fractional␣Brownian␣Motion␣for␣Different␣Hurst␣Indices')
    plt.xlabel('Time')
    plt.ylabel('Value')
    plt.legend()
    #The next command is only needed to run the code in Texmacs.
    #Otherwise replace it by plt.show()
    pdf_out(plt.gcf())
```



>>>

We see that the roughness of the trajectory increases as H decreases. We would like to solve (multidimensional!) stochastic differential equations with fractional Brownian noise,

$$\mathrm{d}Y_t = b(Y_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}B_t, \qquad Y_0 = x.$$

The paths of B are almost surely not of finite variation, and therefore the meaning of the integral $\int_0^t \sigma(Y_s)\mathrm{d}B_s$ is unclear. If B is a Brownian motion $\left(H = \frac{1}{2}\right)$, then of course we can interpret the integral by Itô calculus or as a Stratonovich integral. But for $H \neq \frac{1}{2}$, B is no semimartingale and therefore we cannot use Itô calculus. We thus require an integration theory that allows to make sense of $\int_0^t \sigma(Y_s)\mathrm{d}B_s$ for rough processes B that are not semimartingales (or Dirichlet processes, if you know what that is).

Example 2. (Homogenization) Consider a sufficiently chaotic dynamical system (deterministic) of the form

$$\dot{X} = f(X), \qquad X(0) \sim \mu,$$

where μ is an invariant measure (i.e. $X(t) \sim \mu$ for all $t \geqslant 0$), and let

$$\dot{Y}^{\varepsilon}(t) = b(Y^{\varepsilon}(t), X(t/\varepsilon)), \qquad Y^{\varepsilon}(0) = x_0,$$

i.e. the slow dynamics of Y^{ε} depend on the fast, chaotic dynamics $X(t/\varepsilon)$. We could for example think of Y^{ε} as ocean temperatures and $X(t/\varepsilon)$ as air temperatures, in different places across earth. On the scale in which ocean temperatures change, the air temperature fluctuates very rapidly. Under suitable assumptions we have $\lim_{\varepsilon \to 0} Y^{\varepsilon} = Y^0$, where

$$\dot{Y}^0(t) = \bar{b}(Y^0(t)), \qquad Y^0(0) = x_0,$$

with

$$\bar{b}(y) = \int b(y, x)\mu(\mathrm{d}x).$$

If $\bar{b} \equiv 0$, we can consider longer time scales by going to $Z^{\varepsilon} = Y^{\varepsilon}(t/\sqrt{\varepsilon})$. Assume from now on for simplicity[1] that $b(y, x) = b(y)x$ (abuse of notation) is linear in x. Then we could consider the case of constant $b(x) \equiv b$ first, for which

$$Z^{\varepsilon}(t) = x_0 + b\frac{1}{\sqrt{\varepsilon}}\int_0^t X(s/\varepsilon)\mathrm{d}s.$$

If X is sufficiently mixing, we can show that the integral on the right hand side converges to a Brownian motion B [Liv96]. What if b is not constant? In that case we might expect that $Z^{\varepsilon} \to Z^0$ solving

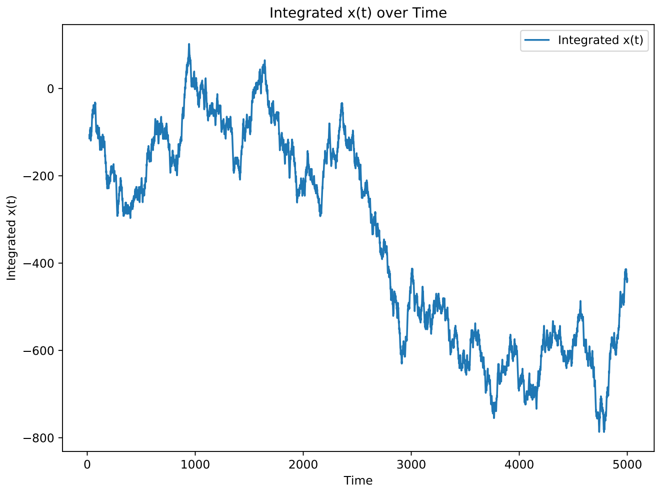
$$Z^0(t) = x_0 + \int_0^t b(Z_s^0)\mathrm{d}B_s.$$

In certain situations this is nearly correct, although there will typically be a correction in the drift [KM17]. But how can we prove this? Since Z^{ε} is purely deterministic up the randomness in the initial condition for X, we cannot make use of Itô calculus for this problem. Therefore, we would like a theory which "lifts" the convergence $\frac{1}{\sqrt{\varepsilon}}\int_0^{\cdot} X(s/\varepsilon)\mathrm{d}s \to B$ to the convergence $Z^{\varepsilon} \to Z^0$.

Simulation of $\int_0^t X_1(s)\mathrm{d}s$ for $t \gg 1$, where $X = (X_1, X_2, X_3)$ is the Lorenz 63 system, a well known chaotic dynamical system. Simulation provided by ChatGPT.

1. The techniques we develop allow to deal with general $b(x, y)$ by considering "infinite-dimensional rough paths".

```
>>> import numpy as np
    import matplotlib.pyplot as plt
    from scipy.integrate import solve_ivp

    # Parameters for the Lorenz system
    sigma = 10
    rho = 28
    beta = 8/3

    # Lorenz system equations
    def lorenz(t, state):
        x, y, z = state
        dxdt = sigma * (y - x)
        dydt = x * (rho - z) - y
        dzdt = x * y - beta * z
        return [dxdt, dydt, dzdt]

    # Time settings
    t_start = 0
    t_end = 5000
    dt = 0.01
    t = np.arange(t_start, t_end, dt)
    transient_cutoff = int(20 / dt)  # First 20 time steps to omit

    # Initial conditions
    initial_state = [0.1, 0.1, 0.1]

    # Solve the system
    solution = solve_ivp(lorenz, [t_start, t_end], initial_state, t_eval=t,
    method='RK45')

    # Extract the x-component
    x = solution.y[0]

    # Integrate x(t) over time
    integrated_x = np.cumsum(x) * dt

    # Omit the first 20 time steps
    t = t[transient_cutoff:]
    integrated_x = integrated_x[transient_cutoff:]

    # Plot integrated x(t)
    plt.figure(figsize=(8, 6))
    plt.plot(t, integrated_x, label='Integrated x(t)')
    plt.title('Integrated x(t) over Time')
    plt.xlabel('Time')
    plt.ylabel('Integrated x(t)')
    plt.legend()
    plt.tight_layout()
    pdf_out(plt.gcf())
```



>>>

Observe the similarity to the plot of Brownain motion $\left(H=\frac{1}{2}\right)$ in the first example, up to the scale on x- and y-axis, which we did not rescale here.

Example 3. (Machine learning with time series data) The goal in machine learning is to learn a function F from given data points

$$(X_i, F(X_i))_{i=1,\ldots,n},$$

or possibly only given noisy observations, say $F(X_i)+\xi_i$. If the data points X_i consist of time series, we could enhance them to paths in a suitable way and then make the ansatz

$$F(X)=\Phi_\theta(y, X),$$

where $\Phi_\theta(y, S(X))$ is the solution Y at time 1 of the controlled differential equation

$$Y_t=y+\int_0^t f_\theta(Y_s)\mathrm{d}X_s,$$

where f_θ is a nonlinear function that is parametrized by a neural network and that can be learned. We optimize f_θ by (stochastic) gradient descent methods, with the goal of minimizing some loss $\sum_{i=1}^n \ell(\Phi_\theta(y, S(X_i)), Y_i)$.

Example 4. (Interface growth) In the pictures below you see different growing interfaces. In a 1986 landmark paper in physics by Kardar, Parisi and Zhang [KPZ86] it was conjectured that the fluctuations in such interface grofwth can, in a certain regime, be modelled by an SPDE which now is called the KPZ equation: $h:\mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}$,

$$\partial_t h=\Delta h+|\nabla h|^2+\xi,$$

where ξ is a space-time white noise, i.e. a centered generalized Gaussian process with

$\mathbb{E}[\xi(f)\xi(g)] = \langle f, g\rangle_{L^2(\mathbb{R}_+\times\mathbb{R}^d)}$. This is a singular SPDE, because the noise makes the solution irregular and only in $d=1$ it is even a function, in higher dimensions it could only be a generalized function (Schwartz distribution)[2]. And even in $d=1$, which corresponds to the pictures below (two-dimensional phases, one-dimensional interface), h is non-smooth and $x\mapsto h(t,x)$ is only as regular as a Brownian motion and therefore $|\nabla h|^2$ makes no sense. Itô theory does not help with this problem, because the singularity appears in the space variable and there is no useful flow of information (filtration).

In that case there is a simple trick to make sense of h: If we define $w=e^h$ ("Cole-Hopf transformation"), then w formally solves the stochastic heat equation

$$\partial_t w = \Delta w + w\xi,$$

which is linear and well-posed as an Itô SPDE. Therefore, we can simply define $h := \log w$ (luckily w is strictly positive for positive initial conditions) and this gives us the right object to work with. But in this way we do not get an equation for h. We will find a way of dealing directly with the nonlinearity $|\nabla h|^2$ and to solve an equation for h. This has the advantage that it directly generalizes to other situations where the Cole-Hopf transformation is not available, and also for certain problems (for example scaling limits, nonlinear coercive bounds) it is more useful to work with h than with e^h.

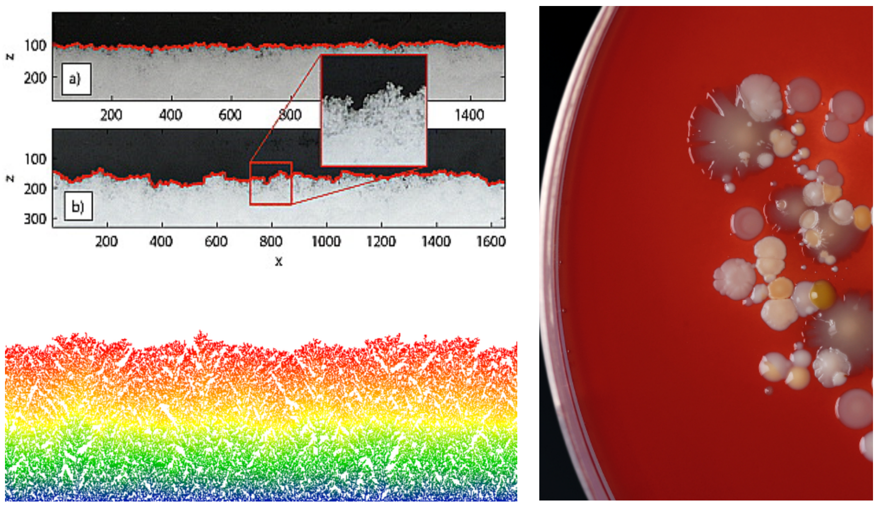


Figure 1. Growing interfaces. Image credit: Löwe et al., Geophys. Res. Letters, Vol. 34, L21507, 2007 (upper left), Nils Berglund (lower left), iStockphoto.com/rudigobbo (right)

Example 5. (Branching population in random environment) SPDEs also arise as scaling limits of population models: Consider independent continuous time random walkers on $\mathbb{Z}^d$, which can branch into new particles or die, according to a random landscape. More precisely, let $(\eta(x))_{x\in\mathbb{Z}^d}$ be an i.i.d. family of centered random variables with sufficiently many moments. If a particle is at site x and $\eta(x)$ is positive, then we interpret this as a

2. It is expected/in some sense shown that in $d\geqslant 3$ there is no nontrivial solution to the SPDE, and "h is Gaussian and a solution of $\partial_t h = \nu\Delta h + \sigma\xi$ for some effective parameters $\nu,\sigma>0$". The physically most relevant case $d=2$ (three-dimensional phases) is more subtle and finer details of the equation should determine whether solutions are Gaussian or not.

favorable environment and the particle can reproduce with rate $\eta(x)^+ = \max\{\eta(x), 0\}$. If the reproduction event happens, then the particle splits into two new independent particles, which both follow the same dynamics as the other particles. If however $\eta(x) < 0$, then the environment is unfavorable and the particle is being killed with rate $\eta(x)^-$. On large scales and for large numbers of particles this system approaches the *parabolic Anderson model*

$$\partial_t u = \frac{1}{2}\Delta u + u\xi,$$

on $\mathbb{R}_+ \times \mathbb{R}^d$, where ξ is a *space white noise* and independent of time. The issue with this equation is that ξ is a Schwartz distribution of low regularity, and u also does not have much regularity and in particular the product $u\xi$ is ill-posed in dimension $d \geqslant 2$.

Maybe also discuss Anderson Hamiltonian.

Example 6. (Regularization by noise) Consider the SDE

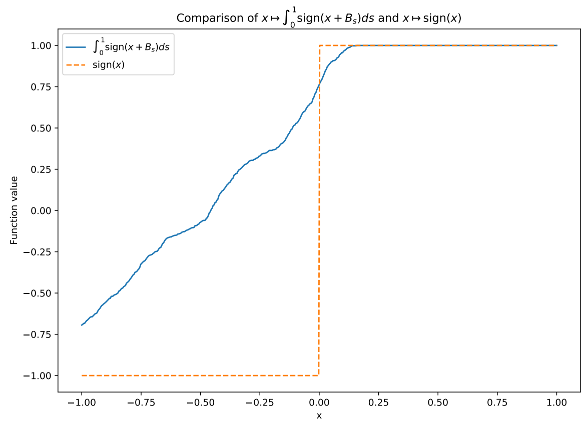
$$\mathrm{d}Y_t = b(Y_t)\mathrm{d}t + \varepsilon \mathrm{d}B_t, \qquad Y_0 = x,$$

where again B is a fractional Brownian motion. Since the noise is additive ($\sigma(x) = \varepsilon \cdot \mathrm{Id}$ in the notation of the first example), there are no issues with interpreting this equation if b is a measurable function. For $\varepsilon = 0$ we know from Cauchy-Lipschitz theory of ODEs that essentially we need Lipschitz continuous b to solve this equation[3]. But if $H = \frac{1}{2}$, i.e. B is Brownian motion, then this SDE is (strongly) well-posed even if b is only bounded and measurable [Zvo74, Ver81]. This is because B regularizes b, for example $x \mapsto \int_0^t b(x + B_s)\mathrm{d}s$ has higher regularity than b because of the oscillatory nature of B which means that

3. For divergence free b, $\operatorname{div} b = 0$, the theory of renormalized solutions [DL89] provides a weaker form of well-posedness if for example $b \in W^{1,1}$.

the path spends little time at singularities of b. Here is an illustration for $d=1$ and $b(x)=\operatorname{sgn}(x)$, provided by ChatGPT:

```
>>> # Correcting the LaTeX syntax for matplotlib
    import numpy as np
    import matplotlib.pyplot as plt

    # Define the integrand
    def f(x, B_t):
        return np.mean(np.sign(x + B_t))

    # Generate a fixed sample of Brownian motion on [0, 1]
    t = np.linspace(0, 1, 1000)
    B_t = np.cumsum(np.sqrt(1/1000) * np.random.randn(1000))

    # Compute the values for the function x -> integral(sign(x + B_s) ds)
    x_values = np.linspace(-1, 1, 500)
    f_values = [f(x, B_t) for x in x_values]

    # Compute the values for the function x -> sign(x)
    sign_values = np.sign(x_values)

    # Plot the results
    plt.figure(figsize=(10, 7))
    plt.plot(x_values, f_values,
    label=r'$\int_0^1␣\mathrm{sign}(x+B_s)␣ds$')
    plt.plot(x_values, sign_values, label=r'$\mathrm{sign}(x)$',
    linestyle='--')
    plt.title(r'Comparison␣of␣$x␣\mapsto␣\int_0^1␣\mathrm{sign}(x+B_s)␣ds$␣and␣$x␣\mapsto␣
    plt.xlabel('x')
    plt.ylabel('Function␣value')
    plt.legend()
    pdf_out(plt.gcf())
```




```
>>>
```

See also this picture by Flandoli:

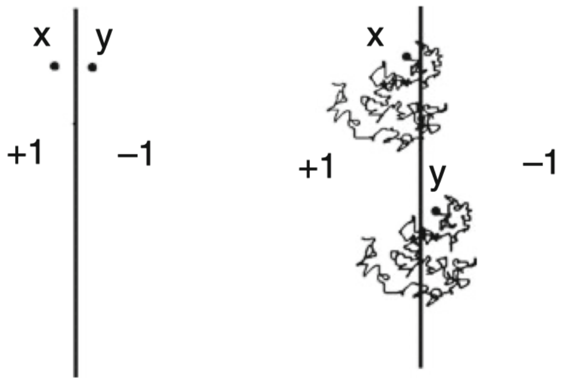


Figure 2. Regularizing effect of Brownian motion. By Flandoli [Fla11].

Since fractional Brownian motion for $H<\frac{1}{2}$ is even more oscillatory, we would hope to see even better regularization properties, and maybe even allow generalized functions b, such as a Dirac delta or a white noise. But the proofs in [Zvo74, Ver81] are based on the relation between Brownian motion and heat equation, and they completely break down in the fractional case $H\neq\frac{1}{2}$.

In this course we will learn a set of tools with a common "philosophy" that will allow us to solve all of the above and many other problems.

In Examples 1, 2, 4, 5 the common problem is a lack of regularity. This is somewhat similar to the problem that we have when constructing the Itô integral: Since the Brownian motion is not of finite variation, it is not immediately clear how to make sense of $\int_0^t H_s \mathrm{d}B_s$. But then we can use the flow of information generated by the underlying filtration together with the martingale structure of B to make sense of the integral for adapted H. But in our examples martingale arguments seem quite hopeless. Instead, we will introduce regularity based pathwise arguments to make sense of the first two examples. For that purpose we have to go what we could do with classical analysis tools and find new arguments and techniques to deal with very irregular equations (because we want to study very irregular noise), which requires us to shift perspective on what is an oscillatory path. To learn the philosophy of the pathwise approach, we first focus on stochastic ordinary equations and Lyons's rough path theory [Lyo98]. Then we will discuss ramifications and applications of our new perspective that go well beyond making sense of and studying convergence results for differential equations with irregular noise, and more precisely we will cover Example 3. Then we will briefly discuss extensions from paths functions of multidimensional variables, giving a glimpose into Hairer's theory of regularity structures [Hai14]. We focus on regularity structures due to their similarity with the rough path approach; the alternative paracontrolled approach [GIP15] may be easier to learn at first, but would require us to introduce too many tools from Fourier analysis. Finally, we will see that the main workhorse of rough path theory, the sewing lemma, has a stochastic extension due to Lê [Lê20] which allows us, among many other applications, to see the regularizing effect of fractional Brownian motion.

The material in these notes is too much to cover during the school and, depending on the audience's wishes, we can focus on certain aspects and discuss other parts more briefly.

Conventions and notation Throughout these notes, $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ is a filtered probability space satisfying the usual conditions.

We write $a\lesssim b$ if there exists a constant $C>0$, independent of a and b, such that $a\leqslant Cb$. Similarly for $\gtrsim$ and $\simeq$. For example, it follows from Hölder's inequality that $(x+y)^p\leqslant 2^{p-1}(x^p+y^p)$ for $p\geqslant 1$ and $x,y\geqslant 0$, so we would write $(x+y)^p\lesssim x^p+y^p$. If we want to stress that the implicit constant depends on one of the (unimportant) variables, we denote it with a subscript. For example $(x+y)^p \lesssim_p x^p + y^p$.

$\lfloor x \rfloor = \max\{k \in \mathbb{Z}: k \leqslant x\}$.

Multi-index notation: $\mu \in \mathbb{N}_0^d$, then $\partial^\mu = \partial_1^{\mu_1} \cdots \partial_d^{\mu_d}$ and $x^\mu = x_1^{\mu_1} \cdot \ldots \cdot x_d^{\mu_d}$ and $|\mu| = \mu_1 + \cdots + \mu_d$ and $\mu! = \mu_1! \cdot \ldots \cdot \mu_d!$.

1 The sewing lemma and Young integration

We start with the simpler example of the stochastic differential equation

$$\mathrm{d}Y_t = b(Y_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}B_t, \qquad Y_0 = x, \tag{1.1}$$

where B is a fractional Brownian motion of Hurst index $H > \frac{1}{2}$. Let us first derive the Hölder regularity of the fractional Brownian motion:

Lemma 1.1. *If B is a fractional Brownian motion of Hurst index H, then almost surely $(B_t)_{t \in [0,T]} \in C^\alpha([0,T])$ for all $\alpha \in (0,H)$ and all $T > 0$, where*

$$C^\alpha([0,T]) = C^\alpha([0,T],\mathbb{R}) = \left\{ f \in C([0,T]) : \|f\|_\alpha = \sup_{0 \leqslant s < t \leqslant T} \frac{|f(t) - f(s)|}{|t-s|^\alpha} < \infty \right\}$$

is the space of α-Hölder continuous functions.

Proof. We have

$$\begin{aligned} \mathbb{E}[|B_t - B_s|^2] &= \Gamma(t,t) + \Gamma(s,s) - 2\Gamma(s,t) \\ &= t^{2H} + s^{2H} - (s^{2H} + t^{2H} - |t-s|^{2H}) \\ &= |t-s|^{2H}. \end{aligned}$$

Since $B_t - B_s$ is Gaussian, we get for $Z \sim \mathcal{N}(0,1)$ and $p > 0$

$$\begin{aligned} \mathbb{E}[|B_t - B_s|^p]^{1/p} &= \mathbb{E}[||t-s|^H Z|^p]^{1/p} \\ &= |t-s|^H \mathbb{E}[|Z|^p]^{1/p} \\ &\simeq |t-s|^H. \end{aligned}$$

For $p > 1/H$ we thus obtain from Kolmogorov's continuity criterion [KS91, RY99] that B has an α-Hölder continuous modification for any $\alpha \in (0, H - 1/p)$. Since B itself is continuous, it is indistinguishable from this modification, and thus B is a.s. α-Hölder continuous for any $\alpha \in (0, H - 1/p)$. Since $p > 0$ is arbitrary, the claim follows. □

The Young integral [You36] provides a way of defining integrals

$$\int_0^t X_s \mathrm{d}Z_s$$

for X, Z of infinite variation, as long as X and Z have "compatible regularities". For $H > \frac{1}{2}$ we will use the Young integral to solve SDEs driven by fractional Brownian motion [Lyo94].

We will also see that for $H \leqslant 1/2$ the Young approach no longer works and in that case we will use rough path integration to solve the SDE (at least if $H > 1/4$). Both rough path integration and Young integration rely on the following fundamental and beautiful result:

Theorem 1.2. (Sewing lemma) *Let $(\mathcal{X}, \|\cdot\|)$ be a Banach space, let $T > 0$ and consider the simplex*

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

and let $\Xi: \Delta_T \to \mathcal{X}$ be continuous[1.1] *and such that $\Xi_{t,t} = 0$ for all $t \in [0,T]$. We assume that there exist constants $C, \varepsilon > 0$ such that for all $0 \leqslant s < u < t \leqslant T$:*

$$\|\delta \Xi_{s,u,t}\| := \|\Xi_{s,t} - \Xi_{s,u} - \Xi_{u,t}\| \leqslant C|t-s|^{1+\varepsilon}.$$

Then, there exists a unique function $\mathcal{I}\Xi: [0,T] \to \mathcal{X}$ with $\mathcal{I}\Xi_0 = 0$ and such that

$$\|\mathcal{I}\Xi_{s,t} - \Xi_{s,t}\| \lesssim C|t-s|^{1+\varepsilon}, \qquad 0 \leqslant s < t \leqslant T, \tag{1.2}$$

where the notation $\mathcal{I}\Xi_{s,t}$ is introduced below. Moreover, $\mathcal{I}\Xi$ is continuous and

$$\mathcal{I}\Xi_t = \lim_{n\to\infty} \sum_{k=0}^{K_n - 1} \Xi_{t_k^n \wedge t, t_{k+1}^n \wedge t}, \qquad t \in [0,T], \tag{1.3}$$

where the convergence is uniform in t and $\{0 = t_0^n < t_1^n < \cdots < t_{K^n}^n = T\}$ is any sequence of partitions with mesh size going to 0, i.e. $\max_k |t_{k+1}^n - t_k^n| \to 0$.

Notation. *For $X: [0,T] \to \mathcal{X}$ we write*

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

Proof.

1. Construction of $\mathcal{I}\Xi$: We define with $t_k^n = k 2^{-n} T$

$$\mathcal{I}_t^n = \sum_{k=0}^{2^n - 1} \Xi_{t_k^n \wedge t, t_{k+1}^n \wedge t}.$$

Since $t_k^n = t_{2k}^{n+1} < t_{2k+1}^{n+1} < t_{2k+2}^{n+1} = t_{k+1}^n$, we have

$$\begin{aligned} &\|\mathcal{I}_t^n - \mathcal{I}_t^{n+1}\| \\ &= \left\| \sum_{k=0}^{2^n-1} \left(\Xi_{t_k^n \wedge t, t_{k+1}^n \wedge t} - \Xi_{t_{2k}^{n+1} \wedge t, t_{2k+1}^{n+1} \wedge t} - \Xi_{t_{2k+1}^{n+1} \wedge t, t_{2k+2}^{n+1} \wedge t} \right) \right\| \\ &\leqslant \sum_{k=0}^{2^n-1} \left\| \Xi_{t_{2k}^{n+1} \wedge t, t_{2k+2}^{n+1} \wedge t} - \Xi_{t_{2k}^{n+1} \wedge t, t_{2k+1}^{n+1} \wedge t} - \Xi_{t_{2k+1}^{n+1} \wedge t, t_{2k+2}^{n+1} \wedge t} \right\| \\ &= \sum_{k=0}^{2^n-1} \left\| \delta \Xi_{t_{2k}^{n+1} \wedge t, t_{2k+1}^{n+1} \wedge t, t_{2k+2}^{n+1} \wedge t} \right\| \\ &\leqslant \sum_{k=0}^{2^n-1} C(T 2^{-n})^{1+\varepsilon} = C T^{1+\varepsilon} 2^{-n\varepsilon}, \end{aligned}$$

and the right hand side does not depend on t and it is summable in n. Therefore, $(\mathcal{I}^n)_n$ is a Cauchy sequence in $C([0,T], \mathcal{X})$ (continuity of $\mathcal{I}^n$ follows from continuity of Ξ and since $\Xi_{s,s} = 0$), and thus it converges to a limit $\mathcal{I}\Xi \in C([0,T], \mathcal{X})$. This essentially concludes the main part of the proof. Everything which follows now is only needed for proving that $\mathcal{I}\Xi$ satisfies (1.2) and that this characterizes $\mathcal{I}\Xi$.

2. We can slightly improve the estimate above by noting that for k with $t_k^n \geqslant t$ we have

$$\Xi_{t_k^n \wedge t, t_{k+1}^n \wedge t} - \Xi_{t_k^n \wedge t, t_{2k+1}^{n+1} \wedge t} - \Xi_{t_{2k+1}^{n+1} \wedge t, t_{k+1}^n \wedge t} = 0,$$

1.1. We equip Δ_T with the subspace topology of $[0,T]^2$.

and therefore we only have to sum $\lceil 2^n t/T \rceil$ terms and this gives the bound

$$\|\mathcal{I}_t^n - \mathcal{I}_t^{n+1}\| \leqslant \left\lceil \frac{2^n t}{T} \right\rceil C(T2^{-n})^{1+\varepsilon} \lesssim CtT^{\varepsilon}2^{-n\varepsilon}$$

whenever $\frac{2^n t}{T} > 1/2$. Similarly we get for $s<t$ with $\frac{2^n|t-s|}{T} > 1/2$ ($\Leftrightarrow 2^n|t-s| > T/2$)

$$\|\mathcal{I}_{s,t}^n - \mathcal{I}_{s,t}^{n+1}\| \lesssim C|t-s|T^{\varepsilon}2^{-n\varepsilon}. \tag{1.4}$$

3. $\mathcal{I}\Xi$ satisfies (1.2): Let $0 \leqslant s < t \leqslant T$ and let $n \in \mathbb{N}$ be maximal such that $2^{-n}T > |t-s|$ (so in particular $2^{-n}T < 2|t-s| \Leftrightarrow 2^n|t-s| > T/2$). Then

$$\|\mathcal{I}\Xi_{s,t} - \Xi_{s,t}\| \leqslant \|\mathcal{I}\Xi_{s,t} - \mathcal{I}_{s,t}^n\| + \|\mathcal{I}_{s,t}^n - \Xi_{s,t}\|.$$

We treat the two terms on the right hand side separately. For the first one we apply the bound (1.4) and obtain

$$\|\mathcal{I}\Xi_{s,t} - \mathcal{I}_{s,t}^n\| \leqslant \sum_{k=n}^{\infty} \|\mathcal{I}_{s,t}^{k+1} - \mathcal{I}_{s,t}^k\| \lesssim \sum_{k=n}^{\infty} C|t-s|T^{\varepsilon}2^{-n\varepsilon} \lesssim C|t-s|^{1+\varepsilon},$$

where we applied the simple lemma which follows after this proof. To bound the remaining term we note that if k is such that $s \in (t_k^n, t_{k+1}^n]$, then $t \in (t_k^n, t_{k+2}^n)$ and thus

$$\begin{aligned}
\|\mathcal{I}_{s,t}^n - \Xi_{s,t}\| &= \|\Xi_{t_k^n, t_{k+1}^n \wedge t} + \Xi_{t_{k+1}^n \wedge t, t} - \Xi_{t_k^n, s} - \Xi_{s,t}\| \\
&\leqslant \|\Xi_{t_k^n, t} - \Xi_{t_k^n, t_{k+1}^n \wedge t} - \Xi_{t_{k+1}^n \wedge t, t}\| + \|\Xi_{t_k^n, t} - \Xi_{t_k^n, s} - \Xi_{s,t}\| \\
&= \|\delta\Xi_{t_k^n, t_{k+1}^n \wedge t, t}\| + \|\delta\Xi_{t_k^n, s, t}\| \\
&\lesssim C|t-s+2^{-n}T|^{1+\varepsilon} \lesssim C|t-s|^{1+\varepsilon}.
\end{aligned}$$

Therefore, $\mathcal{I}\Xi$ satisfies (1.2).

4. Finally we show that (1.2) uniquely characterizes $\mathcal{I}\Xi$ as the limit in (1.3). Indeed, we have

$$\begin{aligned}
\left\| \mathcal{I}\Xi_t - \sum_{k=0}^{K_n-1} \Xi_{t_k^n \wedge t, t_{k+1}^n \wedge t} \right\| &\leqslant \sum_{k=0}^{K_n-1} \|\mathcal{I}\Xi_{t_k^n \wedge t, t_{k+1}^n \wedge t} - \Xi_{t_k^n \wedge t, t_{k+1}^n \wedge t}\| \\
&\lesssim C \sum_{k=0}^{K_n-1} |t_{k+1}^n \wedge t - t_k^n \wedge t|^{1+\varepsilon} \\
&\leqslant C \max_k |t_{k+1}^n - t_k^n|^{\varepsilon} \sum_{k=0}^{K_n-1} |t_{k+1}^n - t_k^n| \\
&= C \max_k |t_{k+1}^n - t_k^n|^{\varepsilon} T,
\end{aligned}$$

and by assumption the right hand side converges to zero as $n \to \infty$. This concludes the proof. □

Lemma 1.3. (Geometric series) *Let $\alpha > 0$. Then for all $n \in \mathbb{N}$*

$$\sum_{k=0}^{n} 2^{k\alpha} = \frac{2^{(n+1)\alpha} - 1}{2^{\alpha} - 1} \simeq 2^{n\alpha}, \qquad \sum_{k=n}^{\infty} 2^{-k\alpha} = 2^{-n\alpha} \sum_{k=0}^{\infty} 2^{-k\alpha} = \frac{2^{-n\alpha}}{1 - 2^{-\alpha}} \simeq 2^{-n\alpha}.$$

Exercise 1.1. Show that $\delta\Xi_{s,u,t} \equiv 0$ for all $s < u < t$ if and only if there exists a function $f: [0,T] \to \mathcal{X}$ such that $\Xi_{s,t} = f_t - f_s$.

Remark 1.4. (Additive function) If Ξ is given by the increments of a function, i.e. $\Xi_{s,t}=x_t-x_s$ for some $x\colon[0,T]\to\mathcal{X}$, then $\Xi_{s,t}=\Xi_{s,u}+\Xi_{u,t}$, and we call Ξ an *additive function*. In that case $\delta\Xi_{s,u,t}=x_{s,t}-x_{s,u}-x_{u,t}=0$ and $\mathcal{I}^n_t=x_t-x_0$ for all n and thus $\mathcal{I}\Xi_t=x_t-x_0$.

So, $\|\delta\Xi_{s,u,t}\|$ measures "how far Ξ is from being an additive function". If $\|\delta\Xi_{s,u,t}\|\lesssim|t-s|^{1+\varepsilon}$, then in general Ξ is not an additive function, but the sewing lemma shows that there exists a unique (up to addition of constants) additive function $\mathcal{I}\Xi$ such that $\|\mathcal{I}\Xi_{s,t}-\Xi_{s,t}\|\lesssim|t-s|^{1+\varepsilon}$. This additive function is obtained by "sewing together" Ξ, since $\mathcal{I}\Xi_t=\lim_{n\to\infty}\sum_{k=0}^{K_n-1}\Xi_{t^n_k\wedge t,t^n_{k+1}\wedge t}$.

Exercise 1.2. Show that if $x\colon[0,T]\to\mathbb{R}$ is α-Hölder continuous for $\alpha>1$, then x is constant: $x_t=x_0$ for all $t\in[0,T]$.

Corollary 1.5. (Young integral) *Let $T>0$ and let $\alpha,\beta\in(0,1]$ be such that $\alpha+\beta>1$. Let $Y\in C^\alpha([0,T])$ and $X\in C^\beta([0,T])$. Then the Young integral*

$$\int_0^t Y_s\mathrm{d}X_s=\lim_{n\to\infty}\sum_{k=0}^{2^n-1}Y_{t^n_k}X_{t^n_k\wedge t,t^n_{k+1}\wedge t},\qquad t^n_k=k2^{-n}T,$$

exists and it is the unique function starting from 0 at $t=0$ and such that for all $[s,t]\subset[0,T]$:

$$\left|\int_s^t Y_r\mathrm{d}X_r-Y_sX_{s,t}\right|\lesssim\|Y\|_\alpha\|X\|_\beta|t-s|^{\alpha+\beta}.\tag{1.5}$$

Moreover,

$$\left\|\int_0^\cdot Y_s\mathrm{d}X_s\right\|_\beta\lesssim(1+T^\alpha)(|Y_0|+\|Y\|_\alpha)\|X\|_\beta.\tag{1.6}$$

Proof. We define

$$\Xi_{s,t}:=Y_s(X_t-X_s),$$

which is continuous in (s,t), satisfies $\Xi_{t,t}=Y_t(X_t-X_t)=0$, and

$$\begin{aligned}|\delta\Xi_{s,u,t}|&=|Y_s(X_t-X_s)-Y_s(X_u-X_s)-Y_u(X_t-X_u)|\\&=|(Y_s-Y_u)(X_t-X_u)|\\&\leqslant\|Y\|_\alpha\|X\|_\beta|u-s|^\alpha|t-u|^\beta\\&\leqslant\|Y\|_\alpha\|X\|_\beta|t-s|^{\alpha+\beta},\end{aligned}$$

and since $\alpha+\beta>1$ we can apply the sewing lemma. To prove (1.6) we use (1.5) and the β-Hölder continuity of X:

$$\begin{aligned}\left|\int_s^t Y_r\mathrm{d}X_r\right|&\leqslant\left|\int_s^t Y_r\mathrm{d}X_r-Y_sX_{s,t}\right|+|Y_sX_{s,t}|\\&\lesssim\|Y\|_\alpha\|X\|_\beta|t-s|^{\alpha+\beta}+|Y_{0,s}X_{s,t}|+|Y_0X_{s,t}|\\&\leqslant T^\alpha\|Y\|_\alpha\|X\|_\beta|t-s|^\beta+\|Y\|_\alpha T^\alpha\|X\|_\beta|t-s|^\beta+|Y_0|\|X\|_\beta|t-s|^\beta\\&\lesssim(1+T^\alpha)(|Y_0|+\|Y\|_\alpha)\|X\|_\beta|t-s|^\beta,\end{aligned}$$

which concludes the proof. □

Exercise 1.3. Let $X_t=B_t(\omega)$ for a typical sample path of a Brownian motion. Which Y can we integrate using the Young integral? Is $Y_t=f(B_t(\omega))$ for a smooth function f ok? Or $Y_t=Y_t(\omega)$, where Y solves the Itô SDE $\mathrm{d}Y_t=b(Y_t)\mathrm{d}t+\sigma(Y_t)\mathrm{d}B_t$?

Remark 1.6. If X is of finite variation (e.g. Lipschitz continuous), then by definition the Young integral agrees with the Lebesgue-Stieltjes integral $\int_0^\cdot Y_s \mathrm{d}X_s$ (since both are limits of the same sums).

Remark 1.7. (Young integral in Banach spaces) We constructed the Young integral for real-valued Y, X. But let $\mathcal{X}, \mathcal{Y}$ be Banach spaces and let $L(\mathcal{X}, \mathcal{Y})$ be the space of bounded linear operators from $\mathcal{X}$ to $\mathcal{Y}$. Then for $Y \in C^\alpha([0,T], L(\mathcal{X}, \mathcal{Y}))$ and $X \in C^\beta([0,T], \mathcal{X})$ the Young integral $\int_0^\cdot Y_s \mathrm{d}X_s \in C^\beta([0,T], \mathcal{Y})$ can be constructed in the same way and it satisfies the same estimates.

If we want use Young integration to solve the equation

$$dY_t = b(Y_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}X_t$$

for $X \in C^\alpha([0,T])$ (for example $X = B(\omega)$ for a fractional Brownian motion with Hurst index $H > \alpha$), then we should assume that $\sigma(Y) \in C^\beta([0,T])$ with $\beta > 1 - \alpha$. We will see that for nice functions σ the path $\sigma(Y)$ has the same regularity as Y, and therefore we need $Y \in C^\beta([0,T])$. Then $\int_0^\cdot \sigma(Y_s)\mathrm{d}X_s \in C^\alpha([0,T])$, and since $\int_0^\cdot b(Y_s)\mathrm{d}s$ is Lipschitz, also $Y \in C^\alpha([0,T])$ and this means that we could at best take $\beta = \alpha$, which leads to the requirement $\alpha > 1 - \alpha$ or equivalently $\alpha > 1/2$. We call this the *Young regime*.

The Young integral estimates are already sufficient to solve linear equations. To handle nonlinear equations we need to understand how nonlinearities interact with Hölder regularity. To state the result we need the following function space and norm:

$$C_b^\gamma(\mathbb{R}^d, \mathbb{R}^n) := \{f \in C^{\lfloor \gamma \rfloor}(\mathbb{R}^d, \mathbb{R}^n) : \|f\|_{C_b^\gamma} < \infty\},$$

where we recall that $\lfloor \gamma \rfloor = \max\{k \in \mathbb{Z} : k \leqslant \gamma\}$ and that $\partial^\mu = \partial_1^{\mu_1} \cdots \partial_d^{\mu_d}$ and $|\mu| = \mu_1 + \cdots + \mu_d$ for $\mu \in \mathbb{N}_0^d$, with which

$$\|f\|_{C_b^\gamma} := \sum_{|\mu| \leqslant \lfloor \gamma \rfloor} \|\partial^\mu f\|_\infty + \max_{|\mu| = \lfloor \gamma \rfloor} \sup_{x \neq y} \frac{|\partial^\mu f(x) - \partial^\mu f(y)|}{|x - y|^{\gamma - \lfloor \gamma \rfloor}}.$$

In words, $C_b^\gamma(\mathbb{R}^d, \mathbb{R}^n)$ is the space of functions which are $\lfloor \gamma \rfloor$ times continuously differentiable with bounded partial derivatives, and the partial derivatives of order $\lfloor \gamma \rfloor$ are $\gamma - \lfloor \gamma \rfloor$ Hölder continuous. For $\gamma = 1$ this means that f is Lipschitz continuous. For $\gamma = \lfloor \gamma \rfloor$ this means that the partial derivatives of order $\lfloor \gamma \rfloor$ are continuous. We also write

$$\|Y\|_{C_b^\alpha} := \|Y\|_\alpha + \|Y\|_\infty$$

for $\alpha \in (0,1]$ and $Y \in C^\alpha([0,T])$.

Lemma 1.8. *Let $\alpha \in (0,1]$ and let $Y, \tilde{Y} \in C^\alpha([0,T], \mathbb{R}^d)$. Let $\sigma \in C_b^1(\mathbb{R}^d, \mathbb{R}^n)$. Then $\sigma(Y) \in C^\alpha([0,T], \mathbb{R}^n)$ and*

$$\|\sigma(Y)\|_\alpha \leqslant \|\sigma\|_{C_b^1} \|Y\|_\alpha.$$

For $\sigma \in C_b^2(\mathbb{R}^d, \mathbb{R}^n)$ we have

$$\|\sigma(Y) - \sigma(\tilde{Y})\|_\alpha \lesssim \|\sigma\|_{C_b^2} (1 + \|Y - \tilde{Y}\|_\infty) \|Y - \tilde{Y}\|_\alpha.$$

Proof. The first bound is easy:

$$|\sigma(Y)_{s,t}| \leqslant \|\sigma\|_{C_b^1} |Y_{s,t}| \leqslant \|\sigma\|_{C_b^1} \|Y\|_\alpha |t - s|^\alpha.$$

The bound for the difference is a bit more technical. Taylor's formula gives

$$\begin{aligned}|\sigma(Y)_{s,t}-\sigma(\tilde{Y})_{s,t}| &= \left|\int_0^1 \nabla\sigma(\tilde{Y}_t+\lambda(Y_t-\tilde{Y}_t))\cdot(Y_t-\tilde{Y}_t)d\lambda - \int_0^1 \nabla\sigma(\tilde{Y}_s+\lambda(Y_s-\tilde{Y}_s))\cdot(Y_s-\tilde{Y}_s)d\lambda\right| \\ &\leqslant \left|\int_0^1 [\nabla\sigma(\tilde{Y}_t+\lambda(Y_t-\tilde{Y}_t))-\nabla\sigma(\tilde{Y}_s+\lambda(Y_s-\tilde{Y}_s))]\cdot(Y_t-\tilde{Y}_t)d\lambda\right| \\ &\quad + \left|\int_0^1 \nabla\sigma(\tilde{Y}_s+\lambda(Y_s-\tilde{Y}_s))\cdot(Y_{s,t}-\tilde{Y}_{s,t})d\lambda\right| \\ &\leqslant \|\sigma\|_{C_b^2}|\tilde{Y}_t+\lambda(Y_t-\tilde{Y}_t)-(\tilde{Y}_s+\lambda(Y_s-\tilde{Y}_s))|\|Y-\tilde{Y}\|_\infty \\ &\quad +2\|\sigma\|_{C_b^1}\|Y-\tilde{Y}\|_\alpha|t-s|^\alpha \\ &\lesssim \|\sigma\|_{C_b^2}\|Y-\tilde{Y}\|_\alpha\|Y-\tilde{Y}\|_\infty|t-s|^\alpha+\|\sigma\|_{C_b^1}\|Y-\tilde{Y}\|_\alpha|t-s|^\alpha \\ &\lesssim \|\sigma\|_{C_b^2}(1+\|Y-\tilde{Y}\|_\infty)\|Y-\tilde{Y}\|_\alpha|t-s|^\alpha.\end{aligned}$$

□

Exercise 1.4. Show that for $\sigma\in C_b^1$ and $Y\in C^\alpha$ we have

$$|\sigma(Y)_{s,t}-\sigma'(Y_s)Y_{s,t}|\leqslant\|\sigma\|_{C_b^2}\|Y\|_\alpha^2|t-s|^{2\alpha}.$$

Deduce that for $\alpha>\frac{1}{2}$ we have

$$\sigma(Y_t)=\sigma(Y_0)+\int_0^t\sigma'(Y_s)dY_s.$$

Theorem 1.9. (Young equation) *Let $\alpha\in(\frac{1}{2},1]$, let $X\in C^\alpha([0,T],\mathbb{R}^n)$ and let $b\in C_b^1(\mathbb{R}^d,\mathbb{R}^d)$ and $\sigma\in C_b^2(\mathbb{R}^d,\mathbb{R}^{d\times n})$. Then for all $x_0\in\mathbb{R}^d$ there exists a unique solution $Y\in C^\alpha([0,T],\mathbb{R}^d)$ to the Young integral equation*

$$\begin{aligned}Y_t &= x_0+\int_0^t b(Y_s)\mathrm{d}s+\int_0^t\sigma(Y_s)\mathrm{d}X_s,\\ &:= x_0+\int_0^t b(Y_s)\mathrm{d}s+\sum_{j=1}^n\int_0^t\sigma_{\cdot,j}(Y_s)\mathrm{d}X_s^j,\qquad t\in[0,T].\end{aligned}$$

Moreover, Y depends locally Lipschitz continuously on $(x_0,X)\in\mathbb{R}^d\times C^\alpha([0,T],\mathbb{R}^n)$: If $\tilde{Y}$ solves the same equation for $\tilde{x}_0,\tilde{X}$, then there exists $K>0$ depending only on b,σ,T and $|x_0|,|\tilde{x}_0|$ and $\|X\|_\alpha,\|\tilde{X}\|_\alpha$, such that for all $\alpha'\in(\frac{1}{2},\alpha)$:

$$\|Y-\tilde{Y}\|_{\alpha'}\leqslant K(|x_0-\tilde{x}_0|+\|X-\tilde{X}\|_\alpha).$$

Proof.

1. We use a Picard iteration on a small time interval. Let $\alpha'\in(\frac{1}{2},\alpha)$. For $\tau\in(0,1\wedge T]$ we consider

$$B_\tau:=\{Y\in C([0,\tau],\mathbb{R}^d):Y(0)=x_0,\|Y\|_{\alpha'}\leqslant 1\},$$

where we write $\|Y\|_{\alpha'}$ and $\|X\|_\alpha$ for the norms restricted to $[0,\tau]$. For $Y\in B_\tau$ we define

$$\Phi(Y)_t:=x_0+\int_0^t b(Y_s)\mathrm{d}s+\int_0^t\sigma(Y_s)\mathrm{d}X_s,\qquad t\in[0,\tau].$$

2. We first show that Φ leaves B_τ invariant if $\tau>0$ is sufficiently small. We have $\Phi(Y)_0=x_0$ and by the bound (1.6) for the Young integral

$$\begin{aligned}\|\Phi(Y)\|_{\alpha'} &\overset{(1.6)}{\lesssim} \tau^{1-\alpha'}\|b\|_\infty+(1+\tau^\alpha)(|\sigma(x_0)|+\|\sigma(Y)\|_{\alpha'})\|X\|_{\alpha'}\\ &\overset{\text{Lem. 1.8}}{\lesssim} \tau^{1-\alpha'}\|b\|_\infty+(1+\tau^\alpha)(\|\sigma\|_\infty+\|\sigma\|_{C_b^1}\|Y\|_{\alpha'})\|X\|_{\alpha'}\\ &\overset{(1.7)}{\lesssim} \tau^{1-\alpha'}\|b\|_\infty+\tau^{\alpha-\alpha'}\|\sigma\|_{C_b^1}\|X\|_\alpha,\end{aligned}$$

where we used that

$$\|X\|_{\alpha'}=\sup_{0\leqslant s<t\leqslant\tau}\frac{|X_{s,t}|}{|t-s|^{\alpha'}}\leqslant\sup_{0\leqslant s<t\leqslant\tau}\frac{|X_{s,t}|}{|t-s|^{\alpha}}\sup_{0\leqslant s<t\leqslant\tau}\frac{|t-s|^\alpha}{|t-s|^{\alpha'}}=\|X\|_\alpha\tau^{\alpha-\alpha'}, \tag{1.7}$$

and also that $\tau\leqslant 1$ so any positive power of τ can be bounded by 1. So if $\tau\in(0,T\wedge 1]$ is small enough (depending only on b,σ,X but not on x_0), then Φ leaves B_τ invariant.

3. Next, we show that Φ is a contraction on the complete metric space $(B_\tau,\|\cdot\|_\alpha)$ (possibly after further decreasing the value of τ). Note that $\|\cdot\|_\alpha$ is only a seminorm because $\|c\|_\alpha=0$ for any constant function c. But since in B_τ we fix the initial value x_0, $\|\cdot\|_\alpha$ becomes a norm. Using the completeness of $\mathbb{R}^d$ it is not difficult to show that $(B_\tau,\|\cdot\|_\alpha)$ is indeed complete.

   To see the contraction property of Φ, note that $\alpha+\alpha'>1$ by assumption (since $\alpha,\alpha'>\frac{1}{2}$), and therefore the bound for the Young integral below is justified:

$$\begin{aligned}\|\Phi(Y)-\Phi(\tilde{Y})\|_{\alpha'} &\leqslant \left\|\int_0^\cdot (b(Y_s)-b(\tilde{Y}_s))\mathrm{d}s\right\|_{\alpha'}+\left\|\int_0^\cdot(\sigma(Y_s)-\sigma(\tilde{Y}_s))\mathrm{d}X_s\right\|_{\alpha'}\\ &\leqslant \tau^{1-\alpha'}\|b(Y)-b(\tilde{Y})\|_\infty+\tau^{\alpha-\alpha'}\left\|\int_0^\cdot(\sigma(Y_s)-\sigma(\tilde{Y}_s))\mathrm{d}X_s\right\|_\alpha\\ &\overset{(1.6)}{\lesssim} \tau^{1-\alpha'}\|b(Y)-b(\tilde{Y})\|_\infty+\tau^{\alpha-\alpha'}(1+\tau^\alpha)\|\sigma(Y)-\sigma(\tilde{Y})\|_{\alpha'}\|X\|_\alpha\\ &\overset{\text{Lem. 1.8}}{\lesssim} \tau^{1-\alpha'}\|b\|_{C_b^1}\|Y-\tilde{Y}\|_\infty+\tau^{\alpha-\alpha'}\|\sigma\|_{C_b^2}(1+\|Y-\tilde{Y}\|_\infty)\\ &\qquad\qquad\times\|Y-\tilde{Y}\|_{\alpha'}\|X\|_\alpha\\ &\leqslant \tau^{1-\alpha'}\|b\|_{C_b^1}\tau^{\alpha'}\|Y-\tilde{Y}\|_{\alpha'}+\tau^{\alpha-\alpha'}\|\sigma\|_{C_b^2}\|Y-\tilde{Y}\|_{\alpha'}\|X\|_\alpha\\ &\lesssim (\tau\|b\|_{C_b^1}+\tau^{\alpha-\alpha'}\|\sigma\|_{C_b^2}\|X\|_{\alpha'})\|Y-\tilde{Y}\|_{\alpha'},\end{aligned}$$

   where we used that since Y and $\tilde{Y}$ both start in x_0:

$$\|Y-\tilde{Y}\|_\infty\leqslant\tau^{\alpha'}\|Y-\tilde{Y}\|_{\alpha'}\lesssim 1.$$

   So if τ is sufficiently small, then Φ is indeed a contraction and there exists a unique fixed point.

4. The length τ of the interval on which Φ is a contraction does not depend on x_0, so now we can interate the construction on $[\tau,2\tau]$, then on $[2\tau,3\tau]$, etc., until we reach $[0,T]$. This concludes the proof of existence and uniqueness of solutions. However, so far we only showed that the solution Y is in $C^{\alpha'}([0,T],\mathbb{R}^d)$. To show that Y is even α-Hölder continuous we use the fact that Y solves the equation and that the right hand side of the equation is in C^α for any $Y\in C^{\alpha'}$.

5. The continuous dependence of Y on (x_0,X) is left as an exercise. □

Exercise 1.5. $C^\alpha([0,T],\mathbb{R}^d)$ is compactly embedded in $C^{\alpha'}([0,T],\mathbb{R}^d)$ for all $\alpha'<\alpha$. Use this to show existence of solutions to Young equations under the assumption that $\sigma\in C_b^1$ (instead of $\sigma\in C_b^2$).

Corollary 1.10. *Let $(B_t)_{t\in[0,T]}$ be an n-dimensional fractional Brownian motion of Hurst parameters $H>\frac{1}{2}$ (i.e. the components $(B^1,\ldots,B^n)$ are i.i.d. and each component is a fractional Brownian motion). Let $b\in C_b^1(\mathbb{R}^d,\mathbb{R}^d)$ and $\sigma\in C_b^2(\mathbb{R}^d,\mathbb{R}^{d\times n})$ and let $x_0\in\mathbb{R}^d$. Let $\alpha\in(\frac{1}{2},H)$ Then there exists a unique (up to indistinguishability) process Y such that almost surely $(Y_t)_{t\in[0,T]}\in C^\alpha([0,T],\mathbb{R}^d)$ and*

$$Y_t=Y_0+\int_0^t b(Y_s)\mathrm{d}s+\int_0^t\sigma(Y_s)\mathrm{d}B_s,$$

where the integral against B is a Young integral which is well defined almost surely.

If $(B^m)_{m\in\mathbb{N}}\subset C^1([0,T],\mathbb{R}^d)$ is a sequence of paths such that almost surely $\|B-B^m\|_{C_b^\alpha}\to 0$, then $Y=\lim_{m\to\infty}Y^m$, where

$$\partial_t Y_t^m=b(Y_t^m)+\sigma(Y_t^m)\partial_t B_t^m,\qquad Y_0^m=Y_0.$$

Remark 1.11. In this result we first freeze the realization $B(\omega)$ of the noise, and then we perform deterministic analysis with this given path. This is very different from Itô stochastic differential equations, which do not make sense for a fixed ω and for which the solution is only defined "modulo null sets". Also, if for the Young differential equation we take the canonical probability space of fractional Brownian motion,

$$\Omega=C^\alpha([0,T],\mathbb{R}^n),$$

for $\alpha\in(\frac{1}{2},H)$, then the map $\omega\mapsto Y(\omega)$ is continuous. While for Itô SDEs it is only measurable.

Our theorem excludes the most interesting case $H=\frac{1}{2}$, which corresponds to the Brownian motion. On the exercise sheet we will see that the conditions for the Young integral are sharp. So, to treat the Brownian case with a similar philosophy and to cover the case $H<\frac{1}{2}$ we need to do more. The solution is to "enrich" the path X by equipping it with more information. In that way we can construct a continuous pathwise integral which applies to paths of regularity $<\frac{1}{2}$.

2 A crash course in rough path theory

Here we give a very brief introduction to the main ideas and techniques of Terry Lyons's rough path theory [Lyo98]. We use Gubinelli's approach [Gub04], which is beautifully exposed in the monograph [FH14], see also [LCL07] for nice lecture notes on rough paths. We first focus on "mildly rough" noise, such as Brownian motion and, more generally, fractional Brownian motion with $H>\frac{1}{3}$. Then we discuss the extension to general rough paths, which is similar in spirit but technically more involved and which makes appear certain algebraic structures.

2.1 Idea and definition of a rough path

To be able to treat the Brownian motion, we would like to extend the theory of Young equations to driving signals $X\in C^\alpha([0,T],\mathbb{R}^n)$ with $\alpha\leqslant\frac{1}{2}$. Unfortunately, a naive solution is not possible.

Example 2.1. Consider the case $d=n=2$ and

$$Y_t^1 = \int_0^t \mathrm{d}X_s^1, \qquad Y_t^2 = \int_0^t Y_s^1 \mathrm{d}X_s^2, \tag{2.1}$$

with initial condition $X(0)=0$. This equation has the explicit solution $Y^1=X^1$ and $Y^2=\int_0^\cdot X^1 \mathrm{d}X^2$. For $m\in\mathbb{N}$ we set

$$X_t^m = \begin{pmatrix} \frac{1}{m}\cos(m^2 t) \\ \frac{1}{m}\sin(m^2 t) \end{pmatrix}.$$

From the exercise sheet we know that (X^m) converges to 0 uniformly and in C^α for all $\alpha<1/2$, and that $Y_t^{2,m}\to\frac{t}{2}$ as $m\to\infty$. But of course, the solution to (2.1) with $X\equiv 0$ is equal to $(0,0)$ and not $(0,t/2)$, and therefore Y does not depend continuously on X in C^α-norm if $\alpha<1/2$.

The problem is that the fast oscillations of X^m interact with the nonlinearity in our integral equation and this interaction creates nontrivial effects, even though the amplitude of X^m is very small.

Note that all paths involved in this example are smooth (C^∞), so the problem is not the lack of regularity but the topology in which the sequence $(X^m)_{m\in\mathbb{N}}$ converges to 0. In the following we will introduce *rough path topologies* which help us to overcome this lack of continuity.

We can imagine two possible approaches for doing so. The naive one would be to try to find a better suited function space, which contains Brownian paths and paths of solutions to SDEs and in which the integral $\int_0^\cdot Y_s \mathrm{d}X_s$ becomes a continuous functional (i.e. not to work with Hölder norms). However, this is impossible! A counterexample by Lyons shows that there cannot exist a Banach space $\mathcal{X}$ of real-valued functions on $[0,1]$ such that $\mathcal{X}$ contains almost all sample paths of the Brownian motion and such that there exists a continuous functional $\mathcal{X}^2\ni(Y,X)\mapsto I(Y,X)$ with $I(Y,X)=\int_0^1 Y_s\partial_s X_s \mathrm{d}s$ whenever $X\in C^1$; see Section 1.5.1 of [LCL07].

The second approach is to accept this lack of continuity, and to *enhance* the path X to make the map $X\mapsto Y$ continuous. To understand this philosophy, let us consider the following trivial example:

Example 2.2. The map $f:\mathbb{R}\to\mathbb{R}$,

$$f(x) = \begin{cases} -1, & x<0, \\ +1, & x\geqslant 0, \end{cases}$$

is obviously discontinuous in 0. The problem is that we can approach 0 from the left or from the right, and $\mathbb{R}$ is not rich enough to encode the information "from where we are coming". To obtain a continuous map, we could enhance the input space to encode the information whether we approach 0 from the left or from the right. More precisely, we consider

$$\mathcal{X} := ((-\infty,0)\times\{-\})\cup(\{0\}\times\{-,+\})\cup((0,\infty)\times\{+\})\subset\mathbb{R}\times\{-,+\}$$

where $\mathbb{R}\times\{-,+\}$ is equipped with the product topology (and $\{-,+\}$ is equipped with the discrete topology). Note that $\mathbb{R}\subset\mathcal{X}$, i.e. there exists a (non-canonical) injection Φ from $\mathbb{R}$ to $\mathcal{X}$ by setting $\Phi(x)=(x,+)$ if $x\geqslant 0$ and $\Phi(x)=(x,-)$ if $x<0$. We define

$$g(x,+)=+1, \qquad g(x,-)=-1.$$

Then g is continuous, and we have

$$f(x)=g(\Phi(x)).$$

In other words, we have decomposed f as the concatenation of the continuous map g with the discontinuous and non-canonical map Φ.

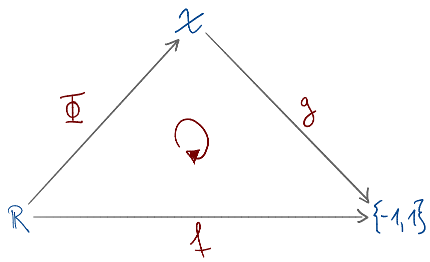


Figure 2.1. Commuting diagram.

Example 2.3. To understand how we should enhance our path space let $X \in C^1([0,T],\mathbb{R}^d)$ and let us try to construct $\int_0^1 f(X_s)\mathrm{d}X_s$ in a way that depends continuously on X in C^α topology for $\alpha < 1/2$, where $f:\mathbb{R}^d \to L(\mathbb{R}^d,\mathbb{R})$ is a smooth bounded function with bounded derivatives. Since X and $f(X)$ are both Lipschitz continuous, we have along a sequence of partitions with mesh size going to zero:

$$\int_0^1 f(X_s)\mathrm{d}X_s = \lim_{n\to\infty}\sum_k f(X_{t_k^n})X_{t_k^n,t_{k+1}^n}.$$

We would like to control this integral using only the C^α-norm of X, and for that purpose we want to apply the sewing lemma. With $\Xi_{s,t} = f(X(s))X_{s,t}$ we get

$$|\delta\Xi_{s,u,t}| = |-f(X)_{s,u}X_{u,t}| \leqslant \|f\|_{C_b^1}\|X\|_\alpha^2|t-s|^{2\alpha}.$$

Since $\alpha < 1/2$, this is not good enough to apply the sewing lemma, and therefore we are stuck. But we can try to improve the approximation:

$$\begin{aligned}
\int_0^1 f(X_s)\mathrm{d}X_s &= \sum_{i=1}^d \int_0^1 f_i(X_s)\mathrm{d}X_s^i \\
&= \sum_k \left(\sum_{i=1}^d \int_{t_k^n}^{t_{k+1}^n} f_i(X_{t_k^n})\mathrm{d}X_s^i + \sum_{i=1}^d \int_{t_k^n}^{t_{k+1}^n} (f_i(X_s) - f_i(X_{t_k^n}))\mathrm{d}X_s^i\right) \\
&= \sum_k \left[\sum_{i=1}^d f_i(X_{t_k^n})X^i_{t_k^n,t_{k+1}^n} + \sum_{j=1}^d \partial_j f_i(X_{t_k^n})\sum_{i=1}^d \int_{t_k^n}^{t_{k+1}^n} (X_s^j - X_{t_k^n}^j)\mathrm{d}X_s^i\right] \\
&\quad + \sum_k \sum_{i=1}^d \int_{t_k^n}^{t_{k+1}^n}\left(f_i(X_s) - f_i(X_{t_k^n}) - \sum_{j=1}^d \partial_j f_i(X_{t_k^n})(X_s^j - X_{t_k^n}^j)\right)\mathrm{d}X_s^i.
\end{aligned}$$

For the last term on the right hand side we expect to get

$$\sum_k \sum_{i=1}^d \int_{t_k^n}^{t_{k+1}^n} O(|t_{k+1}^n - t_k^n|^{2\alpha})\mathrm{d}X_s^i = \sum_k O(|t_{k+1}^n - t_k^n|^{3\alpha}) \leqslant \max_k |t_{k+1}^n - t_k^n|^{3\alpha-1},$$

which converges to 0 if $\alpha > \frac{1}{3}$ (we have not shown fully rigorously that this term is really of order $O(|t_{k+1}^n - t_k^n|^{3\alpha})$, but it can be justified).

On the other hand, for the second term on the right hand side we expect to have

$$\sum_k \sum_{i=1}^d \int_{t_k^n}^{t_{k+1}^n} \sum_{j=1}^d \partial_j f_i(X_{t_k^n})(X_s^j - X_{t_k^n}^j)\mathrm{d}X_s^i = \sum_k O(|t_{k+1}^n - t_k^n|^{2\alpha}),$$

and if $\alpha \leqslant \frac{1}{2}$, then this is not negligible (unlike for the Young case $\alpha > \frac{1}{2}$). This suggests to consider a different Ξ in the sewing lemma, namely

$$\Xi_{s,t} := f(X_s)X_{s,t} + \mathrm{D}f(X_s)\mathbb{X}_{s,t},$$

where $\mathrm{D}f = \begin{pmatrix} \partial_1 f \\ \vdots \\ \partial_d f \end{pmatrix}$ is the derivative of f and

$$\mathbb{X}_{s,t} := \int_s^t (X_r - X_s) \otimes \mathrm{d}X_r = \left(\int_s^t (X_r^i - X_s^i)\mathrm{d}X_r^j \right)_{i,j=1,\ldots,d} \in \mathbb{R}^{d\otimes d}.$$

and

$$\mathrm{D}f(X)\mathbb{X} = \sum_{i=1}^d \partial_i f(X)\mathbb{X}^{i,\cdot} = \sum_{i,j=1}^d \partial_i f_j(X)\mathbb{X}^{i,j}, \qquad \mathbb{X} \in \mathbb{R}^{d\otimes d}.$$

Then

$$\delta\Xi_{s,u,t} = -f(X)_{s,u}X_{u,t} + \mathrm{D}f(X_s)(\mathbb{X}_{s,t} - \mathbb{X}_{s,u}) - \mathrm{D}f(X_u)\mathbb{X}_{u,t}, \tag{2.2}$$

and

$$\begin{aligned} \mathbb{X}_{s,t} - \mathbb{X}_{s,u} &= \int_s^t (X_r - X_s) \otimes \mathrm{d}X_s - \int_s^u (X_r - X_s) \otimes \mathrm{d}X_r \\ &= \int_u^t (X_r - X_u) \otimes \mathrm{d}X_r + X_u \otimes X_{u,t} - X_s \otimes X_{s,t} + X_s \otimes X_{s,u} \\ &= \mathbb{X}_{u,t} + X_{s,u} \otimes X_{u,t}. \end{aligned}$$

Therefore, we obtain in (2.2)

$$\begin{aligned} |\delta\Xi_{s,u,t}| &\leqslant |-f(X)_{s,u}X_{u,t} + \mathrm{D}f(X_s)X_{s,u} \otimes X_{u,t}| + |-\mathrm{D}f(X)_{s,u}\mathbb{X}_{u,t}| \\ &\leqslant \|f\|_{C_b^2} \|X\|_\alpha^3 |t-s|^{3\alpha} + \|\sigma\|_{C_b^2} \|X\|_\alpha \|\mathbb{X}\|_{2\alpha} |t-s|^{3\alpha} = O(|t-s|^{3\alpha}). \end{aligned} \tag{2.3}$$

So if $\alpha > \frac{1}{3}$, we can apply the sewing lemma to bound the integral in terms of f, X, and $\mathbb{X}$. In other words, the knowledge of the functional $\mathbb{X}$ allows us to construct the integral $\int f(X)\mathrm{d}X$ for all $f \in C_b^2(\mathbb{R}^d, L(\mathbb{R}^d, \mathbb{R}))$ as a continuous functional of $(X, \mathbb{X})$. As we will see soon, it also allows us to solve differential equations driven by $(X, \mathbb{X})$ and that the solution depends continuously on the signal.

Exercise 2.1. (Difficult) What could we do if $\alpha \in (\frac{1}{4}, \frac{1}{3}]$?

Let us write $\Delta_T = \{(s,t) \in [0,T]^2 : s \leqslant t\}$ and

$$C_2^{2\alpha}(\Delta_T, \mathbb{R}^{d\otimes d}) := \{f : \Delta_T \to \mathbb{R}^{d\otimes d} : \|f\|_{2\alpha} < \infty\},$$

where

$$\|f\|_{2\alpha} := \sup_{0 \leqslant s < t \leqslant T} \frac{|f_{s,t}|}{|t-s|^{2\alpha}}.$$

Definition 2.4. (Rough path) *Let $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ and $d \in \mathbb{N}$. A d-dimensional α-*rough path *is a pair $(X, \mathbb{X}) =: \boldsymbol{X}$ with $X \in C^\alpha([0,T], \mathbb{R}^d)$ and $\mathbb{X} \in C_2^{2\alpha}(\Delta_T, \mathbb{R}^{d\otimes d})$, such that* Chen's relation

$$\delta\mathbb{X}_{s,u,t} = X_{s,u} \otimes X_{u,t} \tag{2.4}$$

holds for all $0 \leqslant s \leqslant u \leqslant t \leqslant T$. *We define*

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

We say that a sequence of α*-rough paths* $(\boldsymbol{X}^m)$ *converges to* $\boldsymbol{X}$ *in* α*-rough path topology if*

$$\lim_{m\to\infty} |||\boldsymbol{X}^m - \boldsymbol{X}|||_\alpha := \lim_{m\to\infty} \left(\|X^m - X\|_\alpha + \sqrt{\|\mathbb{X}^m - \mathbb{X}\|_{2\alpha}}\right) = 0.$$

Remark 2.5.

i. We think of $\mathbb{X}$ as postulating “iterated integrals” of X,

$$\mathbb{X}_{s,t} = \int_s^t \int_s^{r_2} \mathrm{d}X_{r_1} \otimes \mathrm{d}X_{r_2} = \int_s^t X_r \otimes \mathrm{d}X_r - X_s \otimes X_{s,t}.$$

Since $X \in C^\alpha$ for $\alpha \leqslant \frac{1}{2}$, the right hand side is not well defined in general, so the left hand side should be read as its definition.

ii. At this point, Chen's rule is simply saying that there exists a function $I\colon [0,T] \to \mathbb{R}^{d\otimes d}$ such that

$$\mathbb{X}_{s,t} = I_{s,t} - X_s \otimes X_{s,t},$$

see the lemma below.

iii. The space of rough paths is not a linear space, because Chen's relation (2.4) is not preserved under linear operations. Intuitively, knowing $\int_s^t X_r \otimes \mathrm{d}X_r$ and $\int_s^t \tilde{X}_r \otimes \mathrm{d}\tilde{X}_r$ does not mean that we know $\int_s^t (X_r + \tilde{X}_r) \otimes \mathrm{d}(X_r + \tilde{X}_r)$, and even if we did, it would not equal to $\int_s^t X_r \otimes \mathrm{d}X_r + \int_s^t \tilde{X}_r \otimes \mathrm{d}\tilde{X}_r$.

iv. Also, $|||\boldsymbol{X}|||_\alpha$ is of course not a norm. The reason for considering $\sqrt{\|\mathbb{X}\|_{2\alpha}}$ rather than $\|\mathbb{X}\|_{2\alpha}$ is that the natural dilation on rough path space is $(X, \mathbb{X}) \mapsto (\lambda X, \lambda^2 \mathbb{X})$. Indeed,

$$\int_s^t (\lambda X)_{r_1} \otimes \mathrm{d}(\lambda X)_{r_1} - (\lambda X)_s \otimes (\lambda X)_{s,t} = \lambda^2 \mathbb{X}_{s,t}.$$

So by taking $\sqrt{\|\mathbb{X}\|_{2\alpha}}$, we make $|||\cdot|||_\alpha$ homogeneous under dilations.

Lemma 2.6. *Let* $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ *and* $d \in \mathbb{N}$. *An alternative definition of a* d*-dimensional* α*-rough path is as follows: It is a pair* (X, I) *with* $X \in C^\alpha([0,T], \mathbb{R}^d)$, $I \in C^\alpha([0,T], \mathbb{R}^{d\otimes d})$, *such that*

$$\sup_{(s,t)\in\Delta_T} \frac{|I_{s,t} - X_s \otimes X_{s,t}|}{|t-s|^{2\alpha}} < \infty. \tag{2.5}$$

The link with Definition 2.4 is

$$I_{s,t} = \mathbb{X}_{s,t} + X_s \otimes X_{s,t}.$$

Proof. If (X, I) are as claimed, then the function $\mathbb{X}_{s,t} := I_{s,t} - X_s \otimes X_{s,t}$ satisfies Chen's relation:

$$\begin{aligned}\delta\mathbb{X}_{s,u,t} &= \delta I_{s,u,t} - X_s \otimes X_{s,t} + X_s \otimes X_{s,u} + X_u \otimes X_{u,t}\\ &= 0 + X_{s,u} \otimes X_{u,t},\end{aligned}$$

since I is an additive function. By assumption (2.5), $\mathbb{X} \in C_2^{2\alpha}$.

Conversely, if $(X, \mathbb{X})$ is an α-rough path, then we define

$$I_t := \mathbb{X}_{0,t} + X_0 \otimes X_{0,t}.$$

Then

$$\begin{aligned} I_{s,t} &= \mathbb{X}_{0,t} + X_0 \otimes X_{0,t} - (\mathbb{X}_{0,s} + X_0 \otimes X_{0,s}) \\ &\overset{\text{Chen}}{=} \mathbb{X}_{s,t} + X_{0,s} \otimes X_{s,t} + X_0 \otimes X_{0,t} - X_0 \otimes X_{0,s} \\ &= \mathbb{X}_{s,t} + X_s \otimes X_{s,t}, \end{aligned}$$

and therefore $|I_{s,t} - X_s \otimes X_{s,t}| \lesssim |t-s|^{2\alpha}$. □

Exercise 2.2. Show that if B is a d-dimensional Brownian motion, then $\mathbb{B}_{s,t} := \int_s^t B_{s,r} \otimes \mathrm{d}B_r$ (Itô integral) and $\tilde{\mathbb{B}}_{s,t} := \int_s^t B_{s,r} \otimes \circ \mathrm{d}B_r$ (Stratonovich integral) both satisfy Chen's relation.

Example 2.7. Let $\alpha \in (\frac{1}{3}, \frac{1}{2})$.

i. Let $\beta > \frac{1}{2}$ and $X \in C^\beta([0,T], \mathbb{R}^d)$. Then we could define

$$I_t := \int_0^t X_s \otimes \mathrm{d}X_s,$$

and by the estimate (1.5) for the Young integral we have

$$|I_{s,t} - X_s \otimes X_{s,t}| \lesssim \|X\|_\beta^2 |t-s|^{2\beta} \leqslant T^{2(\beta-\alpha)} \|X\|_\beta^2 |t-s|^{2\alpha}.$$

Therefore, (X, I) is an α-rough path in the sense of Lemma 2.6. However, while $I = \int_0^\cdot X_s \otimes \mathrm{d}X_s$ is a canonical choice, it is by far not the only option: Indeed, for any $Z \in C^{2\alpha}([0,T], \mathbb{R}^{d\otimes d})$ we could also define

$$I_t := \int_0^t X_s \otimes \mathrm{d}X_s + Z_t.$$

Indeed, since $Z \in C^{2\alpha}$ this I obviously still satisfies the estimate (2.5).

ii. In fact, this example shows that whenever $(X, \mathbb{X})$ is an α-rough path and a path (additive function) $Z \in C^{2\alpha}([0,T], \mathbb{R}^{d\otimes d})$, then $(X, \tilde{\mathbb{X}})$ also is an α-rough path, where

$$\tilde{\mathbb{X}}_{s,t} = \mathbb{X}_{s,t} + Z_{s,t}.$$

Conversely, if $(X, \mathbb{X})$ and $(X, \tilde{\mathbb{X}})$ are two α-rough paths with the same "first level", then

$$\delta(\mathbb{X} - \tilde{\mathbb{X}})_{s,u,t} = X_{s,u} \otimes X_{u,t} - X_{s,u} \otimes X_{u,t} = 0,$$

and therefore there exists some path (additive function) Z such that

$$\mathbb{X}_{s,t} - \tilde{\mathbb{X}}_{s,t} = Z_{s,t}$$

iii. More concretely, let us take $d = 2$ and $X \equiv 0$. Then a possible choice for I would be $I_t \equiv 0$, but we could also take

$$I_t = \begin{pmatrix} 0 & \frac{t}{2} \\ -\frac{t}{2} & 0 \end{pmatrix}.$$

If we consider

$$X_t^m = \begin{pmatrix} \frac{1}{m}\cos(m^2 t) \\ \frac{1}{m}\sin(m^2 t) \end{pmatrix}, \qquad I_t^m = \int_0^t X_s^m \otimes \mathrm{d}X_s^m,$$

then $X^m \to X$ in C^α and by Example 2.1 $I_t^m \to I_t$ for all $t \in [0,T]$. In fact one can strengthen this result and show that $(\boldsymbol{Y}^m)$ converges to $\boldsymbol{Y}$ in α-rough path topology. So by keeping track of I_t, we remember that we approximated $X \equiv 0$ by the oscillatory paths (X^m). This is reminiscent of Example 2.2, where by enhancing the state 0 to $(0,-)$ and $(0,+)$ we could keep track whether we had approached 0 from the left or from the right, respectively.

Lemma 2.8. *Let $f \in C_b^2(\mathbb{R}^d, L(\mathbb{R}^d, \mathbb{R}))$ and let $\boldsymbol{X} = (X, \mathbb{X})$ be a d-dimensional α-rough path for $\alpha \in (\frac{1}{3}, \frac{1}{2}]$. Then for all $t \in [0,T]$ the integral*

$$\int_0^t f(X_s)\mathrm{d}\boldsymbol{X}_s := \mathcal{I}\Xi_t, \qquad \textit{for } \Xi_{s,t} := f(X_s)X_{s,t} + \mathrm{D}f(X_s)\mathbb{X}_{s,t},$$

is well defined, and it is the unique function such that

$$\left|\int_s^t f(X_r)\mathrm{d}\boldsymbol{X}_r - f(X_s)X_{s,t} - \mathrm{D}f(X_s)\mathbb{X}_{s,t}\right| \lesssim \|f\|_{C_b^2}(\|X\|_\alpha^3 + \|X\|_\alpha \|\mathbb{X}\|_{2\alpha})|t-s|^{3\alpha}$$

for all $(s,t) \in \Delta_T$. If $(\boldsymbol{X}^m)_{m \in \mathbb{N}}$ is a sequence of α-rough paths converging to $\boldsymbol{Y}$ in rough path topology, then

$$\int_0^t f(X_s^m)\mathrm{d}\boldsymbol{X}_s^m \longrightarrow \int_0^t f(X_s)\mathrm{d}\boldsymbol{X}_s.$$

Proof. For the first part of the statement it suffices to combine the sewing lemma, Theorem 1.2, with the estimate (2.3). To obtain (2.3) we did not use that X was a smooth path, but only that its iterated integrals satisfy Chen's relation.

For the continuity statement we note that $\int_0^t f(X_s^m)\mathrm{d}\boldsymbol{X}_s^m - \int_0^t f(X_s)\mathrm{d}\boldsymbol{X}_s = \mathcal{I}\Xi_t^m$, where

$$\Xi_{s,t}^m = (f(X_s^m)X_{s,t}^m + \mathrm{D}f(X_s^m)\mathbb{X}_{s,t}^m - f(X_s)X_{s,t} - \mathrm{D}f(X_s)\mathbb{X}_{s,t}),$$

and therefore as in (2.3)

$$\begin{aligned}|\delta\Xi_{s,u,t}^m| \leqslant & |-f(X^m)_{s,u}X_{u,t}^m + f(X)_{s,u}X_{u,t} + \mathrm{D}f(X_s^m)X_{s,u}^m \otimes X_{u,t}^m - \mathrm{D}f(X_s)X_{s,u} \otimes X_{u,t}| \\ & + |-\mathrm{D}f(X^m)_{s,u}\mathbb{X}_{u,t}^m + \mathrm{D}f(X)_{s,u}\mathbb{X}_{u,t}|.\end{aligned}$$

By using a Taylor expansion and rebracketing like $ab - cd = a(b-d) + (a-c)d$ we obtain that

$$\begin{aligned}|\delta\Xi_{s,u,t}^m| \lesssim & \|f\|_{C_b^2}(\|X\|_\alpha^2 + \|X^m\|_\alpha^2)\|X - X^m\|_\alpha |t-s|^{3\alpha} \\ & + (\|\mathbb{X}^m\|_{2\alpha} + \|\mathbb{X}\|_{2\alpha})\|f\|_{C_b^1}\|X - X^m\|_{2\alpha}|t-s|^{3\alpha} \\ & + \|f\|_{C_b^1}(\|X\|_\alpha + \|X^m\|_\alpha)\|\mathbb{X} - \mathbb{X}^m\|_{2\alpha}|t-s|^{3\alpha},\end{aligned}$$

and the right hand side converges to 0 as $m \to \infty$. □

Example 2.9. As an application, we obtain that in the setting of Example 2.7

$$\lim_{m\to\infty}\int_0^t f\left(\frac{1}{m}\cos(m^2 s)\right)m\cos(m^2 s)\mathrm{d}s = \lim_{m\to\infty}\int_0^t f(X_s^{m,1})\mathrm{d}X_s^{m,2} = \lim_{m\to\infty}\int_0^t F(X_s^m)\mathrm{d}X_s^m,$$

where $F(x) = (\,0 \;\; f(x^1)\,)$, i.e. $F(x)x = f(x^1)x^2$. Now observe that

$$\int_0^t F(X_s^m)\mathrm{d}X_s^m = \int_0^t F(X_s^m)\mathrm{d}\boldsymbol{X}_s^m$$

because the difference in the Riemann sum approximations is $F'(X^m_s)\mathbb{X}^m_{s,t}$ which, as m is fixed and X^m is smooth, is of order $O(|t-s|^2)$ and (not uniformly in m) and therefore this contribution disappears for vanishing mesh size. Since $\boldsymbol{X}^m \to \boldsymbol{X}$ in rough path topology, where $\boldsymbol{X} = (0, I)$ with $I_t = \begin{pmatrix} 0 & \frac{t}{2} \\ -\frac{t}{2} & 0 \end{pmatrix}$, we get

$$\lim_{m\to\infty} \int_0^t f\left(\frac{1}{m}\cos(m^2 s)\right) m\cos(m^2 s)\mathrm{d}s = \int_0^t F(X_s)\mathrm{d}\boldsymbol{X}_s,$$

with $\int_0^t F(X_s)\mathrm{d}\boldsymbol{X}_s = \mathcal{I}\Xi_t$ for

$$\Xi_{s,t} = f(X^1_s)X^2_{s,t} + f'(X^1_s)\mathbb{X}^{1,2}_{s,t} = 0 = 0 + f'(0)\frac{t-s}{2}.$$

Therefore,

$$\int_0^t F(X_s)\mathrm{d}\boldsymbol{X}_s = \mathcal{I}\Xi_t = f'(0)\frac{t}{2}.$$

Of course, it would also not be difficult to compute this directly. But for differential equations driven by X^m it is more difficult to derive the limit.

Exercise 2.3. Consider some smooth paths (Z^m) with $Z^m \to Z$ in $C^{\frac{1}{2}+\varepsilon}$, where $Z \in C^{\frac{1}{2}+\varepsilon}$ is an arbitrary path. Show that $\int_0^t f(Z^m_s) m\cos(m^2 s)\mathrm{d}s \to 0$.

2.2 Controlled paths

Throughout this section we fix $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ and $T > 0$. In the previous section we defined rough paths and we showed that for any α-rough path $\boldsymbol{Y}$ we can construct the integral $\int_0^\cdot f(X_s)\mathrm{d}\boldsymbol{X}_s$ as a continuous map in α-rough path topology. But ultimately our goal is to solve integral equations

$$\mathrm{d}Y_t = b(Y_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}\boldsymbol{X}_t,$$

and the integral $\int_0^\cdot \sigma(Y_s)\mathrm{d}\boldsymbol{X}_s$ is of a different form than $\int_0^\cdot f(X_s)\mathrm{d}\boldsymbol{X}_s$, because the integrand is not just a function of X_s.

A potential solution would be to enhance our rough path $\boldsymbol{Y}$ so that it also "contains Y". This strategy works and Terry Lyons originally used it in [Lyo98], when he developed the general theory of rough paths. Here we follow instead the later approach of Gubinelli [Gub04], who extends the integral $\int_0^\cdot Y_s\mathrm{d}\boldsymbol{X}_s$ to more general integrands Y, while still keeping its continuity properties. The space of integrands should include functions $Y_s = f(X_s)$ for $f \in C^2_b$, and it should include $\sigma(Y_s)$, where Y solves our integral equation. So in particular it has to contain functions of regularity C^α. But as Example 2.1 shows, we cannot hope to have a continuous integral $\int_0^t Y_s\mathrm{d}\boldsymbol{X}_s$ for generic $Y \in C^\alpha$.

So we need to impose some structure on Y, and this structure should be richer than just requiring sufficient regularity. To understand what we need, let us recall what we used to derive the estimate (2.3) which allowed us to construct $\int_0^\cdot f(X_s)d\boldsymbol{Y}_s$ as a continuous map:

- $f(X), \mathrm{D}f(X) \in C^\alpha$;
- $|f(X)_{s,u} - \mathrm{D}f(X_s)X_{s,u}| \lesssim |u-s|^{2\alpha}$;
- $(X, \mathbb{X})$ is an α-rough path (in particular Chen's relation holds).

So whenever similar conditions hold, we could hope to apply the sewing lemma. Note that the second condition simply says that the increments of $f(X)$ are well approximated by the increments of X, times a "derivative" $\mathrm{D}f(X_s)$. This motivates the following definition:

Definition 2.10. *Let $X \in C^\alpha([0,T],\mathbb{R}^d)$. A path $Y \in C^\alpha([0,T],\mathbb{R}^m)$ is* controlled by X *if there exists $Y' \in C^\alpha([0,T], L(\mathbb{R}^d,\mathbb{R}^m))$, such that $R^Y \in C_2^{2\alpha}([0,T],\mathbb{R}^m)$, where*

$$R^Y_{s,t} := Y_{s,t} - Y'_s X_{s,t}.$$

In that case we write

$$(Y, Y') \in \mathscr{D}_X^{2\alpha}([0,T],\mathbb{R}^m)$$

or simply $(Y,Y') \in \mathscr{D}_X^{2\alpha}$, and we define

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

$\mathscr{D}_X^{2\alpha}$ is a Banach space with respect to the norm $|Y_0| + |Y_0'| + \|Y,Y'\|_{Y,2\alpha}$.

Exercise 2.4. Find an example where Y is controlled by X but Y' is not unique, i.e. there exist $Y' \neq \tilde{Y}'$ such that that $(Y,Y'), (Y,\tilde{Y}') \in \mathscr{D}_X^{2\alpha}$. (Hint: what if X is actually $C^{2\alpha}$ and not just C^α)?

Notation. *In the following we will often have estimates up to T-dependent constants. To simplify the presentation we do not keep track of them explicitly and write $\lesssim_T$ instead. But later it will be important to have a locally uniform control of the T-dependence, so by convention $a \lesssim_T b$ means $a \leqslant C(T) b$ for an increasing function $C: \mathbb{R}_+ \to \mathbb{R}_+$. For example, $|t-s|^{3\alpha} \lesssim_T |t-s|^{2\alpha}$ for $s,t \in [0,T]$*

Theorem 2.11. *Let $\boldsymbol{X}$ be a d-dimensional α-rough path and let $(Y,Y') \in \mathscr{D}_X^{2\alpha}([0,T], L(\mathbb{R}^d,\mathbb{R}^m))$. Then for all $t \in [0,T]$ the sewing integral*

$$\int_0^t Y_s \mathrm{d}\boldsymbol{X}_s := \mathcal{I}\Xi_t, \qquad \Xi_{s,t} := Y_s X_{s,t} + Y_s' \mathbb{X}_{s,t}$$

is well defined and satisfies

$$\left| \int_s^t Y_r \mathrm{d}\boldsymbol{X}_r - Y_s X_{s,t} - Y_s' \mathbb{X}_{s,t} \right| \lesssim (\|R^Y\|_{2\alpha} \|X\|_\alpha + \|Y'\|_\alpha \|\mathbb{X}\|_{2\alpha}) |t-s|^{3\alpha} \tag{2.6}$$

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

$$\mathscr{D}_X^{2\alpha}([0,T], L(\mathbb{R}^d,\mathbb{R}^m)) \ni (Y,Y') \mapsto \left(\int_0^\cdot Y_s \mathrm{d}\boldsymbol{X}_s, Y \right) \in \mathscr{D}_X^{2\alpha}([0,T],\mathbb{R}^m)$$

is a continuous linear operator and satisfies

$$\left\| \int_0^\cdot Y_s \mathrm{d}\boldsymbol{X}_s, Y \right\|_{Y,2\alpha} \lesssim_T (\|R^Y\|_{2\alpha} \|X\|_\alpha + \|Y'\|_\alpha \|\mathbb{X}\|_{2\alpha}) + \|Y'\|_\infty \|\mathbb{X}\|_{2\alpha} + \|Y\|_\alpha. \tag{2.7}$$

Proof. Estimate (2.6) follows easily from the sewing lemma and we leave its proof as an exercise (see also the derivation of (2.3)). Given (2.6), we get

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

and now we simply estimate $|t-s|^{3\alpha} \lesssim_T |t-s|^{2\alpha}$. The estimate for the derivative is trivial: $\|Y\|_\alpha \leqslant \|Y\|_\alpha$. □

Exercise 2.5. Let $\boldsymbol{X}=(X,\mathbb{X})$ be an α-rough path and let $Z\in C^{2\alpha}([0,T],\mathbb{R}^{d\otimes d})$ and $\tilde{\boldsymbol{X}}=(X,\tilde{\mathbb{X}})$ with $\tilde{\mathbb{X}}_{s,t}=\mathbb{X}_{s,t}+Z_{s,t}$. Let $(Y,Y')\in\mathscr{D}_X^{2\alpha}$ and compute

$$\int_0^t Y_s \mathrm{d}\tilde{\boldsymbol{X}}_s-\int_0^t Y_s \mathrm{d}\boldsymbol{X}_s.$$

Remark 2.12. In the setting of Theorem 2.11 let $\tilde{\boldsymbol{X}}=(\tilde{X},\tilde{\mathbb{X}})$ be another d-dimensional α-rough path, and let $(\tilde{Y},\tilde{Y}')\in\mathcal{D}_{\tilde{X}}^{2\alpha}([0,T],L(\mathbb{R}^d,\mathbb{R}^m))$. Define

$$\rho_\alpha(\boldsymbol{X},\tilde{\boldsymbol{X}})=\|X-\tilde{X}\|_\alpha+\|\mathbb{X}-\tilde{\mathbb{X}}\|_{2\alpha},$$
$$d_{X,\tilde{X},2\alpha}(Y,Y',\tilde{Y},\tilde{Y}')=\|Y'-\tilde{Y}'\|_\alpha+\|R^Y-R^{\tilde{Y}}\|_{2\alpha},$$

and $M=\max\{\|X\|_\alpha,\|\mathbb{X}\|_{2\alpha},|Y_0'|,\|Y,Y'\|_{X,2\alpha},\|\tilde{X}\|_\alpha,\|\tilde{\mathbb{X}}\|_{2\alpha},|\tilde{Y}_0'|,\|\tilde{Y},\tilde{Y}'\|_{X,2\alpha}\}$. Set

$$(Z,Z')=\left(\int_0^\cdot Y_s\mathrm{d}\boldsymbol{X}_s,Y\right),\qquad(\tilde{Z},\tilde{Z}')=\left(\int_0^\cdot \tilde{Y}_s\mathrm{d}\tilde{\boldsymbol{X}}_s,\tilde{Y}\right).$$

You show as an exercise that

$$d_{X,\tilde{X},2\alpha}(Z,Z',\tilde{Z},\tilde{Z}')\lesssim_T M(\rho_\alpha(\boldsymbol{X},\tilde{\boldsymbol{X}})+|Y_0'-\tilde{Y}_0'|+d_{X,\tilde{X},2\alpha}(Y,Y',\tilde{Y},\tilde{Y}')).$$

Exercise 2.6. Let $(Y,Y')\in\mathscr{D}_X^{2\alpha}([0,T],\mathbb{R}^m)$ and let $A\in\mathbb{R}^{k\times m}$. Show that AY is a controlled path. What is its derivative?

We have shown that controlled paths are stable under integration against $\boldsymbol{X}$. When solving an equation of the type

$$\mathrm{d}Y_t=b(Y_t)\mathrm{d}t+\sigma(Y_t)\mathrm{d}\boldsymbol{X}_t,$$

we not only need to integrate against $\boldsymbol{X}$, but we also need to apply a nonlinear map σ to a controlled path. The next theorem shows that controlled paths are stable under the application of nonlinear maps.

Theorem 2.13. *Let $X\in C^\alpha([0,T],\mathbb{R}^d)$ and let $(Y,Y')\in\mathscr{D}_X^{2\alpha}([0,T],\mathbb{R}^m)$. Let $f\in C_b^2(\mathbb{R}^m,\mathbb{R}^n)$. Then*

$$(f(Y),\mathrm{D}f(Y)Y')\in\mathscr{D}_X^{2\alpha}([0,T],\mathbb{R}^n),$$

and

$$\|f(Y),\mathrm{D}f(Y)Y'\|_{X,2\alpha}\lesssim_T(1+M)\|f\|_{C_b^2}(1+\|X\|_\alpha)^2(|Y_0'|+\|Y,Y'\|_{X,2\alpha}),\tag{2.8}$$

with $M=|Y_0'|+\|Y,Y'\|_{X,2\alpha}$. If $(\tilde{Y},\tilde{Y}')\in\mathscr{D}_X^{2\alpha}([0,T],\mathbb{R}^m)$ is another controlled path with $|\tilde{Y}_0'|+\|\tilde{Y},\tilde{Y}'\|_{Y,2\alpha}\leqslant M$ and if $f\in C_b^3$, then

$$\begin{aligned}&\|(f(Y),\mathrm{D}f(Y)Y')-(f(\tilde{Y}),\mathrm{D}f(\tilde{Y})\tilde{Y}')\|_{X,2\alpha}\\&\quad\lesssim_{T,M}\|f\|_{C_b^3}(1+\|X\|_\alpha)^2(|Y_0-\tilde{Y}_0|+|Y_0'-\tilde{Y}_0'|+\|(Y,Y')-(\tilde{Y},\tilde{Y}')\|_{X,2\alpha}).\end{aligned}\tag{2.9}$$

Proof. We show (2.8). First, we control the derivative $f(Y)'=\mathrm{D}f(Y)Y'$:

$$\begin{aligned}|(\mathrm{D}f(Y)Y')_{s,t}|&\leqslant|\mathrm{D}f(Y)_{s,t}Y_t'|+|\mathrm{D}f(Y_s)Y_{s,t}'|\\&\leqslant\|f\|_{C_b^2}\|Y\|_\alpha|t-s|^\alpha\|Y'\|_\infty+\|f\|_{C_b^1}\|Y'\|_\alpha|t-s|^\alpha\\&\leqslant\|f\|_{C_b^2}\|Y\|_\alpha|t-s|^\alpha(|Y_0'|+T^\alpha\|Y'\|_\alpha)+\|f\|_{C_b^1}\|Y'\|_\alpha|t-s|^\alpha\\&\lesssim_T\|f\|_{C_b^2}(1+\|Y\|_\alpha)(|Y_0'|+\|Y'\|_\alpha)|t-s|^\alpha.\end{aligned}$$

To bound $\|Y\|_\alpha$ note that

$$\begin{aligned}|Y_{s,t}| &\leqslant |Y_{s,t} - Y'_{s,t}X_{s,t}| + |Y'_s X_{s,t}|\\ &\lesssim_T \|R^Y\|_{2\alpha}|t-s|^\alpha + \|Y'\|_\infty \|X\|_\alpha |t-s|^\alpha\\ &\lesssim_T (|Y'_0| + \|Y,Y'\|_{X,2\alpha})(1+\|X\|_\alpha)|t-s|^\alpha\\ &\lesssim_T M(1+\|X\|_\alpha)|t-s|^\alpha.\end{aligned}$$

Next, we show that $f(Y)$ is controlled with derivative $\mathrm{D}f(Y_s)Y'_s$:

$$\begin{aligned}|f(Y)_{s,t} - \mathrm{D}f(Y_s)Y'_s X_{s,t}| &\leqslant |f(Y)_{s,t} - \mathrm{D}f(Y_s)Y_{s,t}| + |\mathrm{D}f(Y_s)Y_{s,t} - \mathrm{D}f(Y_s)Y'_s X_{s,t}|\\ &\leqslant (\|f\|_{C_b^2}\|Y\|_\alpha^2|t-s|^{2\alpha} + \|f\|_{C_b^1}\|R^Y\|_{2\alpha})|t-s|^{2\alpha}\\ &\lesssim \|f\|_{C_b^2}(1+M)(1+\|X\|_\alpha)^2(|Y'_0| + \|Y,Y'\|_{X,2\alpha})|t-s|^{2\alpha}.\end{aligned}$$

The derivation of (2.9) is more involved. Conceptually it is similar to the proof of the second estimate in Lemma 1.8, but technically it is more complex. See Lemma 7.3 in Friz-Hairer [FH14]. □

Remark 2.14. In the setting of Theorem 2.13 let $\tilde{\boldsymbol{X}} = (\tilde{X}, \tilde{\mathbb{X}})$ be another d-dimensional α-rough path, and let $(\tilde{Y}, \tilde{Y}') \in \mathscr{D}_{\tilde{X}}^{2\alpha}([0,T], L(\mathbb{R}^m))$. Let $\rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}})$, $d_{X,\tilde{X},2\alpha}(Y,Y',\tilde{Y},\tilde{Y}')$, and M be as in Remark 2.12. Set

$$(Z,Z') = (f(Y), \mathrm{D}f(Y)Y'), \qquad (\tilde{Z}, \tilde{Z}') = (f(\tilde{Y}), \mathrm{D}F(\tilde{Y})\tilde{Y}').$$

Then Theorem 7.5 of Friz-Hairer [FH14] shows that

$$d_{X,\tilde{X},2\alpha}(Z,Z',\tilde{Z},\tilde{Z}') \leqslant_{T,M} (\rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}}) + |Y_0 - \tilde{Y}_0| + |Y'_0 - \tilde{Y}'_0| + d_{X,\tilde{X},2\alpha}(Y,Y',\tilde{Y},\tilde{Y}')).$$

We now have all the ingredients that we need in order to solve rough differential equations of the type $\mathrm{d}Y_t = b(Y_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}\boldsymbol{X}_t$, $Y_0 = x$, where $\boldsymbol{X} = (X, \mathbb{X})$ is an α-rough path and we look for solutions $(Y,Y') \in \mathscr{D}_X^{2\alpha}$. For simplicity of notation we will take $b=0$ from now on, but it is not difficult to adapt the arguments to include a drift b. By definition, (Y,Y') solves the equation if

$$Y' = \sigma(Y), \qquad Y_t = x + \int_0^t \sigma(Y_s)\mathrm{d}\boldsymbol{X}_s, \qquad t \in [0,T].$$

From Theorem 2.11 we know that $\int_0^\cdot \sigma(Y_s)\mathrm{d}\boldsymbol{X}_s$ is the unique function which satisfies

$$\left|\int_s^t \sigma(Y_s)\mathrm{d}\boldsymbol{X}_s - \sigma(Y_s)X_{s,t} - \mathrm{D}\sigma(Y_s)\sigma(Y_s)\mathbb{X}_{s,t}\right| \lesssim |t-s|^{3\alpha}$$

for all $(s,t) \in \Delta_T$, where we used that $(\sigma(Y))' = \mathrm{D}\sigma(Y)Y' = \mathrm{D}\sigma(Y)\sigma(Y)$. In other words, we have the following simple observation, which often is useful:

Lemma 2.15. (Davie's formulation of rough differential equations [Dav07]) *Let $\boldsymbol{Y} = (X, \mathbb{X})$ be a d-dimensional α-rough path, let $\sigma \in C_b^2(\mathbb{R}^m, L(\mathbb{R}^d, \mathbb{R}^m))$, and let $x \in \mathbb{R}^m$. Let $Y: [0,T] \to \mathbb{R}^m$. Then $(Y, \sigma(Y)) \in \mathscr{D}_X^{2\alpha}$ and*

$$Y_t = x + \int_0^t \sigma(Y_s)\mathrm{d}\boldsymbol{X}_s, \qquad t \in [0,T],$$

if and only if $Y_0 = x$ *and for all* $(s,t) \in \Delta_T$

$$|Y_{s,t} - \sigma(Y_s)X_{s,t} - \mathrm{D}\sigma(Y_s)\sigma(Y_s)\mathbb{X}_{s,t}| \lesssim |t-s|^{3\alpha}.$$

Exercise 2.7. How does this lemma look like in the Young case? Can you find a formulation for classical ODEs which is equivalent to the formulation as a differential equation?

Theorem 2.16. *Let* $\boldsymbol{X} = (X, \mathbb{X})$ *be a d-dimensional* α*-rough path, let* $\sigma \in C_b^3(\mathbb{R}^m, L(\mathbb{R}^d, \mathbb{R}^m))$*, and let* $x \in \mathbb{R}^m$*. Then there exists a unique solution* $(Y, Y') \in \mathscr{D}_X^{2\alpha}([0,T], \mathbb{R}^m)$ *to the equation*

$$Y_t = x + \int_0^t \sigma(Y_s)\mathrm{d}\boldsymbol{X}_s, \qquad Y_t' = \sigma(Y_t), \qquad t \in [0,T].$$

Proof. Now that we know that the maps $(Y, Y') \mapsto (\sigma(Y), \mathrm{D}\sigma(Y)Y')$ and $(\sigma(Y), \mathrm{D}\sigma(Y)Y') \mapsto (\int_0^\cdot \sigma(Y_s)\mathrm{d}\boldsymbol{X}_s, \sigma(Y))$ are bounded and continuous, the proof is conceptually very similar to the one in the Young case (Theorem 1.9), although of course more technical. See Theorem 8.4 of Friz-Hairer [FH14]. □

Remark 2.17. One of the key results of rough path theory is the continuity of the Itô-Lyons map: In the setting of Theorem 2.16, write

$$Y = \Phi(x, \boldsymbol{X}).$$

It follows from Remarks 2.12 and 2.14 that if $\tilde{\boldsymbol{X}}$ is another α-rough path and if $\tilde{x} \in \mathbb{R}^m$, then

$$\|\Phi(x, \boldsymbol{X}) - \Phi(\tilde{x}, \tilde{\boldsymbol{X}})\|_{\alpha'} \lesssim_{T,M} (|x - \tilde{x}| + \rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}})),$$

where $M = \max\{|x|, |\tilde{x}|, |||\boldsymbol{X}|||_\alpha, |||\tilde{\boldsymbol{X}}|||_\alpha\}$. See Theorem 8.5 of Friz-Hairer [FH14].

Exercise 2.8. Let as in Example 2.7

$$X_t^m = \begin{pmatrix} \frac{1}{m}\cos(m^2 t) \\ \frac{1}{m}\sin(m^2 t) \end{pmatrix}, \qquad I_t^m = \int_0^t X_s^m \otimes \mathrm{d}X_s^m.$$

Let $\sigma \in C_b^3(\mathbb{R}^d, L(\mathbb{R}^2, \mathbb{R}^d))$. Let

$$Y_t^m = x + \int_0^t \sigma(Y_s^m)\partial_s X_s^m \mathrm{d}s, \qquad t \in [0,T].$$

Which equation does $Y = \lim_{m \to \infty} Y^m$ solve?

2.3 Less regular rough paths and their controlled paths

Given the definition of an α-rough path for $\alpha \in (\frac{1}{3}, \frac{1}{2}]$, there is a natural extension to less regular paths: For $\alpha > 0$ we should postulate the iterated integrals

$$\int_s^t \int_s^{r_n} \cdots \int_s^{r_2} \mathrm{d}X_{r_1} \otimes \mathrm{d}X_{r_2} \otimes \cdots \otimes \mathrm{d}X_{r_n} \in (\mathbb{R}^d)^{\otimes n},$$

for $n \leqslant \left\lfloor \frac{1}{\alpha} \right\rfloor$ and $0 \leqslant s < t \leqslant T$. Next to the canonical analytic bound

$$\left| \int_s^t \int_s^{r_n} \cdots \int_s^{r_2} \mathrm{d}X_{r_1} \otimes \mathrm{d}X_{r_2} \otimes \cdots \otimes \mathrm{d}X_{r_n} \right| \lesssim |t-s|^{n\alpha}$$

there are also algebraic considerations to take into account (for example Chen's identity). For that purpose, we first introduce the (truncated) tensor algebra. To shorten the notation we consider paths taking values in a finite-dimensional normed real vector space V from now on (because $V^{\otimes n}$ is more convenient to write than $(\mathbb{R}^d)^{\otimes n}$), and, up to choosing suitable tensor norms, all of the following can be extended[2.1] to infinite-dimensional Banach spaces V.

Definition 2.18. ((Truncated) tensor algebra)

i. The tensor algebra over V is the direct sum

$$T(V) := \bigoplus_{n=0}^{\infty} V^{\otimes n},$$

with the convention $V^{\otimes 0} := \mathbb{R}$. Recall that by definition of the direct sum, every $\boldsymbol{a} \in T(V)$ only has finitely many non-zero entries, i.e. there exists $n \in \mathbb{N}$ with $\boldsymbol{a} = (a_0, a_1, \dots, a_n, 0, 0, \dots)$. The tensor algebra comes with two operations addition and product,

$$\begin{aligned}\boldsymbol{a} + \boldsymbol{b} &= (a_0 + b_0, a_1 + b_1, \dots),\\ \boldsymbol{a} \otimes \boldsymbol{b} &= \left(a_0 \otimes b_0, a_0 \otimes b_1 + a_1 \otimes b_0, \dots, \sum_{k=0}^{n} a_k \otimes b_{n-k}, \dots\right),\end{aligned}$$

and also with scalar multiplication

$$\lambda \boldsymbol{a} = (\lambda a_0, \lambda a_1, \dots),$$

so that it is indeed a (non-commutative) algebra over $\mathbb{R}$.

ii. For $N \in \mathbb{N}_0$, the truncated tensor algebra over V is

$$T^{(N)}(V) := \bigoplus_{n=0}^{N} V^{\otimes n} \simeq \{\boldsymbol{a} \in T(V) : a_k = 0 \text{ for all } k > N\}.$$

It also comes with an addition and a product, which are defined in the same way as on $T(V)$, in the sense that $a_k \otimes b_\ell = 0$ if $k + \ell > N$.

iii. Sometimes it is also useful to consider the space of formal tensor series

$$T((V)) := \{\boldsymbol{x} = (x_0, x_1, \dots) \in \Pi_{n=0}^{\infty} V^{\otimes n}\},$$

which again is equipped with the same operations.

Lemma 2.19.

i. $\mathbf{1} = (1, 0, 0, \cdots) \in T(V)$ is a unit for multiplication.

ii. The set

$$\{\boldsymbol{a} \in T(V) : a_0 = 1\}$$

is a group with respect to multiplication, with inverse

$$\boldsymbol{a}^{-1} = \sum_{n=0}^{\infty} (\mathbf{1} - \boldsymbol{a})^{\otimes n} = (1, -a_1, -a_2, \dots).$$

But $T(V)$ is not a field, because for $v \in V \setminus \{0\}$ the element $(0, v, 0, \dots) \neq 0$ does not have an inverse with respect to multiplication.

2.1. The definition of geometric rough paths has to be done differently in that case, see Section 2.2 in [LCL07].

iii. i. and ii. also hold on $T^{(N)}(V)$.

Proof.

i.

$$\boldsymbol{a} \otimes \mathbf{1} = (a_0 \otimes 1, a_1 \otimes 1 + 0, a_2 \otimes 1 + 0, \ldots) = \boldsymbol{a}$$

and

$$\mathbf{1} \otimes \boldsymbol{a} = (1 \otimes a_0, 0 + 1 \otimes a_1, 0 + 1 \otimes a_2, \ldots) = \boldsymbol{a}.$$

ii.

$$\begin{aligned} \boldsymbol{a} \otimes \boldsymbol{a}^{-1} &= (1 \otimes 1, a_1 \otimes 1 + 1 \otimes (-a_1), a_2 \otimes 1 + 1 \otimes (-a_2), \ldots) \\ &= (1, 0, 0, \ldots) = \mathbf{1}. \end{aligned}$$

Moreover,

$$\boldsymbol{a} \otimes (0, v, 0, \ldots) = (a_0 \otimes 0, a_1 \otimes 0 + a_0 \otimes v, \ldots) = (0, a_0 \otimes v, \ldots) \neq \mathbf{1},$$

and therefore $T(V)$ is not a field.

iii. The proofs are exactly the same. □

We could equivalently[2.2] define an α-rough path as a suitable two-parameter function with values in $T((V))$ or in $T^{(N)}(V)$ with $N = \lfloor \alpha^{-1} \rfloor$. Here we will use $T^{(N)}(V)$ for simplicity, but in certain applications the perspective based on $T((V))$ is more useful.

Definition 2.20. (Homogeneous Hölder norm) *For*

$$\boldsymbol{X} = (1, \mathbb{X}^{(1)}, \ldots, \mathbb{X}^{(N)}) \colon \Delta_T \rightarrow T^{(N)}(V)$$

and $\alpha \in (0, 1]$ we define

$$|||\boldsymbol{X}|||_{\alpha} := \sum_{n=1}^{N} \|\mathbb{X}^{(n)}\|_{n\alpha}^{1/n} = \sum_{n=1}^{N} \sup_{0 \leqslant s < t \leqslant T} \frac{|\mathbb{X}_{s,t}^{(n)}|^{1/n}}{|t - s|^{\alpha}} \in [0, \infty].$$

We say that $\boldsymbol{X}$ is α-Hölder continuous if $|||\boldsymbol{X}|||_{\alpha} < \infty$.

Definition 2.21. (α-rough path) *Let $\alpha \in (0, 1]$ and $N = \lfloor \alpha^{-1} \rfloor$. A V-valued* α-rough path *is a map*

$$\boldsymbol{X} = (1, \mathbb{X}^{(1)}, \mathbb{X}^{(2)}, \cdots, \mathbb{X}^{(N)}) \colon \Delta_T \rightarrow T^{(N)}(V)$$

satisfying

i. Chen's relation:

$$\boldsymbol{X}_{s,t} = \boldsymbol{X}_{s,u} \otimes \boldsymbol{X}_{u,t}, \qquad 0 \leqslant s \leqslant u \leqslant t \leqslant T,$$

ii. α-Hölder continuity:

$$|||\boldsymbol{X}|||_{\alpha} < \infty.$$

We say that a sequence of α-rough paths $(\boldsymbol{X}^m)$ converges to $\boldsymbol{X}$ in α-rough path topology if

$$\lim_{m \rightarrow \infty} |||\boldsymbol{X}^m - \boldsymbol{X}|||_{\alpha} = 0.$$

2.2. This equivalence is far from obvious and it is related to the fact that given the first N iterated integrals, we can uniquely find all the higher order iterated integrals, a result which is known as Lyons universal extension theorem [LCL07].

Exercise 2.9. Show that if $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ and $\boldsymbol{X} = (X, \mathbb{X})$, then setting

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

defines an α-rough path in the sense of our new definition if and only if $\boldsymbol{X}$ is an α-rough path in the sense of our old definition.

Exercise 2.10. Let $X \in C^1([0,T], V)$ and define $\mathbb{X}_{s,t}^{(1)} = X_{s,t}$ and

$$\mathbb{X}_{s,t}^{(n)} = \int_s^t \int_s^{r_n} \dots \int_s^{r_2} \mathrm{d}X_{r_1} \otimes \dots \otimes \mathrm{d}X_{r_n}.$$

Show that $\boldsymbol{X} = (1, \mathbb{X}^{(1)}, \dots, \mathbb{X}^{(N)})$ satisfies Chen's relation. (See the next section for the solution).

To solve differential equations driven by $\boldsymbol{X}$, we also need to extend the notion of controlled paths to our new setting. For simplicity, we focus on real-valued paths. But this extends immediately to paths taking values in finite-dimensional vector spaces by arguing componentwise. Intuitively, we might expect that we need a higher order description of controlled paths in terms of the controlling rough paths, which may also involve higher order levels of the rough path. Also, we need to assume that the derivative paths are themselves controlled, to a lesser degree.

To simplify the bookkeeping, it now becomes useful to identify a basis of $T^{(N)}(V^*)$ with words: The set of all *words* (over the *alphabet* $\{1, \dots, d\}$) is the set of ordered tuples

$$\mathcal{W} = \{(i_1, \dots, i_k) : k \geqslant 1, i_1, \dots, i_k \in \{1, \dots, d\}\} \cup \{\emptyset\},$$

and we say that a word $w = (i_1, \dots, i_k)$ has *length* k and we write $|w| = k$, with $|\emptyset| = 0$. For example, $(1,2,3)$, $(1,1)$ and $(2,1,1,3,1)$ are words over the alphabet $\{1,2,3\}$, of length 3, 2 and 5, respectively.

Let now $e_1, \dots, e_d$ be a basis of V. Given a word $w = (i_1, \dots, i_k) \in \mathcal{W}$ we write

$$e_w := e_{i_1} \otimes \cdots \otimes e_{i_k} \in V^{\otimes k}, \qquad e_\emptyset = 1 \in V^{\otimes 0}.$$

We can then identify w with an element of $T^{(N)}(V^*)$ by setting

$$\langle w, e_{w'} \rangle = \delta_{w,w'}$$

for all words w' of length $\leqslant N$. Given two words $w = (i_1, \dots, i_k)$ and $w' = (j_1, \dots, j_\ell)$ we define the concatenation

$$w w' := (i_1, \dots, i_k, j_1, \dots, j_\ell).$$

With this notation we have for example

$$\langle w, \boldsymbol{a} \otimes \boldsymbol{b} \rangle = \sum_{\substack{v, v' \in \mathcal{W}: \\ v v' = w}} \langle v, \boldsymbol{a} \rangle \langle v', \boldsymbol{b} \rangle. \tag{2.10}$$

We also write

$$\mathbb{X}^w := \langle w, \mathbb{X} \rangle.$$

This notation allows for the the following short and elegant definition:

Definition 2.22. (Controlled path) *Let $\boldsymbol{X} = (1, \mathbb{X}^{(1)}, \mathbb{X}^{(2)}, \cdots, \mathbb{X}^{(N)}) : \Delta_T \to T^{(N)}(V)$ be a V-valued α-rough path. A function $\boldsymbol{Y} : [0,T] \to T^{(N-1)}(V^*)$ is called* controlled *by $\boldsymbol{X}$ if there exists $C > 0$ such that for all words w of length $|w| \leqslant N - 1$ well as for all $(s,t) \in \Delta_T$:*

$$|\langle \boldsymbol{Y}_t, e_w \rangle - \langle \boldsymbol{Y}_s, \boldsymbol{X}_{s,t} \otimes e_w \rangle| \leqslant C |t - s|^{(N - |w|)\alpha},$$

where by truncation we interpret $\boldsymbol{X}_{s,t} \otimes e_w$ *as an element of* $(T^{(N-1)}(V))^* \simeq T^{(N-1)}(V^*)$. *In that case we write*

$$\|\boldsymbol{Y}\|_{\mathscr{D}_{\boldsymbol{X}}^{N\alpha}}$$

for the smallest possible C *that works for all* $|w| \leqslant N-1$ *and* $(s,t) \in \Delta_T$, *and we also write*

$$\boldsymbol{Y} \in \mathscr{D}_{\boldsymbol{X}}^{N\alpha} := \mathscr{D}_{\boldsymbol{X}}^{N\alpha}([0,T], \mathbb{R}).$$

Exercise 2.11. Show that for $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ and $N = 2$ this definition is consistent with our first definition, if given (Y, Y') and $(X, \mathbb{X})$ we set $\boldsymbol{Y}_t = (Y_t, Y_t') \in (V^*)^{\otimes 0} \oplus (V^*)^{\otimes 1}$ and $\boldsymbol{X}_{s,t} = (1, X_{s,t}, \mathbb{X}_{s,t})$.

It is obvious from the linearity of the definition that $\mathscr{D}_{\boldsymbol{X}}^{N\alpha}$ is a vector space. Moreover, for all $\boldsymbol{Y} \in \mathscr{D}_{\boldsymbol{X}}^{N\alpha}$ we can define the controlled rough path integral as another element of $\mathscr{D}_{\boldsymbol{X}}^{N\alpha}$.

Proposition 2.23. (Controlled rough path integral) *Let* $\alpha \in (0,1]$, *let* $N = \lfloor \alpha^{-1} \rfloor$ *and let* $\boldsymbol{X}$ *be an* α*-rough path and let* $\boldsymbol{Y} \in \mathscr{D}_{\boldsymbol{X}}^{N\alpha}$. *Define for* $(s,t) \in \Delta_T$ *and* $i \in \{1, \ldots, d\}$

$$\Xi_{s,t}^i := \langle \boldsymbol{Y}_s i, \boldsymbol{X}_{s,t} \rangle := \sum_{\substack{w \in \mathcal{W}: \\ |w| \leqslant N-1}} \langle \boldsymbol{Y}_s, e_w \rangle \langle wi, \boldsymbol{X}_{s,t} \rangle.$$

Then Ξ *satisfies the assumptions of the sewing lemma and we call*

$$\int_0^t \boldsymbol{Y}_s \mathrm{d}\boldsymbol{X}_s^i := \mathcal{I}\Xi_t^i \in \mathbb{R}, \qquad t \in [0,T]$$

the controlled rough path integral of $\boldsymbol{Y}$ *against* $\boldsymbol{X}^i$. *Setting*

$$\langle \boldsymbol{I}_t^i, e_w \rangle := \begin{cases} \int_0^t \boldsymbol{Y}_s \mathrm{d}\boldsymbol{X}_s^i, & w = \emptyset, \\ \langle \boldsymbol{Y}_s, e_{w'} \rangle, & w = w'i, \\ 0, & w = w'j \text{ with } j \neq i, \end{cases}$$

we obtain a new element of $\mathscr{D}_{\boldsymbol{X}}^{N\alpha}$ *and*

$$\|\boldsymbol{I}^i\|_{\mathscr{D}_{\boldsymbol{X}}^{N\alpha}} \lesssim_T (\|\boldsymbol{Y}\|_\infty + \|\boldsymbol{Y}\|_{\mathscr{D}_{\boldsymbol{X}}^{N\alpha}}) \|\!| X |\!\|_\alpha.$$

Proof. Clearly Ξ^i is continuous and $\Xi_{t,t}^i = 0$ because $\boldsymbol{X}_{t,t} = (1, 0, \ldots, 0)$ (which follows from the Hölder assumption on $\boldsymbol{X}$). Moreover,

$$\delta\Xi_{s,u,t}^i = \sum_{\substack{w \in \mathcal{W}: \\ |w| \leqslant N-1}} (\langle \boldsymbol{Y}_s, e_w \rangle \langle wi, \boldsymbol{X}_{s,t} \rangle - \langle \boldsymbol{Y}_s, e_w \rangle \langle wi, \boldsymbol{X}_{s,u} \rangle - \langle \boldsymbol{Y}_u, e_w \rangle \langle wi, \boldsymbol{X}_{u,t} \rangle)$$

and using the controlled structure of $\boldsymbol{Y}$

$$\begin{aligned} |(\langle \boldsymbol{Y}_u, e_w \rangle - \langle \boldsymbol{Y}_s, \boldsymbol{X}_{s,u} \otimes e_w \rangle) \langle wi, \boldsymbol{X}_{u,t} \rangle| &\leqslant \|\boldsymbol{Y}\|_{\mathscr{D}_{\boldsymbol{X}}^{N\alpha}} |u-s|^{(N-|w|)\alpha} |\langle wi, \boldsymbol{X}_{u,t} \rangle| \\ &\leqslant \|\boldsymbol{Y}\|_{\mathscr{D}_{\boldsymbol{X}}^{N\alpha}} |u-s|^{(N-|w|)\alpha} \|\!| \boldsymbol{X} |\!\|_\alpha |t-u|^{|w|+1} \\ &\leqslant \|\boldsymbol{Y}\|_{\mathscr{D}_{\boldsymbol{X}}^{N\alpha}} \|\!| \boldsymbol{X} |\!\|_\alpha |t-s|^{(N+1)\alpha} \end{aligned}$$

with $(N+1)\alpha > 1$. Next, we will show that

$$\sum_{\substack{w \in \mathcal{W}: \\ |w| \leqslant N-1}} (\langle \boldsymbol{Y}_s, e_w \rangle \langle wi, \boldsymbol{X}_{s,t} \rangle - \langle \boldsymbol{Y}_s, e_w \rangle \langle wi, \boldsymbol{X}_{s,u} \rangle - \langle \boldsymbol{Y}_s, \boldsymbol{X}_{s,u} \otimes e_w \rangle \langle wi, \boldsymbol{X}_{u,t} \rangle) = 0,$$

from where the existence of the controlled rough path integral follows with the sewing lemma. Indeed, observe that

$$\begin{aligned}
&\sum_{\substack{w\in\mathcal{W}:\\|w|\leqslant N-1}} \langle \boldsymbol{Y}_s, \boldsymbol{X}_{s,u}\otimes e_w\rangle\langle wi, \boldsymbol{X}_{u,t}\rangle\\
&= \sum_{\substack{w\in\mathcal{W}:\\|w|\leqslant N-1}} \sum_{\substack{w'\in\mathcal{W}:\\|w'|\leqslant N-1-|w|}} \langle \boldsymbol{Y}_s, e_{w'}\otimes e_w\rangle\langle w', \boldsymbol{X}_{s,u}\rangle\langle wi, \boldsymbol{X}_{u,t}\rangle\\
&= \sum_{\substack{w\in\mathcal{W}:\\|w|\leqslant N-1}} \langle \boldsymbol{Y}_s, e_w\rangle \sum_{\substack{v,v'\in\mathcal{W}:\\vv'=w}} \langle v, \boldsymbol{X}_{s,u}\rangle\langle v'i, \boldsymbol{X}_{u,t}\rangle\\
&= \sum_{\substack{w\in\mathcal{W}:\\|w|\leqslant N-1}} \langle \boldsymbol{Y}_s, e_w\rangle\left(\left(\sum_{\substack{v,v'\in\mathcal{W}:\\vv'=wi}} \langle v, \boldsymbol{X}_{s,u}\rangle\langle v'i, \boldsymbol{X}_{u,t}\rangle\right) - \langle wi, \boldsymbol{X}_{s,u}\rangle\right)\\
&= \sum_{\substack{w\in\mathcal{W}:\\|w|\leqslant N-1}} \langle \boldsymbol{Y}_s, e_w\rangle(\langle wi, \boldsymbol{X}_{s,u}\otimes\boldsymbol{X}_{u,t}\rangle - \langle wi, \boldsymbol{X}_{s,u}\rangle)
\end{aligned}$$

by (2.10), which yields with Chen's relation $\boldsymbol{X}_{s,u}\otimes\boldsymbol{X}_{u,t}=\boldsymbol{X}_{s,t}$:

$$\begin{aligned}
\delta\Xi^i_{s,u,t} &= \sum_{\substack{w\in\mathcal{W}:\\|w|\leqslant N-1}} \langle \boldsymbol{Y}_s, e_w\rangle(\langle wi, \boldsymbol{X}_{s,t}\rangle - \langle wi, \boldsymbol{X}_{s,u}\rangle - (\langle wi, \boldsymbol{X}_{s,u}\otimes\boldsymbol{X}_{u,t}\rangle - \langle wi, \boldsymbol{X}_{s,u}\rangle))\\
&= 0.
\end{aligned}$$

The sewing lemma now yields the bound

$$\begin{aligned}
|\mathcal{I}\Xi^i_t - \mathcal{I}\Xi^i_s - \Xi^i_{s,t}| &\lesssim \sup_{0\leqslant u_1<u_2<u_3\leqslant T} \frac{|\delta\Xi^i_{u_1,u_2,u_3}|}{|u_3-u_1|^{(N+1)\alpha}}|t-s|^{(N+1)\alpha}\\
&\lesssim \|\boldsymbol{Y}\|_{\mathscr{D}^{N\alpha}_{\boldsymbol{X}}}|||\boldsymbol{X}|||_\alpha|t-s|^{(N+1)\alpha},
\end{aligned}$$

and plugging in our definitions, this means

$$\begin{aligned}
|\langle \boldsymbol{I}^i_t, e_\emptyset\rangle - \langle \boldsymbol{I}^i_s, \boldsymbol{X}_{s,t}\otimes e_\emptyset\rangle| &= |\langle \boldsymbol{I}^i_t, e_\emptyset\rangle - \langle \boldsymbol{I}^i_s, \boldsymbol{X}_{s,t}\rangle|\\
&= \left|\mathcal{I}\Xi^i_t - \mathcal{I}\Xi^i_s - \sum_{\substack{w'\in\mathcal{W}:\\1\leqslant|w'|\leqslant N-2}} \langle \boldsymbol{Y}_s, e_{w'}\rangle\langle w'i, \boldsymbol{X}_{s,t}\rangle\right|\\
&\leqslant |\mathcal{I}\Xi^i_t - \mathcal{I}\Xi^i_s - \Xi^i_{s,t}| + \left|\sum_{\substack{w'\in\mathcal{W}:\\|w'|=N-1}} \langle \boldsymbol{Y}_s, e_{w'}\rangle\langle w'i, \boldsymbol{X}_{s,t}\rangle\right|\\
&\lesssim \|\boldsymbol{Y}\|_{\mathscr{D}^{N\alpha}_{\boldsymbol{X}}}|||\boldsymbol{X}|||_\alpha|t-s|^{(N+1)\alpha} + \|\boldsymbol{Y}\|_\infty|||\boldsymbol{X}|||_\alpha|t-s|^{N\alpha}.
\end{aligned}$$

This is the claimed bound for $w=\emptyset$. For $w=w'j$ with $j\neq i$ it suffices to note that

$$|\langle \boldsymbol{I}^i_t, e_{wj}\rangle - \langle \boldsymbol{I}^i_s, \boldsymbol{X}_{s,t}\otimes e_{wj}\rangle| = 0,$$

while for $w=w'i$

$$\begin{aligned}
|\langle \boldsymbol{I}^i_t, e_{w'i}\rangle - \langle \boldsymbol{I}^i_s, \boldsymbol{X}_{s,t}\otimes e_{w'i}\rangle| &= |\langle \boldsymbol{Y}_s, e_{w'}\rangle - \langle \boldsymbol{Y}_s, \boldsymbol{X}_{s,t}\otimes e_{w'}\rangle|\\
&\leqslant \|\boldsymbol{Y}\|_{\mathscr{D}^{N\alpha}_{\boldsymbol{X}}}|t-s|^{(N-|w'|)\alpha}\\
&\leqslant T^\alpha\|\boldsymbol{Y}\|_{\mathscr{D}^{N\alpha}_{\boldsymbol{X}}}|t-s|^{(N-|w|)\alpha}.
\end{aligned}$$

□

With these definitions and results we can solve *linear* rough differential equations driven by low regularity rough paths: For another finite-dimensional (say m-dimensional) normed vector space W let $A \in L(V, L(W, W))$. Then we can solve the rough differential equation

$$\mathrm{d}Y_t = A Y_t \mathrm{d}\boldsymbol{X}_t, \qquad Y_0 = x \in W,$$

which we interpret as $A(\mathrm{d}\boldsymbol{X}_t)Y_t$ or, coordinatewise, as $A \in \mathbb{R}^{m\times m\times d}$ with

$$\mathrm{d}Y_t^i = \sum_{j=1}^{m} \sum_{k=1}^{d} A_{i,j,k} Y_t^j \mathrm{d}\boldsymbol{X}_t^k.$$

Similarly as before, we can set up a Picard iteration and solve this equation globally in time because the length of the time interval for the contraction does not depend on the initial condition.

But what if we want to solve nonlinear rough differential equations? The problem already arises in one dimension, so say we are given $\sigma \in C_b^{N+1}(\mathbb{R}, \mathbb{R})$ and $x \in \mathbb{R}$. Can we solve the equation

$$Y_t = x + \int_0^t \sigma(\boldsymbol{Y}_s) \mathrm{d}\boldsymbol{X}_s, \qquad t \in [0, T],$$

with an appropriate interpretation? Maybe surprisingly, the answer to this question is *in general no*. The problem is that if $\boldsymbol{Y} \in \mathscr{D}_{\boldsymbol{X}}^{N\alpha}$, then we do not know if $\sigma(\boldsymbol{Y}) \in \mathscr{D}_{\boldsymbol{X}}^{N\alpha}$. Indeed, writing $\boldsymbol{Y} = (Y, \mathbb{Y}^{(1)}, \ldots, \mathbb{Y}^{(N-1)})$ and $\boldsymbol{X} = (1, X, \mathbb{X}^{(2)}, \ldots, \mathbb{X}^{(N)})$, we would for example like to set

$$\langle \sigma(\boldsymbol{Y}), e_\emptyset \rangle = \sigma(Y),$$

which satisfies by a Taylor expansion

$$\sigma(Y_t) \;=\; \sum_{k \leqslant N-1} \frac{1}{k!} \sigma^{(k)}(Y_s)(Y_{s,t})^k + \underbrace{O(|Y_{s,t}|^N)}_{=O(|t-s|^{N\alpha})}.$$

But even if $N = 3$ and $k = 2$ we usually cannot express $(Y_{s,t})^2$ as a linear function of $\boldsymbol{X}_{s,t}$ plus a remainder which vanishes at the right order. Indeed,

$$Y_{s,t} = \mathbb{Y}_s^{(1)} X_{s,t} + O(|t-s|^{2\alpha}),$$

so that

$$(Y_{s,t})^2 = \big(\mathbb{Y}_s^{(1)}\big)^2 (X_{s,t})^2 + O(|t-s|^{3\alpha}),$$

and there is no reason why this should be expressible as a linear function of $\mathbb{X}$, which we would need to make $\sigma(Y)$ a controlled path.

Of course, since $\boldsymbol{X} = (1, X, \mathbb{X}^{(2)}, \ldots, \mathbb{X}^{(N)})$ corresponds to a one-dimensional path and morally $\mathbb{X}_{s,t}^{(2)} = \int_s^t X_{s,r} \mathrm{d}X_r$, we might suspect an integration by parts rule,

$$\frac{1}{2}(X_{s,t})^2 = \int_s^t X_{s,r} \mathrm{d}X_r = \mathbb{X}_{s,t}^{(2)}.$$

While this is true if $\boldsymbol{X}$ is built from a smooth path X, we did not make such an assumption in our definition of a rough path. And indeed we discussed Brownian motion with its Itô iterated integrals as an example of a rough path, for which the integration by parts rule fails due to the Itô corrector,

$$\frac{1}{2}(B_{s,t})^2 = \int_s^t B_{s,r} \mathrm{d}B_r + \frac{1}{2}(t-s).$$

To proceed, we need to either enhance the rough path with more data, postulating for example also

$$\int_s^t (X_{s,r})^2 \mathrm{d}X_r.$$

This leads to *branched* rough paths which are no longer indexed by words but by decorated rooted trees [Gub10] and which have a more complicated (Hopf) algebraic structure. Or we give up the example of Itô Brownian motion and we postulate an integration by parts rule. This is the approach chosen in the original work by Lyons [Lyo98] and it leads to geometric rough paths.

2.4 Geometric rough paths

To understand how to formulate the integration by parts rule for the iterated integrals, we introduce the shuffle product on words[2.3]: The shuffle product

$$w \sqcup w'$$

consists of the sum over all possible ways of interweaving w and w' in such a way that the order of letters in each word remains unchanged. For example, (writing $ab := (a, b)$)

$$a \sqcup x = ax + xa, \qquad ab \sqcup x = abx + axb + xab,$$
$$ab \sqcup xy = abxy + axby + axyb + xaby + xayb + xyab.$$

The first realization of the integration by parts rule would be

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

which can be expressed with the shuffle product as

$$\langle \boldsymbol{X}_{s,t}, i \rangle \langle \boldsymbol{X}_{s,t}, j \rangle = \langle \boldsymbol{X}_{s,t}, i \sqcup j \rangle.$$

Thus, we have for all words w, w' of length $|w|, |w'| \leqslant 1$ (the case $|w| = 0$ or $|w'| = 0$ is easy as $\langle \boldsymbol{X}_{s,t}, \emptyset \rangle = 1$):

$$\langle \boldsymbol{X}_{s,t}, w \rangle \langle \boldsymbol{X}_{s,t}, w' \rangle = \langle \boldsymbol{X}_{s,t}, w \sqcup w' \rangle.$$

Exercise 2.12. Let $X \in C^1([0, T], \mathbb{R}^d)$ and let $\boldsymbol{X} = (1, \mathbb{X}^{(1)}, \ldots, \mathbb{X}^{(N)})$ with

$$\mathbb{X}^{i_1 \ldots i_k}_{s,t} := \int_{0 < r_1 < \ldots < r_k < t} \mathrm{d}X^{i_1}_{r_1} \cdots \mathrm{d}X^{i_k}_{r_k}.$$

Show that $\boldsymbol{X}$ satisfies for all words w, w' with $|w| + |w'| \leqslant N$:

$$\langle \boldsymbol{X}_{s,t}, w \rangle \langle \boldsymbol{X}_{s,t}, w' \rangle = \langle \boldsymbol{X}_{s,t}, w \sqcup w' \rangle.$$

Hint: Use induction over $|w| + |w'|$.

Definition 2.24. (Geometric rough path) *Let* $\alpha \in (0, 1]$ *and* $N = \lfloor \alpha^{-1} \rfloor$. *A V-valued* α*-rough path* $\boldsymbol{X} = (1, \mathbb{X}^{(1)}, \mathbb{X}^{(2)}, \cdots, \mathbb{X}^{(N)}) : \Delta_T \to T^{(N)}(V)$ *is called* geometric *if for all* w, $w' \in W$ *with* $|w| + |w'| \leqslant N$:

$$\langle \boldsymbol{X}_{s,t}, w \rangle \langle \boldsymbol{X}_{s,t}, w' \rangle = \langle \boldsymbol{X}_{s,t}, w \sqcup w' \rangle.$$

2.3. The symbol should be different: $\sqcup$ should have a third upward line in the middle. I do not know at the moment how to produce such a symbol in Texmacs.

Remark.

i. One can show that geometric rough paths take their values in a Lie group $G^{(N)}(V) \subset T^{(N)}(V)$. The monograph [FV10] emphasizes this perspective.

ii. In the literature, our geometric rough paths are usually called weakly geometric. Strongly geometric paths are those which can be approximated by smooth paths. One can show that if $\alpha' < \alpha$ is such that $\lfloor (\alpha')^{-1} \rfloor = \lfloor \alpha^{-1} \rfloor$, then every weakly geometric α-rough path is a strongly geometric α'-rough path. Therefore, we do not distinguish the two notions here.

Using Taylor expansions, we could now show that if $\boldsymbol{X}$ is a geometric α-rough path, if $\boldsymbol{Y} \in \mathscr{D}_{\boldsymbol{X}}^{N\alpha}([0,T], \mathbb{R}^m)$, and if $\sigma \in C_b^{N-1}(\mathbb{R}^m, \mathbb{R}^n)$, then $f(\boldsymbol{Y}^{(0)})$ can be naturally lifted to an element of $\mathscr{D}_{\boldsymbol{X}}^{N\alpha}([0,T], \mathbb{R}^n)$. If even $\sigma \in C_b^N(\mathbb{R}^m, \mathbb{R}^n)$, then we have similar continuity estimates as in Theorem 2.13. The key point in the proof is that due to the shuffle product rule, every polynomial function of $\boldsymbol{X}_{s,t}$ (such as the ones appearing in the Taylor expansion of $f(\boldsymbol{Y}_t^{(0)})$) can be expressed as a *linear* function of $\boldsymbol{X}_{s,t}$. Then the solution theory for nonlinear rough differential equations is very similar to the case $\alpha \in (\frac{1}{3}, \frac{1}{2}]$

3 First applications of rough paths

3.1 Stochastic processes as rough paths

We have seen that if $X \in C^\alpha([0,T], \mathbb{R}^d)$, then there is no unique choice of a second order process $\mathbb{X}$ which turns $(X, \mathbb{X})$ into an α-rough path. Indeed it is not even obvious whether such a $\mathbb{X}$ exists at all (by the *Lyons-Victoir extension theorem* it does, but we will not prove this). However, if X is a stochastic process, then there often is a canonical choice for $\mathbb{X}$.

3.1.1 Brownian motion

Let us start with the easiest example, where B is a d-dimensional standard Brownian motion on $[0,T]$. In that case we have almost surely $B \in C^\alpha([0,T], \mathbb{R}^d)$ whenever $\alpha < 1/2$, and we define

$$\mathbb{B}_{s,t}^{\mathrm{Ito}} = \int_s^t B_r \otimes \mathrm{d}B_r - B_s \otimes B_{s,t},$$

where the stochastic integral $\mathrm{d}B_r$ is understood in the Itô sense; in particular, $\mathbb{B}$ is continuous. By construction we also have Chen's relation. It thus only remains to show that almost surely $|\mathbb{B}_{s,t}^{\mathrm{Ito}}| \lesssim |t-s|^{2\alpha}$. This is a consequence of the following result.

Theorem 3.1. (Kolmogorov's continuity criterion for rough paths) *Let $(X, \mathbb{X})$ be a stochastic process which almost surely satisfies Chen's relation. Assume that there exist $p \geqslant 2$, $\beta > \frac{1}{p}$, $C > 0$, such that for all $(s,t) \in \Delta_T$*

$$\mathbb{E}[|X_{s,t}|^p]^{1/p} \leqslant C|t-s|^\beta, \qquad \mathbb{E}[|\mathbb{X}_{s,t}|^{p/2}]^{2/p} \leqslant C|t-s|^{2\beta}.$$

Then there exists a modification $\boldsymbol{X} = (\tilde{X}, \tilde{\mathbb{X}})$ of $(X, \mathbb{X})$ satisfying Chen's relation, and such that

$$\mathbb{E}[|||\tilde{\boldsymbol{X}}|||_\alpha^p]^{1/p} \lesssim C$$

for all $\alpha \in (0, \beta - \frac{1}{p})$.

Proof. See Friz-Hairer [FH14], Theorem 3.1. □

To apply this to $(B,\mathbb{B}^{\text{Ito}})$ we only need to bound $\mathbb{E}[|\mathbb{B}^{\text{Ito}}_{s,t}|^{p/2}]$ for sufficiently large p. We apply the Burkholder-Davis-Gundy inequality twice to obtain

$$\mathbb{E}[|\mathbb{B}^{\text{Ito}}_{s,t}|^{p/2}] \simeq \mathbb{E}\bigg[\bigg(\int_s^t |B_r - B_s|^2 \mathrm{d}s\bigg)^{p/4}\bigg] \leqslant \mathbb{E}\bigg[\sup_{r\in[s,t]} |B_r - B_s|^{p/2}\bigg]|t-s|^{p/4}$$
$$\simeq |t-s|^{p/4} \times |t-s|^{p/4} = |t-s|^{p/2}.$$

Moreover, $\mathbb{E}[|B_{s,t}|^p] \simeq |t-s|^{p/2}$, by another application of the Burkholder-Davis-Gundy inequality or alternatively because B is Gaussian. Taking $p>6$ we obtain that $\boldsymbol{B}^{\text{Ito}} = (B, \mathbb{B}^{\text{Ito}})$ is almost surely an α-rough path for any $\alpha \in (\frac{1}{3}, \frac{1}{2} - \frac{1}{p})$, which is a non-empty interval because $p>6$. Note that we do not have to take a modification of $(B,\mathbb{B}^{\text{Ito}})$, because the pair is already continuous.

We also define the Stratonovich iterated integrals of $\mathbb{B}$ by

$$\mathbb{B}^{\text{Strat}}_{s,t} := \int_s^t (B_r - B_s) \otimes \circ \mathrm{d}B_r = \mathbb{B}^{\text{Ito}}_{s,t} + \frac{1}{2}\mathbb{I}_d (t-s),$$

where $\mathbb{I}_d$ is the identity matrix on $\mathbb{R}^d$ and $\circ \mathrm{d}B_r$ denotes the Stratonovich integral. We set $\boldsymbol{B}^{\text{Strat}} = (B, \mathbb{B}^{\text{Strat}})$.

Theorem 3.2. *If (Y,Y') is an adapted process such that almost surely $(Y,Y') \in \mathscr{D}^{2\alpha}_B$, then almost surely*

$$\int_0^\cdot Y_s \mathrm{d}\boldsymbol{B}^{\text{Ito}}_s = \int_0^\cdot Y_s \mathrm{d}B_s, \qquad \int_0^\cdot Y_s \mathrm{d}\boldsymbol{B}^{\text{Strat}}_s = \int_0^\cdot Y_s \circ \mathrm{d}B_s,$$

where the left hand sides denote the rough path integrals of Y with respect to $\boldsymbol{B}^{\text{Ito}}$ and $\boldsymbol{B}^{\text{Strat}}$ respectively, and the right hand sides denote the Itô and the Stratonovich integral (for the Stratonovich integral we should also assume that Y is a semimartingale).

Proof. By stopping in

$$\tau_n = \inf\{t \geqslant 0 : \|Y,Y'\|_{\mathscr{D}^{2\alpha}_B([0,t])} \geqslant n\},$$

we may assume that $\mathbb{E}[\|Y,Y'\|^2_{B,2\alpha}] < \infty$. For the Itô integral, we estimate

$$\mathbb{E}\bigg[\bigg|\int_s^t Y_r \mathrm{d}B_r - Y_s B_{s,t} - Y'_s \mathbb{B}_{s,t}\bigg|^2\bigg]^{1/2} = \mathbb{E}\bigg[\bigg|\int_s^t (Y_r - Y_s - Y'_s B_{s,r}) \mathrm{d}B_r\bigg|^2\bigg]^{1/2}$$
$$= \mathbb{E}\bigg[\int_s^t |Y_r - Y_s - Y'_s B_{s,r}|^2 dr\bigg]^{1/2}$$
$$\leqslant \mathbb{E}[\|Y,Y'\|^2_{B,2\alpha}]^{1/2} \bigg(\int_s^t |r-s|^{4\alpha} dr\bigg)^{1/2}$$
$$\simeq \mathbb{E}[\|Y,Y'\|^2_{B,2\alpha}]^{1/2} |t-s|^{\frac{1}{2}+2\alpha}.$$

Since $\frac{1}{2} + 2\alpha > 1$, this gives for fixed $t \in [0,T]$ and with $t^n_k = kt/n$:

$$\mathbb{E}\bigg[\bigg|\int_0^t Y_s \mathrm{d}B_s - \sum_{k=0}^{n-1} (Y_{t^n_k} B_{t^n_k, t^n_{k+1}} + Y'_{t^n_k} \mathbb{B}_{t^n_k, t^n_{k+1}})\bigg|^2\bigg]^{1/2} \lesssim \mathbb{E}[\|Y,Y'\|^2_{B,2\alpha}]^{1/2} n \bigg|\frac{t}{n}\bigg|^{\frac{1}{2}+2\alpha} \to 0.$$

Therefore, $\mathcal{I}_n = \sum_{k=0}^{n-1} (Y_{t^n_k} B_{t^n_k, t^n_{k+1}} + Y'_{t^n_k} \mathbb{B}_{t^n_k, t^n_{k+1}})$ converges in L^2 to $\int_0^t Y_s \mathrm{d}B_s$. But it also converges almost surely to $\int_0^t Y_s \mathrm{d}\boldsymbol{B}^{\text{Ito}}_s$, and therefore almost surely $\int_0^t Y_s \mathrm{d}B_s = \int_0^t Y_s \mathrm{d}\boldsymbol{B}^{\text{Ito}}_s$. A priori the null set depends on t, but both processes are continuous, and therefore they are indistinguishable.

The Stratonovich integral can be rewritten in terms of the Itô integral, and this can be used to establish the claim also follows in that case. See Section 5 of Friz-Hairer [FH14] for details. □

Exercise 3.1. Let $Y_t = x + \int_0^t \sigma(Y_s) \mathrm{d}\boldsymbol{B}_s^{\mathrm{Ito}}$ and determine which "Stratonovich equation" Y solves (i.e. derive an equation for Y which involves only an integral against $\boldsymbol{B}^{\mathrm{Strat}}$)

3.1.2 Fractional Brownian motion

What follows is not part of the videos and not relevant for the exam.

Recall that B is a d-dimensional fractional Brownian motion with Hurst parameter $H \in (0,1)$ if $B_0 = 0$ and if the components of $(B^1, \ldots, B^d)$ are independent continuous centered Gaussian processes with covariance

$$\mathbb{E}[B_s^i B_t^i] = \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H}), \qquad i \in \{1, \ldots, d\}, s, t \in [0, T].$$

Our aim is to construct the interated integrals $\mathbb{B}_{s,t} = (\int_s^t B_{s,r}^i \mathrm{d}B_r^j)_{i,j}$. First observe that for $i = j$ we could simply set

$$\int_s^t B_{s,r}^i \mathrm{d}B_r^i := \frac{1}{2}(B_{s,t}^i)^2.$$

This is the only possible choice under which we have the integration by parts rule from classical calculus, and it satisfies Chen's relation on the diagonal::

$$\frac{1}{2}(B_{s,t}^i)^2 - \frac{1}{2}(B_{s,u}^i)^2 - \frac{1}{2}(B_{u,t}^i)^2 = \frac{1}{2}(B_{u,t}^i B_{s,u}^i + B_{s,u}^i B_{u,t}^i) = (B_{s,u} \otimes B_{u,t})^{i,i}.$$

Moreover,

$$\left|\frac{1}{2}(B_{s,t}^i)^2\right| \leqslant \frac{|t-s|^{2\alpha}}{2} \|B\|_\alpha^2,$$

so we get the right regularity.

Remark 3.3. For any sequence of partitions (t_k) of $[0,t]$ we have

$$\frac{1}{2}(B_t^i)^2 = \frac{1}{2}\sum_k [(B_{t_{k+1}}^i)^2 - (B_{t_k}^i)^2] = \sum_k \frac{B_{t_k}^i + B_{t_{k+1}}^i}{2} B_{t_k, t_{k+1}}^i,$$

so our definition of $\int_s^t B_{s,r}^i \mathrm{d}B_r^i$ corresponds to Riemann sums taking the average of the left point and the right point of the integrand. In the Itô case we can also take left-point Riemann sums (replacing $\frac{B_{t_k}^i + B_{t_{k+1}}^i}{2}$ by $B_{t_k}^i$). We could try to do the same for $H < \frac{1}{2}$. The difference between the two Riemann sums is $\frac{1}{2}\sum_k (B_{t_k, t_{k+1}}^i)^2$, which should converge to the quadratic variation. But for $H < \frac{1}{2}$ the quadratic variation does not exist, because

$$\mathbb{E}\left[\sum_k (B_{t_k, t_{k+1}}^i)^2\right] = \sum_k |t_{k+1} - t_k|^{2H} \to \infty$$

if $H < \frac{1}{2}$.

The off-diagonal terms are more complicated.

Lemma 3.4. *Let $(B_t)_{t \in [0,1]}$ and $(W_t)_{t \in [0,1]}$ be indepdendent one-dimensional fractional Brownian motions with Hurst parameter $H \in (0, \frac{1}{2})$. Define for $n \in \mathbb{N}$*

$$I_n(B, \mathrm{d}W)(t) := \sum_{k=0}^{k_t^n - 1} B_{\tau_k^n} W_{\tau_k^n, \tau_{k+1}^n}, \qquad t \in [0,1]$$

where $\tau_k^n := k2^{-n}$ *and* $k_t^n = \max\{k : \tau_k^n \leqslant t\}$. *Then for all* $t \in [0,1]$ *and* $p \in (1,\infty)$ *we have*

$$\mathbb{E}[|I_{n+1}(B,\mathrm{d}W)(t) - I_n(B,\mathrm{d}W)(t)|^2]^{1/2} \lesssim 2^{-n(2H-1/2)}\sqrt{t}.$$

So for $H > \frac{1}{4}$ *the sequence* $(I_n(B,\mathrm{d}W)(t))_n$ *converges in* $L^2(\Omega)$ *to a limit* $I(B,\mathrm{d}W)_t$.

Proof. Let us write $I_n(t) = I_n(B,\mathrm{d}W)(t)$. Then

$$I_{n+1}(t) - I_n(t) = -\sum_{k=0}^{k_t^n - 1} B_{\tau_{2k}^{n+1},\tau_{2k+1}^{n+1}} W_{\tau_{2k+1}^{n+1},\tau_{2k+2}^{n+1}},$$

and therefore by independence of B and W

$$\begin{aligned}
&\mathbb{E}[|I_{n+1}(t) - I_n(t)|^2]\\
&= \sum_{k,\ell=0}^{k_t^n-1} \mathbb{E}[B_{\tau_{2k}^{n+1},\tau_{2k+1}^{n+1}} B_{\tau_{2\ell}^{n+1},\tau_{2\ell+1}^{n+1}}]\mathbb{E}[W_{\tau_{2k+1}^{n+1},\tau_{2k+2}^{n+1}} W_{\tau_{2\ell+1}^{n+1},\tau_{2\ell+2}^{n+1}}]\\
&\leqslant \sum_{k=0}^{k_t^n-1} (2^{-n})^{2H}(2^{-n})^{2H}\\
&+2\sum_{k=0}^{k_t^n-1}\sum_{\ell=0}^{k-1}\Big(\mathbb{E}[|B_{\tau_{2k}^{n+1},\tau_{2k+1}^{n+1}}|^2]\mathbb{E}[|B_{\tau_{2\ell}^{n+1},\tau_{2\ell+1}^{n+1}}|^2]\Big)^{1/2}|\mathbb{E}[W_{\tau_{2k+1}^{n+1},\tau_{2k+2}^{n+1}} W_{\tau_{2\ell+1}^{n+1},\tau_{2\ell+2}^{n+1}}]|.
\end{aligned}$$

The first term on the right hand side is clearly bounded by $\lesssim 2^n t 2^{-n4H} = (2^{-n(2H-1/2)}\sqrt{t})^2$. To bound the second term we need the following estimate, which we leave as an exercise:

$$|\mathbb{E}[W_{s,s+h}W_{t,t+h}]| \lesssim (t-s)^{2H-2}h^2, \qquad 0 \leqslant s < t, \qquad 0 < h \leqslant t-s.$$

This leads to

$$\begin{aligned}
&\sum_{k=0}^{k_t^n-1}\sum_{\ell=0}^{k-1}\Big(\mathbb{E}[|B_{\tau_{2k}^{n+1},\tau_{2k+1}^{n+1}}|^2]\mathbb{E}[|B_{\tau_{2\ell}^{n+1},\tau_{2\ell+1}^{n+1}}|^2]\Big)^{1/2}|\mathbb{E}[W_{\tau_{2k+1}^{n+1},\tau_{2k+2}^{n+1}} W_{\tau_{2\ell+1}^{n+1},\tau_{2\ell+2}^{n+1}}]|\\
&\lesssim 2^{-n2H}\sum_{k=0}^{k_t^n-1}\sum_{\ell=0}^{k-1}|\tau_{2k+1}^{n+1} - \tau_{2\ell+1}^{n+1}|^{2H-2}|2^{-n}|^2 \lesssim 2^{-n2H}\sum_{k=1}^{k_t^n-1}\sum_{\ell=0}^{k-1}|(k-\ell)2^{-n}|^{2H-2}|2^{-n}|^2\\
&= 2^{-n4H}\sum_{k=1}^{k_t^n-1}\sum_{\ell=0}^{k-1}|k-\ell|^{2H-2} \lesssim 2^{-n4H}\int_1^{k_t^n-1}\int_0^{x-1}|x-y|^{2H-2}\mathrm{d}y\,\mathrm{d}x.
\end{aligned}$$

The integral on the right hand side is bounded by

$$\int_1^{k_t^n-1}\int_0^{x-1}|x-y|^{2H-2}\mathrm{d}y\,\mathrm{d}x = \int_1^{k_t^n-1}\frac{1-|x|^{2H-1}}{1-2H}\mathrm{d}x \lesssim |k_t^n| \leqslant 2^n t,$$

and this completes the proof. □

Remark 3.5. The threshold $H > \frac{1}{4}$ does not appear because our estimates are inadequate. For $H \leqslant \frac{1}{4}$ the sequence $(I_n(B,\mathrm{d}W))$ does not converge (see [CQ02]) and there is no known canonical definition of $\int_0^\cdot B_s \mathrm{d}W_s$. See [Unt10, NT11] for two non-canonical constructions based on renormalization arguments (roughly speaking based on subtracting random diverging counterterms from the diverging sequence $I_n(B,\mathrm{d}W)$).

In fact, considering say a mollification $\rho_n(t) = n\rho(nt)$ for $\rho \in C_c^\infty(\mathbb{R},\mathbb{R})$ and $B^n = \rho_n * B$, the solution Y^n to

$$\partial_t Y_t^n = n^{\frac{1}{4}-H}\sigma(Y_t^n)\partial_t B_t^n, \qquad Y_0^n = x,$$

converges for $H<\frac{1}{4}$ weakly to

$$\mathrm{d}Y_t=\frac{\sigma}{2}\sum_{i,j=1}^{d}[\sigma_{i\cdot},\sigma_{j\cdot}]\mathrm{d}W^{ij},$$

where $\sigma>0$ is a constant, $[f,g]=f\cdot\nabla g-g\cdot\nabla f$ is the Lie bracket, and $(W^{ij})_{i,j=1,\ldots,d}$ are independent Brownian motions; see [Hai24].

So far we only control $I(B,\mathrm{d}W)$ in $L^2(\Omega)$. To show that it has sufficient regularity we need the following deep result on moments of polynomials of Gaussian random variables.

Theorem 3.6. (Gaussian hypercontractivity) *Let I be an index set and let $(Y_i)_{i\in I}$ be a centered Gaussian process. Let $P\colon\mathbb{R}^m\to\mathbb{R}$ be a polynomial of degree n. Then for all $0<p<\infty$ there exists a constant $C_{n,p}>0$ (which is independent of m) such that*

$$C_{n,p}^{-1}\mathbb{E}[|P(Y_{i_1},\ldots,Y_{i_m})|^p]^{1/p}\leqslant\mathbb{E}[|P(Y_{i_1},\ldots,Y_{i_m})|^2]^{1/2}\leqslant C_{n,p}\mathbb{E}[|P(Y_{i_1},\ldots,Y_{i_m})|^p]^{1/p}.$$

Proof. See Janson [Jan97], Theorem 3.50. □

Theorem 3.7. *Let $(B_t)_{t\in[0,1]}$ be a d-dimensional fractional Brownian motion with Hurst index $H\in(\frac{1}{3},\frac{1}{2})$. For $i\in\{1,\ldots,d\}$ we set $\mathbb{B}_{s,t}^{ii}:=\frac{1}{2}(B_{s,t}^i)^2$ and for $i\neq j$ let*

$$\mathbb{B}_{s,t}^{ij}:=I(B^i,\mathrm{d}B^j)_{s,t}-B_s^iB_{s,t}^j.$$

Then

$$\mathbb{E}[|\mathbb{B}_{s,t}|^p]^{1/p}\lesssim|t-s|^{2H}$$

for all $p\in(0,\infty)$, and in particular we can apply Theorem 3.1 to obtain a modification $\tilde{\mathbb{B}}$ of $\mathbb{B}$ such that $(B,\tilde{\mathbb{B}})$ is an α-rough path for all $\alpha\in(1/3,H)$.

Proof. Let $i\neq j$ and $0\leqslant s<t\leqslant 1$. Define

$$J_n(t):=I_n(B^i,\mathrm{d}B^j)(t)+B_{\tau_{k_t^n}^n}^iB_{\tau_{k_t^n}^n,\tau_{k_t^n}^n+t}^j,$$

where I_n is as in Lemma 3.4. Using similar arguments as in the proof of Lemma 3.4, one can show that

$$\mathbb{E}[|(J_n)_{s,t}-I(B^i,\mathrm{d}B^j)_{s,t}|^2]^{1/2}\lesssim 2^{-n(2H-1/2)}\sqrt{t-s}.$$

The extra term makes the calculation longer but not more difficult; similarly estimating the difference of the time increments is more technical but not more difficult than estimating the difference at a fixed time.

Pick now n_0 with $2^{-n_0-1}\leqslant|t-s|<2^{-n_0}$. Using the same decomposition that appeared in the proof of the sewing lemma, Theorem 1.2, we get

$$\begin{aligned}\mathbb{E}[|I(B^i,\mathrm{d}B^j)_{s,t}-B_s^iB_{s,t}^j|^p]^{1/p}&\leqslant\mathbb{E}[|I(B^i,\mathrm{d}B^j)_{s,t}-(J_{n_0})_{s,t}|^p]^{1/p}+\mathbb{E}[|(J_{n_0})_{s,t}-B_s^iB_{s,t}^j|^p]^{1/p}\\&\lesssim|t-s|^{2H},\end{aligned}$$

where the last step used Gaussian hypercontractivity. Now the claim follows from Kolmogorov's continuity criterion for rough paths. □

3.2 A glimpse into rough path homogenization

Recall the example from the introduction: X is given by a sufficiently chaotic dynamical system

$$\dot{X}=f(X),\qquad X(0)\sim\mu,$$

where μ is an invariant measure (i.e. $X(t)\sim\mu$ for all $t\geqslant 0$), and

$$\dot{Y}^{\varepsilon}(t)=\frac{1}{\sqrt{\varepsilon}}\sigma(Y^{\varepsilon}(t))X(t/\varepsilon), \qquad Y^{\varepsilon}(0)=x_0,$$

where $\sigma\in C_b^3(\mathbb{R}^m, L(\mathbb{R}^d,\mathbb{R}^m))$ and we assume that

$$\int y\mu(\mathrm{d}y)=0.$$

Based on what we understand now, we know that if

$$X^{\varepsilon}_{s,t}=\frac{1}{\sqrt{\varepsilon}}\int_s^t X(r/\varepsilon)\mathrm{d}r, \qquad \mathbb{X}^{\varepsilon}_{s,t}=\frac{1}{\varepsilon}\int_s^t\int_s^{r_1} X(r_2/\varepsilon)\mathrm{d}r_2\otimes X(r_1/\varepsilon)\mathrm{d}r_1$$

converges weakly in α-rough path topology for $\alpha\in(\frac{1}{3},\frac{1}{2})$ to

$$\left(B_{s,t}, \int_s^t B_{s,r_1}\otimes\circ\mathrm{d}B_{r_1}+A(t-s)\right),$$

where $A\in\mathbb{R}^{d\otimes d}$ takes values in the antisymmetric matrices (which has to be true because $(X^{\varepsilon},\mathbb{X}^{\varepsilon})$ is a geometric rough path and therefore the limit is, too), then Y^{ε} converges weakly to Y^0 solving $\sigma(Y)\in\mathbb{R}^m$

$$\mathrm{d}Y^0_t=\sigma(Y^0_t)\circ\mathrm{d}B_t+D\sigma(Y^0_t)\sigma(Y^0_t)A\mathrm{d}t.$$

And indeed, the rough path convergence can be shown under suitable assumptions on X, see for example the foundational work [KM17], where also σ that are nonlinear in X are studied.

4 Signatures and applications in machine learning

The discussion of low regularity rough paths showed the importance of higher iterated integrals. The *signature* of a path consists of all of its iterated integrals, seen not as functions but evaluated at the terminal time. This can be interpreted as a nonlinear type of "Fourier transform", which encodes important information about a path and its nonlinear effects. Therefore, the signature is a powerful tool in machine learning applications.

4.1 The signature of a smooth path

We start by introducing the signature of a differentiable path $X\in C^1([0,T],V)$, where as before V is a d-dimensional normed vector space with basis $(e_1,\ldots,e_d)$.

Definition 4.1. (Signature of non-rough paths) *Let $X\in C^1([s,t],V)$. The* signature *of X is defined as*

$$S(X):=\big(1,\mathbb{X}^{(1)}_{s,t},\mathbb{X}^{(2)}_{s,t},\ldots\big)\in T((V)),$$

where

$$\mathbb{X}^{(n)}_{s,t}:=\int_{s<r_1<\cdots<r_n<t}\mathrm{d}X_{r_1}\otimes\cdots\otimes\mathrm{d}X_{r_n}\in V^{\otimes n},$$

or, equivalently,

$$\mathbb{X}^{i_1\cdots i_n}_{s,t}:=\int_{s<r_1<\cdots<r_n<t}\mathrm{d}X^{i_1}_{r_1}\cdots\mathrm{d}X^{i_n}_{r_n}.$$

Sometimes we also write $S(X)_{s,t}$ to emphasize the time interval under consideration, so that $r\mapsto S(X)_{s,r}$ is a $T((V))$-valued path.

The signature is just the collection of all iterated integrals of a path, evaluated at a fixed time. It has several useful properties, and to formulate one of those we need the concatenation of paths:

Definition 4.2. (Concatenation of paths) *Let* $0 \leqslant s \leqslant u \leqslant t$ *and let* $X: [s,u] \to V$ *and* $Y: [u,t] \to V$. *The concatenation of* X *and* Y *is*

$$X\star Y: [s,t] \to V, \qquad (X\star Y)_r := \begin{cases} X_r, & r \in [s,u], \\ Y_r - Y_u + X_u, & r \in (u,t]. \end{cases}$$

Note that $X\star Y$ is not necessarily C^1, but it is always continuous and of bounded variation (actually piecewise C^1 and Lipschitz continuous). All of the discussion in this section extends to continuous paths of bounded variation, and therefore this loss of continuous differentiability does not pose a problem.

Lemma 4.3. (Properties of the signature) *Let* $X \in C^1([s,u],V)$ *and* $Y \in C^1([u,t],V)$.

i. Invariance under time reparametrization: *For* $\tau \in C^1([a,b],[s,u])$ *strictly increasing and surjective (a "time change") consider* $X \circ \tau \in C^1([a,b],V)$. *Then*

$$S(X \circ \tau)_{a,b} = S(X)_{s,u}.$$

ii. Homomorphism property / Chen's relation:

$$S(X\star Y) = S(X) \otimes S(Y).$$

iii. Time reversal: $\overleftarrow{X}: [s,u] \to V$ *with* $\overleftarrow{X}_r := X_{s+u-r}$ *satisfies*

$$S(\overleftarrow{X})_{s,u} = S(X)_{s,u}^{-1},$$

i.e.

$$S(X)_{s,u} \otimes S(\overleftarrow{X})_{s,u} = S(\overleftarrow{X})_{s,u} \otimes S(X)_{s,u} = \mathbf{1} = (1,0,0,\ldots).$$

iv. Shuffle product: *For all words* $w, w' \in \mathcal{W}$ *we have*

$$\langle S(X), w\rangle \langle S(X), w'\rangle = \langle S(X), w \sqcup w'\rangle.$$

Proof.

i. Note that $(X \circ \tau)' = (X' \circ \tau)\tau'$ by the chain rule, and thus a change of variables yields for any continuous function $f: [a,b] \to \mathbb{R}$

$$\int_a^c f(r) \mathrm{d}(X \circ \tau)_r^i = \int_s^{\tau(c)} (f \circ \tau^{-1})(r) \mathrm{d}X_r^i. \tag{4.1}$$

With $f = 1$ we obtain $S(X \circ \tau)_{a,c}^i = S(X)_{s,\tau(c)}^i$. Now assume by induction that $S(X \circ \tau)_{a,c}^{i_1 \ldots i_{n-1}} = S(X)_{s,\tau(c)}^{i_1 \ldots i_{n-1}}$. Then

$$\begin{aligned} S(X \circ \tau)_{a,c}^{i_1 \ldots i_n} &= \int_a^c S(X \circ \tau)_{a,r}^{i_1 \ldots i_{n-1}} \mathrm{d}(X \circ \tau)_r^{i_n} \\ &\overset{\text{induct.}}{=} \int_a^c S(X)_{s,\tau(r)}^{i_1 \ldots i_{n-1}} \mathrm{d}(X \circ \tau)_r^{i_n} \\ &\overset{(4.1)}{=} \int_s^{\tau(c)} S(X)_{s,r}^{i_1 \ldots i_{n-1}} \mathrm{d}X_r^i \\ &= S(X)_{s,\tau(c)}^{i_1 \ldots i_n}. \end{aligned}$$

Taking $c = b$ with $\tau(c) = u$ by surjectivity, the claim follows.

ii. Let $Z := X \star Y$. We have with $r_0 := 0$ and $r_{n+1} := t$

$$
\begin{aligned}
\mathbb{Z}^{i_1 \cdots i_n}_{s,t} &= \int_{s<r_1<\cdots<r_n<t} \mathrm{d}Z^{i_1}_{r_1} \cdots \mathrm{d}Z^{i_n}_{r_n} \\
&= \sum_{k=0}^{n} \int_{s<r_1<\cdots<r_k<u<r_{k+1}<\cdots<r_n<t} \mathrm{d}Z^{i_1}_{r_1} \cdots \mathrm{d}Z^{i_n}_{r_n} \\
&\overset{\text{Fubini}}{=} \sum_{k=0}^{n} \int_{s<r_1<\cdots<r_k<u} \mathrm{d}Z^{i_1}_{r_1} \cdots \mathrm{d}Z^{i_k}_{r_k} \int_{u<r_{k+1}<\cdots<r_n<t} \mathrm{d}Z^{i_{k+1}}_{r_{k+1}} \cdots \mathrm{d}Z^{i_n}_{r_n} \\
&\overset{Z=Y\star X}{=} \sum_{k=0}^{n} \int_{s<r_1<\cdots<r_k<u} \mathrm{d}X^{i_1}_{r_1} \cdots \mathrm{d}X^{i_k}_{r_k} \int_{u<r_{k+1}<\cdots<r_n<t} \mathrm{d}Y^{i_{k+1}}_{r_{k+1}} \cdots \mathrm{d}Y^{i_n}_{r_n} \\
&= \sum_{k=0}^{n} \mathbb{X}^{i_1 \ldots i_k}_{s,u} \mathbb{Y}^{i_{k+1} \ldots i_n}_{u,t} \\
&= (S(X) \otimes S(Y))^{i_1 \ldots i_n}_{s,t}.
\end{aligned}
$$

iii. This follows by observing that the signature solves the differential equation $\mathrm{d}S(X)_{s,r} = S(X)_{s,r} \otimes \mathrm{d}X_r$. See Proposition 2.14 in [LCL07] for details.

iv. The claim clearly holds for $|w|, |w'| \leqslant 1$ by the integration by parts rule. Assume it holds for $|w| + |w'| \leqslant n$ and consider wi and $w'j$ with $|wi| + |w'j| \leqslant n + 1$. Then $|wi| + |w'| \leqslant n$ and $|w| + |w'j| \leqslant n$ and therefore

$$
\begin{aligned}
\langle S(X), wi \rangle \langle S(X), w'j \rangle &= \int_{s<r_1<t} S(X)^{w}_{s,r_1} \mathrm{d}X^{i}_{r_1} \int_{s<r_2<t} S(X)^{w'}_{s,r_2} \mathrm{d}X^{j}_{r_2} \\
&= \int_{s<r_1<r_r<t} S(X)^{w}_{s,r_1} S(X)^{w'}_{s,r_2} \mathrm{d}X^{i}_{r_1} \mathrm{d}X^{j}_{r_2} \\
&\quad + \int_{s<r_2<r_1<t} S(X)^{w}_{s,r_1} S(X)^{w'}_{s,r_2} \mathrm{d}X^{i}_{r_1} \mathrm{d}X^{j}_{r_2} \\
&= \int_{s<r_2<t} S(X)^{wi}_{s,r_2} S(X)^{w'}_{s,r_2} \mathrm{d}X^{j}_{r_2} \\
&\quad + \int_{s<r_1<t} S(X)^{w}_{s,r_1} S(X)^{w'j}_{s,r_1} \mathrm{d}X^{i}_{r_1} \\
&\overset{\text{induction}}{=} \int_{s<r_2<t} S(X)^{wi \sqcup w'}_{s,r_2} \mathrm{d}X^{j}_{r_2} + \int_{s<r_1<t} S(X)^{w \sqcup w'j}_{s,r_1} \mathrm{d}X^{i}_{r_1} \\
&= S(X)^{(wi \sqcup w')j}_{s,t} + S(X)^{(w \sqcup w'j)i}_{s,t} \\
&= S(X)^{wi \sqcup w'j}_{s,t}.
\end{aligned}
$$

□

From the signature of a C^1 (or bounded variation) path X, we can read off any solution to a linear equation driven by X: Let $A \in L(V, L(W, W))$ for an m-dimensional normed vector space W and consider the linear differential equation

$$\mathrm{d}Y_t = A Y_t \mathrm{d}X_t, \qquad Y_0 = x \in W,$$

which we interpret as $A(\mathrm{d}X_t) Y_t$ or, coordinatewise, as $A \in \mathbb{R}^{m \times m \times d}$ with

$$\mathrm{d}Y^i_t = \sum_{j=1}^{m} \sum_{\ell=1}^{d} A_{i,j,k} Y^j_t \mathrm{d}X^k_t.$$

Let us set up a Picard iteration with

$$\Phi(Y)^i_t = x^i + \int_0^t \sum_{j=1}^m \sum_{k=1}^d A_{i,k,\ell} Y^j_s \mathrm{d}X^k_s,$$

which we start in the zero path. Then

$$\begin{aligned}
\Phi(0)^i_t &= x^i,\\
\Phi\circ\Phi(0)^i_t &= x^i + \int_0^t \sum_{j=1}^m \sum_{k=1}^d A_{i,j,k}\, x^j \mathrm{d}X^k_s\\
&= x^i + \sum_{j=1}^m \sum_{k=1}^d A_{i,j,k}\, x^j \mathbb{X}^k_{0,t},\\
\Phi^{\circ n+1}(0)^i_t &= x^i + \sum_{j=1}^m \sum_{k=1}^d A_{i,j,k} \int_0^t \Phi^{\circ n}(0)^j_s \mathrm{d}X^k_s.
\end{aligned}$$

This suggests defining $A^{\otimes 0}\colon W\to W$

$$A^{\otimes 0}(x)^i := x^i,$$

and then inductively $A^{\otimes n}\colon W\otimes V^{\otimes n}\to W$,

$$A^{\otimes n+1}(x\otimes y)^i = \sum_{j_{n+1}=1}^m \sum_{k_{n+1}=1}^d A_{i,j_{n+1},k_{n+1}} A^{\otimes n}(x, y^{\cdot k_{n+1}})^{j_{n+1}},$$

so that

$$\Phi(0)_t = A^{\otimes 0}(x), \qquad \Phi^{\circ 2}(0)_t = A^{\otimes 0}(x) + A^{\otimes 1}(x\otimes \mathbb{X}_{0,t}),$$

and then with the induction assumption $\Phi^{\circ n}(0)_t = \sum_{r=0}^{n-1} A^{\otimes r}\big(x\otimes \mathbb{X}^{(r)}_{0,t}\big)$:

$$\begin{aligned}
\Phi^{\circ n+1}(0)^i_t &= x^i + \int_0^t \sum_{j_{n+1}=1}^m \sum_{k_{n+1}=1}^d A_{i,j_{n+1},k_{n+1}} \sum_{r=0}^{n-1} A^{\otimes r}\big(x\otimes \mathbb{X}^{(r)}_{0,s}\big)^{j_{n+1}} \mathrm{d}X^{j_{n+1}}_s\\
&= x^i + \sum_{r=1}^n A^{\otimes r}\big(x\otimes \mathbb{X}^{(r)}_{0,t}\big)^i.
\end{aligned}$$

Corollary 4.4. *Let $X\in C^1([0,T],V)$ and let $A\in L(V,L(W,W))$ and $x\in W$. The solution to the differential equation*

$$\mathrm{d}Y_t = AY_t \mathrm{d}X_t, \qquad Y_0 = x\in W,$$

is given by

$$Y_t = \sum_{r=0}^\infty A^{\otimes r}\big(x\otimes \mathbb{X}^{(r)}_{0,t}\big).$$

Proof. The only thing left to prove is the convergence of the series for all times (for small times it follows from the contraction property of the Picard iteration). But

$$\big|A^{\otimes r}\big(x\otimes \mathbb{X}^{(r)}_{0,t}\big)\big| \lesssim |A|^r |x| \big|\mathbb{X}^{(r)}_{0,t}\big|$$

and

$$\begin{aligned}
\big|\mathbb{X}^{(r)}_{0,t}\big| &= \left|\int_{0<s_1<\cdots<s_r<t} \mathrm{d}X_{s_1}\otimes\cdots\otimes \mathrm{d}X_{s_r}\right|\\
&\leqslant \|X'\|_\infty C(d)^r \left|\int_{0<s_1<\cdots<s_r<t} \mathrm{d}s_1 \ldots \mathrm{d}s_r\right|\\
&= \|X'\|_\infty C(d)^r \frac{t^r}{r!},
\end{aligned}$$

which proves the summability of the series. □

Thus, if X and $\tilde{X}$ are two continuously differentiable or bounded variation paths with $S(X)_{0,T} = S(\tilde{X})_{0,T}$, then the solutions Y and $\tilde{Y}$ to the linear equations

$$\mathrm{d}Y_t = A Y_t \mathrm{d}X_t, \qquad Y_0 = x \in W,$$

respectively

$$\mathrm{d}\tilde{Y}_t = A \tilde{Y}_t \mathrm{d}\tilde{X}_t, \qquad Y_0 = x \in W,$$

agree at the end time: $Y_T = \tilde{Y}_T$.

What about nonlinear equations? Can we determine terminal value of the solution to a nonlinear ODE driven by X from the signature $S(X)_{0,T}$? Or, more ambitiously, is it maybe possible to recover the entire path $(X_t)_{t\in[0,T]}$ from knowing $S(X)_{0,T}$? Due to invariance under time reparametrization and Chen's rule together with the time reversal property, the answer is clearly no: We can only hope to obtain X up to an unknown time parametrization, and we cannot tell apart the signature of $X \star \overleftarrow{X}$ and of the zero path $0_t = 0$:

$$S\big(X \star \overleftarrow{X}\big)_{0,T} = S(X)_{0,T} \otimes S\big(\overleftarrow{X}\big)_{0,T} = \mathbf{1} = S(0)_{0,T}.$$

But remarkably this is the only loss of information when passing from a path to its signature: As shown by Hambly and Lyons [HL10] (with previous work by Chen on piecewise smooth paths and later work by Beodihardjo-Geng-Lyons-Yang on rough paths), the signature of a bounded variation path determines up to time parametrization a unique path that never goes directly back on itself, in the sense that does not have any *tree-like* parts, where *tree-like path* is a notion introduced in [HL10].

Since the time-parametrization and the insertion of tree-like paths do not affect solutions to differential equations driven by a given path, we obtain as a corollary that if $S(X)_{0,T} = S(\tilde{X})_{0,T}$, then the solutions to any (linear or nonlinear) differential equation driven by X respectively $\tilde{X}$ agree at the terminal time.

Maybe discuss log signature.

4.2 The signature of a rough path

It is natural to also consider the signature of a rough path. Its existence is less obvious than in the smooth case, but it follows from the existence of the controlled rough path integral.

Definition 4.5. (Signature of non-rough paths) *Let $\boldsymbol{X}$ be an α-rough path and let $N = \lfloor \alpha^{-1} \rfloor$. The* signature *of $\boldsymbol{X}$ is defined as*

$$S(\boldsymbol{X})_{s,t} := \big(1, \mathbb{X}_{s,t}^{(1)}, \ldots, \mathbb{X}_{s,t}^{(N)}, \mathbb{X}_{s,t}^{(N+1)}, \ldots\big) \in T((V)),$$

where for $n > N$ we define inductively as controlled rough path integral

$$\mathbb{X}_{s,t}^{(n)} := \int_{s<r<t} \mathbb{X}_{s,r}^{(n-1)} \otimes \mathrm{d}X_r \in V^{\otimes n},$$

or, equivalently,

$$\mathbb{X}_{s,t}^{i_1 \cdots i_n} := \int_{s<r<t} \mathbb{X}_{s,r}^{i_1 \ldots i_{n-1}} \mathrm{d}X_r^{i_n}.$$

Since we saw that the rough path integral maps controlled paths to controlled paths, the existence of all higher order integrals follows from that of

$$\mathbb{X}_{s,t}^{i_1 \ldots i_{N+1}} = \int_{s<r<t} \mathbb{X}_{s,r}^{i_1 \ldots i_N} \mathrm{d}X_r^{i_{N+1}},$$

for which we have to show that $[s,T]\ni r\mapsto \mathbb{X}_{s,r}^{i_1\dots i_N}\in\mathbb{R}$ is in $\mathscr{D}_{\boldsymbol{X}}^{N\alpha}([s,T],\mathbb{R})$. We leave this as an exercise.

Then, one can derive decay properties of the iterated integrals which allow to conclude the same expansion formula for solutions to linear rough differential equations as in the smooth case: The solution Y to

$$\mathrm{d}Y_t = AY_t\mathrm{d}\boldsymbol{X}_t, \qquad Y_0 = x\in W,$$

is given by

$$Y_t = \sum_{r=0}^{\infty} A^{\otimes r}\big(x\otimes \mathbb{X}_{0,t}^{(r)}\big).$$

Also, $S(\boldsymbol{X})_{0,T}$ determines the terminal value of any nonlinear rough differential equation driven by $\boldsymbol{X}$.

4.3 Applications of signatures in machine learning

In recent years, the signature transform has emerged as a powerful tool in machine learning, particularly for analyzing sequential and time-series data. The essence of rough path theory lies in its ability to provide a robust mathematical framework for handling paths that exhibit irregular and oscillatory behavior, which are commonly encountered in real-world data streams. The signature of a path captures the essential features of the path through its iterated integrals, enabling a compact and efficient representation of complex data. An important aspect of the signature is that it captures the order of events, which distinguishes it from linear signal processing transforms such as wavelets and Fourier transforms.

Signature-based methods have shown impressive performance in tasks such as sequence classification, prediction, and anomaly detection. But I am not an expert and am only able to give a quite superficial glimpse into these applications. See [Lyo14, CK16, FLMS23, LM22, CS24] for more extensive introductions.

Broadly seen, the effectiveness of signature methods can be explained by the following meta result, where we write

$$S^{(N)}(X) := \big(1, \mathbb{X}_{0,T}^{(1)}, \dots, \mathbb{X}_{0,T}^{(N)}\big)$$

for the truncated signature of X.

Theorem 4.6. *Let* $\mathrm{BV}([0,T],V)$ *be the space of continuous bounded variation functions with norm* $\|X\| := \|X\|_\infty + \|X\|_{\mathrm{TV}}$*, where*

$$\|X\|_{\mathrm{TV}} := \sup_{n\in\mathbb{N}} \sup_{0=t_0<\dots<t_n=T} \sum_{k=0}^{n-1} |X_{t_k,t_{k+1}}|.$$

Let $F\colon \mathrm{BV}([0,T],V)\to W$ *be a function which does not depend on the time reparametrization of* X *and which assigns the same value to all paths that are "equivalent up to tree-like paths". Then for every compact set* $K\subset \mathrm{BV}([0,T],V)$ *and every* $\varepsilon>0$ *there exist* $N\in\mathbb{N}_0$ *and an* affine linear *function* $L\colon T^{(N)}(V)\to W$ *such that*

$$\sup_{X\in K} |F(X) - LS^{(N)}(X)| \leqslant \varepsilon.$$

Proof. By arguing componentwise, we may assume that $W=\mathbb{R}$.

We consider an equivalence relation on $\mathrm{BV}([0,T],V)$ where two paths are equivalent if they differ by time-parametrization and the insertion of tree-like paths. Write $\overline{\mathrm{BV}}$ for the equivalence classes. By assumption we can interpret F as a map from $\overline{\mathrm{BV}}$ to W, and the equivalence classes corresponding to elements of K is a compact subset $\overline{K}\subset\overline{\mathrm{BV}}$ in the quotient space topology.

Consider the following space of functions on $\overline{\mathrm{BV}}$:

$$\mathcal{S} = \{X \mapsto L S^{(N)}(X) \colon N \in \mathbb{N}_0, L \colon T^{(N)}(V) \to W \text{ affine linear}\}.$$

By the Hambly-Lyons result, $\mathcal{S}$ separates points. Moreover, $\mathcal{S}$ is an algebra: For linear φ, $\psi \in \mathcal{S}$ we have

$$\begin{aligned}
\varphi(X)\psi(X) &= \left(\sum_{k=0}^{N} \sum_{i_1 \ldots i_k = 1}^{d} \varphi_{i_1 \cdots i_k} \mathbb{X}_{0,T}^{i_1 \ldots i_k}\right)\left(\sum_{\ell=0}^{M} \sum_{j_1 \ldots j_\ell = 1}^{d} \psi_{j_1 \cdots j_\ell} \mathbb{X}_{0,T}^{j_1 \ldots j_\ell}\right) \\
&= \sum_{k=0}^{N} \sum_{\ell=0}^{M} \sum_{i_1 \ldots i_k, j_1 \ldots j_\ell = 1}^{d} \varphi_{i_1 \cdots i_k} \psi_{j_1 \cdots j_\ell} \mathbb{X}_{0,T}^{i_1 \ldots i_k} \mathbb{X}_{0,T}^{j_1 \ldots j_\ell} \\
&= \sum_{k=0}^{N} \sum_{\ell=0}^{M} \sum_{i_1 \ldots i_k, j_1 \ldots j_\ell = 1}^{d} \varphi_{i_1 \cdots i_k} \psi_{j_1 \cdots j_\ell} \mathbb{X}_{0,T}^{i_1 \ldots i_k \sqcup j_1 \ldots j_\ell},
\end{aligned}$$

which is again an element of $\mathcal{S}$. Making φ and/or ψ affine linear by adding a constant does not change anything.

Now the claim follows by the Stone-Weierstraß theorem. □

Exercise 4.1. Formulate a similar result for functions of rough paths.

The goal in machine learning is to learn a function F from given data points

$$(X_i, F(X_i))_{i=1,\ldots,n}.$$

If F has the invariance in the previous theorem, we can achieve this by learning an *affine linear* function on signatures. For example, we could use linear regression methods on signature features.

In recent years, this approach has been implemented by imposing more structure on the function F: The neural controlled differential equation (neural CDE) approach assumes that the data points are given by

$$(X_i, Y_i)_{i=1,\ldots,n},$$

where each X_i is a time series, which we enhance to a continuous path of bounded variation and from now on identify with this function. Then we make the ansatz

$$Y_i = \Phi_\theta(y, S(X_i)),$$

where y is a fixed[4.1] initial value and $\Phi_\theta(y, S(X))$ is the solution Y at time 1 of the controlled differential equation

$$Y_t = y + \int_0^t f_\theta(Y_s) \mathrm{d}X_s,$$

where f_θ is a nonlinear function that is parametrized by a neural network. Numerically, we can approximate the solution by a function of the truncated signature $S^{(N)}(X)$.

Then we learn f_θ by tuning the weights in the neural network via usual (stochastic) gradient descent methods, with the goal of minimizing some loss function ℓ,

$$\sum_{i=1}^{n} \ell(\Phi_\theta(y, S(X_i)), Y_i).$$

4.1. Or maybe also learnable? This is unclear to me.

This method has been implemented and tested on many concrete data sets, with strong performance on complex problems. Another successful installation of the signature method in learning are signature kernels, which integrate signatures with kernel methods to facilitate computations. See the references listed above for further details.

5 Applications to SPDEs

5.1 Solving the KPZ equation with rough paths

To be done, I will not manage to write anything but we can discuss this during the lecture.

5.2 Regularity structures and the parabolic Anderson model

To be done. I will not manage to prepare a "light" version of regularity structures as originally planned. But we can discuss the general theory of regularity structures and specifically the example of the 2d parabolic Anderson model.

6 Stochastic sewing and regularization by noise

The rough path approach bypasses most of stochastic analysis and allows handling stochastic processes with tools from deterministic, pathwise analysis. But in recent years there is a strong trend rough analysis to combine rough path ideas with probabilistic methods. This for example underlies the probabilistic theory of energy solutions to singular SPDEs such as the KPZ equation [GJ14, GP18, GP20]. A particularly elegant combination of rough path and probabilistic ideas is Khoa Lê's stochastic sewing lemma, which has proven to be a powerful tool in stochastic analysis, with applications in[6.1] regularization by noise [Lê20, HP21] (and many more), numerics [LL21] (and many more), homogenization [HL20], hybrid rough-stochastic dynamics [FHL21] and stochastic calculus for fractional Brownian motion [MP24, DŁMP23].

The main idea of the stochastic sewing lemma is to combine the argument of the sewing lemma with the Burkholder-Davis Gundy inequality to make use of stochastic cancellations.

Theorem 6.1. (Lê, [Lê20]) *Let $p \geqslant 2$ and let $\Xi : \Delta_T \to L^p := L^p(\Omega, \mathbb{P})$ be adapted (i.e. $\Xi_{s,t}$ is $\mathcal{F}_t$-measurable for all $(s,t) \in \Delta_T$) and continuous as an L^p-valued map, with $\Xi_{t,t} = 0$ for all $t \in [0,T]$. Assume that there exist $C > 0$ and $\varepsilon > 0$ such that for all $0 \leqslant s \leqslant u \leqslant t \leqslant T$:*

$$\begin{aligned} \|\delta \Xi_{s,u,t}\|_{L^p} &\leqslant C|t-s|^{\frac{1}{2}+\varepsilon}, \\ \|\mathbb{E}[\delta \Xi_{s,u,t}|\mathcal{F}_s]\|_{L^p} &\leqslant C|t-s|^{1+\varepsilon}. \end{aligned}$$

Then there exists a unique stochastic process $\mathcal{I}\Xi : [0,T] \to L^p$ such that $\mathcal{I}\Xi$ is continuous as an L^p-valued map, $\mathcal{I}\Xi_0 = 0$, and such that for all $0 \leqslant s \leqslant t \leqslant T$

$$\begin{aligned} \|\mathcal{I}\Xi_{s,t} - \Xi_{s,t}\|_{L^p} &\lesssim_T C|t-s|^{\frac{1}{2}+\varepsilon}, \\ \|\mathbb{E}[\mathcal{I}\Xi_{s,t} - \Xi_{s,t}|\mathcal{F}_s]\|_{L^p} &\lesssim C|t-s|^{1+\varepsilon}. \end{aligned}$$

6.1. This list is skewed towards my own work, and misses important references by Butkovsky and Gerencser, among others.

Proof.

- Existence: We only construct $\mathcal{I}\Xi$ and show the estimate for $s=0$. The estimate for $s>0$ needs some additional work, as in the case of the deterministic sewing lemma.

  Let $t_k^n := k2^{-n}t$ for $n \in \mathbb{N}_0$ and $k \in \{0, \ldots, 2^n\}$. Then define

$$\mathcal{I}^n\Xi_t := \sum_{k=0}^{2^n-1} \Xi_{t_k^n, t_{k+1}^n}.$$

  The first step is now as in the deterministic case, but then we use a Doob-Meyer type decomposition to facilitate the further analysis:

$$\begin{aligned}
\mathcal{I}^{n+1}\Xi_t - \mathcal{I}^n\Xi_t &= \sum_{k=0}^{2^n-1} \delta\Xi_{t_{2k}^{n+1}, t_{2k+1}^{n+1}, t_{2k+2}^{n+1}} \\
&= \sum_{k=0}^{2^n-1} \left(\delta\Xi_{t_{2k}^{n+1}, t_{2k+1}^{n+1}, t_{2k+2}^{n+1}} - \mathbb{E}\Big[\delta\Xi_{t_{2k}^{n+1}, t_{2k+1}^{n+1}, t_{2k+2}^{n+1}} | \mathcal{F}_{t_{2k}^{n+1}}\Big]\right) \\
&\quad + \sum_{k=0}^{2^n-1} \mathbb{E}\Big[\delta\Xi_{t_{2k}^{n+1}, t_{2k+1}^{n+1}, t_{2k+2}^{n+1}} | \mathcal{F}_{t_{2k}^{n+1}}\Big].
\end{aligned}$$

  The first term on the right hand side consists of martingale increments. Therefore, we apply the Burkholder-Davis Gundy inequality and Minkowski's inequality to obtain

$$\begin{aligned}
&\left\| \sum_{k=0}^{2^n-1} \left(\delta\Xi_{t_{2k}^{n+1}, t_{2k+1}^{n+1}, t_{2k+2}^{n+1}} - \mathbb{E}\Big[\delta\Xi_{t_{2k}^{n+1}, t_{2k+1}^{n+1}, t_{2k+2}^{n+1}} | \mathcal{F}_{t_{2k}^{n+1}}\Big]\right) \right\|_{L^p} \\
&\overset{\text{BDG}}{\lesssim} \left\| \sum_{k=0}^{2^n-1} \left(\delta\Xi_{t_{2k}^{n+1}, t_{2k+1}^{n+1}, t_{2k+2}^{n+1}} - \mathbb{E}\Big[\delta\Xi_{t_{2k}^{n+1}, t_{2k+1}^{n+1}, t_{2k+2}^{n+1}} | \mathcal{F}_{t_{2k}^{n+1}}\Big]\right)^2 \right\|_{L^{p/2}}^{1/2} \\
&\overset{\text{Minkowski}}{\leqslant} \left(2\sum_{k=0}^{2^n-1} \left\| \delta\Xi_{t_{2k}^{n+1}, t_{2k+1}^{n+1}, t_{2k+2}^{n+1}} - \mathbb{E}\Big[\delta\Xi_{t_{2k}^{n+1}, t_{2k+1}^{n+1}, t_{2k+2}^{n+1}} | \mathcal{F}_{t_{2k}^{n+1}}\Big] \right\|_{L^p}^2 \right)^{1/2} \\
&\lesssim \left(\sum_{k=0}^{2^n-1} \|\delta\Xi_{t_{2k}^{n+1}, t_{2k+1}^{n+1}, t_{2k+2}^{n+1}}\|_{L^p}^2 \right)^{1/2} \\
&\leqslant \left(2^n C^2 2^{-n\left(\frac{1}{2}+\varepsilon\right)2} t^{\left(\frac{1}{2}+\varepsilon\right)2}\right)^{1/2} \\
&= C 2^{-n\varepsilon} t^{\frac{1}{2}+\varepsilon},
\end{aligned}$$

  which is summable in n. The remaining term is simply bounded by the triangle inequality:

$$\begin{aligned}
\left\| \sum_{k=0}^{2^n-1} \mathbb{E}\Big[\delta\Xi_{t_{2k}^{n+1}, t_{2k+1}^{n+1}, t_{2k+2}^{n+1}} | \mathcal{F}_{t_{2k}^{n+1}}\Big] \right\|_{L^p} &\leqslant \sum_{k=0}^{2^n-1} \left\| \mathbb{E}\Big[\delta\Xi_{t_{2k}^{n+1}, t_{2k+1}^{n+1}, t_{2k+2}^{n+1}} | \mathcal{F}_{t_{2k}^{n+1}}\Big] \right\|_{L^p} \\
&\leqslant 2^n C 2^{-n(1+\varepsilon)} t^{1+\varepsilon} = C 2^{-n\varepsilon} t^{1+\varepsilon},
\end{aligned}$$

  which is also summable in n. From here we get the existence of $\mathcal{I}\Xi_t$ and the bound for $s=0$.

- Uniqueness: If $\tilde{\mathcal{I}}\Xi$ is another such process, let us write $\Delta_t = \mathcal{I}\Xi_t - \tilde{\mathcal{I}}\Xi_t$. Then Δ is a stochastic process satisfying

$$\|\Delta_{s,t}\|_{L^p} \lesssim |t-s|^{\frac{1}{2}+\varepsilon}, \qquad \|\mathbb{E}[\Delta_{s,t}|\mathcal{F}_s]\|_{L^p} \lesssim |t-s|^{1+\varepsilon}.$$

With the same dyadic times t_k^n as in the existence proof we obtain again with Burkholder-Davis-Gundy and triangle inequality

$$\begin{aligned}\|\Delta_t\|_{L^p} &= \left\|\sum_{k=0}^{2^n-1}\Delta_{t_k^n,t_{k+1}^n}\right\|_{L^p}\\ &\leqslant \left\|\sum_{k=0}^{2^n-1}\Delta_{t_k^n,t_{k+1}^n}-\mathbb{E}[\Delta_{t_k^n,t_{k+1}^n}|\mathcal{F}_{t_k^n}]\right\|_{L^p}+\left\|\sum_{k=0}^{2^n-1}\mathbb{E}[\Delta_{t_k^n,t_{k+1}^n}|\mathcal{F}_{t_k^n}]\right\|_{L^p}\\ &\lesssim \left(\sum_{k=0}^{2^n-1}\|(\Delta_{t_k^n,t_{k+1}^n}-\mathbb{E}[\Delta_{t_k^n,t_{k+1}^n}|\mathcal{F}_{t_k^n}])\|_{L^p}^2\right)^{1/2}+\sum_{k=0}^{2^n-1}\|\mathbb{E}[\Delta_{t_k^n,t_{k+1}^n}|\mathcal{F}_{t_k^n}]\|_{L^p}\\ &\lesssim \left(2^n 2^{-n\left(\frac{1}{2}+\varepsilon\right)2}\right)^{1/2}+2^n 2^{-n(1+\varepsilon)}\lesssim 2^{-n\varepsilon}.\end{aligned}$$

Since n is arbitrary, we must have $\Delta_t=0$. □

As an application of the stochastic sewing lemma, let us discuss a regularization by noise result which in this formulation is inspired by Lê [Lê20], but see also [CG16] and others. To simplify the presentation we restrict to one-dimensional paths, but everything works in exactly the same way in higher dimensions. We start with a lemma, in which $B^{-\alpha}_{\infty,\infty}$ is a Besov space, see [BCD11] or simply imagine a space consisting of distributional derivatives of Hölder continuous functions.

Lemma 6.2. *Let B be a one-dimensional fractional Brownian motion with Hurst index $H\in(0,1)$ and let $f\in C_b^\infty(\mathbb{R},\mathbb{R})$ and $\alpha<\frac{1}{2H}$ and $p\geqslant 2$. Then*

$$\left\|\int_s^t f(B_r)\mathrm{d}r\right\|_{L^p}\lesssim \|f\|_{B^{-\alpha}_{\infty,\infty}}|t-s|^{1-\alpha H}.$$

Proof. Define

$$\Xi_{s,t}:=\mathbb{E}\left[\int_s^t f(B_r)\mathrm{d}r\middle|\mathcal{F}_s\right].$$

Since f is smooth we have

$$\|\Xi_{s,t}-f(B_s)(t-s)\|_{L^p}\lesssim|t-s|^{1+H}$$

and therefore the Riemann sums converge to the same limit and

$$\mathcal{I}\Xi_t=\int_0^t f(B_r)\mathrm{d}r.$$

But working with Ξ will give us better estimates. Indeed, observe that by the tower property of conditional expectation

$$\mathbb{E}[\delta\Xi_{s,u,t}|\mathcal{F}_s]=\mathbb{E}\left[\int_s^t f(B_r)\mathrm{d}r-\int_s^u f(B_r)\mathrm{d}r-\int_u^t f(B_r)\mathrm{d}r\middle|\mathcal{F}_s\right]=0.$$

Since $\delta\Xi_{s,u,t}=\Xi_{s,t}-\Xi_{s,u}-\Xi_{u,t}$, it now suffices to bound $\|\Xi_{s,t}\|_{L^p}\lesssim|t-s|^{\frac{1}{2}+\varepsilon}$ to apply stochastic sewing. We can write the fractional Brownian motion as

$$B_r=\int_{-\infty}^s K(r,u)\mathrm{d}W_u+\int_s^u K(r,u)\mathrm{d}W_u,$$

where W is a Brownian motion and the kernel K satisfies

$$|K(r,u)| \lesssim \begin{cases} (r-u)^{H-\frac{1}{2}}, & |r-u| \ll 1, \\ (r-u)^{H-\frac{3}{2}}, & |r-u| \gg 1. \end{cases}$$

Then[6.2] $\int_{-\infty}^{s} K(r,u)\mathrm{d}W_u$ is $\mathcal{F}_s$-measurable and $\int_s^u K(r,u)\mathrm{d}W_u$ is independent of $\mathcal{F}_s$ and therefore

$$\mathbb{E}\left[\int_s^t f(B_r)\mathrm{d}r \middle| \mathcal{F}_s\right] = \int_s^t \mathbb{E}\left[f\left(x + \int_s^r K(r,u)\mathrm{d}W_u\right)\right]\bigg|_{x=\int_{-\infty}^{s} K(r,u)\mathrm{d}W_u} \mathrm{d}r,$$

where

$$\begin{aligned} \int_s^r K(r,u)\mathrm{d}W_u &\sim \mathcal{N}\left(0, \int_s^r K(r,u)^2\mathrm{d}u\right) \\ &= \mathcal{N}\left(0, \int_s^r (r-u)^{2H-1}\mathrm{d}r\right) \\ &= \mathcal{N}\left(0, \frac{1}{2H}(r-s)^{2H}\right). \end{aligned}$$

Therefore,

$$\mathbb{E}\left[f\left(x + \int_s^r K(r,u)\mathrm{d}W_u\right)\right] = \int f(x+y) P_{\frac{1}{2H}(r-s)^{2H}}(y)\mathrm{d}y = f * P_{\frac{1}{2H}(r-s)^{2H}},$$

where P_{σ^2} is the heat kernel with variance σ^2. It is one of the standard results in Besov spaces and PDEs (see e.g. [GIP15] for a presentation that fits here) that

$$\|g * P_t\|_\infty \lesssim t^{-\frac{\alpha}{2}} \|g\|_{B_{\infty,\infty}^{-\alpha}},$$

so we obtain

$$\left|\mathbb{E}\left[f\left(x + \int_s^r K(r,u)\mathrm{d}W_u\right)\right]\right| \lesssim \|f\|_{B_{\infty,\infty}^{-\alpha}} (r-s)^{-\alpha H}$$

and then, using that $\alpha < \frac{1}{2H}$ so that $\alpha H < \frac{1}{2} < 1$,

$$\left\|\mathbb{E}\left[\int_s^t f(B_r)\mathrm{d}r \middle| \mathcal{F}_s\right]\right\|_{L^p} \lesssim \|f\|_{B_{\infty,\infty}^{-\alpha}} \int_s^t (r-s)^{-\alpha H}\mathrm{d}r \lesssim \|f\|_{B_{\infty,\infty}^{-\alpha}} (t-s)^{1-\alpha H}.$$

Since $\alpha H < \frac{1}{2}$, the right hand side is of the form $|t-s|^{\frac{1}{2}+\varepsilon}$, and the stochastic sewing lemma applies and shows that

$$\begin{aligned} \left\|\int_s^t f(B_r)\mathrm{d}r\right\|_{L^p} &\leqslant \left\|\int_s^t f(B_r)\mathrm{d}r - \Xi_{s,t}\right\|_{L^p} + \|\Xi_{s,t}\|_{L^p} \\ &\lesssim \|f\|_{B_{\infty,\infty}^{-\alpha}} (t-s)^{1-\alpha H}. \end{aligned}$$

□

Together with an approximation argument, this lemma allows us to make sense of

$$\int_0^t f(B_r)\mathrm{d}r$$

6.2. One can show that $\mathcal{F}_t^B = \mathcal{F}_t^W$.

whenever $f \in B^{-\alpha}_{\infty,\infty}$ for $\alpha < \frac{1}{2H}$, and in that case the integral is almost surely $1 - \alpha H - \delta$ Hölder continuous for any $\delta > 0$, so in particular $\frac{1}{2} + \varepsilon$ Hölder continuous for some $\varepsilon > 0$.

With (a lot) more work we can show that if Y is a stochastic process satisfying

$$Y_t = x + \lim_{n \to \infty} \int_0^t b_n(Y_s) \mathrm{d}s + B_t,$$

where $b_n = b * \rho_n$ with a mollifier sequence (ρ_n) and with $b \in B^{1-\alpha}_{\infty,\infty}$ (plus additional technical conditions), then we have the same estimate for

$$\int_0^t f(Y_r) \mathrm{d}r$$

as for $\int_0^t f(B_r) \mathrm{d}r$. In particular, $Y_t = x + \psi_t + B_t$, where

$$\psi_t = \int_0^t b(Y_s) \mathrm{d}s := \lim_{n \to \infty} \int_0^t b_n(Y_s) \mathrm{d}s$$

is $\frac{1}{2} + \varepsilon$ Hölder continuous.

Theorem 6.3. *Let $H < \frac{1}{2}$ and*[6.3] $\alpha < \frac{1}{2H} - 1$*. Let Y and $\tilde{Y}$ be adapted solutions to*

$$Y_t = x + \psi_t + B_t, \qquad \tilde{Y}_t = x + \tilde{\psi}_t + B_t,$$

where

$$\psi_t = \int_0^t b(Y_s) \mathrm{d}s := \lim_{n \to \infty} \int_0^t b_n(Y_s) \mathrm{d}s, \qquad \tilde{\psi}_t = \int_0^t b(\tilde{Y}_s) \mathrm{d}s := \lim_{n \to \infty} \int_0^t b_n(\tilde{Y}_s) \mathrm{d}s$$

are such that $\mathbb{E}[\| \psi \|^p_{1/2+\varepsilon}] + \mathbb{E}[\| \tilde{\psi} \|^p_{1/2+\varepsilon}] < \infty$ for all $p \geqslant 2$. Assume that $b \in B^{1-\alpha}_{\infty,\infty}$. Then Y and $\tilde{Y}$ are indistinguishable.

Proof. Lê's idea is to write with a Taylor expansion

$$\begin{aligned} Y_t - \tilde{Y}_t &= \psi_t - \tilde{\psi}_t \\ &= \lim_{n \to \infty} \int_0^t (b_n(x + \psi_s + B_s) - b_n(x + \tilde{\psi}_s + B_s)) \mathrm{d}s \\ &= \lim_{n \to \infty} \int_0^t (\psi_s - \tilde{\psi}_s) \underbrace{\int_0^1 b_n'(x + \psi_s + \lambda(\tilde{\psi}_s - \psi_s) + B_s) \mathrm{d}\lambda \mathrm{d}s}_{=: \mathrm{d}X^n_s}, \end{aligned}$$

where

$$X^n_t = \int_0^t \int_0^1 b_n'(x + \psi_s + \lambda(\tilde{\psi}_s - \psi_s) + B_s) \mathrm{d}\lambda \mathrm{d}s = \int_0^1 \int_0^t b_n'(\zeta^\lambda_s + B_s) \mathrm{d}s \mathrm{d}\lambda,$$

where ζ^λ is an adapted stochastic process with $\| \| \zeta^\lambda \|_{1/2+\varepsilon} \|_{L^p} \lesssim 1$, uniformly in λ.

Once we show that X^n converges in probability to a process that is almost surely $\frac{1}{2} + \varepsilon$ Hölder continuous, we are done: Then $\psi_t - \tilde{\psi}_t$ solves a linear Young differential equation with zero initial data, and thus is $\equiv 0$ by uniqueness. As $b \in B^{1-\alpha}_{\infty,\infty}$ we get $b' \in B^{-\alpha}_{\infty,\infty}$ and thus the integral $\int_0^t b_n'(\zeta^\lambda_s + B_s) \mathrm{d}s$ is of a similar type as in the previous lemma, except for the Hölder continuous perturbation ζ^λ.

We define similarly to the previous proof

$$\Xi_{s,t} := \Xi^{n,\lambda}_{s,t} := \mathbb{E}\left[\int_s^t b_n'(\zeta^\lambda_s + B_r) \mathrm{d}r \middle| \mathcal{F}_s \right],$$

6.3. $\alpha < \frac{1}{2H}$ should be sufficient, I may have made a mistake somewhere in the computations.

which satisfies

$$\|\Xi_{s,t} - b_n'(\zeta_s^\lambda + B_s)(t-s)\|_{L^p} \lesssim \|b_n''\|_\infty |t-s|^{1+H}$$

and thus $\mathcal{I}\Xi_t = \int_0^t b_n'(\zeta_s^\lambda + B_s)\mathrm{d}s$ (here n is fixed). However, now we unfortunately no longer have $\mathbb{E}[\delta\Xi_{s,u,t}|\mathcal{F}_s] = 0$ and instead

$$\begin{aligned}\delta\Xi_{s,u,t} &= \mathbb{E}\bigg[\int_s^t b_n'(\zeta_s^\lambda + B_r)\mathrm{d}r\bigg|\mathcal{F}_s\bigg] - \mathbb{E}\bigg[\int_s^u b_n'(\zeta_s^\lambda + B_r)\mathrm{d}r\bigg|\mathcal{F}_s\bigg] - \mathbb{E}\bigg[\int_u^t b_n'(\zeta_u^\lambda + B_r)\mathrm{d}r\bigg|\mathcal{F}_u\bigg]\\ &= \mathbb{E}\bigg[\int_u^t b_n'(\zeta_s^\lambda + B_r)\mathrm{d}r\bigg|\mathcal{F}_s\bigg] - \mathbb{E}\bigg[\int_u^t b_n'(\zeta_u^\lambda + B_r)\mathrm{d}r\bigg|\mathcal{F}_u\bigg]\end{aligned}$$

and thus

$$\begin{aligned}\mathbb{E}[\delta\Xi_{s,u,t}|\mathcal{F}_s] &= \mathbb{E}\bigg[\int_u^t (b_n'(\zeta_s^\lambda + B_r) - b_n'(\zeta_u^\lambda + B_r))\mathrm{d}r\bigg|\mathcal{F}_s\bigg]\\ &= \mathbb{E}\bigg[\mathbb{E}\bigg[\int_u^t (b_n'(\zeta_s^\lambda + B_r) - b_n'(\zeta_u^\lambda + B_r))\mathrm{d}r\bigg|\mathcal{F}_u\bigg]\bigg|\mathcal{F}_s\bigg]\end{aligned}$$

by the tower property. The inner conditional expectation is with $\tilde{B}_{u,r} := \int_{-\infty}^u K(r,s)\mathrm{d}W_s$

$$\begin{aligned}&\bigg|\mathbb{E}\bigg[\int_u^t (b_n'(\zeta_s^\lambda + B_r) - b_n'(\zeta_u^\lambda + B_r))\mathrm{d}r\bigg|\mathcal{F}_u\bigg]\bigg|\\ &= \bigg|\int_u^t \Big(P_{\frac{1}{2H}(r-s)^{2H}} * b_n'\Big)(\zeta_s^\lambda + \tilde{B}_{u,r}) - \Big(P_{\frac{1}{2H}(r-s)^{2H}} * b_n'\Big)(\zeta_u^\lambda + \tilde{B}_{u,r})\mathrm{d}r\bigg|\\ &\lesssim \int_u^t \Big\|P_{\frac{1}{2H}(r-s)^{2H}} * b_n''\Big\|_\infty |\zeta_s^\lambda - \zeta_u^\lambda|\mathrm{d}r\\ &\lesssim \int_u^t (r-s)^{-(\alpha+1)H}\|b''\|_{B_{\infty,\infty}^{-\alpha-1}}(s-u)^{\frac{1}{2}+\varepsilon}\|\zeta^\lambda\|_{1/2+\varepsilon}\mathrm{d}r\\ &\lesssim (t-s)^{1-(\alpha+1)H+\frac{1}{2}+\varepsilon}\|b''\|_{B_{\infty,\infty}^{-\alpha-1}}\|\zeta^\lambda\|_{1/2+\varepsilon},\end{aligned}$$

where we used that $\alpha < \frac{1}{2H}$ and thus $(\alpha+1)H < \frac{1}{2} + H < 1$. Since even $\alpha < \frac{1}{2H} - 1$, the exponent on $(t-s)$ is greater than 1 and together with the bound

$$\|\delta\Xi_{s,u,t}\|_{L^p} \lesssim \|b'\|_{B_{\infty,\infty}^{-\alpha}}(t-s)^{1-\alpha H}$$

we obtain that Ξ can be sewed and then that the claimed uniqueness holds. □

Bibliography

[BCD11] Hajer Bahouri, Jean-Yves Chemin, and Raphael Danchin. *Fourier analysis and nonlinear partial differential equations*. Springer, 2011.

[CG16] Rémi Catellier and Massimiliano Gubinelli. Averaging along irregular curves and regularisation of ODEs. *Stochastic Process. Appl.*, 126(8):2323–2366, 2016.

[CK16] Ilya Chevyrev and Andrey Kormilitzin. A primer on the signature method in machine learning. *ArXiv preprint arXiv:1603.03788*, 2016.

[CQ02] Laure Coutin and Zhongmin Qian. Stochastic analysis, rough path analysis and fractional Brownian motions. *Probab. Theory Related Fields*, 122(1):108–140, 2002.

[CS24] Thomas Cass and Cristopher Salvi. Lecture notes on rough paths and applications to machine learning. *ArXiv preprint arXiv:2404.06583*, 2024.

[DŁMP23] Purba Das, Rafał Łochowski, Toyomu Matsuda, and Nicolas Perkowski. Level crossings of fractional brownian motion. *ArXiv preprint arXiv:2308.08274*, 2023.

[Dav07] A. M. Davie. Differential equations driven by rough paths: an approach via discrete approximation. *Appl. Math. Res. Express. AMRX*, (2):0, 2007.

[DL89] R. J. DiPerna and P.-L. Lions. Ordinary differential equations, transport theory and Sobolev spaces. *Invent. Math.*, 98(3):511–547, 1989.

[FH14] Peter K. Friz and Martin Hairer. *A course on rough paths*. Universitext. Springer, Cham, 2014. With an introduction to regularity structures.

[FHL21] Peter K Friz, Antoine Hocquet, and Khoa Lê. Rough stochastic differential equations. *ArXiv preprint arXiv:2106.10340*, 2021.

[Fla11] Franco Flandoli. *Random Perturbation of PDEs and Fluid Dynamic Models: École d'été de Probabilités de Saint-Flour XL–2010*, volume 2015. Springer Science & Business Media, 2011.

[FLMS23] Adeline Fermanian, Terry Lyons, James Morrill, and Cristopher Salvi. New directions in the applications of rough path theory. *IEEE BITS the Information Theory Magazine*, 2023.

[FV10] Peter K. Friz and Nicolas B. Victoir. *Multidimensional stochastic processes as rough paths*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Theory and applications.

[GIP15] Massimiliano Gubinelli, Peter Imkeller, and Nicolas Perkowski. Paracontrolled distributions and singular PDEs. *Forum of Mathematics, Pi*, 3(e6), 2015.

[GJ14] Patrícia Gonçalves and Milton Jara. Nonlinear fluctuations of weakly asymmetric interacting particle systems. *Arch. Ration. Mech. Anal.*, 212(2):597–644, 2014.

[GP18] Massimiliano Gubinelli and Nicolas Perkowski. Energy solutions of KPZ are unique. *J. Amer. Math. Soc.*, 31(2):427–471, 2018.

[GP20] Massimiliano Gubinelli and Nicolas Perkowski. The infinitesimal generator of the stochastic Burgers equation. *Probab. Theory Related Fields*, 178(3-4):1067–1124, 2020.

[Gub04] Massimiliano Gubinelli. Controlling rough paths. *J. Funct. Anal.*, 216(1):86–140, nov 2004.

[Gub10] Massimiliano Gubinelli. Ramification of rough paths. *J. Differential Equations*, 248(4):693–721, 2010.

[H+98] J. van der Hoeven et al. GNU TeXmacs. https://www.texmacs.org, 1998.

[Hai14] Martin Hairer. A theory of regularity structures. *Invent. Math.*, 198(2):269–504, 2014.

[Hai24] Martin Hairer. Renormalisation in the presence of variance blowup. *ArXiv:2401.10868*, 2024.

[HL10] Ben Hambly and Terry Lyons. Uniqueness for the signature of a path of bounded variation and the reduced path group. *Ann. of Math. (2)*, 171(1):109–167, 2010.

[HL20] Martin Hairer and Xue-Mei Li. Averaging dynamics driven by fractional Brownian motion. *Ann. Probab.*, 48(4):1826–1860, 2020.

[HP21] Fabian Andsem Harang and Nicolas Perkowski. C^∞-regularization of ODEs perturbed by noise. *Stoch. Dyn.*, 21(8):0, 2021.

[Jan97] Svante Janson. *Gaussian Hilbert spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.

[KM17] David Kelly and Ian Melbourne. Deterministic homogenization for fast-slow systems with chaotic noise. *J. Funct. Anal.*, 272(10):4063–4102, 2017.

[KPZ86] Mehran Kardar, Giorgio Parisi, and Yi-Cheng Zhang. Dynamic scaling of growing interfaces. *Physical Review Letters*, 56(9):889–892, 1986.

[KS91] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, Second edition, 1991.

[LCL07] Terry J. Lyons, Michael Caruana, and Thierry Lévy. *Differential equations driven by rough paths*, volume 1908 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.

[Lê20] Khoa Lê. A stochastic sewing lemma and applications. *Electron. J. Probab.*, 25:0, 2020.

[Liv96] Carlangelo Liverani. Central limit theorem for deterministic systems. In *International Conference on Dynamical Systems (Montevideo, 1995)*, volume 362 of *Pitman Res. Notes Math. Ser.*, pages 56–75. Longman, Harlow, 1996.

[LL21] Khoa Lê and Chengcheng Ling. Taming singular stochastic differential equations: a numerical method. *ArXiv preprint arXiv:2110.01343*, 2021.

[LM22] Terry Lyons and Andrew D McLeod. Signature methods in machine learning. *ArXiv preprint arXiv:2206.14674*, 2022.

[Lyo94] Terry Lyons. Differential equations driven by rough signals. I. An extension of an inequality of L. C. Young. *Math. Res. Lett.*, 1(4):451–464, 1994.

[Lyo98] Terry J. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoam.*, 14(2):215–310, 1998.

[Lyo14] Terry Lyons. Rough paths, signatures and the modelling of functions on streams. In *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. IV*, pages 163–184. Kyung Moon Sa, Seoul, 2014.

[MP24] Toyomu Matsuda and Nicolas Perkowski. An extension of the stochastic sewing lemma and applications to fractional stochastic calculus. *Forum Math. Sigma*, 12:0, 2024.

[MvN68] Benoit B. Mandelbrot and John W. van Ness. Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Rev.*, 10(4):422–437, 1968.

[NT11] David Nualart and Samy Tindel. A construction of the rough path above fractional Brownian motion using Volterra's representation. *The Annals of Probability*, 39(3):1061–1096, 2011.

[RY99] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, Third edition, 1999.

[Unt10] Jérémie Unterberger. A rough path over multidimensional fractional Brownian motion with arbitrary Hurst index by Fourier normal ordering. *Stochastic Processes and their Applications*, 120(8):1444–1472, 2010.

[Ver81] Alexander Yu. Veretennikov. On strong solution and explicit formulas for solutions of stochastic integral equations. *Math. USSR Sb.*, 39:387–403, 1981.

[You36] Laurence C. Young. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Math.*, 67(1):251–282, 1936.

[Zvo74] A. K. Zvonkin. A transformation of the phase space of a diffusion process that will remove the drift. *Mat. Sb. (N.S.)*, 93(135):129–149, 1974.