Leibniz Universität Hannover

Fakultät für Mathematik und Physik
Institut für Analysis

Stochastic Optimal Control using Signatures

Master thesis

In partial fulfillment of the requirements for the degree of Master of Science in Mathematics.

Author:	Tobias Christian Nauen
Matriculation Number:	10006435
First Examiner:	Prof. Dr. Sebastian Riedel
Second Examiner:	Prof. Dr. Thomas Wick
Date:	June 9, 2022

Contents

1 Introduction

In this thesis, we consider a stochastic control problem of the form

$$dY_t = \mu_t b(Y_t)dt + \sigma(Y_t)dB_t, \tag{1}$$

where μ_t is an $\mathcal{F}_t = \sigma(B_s|s \leq t)$-measurable, continuous process we have some control over. An SDE of this form can be found when one considers a noisy process, where only some control on the drift, i.e. the average direction is given. This control manifests itself in the function $\mu_t : [0,T] \to \mathbb{R}$.
A toy example for a problem of this kind could be modeling navigating on the seas or in space, where the random part is the combined influence of winds and currents on a boat and μ_t represents the direction of the rudder, or in the space example, the randomness represents course altering events like solar winds and μ_t is the direction or strength of thrust. A similar optimal control problem with control in the drift was considered in [DFG17], which investigates the value function to find a dual problem. This was the first paper on stochastic optimal control, using rough path analysis.
We now use the ansatz

$$\mu_t = \Theta(B|_{[0,t]}) \qquad\qquad \Theta \in C(\Lambda_T, \mathbb{R}) =: \mathcal{T},$$

with Λ_T being the space of stopped rough paths up to time T (see Definition 5.2).
This gives the SDE

$$dY_t^\mu = \underbrace{\Theta(\hat{B}|_{[0,t]})}_{=\mu_t} b(Y_t^\mu)dt + \sigma(Y_t^\mu)dB_t. \tag{2}$$

We can now define a loss-function like

$$L(Y^\mu) := \mathbb{E}(Y_T^\mu)^2 + \mathbb{E}(\left|Y_T^\mu\right|^2),$$

but in general all losses $L : C([0,T], \mathbb{R}^m) \to \mathbb{R}^+$ Lipschitz or Hölder continuous are possible.
The question we want to answer is: What is $\inf_\mu \mathbb{E}[L(Y^\mu)]$ and what does the corresponding μ (and Θ) look like? It is the question of an optimal way to act, while counteracting random noise.
First, we need to understand our main problem in Equation (1). This is a shorthand notation for

$$Y_t - Y_0 = \int_0^t \mu_\tau b(Y_\tau)d\tau + \int_0^t \sigma(Y_\tau)dB_\tau. \tag{3}$$

The second integral in Equation (3) can be seen as an Itô integral. However, we will view it as the integral over a rough path, a so-called rough integral. This is a generalization of the Itô-Map, which can also incorporate other types of stochastic integrals, like the Stratonovich-Integral. This change of perspective is useful since we want to look at the so-called signatures of some processes, which are defined naturally in the context of rough paths.
The theory of rough paths was first introduced in the 1990s by Terry Lyons [Lyo98]. It is an elegant framework for path-wise integration with rough driving signals and is therefore suited to a general class of stochastic processes, like Brownian motion or fractional Brownian motion. In particular rough integrals are a generalization of Young's theory of integration. An important aspect of the theory is the continuity of the solution map of rough differential equations, which is not given in the classical case of Itô SDEs, where the solution map is measurable, but not continuous.

In addition to theoretical advances in SDEs, there were additional tools developed for rough paths, most notably the signature. The signature $\mathbb{X}^{<\infty}$ of a path $x : [0,T] \to \mathbb{R}^n$ is a collection of iterated integrals of all components of the path against each other;

$$\int_{0\leq t_1\leq ...\leq t_k\leq t} dx_{t_1}^{i_1}...dx_{t_k}^{i_k}$$

for $k \in \mathbb{N}$ and $i_1, ..., i_k \in \{1, ..., n\}$. Now, the values of the signature have to be defined up to a certain level k, which depends on the roughness of the underlying path. To see why this is true, one can consider the differences between the Itô and Stratonovich integrals, which both are fair definitions of integrals with respect to Brownian motion. We have

$$\int_{0\leq t_1\leq t_2\leq T} dB_{t_1}dB_{t_2} = \int_0^T B_t dB_t = \frac{B_T^2}{2} + \frac{T}{2},$$

but also

$$\int_{0\leq t_1\leq t_2\leq T} \circ dB_{t_1} \circ dB_{t_2} = \int_0^T B_t \circ dB_t = \frac{B_T^2}{2},$$

which makes it clear, that there is not one single way of defining the signature of a process. This is why, when working with iterated integrals, one has to set one way of calculation. The theory of rough paths gives a framework for doing exactly that. The signature of a path is important because the signature at time t determines the whole path up to time t up to so-called tree-like extensions. In particular the signature of an augmented rough path, i.e. a path $x_t = (x_t^{(1)}, ..., x_t^{(n)})$ with an additional dimension that represents the time

$$\hat{x}_t = (x_t^{(1)}, ..., x_t^{(n)}, t) \in \mathbb{R}^{n+1}$$

is unique. This makes the signature an important tool in machine learning as a model-free way to extract features from time-series data, like audio, speech, or character drawing. As such it has been used successfully in several machine learning applications [CK16] including Chinese character recognition [Gra13] or even medical tasks like the recognition of mental disorders [Arr+18].

The property of injectiveness of the signature map also makes it important to us and is why we take the following ansatz for answering the question from above:

$$\Theta(\hat{B}|_{[0,t]}) = \langle \ell, \hat{B}_{0,t}^{<\infty} \rangle.$$

Here, $\hat{B}_{0,t}^{<\infty}$ is the signature of the augmented path of Brownian motion. In this, we will follow the reasoning of [KLA20] and [Bay+22], where it was shown that similar control problems of optimal trading speed and optimal stopping can be solved by just using linear maps of the path signature.

The main result of this thesis will be

Theorem 5.6:
Let $2 \leq p < 3$ and let $\mathbb{P}$ be a probability measure on $\left(\hat{\Omega}_T^p, \mathcal{B}(\hat{\Omega}_T^p)\right)$. Let Y^μ be the unique solution to

$$dY = \mu_t b(Y_t)dt + \sigma(Y_t)d\mathbf{x}$$

started at $\xi \in \mathbb{R}^m$, with $\mu \in \mathcal{T}$, b Lipschitz, and $\sigma \in C_b^3(\mathbb{R}^m, \mathbb{R}^{m\times n})$. Here, the $\mathbf{x}$ is a random geometric p-rough path with distribution determined by $\mathbb{P}$. It holds

$$\inf_{\mu\in\mathcal{T}} \mathbb{E}[L(Y^\mu)] = \inf_{\mu\in\mathcal{T}_{sig}} \mathbb{E}[L(Y^\mu)]$$

for a loss function $L : C([0,T], \mathbb{R}^m) \to \mathbb{R}$ bounded and α-Hölder for some $\alpha > 0$.

Here $\mathcal{T} = C(\Lambda_T, \mathbb{R})$ is the set of all continuous functions of the path up to some time $t \in [0,T]$, while $\mathcal{T}_{sig}$ is the set of all functions of the form $\langle \ell, \hat{\mathbb{X}}_{0,t}^{<\infty} \rangle$. The theorem, therefore, says, that the optimal control problem can be solved by considering just linear maps of the signature of the augmented path. The statement will then also be extended to Itô integrals, as considered in Equation (3), in Theorem 5.8.

Using these theorems, we can tackle our question numerically by modeling $\mu_t = \langle \ell, \hat{B}_{0,t}^{\leq k} \rangle$ to be a linear map. Here, we drop from the infinite-dimensional, full signature $\hat{B}_{0,t}^{<\infty}$ to the finite-dimensional, truncated signature $\hat{B}_{0,t}^{\leq k}$ for numerical reasons. This is a good approximation, as

$$\left\| \hat{B}_{s,t}^{k} \right\| \leq C \frac{\omega(s,t)^{\frac{k}{p}}}{\left(\frac{k}{p}\right)!}$$

(see [LCL07, Theorem 3.7]), i.e. the norms of additional signature levels decrease like $\frac{1}{k!}$. We can approximate the RDE's solution by using a Milstein scheme (Algorithm 3) on a discrete time-grid

$$0 = t_0 < t_1 < ... < t_k = T$$

and estimating the expected loss $\mathbb{E}[L(\theta)]$ after many such simulations. Using the backpropagation algorithm then can lead us arbitrarily close to the optimal solution μ_t.

At first, we will introduce the theory of rough paths with its basic facts and definitions and derive rough integrals as a limit of Riemann-like sums in Section 2. Throughout the thesis, we will work with general rough paths with finite p-variation for $p \in [2,3)$, where Young integration breaks down. For ease of notation, we will introduce a tensor calculus. In this section, a general setting of controlled rough paths is also established that deals with all kinds of rough paths as opposed to the theory of [FH20] only considering α-Hölder paths. After that, in Section 3, we will deal with rough differential equations (RDEs). We will prove the existence and uniqueness of solutions in the usual way via Picard iteration, but then extend the theory to RDEs with drift term, where we will only require very mild assumptions on the drift term, such that we can incorporate all RDEs of the form seen in Equation (1) for b Lipschitz and μ continuous. We also investigate the stability of RDEs in the drift term. After having introduced RDEs , we will move on to signatures in Section 4, where we will see the basic definitions, along with a proof of the shuffle identity for geometric rough paths. This is directly followed by the proof of our main theorem, Theorem 5.6, in Section 5. Here, we will exploit the notion of stopped rough paths, as well as Lemma 5.5 which has also been used in [KLA20] and [Bay+22] to show the density of signature controls on compact sets of arbitrary high probability (< 1). We then expand the main theorem to work with Itô-integrals. After proving the theoretical results, we will go on to state numerical algorithms which can be used for approximation and which are also implemented and can be viewed on GitHub[1], as well as some convergence results for said algorithms in Section 6. Then, in Section 7, we test our implementation against a julia reference implementation [RN17] based on two SDE problems. We also use our framework to solve an optimal asset allocation problem in the Black-Scholes model. The SDE of this problem is of a different structure than we had before and we argue why the same approach we took (approximating $\mu \in \mathcal{T}$ by $\mu \in \mathcal{T}_{sig}$) can also be done when one has combined control over the drift and volatility terms. Here, we use the Markov property of Brownian motion and neural networks to choose the control term to be $C(\mathbb{R}^{m+1}, \mathbb{R})$ instead of a linear function of the signature of the process. In the end (Section 8) we will discuss some extensions of the problem, as well as different possibilities of defining the Gubinelli derivative of RDE solutions when dealing with a drift term.

[1] `https://github.com/tobna/DeepRoughPaths`

2 Rough Paths & Rough Integrals

The theory of rough paths deals with paths

$$x : [0,T] \to \mathbb{R}^n.$$

An important set of features is the finiteness of semi-norms of p-variation for $p > 0$, which measure the smoothness of such a path. Now the finding is, that depending on their finiteness one needs additional pieces of information to define certain integrals with respect to the path. The theory of rough paths is well suited to stochastic processes because their sample paths are oftentimes too rough to work with other types of integrals (Young integrals). In this thesis, we will always deal with the fixed time horizon $[0,T]$ for some $T > 0$. This section is based on [Lej03] and [FH20], taking the general approach of [Lej03] and combining it with some of the results and definitions of [FH20], which only deals with α-Hölder rough paths.

2.1 p-variation

To classify the roughness of paths, we introduce the semi-norm of p-variation.

Definition 2.1 (p-variation)**:**
Let $x : [0,T] \to \mathbb{R}^n$ be a path. The p-variation of x on $[s,t] \subset [0,T]$ is defined by

$$\mathrm{Var}_{p,[s,t]}(x) := \mathrm{Var}_{p,[s,t]}((t_1 \le t_2) \mapsto x_{t_2} - x_{t_1}) := \sup_{\{s \le t_0 < \ldots < t_k \le t\}} \left(\sum_{j=1}^{k} \left| x_{t_j} - x_{t_{j-1}} \right|^p \right)^{\frac{1}{p}}.$$

We further define

$$\mathrm{Var}_p(x) := \mathrm{Var}_{p,[0,T]}(x).$$

First of all, we show some basic properties of the p-variation. The following lemma is a set of statements taken from [Lej03, Section 2.2].

Lemma 2.2:
Let $q > p \ge 1$, $0 \le s < t \le 1$, and $x : [0,1] \to \mathbb{R}^n$. Then $\mathrm{Var}_p(\cdot)$ is a semi-norm on the space of continuous $\mathbb{R}^n$-valued paths, and

$$\mathrm{Var}_{p,[s,t]}(x) \ge \mathrm{Var}_{q,[s,t]}(x).$$

Proof. Essentially all these statements come from the corresponding statements on p-norms $\|\cdot\|_p$ of $\mathbb{R}^n$. First, we quickly show that the p-variation is a semi-norm. For this, let $\lambda \in \mathbb{R}$, and $x, y : [0,T] \to \mathbb{R}^n$. We have

$$\begin{aligned}
\mathrm{Var}_{p,[s,t]}(\lambda x) &= \sup_{\{s \le t_0 < \ldots < t_k \le t\}} \left(\sum_{j=1}^{k} \left| \lambda x_{t_j} - \lambda x_{t_{j-1}} \right|^p \right)^{\frac{1}{p}} \\
&= \sup_{\{s \le t_0 < \ldots < t_k \le t\}} \left(\sum_{j=1}^{k} |\lambda|^p \left| x_{t_j} - x_{t_{j-1}} \right|^p \right)^{\frac{1}{p}} \\
&= |\lambda| \sup_{\{s \le t_0 < \ldots < t_k \le t\}} \left(\sum_{j=1}^{k} \left| x_{t_j} - x_{t_{j-1}} \right|^p \right)^{\frac{1}{p}} = |\lambda| \, \mathrm{Var}_{p,[s,t]}(x)
\end{aligned}$$

and

$$\begin{aligned}\mathrm{Var}_{p,[s,t]}(x+y) &= \sup_{\{s\le t_0<\ldots<t_k\le t\}} \left(\sum_{j=1}^{k} \left|x_{t_j}-x_{t_{j-1}}+y_{t_j}-y_{t_{j-1}}\right|^p\right)^{\frac{1}{p}} \\ &\underset{\textit{inequality}}{\overset{\textit{Minkowski}}{\le}} \sup_{\{s\le t_0<\ldots<t_k\le t\}} \left(\sum_{j=1}^{k} \left|x_{t_j}-x_{t_{j-1}}\right|^p\right)^{\frac{1}{p}} + \left(\sum_{j=1}^{k} \left|y_{t_j}-y_{t_{j-1}}\right|^p\right)^{\frac{1}{p}} \\ &\le \mathrm{Var}_{p,[s,t]}(x)+\mathrm{Var}_{p,[s,t]}(y).\end{aligned}$$

For $\xi \in \mathbb{R}^n$, we have

$$\begin{aligned}\|\xi\|_q = \left(\sum_{j=1}^{n}|\xi_j|^q\right)^{\frac{1}{q}} &= \|\xi\|_p \left(\sum_{j=1}^{n}\left(\frac{|\xi_j|}{\|\xi\|_p}\right)^q\right)^{\frac{1}{q}} \le \|\xi\|_p \left(\sum_{j=1}^{n}\left(\frac{|\xi_j|}{\|\xi\|_p}\right)^p\right)^{\frac{1}{q}} \\ &= \|\xi\|_p \left(\frac{\|\xi\|_p^p}{\|\xi\|_p^p}\right)^{\frac{1}{q}} = \|\xi\|_p\,,\end{aligned}$$

where the inequality comes from the fact that $t \mapsto \lambda^t$ is non-increasing for $\lambda \in [0,1]$. With that it follows

$$\mathrm{Var}_{q,[s,t]}(x) \le \mathrm{Var}_{p,[s,t]}(x).$$

□

Lemma 2.3:
Let $x : [0,T] \to \mathbb{R}^n$ be a path and $0 \le s < u < t \le 1$. Then it holds

$$\mathrm{Var}^p_{p,[s,u]}(x) + \mathrm{Var}^p_{p,[u,t]}(x) \le \mathrm{Var}^p_{p,[s,t]}(x).$$

Proof. Let $\{s \le t_0 < \ldots < t_k \le u\}$ be a partition of $[s,u]$ and let $\{u \le \tau_0 < \ldots < \tau_\kappa \le t\}$ be a partition of $[u,t]$. Then

$$\begin{aligned}\sum_{j=1}^{k}\left|x_{t_j}-x_{t_{j-1}}\right|^p + \sum_{\ell=1}^{\kappa}\left|x_{\tau_\ell}-x_{\tau_{\ell-1}}\right|^p &\le \sum_{j=1}^{k}\left|x_{t_j}-x_{t_{j-1}}\right|^p + |x_{\tau_0}-x_{t_k}|^p + \sum_{\ell=1}^{\kappa}\left|x_{\tau_\ell}-x_{\tau_{\ell-1}}\right|^p \\ &\le \mathrm{Var}^p_{p,[s,t]}(x).\end{aligned}$$

Now, going to the supremum over partitions of $[s,u]$ and $[u,t]$ gives the assertion. □

Lemma 2.4:
If $x : [0,T] \to \mathbb{R}^n$ is continuous with $\mathrm{Var}_{p,[0,T]}(x) < \infty$, then

$$(s,t) \mapsto \mathrm{Var}_{p,[s,t]}(x)$$

is uniformly continuous on $\Delta_T = \{(s,t)|0 \le s \le t \le T\}$.

Proof. We only show continuity in t. Continuity in s then follows by reflection of the underlying path. Let $\varepsilon > 0$. By continuity of x, there exists a $\delta > 0$, such that

$$|x_t - x_{t+\xi}| < \frac{\varepsilon}{2} \qquad\qquad \forall\, |\xi| < \delta$$

and then, we can find a partition $\{s \le t_0 < \ldots < t_k = t\}$ of $[s,t]$, such that

$$\mathrm{Var}_{p,[s,t]}(x) < \frac{\varepsilon}{2} + \left(\sum_{j=1}^{k}\left|x_{t_j}-x_{t_{j-1}}\right|^p\right)^{\frac{1}{p}},$$

since $\mathrm{Var}_{p,[s,t]}(x)$ is the supremum over partitions of $[s,t]$. Let $\max(t_{k-1}, t-\delta) < u_l < t = t_k$. By the triangle inequality of the p-norm on $\mathbb{R}^n$, it holds

$$\begin{aligned}\mathrm{Var}_{p,[s,t]}(x) <& \frac{\varepsilon}{2} + \left(\sum_{j=1}^{k-1} \left|x_{t_j} - x_{t_{j-1}}\right|^p + \left|x_{t_k} - x_{t_{k-1}}\right|^p\right)^{\frac{1}{p}} \\ \leq& \underbrace{\frac{\varepsilon}{2} + \left(\sum_{j=1}^{k-1} \left|x_{t_j} - x_{t_{j-1}}\right|^p + \left|x_{u_l} - x_{t_{k-1}}\right|^p\right)^{\frac{1}{p}}}_{\leq \mathrm{Var}_{p,[s,u_l]}(x)} + \underbrace{\left|x_{t_k} - x_{u_l}\right|}_{\leq \frac{\varepsilon}{2}}.\end{aligned}$$

We have shown that

$$0 \leq \mathrm{Var}_{p,[s,t]}(x) - \mathrm{Var}_{p,[s,u_l]}(x) < \varepsilon$$

for all $\max(t_{k-1}, t-\delta) < u_l < t$, which is the left-continuity of $t \mapsto \mathrm{Var}_{p,[s,t]}(x)$.
For the right-continuity, take a partition $\{t \leq t_0 < ... < t_k \leq t+\delta\}$ of $[t, t+\delta]$ with

$$\mathrm{Var}^p_{p,[t,t+\delta]}(x) < \frac{\varepsilon}{2} + \sum_{j=1}^{k} \left|x_{t_j} - x_{t_{j-1}}\right|^p.$$

Then

$$\begin{aligned}\mathrm{Var}^p_{p,[t,t+\delta]}(x) <& \frac{\varepsilon}{2} + \sum_{j=1}^{k} \left|x_{t_j} - x_{t_{j-1}}\right|^p \\ =& \frac{\varepsilon}{2} + \underbrace{\left|x_{t_1} - x_{t_0}\right|^p}_{\leq \frac{\varepsilon}{2}} + \underbrace{\sum_{j=2}^{k} \left|x_{t_j} - x_{t_{j-1}}\right|^p}_{\leq \mathrm{Var}^p_{p,[t_1,t+\delta]}(x)} \\ \leq& \varepsilon + \mathrm{Var}^p_{p,[t_1,t+\delta]}(x).\end{aligned}$$

By Lemma 2.3

$$\mathrm{Var}^p_{p,[t,u_r]}(x) \leq \mathrm{Var}^p_{p,[t,t_1]}(x) \leq \mathrm{Var}^p_{p,[t,t+\delta]}(x) - \mathrm{Var}^p_{p,[t_1,t+\delta]}(x) \leq \varepsilon,$$

for all $t < u_r < t_1$. Now we need to incorporate the left point s. We therefore consider two new, continuous paths $x^{(1)}, x^{(2)} : [0,1] \to \mathbb{R}^n$ by setting

$$x^{(1)}_\tau = \begin{cases} x_\tau & \tau < t \\ x_t & \tau \geq t \end{cases} \tag{4}$$

and

$$x^{(2)}_\tau = \begin{cases} 0 & \tau < t \\ x_\tau - x_t & \tau \geq t \end{cases}.$$

Conveniently $x^{(1)} + x^{(2)} \equiv x$ on $[0,T]$ and $x^{(1)} \equiv x$ on $[0,t] \supset [s,t]$. By the triangle inequality (in Lemma 2.2)

$$\begin{aligned}0 \leq \mathrm{Var}_{p,[s,u_r]}(x) - \mathrm{Var}_{p,[s,t]}(x) \leq& \underbrace{\mathrm{Var}_{p,[s,u_r]}(x^{(1)})}_{=\mathrm{Var}_{p,[s,t]}(x)} + \underbrace{\mathrm{Var}_{p,[s,u_r]}(x^{(2)})}_{=\mathrm{Var}_{p,[t,u_r]}(x)} - \mathrm{Var}_{p,[s,t]}(x) \\ =& \mathrm{Var}_{p,[t,u_r]}(x) \leq \sqrt[p]{\varepsilon}\end{aligned}$$

for all $t < u_r < t_1$, which is the right-continuity of $t \mapsto \mathrm{Var}_{p,[s,t]}(x)$. Since $\Delta_T \subset \mathbb{R}^2$ is compact, $\Delta_T \ni (s,t) \mapsto \mathrm{Var}_{p,[s,t]}(x)$ is also uniformly continuous. $\square$

Example 2.5 (p-variation of Brownian Motion)**:**
It is known, that Brownian motion has infinite 2-variation almost surely [FV10, Section 13.9]. Using Lemma 2.2, we see that it also has infinite p-variation for $1 \leq p \leq 2$ almost surely. Since Brownian motion has α-Hölder paths for every $0 < \alpha < \frac{1}{2}$ almost surely [SP14, Section 10.1], we see that it has finite p-variation for $p > 2$ almost surely: Let $\Pi = \{s \leq t_0 \leq ... \leq t_k \leq t\}$ be a partition of $[s,t]$ and let $p > 2$. Then

$$\sum_{j=1}^{k} \left|B_{t_j} - B_{t_{j-1}}\right|^p \leq \sum_{j=1}^{k} \|B\|^p_{\frac{1}{p}-\text{Höl}} |t_j - t_{j-1}| = \|B\|^p_{\frac{1}{p}-\text{Höl}} |t-s| \qquad a.s.$$

and the p-variation is finite a.s.
We could also use this property instead of p-variation to define α-Hölder rough paths ($\alpha = \frac{1}{p}$), substituting $\|x\|^p_{\alpha-\text{Höl}} |t-s|$ for $\operatorname{Var}^p_{p,[s,t]}(x)$ everywhere, but the p-variation version is a little more general.

Remark 2.6:
One should not confuse 2-variation with quadratic variation. The quadratic variation of a function $f : [0,T] \to \mathbb{R}^n$ is defined by

$$[f,f](T) = \lim_{\|\Pi\| \to 0} \sum_{j=1}^{n} |f(t_j) - f(t_{j-1})|^2,$$

where the limit is over a sequence of partitions of $[0,T]$ with mesh going to zero.
In particular, we have that for a standard Brownian Motion $(B_t)_{0 \leq t \leq T}$ in $\mathbb{R}$

$$\operatorname{Var}_{p,[0,T]}(B_t) = \infty \qquad a.s.$$

as we saw in Example 2.5, but

$$[B_t, B_t](T) = T \qquad a.s.$$

if the meshes $\|\Pi_j\|$ go to zero fast enough, i.e. $j^2 \|\Pi_j\| \xrightarrow{j \to \infty} 0$, otherwise we only have convergence in L^2. [SP14, Section 9.1]

The p-variation is one example of a more general class of functions, we will call controlling functions.

Definition 2.7 (Controlling Function)**:**
Let $\omega : \Delta_T := \{(s,t) | 0 \leq s \leq t \leq T\} \to \mathbb{R}^+$. We call ω a *controlling function*, if

(i) ω is bounded,

(ii) ω is uniformly continuous,

(iii) ω is super-additive, i.e.

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

for all $0 \leq s \leq u \leq t \leq T$, and

(iv) $\omega(t,t) = 0$ for all $t \in [0,T]$.

This definition is similar to [Lej03, Assumption 1].

Example 2.8 (Controlling Functions)**:**
We consider some examples of controlling functions to use later on.

(i) Let x be a process with finite p-variation. Then the mapping

$$(s,t) \mapsto \mathrm{Var}^p_{p,[s,t]}(x)$$

is a controlling function by Lemma 2.2, Lemma 2.3, and Lemma 2.4.

(ii) The map

$$(s,t) \mapsto |t-s|$$

is a controlling function.

(iii) Let ω_1 and ω_2 be controlling functions, then so is $\omega_1 + \omega_2$. In particular

$$(s,t) \mapsto \mathrm{Var}^p_{p,[s,t]}(x) + |t-s|$$

is a controlling function.

Lemma 2.9:
Let ω be a controlling function and $\gamma > 1$. Then ω^γ is super-additive.

Proof. Using the super-additivity of ω, we have

$$\omega^\gamma(s,t) \geq (\omega(s,u) + \omega(u,t))^\gamma \geq \omega^\gamma(s,u) + \omega^\gamma(u,t)$$

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

2.2 The Case $p < 2$

To define rough paths, we first look at the most simple case ($p < 2$). Here we can define certain integrals without further information, using the standard way of Riemann sums.
Therefore, let $1 \leq p < 2$ and $f : \mathbb{R}^n \to \mathbb{R}^{m \times n}$ be a bounded, α-Hölder continuous function, with $\alpha > p - 1$. Let $x : [0,T] \to \mathbb{R}^n$. For our summands of the Riemann sum, we approximate small parts of the integral as follows. For $0 \leq s \leq t \leq T$, define

$$y_{s,t} := f(x_s)(x_t - x_s) \in \mathbb{R}^m.$$

We then have

$$y_{s,t} \approx \int_s^t f(x_\tau) dx_\tau$$

for small $|t-s|$. Now, for some partition $\Pi = \{s \leq t_0 < ... < t_k \leq t\}$ of $[s,t]$ we additionally define

$$z^\Pi_{s,t} := \sum_{j=1}^k y_{t_{j-1},t_j}$$

and the meshsize

$$\|\Pi\| := \sup_{j=1,\dots,k} |t_j - t_{j-1}|.$$

Theorem 2.10:
For a sequence $(\Pi_j)_j$ of partitions of $[s,t]$ with vanishing mesh sizes, $z^{\Pi_j}_{s,t}$ admits a unique limit $z_{s,t}$ for every $0 \leq s \leq t \leq T$. The map $(s,t) \mapsto z_{s,t}$ is continuous and we have the relation

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

for all $0 \leq s \leq u \leq t \leq T$. Furthermore, the map $t \mapsto z_{0,t}$ has finite p-variation with

$$|z_{s,t}| \leq K \, \mathrm{Var}_{p,[s,t]}(x)$$

for some $K > 0$.

At this point, we will outsource some of the technicalities of the proof of Theorem 2.10 into another lemma. This is so that we can use it later on for the definition of the rough integral. Both Theorem 2.10 and Lemma 2.11 are based on [Lej03, Proposition 1], while the sewing lemma incorporates some arguments from [FH20, Lemma 4.2].

Lemma 2.11 (Sewing Lemma)**:**
Let $\alpha, p > 1$. Let ω be a controlling function. Let $\mathcal{A}_p^\omega$ be the space

$$\mathcal{A}_p^\omega := \{\mathbb{Y} : \{0 \le s \le t \le 1\} \to \mathbb{R}^m \mid \exists C_1 > 0 : |\mathbb{Y}_{s,t}|^p \le C_1 \omega(s,t) \ \forall 0 \le s \le t \le T\}$$

and

$$\mathcal{A}_{p,\alpha}^\omega := \left\{\mathbb{Y} \in \mathcal{A}_p^\omega \mid \exists C_2 > 0 : |\mathbb{Y}_{s,t} - \mathbb{Y}_{s,u} - \mathbb{Y}_{u,t}| \le C_2 \omega^\alpha(s,t) \ \forall 0 \le s \le u \le t \le T\right\}.$$

There exists a unique continuous linear map $\mathcal{I} : \mathcal{A}_{p,\alpha}^\omega \to \mathcal{A}_p^\omega$, such that there is $C = C_2 2^\alpha \zeta(\alpha) > 0$, with

$$|(\mathcal{I}\mathbb{Y})_{s,t} - \mathbb{Y}_{s,t}| \le C\omega^\alpha(s,t);$$

this is called the maximal inequality*.*

In the literature, Lemma 2.11 is known as the sewing lemma, because it tells us how to sew the increments $\mathbb{Y}_{t,t+\Delta t}$ together, to gain an additive path $\mathbb{Z}_{s,t}$.

Proof. Let $\mathbb{Y} \in \mathcal{A}_{p,\alpha}^\omega$,

$$C_1 := \sup_{0 \le s \le t \le 1} \frac{|\mathbb{Y}_{s,t}|^p}{\omega(s,t)},$$

and

$$C_2 := \sup_{0 \le s \le u \le t \le 1} \frac{|\mathbb{Y}_{s,t} - \mathbb{Y}_{s,u} - \mathbb{Y}_{u,t}|}{\omega^\alpha(s,t)}.$$

Step 1: Let $0 \le s < t \le T$, $\{s \le t_1 < ... < t_k \le t\}$ be a partition of $[s,t]$ and set $t_0 = s$ and $t_{k+1} = t$. If $k \ge 2$, there exists an $\ell \in \{1, ..., k\}$, such that

$$\omega(t_{\ell-1}, t_{\ell+1}) \le \frac{2}{k}\omega(s,t).$$

Using the super-additivity of ω, we see

$$\sum_{j=1}^k \omega(t_{j-1}, t_{j+1}) \le 2\omega(s,t).$$

Therefore, at least one summand has to fulfill the condition. For $k = 1$ the result holds with the obvious choice $\ell = 1$.
Step 2: Now, we fix a partition $\Pi = \{0 \le t_1 < ... < t_k \le T\}$ of $[0,T]$ and for $0 \le s < t \le T$ let $\Pi_{[s,t]} = \Pi \cap [s,t]$. If $\Pi_{[s,t]} = \{s \le t_j < ... < t_\Gamma \le t\}$ contains at least one point, we can use the above to find an index $\ell \in \mathbb{N}$, such that

$$\omega(t_{\ell-1}, t_{\ell+1}) \le \frac{2}{\#\Pi_{[s,t]}}\omega(s,t).$$

Let

$$\mathbb{Z}_{s,t}^\Pi := \sum_{i=j+1}^{\Gamma} \mathbb{Y}_{t_{i-1},t_i}.$$

Then

$$\left|\mathbb{Z}^{\Pi}_{s,t} - \mathbb{Z}^{\Pi\setminus\{t_\ell\}}_{s,t}\right| = \left|\mathbb{Y}_{t_{\ell-1},t_{\ell+1}} - \mathbb{Y}_{t_{\ell-1},t_\ell} - \mathbb{Y}_{t_\ell,t_{\ell+1}}\right|$$
$$\leq C_2\omega^\alpha(t_{\ell-1},t_{\ell+1}) \leq C_2\left(\frac{2}{\#\Pi_{[s,t]}}\right)^\alpha \omega^\alpha(s,t)$$

and by iterating this process, we gain the *maximal inequality*

$$\left|\mathbb{Z}^{\Pi}_{s,t} - \mathbb{Y}_{s,t}\right| \leq \underbrace{C_2 2^\alpha \zeta(\alpha)}_{=:C_3}\omega^\alpha(s,t),$$

where $\zeta(\alpha) = \sum_{n=1}^\infty \frac{1}{n^\alpha}$ denotes the Riemann zeta function, which converges, since $\alpha > 1$. The constant C_3 depends on p, α, and C_2, but crucially *not* on s, t, or the partition Π.
Step 3: Let $\tilde{\Pi}$ be another partition of $[0,T]$. We first assume $\Pi \subset \tilde{\Pi}$, then

$$\left|\mathbb{Z}^{\Pi}_{s,t} - \mathbb{Z}^{\tilde{\Pi}}_{s,t}\right| \leq \sum_{i=j+1}^{\Gamma}\left|\mathbb{Y}_{t_{i-1},t_i} - \mathbb{Z}^{\tilde{\Pi}}_{t_{i-1},t_i}\right| \leq C_3\sum_{i=j+1}^{\Gamma}\omega^\alpha(t_{i-1},t_i)$$
$$\leq C_3\omega(s,t)\sup_{i=j+1,\dots,\Gamma}\omega^{\alpha-1}(t_{i-1},t_i) \leq C_3\omega(0,T)\sup_{\substack{0\leq\tau_1\leq\tau_2\leq T\\ |\tau_1-\tau_2|\leq\|\Pi\|}}\omega^{\alpha-1}(\tau_1,\tau_2).$$

For an arbitrary partition $\tilde{\Pi}$, we then get

$$\left|\mathbb{Z}^{\Pi}_{s,t} - \mathbb{Z}^{\tilde{\Pi}}_{s,t}\right| \leq \left|\mathbb{Z}^{\Pi}_{s,t} - \mathbb{Z}^{\Pi\cup\tilde{\Pi}}_{s,t}\right| + \left|\mathbb{Z}^{\tilde{\Pi}}_{s,t} - \mathbb{Z}^{\Pi\cup\tilde{\Pi}}_{s,t}\right|$$
$$\leq 2C_3\omega(0,T)\sup_{\substack{0\leq\tau_1\leq\tau_2\leq T\\ |\tau_1-\tau_2|\leq\|\Pi\|\vee\|\tilde{\Pi}\|}}\omega^{\alpha-1}(\tau_1,\tau_2)$$

uniformly in s and t. By the uniform continuity of ω, $\alpha-1>0$, and $\omega(t,t)=0$ for all $0\leq t\leq T$, the right hand side goes to zero, as $\|\Pi\|\vee\left\|\tilde{\Pi}\right\| \to 0$. Therefore, the limit

$$\mathbb{Z}_{s,t} := \lim_{\|\Pi\|\to 0}\mathbb{Z}^{\Pi}_{s,t}$$

exists and is unique.
Step 4: By the maximal inequality,

$$|\mathbb{Z}_{s,t} - \mathbb{Y}_{s,t}| \leq C_3\omega^\alpha(s,t).$$

To show that indeed $\mathbb{Z}_{s,t} \in \mathcal{A}^\omega_p$, we use

$$|\mathbb{Z}_{s,t}|^p \leq 2^p|\mathbb{Z}_{s,t} - \mathbb{Y}_{s,t}|^p + 2^p|\mathbb{Y}_{s,t}|^p \leq \left(2^pC_3^p\omega^{\alpha p-1}(0,T) + 2^pC_1\right)\omega(s,t),$$

as $\mathbb{Y} \in \mathcal{A}^\omega_p$. This also shows the boundedness of $\mathcal{I}: \mathbb{Y} \mapsto \mathbb{Z}$ as an operator. By the definition of the $\mathbb{Z}^{\Pi}_{s,t}$ and the linearity of the limit, $\mathcal{I}$ is linear and continuous, as a bounded, linear operator. □

Now, we can apply the sewing lemma to the theorem, we want to prove.

Proof of Theorem 2.10. We show that there exists a constant $C > 0$, such that

$$|y_{s,t} - y_{s,u} - y_{u,t}| \leq C\operatorname{Var}^{1+\alpha}_{p,[s,t]}(x).$$

Since f is α-Hölder, we can choose

$$C = \sup_{x\neq y}\frac{|f(x)-f(y)|}{|x-y|^\alpha} < \infty.$$

Then we have

$$\begin{aligned}|y_{s,t} - y_{s,u} - y_{u,t}| &= |f(x_s)(x_t - x_s) - f(x_s)(x_u - x_s) - f(x_u)(x_t - x_u)| \\ &= |(f(x_s) - f(x_u))(x_t - x_u)| \leq C\,|x_s - x_u|^\alpha\,|x_t - x_u| \\ &\leq C\,\mathrm{Var}^\alpha_{p,[s,u]}(x)\,\mathrm{Var}_{p,[u,t]}(x) \leq C\,\mathrm{Var}^{1+\alpha}_{p,[s,t]}(x).\end{aligned} \tag{5}$$

Additionally it holds

$$|y_{s,t}| = |f(x_s)(x_t - x_s)| \leq \|f\|_\infty\,|x_t - x_s| \leq \|f\|_\infty\,\mathrm{Var}_{p,[s,t]}(x).$$

Now, we can simply use the sewing lemma (Lemma 2.11) with $\omega(s,t) = \mathrm{Var}^p_{p,[s,t]}(x)$ to obtain $z_{s,t}$.

□

We can now define the value of the integral over one forms $f(x_t)dx_t$.

Definition 2.12:
Let $2 > p > 1$. Let $x : [0,T] \to \mathbb{R}^n$ be a continuous path with finite p-variation and let $f : \mathbb{R}^n \to \mathbb{R}^{m\times n}$ be α-Hölder continuous with $\alpha > p-1$ and bounded. Let $0 \leq s \leq t \leq T$. We then define

$$\int_s^t f(x_\tau)dx_\tau := z_{s,t},$$

where $z_{s,t}$ is obtained from Theorem 2.10.

Remark 2.13:
The proof of Theorem 2.10 also works for $f : [0,T] \to \mathbb{R}^{m\times n}$ with

$$|f_t - f_s| \leq C\,\mathrm{Var}^\alpha_{p,[s,t]}(x)$$

and $\alpha > p-1$. In that case, we replace Equation (5) by

$$|y_{s,t} - y_{s,u} - y_{u,t}| \leq |f_u - f_s|\,|x_t - x_u| \leq C\,\mathrm{Var}^\alpha_{p,[s,u]}(x)\,\mathrm{Var}_{p,[u,t]}(x) \leq C\,\mathrm{Var}^{1+\alpha}_{p,[s,t]}(x).$$

The function f_t is also automatically bounded, since

$$|f_t| \leq |f_0| + |f_t - f_0| \leq |f_0| + C\,\mathrm{Var}^\alpha_{p,[0,t]}(x) \leq |f_0| + C\,\mathrm{Var}^\alpha_{p,[0,T]}(x) < \infty.$$

Using this, we can define the integral

$$\int_s^t f_\tau dx_\tau := z_{s,t},$$

again by Theorem 2.10.

Remark 2.14:
The above lets us construct the iterated integrals of $x : [0,T] \to \mathbb{R}^n$. Consider $f_j : [0,T] \to \mathbb{R}^{1\times n}$

$$f_j(t) := (0, ..., 0, 1, 0, ..., 0),$$

where the 1 is in the j-th component of f_j. Then f_j obviously fulfills the conditions of Remark 2.13, and we can define

$$\int_s^t dx^{(j)}_\tau := \int_s^t f_j dx_\tau (= x^{(j)}_t - x^{(j)}_s).$$

We also get the rather obvious estimate $\left|x_t^{(j)} - x_s^{(j)}\right| \leq C \operatorname{Var}_{p,[s,t]}(x)$. Now, we can define the further iterated integrals by induction. Let $J = (j_1, ..., j_m) \in \{1, ..., n\}^m$ and let

$$z_{s,t}^J = \int_{s \leq t_1 \leq ... \leq t_n \leq t} dx_{t_1}^{j_1} ... dx_{t_n}^{j_n},$$

together with a constant K_J, such that

$$\left|z_{s,t}^J\right| \leq K_J \operatorname{Var}_{p,[s,t]}(x)$$

Then we define for $j_{n+1} \in \{1, ..., n\}$

$$z_{s,t}^{(j_1,...,j_{n+1})} := \int_{s \leq t_1 \leq ... \leq t_n \leq t_{n+1} \leq t} dx_{t_1}^{j_1} ... dx_{t_n}^{j_n} dx_{t_{n+1}}^{j_{n+1}} := \int_s^t f_{j_{n+1}} z_{s,\tau}^J dx_\tau,$$

using Remark 2.13 with $\alpha = 1$. We also get the existence of $K_{(j_1,...,j_{n+1})}$ with

$$\left|z_{s,t}^{(j_1,...,j_{n+1})}\right| \leq K_{(j_1,...,j_{n+1})} \operatorname{Var}_{p,[s,t]}(x).$$

2.3 Tensor Calculus

To expand these calculations to $2 \leq p < 3$ and be able to rigorously define the iterated integrals, later on, we first need to define the tensor calculus.

Definition 2.15 (Tensor Calculus)**:**
Let $n, k \in \mathbb{N}$ and $\{e^{(i)}\}_{i=1}^n$ be the canonical basis of $\mathbb{R}^n$. Let

$$\begin{aligned} W_\infty(\mathbb{R}^n) &:= \operatorname{span}\left\{\mathbb{R}, \mathbb{R}^n, (\mathbb{R}^n)^{\otimes 2}, ...\right\}, \\ W_k(\mathbb{R}^n) &:= \operatorname{span}\left\{\mathbb{R}, \mathbb{R}^n, (\mathbb{R}^n)^{\otimes 2}, ..., (\mathbb{R}^n)^{\otimes k}\right\} \end{aligned}$$

and let

$$\begin{aligned} T_\infty(\mathbb{R}^n) &:= 1 + \operatorname{span}\{\mathbb{R}^n, \mathbb{R}^n \otimes \mathbb{R}^n, ...\} \subset W_\infty(\mathbb{R}^n), \\ T_k(\mathbb{R}^n) &:= 1 + \operatorname{span}\left\{\mathbb{R}^n, \mathbb{R}^n \otimes \mathbb{R}^n, ..., (\mathbb{R}^n)^{\otimes k}\right\} \subset W_k(\mathbb{R}^n). \end{aligned}$$

These are all understood to be formal sums. For example, $T_\infty(\mathbb{R}^n)$ is the set of formal sums

$$1 + x_1 + ... + x_k + ... = 1 + \sum_{i=1}^n \underbrace{x_1^{(i)}}_{\in \mathbb{R}} e^{(i)} + ... + \sum_{i_1,...,i_k=1}^n \underbrace{x_k^{(i_1,...,i_k)}}_{\in \mathbb{R}} e^{(i_1)} \otimes ... \otimes e^{(i_k)} + ...$$

for $x_j \in (\mathbb{R}^n)^{\otimes j}$.
Let $\mathbf{x} = x_0 + x_1 + ... + x_k + ..., \mathbf{y} = y_0 + y_1 + ... + y_k + ... \in W_\infty(\mathbb{R}^n)$ with $x_j, y_j \in (\mathbb{R}^n)^{\otimes j}$ and $x_0 = y_0 = 0$. Using the convention $1 \otimes x = x \otimes 1 = x$ for $x \in (\mathbb{R}^n)^{\otimes j}$, we can now define the tensor product of $\mathbf{x}$ and $\mathbf{y}$ to be

$$\mathbf{x} \otimes \mathbf{y} := \sum_{i,j=1}^\infty \underbrace{x_i \otimes y_j}_{\in (\mathbb{R}^n)^{\otimes (i+j)}}.$$

Here, we just extend the tensor product bilinearly. We do not run into problems regarding the convergence of this sum, since for each tensor level, there are only finitely many summands. We define the tensor product on $W_k(\mathbb{R}^n)$, by first applying the tensor product on $W_\infty(\mathbb{R}^n)$ and after that applying the canonical projection from $W_\infty(\mathbb{R}^n)$ to $W_k(\mathbb{R}^n)$, i.e. ignoring summands of tensor level $> k$. When taking the tensor product in $W_k(\mathbb{R}^n)$, we can safely take x_0 and y_0

to be something different than zero, as we only have finitely many summands anyways. We will also write

$$\mathbf{x}^{\otimes j} := \underbrace{\mathbf{x} \otimes \ldots \otimes \mathbf{x}}_{j \text{ times}}$$

for $j \in \mathbb{N}$.

The following lemma is a generalization of statements from [FH20, Section 2.3] to $k > 2$.

Lemma 2.16:
Let $\mathbf{x} \in T_\infty(\mathbb{R}^n)$ or $\mathbf{x} \in T_k(\mathbb{R}^n)$. Then $\mathbf{x}$ has an inverse element $\mathbf{x}^{\otimes -1}$, such that

$$\mathbf{x} \otimes \mathbf{x}^{\otimes -1} = \mathbf{x}^{\otimes -1} \otimes \mathbf{x} = 1.$$

Proof. Let $a := \mathbf{x} - 1 \in W_\infty(\mathbb{R}^n)$. Then all summands of a are of tensor level ≥ 1. We set

$$\mathbf{x}^{\otimes -1} := 1 + \sum_{j=1}^{\infty} (-1)^j \, a^{\otimes j}.$$

This sum converges, as we only have finitely many summands for each tensor level (all summands of $a^{\otimes j}$ have tensor level $\geq j$). Then

$$\begin{aligned}\mathbf{x} \otimes \mathbf{x}^{\otimes -1} = (1 + a) \otimes \left(1 + \sum_{j=1}^{\infty} (-1)^j \, a^{\otimes j}\right) &= 1 + \sum_{j=1}^{\infty} (-1)^j \, a^{\otimes j} + a + \sum_{j=1}^{\infty} (-1)^j \, a^{\otimes j+1} \\ &= 1 - a + \sum_{j=2}^{\infty} (-1)^j \, a^{\otimes j} + a + \sum_{j=2}^{\infty} (-1)^{j-1} \, a^{\otimes j} = 1.\end{aligned}$$

Analogously, one obtains $\mathbf{x}^{\otimes -1} \otimes \mathbf{x} = 1$. In the case of T_k, the calculation is the same, just stopped at tensor level k, as in this case all tensors of higher levels just become zero by definition. □

We will also need the following:

Lemma 2.17:
Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. Then

$$|x \otimes y| = |x| \, |y| \, .$$

Proof. Let $x = \sum_{i=1}^n x_i e^{(i)}$ and $y = \sum_{j=1}^m y_j e^{(j)}$. By Parseval's identity, we have

$$|x \otimes y|^2 = \left| \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j e^{(i)} \otimes e^{(j)} \right|^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} |x_i|^2 \, |y_j|^2 = \left(\sum_{i=1}^{n} |x_i|^2 \right) \left(\sum_{j=1}^{m} |y_j|^2 \right) = |x|^2 \, |y|^2 \, ,$$

as $\{e^{(i)} \otimes e^{(j)} | i = 1, ..., n, \;\; j = 1, ..., m\}$ is an orthonormal basis of $\mathbb{R}^n \otimes \mathbb{R}^m$. □

To make notation simpler, let us introduce the following: For

$$\mathbf{y}_t = \sum_{1 \leq i_1, \ldots, i_k \leq n} y_t^{(i_1, \ldots, i_k)} \;\; e^{(i_1)} \otimes \ldots \otimes e^{(i_k)},$$

a $(\mathbb{R}^m)^{\otimes k}$ valued function, and the $\mathbb{R}^n$-valued path $x : [0, T] \to \mathbb{R}^n$, we define the tensor integral to be the shorthand notation

$$\int_s^t \mathbf{y}_\tau \otimes dx_\tau := \sum_{1 \leq i_1, \ldots, i_k, j \leq n} \int_s^t y_\tau^{(i_1, \ldots, i_k)} dx_\tau^{(j)} \;\; e^{(i_1)} \otimes \ldots \otimes e^{(i_k)} \otimes e^{(j)} \in (\mathbb{R}^m)^{\otimes k} \otimes \mathbb{R}^n.$$

We can now adapt the result of Remark 2.14 to this notation:

$$\int_{s\le t_1\le\ldots\le t_k\le t} dx_{t_1}\otimes\ldots\otimes dx_{t_k} := \sum_{1\le i_1,\ldots,i_k\le n} z_{s,t}^{(i_1,\ldots,i_k)}\ \ e^{(i_1)}\otimes\ldots\otimes e^{(i_k)}\in(\mathbb{R}^n)^{\otimes k},$$

where the $z_{s,t}^{(i_1,\ldots,i_k)}$ are defined by Remark 2.14.
To generalize our results from Section 2.2, we need to introduce multiplicative functionals, based on the notation of [Lej03, Section 3.3]:

Definition 2.18:
Let $k\in\mathbb{N}$ and $\mathbf{x}:\{0\le s\le t\le T\}\to T_k(\mathbb{R}^n)$. We call $\mathbf{x}$ a *multiplicative functional* if

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

for all $0\le s\le u\le t\le T$. $\mathcal{S}^k(\mathbb{R}^n)$ is the set of all multiplicative functionals mapping to $T_k(\mathbb{R}^n)$. This is *not* a linear space.

Remark 2.19:
We want the k-th tensor level of a multiplicative functional to represent the k-th iterated integrals of a given path $x:[0,T]\to\mathbb{R}^n$, i.e.

$$\mathbf{x}_{s,t}^k\text{“}=\text{”}\int_{s\le t_1\le\ldots\le t_k\le t} d\mathbf{x}_{t_1}\otimes\ldots\otimes d\mathbf{x}_{t_k}=\int_s^t \mathbf{x}_{s,\tau}^{k-1}\otimes d\mathbf{x}_\tau.$$

This makes the rule from Definition 2.18 natural, since

$$\mathbf{x}_{s,u}^1+\mathbf{x}_{u,t}^1=\int_s^u dx_\tau+\int_u^t dx_\tau=\int_s^t dx_\tau=\mathbf{x}_{s,t}^1,$$

as well as

$$\begin{aligned}\mathbf{x}_{s,u}^2+\mathbf{x}_{u,t}^2\text{“}=\text{”}&\int_s^u (x_\tau-x_s)\otimes dx_\tau+\int_u^t (x_\tau-x_u)\otimes dx_\tau\\ =&\int_s^t x_\tau\otimes dx_\tau-x_s\otimes(x_u-x_s)-x_u\otimes(x_t-x_u)\\ =&\int_s^t x_\tau\otimes dx_\tau-x_s\otimes(x_t-x_s)+x_s\otimes(x_t-x_u)-x_u\otimes(x_t-x_u)\\ =&\int_s^t (x_\tau-x_s)\otimes dx_\tau-(x_u-x_s)\otimes(x_t-x_u)\text{“}=\text{”}\mathbf{x}_{s,t}^2-x_{s,u}\otimes x_{u,t}.\end{aligned}\tag{6}$$

for $\mathbf{x}_{s,t}=1+\mathbf{x}_{s,t}^1+\mathbf{x}_{s,t}^2\in\mathcal{S}^2(\mathbb{R}^n)$ with $\mathbf{x}_{s,t}^1:=x_t-x_s$ and $0\le s\le u\le t\le 1$. Then

$$\begin{aligned}\mathbf{x}_{s,u}\otimes\mathbf{x}_{u,t}=&\left(1+\mathbf{x}_{s,u}^1+\mathbf{x}_{s,u}^2\right)\otimes\left(1+\mathbf{x}_{u,t}^1+\mathbf{x}_{u,t}^2\right)\\ =&1+\underbrace{\mathbf{x}_{s,u}^1+\mathbf{x}_{u,t}^1}_{=\mathbf{x}_{s,t}^1}+\underbrace{\mathbf{x}_{s,u}^2+\mathbf{x}_{u,t}^2+\mathbf{x}_{s,u}^1\otimes\mathbf{x}_{u,t}^1}_{=\mathbf{x}_{s,t}^2}=\mathbf{x}_{s,t}.\end{aligned}$$

We can now extend the semi-norm of p-variation to $\mathcal{S}^k(\mathbb{R}^n)$:

Definition 2.20:
Let $p\ge 1$. Let $\mathbf{x}=1+\mathbf{x}^1+\ldots+\mathbf{x}^k\in\mathcal{S}^k(\mathbb{R}^n)$ with $\mathbf{x}^j:\{0\le s\le t\le T\}\to(\mathbb{R}^n)^{\otimes j}$. We define the p-variation of $\mathbf{x}$ to be

$$\mathrm{Var}_{p,[s,t]}(\mathbf{x}):=\sum_{j=1}^k\left(\mathrm{Var}_{\frac{p}{j},[s,t]}(\mathbf{x}^j)\right)^{\frac{1}{j}}.$$

Here

$$\mathrm{Var}_{q,[s,t]}(\mathbf{x}^j) := \sup_{\{s\le t_0<\ldots<t_k\le t\}} \left(\sum_{i=1}^{k} \left| \mathbf{x}^j_{t_{i-1},t_i} \right|^q \right)^{\frac{1}{q}}$$

for $q > 0$. In particular, we then also have

$$\left| \mathbf{x}^j_{s,t} \right| \le \mathrm{Var}^j_{p,[s,t]}(\mathbf{x})$$

for all $j = 1, ..., k$. We also define the p-variation distance of $\mathbf{x}, \mathbf{y} \in \mathcal{S}^k(\mathbb{R}^n)$ to be

$$d_{p-var;[s,t]}(\mathbf{x}, \mathbf{y}) := \mathrm{Var}_{p,[s,t]}(\mathbf{x} - \mathbf{y}).$$

This is only one of the possible definitions for combining the p-variations of the multiple tensor levels, which is based on [Lej03, Section 6.1].
For a path $x : [0, T] \to \mathbb{R}^n$ and $[(s, t) \mapsto \mathbf{x}_{s,t} := 1 + x_t - x_s] \in \mathcal{S}^1(\mathbb{R}^n)$, we then have

$$\mathrm{Var}_{p,[s,t]}(x) = \mathrm{Var}_{p,[s,t]}(\mathbf{x}).$$

Now using these definitions, we can finally define rough paths.

Definition 2.21 (Rough Path)**:**
Let $p \ge 1$. A *p-rough path* is a continuous multiplicative functional $\mathbf{x} \in \mathcal{S}^{\lfloor p \rfloor}(\mathbb{R}^n)$, such that

$$\mathrm{Var}_{p,[0,T]}(\mathbf{x}) < \infty.$$

We will also identify the rough path $\mathbf{x}$ with the function $[0, T] \ni t \mapsto \mathbf{x}_{0,t} \in T_{\lfloor p \rfloor}(\mathbb{R}^n)$, as we can simply get back the other values using the identity

$$\mathbf{x}_{s,t} = \mathbf{x}^{\otimes -1}_{0,s} \otimes \mathbf{x}_{0,t}.$$

This makes $\mathbf{x}$ a path and not some two-variable function. Let $\Omega^p = \Omega^p(\mathbb{R}^n)$ be the space of $\mathbb{R}^n$-valued p-rough paths, equipped with the topology induced by the p-variation semi-norm.
An important subclass of rough paths are geometric rough paths, which we will need later on:

Definition 2.22 (Geometric Rough Path)**:**
Let $p > 1$. A p-rough path $\mathbf{x} \in \Omega^p(\mathbb{R}^n)$ is called a *geometric p-rough path* or just *geometric rough path*, if there is a sequence $(\mathbf{x}_n) \subset \Omega^p(\mathbb{R}^n)$ of piecewise smooth paths with iterated integrals defined by the usual path integral, i.e.

$$(\mathbf{x}^k_n)_{s,t} = \int_s^t (\mathbf{x}^{k-1}_n)_{s,\tau} \otimes d(\mathbf{x}_n)_\tau = \int_s^t (\mathbf{x}^{k-1}_n)_{s,\tau} \otimes \frac{\partial}{\partial t}(\mathbf{x}^1_n)_\tau d\tau$$

for $k = 2, ..., \lfloor p \rfloor$, such that

$$d_{p-var}(\mathbf{x}, \mathbf{x}_n) \xrightarrow{n\to\infty} 0.$$

$G\Omega^p = G\Omega^p(\mathbb{R}^n)$ is the set of geometric p-rough paths.

Remark 2.23:
A stochastic process $\mathbf{x}_t$ with finite p-variation for $3 > p$ and the first iterated integral defined by the Stratonovich integral is a geometric rough path. In particular

$$\mathbf{B}^{\mathrm{Strat}}_t := 1 + B_t + \underbrace{\int_0^t B_\tau \otimes \circ dB_t}_{=\frac{B_t^2}{2} \text{ for } n=1}$$

is a geometric p-rough path for all $p > 2$ [FH20, Section 3.3].

Example 2.24 (Geometric Rough Path Lifts)**:**
A rough path lift of a process $x : [0, T] \to \mathbb{R}^n$ is a p-rough path $\mathbf{x} = 1 + \mathbf{x}^1 + \mathbf{x}^2 + ... + \mathbf{x}^{\lfloor p \rfloor}$, such that $\mathbf{x}^1_{s,t} = x_t - x_s$ for all $s < t$, i.e. it is a rough path $\mathbf{x}$ with $\mathbb{R}^n$-component corresponding to the original process x. Some examples of geometric rough path lifts include the following:

- **Semi-martingales:** Semi-martingales can be lifted to geometric p-rough paths for $p \in (2, 3)$ [Lyo98; FV10; CL05; KLA20].
- **Lévy-processes:** Lévy-processes can be lifted to geometric p-rough paths [FS17; Che17; KLA20]. It can of course happen that we need $p \geq 3$, which we do not cover explicitly, but for which these results also hold with analogous proofs.
- **Fractional Brownian motion:** Fractional Brownian motion with Hurst parameter $H > \frac{1}{4}$ can be lifted to a geometric p-rough path for $p > \frac{1}{H}$ [CQ02; KLA20].

2.4 The Case $2 \leq p < 3$

For $p \geq 2$, the above construction of the integral (Theorem 2.10) still works, if we use α-Hölder functions, with α large enough, but this rules out more and more functions, with increasing roughness. In particular, we can't define the iterated integrals anymore, but these will be essential for defining the signature of a process, later on.
Therefore, we need to refine our process. The problem we have to face is, that the $y_{s,t}$ do not converge fast enough, and so the upper bound in Equation (5) will be too large. Before, we approximated the integral by

$$\int_s^t f(x_\tau) dx_\tau \approx \int_s^t f(x_s) dx_\tau = f(x_s)(x_t - x_s).$$

Now, for faster convergence, we can approximate f by a two-step Taylor series, instead of only using the first term. For $f \in C^1(\mathbb{R}^n, \mathbb{R}^{m \times n})$ we let $f_1, ..., f_n \in C(\mathbb{R}^n, \mathbb{R}^m)$ and

$$\begin{aligned}\int_s^t f_i(x_\tau) dx_\tau^{(i)} \approx & \int_s^t f_i(x_s) + \nabla f_i(x_s)^T (x_\tau - x_s) dx_\tau^{(i)} \\ =& f_i(x_s)(x_t^{(i)} - x_s^{(i)}) + \sum_{j=1}^n \frac{\partial f_i}{\partial x^{(j)}}(x_s) \int_s^t (x_\tau^{(j)} - x_s^{(j)}) dx_\tau^{(i)}.\end{aligned}$$

Let

$$\mathbf{x}^1_{s,t} := x_t - x_s = \int_s^t dx_\tau.$$

Now suppose, we know the second order iterated integrals of x

$$\mathbf{x}^2_{s,t} = \int_{s \leq t_1 \leq t_2 \leq t} dx_{t_1} \otimes dx_{t_2} \in \mathbb{R}^n \otimes \mathbb{R}^n \tag{7}$$

with

$$\left|\mathbf{x}^2_{s,t}\right|^{\frac{p}{2}} \leq \mathrm{Var}^p_{p,[s,t]}(x). \tag{8}$$

Now, $\mathbf{x}_{s,t} := 1 + \mathbf{x}^1_{s,t} + \mathbf{x}^2_{s,t}$ will be the rough path, that we can use to construct the integral. One can choose any $\mathbf{x}^2_{s,t}$, as long as it fulfills the conditions of Equation (8) and Definition 2.18. One possible choice always is

$$\mathbf{x}^k_{s,t} := \frac{(x_t - x_s)^{\otimes k}}{k!}.$$

Let

$$\mathbf{y}_{s,t} := f(x_s)\mathbf{x}^1_{s,t} + Df(x_s)\mathbf{x}^2_{s,t},$$

where $Df(x) \in \mathcal{L}(\mathbb{R}^n \otimes \mathbb{R}^n, \mathbb{R}^m)$ is the biliniear map defined by

$$Df(x)e^{(i)} \otimes e^{(j)} = \frac{\partial f_j}{\partial x^{(i)}}(x) \in \mathbb{R}^m.$$

For a partition $\Pi = \{s \leq t_0 < ..., t_k \leq t\}$ of $[s,t]$ again set

$$z^{\Pi}_{s,t} := \sum_{j=1}^{k} \mathbf{y}_{t_{j-1},t_j}.$$

Theorem 2.25:
Let $2 \leq p < 3$. Let $\mathbf{x}$ be defined as above, i.e. $\mathbf{x}$ is a p-rough path, and f be a continuous, bounded function, with α-Hölder continuous, bounded derivatives for some $\alpha > p-2$. For any sequence $(\Pi_j)_j$ of partitions of $[s,t]$ with vanishing meshes, $z^{\Pi_j}_{s,t}$ admits a unique limit $z_{s,t}$ for every $0 \leq s \leq t \leq 1$. The map $(s,t) \mapsto z_{s,t}$ is continuous and we have the relation

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

for all $0 \leq s \leq u \leq t \leq T$. Furthermore, $z_{s,t}$ has finite p-variation with

$$\operatorname{Var}_{p,[s,t]}(z) \leq K \operatorname{Var}_{p,[s,t]}(x).$$

This theorem for defining the integral over one-forms is from [Lej03, Section 3.2].

Proof. We need to show that for $0 \leq s \leq u \leq t \leq 1$, there is some constant C such that

$$|\mathbf{y}_{s,t} - \mathbf{y}_{s,u} - \mathbf{y}_{u,t}| \leq C \operatorname{Var}^{2+\alpha}_{p,[s,t]}(x),$$

to apply Lemma 2.11. Let

$$N(f) := \inf \left\{ M > 0 \,\middle|\, \|f\|_{\infty, \operatorname{Im} x} \leq M \text{ and } \|Df\|_{\infty, \operatorname{Im} x} \leq M \text{ and } \sup_{x \neq y} \frac{|Df(x) - Df(y)|}{|x-y|^\alpha} \leq M \right\}.$$

By the requirements we placed on f, $N(f)$ is necessarily finite. Then, by the Taylor-Formula, we have

$$f(b) = f(a) + \sum_{j=1}^{n} \frac{\partial f}{\partial x^{(j)}}(a)(b^j - a^j) + R(a,b)$$

for $a, b \in \mathbb{R}^n$. Here

$$|R(a,b)| = \left| \sum_{j=1}^{n} \int_0^1 \underbrace{\left(\frac{\partial f}{\partial x^{(j)}}(a) - \frac{\partial f}{\partial x^{(j)}}(a + (b-a)r) \right)}_{\|\cdot\| \leq N(f)|b-a|^\alpha r^\alpha} (b^j - a^j) dr \right| \leq nN(f)\,|b-a|^{1+\alpha}. \tag{9}$$

Then we can use $\mathcal{L}(\mathbb{R}^n \otimes \mathbb{R}^n, \mathbb{R}^m) \widehat{=} \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$ via $[x \otimes y \mapsto f(x \otimes y)] \widehat{=} [x \mapsto fx = [y \mapsto f(x \otimes y)]]$ to see

$$\begin{aligned}
|\mathbf{y}_{s,t} - \mathbf{y}_{s,u} - \mathbf{y}_{u,t}| =& \big| f(x_s)\mathbf{x}^1_{s,t} + Df(x_s)\mathbf{x}^2_{s,t} - f(x_s)\mathbf{x}^1_{s,u} \\
& \quad - Df(x_s)\mathbf{x}^2_{s,u} - f(x_u)\mathbf{x}^1_{u,t} - Df(x_u)\mathbf{x}^2_{u,t} \big| \\
=& \big| (f(x_s) - f(x_u))\mathbf{x}^1_{u,t} + (Df(x_s) - Df(x_u))\mathbf{x}^2_{u,t} + Df(x_s)\mathbf{x}^1_{s,u} \otimes \mathbf{x}^1_{u,t} \big| \\
\leq& \big| (Df(x_s) - Df(x_u))\mathbf{x}^2_{u,t} \big| + \big| \big(f(x_s) - f(x_u) + Df(x_s)\mathbf{x}^1_{s,u} \big) \mathbf{x}^1_{u,t} \big| \\
\leq& \big| (Df(x_s) - Df(x_u))\mathbf{x}^2_{u,t} \big| + \big| R(x_s, x_u)\mathbf{x}^1_{u,t} \big| \\
\leq& N(f) \left| \mathbf{x}^1_{s,u} \right|^\alpha \left| \mathbf{x}^2_{u,t} \right| + nN(f) \left| \mathbf{x}^1_{s,u} \right|^{1+\alpha} \left| \mathbf{x}^1_{u,t} \right| \\
\leq& (1+n)N(f) \operatorname{Var}^{2+\alpha}_{p,[s,t]}(x).
\end{aligned}$$

Now the sewing lemma (Lemma 2.11) gives the assertion. $\square$

Definition 2.26:
We can now define the integral of the one form $f(x_\tau)d\mathbf{x}_\tau$ for f bounded with bounded, α-Hölder derivatives for $\alpha > p-2$:

$$\int_s^t f(x_\tau)d\mathbf{x}_\tau := \mathbf{z}_{s,t},$$

where $\mathbf{z}_{s,t}$ is constructed in Theorem 2.25. Note that we are no longer integrating with respect to the path x, but with respect to the rough path $\mathbf{x}$.

The essential property, the proof of Theorem 2.25 relies on Equation (9):

$$\left|f(x_t) - f(x_s) - Df(x_s)\mathbf{x}^1_{s,t}\right| \le C \operatorname{Var}^{1+\alpha}_{p,[s,t]}(x).$$

This gives rise to the notion of controlled rough paths. Our notion of controlled paths, as well as Theorem 2.29, are loosely based on [FH20, Section 4.3].

Definition 2.27 (Controlled Rough Paths)**:**
Let $2 \le p < 3$, $\omega : \Delta^+_T \to \mathbb{R}^+$ be a controlling function. Let W be a Banach space and let $\mathbf{x}$ be a $\mathbb{R}^n$-valued p-rough path dominated by ω, i.e.

$$\operatorname{Var}^p_{p,[s,t]}(\mathbf{x}) \le \omega(s,t).$$

We call the bounded map $Y : [0,T] \to W$ *controlled by* $\mathbf{x}$ *(via* ω*)*, if there exists a bounded map $Y' : [0,T] \to \mathcal{L}(\mathbb{R}^n, W)$, $\gamma > p-2$ and $C > 0$ with

$$\begin{aligned}\left\|Y_t - Y_s - Y'_s\mathbf{x}^1_{s,t}\right\|_W \le& C\omega^{\frac{1+\gamma}{p}}(s,t),\\ \left\|Y'_t - Y'_s\right\|_{\mathcal{L}(\mathbb{R}^n,W)} \le& C\omega^{\frac{\gamma}{p}}(s,t).\end{aligned}$$

Y' is called a *Gubinelli derivative* of Y and is not unique. Let $\mathcal{D}^\gamma_{\omega,\mathbf{x}} = \mathcal{D}^\gamma_{\omega,\mathbf{x}}(W)$ be the space of the *controlled rough paths* (Y, Y') endowed with the semi norm

$$\begin{aligned}\left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}} :=& \inf\left\{C>0\middle| \left\|Y_t - Y_s - Y'_s\mathbf{x}^1_{s,t}\right\|_W \le C\omega^{\frac{1+\gamma}{p}}(s,t)\right\}\\ &+ \inf\left\{C>0\middle| \left\|Y'_t - Y'_s\right\|_{\mathcal{L}(\mathbb{R}^n,W)} \le C\omega^{\frac{\gamma}{p}}(s,t)\right\}.\end{aligned}$$

We will use the short-hand $\mathcal{D}^\gamma_{\mathbf{x}} := \mathcal{D}^\gamma_{(s,t)\mapsto \operatorname{Var}^p_{p,[s,t]}(\mathbf{x}),\mathbf{x}}$ and sometimes also $R^Y_{s,t} = Y_t - Y_s - Y'_s\mathbf{x}^1_{s,t}$.

Using the definition, we see that

$$\begin{aligned}\|Y_t - Y_s\|_W &\le \left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}\omega^{\frac{1+\gamma}{p}}(s,t) + \left\|Y'\right\|_\infty \omega^{\frac{1}{p}}(s,t)\\ &\le \left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}\omega^{\frac{1+\gamma}{p}}(s,t) + \left(\left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}\omega^{\frac{\gamma}{p}}(0,T) + \|Y_0\|\right)\omega^{\frac{1}{p}}(s,t)\\ &\le \left(2\left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}\omega^{\frac{\gamma}{p}}(0,T) + \|Y_0\|\right)\omega^{\frac{1}{p}}(s,t).\end{aligned} \tag{10}$$

Even though Ω^p is not even a vector space, the space $\mathcal{D}^\gamma_{\omega,\mathbf{x}}(W)$ is a Banach space (with respect to the norm $\|Y_0\| + \|Y'_0\| + \|(Y,Y')\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}$), that depends on $\mathbf{x} \in \Omega^p$. We have a Leibniz rule for the Gubinelli derivative, based on [FH20, Lemma 7.5], in the following sense:

Lemma 2.28 (Leibniz Rule)**:**
Let $1 \ge \gamma > p-2$ *and let* $(Y,Y') \in \mathcal{D}^\gamma_{\omega,\mathbf{x}}(\mathcal{L}(W,W))$, $(Z,Z') \in \mathcal{D}^\gamma_{\omega,\mathbf{x}}(W)$ *be two controlled rough paths. Then* (U,U') *with*

$$\begin{aligned}U =& YZ \in W\\ U' =& YZ' + Y'Z \in \mathcal{L}(\mathbb{R}^n, W)\end{aligned}$$

is a controlled rough path in $\mathcal{D}^\gamma_{\omega,\mathbf{x}}(W)$. *Where* $Y'Z$ *is the map*

$$Y'Zx := Y'(x)(Z) \in W$$

for $x \in \mathbb{R}^n$, *since* $Y' \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(W,W))$. *We also have*

$$\left\|(U,U')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}} \le C\left(\|Y_0\| + \left\|Y_0'\right\| + \left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}\right)\left(\|Z_0\| + \left\|Z_0'\right\| + \left\|(Z,Z')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}\right).$$

Proof. It holds

$$\begin{aligned}
\left\|U_t - U_s - U_s'\mathbf{x}^1_{s,t}\right\|_W &= \left\|Y_tZ_t - Y_sZ_s - Y_sZ_s'\mathbf{x}^1_{s,t} - Y_s'Z_s\mathbf{x}^1_{s,t}\right\| \\
&\le \left\|Y_sZ_t - Y_sZ_s - Y_sZ_s'\mathbf{x}^1_{s,t}\right\| + \left\|Y_tZ_t - Y_sZ_t - Y_s'Z_t\mathbf{x}^1_{s,t}\right\| \\
&\quad + \left\|Y_s'\left(Z_t - Z_s\right)\mathbf{x}^1_{s,t}\right\| \\
&\le \|Y_s\|\left\|(Z,Z')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}\omega^{\frac{1+\gamma}{p}}(s,t) + \|Z_t\|\left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}\omega^{\frac{1+\gamma}{p}}(s,t) \\
&\quad + \left\|Y_s'\right\|\|Z_t - Z_s\|\,\omega^{\frac{1}{p}}(s,t).
\end{aligned}$$

Using Equation (10), we get

$$\|Y_s\| \le \|Y_0\| + \|Y_s - Y_0\| \le \|Y_0\| + \left(2\left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}\omega^{\frac{\gamma}{p}}(0,T) + \|Y_0\|\right)\omega^{\frac{1}{p}}(0,T)$$

and its analog for Z_s. We can estimate the last summand by

$$\left\|Y_s'\right\| \le \left\|Y_0'\right\| + \left\|Y_s' - Y_0'\right\| \le \left\|Y_0'\right\| + \left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}\omega^{\frac{\gamma}{p}}(0,T)$$

and

$$\|Z_t - Z_s\| \le \left(2\left\|(Z,Z')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}\omega^{\frac{\gamma}{p}}(s,t) + \|Z_0\|\right)\omega^{\frac{1}{p}}(s,t).$$

For the other term, we see that

$$\begin{aligned}
\left\|Y_tZ_t' + Y_t'Z_t - Y_sZ_s' - Y_s'Z_s\right\| &\le \left\|Y_tZ_t' - Y_tZ_s'\right\| + \left\|Y_tZ_s' - Y_sZ_s'\right\| \\
&\quad + \left\|Y_t'Z_t - Y_t'Z_s\right\| + \left\|Y_t'Z_s - Y_s'Z_s\right\| \\
&\le \|Y_t\|\left\|Z_t' - Z_s'\right\| + \left\|Z_s'\right\|\|Y_t - Y_s\| \\
&\quad + \left\|Y_t'\right\|\|Z_t - Z_s\| + \|Z_s\|\left\|Y_t' - Y_s'\right\|
\end{aligned}$$

All in all, we can find $C > 0$ only dependent on $\mathbf{x}$ and γ, such that

$$\left\|(U,U')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}} \le C\left(\|Y_0\| + \left\|Y_0'\right\| + \left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}\right)\left(\|Z_0\| + \left\|Z_0'\right\| + \left\|(Z,Z')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}\right).$$

□

We can now define the integral of a controlled rough path, with respect to the rough path $\mathbf{x}$. This result goes back to [Gub04].

Theorem 2.29:
Let $3 > p \ge 2$ *and* $1 + \mathbf{x}^1_{s,t} + \mathbf{x}^2_{s,t} \in \Omega^p$ *be a rough path with values in* $\mathbb{R}^m$. *Let* $\gamma > p - 2$ *and let* $(Y,Y') \in \mathcal{D}^\gamma_{\omega,\mathbf{x}}(\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m))$ *be a controlled rough path. Then the limit*

$$\mathbf{z}_{s,t} := \lim_{\|\{s \le t_0 < \ldots t_k \le t\}\| \to 0} \sum_{j=1}^{k}\left(Y_{t_{j-1}}\mathbf{x}^1_{t_{j-1},t_j} + Y_{t_{j-1}}'\mathbf{x}^2_{t_{j-1},t_j}\right)$$

exists, is unique, and we have

$$\left|z_{s,t} - Y_s\mathbf{x}^1_{s,t} + Y_s'\mathbf{x}^2_{s,t}\right| \le C\left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}\omega^{\frac{2+\gamma}{p}}(s,t)$$

with $C = 2^{\frac{2+\gamma}{p}+1}\zeta\left(\frac{2+\gamma}{p}\right)$, *as well as* $(\mathbf{z}_{0,\cdot}, Y) \in \mathcal{D}^{\gamma\wedge 1}_{\omega,\mathbf{x}}(\mathbb{R}^m)$.

Proof. To show the existence and uniqueness of $\mathbf{z}$, we again use the sewing lemma (Lemma 2.11). Let

$$\mathbf{y}_{s,t} := Y_s \mathbf{x}^1_{s,t} + Y'_s \mathbf{x}^2_{s,t}.$$

Analogous to the proof of Theorem 2.25, we can then show that

$$\begin{aligned}|\mathbf{y}_{s,t} - \mathbf{y}_{s,u} - \mathbf{y}_{u,t}| &= \left|(Y_s - Y_u)\mathbf{x}^1_{u,t} + (Y'_s - Y'_u)\mathbf{x}^2_{u,t} + Y'_s \mathbf{x}^1_{s,u} \otimes \mathbf{x}^1_{u,t}\right| \\ &\leq \left\|(Y, Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}} \left(\omega^{\frac{2+\gamma}{p}}(s,t) + \omega^{\frac{2+\gamma}{p}}(s,t)\right) \leq 2\left\|(Y, Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}} \omega^{\frac{2+\gamma}{p}}(s,t)\end{aligned}$$

and

$$|\mathbf{y}_{s,t}| \leq \|Y\|_\infty \omega^{\frac{1}{p}}(s,t) + \left\|Y'\right\|_\infty \omega^{\frac{2}{p}}(s,t) \leq \left(\|Y\|_\infty + \left\|Y'\right\|_\infty \omega^{\frac{1}{p}}(0,T)\right)\omega^{\frac{1}{p}}(s,t).$$

Now, the sewing lemma (Lemma 2.11) gives the existence and uniqueness of $\mathbf{z}_{s,t}$. This also gives

$$\left|\mathbf{z}_{s,t} - Y_s \mathbf{x}^1_{s,t} - Y'_s \mathbf{x}^2_{s,t}\right| \leq 2^{\frac{2+\gamma}{p}+1} \zeta\left(\frac{2+\gamma}{p}\right) \left\|(Y, Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}} \omega^{\frac{2+\gamma}{p}}(s,t).$$

Therefore

$$\begin{aligned}\left|\mathbf{z}_{s,t} - Y_s \mathbf{x}^1_{s,t}\right| &\leq \left|\mathbf{z}_{s,t} - Y_s \mathbf{x}^1_{s,t} - Y'_s \mathbf{x}^2_{s,t}\right| + \left|Y'_s \mathbf{x}^2_{s,t}\right| \\ &\leq 2^{\frac{2+\gamma}{p}+1} \zeta\left(\frac{2+\gamma}{p}\right) \left\|(Y, Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}} \omega^{\frac{2+\gamma}{p}}(s,t) + \left\|Y'\right\|_\infty \omega^{\frac{2}{p}}(s,t) \\ &\leq \left(2^{\frac{2+\gamma}{p}+1} \zeta\left(\frac{2+\gamma}{p}\right) \left\|(Y, Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}} \omega^{\frac{\gamma}{p}}(0,T) + \left\|Y'\right\|_\infty\right) \omega^{\frac{2}{p}}(s,t).\end{aligned} \tag{11}$$

Using Equation (10), we get $(z_{0,\cdot}, Y) \in \mathcal{D}^{\gamma\wedge 1}_{\mathbf{x}}(\mathbb{R}^m)$ ☐

Now, we can define the integral of a controlled rough path:

Definition 2.30:
Let $3 > p \geq 2$ and $1 + \mathbf{x}^1_{s,t} + \mathbf{x}^2_{s,t} \in \Omega^p$ be a rough path with values in $\mathbb{R}^m$. Let $\gamma > p - 2$ and let $(Y, Y') \in \mathcal{D}^\gamma_{\omega,\mathbf{x}}(\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$ be a controlled rough path. We define

$$\int_s^t Y_\tau d\mathbf{x}_\tau := \mathbf{z}_{s,t},$$

where $\mathbf{z}$ is obtained from Theorem 2.29. Note that we suppress the dependence on Y' in our notation.

Remark 2.31:
We can get back the Itô-integral from Definition 2.30 in the special case, where we consider the rough path

$$\mathbf{B}^{\text{Itô}}_t := 1 + B_t + \int_0^t B_\tau \otimes dB_\tau,$$

where B_t is a standard Brownian motion in $\mathbb{R}^n$ and the integral on the right hand side is the Itô-integral with symmetric part

$$Sym\left(\int_0^t B_\tau \otimes dB_\tau\right) = \frac{1}{2} B_t \otimes B_t - \frac{t}{2}\mathbb{1},$$

where $\mathbb{1}$ is the identity matrix. The resulting rough integral with respect to $\mathbf{B}^{\text{Itô}}_t$ then coincides with the Itô-integral [FH20, Section 5.1]. Similarly, we can construct the Stratonovich integral of Brownian motion, by integrating with respect to the rough path

$$\mathbf{B}^{\text{Strat}}_t := 1 + B_t + \int_0^t B_\tau \otimes \circ dB_\tau,$$

where the integral on the right hand side is the Stratonovich integral with symmetric part

$$Sym\left(\int_0^t B_\tau \otimes \circ dB_\tau\right) = \frac{1}{2} B_t \otimes B_t$$

[FH20, Section 5.2].

We can finally construct the iterated integrals of a rough path for $3 > p \geq 2$ analogous to Remark 2.14:

Remark 2.32:
Consider again $f_j : [0,T] \to \mathbb{R}^{1\times n}$

$$f_j(t) := (0, ..., 0, 1, 0, ..., 0)$$

with 1 in the j-th component. Then obviously $(f_j, 0) \in \mathcal{D}^1_{\mathbf{x}}(\mathcal{L}(\mathbb{R}^n, \mathbb{R}))$. Now we can use Theorem 2.29, to construct

$$\int_s^t d\mathbf{x}^{(j)}_\tau := \mathbf{z}^{(j)}_{s,t} := \int_s^t f_j d\mathbf{x}_\tau.$$

Then $(\mathbf{z}^{(j)}_{s,t}, f_j) \in \mathcal{D}^1_{\mathbf{x}}(\mathbb{R})$. Now we can define the iterated integrals inductively: Let $J = (j_1, ..., j_m) \in \{1, ..., n\}^m$ and let

$$\mathbf{z}^J_{s,t} = \int_{s\leq t_1 \leq ... \leq t_n \leq t} d\mathbf{x}^{j_1}_{t_1} ... d\mathbf{x}^{j_m}_{t_m},$$

with $(\mathbf{z}^J_{s,\cdot}, (\mathbf{z}^J_{s,\cdot})') \in \mathcal{D}^1_{\mathbf{x}}(\mathbb{R})$. Then for $j_{m+1} \in \{1, ..., n\}$

$$\left(f_{j_{m+1}} \mathbf{z}^J_{s,\cdot}, e^i \mapsto \delta_{i,j_{m+1}} \left(\mathbf{z}^J_{s,\cdot}\right)'\right) \in \mathcal{D}^1_{\mathbf{x}}(\mathcal{L}(\mathbb{R}^n, \mathbb{R})).$$

We then define

$$\int_{s\leq t_1 \leq ... \leq t_{m+1} \leq t} d\mathbf{x}^{j_1}_{t_1} ... d\mathbf{x}^{j_{m+1}}_{t_{m+1}} := \mathbf{z}^{(j_1,...,j_{m+1})}_{s,t} := \int_s^t f_{j_{m+1}} \mathbf{z}^J_{s,t_{m+1}} d\mathbf{x}_{t_{m+1}}$$

and get $\left(\mathbf{z}^{(j_1,...,j_{m+1})}_{s,\cdot}, f_{j_{m+1}} \mathbf{z}^J_{s,\cdot}\right) \in \mathcal{D}^1_{\mathbf{x}}(\mathbb{R})$. Using the notation of the tensor integral, we can write

$$\int_{s\leq t_1 \leq ... \leq t_m \leq t} d\mathbf{x}_{t_1} \otimes ... \otimes d\mathbf{x}_{t_m} = \sum_{j_1,...,j_m=1}^{n} \mathbf{z}^{(j_1,...,j_m)}_{s,t} e^{(j_1)} \otimes ... \otimes e^{(j_m)}.$$

2.5 Stability of rough integration

Since we want to approximate an infimum, we will use sequences of solutions to RDEs. To ensure the convergence of these sequences, we need to study the stability of rough integration. This section is a generalization of [FH20, Theorem 4.17].
Let $3 > p \geq 2$, $\gamma > p-2$, and let W be a Banach space. Consider two p-rough paths $\mathbf{x} \in \Omega^p(\mathbb{R}^n)$ and $\tilde{\mathbf{x}} \in \Omega^p(\mathbb{R}^n)$. Let ω_1 be a controlling function for $\mathbf{x}$ and $(Y, Y') \in \mathcal{D}^\gamma_{\omega_1,\mathbf{x}}$ be an $\mathbf{x}$-controlled rough path. Similarly, let ω_2 be a controlling function for $\tilde{\mathbf{x}}$ and $(\tilde{Y}, \tilde{Y}') \in \mathcal{D}^\gamma_{\omega_2,\tilde{\mathbf{x}}}$ be an $\tilde{\mathbf{x}}$-controlled rough path. In particular, we have $(Y, Y') \in \mathcal{D}^\gamma_{\omega,\mathbf{x}}$ and $(\tilde{Y}, \tilde{Y}') \in \mathcal{D}^\gamma_{\omega,\tilde{\mathbf{x}}}$ for $\omega = \omega_1 + \omega_2$. As always, we consider the fixed time horizon $[0, T]$. For the stability of rough integration, we need a notion of distance between (Y, Y') and $(\tilde{Y}, \tilde{Y}')$, even though in general, they may lie in different Banach spaces. Recall the definition of the semi-norm on $\mathcal{D}^\gamma_{\omega,\mathbf{x}}$. This leads us to the following definition

$$\left\|Y, Y'; \tilde{Y}, \tilde{Y}'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma} := \inf\left\{C > 0 \middle| \left\|Y_t - Y_s - Y'_s \mathbf{x}^1_{s,t} - \tilde{Y}_t + \tilde{Y}_s + \tilde{Y}'_s \tilde{\mathbf{x}}^1_{s,t}\right\|_W \leq C\omega^{\frac{1+\gamma}{p}}(s,t)\right\}$$
$$+ \inf\left\{C > 0 \middle| \left\|Y'_t - Y'_s - \tilde{Y}'_t + \tilde{Y}'_s\right\|_{\mathcal{L}(\mathbb{R}^n,W)} \leq C\omega^{\frac{\gamma}{p}}(s,t)\right\}.$$

We have

$$\left\|Y,Y';\tilde Y,\tilde Y'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma} \le \left\|(Y,Y')\right\|_{\mathcal{D}^{\gamma}_{\omega,\mathbf{x}}} + \left\|(\tilde Y,\tilde Y')\right\|_{\mathcal{D}^{\gamma}_{\omega,\tilde{\mathbf{x}}}}.$$

Let

$$d^p_\omega(\mathbf{x},\tilde{\mathbf{x}}) := \sup_{s<t} \frac{d^p_{p-var;[s,t]}(\mathbf{x},\tilde{\mathbf{x}})}{\omega(s,t)} \le 2.$$

Then it also holds

$$\begin{aligned}\left\|Y_{s,t}-\tilde Y_{s,t}\right\|_W &\le \left\|Y_{s,t}-\tilde Y_{s,t}-Y'_s\mathbf{x}^1_{s,t}+\tilde Y'_s\tilde{\mathbf{x}}^1_{s,t}\right\|_W + \left\|Y'_s\mathbf{x}^1_{s,t}-\tilde Y'_s\tilde{\mathbf{x}}^1_{s,t}\right\| \\ &\le \left\|Y,Y';\tilde Y,\tilde Y'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma}\omega^{\frac{1+\gamma}{p}}(s,t) + \left\|Y'_s-\tilde Y'_s\right\|\omega^{\frac{1}{p}}(s,t) + \left\|\tilde Y'_s\right\|\left\|\mathbf{x}^1_{s,t}-\tilde{\mathbf{x}}^1_{s,t}\right\| \qquad (12)\\ &\le \left(2\omega^{\frac{\gamma}{p}}(0,T)\left\|Y,Y';\tilde Y,\tilde Y'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma} + \left\|Y'_0-\tilde Y'_0\right\| + \left\|\tilde Y'_s\right\| d_\omega(\mathbf{x},\tilde{\mathbf{x}})\right)\omega^{\frac{1}{p}}(s,t).\end{aligned}$$

We can now prove:

Theorem 2.33 (Stability of rough integration)**:**
Let $p,\gamma,(Y,Y'),(\tilde Y,\tilde Y'),\mathbf{x},\tilde{\mathbf{x}}$ be as above. Let $(Y,Y'),(\tilde Y,\tilde Y')$ be uniformly bounded in the sense that

$$\left|Y'_0\right| + \left\|(Y,Y')\right\|_{\mathcal{D}^{\gamma}_{\omega,\mathbf{x}}} \le M, \qquad \left|\tilde Y'_0\right| + \left\|(\tilde Y,\tilde Y')\right\|_{\mathcal{D}^{\gamma}_{\omega,\mathbf{x}}} \le M.$$

Let

$$(\mathbf{z},\mathbf{z}') := \left(\int_0^\cdot Y\,d\mathbf{x}_t, Y\right) \in \mathcal{D}^{\gamma}_{\omega,\mathbf{x}}(\mathbb{R}^m)$$

and $(\tilde{\mathbf{z}},\tilde{\mathbf{z}}')$ analogously. Then it holds

$$\left\|\mathbf{z},\mathbf{z}';\tilde{\mathbf{z}},\tilde{\mathbf{z}}'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma} \le C\left(d_\omega(\mathbf{x},\tilde{\mathbf{x}}) + \left|Y_0-\tilde Y_0\right| + \left\|Y'_0-\tilde Y'_0\right\| + \omega^{\frac{\gamma}{p}}(0,T)\left\|Y,Y';\tilde Y,\tilde Y'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma}\right)$$

for some $C = C(M,p,\omega,\gamma)$.

Proof. To prove the statement, we need to find suitable bounds on $\left|\mathbf{z}'_{s,t}-\tilde{\mathbf{z}}'_{s,t}\right|$ and $\left|R^{\mathbf{z}}_{s,t}-R^{\tilde{\mathbf{z}}}_{s,t}\right|$, where

$$R^{\mathbf{z}}_{s,t} := \mathbf{z}_{s,t} - \mathbf{z}'_s\mathbf{x}^1_{s,t} = \int_s^t Y\,d\mathbf{x} - Y_s\mathbf{x}^1_{s,t} = (\mathcal{I}\Xi)_{s,t} - \Xi_{s,t} + Y'_s\mathbf{x}^2_{s,t}.$$

Here $\Xi_{s,t} := Y_s\mathbf{x}^1_{s,t} + Y'_s\mathbf{x}^2_{s,t}$ and $\mathcal{I}$ is the map constructed in the sewing lemma (Lemma 2.11). For $\Delta_{s,t} := \Xi_{s,t} - \tilde{\Xi}_{s,t}$, we have

$$\left|R^{\mathbf{z}}_{s,t}-R^{\tilde{\mathbf{z}}}_{s,t}\right| \le \left|(\mathcal{I}\Delta)_{s,t}-\Delta_{s,t}\right| + \left|Y'_s\mathbf{x}^2_{s,t}-\tilde Y'_s\tilde{\mathbf{x}}^2_{s,t}\right| \le C_\Delta\omega^{\frac{2+\gamma}{p}}(s,t) + \left|Y'_s\mathbf{x}^2_{s,t}-\tilde Y'_s\tilde{\mathbf{x}}^2_{s,t}\right|,$$

where

$$C_\Delta = 2^{\frac{2+\gamma}{p}+1}\zeta\left(\frac{2+\gamma}{p}\right)\sup_{s<u<t}\frac{\left|R^{\tilde Y}_{s,u}\tilde{\mathbf{x}}^1_{u,t} - R^{Y}_{s,u}\mathbf{x}^1_{u,t} + \tilde Y'_{s,u}\tilde{\mathbf{x}}^2_{u,t} - Y'_{s,u}\mathbf{x}^2_{u,t}\right|}{\omega^{\frac{2+\gamma}{p}}(s,t)}.$$

Using

$$\begin{aligned}
\left|R^{\tilde{Y}}_{s,u}\tilde{\mathbf{x}}^1_{u,t} - R^{Y}_{s,u}\mathbf{x}^1_{u,t} + \tilde{Y}'_{s,u}\tilde{\mathbf{x}}^2_{u,t} - Y'_{s,u}\mathbf{x}^2_{u,t}\right| &\le \left|R^{\tilde{Y}}_{s,u}(\tilde{\mathbf{x}}^1_{u,t} - \mathbf{x}^1_{u,t})\right| + \left|(R^{\tilde{Y}}_{s,u} - R^{Y}_{s,u})\mathbf{x}^1_{u,t}\right| \\
&\quad + \left|\tilde{Y}'_{s,u}(\tilde{\mathbf{x}}^2_{u,t} - \mathbf{x}^2_{u,t})\right| + \left|(\tilde{Y}'_{s,u} - Y'_{s,u})\mathbf{x}^2_{u,t}\right| \\
&\le \underbrace{\left\|(\tilde{Y}, \tilde{Y}')\right\|_{\mathcal{D}^{\gamma}_{\omega,\tilde{\mathbf{x}}}}}_{\le M} \omega^{\frac{1+\gamma}{p}}(s,u) d_{\omega}(\mathbf{x},\tilde{\mathbf{x}}) \omega^{\frac{1}{p}}(u,t) \\
&\quad + \left\|Y, Y'; \tilde{Y}, \tilde{Y}'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega} \omega^{\frac{1+\gamma}{p}}(s,u)\omega^{\frac{1}{p}}(u,t) \\
&\quad + \underbrace{\left\|(\tilde{Y}, \tilde{Y}')\right\|_{\mathcal{D}^{\gamma}_{\omega,\tilde{\mathbf{x}}}}}_{\le M} \omega^{\frac{\gamma}{p}}(s,u) \underbrace{d^2_{\omega}(\mathbf{x},\tilde{\mathbf{x}})}_{\le 2 d_{\omega}(\mathbf{x},\tilde{\mathbf{x}})} \omega^{\frac{2}{p}}(u,t) \\
&\quad + \left\|Y, Y'; \tilde{Y}, \tilde{Y}'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega} \omega^{\frac{\gamma}{p}}(s,u)\omega^{\frac{2}{p}}(u,t)
\end{aligned}$$

and

$$\begin{aligned}
\left|Y'_s\mathbf{x}^2_{s,t} - \tilde{Y}'_s\tilde{\mathbf{x}}^2_{s,t}\right| &\le \left|(Y'_s - \tilde{Y}'_s)\mathbf{x}^2_{s,t}\right| + \left|\tilde{Y}'_s(\mathbf{x}^2_{s,t} - \tilde{\mathbf{x}}^2_{s,t})\right| \\
&\le \left(\left\|Y, Y'; \tilde{Y}, \tilde{Y}'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega} \omega^{\frac{\gamma}{p}}(0,T) + \left\|Y'_0 - \tilde{Y}'_0\right\|\right) \omega^{\frac{2}{p}}(s,t) \\
&\quad + \left(\underbrace{\left\|\tilde{Y}'_0\right\|}_{\le M} + \omega^{\frac{\gamma}{p}}(0,T) \underbrace{\left\|(\tilde{Y}, \tilde{Y}')\right\|_{\mathcal{D}^{\gamma}_{\omega,\tilde{\mathbf{x}}}}}_{\le M}\right) \underbrace{d^2_{\omega}(\mathbf{x},\tilde{\mathbf{x}})}_{\le 2 d_{\omega}(\mathbf{x},\tilde{\mathbf{x}})} \omega^{\frac{2}{p}}(s,t)
\end{aligned}$$

shows the estimate on the $\left|R^{\mathbf{z}}_{s,t} - R^{\tilde{\mathbf{z}}}_{s,t}\right|$ part. For the other part, we simply notice $\mathbf{z}'_{s,t} - \tilde{\mathbf{z}}'_{s,t} = Y_{s,t} - \tilde{Y}_{s,t}$. The assertion then follows from Equation (12) and

$$\left\|\tilde{Y}'_s\right\| \le \left\|\tilde{Y}'_0\right\| + \left\|(\tilde{Y}, \tilde{Y}')\right\|_{\mathcal{D}^{\gamma}_{\omega,\tilde{\mathbf{x}}}} \omega^{\frac{\gamma}{p}}(0,s) \le \left(1 + \omega^{\frac{\gamma}{p}}(0,T)\right) M.$$

□

Remark 2.34:
Now, to show convergence of the rough integral using Theorem 2.33, we need convergence of the rough paths $\mathbf{x}$ and $\tilde{\mathbf{x}}$ in $d_{\omega}(\mathbf{x},\tilde{\mathbf{x}})$. This is a stronger condition than convergence in p-variation. To see this, consider the smooth ($p = 1$) paths $\mathbf{x} : x \mapsto x$ and $\mathbf{x}_{\alpha} : x \mapsto \begin{cases} 0 & x \le \alpha \\ \frac{x-\alpha}{1-\alpha} & x > \alpha \end{cases}$ on $[0,1]$. Then

$$d_{1-var}(\mathbf{x},\mathbf{x}_{\alpha}) = \mathrm{Var}_{1,[0,1]}\left(\left[x \mapsto \begin{cases} x & x \le \alpha \\ \alpha\frac{1-x}{1-\alpha} & x > \alpha \end{cases}\right]\right) = 2\alpha \xrightarrow{\alpha\to 0} 0,$$

but

$$d_{\omega}(\mathbf{x},\mathbf{x}_{\alpha}) \ge \frac{d_{1-var;[0,\alpha]}(\mathbf{x},\mathbf{x}_{\omega})}{\omega(0,\alpha)} = \frac{\alpha}{\alpha} = 1$$

with $\omega(s,t) = |t-s|$.

However, by slightly lowering the parameter of control γ, convergence in p-variation is enough:

Corollary 2.35:
Let $p, \gamma, (Y, Y'), (\tilde{Y}, \tilde{Y}'), \mathbf{x}, \tilde{\mathbf{x}}, \mathbf{z}, \tilde{\mathbf{z}}$ be as in Theorem 2.33. Let $\gamma > \delta > p-2$ with $|\gamma - \delta| \le 1$. Then we have

$$\left\|\mathbf{z}, \mathbf{z}'; \tilde{\mathbf{z}}, \tilde{\mathbf{z}}'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\delta} \le C\left(d^{\gamma-\delta}_{p-var}(\mathbf{x},\tilde{\mathbf{x}}) + \left|Y_0 - \tilde{Y}_0\right| + \left\|Y'_0 - \tilde{Y}'_0\right\| + \omega^{\frac{\gamma}{p}}(0,T)\left\|Y, Y'; \tilde{Y}, \tilde{Y}'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma}\right)$$

for some $C = C(M, p, \omega, \gamma, \delta)$.

Proof. The proof is essentially the same as for Theorem 2.33, but instead of estimating

$$\begin{aligned}\left|\mathbf{x}^1_{s,t} - \tilde{\mathbf{x}}^1_{s,t}\right| \leq & d_{p-var}(\mathbf{x}, \tilde{\mathbf{x}}) \leq d_\omega(\mathbf{x}, \tilde{\mathbf{x}})\omega^{\frac{1}{p}}(s,t), \\ \left|\mathbf{x}^2_{s,t} - \tilde{\mathbf{x}}^2_{s,t}\right| \leq & d^2_{p-var}(\mathbf{x}, \tilde{\mathbf{x}}) \leq d^2_\omega(\mathbf{x}, \tilde{\mathbf{x}})\omega^{\frac{2}{p}}(s,t),\end{aligned}$$

we estimate

$$\begin{aligned}\left|\mathbf{x}^1_{s,t} - \tilde{\mathbf{x}}^1_{s,t}\right| \leq & d_{p-var}(\mathbf{x}, \tilde{\mathbf{x}}) \leq d^{\gamma-\delta}_{p-var}(\mathbf{x}, \tilde{\mathbf{x}}) d^{1-(\gamma-\delta)}_\omega(\mathbf{x}, \tilde{\mathbf{x}})\omega^{\frac{1-(\gamma-\delta)}{p}}(s,t) \\ \leq & 2^{1-(\gamma-\delta)} d^{\gamma-\delta}_{p-var}(\mathbf{x}, \tilde{\mathbf{x}})\omega^{\frac{1-(\gamma-\delta)}{p}}(s,t)\end{aligned}$$

and

$$\begin{aligned}\left|\mathbf{x}^2_{s,t} - \tilde{\mathbf{x}}^2_{s,t}\right| \leq & d^2_{p-var}(\mathbf{x}, \tilde{\mathbf{x}}) \leq d^{\gamma-\delta}_{p-var}(\mathbf{x}, \tilde{\mathbf{x}}) d^{2-(\gamma-\delta)}_\omega(\mathbf{x}, \tilde{\mathbf{x}})\omega^{\frac{2-(\gamma-\delta)}{p}}(s,t) \\ \leq & 2^{2-(\gamma-\delta)} d^{\gamma-\delta}_{p-var}(\mathbf{x}, \tilde{\mathbf{x}})\omega^{\frac{2-(\gamma-\delta)}{p}}(s,t).\end{aligned}$$

□

3 Rough Differential Equations

After defining the rough integral, we can now lay our focus on rough differential equations. From here on out, we will only be concerned with the case $2 \leq p < 3$. This section is a generalization of the RDE-results from [FH20, Chapter 8] to general rough paths and RDEs with drift term. First, we need to understand what $f(Y)$ for a controlled rough path Y and $f : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ means.

Let W be a Banach space and let $\varphi : \mathbb{R}^m \to W$ be a bounded function with bounded, continuous second derivatives ($\varphi \in C_b^2$). Let $(Y, Y') \in \mathcal{D}_{\omega,\mathbf{x}}^{\gamma}(\mathbb{R}^m)$ be a controlled rough path. Then we can define

$$\begin{aligned}\varphi(Y)_t :=& \varphi(Y_t),\\ \varphi(Y)'_t :=& D_Y\varphi(Y_t)Y'_t.\end{aligned}$$

Here $D\varphi(y)$ is the map

$$\mathbb{R}^m \ni x \mapsto D\varphi(y)x := \lim_{h \searrow 0} \frac{\varphi(y + hx) - \varphi(y)}{h} \in W,$$

so $\varphi(Y)'_t \in \mathcal{L}(\mathbb{R}^n, W)$. Now, we want to show that $(\varphi(Y), \varphi(Y)')$ is indeed a controlled rough path.

Lemma 3.1:
Let $2 \leq p < 3$, $\varphi \in C_b^2(\mathbb{R}^m, W)$ and let $(Y, Y') \in \mathcal{D}_{\omega,\mathbf{x}}^{\gamma}(\mathbb{R}^m)$ be a controlled rough path for some rough path $\mathbf{x} \in \Omega^p$, a controlling function ω, and $\gamma > p - 2$. Let $(\varphi(Y), \varphi(Y)')$ be defined as above. Then $(\varphi(Y), \varphi(Y)') \in \mathcal{D}_{\omega,\mathbf{x}}^{\gamma\wedge 1}(W)$ is a controlled rough path. In particular, we have

$$\begin{aligned}\left\|(\varphi(Y), \varphi(Y)')\right\|_{\mathcal{D}_{\omega,\mathbf{x}}^{\gamma}} \leq& \left(2\left\|D\varphi\right\|_\infty + \left\|D^2\varphi\right\|_\infty \left\|Y'\right\|_\infty \omega^{\frac{1}{p}}(0,T)\right) \left\|(Y, Y')\right\|_{\mathcal{D}_{\omega,\mathbf{x}}^{\gamma}}\\ &+ 2\left\|D^2\varphi\right\|_\infty \left(\left\|(Y, Y')\right\|^2_{\mathcal{D}_{\omega,\mathbf{x}}^{\gamma}} \omega^{\frac{2\gamma}{p}}(0,T) + 2\left\|Y'\right\|^2_\infty\right) \omega^{\frac{1-\gamma}{p}}(0,T).\end{aligned}$$

Proof. Without loss of generality, let $\gamma \leq 1$. We have

$$\begin{aligned}\left\|\varphi(Y)'_t - \varphi(Y)'_s\right\| =& \left\|D\varphi(Y_t)Y'_t - D\varphi(Y_t)Y'_s + D\varphi(Y_t)Y'_s - D\varphi(Y_s)Y'_s\right\|\\ \leq& \left\|D\varphi\right\|_\infty \left\|Y'_t - Y'_s\right\| + \left\|D\varphi(Y_t) - D\varphi(Y_s)\right\| \left\|Y'\right\|_\infty\\ \leq& \left\|D\varphi\right\|_\infty \left\|(Y, Y')\right\|_{\mathcal{D}_{\omega,\mathbf{x}}^{\gamma}} \omega^{\frac{\gamma}{p}}(s,t) + \left\|D^2\varphi\right\|_\infty \left\|Y'\right\|_\infty |Y_t - Y_s|\\ \leq& \left\|(Y, Y')\right\|_{\mathcal{D}_{\omega,\mathbf{x}}^{\gamma}} \left(\left\|D\varphi\right\|_\infty + \left\|D^2\varphi\right\|_\infty \left\|Y'\right\|_\infty \omega^{\frac{1}{p}}(0,T)\right) \omega^{\frac{\gamma}{p}}(s,t)\\ &+ \left\|D^2\varphi\right\|_\infty \left\|Y'\right\|^2_\infty \omega^{\frac{1}{p}}(s,t),\end{aligned}$$

where we used the first inequality of Equation (10). For

$$R^{\varphi}_{s,t} := \varphi(Y_t) - \varphi(Y_s) - D\varphi(Y_s)Y'_s\mathbf{x}^1_{s,t} = \varphi(Y_t) - \varphi(Y_s) - D\varphi(Y_s) \underbrace{Y_{s,t}}_{=Y_t - Y_s} + D\varphi(Y_s)R^Y_{s,t}$$

we can use the Taylor formula, to see

$$\begin{aligned}\left\|R^{\varphi}_{s,t}\right\| \leq& \frac{1}{2}\left\|D^2\varphi\right\|_\infty |Y_t - Y_s|^2 + \left\|D\varphi\right\|_\infty \left\|R^Y_{s,t}\right\|\\ \leq& 2\left\|D^2\varphi\right\|_\infty \left(\left\|(Y, Y')\right\|^2_{\mathcal{D}_{\omega,\mathbf{x}}^{\gamma}} \omega^{\frac{2\gamma}{p}}(s,t) + \left\|Y'\right\|^2_\infty\right) \omega^{\frac{2}{p}}(s,t)\\ &+ \left\|D\varphi\right\|_\infty \left\|(Y, Y')\right\|_{\mathcal{D}_{\omega,\mathbf{x}}^{\gamma}} \omega^{\frac{1+\gamma}{p}}(s,t)\end{aligned}$$

with $(|a| + |b|)^2 \leq (2\max(|a|, |b|))^2 = 4\max(|a|^2, |b|^2) \leq 4(|a|^2 + |b|^2)$. Therefore $(\varphi(Y), \varphi(Y)') \in \mathcal{D}_{\omega,\mathbf{x}}^{\gamma}(W)$ with the aforementioned bound. $\square$

Using this, we can define rough differential equations:

Definition 3.2 (RDEs)**:**
Let $f : \mathbb{R}^m \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. A *Rough Differential Equation (RDE)* is an equation of the form

$$dY_t = \mu(Y_t, t)dt + f(Y_t, t)d\mathbf{x}_t, \tag{13}$$

where $\mathbf{x}_t \in \Omega^p(\mathbb{R}^n)$ is the *driving rough path.* A controlled rough path $Y \in \mathcal{D}^\gamma_{\omega,\mathbf{x}}(\mathbb{R}^m)$ with $\gamma > p-2$ and $Y' = f(Y)$ is called a solution to Equation (13), started at $\xi \in \mathbb{R}^m$, given

$$Y_t = \xi + \int_0^t \mu(Y_\tau, \tau)d\tau + \int_0^t f(Y_\tau, \tau)d\mathbf{x}_\tau.$$

The integral is the one we defined in Theorem 2.29 for the controlled rough path $(f(Y), f(Y)') \in \mathcal{D}^\gamma_{\omega,\mathbf{x}}(\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$.

3.1 RDEs without Drift

As usual, one is interested in the existence and uniqueness of solutions to Equation (13). In the case of rough paths, this is done using Picard iteration. First, we only consider the case of no drift ($\mu = 0$), based on [FH20, Theorem 8.3].

Theorem 3.3:
Let $2 \le p < 3$, $\xi \in \mathbb{R}^m$, $f \in C^3(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$ and a rough path $\mathbf{x} \in \Omega^p(\mathbb{R}^n)$. Then there exists a time $0 < T_0 \le T$ and a unique element $(Y, Y') \in \mathcal{D}^1_{\mathbf{x}}(\mathbb{R}^m)$ with $Y' = f(Y)$, such that

$$Y_t = \xi + \int_0^t f(Y_\tau)d\mathbf{x}_\tau$$

for all $0 \le t \le T_0$. T_0 does not dependent on ξ. If we find an element $(\tilde{Y}, \tilde{Y}') \in \mathcal{D}^\gamma_{\mathbf{x}}(\mathbb{R}^m)$ for $1 > \gamma > p-2$, which is a solution on $[0, T_0]$, then it already holds $(Y, Y') = (\tilde{Y}, \tilde{Y}') \in \mathcal{D}^1_{\mathbf{x}}(\mathbb{R}^m)$.

Proof. Let $1 > \gamma > p-2$ and let $(Y, Y') \in \mathcal{D}^\gamma_{\mathbf{x}}$. From Lemma 3.1, we know that

$$(f(Y), f(Y)') := (f(Y), Df(Y)Y') \in \mathcal{D}^\gamma_{\mathbf{x}}(\mathbb{R}^m).$$

This allows us to define the map

$$\mathcal{M}_{T_0}(Y, Y') := \left(\xi + \int_0^{\cdot\wedge T_0} f(Y)d\mathbf{x}, f(Y)_{\cdot\wedge T_0}\right) \in \mathcal{D}^\gamma_{\mathbf{x}}(\mathbb{R}^m).$$

The solution to Equation (13) is a fixed point of this map, which a priori only is in $\mathcal{D}^\gamma_{\mathbf{x}}$. This solution then already lies in $\mathcal{D}^1_{\mathbf{x}}$:
From Equation (11), we see that

$$\left\| \int_s^t f(Y_\tau)d\mathbf{x}_\tau - f(Y_s)\mathbf{x}^1_{s,t} \right\| \le C \left\|(Y, Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}} \operatorname{Var}^2_{p,[s,t]}(\mathbf{x})$$

and from Equation (10), we see that

$$\|f(Y_t) - f(Y_s)\| \le \|Df\|_\infty \|Y_t - Y_s\| \le \|Df\|_\infty \left(2\left\|(Y, Y')\right\|_{\mathcal{D}^\gamma_{\mathbf{x}}} \operatorname{Var}^\gamma_{p,[0,T]}(\mathbf{x}) + \|Y_0\|\right) \operatorname{Var}_{p,[s,t]}(\mathbf{x}).$$

Hence the fixed point is already in $\mathcal{D}^1_{\mathbf{x}}$ and therefore is the solution we are looking for.
$\mathcal{M}_{T_0}$ can be viewed as a map on the space of controlled rough paths started in $(\xi, f(\xi))$:

$$\mathcal{B}_\xi := \left\{(Y, Y') \in \mathcal{D}^\gamma_{\mathbf{x}}(\mathbb{R}^m) \middle| Y_0 = \xi, Y'_0 = f(\xi)\right\}.$$

Since $\mathcal{D}^\gamma_{\mathbf{x}}$ is a Banach space, this is complete under the induced norm. Consider the closed unit ball $\mathbb{D}_\xi$ centered at

$$t \mapsto c(t) = (\xi + f(\xi)\mathbf{x}^1_{0,t}, f(\xi)).$$

in $\mathcal{B}_\xi$. $c(t)$ is an element of $\mathcal{D}^\gamma_{\mathbf{x}}$, since

$$\begin{aligned}\xi + f(\xi)\mathbf{x}^1_{0,t} - \xi - f(\xi)\mathbf{x}^1_{0,s} - f(\xi)\mathbf{x}^1_{s,t} = 0 \text{ and}\\ f(\xi) - f(\xi) = 0.\end{aligned}$$

Also, $\mathbb{D}_\xi$ is the set of all $(Y, Y') \in \mathcal{D}^\gamma_{\mathbf{x}}$ with $Y_0 = \xi$ and $Y'_0 = f(\xi)$, such that

$$\begin{aligned}\left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\mathbf{x}}} =& \inf\left\{C>0 \middle| \left\|Y_t - Y_s - Y'_s\mathbf{x}^1_{s,t}\right\|_W \le C \operatorname{Var}^{\gamma+1}_{p,[s,t]}(\mathbf{x})\right\}\\ &+ \inf\left\{C>0 \middle| \left\|Y'_t - Y'_s\right\|_{\mathcal{L}(\mathbb{R}^n,W)} \le C \operatorname{Var}^{\gamma}_{p,[s,t]}(\mathbf{x})\right\}\\ \le& 1,\end{aligned}$$

since $\|(Y,Y') - c\|_{\mathcal{D}^\gamma_{\mathbf{x}}} = \|(Y,Y')\|_{\mathcal{D}^\gamma_{\mathbf{x}}}$, as the additional terms always cancel each other out.
Invariance: Now we show that for T_0 small enough, $\mathcal{M}_{T_0}$ maps $\mathbb{D}_\xi$ to itself. Obviously $\xi + \int_0^0 f(Y)d\mathbf{x} = \xi$ and $f(Y)_0 = f(\xi)$. We have

$$\begin{aligned}\left\|\int_s^t f(Y_\tau)d\mathbf{x}_\tau - f(Y_s)\mathbf{x}^1_{s,t}\right\| \le& C\left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}} \operatorname{Var}^2_{p,[s,t]}(\mathbf{x})\\ =& C\left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}} \operatorname{Var}^{1-\gamma}_{p,[s,t]}(\mathbf{x}) \operatorname{Var}^{1+\gamma}_{p,[s,t]}(\mathbf{x})\end{aligned}$$

and

$$\begin{aligned}\|f(Y_t) - f(Y_s)\| \le& C\left(\left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\mathbf{x}}} + \left\|Y'\right\|_\infty\right) \operatorname{Var}_{p,[s,t]}(\mathbf{x})\\ =& C\left(\left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\mathbf{x}}} + \left\|Y'\right\|_\infty\right) \operatorname{Var}^{1-\gamma}_{p,[s,t]}(\mathbf{x}) \operatorname{Var}^{\gamma}_{p,[s,t]}(\mathbf{x})\end{aligned}$$

Now since we only consider the process up to time T_0, we have

$$\begin{aligned}\left\|M_{T_0}(Y,Y')\right\|_{\mathcal{D}^\gamma_{\mathbf{x}}} \le& C\left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}} \operatorname{Var}^{1-\gamma}_{p,[s,t]}(\mathbf{x}) + C\left(\left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\mathbf{x}}} + \left\|Y'\right\|_\infty\right) \operatorname{Var}^{1-\gamma}_{p,[s,t]}(\mathbf{x})\\ \le& C\left(2\left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\mathbf{x}}} + \left\|Y'\right\|_\infty\right) \operatorname{Var}^{1-\gamma}_{p,[0,T_0]}(\mathbf{x}).\end{aligned}$$

Therefore, we can choose $T_0 > 0$ small enough, such that this is smaller than 1. Note that this choice can be made independent of ξ, since $\|(Y,Y')\|_{\mathcal{D}^\gamma_{\mathbf{x}}} \le 1$ and the constants in Equations (10) and (11) depend on $\|Y'\|_\infty$, which we can control using

$$\left\|Y'_t\right\| \le \| \underbrace{Y'_0}_{=f(\xi)} \| + \left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\mathbf{x}}} \le \|f\|_\infty + 1$$

independent of ξ. Hence $\mathcal{M}_{T_0}$ leaves $\mathbb{D}_\xi$ invariant for small $T_0 > 0$, independent of ξ.
Contraction: If we can show, that $\mathcal{M}_{T_0}$ is contracting in $\mathbb{D}_\xi$ for small $T_0 > 0$, we know that it possesses a unique fixed point. For this, we consider $(Y, Y'), (\tilde{Y}, \tilde{Y}') \in \mathbb{D}_\xi$ and set

$$\begin{aligned}\Delta_t :=& f(Y_t) - f(\tilde{Y}_t),\\ \Delta'_t :=& f(Y_t)' - f(\tilde{Y}_t)' = D_Y f(Y_t)Y'_t - D_Y f(\tilde{Y}_t)\tilde{Y}'_t.\end{aligned}$$

Then it holds

$$\left\|\mathcal{M}_{T_0}(Y,Y'); \mathcal{M}_{T_0}(\tilde{Y},\tilde{Y}')\right\|_{\mathbf{x},\mathbf{x},\operatorname{Var}^p_p,\gamma} = \left\|\int_0^{\cdot\wedge T_0} \Delta_\tau d\mathbf{x}_\tau, \Delta_{\cdot\wedge T_0}\right\|_{\mathcal{D}^\gamma_{\mathbf{x}}},$$

and by Equation (11)

$$\left|\int_s^t \Delta_\tau d\mathbf{x}_\tau - \Delta'_s \mathbf{x}^1_{s,t}\right| \le C\left(\left\|(\Delta,\Delta')\right\|_{\mathcal{D}^\gamma_{\mathbf{x}}} \operatorname{Var}^\gamma_{p,[0,T]}(\mathbf{x}) + \left\|\Delta'\right\|_\infty\right) \operatorname{Var}^2_{p,[s,t]}(\mathbf{x}),$$

as well as

$$\left\|\Delta'\right\|_\infty \le \underbrace{\left\|\Delta'_0\right\|}_{=0} + \left\|(\Delta,\Delta')\right\|_{\mathcal{D}^\gamma_{\mathbf{x}}} \operatorname{Var}^\gamma_{p,[0,T]}(\mathbf{x}).$$

We know from Equation (10), that

$$\begin{aligned}|\Delta_t - \Delta_s| \le& \big(2\left\|(\Delta,\Delta')\right\|_{\mathcal{D}^\gamma_{\mathbf{x}}} \operatorname{Var}^\gamma_{p,[0,T]}(\mathbf{x}) + \underbrace{\left\|\Delta_0\right\|}_{=0}\big) \operatorname{Var}_{p,[s,t]}(\mathbf{x}) \\ \le& 2\operatorname{Var}^\gamma_{p,[0,T]}(\mathbf{x}) \left\|(\Delta,\Delta')\right\|_{\mathcal{D}^\gamma_{\mathbf{x}}} \operatorname{Var}^{1-\gamma}_{p,[s,t]}(\mathbf{x}) \operatorname{Var}^\gamma_{p,[s,t]}(\mathbf{x}).\end{aligned}$$

If we can show that $\|(\Delta,\Delta')\|_{\mathcal{D}^\gamma_{\mathbf{x}}} \le C\left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\mathbf{x},\operatorname{Var}^p_p,\gamma}$, we see that

$$\begin{aligned}\left\|\mathcal{M}_{T_0}(Y,Y');\mathcal{M}_{T_0}(\tilde{Y},\tilde{Y}')\right\|_{\mathbf{x},\mathbf{x},\operatorname{Var}^p_p,\gamma} \le& \left(\left\|(\Delta,\Delta')\right\|_{\mathcal{D}^\gamma_{\mathbf{x}}} \operatorname{Var}^\gamma_{p,[0,T]}(\mathbf{x}) + \left\|\Delta'\right\|_\infty\right) \operatorname{Var}^{1-\gamma}_{p,[s,t]}(\mathbf{x}) \\ &+ 2\operatorname{Var}^\gamma_{p,[0,T]}(\mathbf{x}) \left\|(\Delta,\Delta')\right\|_{\mathcal{D}^\gamma_{\mathbf{x}}} \operatorname{Var}^{1-\gamma}_{p,[s,t]}(\mathbf{x}) \\ \le& C_{\mathbf{x},\gamma}\left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\mathbf{x},\operatorname{Var}^p_p,\gamma} \operatorname{Var}^{1-\gamma}_{p,[0,T_0]}(\mathbf{x}).\end{aligned}$$

Then, we can choose $T_0 > 0$ small enough, such that

$$\left\|\mathcal{M}_{T_0}(Y,Y');\mathcal{M}_{T_0}(\tilde{Y},\tilde{Y}')\right\|_{\mathbf{x},\mathbf{x},\operatorname{Var}^p_p,\gamma} \le \frac{1}{2}\left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\mathbf{x},\operatorname{Var}^p_p,\gamma}$$

independent of ξ. To show the assertion $\|(\Delta,\Delta')\|_{\mathcal{D}^\gamma_{\mathbf{x}}} \le C\left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\mathbf{x},\operatorname{Var}^p_p,\gamma}$, we construct

$$\Delta_s = G_s H_s$$

with

$$\begin{aligned}H_s =& Y_s - \tilde{Y}_s \in \mathbb{R}^m, \\ H'_s =& Y'_s - \tilde{Y}'_s \in \mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)\end{aligned}$$

and, since $f \in C^3$,

$$\begin{aligned}g(x,y) =& \int_0^1 Df(tx+(1-t)y)dt, \\ G_s =& g(Y_s,\tilde{Y}_s) \in \mathcal{L}(\mathbb{R}^m,\mathbb{R}^m), \\ G'_s =& (D_Y g(Y_s,\tilde{Y}_s))Y'_s + (D_{\tilde{Y}} g(Y_s,\tilde{Y}_s))\tilde{Y}'_s \in \mathcal{L}(\mathbb{R}^n,\mathcal{L}(\mathbb{R}^m,\mathbb{R}^m)).\end{aligned}$$

Then we see

$$\frac{\partial}{\partial t}\left[t \mapsto f(y+t(x-y))\right] = Df(y+t(x-y))(x-y),$$

so indeed

$$G_s H_s = \int_0^1 Df(tY_s+(1-t)\tilde{Y}_s)(Y_s-\tilde{Y}_s)dt = f(Y_s) - f(\tilde{Y}_s) = \Delta_s.$$

Obviously $(H, H') \in \mathcal{D}_{\mathbf{x}}^{\gamma}$ and $(G_s, G_s') \in \mathcal{D}_{\mathbf{x}}^{\gamma}$ analog to Lemma 3.1. Now we can use Lemma 2.28, to see that

$$\left\|(\Delta, \Delta')\right\|_{\mathcal{D}_{\mathbf{x}}^{\gamma}} \leq C(\|G_0\| + \|G_0'\| + \left\|(G, G')\right\|_{\mathcal{D}_{\mathbf{x}}^{\gamma}})(\underbrace{\|H_0\|}_{=0} + \underbrace{\|H_0'\|}_{=0} + \underbrace{\left\|(H, H')\right\|_{\mathcal{D}_{\mathbf{x}}^{\gamma}}}_{=\left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\mathbf{x},\mathrm{Var}_p^p,\gamma}}).$$

Let's estimate the parts containing G and G' independently:

$$\|G_0\| \leq \|g\|_{\infty} \leq \|Df\|_{\infty},$$

$$\|G_0'\| \leq \|Dg\|_{\infty} \Big(\underbrace{\|Y_0'\|}_{=\|f(\xi)\|} + \underbrace{\left\|\tilde{Y}_0'\right\|}_{=\|f(\xi)\|} \Big) \leq 2 \left\|D^2 f\right\|_{\infty} \|f\|_{\infty},$$

and from the proof of Lemma 3.1, we know that

$$\begin{aligned}\left\|(D_Y g(Y_t, \tilde{Y}_t))Y_t' - (D_Y g(Y_s, \tilde{Y}_s))Y_s'\right\| \leq \Bigg[&\underbrace{\left\|(Y, Y')\right\|_{\mathcal{D}_{\mathbf{x}}^{\gamma}}}_{\leq 1} \Bigg(\underbrace{\|D_Y g\|_{\infty}}_{\leq \|D^2 f\|_{\infty}} + \underbrace{\left\|D_Y^2 g\right\|_{\infty}}_{\leq \|D^3 f\|_{\infty}} \|Y'\|_{\infty} \mathrm{Var}_{p,[0,T]}(\mathbf{x}) \Bigg) \\ &+ \underbrace{\left\|D^2 g\right\|_{\infty}}_{\leq \|D^3 f\|_{\infty}} \|Y'\|_{\infty}^2 \mathrm{Var}_{p,[0,T]}^{1-\gamma}(\mathbf{x}) \Bigg] \mathrm{Var}_{p,[s,t]}^{\gamma}(\mathbf{x}),\end{aligned}$$

and analogous for $\left\|(D_{\tilde{Y}} g(Y_t, \tilde{Y}_t))\tilde{Y}_t' - (D_{\tilde{Y}} g(Y_s, \tilde{Y}_s))\tilde{Y}_s'\right\|$, as well as

$$\begin{aligned}\left\|R_{s,t}^g\right\| \leq& \frac{1}{2} \left\|D_Y^2 g\right\|_{\infty} |Y_t - Y_s|^2 + \frac{1}{2} \left\|D_{\tilde{Y}}^2 g\right\|_{\infty} \left|\tilde{Y}_t - \tilde{Y}_s\right|^2 + \|D_Y g\|_{\infty} \left\|R_{s,t}^Y\right\| + \left\|D_{\tilde{Y}} g\right\|_{\infty} \left\|R_{s,t}^{\tilde{Y}}\right\| \\ \leq& \frac{1}{2} \left\|D^3 f\right\|_{\infty} \Bigg[\Bigg(\underbrace{\left\|(Y, Y')\right\|_{\mathcal{D}_{\mathbf{x}}^{\gamma}}}_{\leq 1} \mathrm{Var}_{p,[0,T]}^{\gamma}(\mathbf{x}) + \|Y'\|_{\infty} \Bigg)^2 \\ &+ \Bigg(\underbrace{\left\|(\tilde{Y}, \tilde{Y}')\right\|_{\mathcal{D}_{\mathbf{x}}^{\gamma}}}_{\leq 1} \mathrm{Var}_{p,[0,T]}^{\gamma}(\mathbf{x}) + \|Y'\|_{\infty} \Bigg)^2 \Bigg] \mathrm{Var}_{p,[s,t]}^2(\mathbf{x}) \\ &+ \left\|D^2 f\right\|_{\infty} \Bigg(\underbrace{\left\|(Y, Y')\right\|_{\mathcal{D}_{\mathbf{x}}^{\gamma}}}_{\leq 1} + \underbrace{\left\|(\tilde{Y}, \tilde{Y}')\right\|_{\mathcal{D}_{\mathbf{x}}^{\gamma}}}_{\leq 1} \Bigg) \mathrm{Var}_{p,[s,t]}^{1+\gamma}(\mathbf{x}),\end{aligned}$$

using the first inequality of Equation (10).
Therefore, there is some $C > 0$ only depending on γ and $\mathbf{x}$, such that

$$\left\|(\Delta, \Delta')\right\|_{\mathcal{D}_{\mathbf{x}}^{\gamma}} \leq C \left(1 + \|f\|_{C_b^3}\right)^4 \left\|Y, Y'; \tilde{Y}, \tilde{Y}'\right\|_{\mathbf{x},\mathbf{x},\mathrm{Var}_p^p,\gamma},$$

since

$$\|Y'\|_{\infty} \leq \underbrace{\|Y_0'\|}_{=\|f(\xi)\| \leq \|f\|_{\infty}} + \underbrace{\left\|(Y, Y')\right\|_{\mathcal{D}_{\mathbf{x}}^{\gamma}}}_{\leq 1} \mathrm{Var}_{p,[0,T]}^{\gamma}(\mathbf{x}).$$

Hence we can find $T_0 > 0$ small enough such that

$$\left\|\mathcal{M}_{T_0}(Y, Y'); \mathcal{M}_{T_0}(\tilde{Y}, \tilde{Y}')\right\|_{\mathbf{x},\mathbf{x},\mathrm{Var}_p^p,\gamma} \leq \frac{1}{2} \left\|Y, Y'; \tilde{Y}, \tilde{Y}'\right\|_{\mathbf{x},\mathbf{x},\mathrm{Var}_p^p,\gamma}.$$

Therefore $\mathcal{M}_{T_0}$ admits the unique fixed point $(Y, Y') \in \mathbb{D}_\xi$, which is the unique solution to the RDE on the interval $[0, T_0]$. T_0 does *not* depend on the starting point ξ. By the uniform continuity of the p-variation (Lemma 2.4), we can also choose T_0 such that

$$\left\| \mathcal{M}_{T_0}(Y, Y'); \mathcal{M}_{T_0}(\tilde{Y}, \tilde{Y}') \right\|_{\tilde{\mathbf{x}}, \tilde{\mathbf{x}}, \mathrm{Var}_p^p, \gamma} \leq \frac{1}{2} \left\| Y, Y'; \tilde{Y}, \tilde{Y}' \right\|_{\tilde{\mathbf{x}}, \tilde{\mathbf{x}}, \mathrm{Var}_p^p, \gamma},$$

when considering the path started at a different time: $[s \mapsto \tilde{\mathbf{x}}_s] = [s \mapsto \mathbf{x}_{t+s}]$ for all $t > 0$. □

Now, we can stitch multiple solutions together, to get the full solution:

Corollary 3.4:
Let $2 \leq p < 3$, $\xi \in \mathbb{R}^m$, $f \in C^3(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$ and a rough path $\mathbf{x} \in \Omega^p(\mathbb{R}^n)$. Then there exists a unique element $(Y, Y') \in \mathcal{D}^1_{\mathbf{x}}(\mathbb{R}^m)$ with $Y' = f(Y)$, such that

$$Y_t = \xi + \int_0^t f(Y_\tau) d\mathbf{x}_\tau$$

for all $0 \leq t \leq T$. If we find an element $(\tilde{Y}, \tilde{Y}') \in \mathcal{D}^\gamma_{\mathbf{x}}(\mathbb{R}^m)$ for $1 > \gamma > p - 2$, which fulfills the condition above, then it already holds $(Y, Y') = (\tilde{Y}, \tilde{Y}') \in \mathcal{D}^1_{\mathbf{x}}(\mathbb{R}^m)$.

Proof. Let T_0 be the size of the interval of the unique solution from Theorem 3.3. Let $(Y^{(0)}, Y'^{(0)})$ be the unique solution on $[0, T_0]$. Then we can build up the full solution inductively:
Let $(Y^{(k)}, Y'^{(k)})$ be the unique solution on $[0, kT_0]$. Consider the rough path $\mathbf{x}^{(k)}_t := \mathbf{x}_{kT_0 + t \wedge T}$. Let $(\tilde{Y}_k, \tilde{Y}'_k)$ be the unique solution of

$$\tilde{Y}_t = Y^{(k)}_{kT_0} + \int_0^t f(\tilde{Y}_\tau) d\mathbf{x}^{(k)}_\tau$$

from Theorem 3.3 on the interval $[0, T_0]$. Then

$$(Y^{(k+1)}_t, Y'^{(k+1)}_t) := \begin{cases} (Y^{(k)}_t, Y'^{(k)}_t) & t \leq kT_0 \\ ((\tilde{Y}_k)_{t-kT_0}, (\tilde{Y}'_k)_{t-kT_0}) & t \geq kT_0 \end{cases}$$

is the unique solution of the problem on the interval $[0, (k+1)T_0]$. Therefore, we have constructed the unique solution on the whole interval $[0, T]$ inductively. A solution $(Y, Y') \in \mathcal{D}^\gamma_{\mathbf{x}}$ is already in $\mathcal{D}^1_{\mathbf{x}}$, since the same holds true for its pieces $(\tilde{Y}_k, \tilde{Y}'_k)$. □

3.2 RDEs with Drift

We can extend Theorem 3.3 to RDEs of the form

$$dY_t = \mu(t, Y_t) dt + f(t, Y_t) d\mathbf{x}_t \tag{14}$$

for a rough path $\mathbf{x} \in \Omega^p$ as follows:

Definition 3.5:
For $2 \leq p < 3$ and a rough path $\mathbf{x} = 1 + \mathbf{x}^1 + \mathbf{x}^2 \in \Omega^p(\mathbb{R}^n)$, define the *augmented rough path*

$$\hat{\mathbf{x}} := 1 + \hat{\mathbf{x}}^1 + \hat{\mathbf{x}}^2,$$

where

$$\hat{\mathbf{x}}^1_t := (\mathbf{x}^1_t, t) \in \mathbb{R}^{n+1}$$

and $\hat{\mathbf{x}}^2$ is suitable. $\hat{\Omega}^p(\mathbb{R}^n)$ is the set of augmented rough paths.

Then we can rewrite Equation (14) to be

$$d\hat{Y}_t = \hat{f}(\hat{Y}_t)d\hat{\mathbf{x}}_t,$$

with $\hat{Y}_t = (Y_t, t)$ and $\hat{f}(\hat{Y}_t) \in \mathcal{L}(\mathbb{R}^{n+1}, \mathbb{R}^{m+1})$:

$$\sum_{j=1}^{n+1} \alpha_j e^{(j)} \mapsto \hat{f}(\hat{Y}_t) \left(\sum_{j=1}^{n+1} \alpha_j e^{(j)} \right) := \underbrace{f(\hat{Y}_t) \left(\sum_{j=1}^{n} \alpha_j e^{(j)} \right)}_{\in \mathbb{R}^m \widehat{=} \mathbb{R}^m \times \{0\} \subset \mathbb{R}^{m+1}} + \underbrace{\mu(\hat{Y}_t)\alpha_{n+1}}_{\in \mathbb{R}^m \widehat{=} \mathbb{R}^m \times \{0\} \subset \mathbb{R}^{m+1}} + \alpha_{n+1} e^{(m+1)}.$$

To be able to employ Corollary 3.4, we need $\hat{f} \in C^3$, which means $\mu \in C^3$. It can be shown, that μ continuous in t and Lipschitz in Y_t suffices, by considering the following:

Theorem 3.6:
Let $2 \le p < 3$, $\xi \in \mathbb{R}^m$, $f \in C^3(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$ and a rough path $\mathbf{x} \in \Omega^p(\mathbb{R}^n)$. Let $\mu : \mathbb{R} \times \mathbb{R}^m \ni (t, y) \mapsto \mu(t, y) \in \mathbb{R}^m$ be continuous in t and uniformly Lipschitz in y. Let ω be a controlling function with

$$\omega(s,t) \ge \mathrm{Var}^p_{p,[s,t]}(\mathbf{x}) + |t-s|$$

for all $0 \le s \le t \le T$. Then there exists a unique element $(Y, Y') \in \mathcal{D}^1_{\omega,\mathbf{x}}(\mathbb{R}^m)$ with $Y' = f(Y)$, such that

$$Y_t = \xi + \int_0^t \mu(\tau, Y_\tau)d\tau + \int_0^t f(Y_\tau)d\mathbf{x}_\tau$$

for all $0 \le t \le T$. If we find an element $(\tilde{Y}, \tilde{Y}') \in \mathcal{D}^\gamma_{\omega,\mathbf{x}}(\mathbb{R}^m)$ for $1 > \gamma > p-2$, which fulfills the condition above, then it already holds $(Y, Y') = (\tilde{Y}, \tilde{Y}') \in \mathcal{D}^1_{\omega,\mathbf{x}}(\mathbb{R}^m)$.

Proof. We will modify the proof of Theorem 3.3. Let $1 > \gamma > p-2$. First of all, consider $(Y, Y') \in \mathcal{D}^\gamma_{\omega,\mathbf{x}}$. Then also

$$(f(Y), f(Y)') := (f(Y), Df(Y)Y') \in \mathcal{D}_{\omega,\mathbf{x}}(\mathbb{R}^m).$$

We now consider

$$\mathcal{M}_{T_0}(Y, Y') := \left(\xi + \int_0^{\cdot\wedge T_0} \mu(t, Y_t)dt + \int_0^{\cdot\wedge T_0} f(Y_t)d\mathbf{x}_t, f(Y)_{\cdot\wedge T} \right).$$

Note that the Gubinelli derivative does not change when considering a drift term, as the integration against time is smooth enough on its own. This is in $\mathcal{D}^\gamma_{\omega,\mathbf{x}}(\mathbb{R}^m)$, since

$$\left(\xi + \int_0^{\cdot\wedge T_0} f(Y_t)d\mathbf{x}_t, f(Y)_{\cdot\wedge T_0} \right) \in \mathcal{D}^\gamma_{\omega,\mathbf{x}}(\mathbb{R}^m)$$

by Theorem 2.29 and

$$\begin{aligned}
\left| \int_s^t \mu(\tau, Y_\tau)d\tau \right| &\le |t-s| \, \|\mu(\cdot, Y_\cdot)\|_\infty \\
&\le |t-s| \left(\|\mu(\cdot, Y_\cdot) - \mu(\cdot, 0)\|_\infty + \|\mu(\cdot, 0)\|_\infty \right) \\
&\le |t-s| \left(L_{\mu\text{-Lip}} \|Y_t\|_\infty + \|\mu(\cdot, 0)\|_\infty \right) \\
&\le \left((1 + L_{\mu\text{-Lip}}) \left(\|Y_0\| + 2 \|Y, Y'\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}} \omega^{\frac{1}{p}}(0,T) \right) + \|\mu(\cdot, 0)\|_\infty \right) |t-s| \\
&\le \left((1 + L_{\mu\text{-Lip}}) \left(\|Y_0\| + 2 \|Y, Y'\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}} \omega^{\frac{1}{p}}(0,T) \right) + \|\mu(\cdot, 0)\|_\infty \right) \omega(s,t)
\end{aligned}$$

using Equation (10). Again, a solution in $\mathcal{D}^\gamma_{\omega,\mathbf{x}}$ for $1 \geq \gamma > p-2$ is also in $\mathcal{D}^1_{\omega,\mathbf{x}}$, by the same calculation as before (in Theorem 3.3). This time we consider

$$\mathcal{B}_\xi := \left\{(Y,Y') \in \mathcal{D}^\gamma_{\omega,\mathbf{x}} \middle| Y_0 = \xi, Y'_0 = f(\xi)\right\}$$

and the ball $\mathbb{D}_\xi$ centered in $t \mapsto c(t) = \left(\xi + f(\xi)\mathbf{x}^1_{0,t}, f(\xi)\right)$.
For the invariance step, we gain the additional term

$$\left|\int_s^t \mu(\tau, Y_\tau)d\tau\right| \leq C_{\mu,Y_0,\|Y,Y'\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}},\gamma}\omega(s,t) \leq C_{\mu,Y_0,\|Y,Y'\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}},\gamma}\omega^{1-\frac{1+\gamma}{p}}(0,T_0)\omega^{\frac{1+\gamma}{p}}(s,t)$$

from the triangle inequality. This can be made arbitrarily small for $(Y,Y') \in \mathbb{D}_\xi$ by choosing $T_0 > 0$ small enough.
In the contraction step, we again have the additional term

$$\begin{aligned}\left|\int_s^t \mu(\tau,Y_\tau) - \mu(\tau,\tilde{Y}_\tau)d\tau\right| \leq & \left|t-s\right| L_{\mu\text{-Lip}} \left\|Y_\tau - \tilde{Y}_\tau\right\|_\infty \\ \leq & L_{\mu\text{-Lip}}\Bigg(2\underbrace{\left\|Y_0 - \tilde{Y}_0\right\|_\infty}_{=0} \\ & \qquad + 2\left\|Y-\tilde{Y}, Y'-\tilde{Y}'\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}\omega^{\frac{\gamma}{p}}(0,T)\Bigg)\omega^{\frac{1}{p}}(s,t)\left|t-s\right| \\ \leq & C_{\omega,\gamma,p}L_{\mu\text{-Lip}}\left\|Y-\tilde{Y}, Y'-\tilde{Y}'\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}\omega^{\frac{\gamma+p}{p}}(s,t).\end{aligned}$$

The rest of the proof is the same as in Theorem 3.3. The extension to the full interval follows analogous to Corollary 3.4. □

Remark 3.7:
The solution constructed in this way (Theorem 3.6) then is **different** to a solution of the RDE

$$d\hat{Y}_t = \hat{f}(\hat{Y}_t)d\hat{\mathbf{x}}_t$$

from above in the sense of Theorem 3.3 and Corollary 3.4, since in that case we would have $\hat{Y}'_t = \hat{f}(\hat{Y}_t) \neq f(Y_t)$.

3.3 Bounds on RDE Solutions

To prove the main result, later on, we also need the following lemma. This will allow us to find uniform bounds on a set of RDE solutions. The proof is lengthy and complicated, but we didn't want to omit it for the sake of completeness.

Theorem 3.8:
Let $\mu : [0,T] \times \mathbb{R}^m \to \mathbb{R}^m$ be continuous in the first and uniformly Lipschitz in the second component, let $f \in C^3_b(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$, and let $\mathbf{x}$ be a p-rough path dominated by the controlling function ω with $\omega(s,t) \geq |t-s|$. Let Y be the unique solution to the RDE

$$dY = \mu(t, Y_t)dt + f(Y_t)d\mathbf{x}_t.$$

Then there is a constant $C\left(p, \omega(0,T), L_{a-Lip}, \|a(\cdot,0)\|_\infty, \|f\|_{C^3_b}\right) < \infty$, such that

$$\left\|(Y,Y')\right\|_{\mathcal{D}^1_{\omega,\mathbf{x}}} \leq C\left(p, \omega(0,T), L_{a-Lip}, \|a(\cdot,0)\|_\infty, \|f\|_{C^3_b}\right).$$

Also, the dependencies of C are continuous.

Something similar was shown in [FH20, Proposition 8.2], which only considered α-Hölder paths and, more crucially, only works for RDEs without a drift term, which is why our proof has to take an additional detour to first show the boundedness of $|Y_s|$, which is a term present in an estimate of the drift part.

Proof. Let Y be the unique solution of the RDE above. Recall that $Y'_t = f(Y_t)$. We then can consider

$$\begin{aligned}
\left|R^Y_{s,t}\right| &= \left|Y_{s,t} - f(Y_s)\mathbf{x}^1_{s,t}\right| \\
&\leq \left|\int_s^t f(Y)d\mathbf{x} - f(Y_s)\mathbf{x}^1_{s,t} - Df(Y_s)f(Y_s)\mathbf{x}^2_{s,t}\right| + \left|Df(Y_s)f(Y_s)\mathbf{x}^2_{s,t}\right| + \left|\int_s^t a(t,Y_t)dt\right| \\
&\leq 2^{\frac{3}{p}+1}\zeta\left(\frac{3}{p}\right)\left\|(f(Y),f(Y)')\right\|_{\mathcal{D}^1_{\omega,\mathbf{x}}}\omega^{\frac{3}{p}}(s,t) + \|Df\|_\infty\|f\|_\infty\omega^{\frac{2}{p}}(s,t) \\
&\quad + \left[(1+L_{a-Lip})\left(\|Y_s\| + 2\left\|(Y,Y')\right\|_{\mathcal{D}^1_{\omega,\mathbf{x}}}\omega^{\frac{1}{p}}(0,T)\right) + \|a(t,0)\|_\infty\right]\omega(s,t) \\
&\leq 2^{\frac{3}{p}+1}\zeta\left(\frac{3}{p}\right)\Big[C_{3.1;1}\left(\|f\|_{C^3_b}, \left\|Y'\right\|_\infty, \omega(0,T), p\right)\left\|(Y,Y')\right\|_{\mathcal{D}^1_{\omega,\mathbf{x}}} \\
&\qquad + C_{3.1;2}\left(\|f\|_{C^3_b}\right)\left\|(Y,Y')\right\|^2_{\mathcal{D}^1_{\omega,\mathbf{x}}}\omega^{\frac{2}{p}}(s,t) \\
&\qquad + C_{3.1;3}(\|f\|_{C^3_b}, \left\|Y'\right\|_\infty)\Big]\omega^{\frac{3}{p}}(s,t) + \|Df\|_\infty\|f\|_\infty\omega^{\frac{2}{p}}(s,t) \\
&\quad + \left((1+L_{a-Lip})\|Y_s\| + \|a(t,0)\|_\infty\right)\omega(s,t) \\
&\quad + 2(1+L_{a-Lip})\omega^{\frac{1}{p}}(0,T)\left\|(Y,Y')\right\|_{\mathcal{D}^1_{\omega,\mathbf{x}}}\omega(s,t).
\end{aligned}$$

For the second inequality, we used the Maximal inequality as well as an inequality we used in the proof of Theorem 3.6, while for the third inequality, we used Lemma 3.1 on $\|(f(Y),f(Y)')\|_{\mathcal{D}^1_{\omega,\mathbf{x}}}$. The constants $C_{3.1;1}\left(\|f\|_{C^3_b}, \|Y'\|_\infty, \omega(0,T), p\right), C_{3.1;2}\left(\|f\|_{C^3_b}\right)$, and $C_{3.1;3}(\|f\|_{C^3_b}, \|Y'\|_\infty)$ are taken from Lemma 3.1 and are each continuous in the values they depend on. We additionally note, that each of the inequalities only depends on the values (Y,Y') in the interval $[s,t]$, as well as $\|Y'\|_\infty$, which is bounded by $\|f\|_\infty$. In particular, we only need local estimates in $[s,t]$ when considering $\|(Y,Y')\|_{\mathcal{D}^1_{\omega,\mathbf{x}}}$. This lets us use the notation $\|(Y,Y')\|_{\mathcal{D}^1_{\omega,\mathbf{x};[s,t]}}$, which is defined in the same way as $\|(Y,Y')\|_{\mathcal{D}^1_{\omega,\mathbf{x}}}$ but restricted to the time period $[s,t]$. The simple estimate

$$\left|Y'_t - Y'_s\right| = |f(Y_t) - f(Y_s)| \leq \|Df\|_\infty|Y_t - Y_s| \leq \|Df\|_\infty\left|R^Y_{s,t}\right| + \|Df\|_\infty\|f\|_\infty\omega^{\frac{1}{p}}(s,t)$$

then gives us

$$\begin{aligned}
\left\|(Y,Y')\right\|_{\mathcal{D}^1_{\omega,\mathbf{x};[s,t]}} \leq & C_1\left(p, \|f\|_{C^3_b}, \left\|Y'\right\|_\infty, \omega(0,T)\right)\left\|(Y,Y')\right\|_{\mathcal{D}^1_{\omega,\mathbf{x};[s,t]}}\omega^{\frac{1}{p}}(s,t) \\
&+ C_2\left(\|f\|_{C^3_b}, p, \omega(0,T)\right)\left\|(Y,Y')\right\|^2_{\mathcal{D}^1_{\omega,\mathbf{x};[s,t]}}\omega(s,t) \\
&+ C_3(L_{a-Lip}, p, \omega(0,T))\|Y_s\|\omega^{1-\frac{2}{p}}(s,t) \\
&+ C_4\left(L_{a-Lip}, p, \omega(0,T), \|f\|_{C^3_b}\right)\left\|(Y,Y')\right\|_{\mathcal{D}^1_{\omega,\mathbf{x};[s,t]}}\omega^{1-\frac{2}{p}}(s,t) \\
&+ C_5\left(p, \|f\|_{C^3_b}, \left\|Y'\right\|_\infty, \omega(0,T), \|a(t,0)\|_\infty\right),
\end{aligned}$$

where we combined all expressions in terms we do not need explicitly into the constants $C_1, ..., C_5$, each with their continuous dependencies, which we will not write down explicitly from here on out, to make the notation clearer. We will now only consider special time ranges $[s,t]$. For this, we set

$$\omega_{\max}(|Y_s|) := \min\left(1, \left(\frac{1}{3C_1}\right)^p, \left(\frac{1}{3C_4}\right)^{\frac{p}{p-2}}, \frac{1}{45C_2(C_3|Y_s|\omega^{1-\frac{2}{p}}(0,T) + C_5)}\right).$$

Note that $\omega_{\max}(|Y_s|)$ depends on $p, \|f\|_{C_b^3}, \|Y'\|_\infty, L_{a-Lip}, \omega(0,T), \|a(t,0)\|_\infty$, and $|Y_s|$ and that this dependence is continuous. We suppress the dependence on terms other then $|Y_s|$ in our notation.
Now, let $s<t$ such that $\omega(s,t) \le \omega_{\max}(|Y_s|)$. Then we have

$$\left\|(Y,Y')\right\|_{\mathcal{D}^1_{\omega,\mathbf{x};[s,t]}} \le 3C_2 \left\|(Y,Y')\right\|^2_{\mathcal{D}^1_{\omega,\mathbf{x};[s,t]}} \omega(s,t) + 3C_3 |Y_s| \omega^{1-\frac{2}{p}}(s,t) + 3C_5.$$

To get rid of the $\|(Y,Y')\|^2_{\mathcal{D}^1_{\omega,\mathbf{x};[s,t]}}$ term, we show that with our choice of $\omega_{\max}(|Y_s|)$ we actually already guarantee $3C_2 \|(Y,Y')\|_{\mathcal{D}^1_{\omega,\mathbf{x};[s,t]}} \omega(s,t) < \frac{1}{2}$: Define

$$\begin{aligned}\psi_{s,t} &:= 3C_2 \left\|(Y,Y')\right\|_{\mathcal{D}^1_{\omega,\mathbf{x};[s,t]}} \omega(s,t),\\ \lambda_{s,t} &:= 9C_2 \left(C_3 |Y_s| \omega^{1-\frac{2}{p}}(s,t) + C_5\right) \omega(s,t).\end{aligned}$$

Multiplying the inequality from above by $3C_2\omega(s,t)$ gives

$$\psi_{s,t} \le \psi_{s,t}^2 + \lambda_{s,t} \Leftrightarrow 0 \le \psi_{s,t}^2 - \psi_{s,t} + \lambda_{s,t}.$$

By our choice of $\omega_{\max}(|Y_s|)$, we have $\lambda_{s,t} \le \frac{1}{5}$ and therefore by the p-q-formula one of the following two hold:

$$\begin{aligned}\psi_{s,t} &\le \frac{1}{2} - \sqrt{\frac{1}{4} - \lambda_{s,t}} \le \frac{1}{2} - \sqrt{\frac{1}{20}} < \frac{1}{2}, \text{ or}\\ \psi_{s,t} &\ge \frac{1}{2} + \sqrt{\frac{1}{4} - \lambda_{s,t}} \ge \frac{1}{2} + \sqrt{\frac{1}{20}}.\end{aligned}$$

Since for $t \searrow s$, we have $\psi_{s,t} \to 0 < \frac{1}{2}$, we have to be in the first case for $|t-s|$ small enough, and since $\psi_{s,t}$ is continuous in s and t, we cannot jump from the first to the second case. This means we have to be in the first case for all $s<t$, such that $\omega(s,t) \le \omega_{\max}(|Y_s|)$. In particular $\psi_{s,t} \le \frac{1}{2}$, which implies

$$\psi_{s,t} \le 2\lambda_{s,t},$$

i.e.

$$\left\|(Y,Y')\right\|_{\mathcal{D}^1_{\omega,\mathbf{x};[s,t]}} \le 6(C_3 |Y_s| \omega^{1-\frac{2}{p}}(s,t) + C_5).$$

This is the result for small enough $\omega(s,t)$.
To generalize this, we first show that $|Y_s|$ is bounded: Consider again

$$\omega_{\max}(|Y_s|) = \min\left(1, \left(\frac{1}{3C_1}\right)^p, \left(\frac{1}{3C_4}\right)^{\frac{p}{p-2}}, \frac{1}{45(C_2 |Y_s| \omega^{1-\frac{2}{p}}(0,T) + C_5)}\right).$$

Then (after possibly making C_5 bigger by a term in C_1, C_4, and p)

$$\omega_{\max}(|Y_s|) = \frac{1}{45(C_2 |Y_s| \omega^{1-\frac{2}{p}}(0,T) + C_5)} \le \frac{1}{45C_2 |Y_s| \omega^{1-\frac{2}{p}}(0,T)}.$$

So it holds (for $s < t$, such that $\omega(s,t) \le \omega_{\max}(|Y_s|)$)

$$\begin{aligned}
|Y_t - Y_s| \le & \left\|(Y,Y')\right\|_{\mathcal{D}^1_{\omega,\mathbf{x};[s,t]}} \omega^{\frac{2}{p}}(s,t) + \|f\|_\infty \omega^{\frac{1}{p}}(s,t) \\
\le & 6(C_3 |Y_s| \omega^{1-\frac{2}{p}}(s,t) + C_5)\omega^{\frac{2}{p}}(s,t) + \|f\|_\infty \omega^{\frac{1}{p}}(s,t) \\
\le & 6C_3 |Y_s| \omega_{\max}(|Y_s|) + 6C_5 \omega_{\max}^{\frac{2}{p}}(|Y_s|) + \|f\|_\infty \omega_{\max}^{\frac{1}{p}}(|Y_s|) \\
\le & 6C_3 |Y_s| \left(\frac{1}{45C_2 |Y_s| \omega^{1-\frac{2}{p}}(0,T)}\right) + 6C_5 \left(\frac{1}{45(C_2 |Y_s| \omega^{1-\frac{2}{p}}(0,T) + C_5)}\right)^{\frac{1}{p}} \\
& + \|f\|_\infty \left(\frac{1}{45(C_2 |Y_s| \omega^{1-\frac{2}{p}}(0,T) + C_5)}\right)^{\frac{2}{p}} \\
\le & \frac{2C_3}{15C_5 \omega^{1-\frac{2}{p}}(0,T)} + 6C_5 \left(\frac{1}{45C_5}\right)^{\frac{2}{p}} + \|f\|_\infty \left(\frac{1}{45C_5}\right)^{\frac{1}{p}} =: C_6
\end{aligned}$$

by Equation (12). Now, we can show the boundedness of $|Y_s|$. For this set $t_0 := 0$,

$$t_{k+1} := \max\left\{\tau > t_k | \tau \le T \text{ and } \omega(t_k,\tau) \le \omega_{\max}(Y_{t_k})\right\}.$$

The above then implies

$$|Y_{t_k}| \le \left|Y_{t_{k-1}}\right| + C_6 \le ... \le |Y_0| + kC_6$$

We need to show that there is a $K \in \mathbb{N}$, such that $t_K = T$. With

$$K_0 = \left\lceil \frac{|Y_0|}{C_6} + \frac{C_5}{C_2 C_6 \omega^{1-\frac{1}{p}}}(0,T) \right\rceil,$$

i.e.

$$C_2 C_6 K_0 \omega^{1-\frac{1}{p}}(0,T) \ge C_2 |Y_0| \omega^{1-\frac{1}{p}}(0,T) + C_5.$$

For $t_k < T$, we then have

$$\begin{aligned}
\omega(0,t_k) \ge & \sum_{j=1}^{k} \omega(t_{j-1},t_j) = \sum_{j=0}^{k-1} \omega_{\max}(\left|Y_{t_j}\right|) \\
= & \sum_{j=0}^{k-1} \frac{1}{45(C_2(|Y_0| + jC_6)\omega^{1-\frac{2}{p}}(0,T) + C_5)} \ge \sum_{j=K_0}^{k-1} \frac{1}{45(C_2(|Y_0| + jC_6)\omega^{1-\frac{2}{p}}(0,T) + C_5)} \\
\ge & \sum_{j=K_0}^{k-1} \frac{1}{90C_2 C_6 \omega^{1-\frac{1}{p}}(0,T) j} \ge \frac{1}{90C_2 C_6 \omega^{1-\frac{1}{p}}(0,T)} \sum_{j=K_0}^{k-1} \frac{1}{j} \xrightarrow{k\to\infty} \infty.
\end{aligned}$$

Therefore, there is a $k \in \mathbb{N}$ such that

$$\omega(0,t_k) \ge \omega(0,T).$$

Since $t \mapsto \omega(0,t)$ is (strictly) increasing, we have $t_k = T$.
The map $\omega(0,T) \mapsto k(\omega)$ of course has jumps, but we can find C_8 continuous, such that

$$k(\omega) \le C_8(p, C_1, C_2, C_3, C_4, C_5, \omega(0,T))$$

by making it a little larger. Then we have shown

$$|Y_s| \le (|Y_0| + C_8 C_6)$$

for all $s \in [0,T]$. This then can be used to set

$$\omega_{\max} := \omega_{\max}(|Y_0| + C_8C_6)$$

independently of $|Y_s|$. Hence,

$$\begin{aligned}\left\|(Y,Y')\right\|_{\mathcal{D}^1_{\omega,\mathbf{x};[s,t]}} \le& 6(C_3\,|Y_s|\,\omega^{1-\frac{2}{p}}(s,t) + C_5)\\ \le& 6(C_3\,(|Y_0| + C_8C_6)\,\omega^{1-\frac{2}{p}}(0,T) + C_5) =: C_9\end{aligned}$$

for all $s<t$, such that $\omega(s,t) \le \omega_{\max}$.
This lets us come to general $s<t$ with $\omega(s,t) > \omega_{\max}$. Then we can again define $t_0 := s$ and

$$t_{k+1} := \max\{\tau > t_k | \tau \le t \text{ and } \omega(t_k,\tau) \le \omega_{\max}\}$$

with $t_K = T$ for $K = \left\lceil \frac{\omega(s,t)}{\omega_{\max}} \right\rceil$. It holds

$$\begin{aligned}\left|R^Y_{s,t}\right| = \left|Y_{s,t} - f(Y_s)\mathbf{x}^1_{s,t}\right| &\le \sum_{k=1}^K \left|Y_{t_{k-1},t_k} - f(Y_s)\mathbf{x}^1_{t_{k-1},t_k}\right|\\ &\le \sum_{k=1}^K \left|R^Y_{t_{k-1},t_k}\right| + \left|\left(f(Y_{t_{k-1}}) - f(Y_s)\right)\mathbf{x}^1_{t_{k-1},t_k}\right|\\ &\le \sum_{k=1}^K \left\|(Y,Y')\right\|_{\mathcal{D}^1_{\omega,\mathbf{x};[t_{k-1},t_k]}} \omega^{\frac{2}{p}}(t_{k-1},t_k) + 2\left\|f\right\|_\infty \omega^{\frac{1}{p}}(t_{k-1},t_k)\\ &\le \left\lceil \frac{\omega(s,t)}{\omega_{\max}} \right\rceil \left(C_9\omega_{\max}^{\frac{2}{p}} + 2\left\|f\right\|_\infty \omega_{\max}^{\frac{1}{p}}\right) \le \left(C_9\omega_{\max}^{\frac{2}{p}} + 2\left\|f\right\|_\infty \omega_{\max}^{\frac{1}{p}}\right)\left(\frac{\omega(s,t)}{\omega_{\max}} + 1\right)\\ &\le \left(C_9\omega_{\max}^{\frac{2}{p}} + 2\left\|f\right\|_\infty \omega_{\max}^{\frac{1}{p}}\right)\frac{2}{\omega_{\max}}\omega(s,t)\end{aligned}$$

and

$$\begin{aligned}\left|Y'_t - Y'_s\right| \le& \sum_{k=1}^K \left|Y'_{t_k} - Y'_{t_{k-1}}\right| = \sum_{k=1}^K \left|f(Y_{t_k}) - f(Y_{t_{k-1}})\right| \le \left\|Df\right\|_\infty \sum_{k=1}^K \left|Y_{t_k} - Y_{t_{k-1}}\right|\\ \le& \left\|Df\right\|_\infty \sum_{k=1}^K \left|R^Y_{t_{k-1},t_k}\right| + \left|f(Y_{t_{k-1}})\mathbf{x}^1_{t_{k-1},t_k}\right|\\ \le& \left\|Df\right\|_\infty \left(C_9\omega_{\max}^{\frac{2}{p}} + \left\|f\right\|_\infty \omega_{\max}^{\frac{1}{p}}\right)\frac{2}{\omega_{\max}}\omega(s,t)\end{aligned}$$

analogous to the above, for $\omega(s,t) > \omega_{\max}$. This finally concludes the proof. ☐

3.4 Stability of RDEs

What is still left to show is the continuous dependence of the solution Y on the control μ. For this, we need to consider the stability of the composition of controlled rough paths with functions (see Lemma 3.1). This subsection is again based on [FH20, Theorem 7.6 & Theorem 8.5] First, we prove the following little lemma:

Lemma 3.9:
Let $W, \tilde{W}$ be Banach spaces and $\varphi \in C^2_b(W,\tilde{W})$. Then there exists a constant $C = C_{M,\varphi,\omega,\gamma,p}$, such that for all $(Y,Y') \in \mathcal{D}^\gamma_{\omega,\mathbf{x}}(W)$ and $(\tilde{Y},\tilde{Y}') \in \mathcal{D}^\gamma_{\omega,\tilde{\mathbf{x}}}$ with

$$\left\|Y_0\right\|, \left\|Y'_0\right\|, \left\|(Y,Y')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}, \left\|\tilde{Y}_0\right\|, \left\|\tilde{Y}'_0\right\|, \left\|(\tilde{Y},\tilde{Y}')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}} \le M,$$

it holds

$$\left\|\varphi(Y)_{s,t} - \varphi(\tilde{Y})_{s,t}\right\| \le C\left(\left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma} + \left\|Y'_0 - \tilde{Y}'_0\right\| + \left\|Y_0 - \tilde{Y}_0\right\| + d_\omega(\mathbf{x},\tilde{\mathbf{x}})\right)\omega^{\frac{1}{p}}(s,t).$$

Proof. Similar to the proof of Theorem 3.3, we consider

$$g(x,y) := \int_0^1 D\varphi(tx + (1-t)y)dt.$$

Then

$$\varphi(x) - \varphi(y) = g(x,y)(x-y)$$

and

$$\|g\|_\infty \leq \|D\varphi\|_\infty ,$$

as well as

$$\begin{aligned}\|g(x,y) - g(\tilde{x},\tilde{y})\| &= \left\| \int_0^1 D\varphi(tx + (1-t)y) - D\varphi(t\tilde{x} + (1-t)\tilde{y})dt \right\| \\ &\leq \int_0^1 \left\|D^2\varphi\right\|_\infty (t\|x-\tilde{x}\| + (1-t)\|y-\tilde{y}\|)\,dt \\ &\leq \frac{1}{2}\left\|D^2\varphi\right\|_\infty (\|x-\tilde{x}\| + \|y-\tilde{y}\|).\end{aligned}$$

Now we can set $\Delta_t = Y_t - \tilde{Y}_t$ and see that

$$\begin{aligned}\left\|\varphi(Y)_{s,t} - \varphi(\tilde{Y})\right\| &= \left\|g(Y_t,\tilde{Y}_t)\Delta_t - g(Y_s,\tilde{Y}_s)\Delta_s\right\| \\ &= \left\|g(Y_t,\tilde{Y}_t)(\Delta_t - \Delta_s) + \left(g(Y_t,\tilde{Y}_t) - g(Y_s,\tilde{Y}_s)\right)\Delta_s\right\| \\ &\leq \|g\|_\infty \left\|Y_{s,t} - \tilde{Y}_{s,t}\right\| + \frac{1}{2}\left\|D^2\varphi\right\|_\infty \left(\|Y_{s,t}\| + \left\|\tilde{Y}_{s,t}\right\|\right)\left\|Y_s - \tilde{Y}_s\right\|.\end{aligned}$$

By Equations (10) and (12), we then get

$$\begin{aligned}\left\|Y_s - \tilde{Y}_s\right\|\left(\|Y_{s,t}\| + \left\|\tilde{Y}_{s,t}\right\|\right) \leq & C_{\omega,\gamma,p}\Bigg(\left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma} + \left\|Y_0' - \tilde{Y}_0'\right\| + \left\|Y_0 - \tilde{Y}_0\right\| \\ &+ \underbrace{\left\|\tilde{Y}_0'\right\|}_{\leq M} d_\omega(\mathbf{x},\tilde{\mathbf{x}})\Bigg)\omega^{\frac{1}{p}}(0,T)\Bigg(2\underbrace{\|(Y,Y')\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}}_{\leq M}\omega^{\frac{\gamma}{p}}(0,T) + \underbrace{\|Y_0\|}_{\leq M} \\ &+ 2\underbrace{\left\|(\tilde{Y},\tilde{Y}')\right\|_{\mathcal{D}^\gamma_{\omega,\tilde{\mathbf{x}}}}}_{\leq M}\omega^{\frac{\gamma}{p}}(0,T) + \underbrace{\left\|\tilde{Y}_0\right\|}_{\leq M}\Bigg)\omega^{\frac{1}{p}}(s,t)\end{aligned}$$

and

$$\left\|Y_{s,t} - \tilde{Y}_{s,t}\right\| \leq C_{\omega,\gamma,\varphi,p}\left(\left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma} + \left\|Y_0' - \tilde{Y}_0'\right\| + \left\|\tilde{Y}_s'\right\| d_\omega(\mathbf{x},\tilde{\mathbf{x}})\right)\omega^{\frac{1}{p}}(s,t)$$

by Equation (12). Now,

$$\left\|\tilde{Y}_s'\right\| \leq \left\|\tilde{Y}_0'\right\| + \left\|(\tilde{Y},\tilde{Y}')\right\|_{\mathcal{D}^\gamma_{\omega,\tilde{\mathbf{x}}}}\omega^{\frac{\gamma}{p}}(0,1) \leq \left(1 + \omega^{\frac{\gamma}{p}}(0,1)\right)M$$

gives the assertion. □

This directly allows us to prove:

Lemma 3.10:
Let $3 > p \geq 2$ and $1 \geq \gamma > p-2$. Let $\mathbf{x}, \tilde{\mathbf{x}} \in \Omega^p$ be p-rough paths dominated by the controlling function ω. Let $(Y, Y') \in \mathcal{D}^\gamma_{\omega,\mathbf{x}}$ and $(\tilde{Y}, \tilde{Y}') \in \mathcal{D}^\gamma_{\omega,\tilde{\mathbf{x}}}$. For $\varphi \in C^3_b$ let

$$(Z, Z') := (\varphi(Y), D\varphi(Y)Y') \in \mathcal{D}^\gamma_{\omega,\mathbf{x}}$$

and similar for $(\tilde{Z}, \tilde{Z}')$. Then it holds

$$\left\| Z, Z'; \tilde{Z}, \tilde{Z}' \right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma} \leq C \left(d_\omega(\mathbf{x}, \tilde{\mathbf{x}}) + \left| Y_0 - \tilde{Y}_0 \right| + \left| Y'_0 - \tilde{Y}'_0 \right| + \left\| Y, Y'; \tilde{Y}, \tilde{Y}' \right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma} \right)$$

for some $C = C(M, p, \omega, \varphi, \gamma)$, where

$$|Y_0|, |Y'_0|, \|(Y, Y')\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}}, \left|\tilde{Y}_0\right|, \left|\tilde{Y}'_0\right|, \left\|(\tilde{Y}, \tilde{Y}')\right\|_{\mathcal{D}^\gamma_{\omega,\mathbf{x}}} \leq M.$$

Proof. To show the result, we need the two estimates for $\left\| Z, Z'; \tilde{Z}, \tilde{Z}' \right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma}$ as defined in Section 2.5. We start with

$$\begin{aligned}
\left| Z'_{s,t} - \tilde{Z}'_{s,t} \right| &= \left| (D\varphi(Y)Y')_{s,t} - (D\varphi(\tilde{Y})\tilde{Y}')_{s,t} \right| \\
&= \left| D\varphi(Y_t)Y'_t - D\varphi(Y_s)Y'_s - \left(D\varphi(\tilde{Y}_t)\tilde{Y}'_t - D\varphi(\tilde{Y}_s)\tilde{Y}'_s \right) \right| \\
&= \left| D\varphi(Y)_{s,t}Y'_t + D\varphi(Y_s)Y'_{s,t} - \left(D\varphi(\tilde{Y})_{s,t}\tilde{Y}'_t + D\varphi(\tilde{Y}_s)\tilde{Y}'_{s,t} \right) \right| \\
&\leq \left| \left(D\varphi(Y)_{s,t} - D\varphi(\tilde{Y})_{s,t} \right) Y'_t \right| + \left| D\varphi(\tilde{Y})_{s,t}(Y'_t - \tilde{Y}'_t) \right| \\
&\quad + \left| \left(D\varphi(Y_s) - D\varphi(\tilde{Y}_s) \right) Y'_{s,t} \right| + \left| D\varphi(\tilde{Y}_s) \left(Y'_{s,t} - \tilde{Y}'_{s,t} \right) \right|.
\end{aligned}$$

Now, we can assess each of these terms individually: Using Lemma 3.9 gives

$$\begin{aligned}
\left| \left(D\varphi(Y)_{s,t} - D\varphi(\tilde{Y})_{s,t} \right) Y'_t \right| &\leq \left| D\varphi(Y)_{s,t} - D\varphi(\tilde{Y})_{s,t} \right| \underbrace{|Y'_t|}_{\leq C_{M,\gamma,\omega}} \\
&\leq C_{M,\varphi,\omega,\gamma,p} \Bigg(\left\| Y, Y'; \tilde{Y}, \tilde{Y}' \right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma} + \left| Y'_0 - \tilde{Y}'_0 \right| \\
&\quad + \left| Y_0 - \tilde{Y}_0 \right| + d_\omega(\mathbf{x}, \tilde{\mathbf{x}}) \Bigg) \omega^{\frac{1}{p}}(s,t).
\end{aligned}$$

We also have

$$\begin{aligned}
\left| D\varphi(\tilde{Y})_{s,t}(Y'_t - \tilde{Y}'_t) \right| &\leq \left| D\varphi(\tilde{Y})_{s,t} \right| \left| Y'_t - \tilde{Y}'_t \right| \leq \left\| D^2\varphi \right\|_\infty \left| \tilde{Y}_t - \tilde{Y}_s \right| \left| Y'_t - \tilde{Y}'_t \right| \\
&\leq \left\| D^2\varphi \right\|_\infty \Bigg(2 \underbrace{\left\| (\tilde{Y}, \tilde{Y}') \right\|_{\mathcal{D}^\gamma_{\omega,\tilde{\mathbf{x}}}}}_{\leq M} \omega^{\frac{\gamma}{p}}(0,T) + \underbrace{\left| \tilde{Y}_0 \right|}_{\leq M} \Bigg) \omega^{\frac{1}{p}}(s,t) \\
&\quad \cdot \left(\left| Y'_0 - \tilde{Y}'_0 \right| + \left\| Y, Y'; \tilde{Y}, \tilde{Y}' \right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma} \omega^{\frac{\gamma}{p}}(0,T) \right)
\end{aligned}$$

by Equation (10),

$$\begin{aligned}
\left| \left(D\varphi(Y_s) - D\varphi(\tilde{Y}_s) \right) Y'_{s,t} \right| &\leq \left| D\varphi(Y_s) - D\varphi(\tilde{Y}_s) \right| \left| Y'_{s,t} \right| \leq \left\| D^2\varphi \right\|_\infty \left| Y_s - \tilde{Y}_s \right| \left| Y'_{s,t} \right| \\
&\leq \left\| D^2\varphi \right\|_\infty C_{\omega,\gamma,M} \Bigg(\left\| Y, Y'; \tilde{Y}, \tilde{Y}' \right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma} \\
&\quad + \left| Y'_0 - \tilde{Y}'_0 \right| + \underbrace{\left| \tilde{Y}'_0 \right|}_{\leq M} d_\omega(\mathbf{x}, \tilde{\mathbf{x}}) \Bigg)
\end{aligned}$$

by Equation (12), and

$$\left|D\varphi(\tilde{Y}_s)\left(Y'_{s,t}-\tilde{Y}'_{s,t}\right)\right| \le \|D\varphi\|_\infty \left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma}\omega^{\frac{\gamma}{p}}(s,t),$$

which give the estimate on $\left|Z'_{s,t}-\tilde{Z}'_{s,t}\right|$. For the other term, we write

$$R^Z_{s,t}=\varphi(Y_t)-\varphi(Y_s)-D\varphi(Y_s)Y'_sX_{s,t}=\varphi(Y_t)-\varphi(Y_s)-D\varphi(Y_s)Y_{s,t}+D\varphi(Y_s)R^Y_{s,t}$$

and the same for $\tilde{Z}$, similar to the proof of Lemma 3.1. Therefore

$$\begin{aligned}\left|R^Z_{s,t}-R^{\tilde{Z}}_{s,t}\right| \le &\underbrace{\left|\varphi(Y_t)-\varphi(Y_s)-D\varphi(Y_s)Y_{s,t}-\left(\varphi(\tilde{Y}_t)-\varphi(\tilde{Y}_s)-D\varphi(\tilde{Y}_s)\tilde{Y}_{s,t}\right)\right|}_{=:|T_1|}\\ &+\underbrace{\left|D\varphi(Y_s)R^Y_{s,t}-D\varphi(\tilde{Y}_s)R^{\tilde{Y}}_{s,t}\right|}_{=:|T_2|}.\end{aligned}$$

Then, we have

$$T_1=\int_0^1\left(D^2\varphi(Y_s+\theta Y_{s,t})(Y_{s,t}\otimes Y_{s,t})-D^2\varphi(\tilde{Y}_s+\theta\tilde{Y}_{s,t})(\tilde{Y}_{s,t}\otimes\tilde{Y}_{s,t})\right)(1-\theta)d\theta,$$

since

$$\begin{aligned}\int_0^1 D^2\varphi(Y_s+\theta Y_{s,t})(Y_{s,t}\otimes Y_{s,t})(1-\theta)d\theta =&D\varphi(Y_s+\theta Y_{s,t})Y_{s,t}(1-\theta)\Big|_0^1\\ &+\int_0^1 D\varphi(Y_s+\theta Y_{s,t})Y_{s,t}d\theta\\ =&-D\varphi(Y_s)Y_{s,t}+\varphi(Y_t)-\varphi(Y_s).\end{aligned}$$

We can then use

$$\begin{aligned}\left|D^2\varphi(Y_s+\theta Y_{s,t})-D^2\varphi(\tilde{Y}_s+\theta\tilde{Y}_{s,t})\right| \le& \left\|D^3\varphi\right\|_\infty\left(\left|Y_s-\tilde{Y}_s\right|+\left|Y_{s,t}-\tilde{Y}_{s,t}\right|\right)\\ \le& C_{\varphi,\gamma,p,\omega,M}\Bigg(\left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma}+\left|Y_0-\tilde{Y}_0\right|\\ &+\left\|Y'_0-\tilde{Y}'_0\right\|+d_\omega(\mathbf{x},\tilde{\mathbf{x}})\Bigg)\end{aligned}$$

to see

$$\begin{aligned}&\left|D^2\varphi(Y_s+\theta Y_{s,t})(Y_{s,t}\otimes Y_{s,t})-D^2\varphi(\tilde{Y}_s+\theta\tilde{Y}_{s,t})(\tilde{Y}_{s,t}\otimes\tilde{Y}_{s,t})\right|\\ &\le\left|D^2\varphi(Y_s+\theta Y_{s,t})\left((Y_{s,t}\otimes Y_{s,t})-(\tilde{Y}_{s,t}\otimes\tilde{Y}_{s,t})\right)\right|\\ &\quad+\left|\left(D^2\varphi(Y_s+\theta Y_{s,t})-D^2\varphi(\tilde{Y}_s+\theta\tilde{Y}_{s,t})\right)(\tilde{Y}_{s,t}\otimes\tilde{Y}_{s,t})\right|\\ &\le\left\|D^2\varphi\right\|_\infty\left\|Y_{s,t}\otimes Y_{s,t}-\tilde{Y}_{s,t}\otimes\tilde{Y}_{s,t}\right\|.\\ &\quad+\left\|D^3\varphi\right\|_\infty\left(\left|Y_s-\tilde{Y}_s\right|+\left|Y_{s,t}-\tilde{Y}_{s,t}\right|\right)\left|\tilde{Y}_{s,t}\right|^2\end{aligned}$$

Considering

$$\begin{aligned}\left\|Y_{s,t}\otimes Y_{s,t}-\tilde{Y}_{s,t}\otimes\tilde{Y}_{s,t}\right\| \le& \left\|Y_{s,t}\otimes Y_{s,t}-Y_{s,t}\otimes\tilde{Y}_{s,t}\right\|+\left\|Y_{s,t}\otimes\tilde{Y}_{s,t}-\tilde{Y}_{s,t}\otimes\tilde{Y}_{s,t}\right\|\\ \le&\left\|Y_{s,t}\otimes(Y_{s,t}-\tilde{Y}_{s,t})\right\|+\left\|(Y_{s,t}-\tilde{Y}_{s,t})\otimes\tilde{Y}_{s,t}\right\|\\ \le&\left|Y_{s,t}-\tilde{Y}_{s,t}\right|\left(|Y_{s,t}|+\left|\tilde{Y}_{s,t}\right|\right)\end{aligned}$$

as well as Equations (10) and (12) gives

$$\begin{aligned}\Big|D^2\varphi(Y_s+\theta Y_{s,t})(Y_{s,t}\otimes Y_{s,t})-&D^2\varphi(\tilde{Y}_s+\theta\tilde{Y}_{s,t})(\tilde{Y}_{s,t}\otimes\tilde{Y}_{s,t})\Big| \\ &\le C_{\varphi,M,\omega,\gamma,p}\Bigg(\left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma}+\left|Y_0'-\tilde{Y}_0'\right| \\ &\qquad+\left|Y_0-\tilde{Y}_0\right|+d_\omega(\mathbf{x},\tilde{\mathbf{x}})\Bigg)\omega^{\frac{2}{p}}(s,t)\end{aligned}$$

and therefore

$$|T_1|\le C_{\varphi,M,\omega,\gamma,p}\Bigg(\left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma}+\left|Y_0'-\tilde{Y}_0'\right|+\left|Y_0-\tilde{Y}_0\right|+d_\omega(\mathbf{x},\tilde{\mathbf{x}})\Bigg)\omega^{\frac{2}{p}}(s,t).$$

For the second term, T_2, we have

$$\begin{aligned}|T_2|=&\left|D\varphi(Y_s)R^Y_{s,t}-D\varphi(\tilde{Y}_s)R^{\tilde{Y}}_{s,t}\right|\le\left|D\varphi(Y_s)(R^Y_{s,t}-R^{\tilde{Y}}_{s,t})\right|+\left|\Big(D\varphi(Y_s)-D\varphi(\tilde{Y}_s)\Big)\,R^{\tilde{Y}}_{s,t}\right| \\ &\le\|D\varphi\|_\infty\left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma}\omega^{\frac{1+\gamma}{p}}(s,t)+\left\|D^2\varphi\right\|_\infty\left|Y_s-\tilde{Y}_s\right|\underbrace{\left\|(\tilde{Y},\tilde{Y}')\right\|_{\mathcal{D}^\gamma_{\omega,\tilde{\mathbf{x}}}}}_{\le M}\omega^{\frac{1+\gamma}{p}}(s,t) \\ &\le C_{\varphi,\omega,p,\gamma,M}\left(\left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\tilde{\mathbf{x}},\omega,\gamma}+\left\|Y_0'-\tilde{Y}_0'\right\|+d_\omega(\mathbf{x},\tilde{\mathbf{x}})\right)\omega^{\frac{1+\gamma}{p}}(s,t)\end{aligned}$$

by Equation (12). This concludes the proof. □

Now all prerequisites are met so that we can prove our main stability result.

Theorem 3.11 (Stability of RDEs)**:**
Let $\mathbf{x}\in\Omega^p$ be a rough path. Let $f:\mathbb{R}^m\to\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)$ be C^3 and let $\mu,\tilde{\mu}:\mathbb{R}\otimes\mathbb{R}^m\to\mathbb{R}^m$ be continuous in the first and uniformly Lipschitz in the second component with the same constant $L_{\mu-Lip}$, i.e.

$$|\nu(t,y)-\nu(t,\tilde{y})|\le L_{\mu-Lip}\|y-\tilde{y}\|$$

for all $t\in[0,T]$, and for both $\nu\in\{\mu,\tilde{\mu}\}$. Define

$$\|\mu,\tilde{\mu}\|_{\infty;T,M}:=\sup_{\substack{t\in[0,T]\\ \|\zeta\|\le M}}|\mu(t,\zeta)-\tilde{\mu}(t,\zeta)|\,.$$

Let Y and $\tilde{Y}$ be solutions to

$$dY=\mu(t,Y)dt+f(Y)d\mathbf{x}\qquad\qquad Y_0=\xi$$

and

$$d\tilde{Y}=\tilde{\mu}(t,\tilde{Y})dt+f(\tilde{Y})d\mathbf{x}\qquad\qquad \tilde{Y}_0=\tilde{\xi},$$

respectively, with $\|Y\|_\infty,\left\|\tilde{Y}\right\|_\infty,|\xi|,\left|\tilde{\xi}\right|\le M$. Similar to Theorem 3.6, we take a controlling function ω with

$$\omega(s,t)\ge\mathrm{Var}^p_{p,[s,t]}(\mathbf{x})+|t-s|\,.$$

Then there is a constant $C=C(M,p,\omega,f,\gamma,L_{\mu-Lip})$ with

$$\left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\mathbf{x},\omega,\gamma}\le C\left(\|\mu,\tilde{\mu}\|_{\infty;T,M}+\left|\xi-\tilde{\xi}\right|\right).$$

The required bounds on the RDE solutions can for example be obtained from Theorem 3.8.

Proof. As a first step, we consider only some possibly small time interval $[0, T_0]$. Let

$$(\mathbf{z}, \mathbf{z}') := \left(\xi + \int_0^{\cdot} \mu(t, Y_t)dt + \int_0^{\cdot} f(Y_t)d\mathbf{x}_t, f(Y_\cdot)\right).$$

Then, we have $(Y, f(Y)) = (Y, Y') = (\mathbf{z}, \mathbf{z}')$ and analog for $\tilde{Y}$. Recall the definition of $\left\|Y, Y'; \tilde{Y}, \tilde{Y}'\right\|_{\mathbf{x},\mathbf{x},\omega,\gamma}$ (Section 2.5). It holds

$$\begin{aligned}\left\|Y, Y'; \tilde{Y}, \tilde{Y}'\right\|_{\mathbf{x},\mathbf{x},\omega,\gamma} = \|\mathbf{z}, \mathbf{z}'; \tilde{\mathbf{z}}, \tilde{\mathbf{z}}'\|_{\mathbf{x},\mathbf{x},\omega,\gamma} &\leq \left\|\int_0^{\cdot} f(Y_t)d\mathbf{x}_t, f(Y_\cdot); \int_0^{\cdot} f(\tilde{Y}_t)d\mathbf{x}_t, f(Y_\cdot)\right\|_{\mathbf{x},\mathbf{x},\omega,\gamma} \\ &+ \left\|\int_0^{\cdot} \mu(t, Y)dt, 0; \int_0^{\cdot} \tilde{\mu}(t, \tilde{Y}_t)dt, 0\right\|_{\mathbf{x},\mathbf{x},\omega,\gamma}.\end{aligned}$$

Let's consider one term at a time.
By Theorem 2.33 we have

$$\begin{aligned}\left\|\int_0^{\cdot} f(Y_t)d\mathbf{x}_t, f(Y_\cdot); \int_0^{\cdot} f(\tilde{Y}_t)d\mathbf{x}_t, f(\tilde{Y}_\cdot)\right\|_{\mathbf{x},\mathbf{x},\omega,\gamma} \leq & C\Big(\left|f(\xi) - f(\tilde{\xi})\right| + \left\|f(\xi)' - f(\tilde{\xi})'\right\| \\ &+ \omega^{\frac{\gamma}{p}}(0, T_1)\left\|f(Y), f(Y)'; f(\tilde{Y}), f(\tilde{Y})'\right\|_{\mathbf{x},\mathbf{x},\omega,\gamma}\Big)\end{aligned}$$

with C dependent on M, p, γ and ω. It also holds

$$\left|f(\xi) - f(\tilde{\xi})\right| \leq \|Df\|_\infty \left|\xi - \tilde{\xi}\right|$$

and

$$\begin{aligned}\left\|f(\xi)' - f(\tilde{\xi})'\right\| &= \left\|Df(\xi)f(\xi) - Df(\tilde{\xi})f(\tilde{\xi})\right\| \\ &\leq \left\|Df(\xi)f(\xi) - Df(\tilde{\xi})f(\xi)\right\| + \left\|Df(\tilde{\xi})f(\xi) - Df(\tilde{\xi})f(\tilde{\xi})\right\| \\ &\leq \left\|D^2 f\right\|_\infty \left|\xi - \tilde{\xi}\right| \underbrace{|f(\xi)|}_{\leq \|f\|_\infty} + \|Df\|_\infty^2 \left|\xi - \tilde{\xi}\right|\end{aligned}$$

and therefore

$$\begin{aligned}&\left\|\int_0^{\cdot} f(Y_t)d\mathbf{x}_t, f(Y_\cdot); \int_0^{\cdot} f(\tilde{Y}_t)d\mathbf{x}_t, f(\tilde{Y}_\cdot)\right\|_{\mathbf{x},\mathbf{x},\omega,\gamma} \\ &\qquad \leq C(M, p, \omega, f, \gamma)\left(\left|\xi - \tilde{\xi}\right| + \omega^{\frac{\gamma}{p}}(0, T_1)\left\|f(Y), f(Y)'; f(\tilde{Y}), f(\tilde{Y})'\right\|_{\mathbf{x},\mathbf{x},\omega,\gamma}\right).\end{aligned}$$

From Lemma 3.10, we get

$$\begin{aligned}&\left\|f(Y), f(Y)'; f(\tilde{Y}), f(\tilde{Y})'\right\|_{\mathbf{x},\mathbf{x},\omega,\gamma} \\ &\qquad \leq C(M, p, \omega, \gamma, f)\Big(\left|\xi - \tilde{\xi}\right| + \underbrace{\left\|Y_0' - \tilde{Y}_0'\right\|}_{=\left|f(\xi) - f(\tilde{\xi})\right|} + \left\|Y, Y'; \tilde{Y}, \tilde{Y}'\right\|_{\mathbf{x},\mathbf{x},\omega,\gamma}\Big) \\ &\qquad \leq (1 + \|Df\|_\infty)C(M, p, \omega, \gamma, f)\left(\left|\xi - \tilde{\xi}\right| + \left\|Y, Y'; \tilde{Y}, \tilde{Y}'\right\|_{\mathbf{x},\mathbf{x},\omega,\gamma}\right),\end{aligned}$$

so overall

$$\left\| \int_0^\cdot f(Y_t) d\mathbf{x}_t, f(Y_\cdot); \int_0^\cdot f(\tilde{Y}_t) d\mathbf{x}_t, f(\tilde{Y}_\cdot) \right\|_{\mathbf{x},\mathbf{x},\omega,\gamma}$$
$$\leq C(M,p,\omega,\gamma,f) \left(\left| \xi - \tilde{\xi} \right| + \omega^{\frac{\gamma}{p}}(0,T_1) \left\| Y, Y'; \tilde{Y}, \tilde{Y}' \right\|_{\mathbf{x},\mathbf{x},\omega,\gamma} \right).$$

For $[s,t] \subset [0,T_0]$, we have

$$\left| \int_s^t \mu(\tau, Y_\tau) d\tau - \int_s^t \tilde{\mu}(\tau, \tilde{Y}_\tau) d\tau \right|$$
$$\leq \underbrace{|t-s|}_{\leq \omega(s,t)} \left\| \mu(\cdot, Y_\cdot) - \tilde{\mu}(\cdot, \tilde{Y}_\cdot) \right\|_{\infty;[s,t]}$$
$$\leq \omega(s,t) \left(\| \mu(\cdot, Y_\cdot) - \tilde{\mu}(\cdot, Y_\cdot) \|_{\infty;[s,t]} + \left\| \tilde{\mu}(\cdot, Y_\cdot) - \tilde{\mu}(\cdot, \tilde{Y}_\cdot) \right\|_{\infty;[s,t]} \right)$$
$$\leq \omega(s,t) \left(\| \mu, \tilde{\mu} \|_{\infty;1,M} + L_{\mu-Lip} \left(\left\| Y_{0,\cdot} - \tilde{Y}_{0,\cdot} \right\|_{\infty;[s,t]} + \left| \xi - \tilde{\xi} \right| \right) \right).$$

Using Equation (12), we get

$$\left\| Y_{0,\cdot} - \tilde{Y}_{0,\cdot} \right\|_{\infty;[s,t]} \leq \omega^{\frac{1}{p}}(s,t) \left(2\omega^{\frac{\gamma}{p}}(0,T_1) \left\| Y, Y'; \tilde{Y}, \tilde{Y}' \right\|_{\mathbf{x},\mathbf{x},\omega,\gamma} + \left| Y'_0 - \tilde{Y}'_0 \right| \right)$$

and additionally

$$\left| Y'_0 - \tilde{Y}'_0 \right| = \left| f(Y_0) - f(\tilde{Y}_0) \right| \leq \| Df \|_\infty \left| \xi - \tilde{\xi} \right|.$$

Therefore

$$\left\| \int_0^\cdot \mu(t, Y) dt, 0; \int_0^\cdot \tilde{\mu}(t, \tilde{Y}_t) dt, 0 \right\|_{\mathbf{x},\mathbf{x},\omega,\gamma} \leq C_{\omega,f,L_{\mu-Lip},\gamma,p} \Big(\| \mu, \tilde{\mu} \|_{\infty;1,M} + \left| \xi - \tilde{\xi} \right|$$
$$+ \omega^{\frac{\gamma}{p}}(0,T_1) \left\| Y, Y'; \tilde{Y}, \tilde{Y}' \right\|_{\mathbf{x},\mathbf{x},\omega,\gamma} \Big)$$

on $[0,T_1]$.
Overall we have shown that there exists a $C = C(M,p,\omega,\gamma,f,L_{\mu-Lip})$, such that

$$\left\| Y, Y'; \tilde{Y}, \tilde{Y}' \right\|_{\mathbf{x},\mathbf{x},\omega,\gamma;[0,T_1]} \leq C \left(\| \mu, \tilde{\mu} \|_{\infty;T,M} + \left| \xi - \tilde{\xi} \right| + \omega^{\frac{\gamma}{p}}(0,T_1) \left\| Y, Y'; \tilde{Y}, \tilde{Y}' \right\|_{\mathbf{x},\mathbf{x},\omega,\gamma} \right).$$

Since ω is uniformly continuous by the definition of a controlling function (Definition 2.7), there exists some $0 < T_1 \leq T$, such that

$$\omega(0,T_1) \leq \inf = \omega_{\max} = \inf \left[\left(\frac{1}{2C} \right)^{\frac{p}{\gamma}}, \left(\frac{1}{2} \right)^{\frac{p}{\gamma}}, \left(\frac{1}{2} \right)^{p} \right].$$

Now, if we only consider the interval $[0,T_1]$, we get

$$\left\| Y, Y'; \tilde{Y}, \tilde{Y}' \right\|_{\mathbf{x},\mathbf{x},\omega,\gamma;[0,T_1]} \leq 2C \left(\| \mu, \tilde{\mu} \|_{\infty;T,M} + \left| \xi - \tilde{\xi} \right| \right).$$

To get this estimate to the full interval $[0,T]$, we just need to iterate it. By Equation (12) we have

$$\left| Y_{T_1} - \tilde{Y}_{T_1} \right| \leq \left| Y_{0,T_1} - \tilde{Y}_{0,T_1} \right| + \left| \xi - \tilde{\xi} \right|$$
$$\leq \tilde{C} \left(\left\| Y, Y'; \tilde{Y}, \tilde{Y}' \right\|_{\mathbf{x},\mathbf{x},\omega,\gamma;[0,T_1]} + \| Df \|_\infty \left| \xi - \tilde{\xi} \right| \right) + \left| \xi - \tilde{\xi} \right|$$
$$\leq \left(2C\tilde{C} + 1 + \tilde{C} \| Df \|_\infty \right) \left(\| \mu, \tilde{\mu} \|_{\infty;T,M} + \left| \xi - \tilde{\xi} \right| \right).$$

Now starting the above process at time T_1 gives

$$\left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\mathbf{x},\omega,\gamma;[T_1,T_2]} \le \underbrace{2\left(2C\tilde{C}+1+\tilde{C}\left\|Df\right\|_\infty\right)^2}_{=:C_1}\left(\|\mu,\tilde{\mu}\|_{\infty;T,M}+\left|\xi-\tilde{\xi}\right|\right)$$

on the interval $[T_1,T_2]$ with $\omega(T_1,T_2)\le\omega_{\max}$. We can iterate this up to $k:=\lceil\frac{\omega(0,T)}{\omega_{\max}}\rceil$ to get

$$\left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\mathbf{x},\omega,\gamma;[T_{k-1},T_k=T]} \le \underbrace{2^{k-1}\left(2C\tilde{C}+1+\tilde{C}\left\|Df\right\|_\infty\right)^k}_{=:C_k}\left(\|\mu,\tilde{\mu}\|_{\infty;T,M}+\left|\xi-\tilde{\xi}\right|\right).$$

So overall

$$\left\|Y,Y';\tilde{Y},\tilde{Y}'\right\|_{\mathbf{x},\mathbf{x},\omega,\gamma;[s,t]} \le \max_{j=0,\dots,k-1} C_j\left(\|\mu,\tilde{\mu}\|_{\infty;T,M}+\left|\xi-\tilde{\xi}\right|\right)$$

for all $s<t$, such that $\omega(s,t)\le\omega_{\max}$. The extension to $\omega(s,t)>\omega_{\max}$ then follows analog to the (second) extension to $[0,T]$ in the proof of Theorem 3.8. □

4 The Signature

This section is a collection of statements on the signature, which we will need in Section 5 to prove the main statement of this thesis and solve an optimal stochastic control problem. While it is based on [KLA20, Section 2], we have greatly expanded on the proofs.

Definition 4.1 (Signature)**:**
Let $1 \leq p < 3$ and $\mathbf{x} \in \Omega^p(\mathbb{R}^n)$ be a p-rough path. Then the *signature* of $\mathbf{x}$ is defined to be

$$\mathbb{X}_{s,t}^{<\infty} := 1 + \underbrace{\int_s^t d\mathbf{x}_\tau}_{=:\mathbb{X}_{s,t}^1} + \underbrace{\int_{s \leq t_1 \leq t_2 \leq t} d\mathbf{x}_{t_1} \otimes d\mathbf{x}_{t_2}}_{=:\mathbb{X}_{s,t}^2} + ... \in T_\infty(\mathbb{R}^n).$$

We defined these integrals in Remark 2.14 and Remark 2.32. The *truncated signature (up to level k)* of $\mathbf{x}$ is

$$\mathbb{X}_{s,t}^{\leq k} := 1 + \mathbb{X}_{s,t}^1 + ... + \mathbb{X}_{s,t}^k \in T_k(\mathbb{R}^n);$$

in particular $\mathbf{x} = \mathbb{X}^{\leq \lfloor p \rfloor}$.

Remark 4.2:
Using the definition, Remark 2.14 and Remark 2.32, we see that the signature $\mathbb{X}_{s,t}^{<\infty} = 1 + \mathbb{X}_{s,t}^1 + \mathbb{X}_{s,t}^2 + ...$ of the rough path $\mathbf{x} \in \Omega^p(\mathbb{R}^n)$ is the unique solution to the RDE

$$d\mathbb{X}_{s,\cdot}^{<\infty} = \mathbb{X}_{s,\cdot}^{<\infty} \otimes d\mathbf{x}.$$

started at $\mathbb{X}_{s,s}^{<\infty} = 1 \in \mathbb{R}$; or, for the individual tensor levels of the signature

$$d\mathbb{X}_{s,\cdot}^k = \mathbb{X}_{s,\cdot}^{k-1} \otimes d\mathbf{x}.$$

started at $\mathbb{X}_{s,s}^k = \delta_{k,0}$, where δ is the usual Kronecker delta function.

We can extend the logarithm to the space $T_\infty(\mathbb{R}^n)$, using its series representation

$$\log(1+x) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{x^k}{k}.$$

Let $1 + \mathbf{x} = 1 + \mathbf{x}^1 + \mathbf{x}^2 + ... \in T_\infty(\mathbb{R}^n)$, then its logarithm $\log : T_\infty(\mathbb{R}^n) \to W_\infty(\mathbb{R}^n)$ is defined to be

$$\log(1+\mathbf{x}) := \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\mathbf{x}^{\otimes k}}{k}.$$

There is no issue with convergence here, since for each tensor level, there are only finitely many summands.
Similarly, we can extend the exponential to $W_\infty(\mathbb{R}^n)$, by setting

$$\mathbf{x}^0 + \mathbf{x}^1 + \mathbf{x}^2 + ... = \mathbf{x} \mapsto \exp(\mathbf{x}) := \exp(\mathbf{x}^0) \sum_{n=0}^{\infty} \frac{(\mathbf{x} - \mathbf{x}^0)^{\otimes n}}{n!}.$$

Note that by pulling out the real part $\mathbf{x}^0$, there is no problem with convergence of the remaining power series, since all parts of $\mathbf{x} - \mathbf{x}^0$ are of tensor level 1 or higher. Therefore, for each tensor level of the series, there are only finitely many summands.

Example 4.3 (Expected Signature of Brownian motion)**:**
Let $\mathbf{B}_t^{\text{Strat}}$ be the Stratonovich version of the $\mathbb{R}^n$-valued Brownian motion rough path from Remark 2.31 and let $\mathbb{B}_{s,t}^{<\infty}$ be its signature. Then

$$\mathbb{E}[\mathbb{B}_{0,t}^{<\infty}] = \exp\left(\frac{t}{2}\sum_{j=1}^{n} e^{(j)} \otimes e^{(j)}\right).$$

This result goes back to [Faw03]; see [FH20, Thm. 3.9] for a proof.

Later on, we will make use of a specific property of products of linear maps of the signature of geometric rough paths. For this, let us introduce shuffling.

Definition 4.4 (Words & Shuffle Product)**:**
Let

$$\mathcal{W}_n := \{i_1...i_k | k \in \mathbb{N}_0 \text{ and } i_\ell \in \{1, ..., n\} \text{ for all } \ell = 1, ..., k\}$$

be the set of *words over the alphabet* $\{1, ..., n\}$. Let $\varnothing$ denote the empty word of length zero. For $i_1...i_k \in \mathcal{W}_n$, we define a linear functional on $W_\infty(\mathbb{R}^n)$ by

$$\langle i_1...i_k, \mathbb{X}\rangle := (\mathbb{X}^k)_{i_1,...,i_k}$$

for $\mathbb{X} \in W_\infty(\mathbb{R}^n)$, where $\mathbb{X}^k$ is the component of $\mathbb{X}$ in $(\mathbb{R}^n)^{\otimes k}$ and we evaluate its $e^{(i_1)} \otimes ... \otimes e^{(i_k)}$ entry in basis notation. This definition extends linearly to linear combinations of words. The length of a word $\ell = i_1...i_k \in \mathcal{W}_n$ is

$$|\ell| := k,$$

while the length of a linear combination of words is defined to be the length of its longest word. We can now define the bilinear *shuffle product* $\sqcup\!\sqcup$ on linear combinations of words by setting

$$ui \sqcup\!\sqcup vj := (ui \sqcup\!\sqcup v)j + (u \sqcup\!\sqcup vj)i$$

for words $u, v \in \mathcal{W}_n$ and letters $i, j \in \{1, ..., n\}$ and $\varnothing \sqcup\!\sqcup u = u \sqcup\!\sqcup \varnothing = u$. This also extends bilinearly to linear combinations of words.

Intuitively, the shuffle product of two words is all the possible ways two piles of cards can be shuffled by a riffle shuffle. For two words $i_1...i_k, j_1...j_\ell \in \mathcal{W}_n$ the shuffle product are all the words made up of exactly the letters $i_1, ..., i_k, j_1, ..., j_\ell$ where i_1 comes before i_2 and j_1 comes before j_2 and so on.
Now, for piecewise C^∞ functions, we prove the shuffle identity:

Lemma 4.5:
Let $x : [0, T] \to \mathbb{R}^n$ be a smooth (piecewise C^∞) path. Let $\mathbb{X}^{<\infty} = 1 + \mathbb{X}^1 + \mathbb{X}^2 + ...$ be the signature of x, i.e.

$$\mathbb{X}_{s,t}^k = \int_{s \le t_1 \le ... \le t_k \le t} \dot{x}_{t_1} \otimes ... \otimes \dot{x}_{t_k} dt_1...dt_k$$

as defined by the usual path integral. Let $u, v \in \mathcal{W}_n$ be two words. Then we have

$$\langle u, \mathbb{X}_{s,t}^{<\infty}\rangle\langle v, \mathbb{X}_{s,t}^{<\infty}\rangle = \langle u \sqcup\!\sqcup v, \mathbb{X}_{s,t}^{<\infty}\rangle$$

for any $0 \le s \le t \le T$.

Proof. We will prove this by induction on the combined length $(|u| + |v|)$ of the words u and v. If $u = \varnothing$ or $v = \varnothing$, the assertion holds, since $\langle \varnothing, \mathbb{X}_{s,t}^{<\infty} \rangle = 1$ (the $(\mathbb{R}^n)^{\otimes 0} = \mathbb{R}$ component of the signature is 1) and $\varnothing \sqcup\!\sqcup\, u = u$. Now, let $i, j \in \{1, ..., n\}$ be two letters. Then

$$\begin{aligned}\langle i, \mathbb{X}_{s,t}^{<\infty} \rangle \langle j, \mathbb{X}_{s,t}^{<\infty} \rangle = \int_s^t \dot{x}_{t_i}^i dt_i \int_s^t \dot{x}_{t_j}^j dt_j &= \int_{\substack{s \le t_i \le t \\ s \le t_j \le t}} \dot{x}_{t_i}^i \dot{x}_{t_j}^j dt_i dt_j \\ &= \int_{s \le t_i \le t_j \le t} \dot{x}_{t_i}^i \dot{x}_{t_j}^j dt_i dt_j + \int_{s \le t_j \le t_i \le t} \dot{x}_{t_i}^i \dot{x}_{t_j}^j dt_i dt_j \\ &= \langle ij, \mathbb{X}_{s,t}^{<\infty} \rangle + \langle ji, \mathbb{X}_{s,t}^{<\infty} \rangle = \langle i \sqcup\!\sqcup\, j, \mathbb{X}_{s,t}^{<\infty} \rangle.\end{aligned}$$

We have shown the assertion for all u, v of combined length two or less. Now let $u, v \in \mathcal{W}_n$ be two words and assume the assertion holds for all words of combined length $1 + |u| + |v|$. Let $i, j \in \{1, ..., n\}$ be two letters. Then we have

$$\begin{aligned}\langle ui, \mathbb{X}_{s,t}^{<\infty} \rangle \langle vj, \mathbb{X}_{s,t}^{<\infty} \rangle &= \int_s^t \langle u, \mathbb{X}_{s,t_i}^{<\infty} \rangle \dot{x}_{t_i}^i dt_i \int_s^t \langle v, \mathbb{X}_{s,t_j}^{<\infty} \rangle \dot{x}_{t_j}^j dt_j \\ &= \int_{\substack{s \le t_i \le t \\ s \le t_j \le t}} \langle u, \mathbb{X}_{s,t_i}^{<\infty} \rangle \dot{x}_{t_i}^i \langle v, \mathbb{X}_{s,t_j}^{<\infty} \rangle \dot{x}_{t_j}^j dt_i dt_j \\ &= \int_{s \le t_i \le t_j \le t} \langle u, \mathbb{X}_{s,t_i}^{<\infty} \rangle \dot{x}_{t_i}^i dt_i \langle v, \mathbb{X}_{s,t_j}^{<\infty} \rangle \dot{x}_{t_j}^j dt_j \\ &\quad + \int_{s \le t_j \le t_i \le t} \langle v, \mathbb{X}_{s,t_j}^{<\infty} \rangle \dot{x}_{t_j}^j dt_j \langle u, \mathbb{X}_{s,t_i}^{<\infty} \rangle \dot{x}_{t_i}^i dt_i \\ &= \int_s^t \langle ui, \mathbb{X}_{s,t_j}^{<\infty} \rangle \langle v, \mathbb{X}_{s,t_j}^{<\infty} \rangle \dot{x}_{t_j}^j dt_j + \int_s^t \langle vj, \mathbb{X}_{s,t_i}^{<\infty} \rangle \langle u, \mathbb{X}_{s,t_i}^{<\infty} \rangle \dot{x}_{t_i}^i dt_i \\ &= \int_s^t \langle ui \sqcup\!\sqcup\, v, \mathbb{X}_{s,t_j}^{<\infty} \rangle \dot{x}_{t_j}^j dt_j + \int_s^t \langle u \sqcup\!\sqcup\, vj, \mathbb{X}_{s,t_i}^{<\infty} \rangle \dot{x}_{t_i}^i dt_i \\ &= \langle (ui \sqcup\!\sqcup\, v)j, \mathbb{X}_{s,t}^{<\infty} \rangle + \langle (u \sqcup\!\sqcup\, vj)i, \mathbb{X}_{s,t}^{<\infty} \rangle = \langle ui \sqcup\!\sqcup\, vj, \mathbb{X}_{s,t}^{<\infty} \rangle.\end{aligned}$$

□

Now, to be able to apply Lemma 4.5 to geometric rough paths, we still need to show that when paths converge in p-variation, so do their signatures.

Theorem 4.6:
Let $p \geq 1$. Let $\mathbf{x} \in \Omega^p$ be a p-rough path and let $(\mathbf{x}_n) \subset \Omega^p$ be a sequence of p-rough paths, such that $d_{p-var}(\mathbf{x}, \mathbf{x}_n) \xrightarrow{n\to\infty} 0$. Let $\mathbf{x}$ and all the $\mathbf{x}_n$ be dominated by the controlling function ω. Then

$$\left\| \mathbb{X}_{s,t}^k, \mathbb{X}_{s,t}^{k-1}; (\mathbb{X}_n^k)_{s,t}, (\mathbb{X}_n^{k-1})_{s,t} \right\|_{\mathbf{x},\mathbf{x}_n,\omega,\gamma} \xrightarrow{n\to\infty} 0$$

for every $0 < \gamma < 1$ and $k \in \mathbb{N}$. Here $\mathbb{X}^k$ and $(\mathbb{X}_n^k)$ are the k-th level of the signature of $\mathbf{x}$ and $\mathbf{x}_n$ respectively.

Proof. We proceed by induction and only consider the case $3 > p \geq 2$. First let $k = 1$. We have

$$\mathbb{X}_{s,t}^1 = \mathbf{x}_{s,t}^1 = \int_s^t 1 d\mathbf{x}$$

with $(1, 0) \in \mathcal{D}_{\omega,\mathbf{x}}^1$ (this is true for every rough path $\mathbf{x}$). By Corollary 2.35, for every $1 > \gamma > 0$,

$$\left\| \mathbb{X}_{s,t}^1, 1; (\mathbb{X}_n)_{s,t}^1, 1 \right\|_{\mathbf{x},\mathbf{x}_n,\omega,\gamma} \leq C d_{p-var}^{1-\gamma}(\mathbf{x}, \mathbf{x}_n) \xrightarrow{n\to\infty} 0.$$

Now, let $k \in \mathbb{N}$ be arbitrary and let the assertion hold for $k-1$. Again by Corollary 2.35, we have

$$\left\|\mathbb{X}^k_{s,t}, \mathbb{X}^{k-1}_{s,t}; (\mathbb{X}_n)^k_{s,t}, (\mathbb{X}_n)^{k-1}_{s,t}\right\|_{\mathbf{x},\mathbf{x}_n,\omega,\gamma} \leq C\Bigg(d^{\delta-\gamma}_{p-var}(\mathbf{x},\mathbf{x}_n) + \omega^{\frac{\gamma}{p}}(0,T)\left\|\mathbb{X}^{k-1}_{s,t}, \mathbb{X}^{k-2}_{s,t}; (\mathbb{X}_n)^{k-1}_{s,t}, (\mathbb{X}_n)^{k-2}_{s,t}\right\|_{\mathbf{x},\mathbf{x}_n,\omega,\delta}\Bigg)$$

for every $1 > \delta > \gamma > 0$. By assumption, this goes to zero as $n \to \infty$. □

Corollary 4.7:
Let $\mathbf{x} \in G\Omega^p$ be a geometric rough path. Then it holds

$$\langle u, \mathbb{X}^{<\infty}_{s,t}\rangle\langle v, \mathbb{X}^{<\infty}_{s,t}\rangle = \langle u \sqcup\!\sqcup v, \mathbb{X}^{<\infty}_{s,t}\rangle$$

for every $0 \leq s \leq t \leq 1$ and all words $u, v \in \mathcal{W}_n$.

Proof. As $\mathbf{x}$ is a geometric rough path, there is a sequence $(\mathbf{x}_n) \subset \Omega^p$, that converge to $\mathbf{x}$ in p-variation. For these, the shuffle identity holds by Lemma 4.5. Let

$$\omega(s,t) := \mathrm{Var}^p_{p,[s,t]}(\mathbf{x}) + \sum_{n=1}^{\infty} d^p_{p-var;[s,t]}(\mathbf{x},\mathbf{x}_n).$$

After possibly looking at a subsequence of $(\mathbf{x}_n)$, such that

$$d_{p-var}(\mathbf{x},\mathbf{x}_n)^p \leq \left(\frac{1}{2}\right)^n$$

for all $n \in \mathbb{N}$, ω is a finite controlling function, which dominates both $\mathbf{x}$ and $\mathbf{x}_n$ for every $n \in \mathbb{N}$. By Theorem 4.6 and Equation (12), we have

$$\left|\mathbb{X}^k_{s,t} - (\mathbb{X}_n)^k_{s,t}\right| \xrightarrow{n\to\infty} 0$$

for all $k \in \mathbb{N}$. As the map $\mathbb{X}^{<\infty}_{s,t} \mapsto \langle \ell, \mathbb{X}^{<\infty}_{s,t}\rangle$ for a linear combination of words ℓ is continuous and only depends on the truncated signature $\mathbb{X}^{\leq|\ell|}_{s,t}$ up to the length of ℓ, it holds

$$\begin{aligned}
\langle u, \mathbb{X}^{<\infty}_{s,t}\rangle\langle v, \mathbb{X}^{<\infty}_{s,t}\rangle &= \langle u, \mathbb{X}^{\leq|u|}_{s,t}\rangle\langle v, \mathbb{X}^{\leq|v|}_{s,t}\rangle = \lim_{n\to\infty}\langle u, (\mathbb{X}^{\leq|u|}_n)_{s,t}\rangle\langle v, (\mathbb{X}^{\leq|v|}_n)_{s,t}\rangle \\
&= \lim_{n\to\infty}\langle u, (\mathbb{X}^{<\infty}_n)_{s,t}\rangle\langle v, (\mathbb{X}^{<\infty}_n)_{s,t}\rangle = \lim_{n\to\infty}\langle u \sqcup\!\sqcup v, (\mathbb{X}^{<\infty}_n)_{s,t}\rangle \\
&= \lim_{n\to\infty}\langle u \sqcup\!\sqcup v, (\mathbb{X}^{\leq|u|+|v|}_n)_{s,t}\rangle = \langle u \sqcup\!\sqcup v, \mathbb{X}^{\leq|u|+|v|}_{s,t}\rangle = \langle u \sqcup\!\sqcup v, \mathbb{X}^{<\infty}_{s,t}\rangle,
\end{aligned}$$

since the first $|u+v|$ levels of the signature converge uniformly. □

Remark 4.8:
Brownian motion, with the Itô integral to define the first iterated integral, is *not* a geometric rough path, since we have

$$\langle 1, \mathbf{B}^{<\infty}_{0,t}\rangle\langle 1, \mathbf{B}^{<\infty}_{0,t}\rangle = B_t^2 \neq B_t^2 - t = 2\langle 11, \mathbf{B}^{<\infty}_{0,t}\rangle = \langle 1 \sqcup\!\sqcup 1, \mathbf{B}^{<\infty}_{0,t}\rangle,$$

where $\mathbf{B}^{<\infty}$ is the signature of the rough path $1 + B_t + \int_0^t B_\tau \otimes dB_\tau$.

Remark 4.9:
The signature of a geometric rough path is a very important object, since it determines the path itself up to tree-like extensions [KLA20, Lemma 2.12]. A path $x : [0,T] \to \mathbb{R}^n$ is tree-like, if there exists a function $h : [0,T] \to \mathbb{R}^+$ with $h(0) = h(T) = 0$, such that for all $0 \leq s \leq t \leq T$

$$|x_t - x_s| \leq h(t) + h(s) - 2\inf_{u\in[s,t]} h(u).$$

In particular $x_0 = x_T$. [LCL07, Definition 2.28] This means, that the signature of an augmented geometric rough path with set starting point is unique [KLA20, Corollary 2.13].

5 The Optimal Control Problem

We want to find μ_t, such that the solution Y_t^μ to the RDE

$$dY_t^\mu = \mu_t b(Y_t^\mu)dt + \sigma(Y_t^\mu)dB_t \tag{1}$$

minimizes some expected loss $\mathbb{E}[L(Y^\mu)]$. In particular, we want b to be Lipschitz, σ to be C^3, and μ to be continuous so that we can use Theorem 3.6 to guarantee the existence and uniqueness of the solution Y^μ (more on that later). We choose our set of functions μ, such that $t \mapsto \mu_t$ is a function of the path up to time t, as well as the time t itself

$$\mu_t = \Theta(\hat{B}|_{[0,t]}),$$

where $\hat{B}_t$ is augmented Brownian motion. Then μ_t also is adapted.
First, we will develop a way of comparing two paths up to different points in time.

Definition 5.1:
Let $p \geq 1$ and $t \in [0,T]$. Let

$$\hat{\Omega}_t^p = \left\{\hat{\mathbf{x}} \in G\Omega^p \middle| \mathbf{x} \text{ is a geometric rough path started at } 0 \in \mathbb{R}^n \text{ up to time } t\right\}$$

be the space of augmented geometric rough paths up to time t, so defined on $[0,t]$. We can equip $\hat{\Omega}_t^p$ with the p-variation distance (see Definition 2.20), which is a norm here, since $\hat{\mathbf{x}}_0 = 0$ for all $\hat{\mathbf{x}} \in \hat{\Omega}_t^p$.

Let $0 \leq s < t$. We can extend $\hat{\mathbf{x}} \in \hat{\Omega}_s^p$ to $\hat{\Omega}_t^p$ in the following way. We have $\hat{\mathbf{x}}_\tau^1 = (\mathbf{x}_\tau^1, \tau)$ for $\tau \in [0,s]$, which will be extended to $\tau \in [0,t]$ by setting

$$\tilde{\mathbf{x}}_\tau^1 = (\mathbf{x}_{\tau\wedge s}^1, \tau)$$

and adjusting the iterated integrals accordingly, i.e. augmenting the path $\mathbf{x}_{\tau\wedge s}^1$ on the interval $[0,t]$.
This leads us to the following definition:

Definition 5.2 (Stopped Rough Paths)**:**
Let $T > 0$. Define $\Lambda_T := \bigcup_{t\in[0,T]} \hat{\Omega}_t^p$ to be the space of *stopped rough paths (up to time T)* and equip it with the metric

$$d(\hat{\mathbf{x}}|_{[0,t]}, \hat{\mathbf{y}}|_{[0,s]}) := d_{p-var;[0,t]}(\hat{\mathbf{x}}, \tilde{\mathbf{y}}) + |t-s|$$

for $0 \leq s \leq t \leq T$ and $\hat{\mathbf{x}} \in \hat{\Omega}_t^p$, as well as $\hat{\mathbf{y}} \in \hat{\Omega}_s^p$.

Remark 5.3:
Λ_T is a polish space [Bay+22].

Definition 5.4 (Admissible Controls & Signature Controls)**:**
Let $\mathcal{T} = C(\Lambda_T, \mathbb{R})$ be the space of *admissible controls*. Additionally, we define the space of *signature controls* to be

$$\begin{aligned}\mathcal{T}_{sig} := \Big\{\Theta \in \mathcal{T} \Big| \exists \ell \text{ a linear combination of words in } \mathcal{W}_n, \text{ such that} \\ \Theta(\hat{\mathbf{x}}|_{[0,t]}) = \langle \ell, \hat{\mathbb{X}}_{0,t}^{<\infty}\rangle \text{ for all } \hat{\mathbf{x}} \in \hat{\Omega}_T^p \text{ with signature } \hat{\mathbb{X}}^{<\infty}\Big\}.\end{aligned}$$

Now the goal is to show that one can optimize over $\mu_t \in \mathcal{T}_{sig}$ instead of over $\mu_t \in \mathcal{T}$ when considering Equation (1). We show that there are compact sets $K \subset \Lambda_T$, such that $\mathcal{T}_{sig}$ is in a sense 'likely dense' in $\mathcal{T}$. Lemma 5.5 is the main idea we will use to solve the optimal control problem. It's taken from [KLA20] and was also used in [Bay+22] to prove a similar result.

Lemma 5.5:
Let $\mathbb{P}$ be a probability measure on $\left(\hat{\Omega}_T^p, \mathcal{B}(\hat{\Omega}_T^p)\right)$. Let $\varepsilon > 0$. Then there exists a compact set $K \subset \hat{\Omega}_T^p$, such that

(i) $\mathbb{P}(K) > 1 - \varepsilon$

(ii) $\mathcal{T}_{sig}$ *is dense in* $\mathcal{T}$*, restricted to* K*. In other words, for every* $\Theta \in \mathcal{T}$ *there is a sequence* $(\Theta_n) \subset \mathcal{T}_{sig}$*, such that*

$$\sup_{\hat{\mathbf{x}} \in K; t \in [0,T]} \left|\Theta(\hat{\mathbf{x}}|_{[0,t]}) - \Theta_n(\hat{\mathbf{x}}|_{[0,t]})\right| \xrightarrow{n\to\infty} 0.$$

Proof. Since Λ_T is polish (Remark 5.3), so is $\hat{\Omega}_T^p$. Since $\mathbb{P}$ is a probability measure, i.e. finite, there is a compact set $K \subset \hat{\Omega}_T^p$, such that $\mathbb{P}(K) > 1 - \varepsilon$ [Kle06, Lemma 13.5]. Let $\Theta_1, \Theta_2 \in \mathcal{T}_{sig}$. Then there exist linear combinations of words ℓ_1, ℓ_2, such that

$$\Theta_i(\hat{\mathbf{x}}|_{[0,t]}) = \langle \ell_i, \hat{\mathbb{X}}_{0,t}^{<\infty} \rangle$$

for all $\hat{\mathbf{x}} \in \Lambda_T$, $t \in [0,T]$, and $i = 1, 2$. Let $\Theta(\hat{\mathbf{x}}|_{[0,t]}) := \langle \ell_1 ⧢ \ell_2, \hat{\mathbb{X}}_{0,t}^{<\infty} \rangle$. Then $\Theta \in \mathcal{T}_{sig}$ and we have

$$\Theta_1(\hat{\mathbf{x}}|_{[0,t]})\Theta_2(\hat{\mathbf{x}}|_{[0,t]}) = \langle \ell_1, \hat{\mathbb{X}}_{0,t}^{<\infty} \rangle \langle \ell_2, \hat{\mathbb{X}}_{0,t}^{<\infty} \rangle = \langle \ell_1 ⧢ \ell_2, \hat{\mathbb{X}}_{0,t}^{<\infty} \rangle = \Theta(\hat{\mathbf{x}}|_{[0,t]})$$

by the shuffle identity (Corollary 4.7). Therefore $\mathcal{T}_{sig}$ forms an algebra. On the other hand, $\mathcal{T}_{sig}$ separates the points of $\hat{\Omega}_T^p$ by Remark 4.9. Moreover, $\mathcal{T}_{sig}$ contains the constants, since

$$\langle \varnothing, \hat{\mathbb{X}}_{0,t}^{<\infty} \rangle = 1$$

for each $\hat{\mathbf{x}} \in \hat{\Omega}_T^p$. Therefore, by the Stone-Weierstrass theorem, $\mathcal{T}_{sig}$ is dense in $\mathcal{T}$, restricted to the compact set K, with respect to uniform convergence. $\square$

Overall, we can now show the following, main result of this thesis:

Theorem 5.6:
Let $2 \leq p < 3$ and let $\mathbb{P}$ be a probability measure on $\left(\hat{\Omega}_T^p, \mathcal{B}(\hat{\Omega}_T^p)\right)$. Let Y^μ be the unique solution to

$$dY = \mu_t b(Y_t)dt + \sigma(Y_t)d\mathbf{x}$$

started at $\xi \in \mathbb{R}^m$, with $\mu \in \mathcal{T}$, b Lipschitz, and $\sigma \in C_b^3(\mathbb{R}^m, \mathbb{R}^{m\times n})$. Here, the $\mathbf{x}$ is a random geometric p-rough path with distribution determined by $\mathbb{P}$. It holds

$$\inf_{\mu \in \mathcal{T}} \mathbb{E}[L(Y^\mu)] = \inf_{\mu \in \mathcal{T}_{sig}} \mathbb{E}[L(Y^\mu)]$$

for a loss function $L : C([0,T], \mathbb{R}^m) \to \mathbb{R}$ bounded and α-Hölder for some $\alpha > 0$.

Note that here, we consider a random geometric p-rough path $\hat{\mathbf{x}}$ (chosen by $\mathbb{P}$). Also the controlling function ω is random; i.e. we can find a controlling function per path $\hat{\mathbf{x}}$:

$$\omega_{\hat{\mathbf{x}}}(s,t) := \mathrm{Var}_{p,[s,t]}^p(\hat{\mathbf{x}}) + |t - s|$$

The solution Y^μ of the RDE exists, since with $\mu_t : [0,T] \to \mathbb{R}$ continuous and b Lipschitz we have that

$$a(t,Y) := \mu_t b(Y)$$

is continuous in t and uniformly Lipschitz in Y with Lipschitz-norm $L_{a-Lip} = \|\mu\|_{\infty;[0,T]} L_{b-Lip}$, also $\sigma \in C^3$. Therefore we can use Theorem 3.6 to show the existence and uniqueness of the solution.

Proof. Let $\mu \in \mathcal{T}$ be fixed and let $\varepsilon > 0$. By Lemma 5.5 there is a compact set $K_\varepsilon \subset \hat{\Omega}^p_T$ with probability larger than $1-\varepsilon$, and we have a sequence $(\mu^{(n)}) \subset \mathcal{T}_{sig}$ such that

$$\sup_{\substack{t\in[0,T]\\ \hat{\mathbf{x}}\in K_\varepsilon}} \left|\mu_t^{(n)} - \mu_t\right| \xrightarrow{n\to\infty} 0$$

Note that by μ_t, we actually mean $\Theta(\hat{\mathbf{x}}|_{[0,t]})$ and the same fore $\mu^{(n)}$. Then we have

$$\begin{aligned}\left|\mathbb{E}\left[L(Y^\mu) - L(Y^{\mu^{(n)}})\right]\right| \leq& \left|\mathbb{E}\left[\left(L(Y^\mu) - L(Y^{\mu^{(n)}})\right)\mathbb{1}_{K_\varepsilon}\right]\right| \\ &+ \underbrace{\left|\mathbb{E}\left[\left(L(Y^\mu) - L(Y^{\mu^{(n)}})\right)(1-\mathbb{1}_{K_\varepsilon})\right]\right|}_{\leq 2\|L\|_\infty \varepsilon}.\end{aligned}$$

By Theorem 3.8, we can find $M > 0$ large enough, such that

$$M \geq \max\{\omega_{\hat{\mathbf{x}}}(0,T) | \hat{\mathbf{x}} \in K_\varepsilon\}, |\xi|, |\sigma(\xi)|, \max\{\|Y^\mu\|_\infty | \hat{\mathbf{x}} \in K_\varepsilon\}.$$

This is possible, since K_ε is a compact set and therefore the difference (in p-variation) of rough paths in K_ε is bounded. Also, $(t,\xi) \mapsto \mu_t b(\xi), \mu_t^{(n)} b(\xi)$ are both continuous in t and uniformly Lipschitz with constant $L = L_{b-\mathrm{Lip}}(\|\mu\|_\infty + \varepsilon)$ for n large enough. By Theorem 3.11

$$\left\|Y^\mu, (Y^\mu)'; Y^{\mu^{(n)}}, \left(Y^{\mu^{(n)}}\right)'\right\|_{\mathbf{x},\mathbf{x},\omega_{\hat{\mathbf{x}}},\gamma} \leq C_{\hat{\mathbf{x}}} \left\|\mu b, \mu^{(n)} b\right\|_{\infty;T,M}.$$

Since the constant $C_{\hat{\mathbf{x}}}$ in that equation depends continuously on $\omega_{\hat{\mathbf{x}}}(0,T)$ and K_ε is compact, we can find $C > 0$, such that $C_{\mathbf{x}} \leq C < \infty$ for all $\hat{\mathbf{x}} \in K_\varepsilon$. We also have

$$\left\|\mu b, \mu^{(n)} b\right\|_{\infty;1,M} = \sup_{\substack{t\in[0,T]\\ \|\xi\|<M}} \left|\mu_t b(\xi) - \mu_t^{(n)} b(\xi)\right| \leq \left(\sup_{t\in[0,T]} \left|\mu_t - \mu_t^{(n)}\right|\right) \underbrace{\left(\sup_{\|\xi\|<M} |b(\xi)|\right)}_{\leq |b(0)| + L_{b-Lip} M = C'}.$$

Then, by Equation (12)

$$\begin{aligned}\left|\mathbb{E}\left[\left(L(Y^\mu) - L(Y^{\mu^{(n)}})\right)\mathbb{1}_{K_\varepsilon}\right]\right| \leq& \mathbb{E}\left[\left|L(Y^\mu) - L(Y^{\mu^{(n)}})\right| \mathbb{1}_{K_\varepsilon}\right] \\ \leq& \|L\|_{\alpha-\text{Höl}} \mathbb{E}\left[\left\|Y^\mu - Y^{\mu^{(n)}}\right\|_\infty^\alpha \mathbb{1}_{K_\varepsilon}\right] \\ \leq& C \|L\|_{\alpha-\text{Höl}} \mathbb{E}\left[\left\|Y^\mu, (Y^\mu)'; Y^{\mu^{(n)}}, \left(Y^{\mu^{(n)}}\right)'\right\|_{\mathbf{x},\mathbf{x},\omega,\gamma}^\alpha \mathbb{1}_{K_\varepsilon}\right] \\ \leq& CC' \|L\|_{\alpha-\text{Höl}} \mathbb{E}\left[\sup_{\substack{t\in[0,T]\\ \hat{\mathbf{x}}\in K_\varepsilon}} \left|\mu_t^{(n)} - \mu_t\right|^\alpha \mathbb{1}_{K_\varepsilon}\right] \\ \leq& CC' \|L\|_{\alpha-\text{Höl}} \sup_{\substack{t\in[0,T]\\ \hat{\mathbf{x}}\in K_\varepsilon}} \left|\mu_t^{(n)} - \mu_t\right|^\alpha,\end{aligned}$$

where the constant again depends continuously on $\omega_{\hat{\mathbf{x}}}(0,T)$ and therefore can be chosen uniformly on K_ε, and

$$\left|\mathbb{E}[L(Y^{\mu^{(n)}})] - \mathbb{E}[L(Y^\mu)]\right| \leq CC' \|L\|_{\alpha-\text{Höl}} \sup_{\substack{t\in[0,T]\\ \hat{\mathbf{x}}\in K_\varepsilon}} \left|\mu_t^{(n)} - \mu_t\right|^\alpha + 2\|L\|_\infty \varepsilon \xrightarrow{n\to\infty} 2\|L\|_\infty \varepsilon.$$

As $\varepsilon > 0$ was arbitrary, this shows the existence of a sequence $\left(\mu^{(n)}\right) \subset \mathcal{T}_{sig}$, such that

$$\mathbb{E}\left[L(Y^{\mu^{(n)}})\right] \xrightarrow{n\to\infty} \mathbb{E}\left[L(Y^\mu)\right],$$

by a diagonal sequence argument. □

Now, we want to apply Theorem 5.6 to Equation (1), which is an Itô-SDE, but as noted before in Remark 4.8, the Itô integral does not give us a geometric rough path, but the Stratonovich integral does (Remark 2.23). However, Lemma 5.5 only works on geometric rough paths, as it requires the shuffle identity. That is why we need to convert the Itô-SDE from our problem to a Stratonovich-SDE.

Remark 5.7:
Let $\sigma : \mathbb{R}^m \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and let B_t be n-dimensional Brownian motion. Then it holds

$$\begin{aligned} dY_t &= \mu(Y_t, t)dt + \sigma(Y_t)dB_t \\ \Rightarrow dY_t &= \mu(Y_t, t)dt + \sigma(Y_t) \circ dB_t - \frac{1}{2}c(Y_t)dt \end{aligned}$$

with

$$c(y) = D\sigma(y)\,\sigma(y).$$

[KP92, Chapter 4.1]

Using this, we can apply Theorem 5.6 to Equation (1):

Theorem 5.8:
Let $2 \leq p < 3$ and let B_t be standard Brownian motion. Let Y^μ be the unique solution to the Itô SDE

$$dY = \mu_t b(Y_t)dt + \sigma(Y_t)dB_t$$

started at $\xi \in \mathbb{R}^m$, with $\mu \in \mathcal{T}$, b Lipschitz, and $\sigma \in C_b^3(\mathbb{R}^m, \mathbb{R}^{m\times n})$. Let $L : C([0,T], \mathbb{R}^m) \to \mathbb{R}$ be a bounded loss function and α-Hölder for some $\alpha > 0$. Then we have

$$\inf_{\mu\in\mathcal{T}} \mathbb{E}[L(Y^\mu)] = \inf_{\mu\in\mathcal{T}_{sig}} \mathbb{E}[L(Y^\mu)],$$

where the signature used in $\mathcal{T}_{sig}$ is the Stratonovich version of the Brownian motion signature.

Proof. As $\sigma \in C_b^3$, $D\sigma \in C_b^2$ and therefore our correction term c is Lipschitz. We also have $\left\|\mu b, \mu^{(n)} b\right\|_{\infty;T,M} = \left\|\mu b + \frac{1}{2}c, \mu^{(n)} b + \frac{1}{2}c\right\|_{\infty;T,M}$. Then Theorem 5.6 gives the assertion. ☐

Remark 5.9:
When considering the proof of Theorem 5.6, one notices that we even have

$$Y^{\mu^{(n)}} \xrightarrow{n\to\infty} Y^\mu$$

uniformly with probability $\mathbb{P}(K_\varepsilon) \geq 1 - \varepsilon$ for an arbitrary $\varepsilon > 0$. That means that we have

$$L(Y^{\mu^{(n)}}) \xrightarrow{n\to\infty} L(Y^\mu)$$

with probability $\mathbb{P}(K_\varepsilon) \geq 1 - \varepsilon$ for any $L \in C(C([0,T], \mathbb{R}^m), \mathbb{R})$ in the cases of Theorem 5.6 and Theorem 5.8.

6 Numerics

As part of this thesis, a python library for numerical calculations of rough paths, rough integrals and for the numerical solution of RDEs with $2 \leq p < 3$ was created. This library uses PyTorch [Pas+19], and is accessible on GitHub[2].

6.1 Rough Integrals

We have seen in Theorem 2.29, that the rough integral is the limit of Riemann like sums over partitions of an interval, when the mesh size goes to zero. This can now be used to approximate the integral by one of those sums

$$\int_s^t Y_\tau d\mathbf{x}_\tau \approx \sum_{n=0}^{N-1} Y_{t_n}\mathbf{x}^1_{t_n,t_{n+1}} + Y'_{t_n}\mathbf{x}^2_{t_n,t_{n+1}},$$

with $t_n = s + n\frac{t-s}{N}$. Now for a p-rough path $\mathbf{x}$ with $2 \leq p < 3$, a controlled path $(Y, Y') \in \mathcal{D}^\gamma_{\omega,\mathbf{x}}$ for $\gamma > p-2$, integral bounds $s < t$, and the number of steps N this yields Algorithm 1.

Algorithm 1 Approximation for Rough Integrals

Input: $\mathbf{x}$: rough path, (Y, Y') controlled path, $N \geq 1$ number of steps, $t > s > 0$ integral bounds

Output: z

$z \leftarrow 0$
$\Delta t \leftarrow \frac{t-s}{N}$
for $n \leftarrow 0, ..., N-1$ **do**
  $u \leftarrow s + n\Delta t$
  $\Delta\mathbf{x}^1 \leftarrow \mathbf{x}^1_{u+\Delta t} - \mathbf{x}^1_u$
  $\Delta\mathbf{x}^2 \leftarrow \mathbf{x}^2_{u+\Delta t} - \mathbf{x}^2_u - \mathbf{x}^1_u \otimes \Delta\mathbf{x}^1$
  $z \leftarrow z + Y_u\Delta\mathbf{x}^1 + Y'_u\Delta\mathbf{x}^2$
end for

Lemma 6.1:
For $\mathbf{x}$ a p-rough path with $2 \leq p < 3$ and $Y \in \mathcal{D}^\gamma_{\omega,\mathbf{x}}$ with a controlling function ω and $\gamma > p-2$, we have

$$\left| \int_s^t Y_\tau d\mathbf{x}_\tau - \sum_{n=0}^{N-1} Y_{t_n}\mathbf{x}^1_{t_n,t_{n+1}} + Y'_{t_n}\mathbf{x}^2_{t_n,t_{n+1}} \right| = O\left(\sup_{n=0,...,N-1} \omega^{\frac{\gamma+2}{p}-1}(t_n, t_{n+1}) \right).$$

Proof. In the construction of the integral (Theorem 2.29), we use the sewing lemma (Lemma 2.11) with $\alpha = \frac{\gamma+2}{p}$. Then step 3 of the proof of the sewing lemma gives the assertion. □

Therefore Algorithm 1 converges for $N \to \infty$ to the real solution, up to floating-point errors. The approximation scheme from Algorithm 1 can also be adapted to calculate the signature of a path, the structure of which can be seen in Remark 2.32. We again approximate

$$\mathbb{X}^m_{s,t} = \int_{s \leq \tau_1 \leq ... \leq \tau_m \leq t} d\mathbf{x}_{\tau_1} \otimes ... \otimes d\mathbf{x}_{\tau_m} = \int_s^t \mathbb{X}^{m-1}_{s,\tau} \otimes d\mathbf{x}_\tau$$

by

$$\mathbb{X}^m_{s,t} \approx \sum_{n=0}^{N-1} \mathbb{X}^{m-1}_{s,t_n} \otimes \mathbf{x}^1_{t_n,t_{n+1}} + \mathbb{X}^{m-2}_{s,t_n} \otimes \mathbf{x}^2_{t_n,t_{n+1}}.$$

Algorithm 2 Milstein scheme for the signature

Input: $\mathbf{x}$: rough path, $N \geq 1$: number of steps, $t > 0$: time, m: maximum level
Output: z
  for $k \leftarrow 0, ..., m$ **do**
    initialize $z_{last}[k]$ to be the zero tensor of shape $\underbrace{n \times ... \times n}_{k \text{ times}}$
  end for
  $\Delta t \leftarrow \frac{t}{N}$
  for $k \leftarrow 0, ..., N-1$ **do**
    $u \leftarrow k\Delta t$
    $\Delta \mathbf{x}^1 \leftarrow \mathbf{x}^1_{u+\Delta t} - \mathbf{x}^1_u$
    $\Delta \mathbf{x}^2 \leftarrow \mathbf{x}^2_{u+\Delta t} - \mathbf{x}^2_u - \mathbf{x}^1_u \otimes \Delta \mathbf{x}^1$
    $z_{curr}[0] \leftarrow 1$
    for $k \leftarrow 1, 2$ **do**
      $z_{curr}[k] \leftarrow \mathbf{x}^k_u$
    end for
    for $k \leftarrow 3, ..., m$ **do**
      $z_{curr}[k] \leftarrow z_{last}[k] + z_{last}[k-1] \otimes \Delta \mathbf{x}^1 + z_{last}[k-2] \otimes \Delta \mathbf{x}^2$
    end for
    $z_{last} \leftarrow z_{curr}$
  end for
  $z \leftarrow z_{last}$

We can use Algorithm 2 to calculate the signature of the p-rough path $\mathbf{x}$ at time $t > 0$ ($\mathbb{X}^{\leq m}_{0,t}$) up to level m in N steps.
Similar to before, we can calculate the order of convergence:

Lemma 6.2:
For $\mathbf{x}$ a p-rough path with $2 \leq p < 3$ and a level m, we have

$$\left|\hat{\mathbb{X}}^m_{0,t} - \mathbb{X}^m_{0,t}\right| = O\left(\sup_{n=0,\dots,N-1} \mathrm{Var}^{3-p}_{p,[t_n,t_{n+1}]}(\mathbf{x})\right),$$

where

$$\mathbb{X}^m_{0,t} = \int_{0\leq\tau_1\leq\ldots\leq\tau_m\leq t} d\mathbf{x}_{\tau_1} \otimes ... \otimes d\mathbf{x}_{\tau_m}$$

and $\hat{\mathbb{X}}^m_{0,t}$ is the approximation of Algorithm 2.

Proof. Let $m \geq 3$, so that we actually approximate the solution, since for $m = 0, 1, 2$ the solution is part of the rough path $\mathbf{x}$. Also let $R^k_s := \hat{\mathbb{X}}^k_{0,t} - \mathbb{X}^k_{0,t}$ be the approximation error at level m. Then, for $s = t_n, t = t_{n+1}$ we have

$$\begin{aligned} R^k_t - R^k_s =& \hat{\mathbb{X}}^{k-1}_{0,s} \otimes \mathbf{x}^1_{s,t} + \hat{\mathbb{X}}^{k-2}_{0,s} \otimes \mathbf{x}^2_{s,t} - \int_s^t \mathbb{X}^{k-1}_{0,\tau} \otimes d\mathbf{x}_\tau \\ =& R^{k-1}_s \otimes \mathbf{x}^1_{s,t} + R^{k-2}_s \otimes \mathbf{x}^2_{s,t} + \underbrace{\mathbb{X}^{k-1}_{0,s} \otimes \mathbf{x}^1_{s,t} + \mathbb{X}^{k-2}_{0,s} \otimes \mathbf{x}^2_{s,t} - \int_s^t \mathbb{X}^{k-1}_{0,\tau} \otimes d\mathbf{x}_\tau}_{=O(\mathrm{Var}^3_{p,[s,t]}(\mathbf{x}))} \end{aligned}$$

and therefore (if R^{k-2}_s is bounded)

$$\begin{aligned} R^k_t - R^k_s - R^{k-1}_s \otimes \mathbf{x}^1_{s,t} =& R^{k-2}_s \otimes \mathbf{x}^2_{s,t} + O(\mathrm{Var}^3_{p,[s,t]}(\mathbf{x}) \\ \leq& C\left(\left\|R^{k-2}_\cdot\right\|_\infty + \mathrm{Var}_{p,[s,t]}(\mathbf{x})\right) \mathrm{Var}^2_{p,[s,t]}(\mathbf{x}). \end{aligned}$$

[2] `https://github.com/tobna/DeepRoughPaths`

Analogously to above

$$R_t^{k-1} - R_s^{k-1} = R_s^{k-2} \otimes \mathbf{x}_{s,t}^1 + R_s^{k-3} \otimes \mathbf{x}_{s,t}^2 + O(\mathrm{Var}_{p,[s,t]}^3(\mathbf{x}))$$
$$\leq C\left(\left\|R_\cdot^{k-2}\right\|_\infty + \left\|R_\cdot^{k-3}\right\|_\infty \mathrm{Var}_{p,[s,t]}(\mathbf{x}) + \mathrm{Var}_{p,[s,t]}^2(\mathbf{x})\right) \mathrm{Var}_{p,[s,t]}(\mathbf{x})$$

and $\left(R_\cdot^k, R_\cdot^{k-1}\right) \in \mathcal{D}_{\mathbf{x}}^1$ with

$$\left\|\left(R_\cdot^k, R_\cdot^{k-1}\right)\right\|_{\mathcal{D}_{\mathbf{x}}^1} \leq C\left[\left\|R_\cdot^{k-2}\right\|_\infty + \left(\left\|R_\cdot^{k-3}\right\|_\infty + 1\right) \sup_{n=0,\dots,N-1} \mathrm{Var}_{p,[t_n,t_{n+1}]}(\mathbf{x})\right].$$

When combining all the estimates from Lemma 2.11 and Theorem 2.29, we get

$$\left|\int_0^t R_s^k d\mathbf{x}_s\right| \overset{R_0^k=0}{\leq} \left(2^{\frac{3}{p}+1}\zeta\left(\frac{3}{p}\right)+1\right) \left\|(R_\cdot^k, R_\cdot^{k-1})\right\|_{\mathcal{D}_{\mathbf{x}}^1} \mathrm{Var}_{p,[0,T]}^3(\mathbf{x}).$$

Now, since $\hat{\mathbb{X}}_{0,t}^{k-1} = \mathbb{X}_{0,t}^{k-1} + R_t^{k-1}$ for all k, we have

$$\left|\mathbb{X}_{0,t}^{k+1} - \int_0^t \hat{\mathbb{X}}_{0,s}^k \otimes d\mathbf{x}_s\right| \leq \left|\int_0^t R_s^k d\mathbf{x}_s\right|$$
$$\leq C\,\mathrm{Var}_{p,[0,T]}^3(\mathbf{x})\left[\left\|R_\cdot^{k-2}\right\|_\infty + \left(\left\|R_\cdot^{k-3}\right\|_\infty + 1\right) \sup_{n=0,\dots,N-1} \mathrm{Var}_{p,[t_n,t_{n+1}]}(\mathbf{x})\right]$$

and

$$\left|\hat{\mathbb{X}}_{0,t}^{k+1} - \int_0^t \hat{\mathbb{X}}_{0,s}^k \otimes d\mathbf{x}_s\right| = O\left(\sup_{n=0,\dots,N-1} \mathrm{Var}_{p,[t_n,t_{n+1}]}^{3-p}(\mathbf{x})\right)$$

by Lemma 6.1. Therefore

$$\left|R_t^{k+1}\right| = \left|\mathbb{X}_{0,t}^{k+1} - \hat{\mathbb{X}}_{0,t}^{k+1}\right| \leq O\left(\sup_{n=0,\dots,N-1} \mathrm{Var}_{p,[t_n,t_{n+1}]}^{3-p}(\mathbf{x})\right)$$
$$+ C\,\mathrm{Var}_{p,[0,T]}^3(\mathbf{x})\left[\left\|R_\cdot^{k-2}\right\|_\infty + \left(\left\|R_\cdot^{k-3}\right\|_\infty + 1\right) \sup_{n=0,\dots,N-1} \mathrm{Var}_{p,[t_n,t_{n+1}]}(\mathbf{x})\right]$$

Since we know $\mathbb{X}_{0,t}^k$ exactly for $k = 0, 1, 2$ and therefore $R_s^k = 0$ for these k, and $3 - p < 1$, we have

$$\|R_\cdot^m\|_\infty = O\left(\sup_{n=0,\dots,N-1} \mathrm{Var}_{p,[t_n,t_{n+1}]}^{3-p}(\mathbf{x})\right)$$

by induction. □

Therefore, Algorithm 2 converges to the desired value.

6.2 RDEs

We will take a finite differences approach for approximating the solution of a rough differential equation, while additionally including the second-order information rough paths provide. This is a generalization of the Milstein scheme for SDEs with Brownian motion as the driving path. For a rough differential equation

$$dY = \mu(Y,t)dt + f(Y)d\mathbf{x} \tag{15}$$

we consider its integrated form

$$Y_t = Y_s + \int_s^t \mu(Y_\tau, \tau) d\tau + \int_s^t f(Y_\tau) d\mathbf{x}.$$

Now we approximate the increments of both of the integrals by

$$\begin{aligned}\int_s^t \mu(Y_\tau, \tau) d\tau \approx& \mu(Y_s, s)(t-s),\\ \int_s^t f(Y_\tau) d\mathbf{x} \approx& f(Y_s)\mathbf{x}^1_{s,t} + f(Y_s)'\mathbf{x}^2_{s,t},\end{aligned}$$

for small $|t-s|$. This results in Algorithm 3.

Algorithm 3 Milstein scheme for RDEs

Input: $N \geq 1$, $t > 0$, ξ
Output: y
  $y \leftarrow \xi$
  $\Delta t \leftarrow \frac{t}{N}$
  for $n \leftarrow 0, ..., N-1$ **do**
    $u \leftarrow n\Delta t$
    $m \leftarrow \mu(y, u)$
    $s \leftarrow f(y)$
    $\Delta\mathbf{x}^1 \leftarrow \mathbf{x}^1_{u+\Delta t} - \mathbf{x}^1_u$
    $\Delta\mathbf{x}^2 \leftarrow \mathbf{x}^2_{u+\Delta t} - \mathbf{x}^2_u - \mathbf{x}^1_u \otimes \Delta\mathbf{x}^1$
    $z \leftarrow D_Y f(y) s$
    $y \leftarrow y + m\Delta t + s\Delta\mathbf{x}^1 + z\Delta\mathbf{x}^2$
  end for

Last but not least, we can again consider the order of convergence; the main idea of this proof is based on [FV10, Theorem 10.30].

Lemma 6.3:
Let Y be the solution to Equation (15) and let $Y^{Mil;D}$ be the Milstein approximation from Algorithm 3 based on the partition $D = \{0 = t_0 < t_1 < ... < t_N = T\}$ of $[0,T]$. Let μ be $\frac{1}{p}$-Hölder in t and uniformly Lipschitz in Y. Then we have

$$\left\| Y_\cdot - Y^{Mil;D}_\cdot \right\|_\infty = O\left(\sup_{n=0,...,N-1} \omega^{\frac{3}{p}-1}(t_n, t_{n+1}) \right).$$

In this proof, we again take a controlling function ω with

$$\omega(s,t) \geq \mathrm{Var}^p_{p,[s,t]}(\mathbf{x}) + |t-s|,$$

similar to the proof of existence and uniqueness of the solution (Theorem 3.6).

Proof. First, we again consider the error of the one-step increments: It holds

$$\left| \int_s^t f(Y_\tau) d\mathbf{x} - f(Y_s)\mathbf{x}^1_{s,t} + f(Y_s)'\mathbf{x}^2_{s,t} \right| \leq C\omega^{\frac{3}{p}}(s,t)$$

by Theorem 2.29 and also

$$\begin{aligned}\left|\int_s^t \mu(Y_\tau,\tau)d\tau - \mu(Y_s,s)(t-s)\right| \leq &\int_s^t \left|\mu(Y_\tau,\tau) - a(Y_s,s)\right| d\tau\\ \leq &\int_s^t \left|\mu(Y_\tau,\tau)-\mu(Y_\tau,s)\right| + \left|\mu(Y_\tau,s)-\mu(Y_s,s)\right| d\tau\\ \leq &\int_s^t L_{\mu-Lip}(\tau-s) + L_{\mu-Lip}\left|Y_\tau - Y_s\right| d\tau\\ \leq &\frac{\|\mu\|_{\frac{1}{p}-\text{Höl};t}}{2}(t-s)^{1+\frac{1}{p}}\\ &+\left(2\left\|(Y,Y')\right\|_{\mathcal{D}^1_{\omega,\mathbf{x}}}\omega^{\frac{1}{p}}(s,t)+\|Y\|_\infty\right)\omega^{\frac{1}{p}}(s,t)(t-s)\\ \leq &C_{\mu,Y}\omega^{\frac{1}{p}+1}(s,t)\end{aligned}$$

by Equation (10). Since $p \geq 2$, we have

$$\left|\int_s^t \mu(Y_\tau,\tau)d\tau + \int_s^t f(Y_\tau)d\mathbf{x} - \mu(Y_s,s)(t-s) - f(Y_s)\mathbf{x}^1_{s,t} + f(Y_s)'\mathbf{x}^2_{s,t}\right| \leq C\omega^{\frac{3}{p}}(s,t).$$

Now we extend this to the multi-step Milstein approximation. Let

$$D = \{0 = t_0 < t_1 < ... < t_N = t\}$$

and define z^k to be the value of the process that is the approximate solution in k steps from $0 = t_0$ to t_k (calculated via the Milstein scheme) and the exact solution to Equation (15) in $[t_k, T]$, started at $z^k_{t_k}$. We then have the exact solution $z^0_t = Y_t$ and the approximation $z^N_t = Y^{\text{Mil};D}_t$. Therefore

$$\left|Y_t - Y^{\text{Mil};D}_t\right| = \left|z^0_t - z^N_t\right| \leq \sum_{k=0}^{N-1}\left|z^k_t - z^{k+1}_t\right|.$$

Since in the interval $[t_{k+1}, t]$, both z^k and z^{k+1} are solutions to Equation (15) but started at different values $z^{k+1}_{t_{k+1}}$ and $z^k_{t_{k+1}}$, we can see by Theorem 3.11 that there is a $C > 0$, such that

$$\left\|z^{k+1}, z'^{k+1}; z^k, z'^k\right\|_{\mathbf{x},\mathbf{x},\omega,1} \leq C\left|z^{k+1}_{t_{k+1}} - z^k_{t_{k+1}}\right|$$

and by Equation (12)

$$\left|z^k_{t_{k+1},t} - z^{k+1}_{t_{k+1},t}\right| \leq \left(2\omega^{\frac{1}{p}}(0,T)\left\|z^{k+1}, z'^{k+1}; z^k, z'^k\right\|_{\mathbf{x},\mathbf{x},\omega,1} + \left\|z'^{k+1}_{t_{k+1}} - z'^k_{t_{k+1}}\right\|\right)\omega^{\frac{1}{p}}(t_k,t).$$

We can then use

$$\left\|z'^{k+1}_{t_{k+1}} - z'^k_{t_{k+1}}\right\| = \left\|f(z^k_{t_{k+1}}) - f(z^{k+1}_{t_k})\right\| \leq \|Df\|_\infty\left|z^{k+1}_{t_{k+1}} - z^k_{t_{k+1}}\right|$$

to see

$$\begin{aligned}\left|z^k_t - z^{k+1}_t\right| \leq &\left|z^{k+1}_{t_{k+1}} - z^k_{t_{k+1}}\right| + 2C\omega^{\frac{2}{p}}(0,T)\left|z^{k+1}_{t_{k+1}} - z^k_{t_{k+1}}\right| + \omega^{\frac{1}{p}}(0,T)\|Df\|_\infty\left|z^{k+1}_{t_{k+1}} - z^k_{t_{k+1}}\right|\\ \leq &C'\left|z^{k+1}_{t_{k+1}} - z^k_{t_{k+1}}\right|.\end{aligned}$$

As $z^{k+1}_{t_k} = z^k_{t_k}$ by definition, we can use the calculations from above to show

$$\begin{aligned}
\left|z^{k+1}_{t_{k+1}} - z^k_{t_{k+1}}\right| =& \left|z^{k+1}_{t_{k+1}} - z^{k+1}_{t_k} - z^k_{t_{k+1}} + z^k_{t_k}\right| \\
=& \Big| \int_{t_k}^{t_{k+1}} a(z^{k+1}_\tau, \tau) d\tau + \int_{t_k}^{t_{k+1}} b(z^{k+1}_\tau) d\mathbf{x} \\
&- a(z^{k+1}_{t_k}, t_k)(t_{k+1} - t_k) - b(z^{k+1}_{t_k})\mathbf{x}^1_{t_k,t_{k+1}} + b(z^{k+1}_{t_k})'\mathbf{x}^2_{t_k,t_{k+1}} \Big| \\
\leq& C\omega^{\frac{3}{p}}(t_k, t_{k+1}).
\end{aligned}$$

All in all, we have

$$\begin{aligned}
\left|Y_t - Y^{\mathrm{Mil};D}_t\right| \leq& \sum_{k=0}^{N-1} \left|z^k_t - z^{k+1}_t\right| \leq C' \sum_{k=0}^{N-1} \left|z^k_{t_{k+1}} - z^{k+1}_{t_{k+1}}\right| \\
\leq& \tilde{C} \sum_{k=0}^{N-1} \omega^{\frac{3}{p}}(t_k, t_{k+1}) \leq \tilde{C}\omega(0,T) \sup_{n=0,\dots,N-1} \omega^{\frac{3}{p}-1}(t_n, t_{n+1}),
\end{aligned}$$

which shows the assertion. □

This means, that the Milstein scheme converges path-wise to the correct solution up to floating point errors, when $\|D\| \to 0$.

Remark 6.4:
The solutions of RDEs can be defined in several different ways. Firstly, as controlled rough paths, like we have done in Definition 2.30, as rough paths, like in [LCL07] or [Lej03], or as the limit of exactly such a Milstein approximation. This last one is known in the literature as *Davie's definition* (from [Dav07]).

7 Benchmarks

In order to test the algorithms, we conducted a benchmarking survey against the Differential Equations julia library [RN17]. Here, we took advantage of the fact that our algorithms were easily scalable to large batches, computing multiple samples of a process at a time.
The experiments were conducted on a machine with specs from Table 1 and were executed using the CPU only. We would reckon, that one could get another sizable speed boost by using a GPU for cuda calculations. This is possible with the code we have written.

OS	Linux (5.13.0-35-generic x86_64)
CPU	AMD Ryzen 5 3600 (6 Cores @ 3.6 GHz)
RAM	2x 8 GB DDR4 RAM @ 4200 MHz
Python version	3.9.7
PyTorch version	1.10.2+cu102

Table 1: Specs of the benchmarking machine.

The benchmark was to approximate the solution of a stochastic differential equation, where the exact solution is known, in order to plot the runtime against the strong error

$$E_S = \mathbb{E}\left[\left|\hat{Y}_T - Y_T\right|\right],$$

where Y is the correct solution and $\hat{Y}$ is the approximate solution. This was done for multiple RDEs. In the resulting work-precision-diagrams, our method is called *RDE Solver* and the number is the batch size of the experiment, while all other methods are a selection of methods from [RN17] run with a batch size of 1.

7.1 Scalar Noise Problem

The first goal is to solve the SDE

$$dX_t = \alpha X_t dt + \beta X_t dB_t, \;\; X_0 = \frac{1}{2},$$

where W_t is standard Brownian motion. By the Itô formula, the correct solution is the geometric Brownian motion

$$X_t = X_0 e^{\left(\alpha - \frac{\beta^2}{2}\right)t + \beta B_t}.$$

The experiments were conducted with $\alpha = \frac{1}{10}$ and $\beta = \frac{1}{20}$.
We can see in Figure 1 that our method performs very well with large batch sizes. We found that up to a batch size of 100000, there was basically no difference in computation speed when increasing the batch size. The second-order terms give a huge precision boost of up to a factor of 1000 compared to the Euler scheme (EM) (using the implementation from [RN17]).

7.2 Scalar Wave SDE

A more complex SDE we tested was

$$dX_t = -\alpha^2 \sin(X_t) \cos^3(X_t)\, dt + \alpha \cos^2(X_t)\, dB_t, \;\; X_0 = \frac{1}{2}.$$

Let

$$h(x) := \arctan\left(\alpha x + \tan(X_0)\right).$$

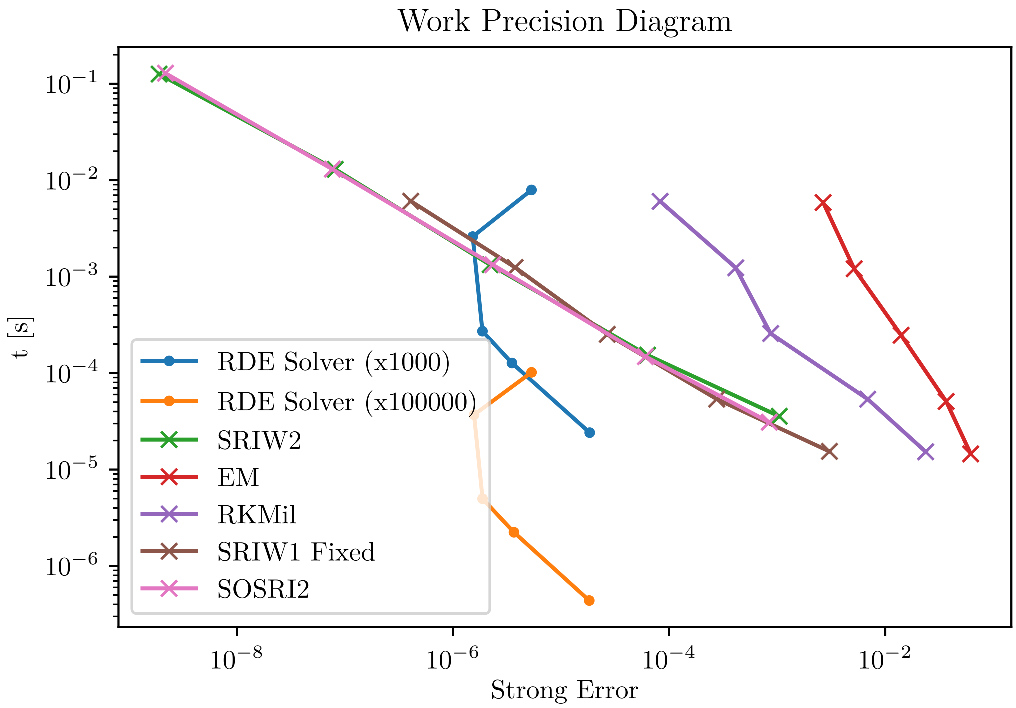

Figure 1: Work precision diagram of different SDE solving methods for the scalar noise problem.

Let $x = \tan(y)$ for $y \in \mathbb{R}$. Then

$$\begin{aligned} x^2 =& \frac{\sin^2(y)}{\cos^2(y)} \\ \Rightarrow x^2 + 1 =& \frac{\sin^2(y) + \cos^2(y)}{\cos^2(y)} = \frac{1}{\cos^2(y)} \\ \Rightarrow \frac{1}{1+x^2} =& \cos^2(y) = \cos^2(\arctan(x)). \end{aligned}$$

Using this, we obtain

$$\frac{\partial h}{\partial x}(x) = \alpha \frac{1}{1 + (\alpha x + \tan(X_0))^2} = \alpha \cos^2(\arctan(\alpha x + \tan(X_0))) = \alpha \cos^2(h(x))$$

and

$$\frac{\partial^2 h}{\partial x^2}(x) = -2\alpha^2 \sin(h(x)) \cos^3(h(x)).$$

Therefore, by the Itô formula, the correct solution is

$$X_t = h(B_t) = \arctan\left(\alpha B_t + \tan\left(X_0\right)\right).$$

The experiment was again conducted using a value of $\alpha = \frac{1}{10}$.
In Figure 2, we see that for the scalar wave problem some of the adaptive algorithms approach lower errors than our RDE method, but using large batch sizes our method still is way faster than all the others. For this SDE, there is no real difference in error between the RDE method and the standard Euler scheme (EM), but using parallelization, our method can still be faster than the julia implementation of the Euler scheme.

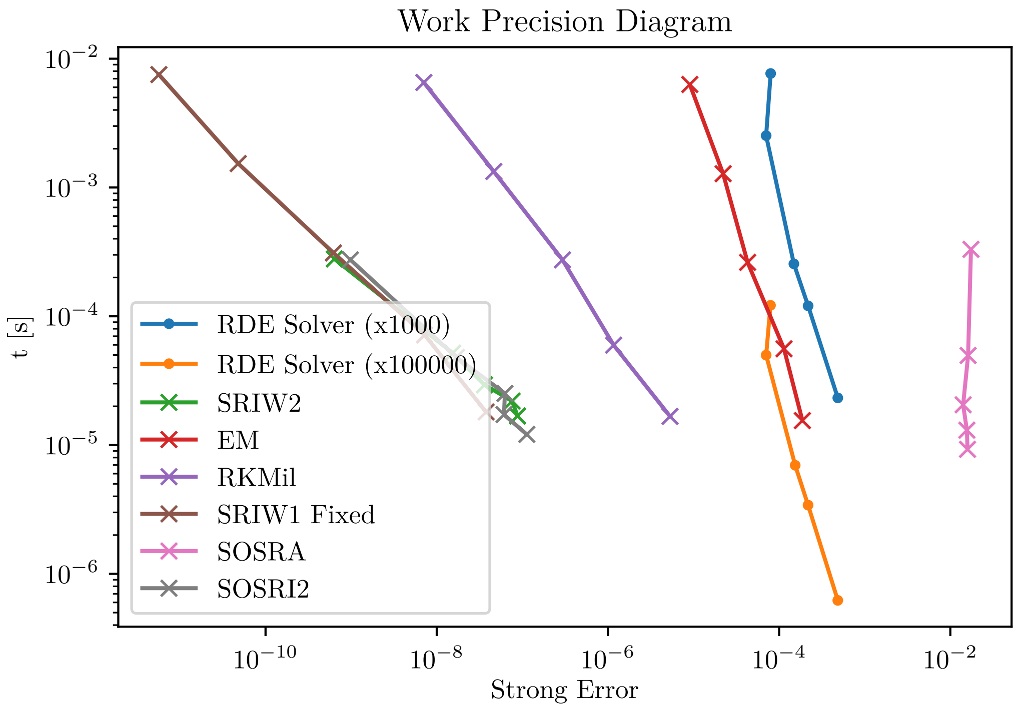

Figure 2: Work precision diagram of different SDE solving methods for the scalar wave problem.

7.3 Optimal Asset Allocation

Consider the Black-Scholes like model of a financial market made up of one stock S_t^1 with

$$dS_t^1 = \mu S_t^1 dt + \sigma S_t^1 dB_t,$$

where B_t is standard Brownian motion and $\mu = r + \xi\sigma$, with ξ being the market price of risk, and a savings account S_t^0 with constant interest rate r

$$dS_t^0 = S_t^0 r dt.$$

Let W_0 be the initial wealth and π be a constant rate of additional investment. Let p_t be the percentage of wealth invested in the stock at time t. Then we can model the wealth process W_t to be

$$\begin{aligned} dW_t =& \frac{p_t W_t}{S_t^1} dS_t^1 + (1-p_t)W_t r dt + \pi dt \\ =& \mu p_t W_t dt + p_t \sigma W_t dB_t + (1-p_t)W_t r dt + \pi dt \\ =& \left[(r + p_t\xi\sigma)\, W_t + \pi\right] dt + p_t \sigma W_t dB_t. \end{aligned}$$

Now, one may notice that this does not strictly fall into our problem setup. The drift part is fine, as we have

$$\begin{aligned} \left\| (r + p_t\xi\sigma)\,\zeta + \pi, (r + \tilde{p}_t\xi\sigma)\,\zeta + \pi \right\|_{\infty;T,M} =& \sup_{\substack{t\in[0,T] \\ \|\zeta\|\le M}} \left\| (r + p_t\xi\sigma)\,\zeta + \pi - (r + \tilde{p}_t\xi\sigma)\,\zeta - \pi \right\| \\ \le& \sup_{t\in[0,T]} \left\| r + p_t\xi\sigma - r - \tilde{p}_t\xi\sigma \right\| M \\ \le& \xi\sigma M \left\| p - \tilde{p} \right\|_\infty \end{aligned}$$

for different asset allocation policies p and $\tilde{p}$. This means that we can control the error in the RDE solution, that comes from the drift term. The error in the volatility term ($p_t\sigma W_t dB_t$) can be controlled using an analog to Theorem 3.11, that incorporates the error that comes from using p_t instead of $\tilde{p}_t$. Using

$$\begin{aligned} f(\xi, t) :=& p_t\sigma\xi, \\ \tilde{f}(\xi, t) :=& \tilde{p}_t\sigma\xi \end{aligned}$$

gives

$$\left|f(\xi, t) - \tilde{f}(\tilde{\xi}, t)\right| = \left|p_t\sigma\xi - \tilde{p}_t\sigma\tilde{\xi}\right| \leq \sigma \left\|p - \tilde{p}\right\|_\infty \underbrace{|\xi|}_{\leq M} + \sigma \left\|\tilde{p}\right\|_\infty \left|\xi - \tilde{\xi}\right|$$

and

$$\begin{aligned} \left\|f(\xi, t)' - \tilde{f}(\tilde{\xi}, t)'\right\| &= \left\|D_\xi f(\xi, t) f(\xi, t) - D_\xi \tilde{f}(\tilde{\xi}, t) f(\tilde{\xi}, t)\right\| = \left\|p_t^2\sigma^2\xi - \tilde{p}_t^2\sigma^2\tilde{\xi}\right\| \\ &\leq \left\|p_t^2\sigma^2\xi - p_t^2\sigma^2\tilde{\xi}\right\| + \left\|p_t^2\sigma^2\tilde{\xi} - \tilde{p}_t^2\sigma^2\tilde{\xi}\right\| \\ &\leq \left\|p_t\right\|_\infty^2 \sigma^2 \left|\xi - \tilde{\xi}\right| + \sigma^2 \underbrace{\left|\tilde{\xi}\right|}_{\leq M} \left\|p^2 - \tilde{p}^2\right\|_\infty \\ &\leq \left\|p\right\|_\infty^2 \sigma^2 \left|\xi - \tilde{\xi}\right| + \sigma^2 M \left\|p - \tilde{p}\right\|_\infty (\left\|p\right\|_\infty + \left\|\tilde{p}\right\|_\infty), \end{aligned}$$

which incorporates the error in the volatility term into the estimate of Theorem 3.11, giving us the additional term $\|p_\cdot - \tilde{p}_\cdot\|_\infty$. This shows that the optimal expected loss for $p \in C(\Lambda_T, \mathbb{R})$ can be approximated by $\tilde{p} \in \mathcal{T}_{sig}$. To finally incorporate the condition $p_t \in [0, 1]$, we can simply set $p_t = s(c_t)$, where s is the sigmoid function

$$s(z) := \frac{1}{1 + e^{-z}}$$

and $c \in C(\Lambda_T, \mathbb{R})$, or $c \in \mathcal{T}_{sig}$, respectively. Then

$$\left\|p - \tilde{p}\right\|_\infty = \left\|s \circ c - s \circ \tilde{c}\right\|_\infty \leq \left\|c - \tilde{c}\right\|_\infty$$

together with

$$\{p \in C(\Lambda_T, \mathbb{R}) | p_t \in [0, 1] \text{ for all } t \in [0, T]\} = \{s \circ c | c \in C(\Lambda_T, \mathbb{R})\}$$

gives the result we want.
We want to maximize the expected returns at some time $T > 0$, while at the same time minimizing the volatility of the returns. For this, we introduce a Lagrange multiplier λ to solve the optimization problem

$$p^* = \operatorname{argmax}_p \left(\mathbb{E}[W_T^p] - \lambda \operatorname{Var}[W_T^p]\right),$$

where W_T^p is the wealth process with asset allocation policy p at time T and the maximum goes over all admissible (continuous and adapted) policies.
For Itô and Stratonovich SDEs involving Brownian motion, we can also choose a different approach of approximating $\Theta \in \mathcal{T} = C(\Lambda_T, \mathbb{R})$. In this case, we make use of the Markov property of Brownian motion, i.e.

$$\sigma(B_u | u \leq s) \underset{B_s}{\perp\!\!\!\perp} B_t$$

for all $s \leq t$. That is why we can simply fall back to $C([0, T] \times \mathbb{R}, \mathbb{R})$ for functions of time and the current value instead of functions of the whole path up to the current time. Since neural networks of large enough depth and width can approximate any continuous function arbitrarily

r	σ	ξ	π	W_0	T	λ
0.03	0.15	0.33	0.1	1	20	0.5

Table 2: Values of constants in experiment.

well (on compact domains) [KL20; Les+93], we model $p_t = p(t, W_t)$ to be a continuous function of the time and wealth at that time, which we approximate using a neural network with two hidden layers of size 50 by solving the differential equation as an RDE using Algorithm 3 and then using backpropagation to update the policy p.
Figure 3 shows the learnt policy for the constant values from Table 2.

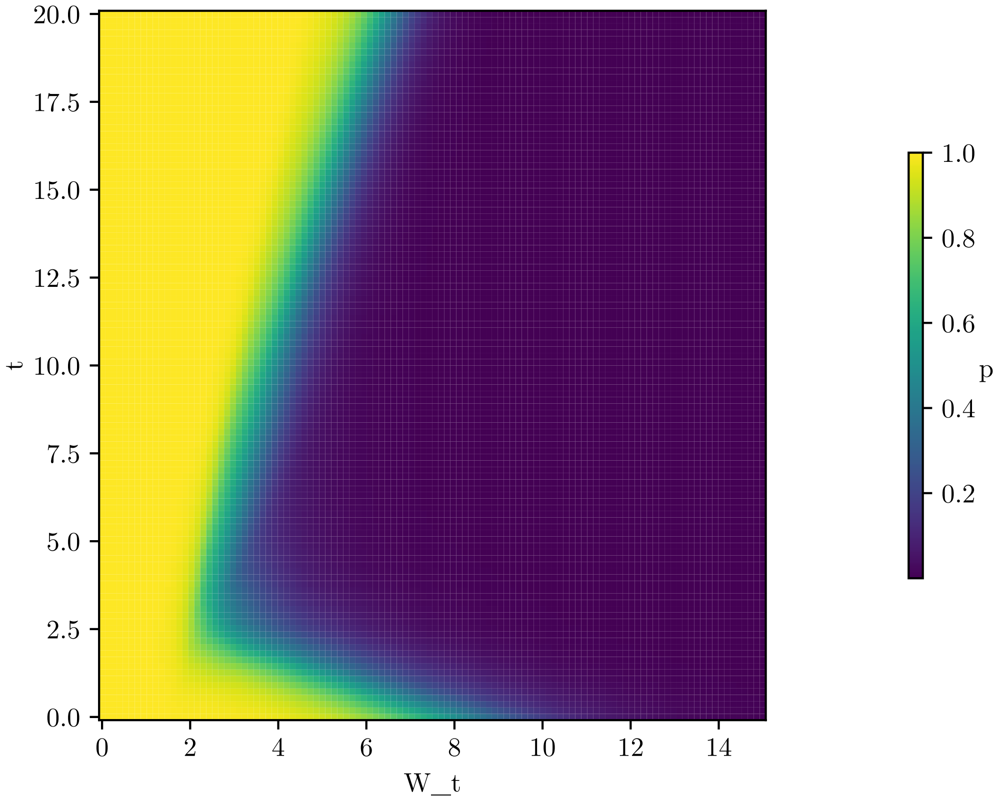

Figure 3: Learnt $p(W_t, t)$ for values from Table 2.

8 Discussion

We would like to discuss a few possible extensions to our problem setup and its solution.

8.1 More General Drift Terms

It would be interesting to know if one can generalize the main result of this thesis (Theorem 5.6) by allowing more general functions for the control. In a first step, one could allow $\Theta(\hat{\mathbf{x}}|_{[0,t]})$ with Θ Borel-measurable instead of continuous. In this case, it can be interesting to note the following:
Looking at [Wis94, Theorem 1], one can find a sequence $(\Theta_n) \subset C(\Lambda_T, \mathbb{R}) = \mathcal{T}$, such that

$$\Theta_n \xrightarrow{n\to\infty} \Theta \qquad\qquad a.s.$$

as $\mathcal{B}(\Lambda_T)$ coincides with the topology induced by the map

$$\varphi : [0, T] \times \hat{\Omega}^p_T \ni (t, \mathbf{x}) \mapsto \mathbf{x}|_{[0,t]} \in \Lambda_T$$

by [Bay+22, Lemma A.1]. Looking at the proof of said theorem, we see that actually $\Theta_n \equiv \Theta$ on a set K_n with

$$\lambda \otimes \mathbb{P}(K_n^c) \xrightarrow{n\to\infty} 0,$$

where λ is the Lebesgue measure, by Lusin's theorem [Coh13, Theorem 7.4.4]. This implies

$$\|\Theta - \Theta_n\|_{L^{\frac{p}{p-2}}(\Lambda_T)} \leq 2 \|\Theta\|_\infty (\lambda \otimes \mathbb{P}(K_n^c))^{\frac{p-2}{p}} \xrightarrow{n\to\infty} 0,$$

which would then allow us to control the error for bounded Θ and give the result.
The most general result of this type would be to show

$$\inf_{\mu_t \text{ adapted}} \mathbb{E}[L(Y^\mu)] = \inf_{\mu_t \in \mathcal{T}_{sig}} \mathbb{E}[L(Y^\mu)]$$

with the same problem setup as before. We would reckon that for this, one needs to have additional knowledge on the structure of adapted functions and how to approximate those.

8.2 Drift & Gubinelli Derivative

As mentioned in Remark 3.7, in the problem setup we have chosen, the drift does not show up as part of the Gubinelli derivative of the solution of an RDE. This would be different when thinking about an RDE with drift simply as an RDE of the augmented path $\hat{\mathbf{x}}$. More accurately, our solution to

$$dY = \mu(t, Y_t)dt + f(Y_t)d\mathbf{x}$$

has Gubinelli derivative $Y'_t = f(Y_t)$, while the solution to

$$d\hat{Y} = \hat{f}(\hat{Y})d\hat{\mathbf{x}}$$

with

$$\hat{f}(\hat{Y}_t) := \begin{pmatrix} f(Y_t) \\ \mu(t, Y_t) \end{pmatrix}$$

has $\hat{Y}' = f(\hat{Y})$. Then $\hat{f}(\hat{Y}_t)d\hat{\mathbf{x}}_t = f(Y_t)d\mathbf{x}_t + \mu(t, Y_t)dt$ and the solution path (Y) is indeed the same (aside from the extra dimension of $\hat{Y}$, that just is the time itself). This difference would not be present if we choose $\mathcal{D}^1_{\omega,\hat{\mathbf{x}}}$ to be the solution space for our RDE, which we could, and also set

$$Y' = \begin{pmatrix} f(Y_t) \\ \mu(t, Y_t) \end{pmatrix}.$$

In the $R^Y_{s,t}$ term, this would be no problem and would work analog to the estimates we used in Theorem 3.6. However, we would additionally need the estimate

$$|\mu(t, Y_t) - \mu(s, Y_s)| \le C\omega^{\frac{1}{p}}(s,t) \tag{16}$$

for the second term of $\|(Y, Y')\|_{\mathcal{D}^1_{\omega,\hat{\mathbf{x}}}}$. This would then allow us to simplify some parts of the theory drastically, starting with the proof of Theorem 3.8. There, the dependence on $|Y_s|$ would be omitted, not only allowing us to skip half of the proof, which is about finding a bound for $\|Y\|_\infty$, but it would also show that the resulting constant is a linear combination of terms independent of $\omega(0,T)$ times $\omega(0,T)^\beta$ for different $\beta > 0$. It would also allow us to omit the demand for $\frac{1}{p}$-Hölder continuity in Lemma 6.3 and just work with the same problem setup as before.
We have omitted this extension as, in the general case of our problem, an estimate of the form of Equation (16) does not hold.

References

[Arr+18] Imanol Perez Arribas et al. "A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder". In: *Translational Psychiatry* 8 (2018). DOI: 10.1038/s41398-018-0334-0.

[Bay+22] Christian Bayer et al. "Optimal stopping with signatures". In: *The Annals of Applied Probability* (2022), to appear. arXiv: 2105.00778.

[Che17] Ilya Chevyrev. "Random walks and Lévy processes as rough paths". In: *Probability Theory and Related Fields* 170.3-4 (2017), pp. 891–932. DOI: 10.1007/s00440-017-0781-1.

[CK16] Ilya Chevyrev and Andrey Kormilitzin. *A Primer on the Signature Method in Machine Learning.* 2016. DOI: 10.48550/ARXIV.1603.03788.

[CL05] Laure Coutin and Antoine Lejay. "Semi-martingales and rough paths theory". In: *Electronic Journal of Probability* 10 (2005), pp. 761–785. DOI: 10.1214/EJP.v10-162.

[Coh13] Donald L. Cohn. *Measure Theory.* Birkhäuser Advanced Texts Basler Lehrbücher. Birkhäuser New York, NY, 2013. DOI: 10.1007/978-1-4614-6956-8.

[CQ02] Laure Coutin and Zhongmin Qian. "Stochastic analysis, rough path analysis and fractional Brownian motions". In: *Probability Theory and Related Fields* 122 (2002), pp. 108–140. DOI: 10.1007/s004400100158.

[Dav07] A. M. Davie. "Differential equations driven by rough paths: an approach via discrete approximation". In: *Applied Mathematics Research Express* (2007).

[DFG17] Joscha Diehl, Peter K. Friz, and Paul Gassiat. "Stochastic control with rough paths". In: *Applied Mathematics & Optimization* 75 (2017), pp. 285–315. DOI: 10.1007/s00245-016-9333-9.

[Faw03] Thomas Fawcett. "Problems in stochastic analysis: Connections between Rough paths and non-commutative harmonic analysis". PhD thesis. University of Oxford, 2003.

[FH20] Peter K. Friz and Martin Hairer. *A course on rough paths.* Springer International Publishing, 2020. DOI: 10.1007/978-3-030-41556-3_2.

[FS17] Peter K. Friz and Atul Shekhar. "General Rough integration, Levy Rough paths and a Levy–Kintchine type formula". In: *The Annals of Probability* 45.4 (2017), pp. 2707–2765. DOI: 10.1214/16-AOP1123.

[FV10] Peter K. Friz and Nicolas B. Victoir. *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2010. DOI: 10.1017/CBO9780511845079.

[Gra13] Benjamin Graham. *Sparse arrays of signatures for online character recognition.* 2013.

[Gub04] Massimiliano Gubinelli. "Controlling Rough Paths". In: *Journal of Functional Analysis* 216.1 (2004), pp. 86–140. DOI: 10.1016/j.jfa.2004.01.002.

[KL20] Patrick Kidger and Terry Lyons. "Universal Approximation with Deep Narrow Networks". In: *Proceedings of Thirty Third Conference on Learning Theory.* Ed. by Jacob Abernethy and Shivani Agarwal. Vol. 125. Proceedings of Machine Learning Research. PMLR, 2020, pp. 2306–2327.

[KLA20] Jasdeep Kalsi, Terry Lyons, and Imanol Perez Arribas. "Optimal execution with rough path signatures". In: *SIAM J. Financial Math.* 11 (2020), pp. 470–493.

[Kle06] Achim Klenke. *Wahrscheinlichkeitstheorie.* Masterclass. Springer Berlin Heidelberg, 2006. DOI: 10.1007/978-3-642-36018-3.

[KP92] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Applications of Mathematics. Springer Berlin Heidelberg, 1992. DOI: 10.1007/978-3-662-12616-5.

[LCL07] Terry J. Lyons, Michael J. Caruana, and Thierry Lévy. *Differential equations driven by rough paths*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2007. DOI: 10.1007/978-3-540-71285-5.

[Lej03] Antoine Lejay. "An Introduction to Rough Paths". In: *Séminaire de Probabilités XXXVII*. Ed. by Jacques Azéma et al. Springer Berlin Heidelberg, 2003, pp. 1–59. DOI: 10.1007/978-3-540-40004-2_1.

[Les+93] Moshe Leshno et al. "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function". In: *Neural Networks* 6.6 (1993), pp. 861–867. DOI: 10.1016/S0893-6080(05)80131-5.

[Lyo98] Terry J. Lyons. "Differential equations driven by rough signals." In: *Revista Matemática Iberoamericana* 14.2 (1998), pp. 215–310. DOI: 10.4171/RMI/240.

[Pas+19] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035.

[RN17] Christopher Rackauckas and Qing Nie. "Differentialequations.jl–a performant and feature-rich ecosystem for solving differential equations in julia". In: *Journal of Open Research Software* 5.1 (2017).

[Rud76] W. Rudin. *Principles of Mathematical Analysis*. International series in pure and applied mathematics. McGraw-Hill, 1976.

[SP14] René L. Schilling and Lothar Partzsch. *Brownian Motion: An Introduction to Stochastic Processes*. De Gruyter, 2014. DOI: 10.1515/9783110307306.

[Wis94] Andrzej Wisniewski. "The structure of measurable mappings on metric spaces". In: *Proceedings of the American Mathematical Society* 122.1 (1994), pp. 147–150.

Selbstständigkeitserklärung

Hiermit versichere ich, Tobias Christian Nauen, dass ich diese Arbeit selbstständig verfasst habe und keine weiteren als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Stellen der Arbeit, die wörtlich oder sinngemäß aus anderen Quellen übernommen wurden, habe ich als solche kenntlich gemacht. Zudem wurde die Arbeit bisher in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegt.

Hannover, 09.06.2022

Ort, Datum