Approximation and Stability in Rough Analysis with Applications to Mathematical Finance

Inauguraldissertation
zur Erlangung des akademischen Grades
des Doktors der Naturwissenschaften
der Universität Mannheim

vorgelegt von
Anna Paula Kwossek
aus Weinheim

Mannheim, 2025

Dekan: Professor Dr. Claus Hertling, Universität Mannheim
Referent: Professor Dr. David J. Prömel, Universität Mannheim
Korreferent: Professor Dr. Paul P. Hager, Universität Wien
Korreferent: Professor Dr. Nicolas Perkowski, Freie Universität Berlin

Tag der mündlichen Prüfung: 03.04.2025

Abstract

In this thesis we use and develop the mathematical framework of rough and stochastic analysis to deal with various aspects of approximation and stability that are particularly relevant for and motivated by applications in mathematical finance, machine learning, and numerical analysis.

We begin with addressing the widespread use of stochastic differential equations in which the drift and diffusion function are represented by neural networks, and provide a rigorous verification of their universal approximation property.
The theory of rough paths is known to provide a fully pathwise and robust solution theory to stochastic differential equations, which we exploit and contribute to as follows:
We study the well-posedness of rough differential equations with path-dependent coefficients and driven by càdlàg rough paths providing a unifying theory for the pathwise analysis of stochastic functional differential equations.
Subsequently, we assume a path property which implies a suitable canonical rough path lift such that the rough integral exists as a limit of left-point Riemann sums. We examine this further and present a transparent pathwise convergence analysis for the first order Euler scheme of stochastic differential equations that has been inexplicable from the rough path perspective so far.
This line of research is continued in the context of mathematical finance under model uncertainty when we investigate the pathwise stability and approximation properties of optimal portfolios.
To gain a deeper understanding, we explain and generalize the aforementioned path property, and prove that the rough integral exists under this assumption as a limit of general Riemann sums.
Based on this approach, we lastly bridge the gap between Itô integration and universal approximation with signatures.

Zusammenfassung

Diese Dissertation befasst sich mit Aspekten der Approximation und Stabilität in der stochastischen und der rauen Analysis, die insbesondere in der Finanzmathematik, im maschinellen Lernen und in der numerischen Analysis thematisiert werden.

Zunächst betrachten wir stochastische Differentialgleichungen, bei denen der Drift- und der Diffusionskoeffizient durch neuronale Netze gegeben sind, und zeigen deren theoretische universelle Approximationseigenschaft.
Die Theorie der rauen Pfade liefert eine vollständig pfadweise und robuste Lösungstheorie für stochastische Differentialgleichungen, der wir uns wie folgt annehmen:
Wir untersuchen die Wohlgestelltheit rauer Differentialgleichungen mit pfadabhängigen Koeffizienten und getrieben von càdlàg rauen Pfaden und formulieren somit einen vereinheitlichenden Ansatz für die pfadweise Analyse stochastischer verzögerter Differentialgleichungen.
Im Folgenden wird eine Pfadeigenschaft angenommen, sodass das raue Integral als Grenzwert linksseitiger Riemannsummen gegeben ist.
Darauf basierend präsentieren wir eine transparente pfadweise Konvergenzanalyse für das Euler-Verfahren erster Ordnung für stochastische Differentialgleichungen, welche die Theorie der rauen Pfade bisher nicht bieten konnte.
Darüber hinaus wenden wir uns dem finanzmathematischen Problem der Modellunsicherheit zu und untersuchen pfadweise Stabilität und Approximationseigenschaften optimaler Portfolios.
Zudem beleuchten und verallgemeinern wir die oben genannte Pfadeigenschaft und zeigen, dass das raue Integral folglich als Grenzwert genereller Riemannsummen existiert.
Schließlich wird dieser Integralbegriff verwendet, um ein Verständnis für den Zusammenhang von Itô-Integration und der universellen Approximation mit Signaturen zu erlangen.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, David Prömel, for opening the doors to life in academia for me and sharing with me his many invaluable insights and knowledge about but not limited to mathematics. With seemingly endless patience and his unassuming, empathetic way of mentoring he smoothed the path to the so educational and enjoyable one it was.

I would like to thank Paul Hager and Nicolas Perkowski for dedicating their time to co-examining this thesis.

This thesis would not have been successfully completed if it were not for my coauthors: I am thankful to Andy Allan and Chong Liu who welcomed me open heartedly into their research team and taught me rough path theory with hilarity and expertise that is deeply appreciated. I would also like to thank Andreas Neuenkirch, and especially Josef Teichmann, for inviting me for a research stay in his working group. Having had the opportunity to be part of such a vibrant environment was truly path shaping.

Financial support by the Baden-Württemberg Stiftung is gratefully acknowledged.

I am further grateful to my colleagues in the office(s) who have created such a pleasant and encouraging work environment and made this time oh so enjoyable, especially to David, Khadija, Martin, Mihriban, and Paul.

Last but not least, to all my friends: knowing to have you by my side feels like the warmest hug, thank you.
And most importantly, to my family: my parents and my brother are what keep me grounded. What their unconditional love and support means to me can not be put into words. Neither can the gratefulness I feel for Clemens for being with me every step of the way.

Contents

Introduction

Lying at the intersection of probability theory and analysis, the purpose of *stochastic analysis* is to provide a rigorous mathematical framework for describing and understanding random evolution in continuous time, for example appearing in real-world phenomena. It has become indispensable in numerous other areas of science, ranging from finance to physics or biology and data science.

In this context, aspects of approximation and stability become of interest. Inevitably, given classical models do not provide an adequate description but only an approximation of reality. One therefore tries to understand how robust they are with respect to changing the underlying model assumptions that we are usually uncertain about; this is what we refer to as *stability*.

As one would further expect, good models should be practically convenient to handle. It is thus often unavoidable to consider an *approximation* by simpler ones that are (more) tractable and analytically and/or numerically solvable, especially if a model is defined implicitly rather than explicitly.

The mathematical foundation on which we want to explore these concepts in this thesis are *rough and stochastic integrals and differential equations*, as they have been established as the fundamental tools for modeling dynamics that evolve randomly in time:

We can think of the time-ordered flow of information of a system, often available in the form of data, e.g. a financial time series, as a path $Y\colon [0,T] \to \mathbb{R}^k$, i.e., a mapping from some time interval $[0,T]$ to $\mathbb{R}^k$. The behavior of Y then is assumed to be affected by an input signal given by $X\colon [0,T] \to \mathbb{R}^d$. Formally, it is assumed to solve a differential equation of the form

$$\mathrm{d}Y_t = f(Y_t)\,\mathrm{d}X_t, \qquad t \in [0,T], \tag{1}$$

for some suitable (non-linear) function f.

When studying a system that exhibits randomness, i.e., when X is assumed to be a stochastic process, prototypically a Brownian motion, which corresponds to white noise, one needs to be careful about how to define the differential $\mathrm{d}X_t$. The classical approach of considering the derivative does not work in the usual way as it would in the case of smooth (deterministic) paths because the behavior of X is too irregular in some sense.

To get around this, the differential equation (1) is rewritten into an integral equation of the form

$$Y_t = y + \int_0^t f(Y_s)\,\mathrm{d}X_s, \qquad t \in [0,T]. \tag{2}$$

This is when the notion of stochastic integration, which has been initiated by Itô [93] and has become a fundamental pillar in stochastic analysis, comes into play. For our purposes, very briefly, let X be a Brownian motion that is defined on a suitable probability space and Y a left-continuous adapted process thereon. The *Itô integral* is then well-defined and admits an intuitive Riemann sum-type approximation:

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

where the limit is, importantly, a limit in probability taken over a sequence of partitions $\mathcal{P}^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$, of $[0,T]$ with vanishing mesh size.

Taking not the left-point but the mid-point of each partition interval in the summation corresponds to the *Stratonovich integral* [66, 153]. While the Stratonovich integral comes more natural as it satisfies classical "first order calculus", the Itô integral is preferred from a modeling perspective because it preserves the martingale property amongst other reasons. Depending on the application one has in mind, these notions of stochastic integration are typically being employed to define stochastic differential equations as in (2), see e.g. [94]. Very significant theoretical advancements are due to [115, 59, 134], and we refer to [97] for an overview of the early historical developments in stochastic integration and mathematical finance.

We point out that in order to define the stochastic integral, one has to postulate a probability space a priori. In modeling terms, this is somehow a leap of faith since the inherent probabilistic structure of the underlying process is not known. The fact that the stochastic integral is not well-posed given only one sample path of the driving signal turns out to be another pitfall from the modeling perspective, since there is usually only one time series of data available describing a particular state of the world.

This has been especially critized in financial modeling, the issues being referred to as model risk and model uncertainty, see [111]. To overcome these limitations of probabilistic modeling, many approaches in mathematical finance have been developed that (partially) discard the probabilistic structure, starting off with, e.g., [128, 13, 89].

Addressing the pathwise, i.e., "state by state", notion of stochastic integration, Föllmer introduced in his seminal paper [67] a first deterministic analog to stochastic Itô integration able to handle sample paths of, e.g., Brownian motion.

Suppose that we have a one-dimensional continuous path X, and X has quadratic variation in some sense along a sequence of partitions of $(\mathcal{P}^n)_{n\in\mathbb{N}}$ of $[0,T]$ with vanishing mesh size. Then for a twice continuously differentiable function f, the limit of left-point Riemann sums

$$\int_0^T \mathrm{D}f(X_s)\,\mathrm{d}X_s := \lim_{n\to\infty} \sum_{k=0}^{N_n-1} \mathrm{D}f(X_{t_k^n})(X_{t_{k+1}^n} - X_{t_k^n}) \tag{3}$$

exists, where $\mathrm{D}f$ denotes the gradient of f, and the integral satisfies a "pathwise Itô formula". The *Föllmer integral* has found many applications and extensions in the pathwise approach to stochastic analysis, see e.g. [39, 9, 40, 34]. Due to the fact that it is approximated by left-point Riemann sums, the Föllmer integral can be interpreted as the capital gains process which is generated by continuous-time trading. This has been vital for its success in the context of model-free approaches to mathematical finance. We refer to [68, 52, 152, 48].

Rough path theory is a popular analytical theory that provides the arguably most general pathwise notion of integration. It was initiated by Terry Lyons [129] with the purpose of rigorously understanding nonlinear systems that are described by (1) and driven by highly oscillatory signals. Since then, it has become an increasingly popular and widely applicable field of research of modern stochastic analysis, reaching into the fields of statistics, mathematical finance and data science.

At the heart of rough path theory is Lyons' significant contribution of having identified the informational structure of a path that is required to define an integral $\int f(X)\,\mathrm{d}X$ against the path, it being of finite p-variation for any $p \geq 1$. As such it extends and generalizes the notions of Riemann–Stieltjes integration ($p = 1$), Young integration ($p \in (1,2)$) [161], and Föllmer integration.

Essentially, the idea is to "enhance" the path by a suitable higher order process that postulates the value of the higher order iterated integrals of the path, thus capturing the exact information that is missing, respective to the regularity of the path.

In this thesis, we will focus primarily on the study of rough paths as a framework for the pathwise analysis of stochastic differential equations in the semimartingale setting, which corresponds to paths of finite p-variation for $p \in (2,3)$. We, therefore, consider a p-rough path $\mathbf{X} = (X, \mathbb{X})$ to be a pair of a path $X\colon [0,T] \to \mathbb{R}^d$ and its "lift" $\mathbb{X}\colon [0,T]^2 \to \mathbb{R}^{d\times d}$ satisfying the following analytical and algebraic conditions: X is of finite p-variation, $\mathbb{X}$ is of finite $\frac{p}{2}$-variation, and Chen's relation holds, i.e.,

$$\mathbb{X}_{s,t} = \mathbb{X}_{s,u} + \mathbb{X}_{u,t} + X_{s,u} \otimes X_{u,t}, \qquad \text{for} \quad 0 \leq s \leq u \leq t \leq T,$$

which mimics exactly the behavior of the iterated integral of X against itself.

Due to Gubinelli [82], the assumption on the integrands can be refined: the rough integral can be defined not only for integrands of the form $f(X)$, where f is suitably nice, but for so-called controlled paths (Y, Y'): heuristically speaking, such a Y locally looks like X. More precisely, a controlled path (Y, Y') is defined as a pair of a path $Y \colon [0,T] \to \mathbb{R}^k$ and its derivative $Y' \colon [0,T] \to \mathbb{R}^{k \times d}$, where Y is of finite p-variation, Y' is of finite q-variation and $R^Y \colon [0,T]^2 \to \mathbb{R}^k$, which is implicitly defined by

$$Y_t - Y_s = Y'_s(X_t - X_s) + R^Y_{s,t}, \qquad (s,t) \in [0,T]^2,$$

is of finite r-variation, for $q \geq p$ and $r \in [\frac{p}{2}, 2)$ such that $\frac{1}{p} + \frac{1}{r} > 1$, and $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$.

Then, the rough integral of (Y, Y') against $\mathbf{X}$ is defined by

$$\int_0^T Y_s \,\mathrm{d}\mathbf{X}_s = \lim_{|\mathcal{P}| \to 0} \sum_{[u,v] \in \mathcal{P}} (Y_u(X_v - X_u) + Y'_u \mathbb{X}_{u,v}), \tag{4}$$

where the limit is taken over any sequence of partitions $\mathcal{P}$ of $[0,T]$ with mesh size $|\mathcal{P}|$ tending to 0.

Considering the integral in (2) to be a rough integral, the rough differential equation of the form

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

is well-posed. The resulting solution map, the Itô–Lyons map

$$\mathbf{X} \mapsto Y, \tag{5}$$

turns out to be, very remarkably, continuous with respect to the driving rough path, and comes with powerful stability estimates. Therefore, rough path theory provides a robust solution theory for the pathwise study of stochastic differential equations.

In this thesis, we shall delve into various aspects of approximation and stability relevant to the dominant modeling paradigms in rough and stochastic analysis as well as in mathematical finance. Roughly speaking, Chapter 2 and 4 tend to stability, while Chapter 1, 3, 4, 5, and 6 tend to approximation.

Every chapter is relatively self-contained and may be read independently. In the following we give an outline of this thesis by summarizing the main contributions of each chapter.

Chapter 1: Universal approximation property of neural stochastic differential equations

Chapter 1 is based on joint work with David J. Prömel and Josef Teichmann, see [117].

Financial modeling in continuous time typically begins with a stochastic differential equation (SDE) of the form

$$Y_t = y + \int_0^t b(s, Y_s)\,\mathrm{d}s + \int_0^t \sigma(s, Y_s)\,\mathrm{d}X_s, \qquad t \in [0, T], \tag{6}$$

where the model parameters b and σ are calibrated to market data. In this context, neural networks have recently been successfully used to approximate, or learn, said model parameters, which extends the idea of neural ordinary differential equations, see e.g. [46, 78, 36]. These so-called neural stochastic differential equations turn out to be very capable, also as continuous-time generative models in machine learning, see e.g. [125, 123, 106].

This idea is also built on the theoretical property of neural networks to approximate any continuous function arbitrarily well on compact subsets of $\mathbb{R}^d$ and in an L^p-sense. However, these classical universal approximation theorems, stated e.g. in [50, 90], do not admit a uniform control of the global growth of the respective neural networks, which therefore poses a question about the theoretical universal approximation property of neural stochastic differential equations. We aim to address this and prove that a number of classes of neural networks are indeed capable of approximating continuous functions locally uniformly subject to a given global linear growth constraint, adapting already proven universal approximation theorems in the literature. Consequently, given that there exists a unique solution to a stochastic differential equation, it can be approximated arbitrarily well by solutions of neural stochastic differential equations in which the neural networks have the above mentioned "universal approximation property under a linear growth constraint".

Chapter 2: Functional differential equations driven by càdlàg rough paths

This chapter is based on joint work with Andreas Neuenkirch and David J. Prömel, see [116].

To model dynamics that evolve not only randomly but also depending on their past values, so-called stochastic functional differential equations, or stochastic delay differential equations, are being considered in the context of stochastic analysis, see e.g. [137, 138]. This motivates the study of the well-posedness of rough differential equations with path-dependent coefficients and driven by càdlàg rough paths that are of the form

$$Y_t = y_0 + \int_0^t F_s(Y)\,\mathrm{d}\mathbf{X}_s, \qquad t \in [0, T],$$

where $\mathbf{X}$ is a càdlàg p-rough path for $p \in (2,3)$, the initial condition (y, y') is a controlled path (with respect to X), and F is a non-anticipative functional mapping controlled paths to controlled paths.

In particular, we show the existence of unique solutions to such rough functional differential equations, and establish that the corresponding Itô–Lyons map

$$((y, y'), F, \mathbf{X}) \mapsto (Y, Y')$$

is locally Lipschitz continuous, thus touching on the aspect of stability. Both of these results rely on a Lipschitz-type condition and a quadratic growth condition on the path-dependent coefficient (F, F').

This covers typical examples of rough differential equations driven by càdlàg rough paths such as classical state dependent RDEs, see e.g. [73, 75], and controlled RDEs, see e.g. [3], and further extends to discrete time dependent RDEs and delayed RDEs, see e.g. [8, 140]. It may therefore be of use for related path-dependent problems, e.g., in control theory, non-linear filtering, and stochastic functional analysis.

To that end, the deterministic theory is then applied to stochastic differential equations with delay. Thus, the continuity of the Itô–Lyons map yields pathwise stability results for these stochastic differential equations. In particular, this allows to resolve an old observation pointed out in [136] about the non-continuity of the flow of stochastic differential equations with delay.

Chapter 3: Pathwise convergence of the Euler scheme for rough and stochastic differential equations

This chapter is based on joint work with Andrew L. Allan, Chong Liu, and David J. Prömel, see [6].

While it is well-known that the stochastic differential equation (6) driven by a semimartingale X admits a unique solution assuming that the coefficients b and σ are suitably regular, it can rarely be solved explicitly. Therefore, numerical methods are being applied to approximate the solution, see e.g. [109]; the most common being the Euler scheme, possibly of higher order. The first order Euler scheme, the so-called Euler–Maruyama scheme, for the SDE (6) is given by

$$Y_t^n = y_0 + \sum_{k\,:\,t_{k+1}^n \le t} b(t_k^n, Y_{t_k^n}^n)(t_{k+1}^n - t_k^n) + \sum_{k\,:\,t_{k+1}^n \le t} \sigma(t_k^n, Y_{t_k^n})(X_{t_{k+1}^n} - X_{t_k^n}), \qquad t \in [0,T],$$

along a sequence of partitions $\mathcal{P}^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$, of $[0,T]$, which is essentially a time-discretized modification of the equation. Higher order Euler schemes, adding higher order terms to the above approximation, converge faster but involve simulating iterated integrals of the driving signal X, which might become a numerically non-trivial task.

Since numerical calibration is carried out path by path, we tend to rough path theory, which provides a fully pathwise solution theory for SDEs. While it has provided a transparent convergence analysis of higher order Euler approximations, see e.g. [70, 75], the pathwise convergence of the first order Euler approximation of stochastic differential equations is inexplicable from the rough path perspective so far. One issue, for example, is that the rough path lift is not unique so that there possibly exist multiple solutions to the corresponding rough differential equation, but the Euler scheme, which does not depend on the rough path lift, can only converge to at most one such solution.

We clarify this gap and investigate rough differential equations that are driven by càdlàg paths satisfying the so-called Property (RIE) along suitable sequences of partitions with vanishing mesh size. Property (RIE) has been introduced in [143] and [7] for applications in mathematical finance considering model uncertainty and notably recovers the rough integral as limit of left-point Riemann sums, cf. (4).

We use this to establish a novel result on the convergence of the Euler scheme (and an approximative variant thereof) of rough differential equations.

It turns out that Property (RIE) is satified by almost all sample paths of Brownian motion, Itô processes, Lévy processes and general càdlàg semimartingales, as well as the driving signals of both so-called mixed and rough stochastic differential equations, relative to different time discretizations. It further ensures the existence of a canonical Itô-type rough path lift, which allows us to treat the SDE (6) as an RDE.

We therefore obtain a rigorous pathwise convergence analysis of the first order Euler approximation of stochastic differential equations driven by various types of stochastic noise, even outside the classical semimartingale setting.

Chapter 4: Pathwise analysis of log-optimal portfolios

This chapter is based on joint work with Andrew L. Allan, Chong Liu and David Prömel.

Classical approaches to optimal portfolio selection problems are based on probabilistic models for asset returns or prices. However, it is now widely recognized that the performance of optimal portfolios is highly sensitive to model misspecifications. To account for various types of model risk, robust and model-free approaches have gained increasing importance in portfolio theory.

In this chapter, we develop a pathwise approach to analyze stability and approximation properties of portfolios for individual price trajectories that are generated by standard models for financial markets.

For this purpose, we rely again on the theory of càdlàg rough paths, as it is a pathwise (stochastic) integration theory which offers powerful stability estimates. To demonstrate that this is an ideal tool, we study, as a prototypical example from portfolio theory, the log-optimal portfolio of a classical investment-consumption optimization problem in a frictionless financial market, modeled by an Itô diffusion process.

However, since the rough integral is defined as a limit of compensated Riemann sums, see (4), we miss the natural financial interpretation of the integral as the capital gains process which is generated by continuous-time trading. We even need to be careful when choosing the rough path lift because the wrong choice might lead to an anticipating integral, that corresponds, e.g., to Stratonovich integration, and thus introduce arbitrage.

To overcome this, when defining an integral, we again assume Property (RIE) to hold for the integrator, which, importantly, recovers the rough integral as limit of left-point, i.e., non-anticipative, Riemann sums. Moreover, in a probabilistic framework, then the rough and the stochastic Itô integral coincide almost surely when both are defined.

This allows us to identify an entirely deterministic framework for constructing the log-optimal portfolio in a pathwise manner, for local volatility models as (6) and Black–Scholes type models (see [24]). In this framework, we derive pathwise stability estimates for the log-optimal portfolio and its associated capital process with respect to the underlying model parameters, i.e., drift and volatility, accounting for model uncertainty.

Furthermore, since trading in reality is not done in continuous time but in (high-frequent) discrete time, we derive pathwise error estimates that result from the time-discretization of the log-optimal portfolio and its associated capital process.

Here we use some of the results obtained in Chapter 2 and Chapter 3.

Chapter 5: Existence of general pathwise stochastic integration

This chapter will be part of a larger joint work with Purba Das and David Prömel.

In the previous two chapters, we have considered the rough path framework assuming Property (RIE) for defining a rough integral and/or a rough differential equation. This has been proven to be a sufficient condition on the integrator for the rough integral to be approximated by left-point Riemann sums, see [143]. We now generalize this to the so-called Property γ-(RIE), for $\gamma \in [0,1]$, relative to a sequence of partitions, which implies that the rough integral is given as limit of general Riemann sums. More precisely, the rough integral is then given by

$$\int_0^t Y_s \,\mathrm{d}\mathbf{X}_s^\gamma = \lim_{n\to\infty} \sum_{k=0}^{N_n-1} (Y_{t_k^n} + \gamma(Y_{t_{k+1}^n} - Y_{t_k^n}))(X_{t_{k+1}^n\wedge t} - X_{t_k^n\wedge t}), \qquad t \in [0,T],$$

where $\mathbf{X}^\gamma = (X, \mathbb{X}^\gamma)$ denotes the canonical rough path lift for a continuous path X satisfying Property γ-(RIE). We note that for $\gamma = 0$ this corresponds to (forward) Itô-type integration, for $\gamma = \frac{1}{2}$ to Stratonovich-type integration, and for $\gamma = 1$ to backward Itô-type integration, these being the most popular choices in applications.

Since Property γ-(RIE) for $\gamma = 0$ is exact Property (RIE), we relate these to one another, depending on the parameter γ which determines the type of Riemann sum approximation. It turns out that for any γ, if a path X satisfies Property (RIE), it also satisfies Property γ-(RIE), and for $\gamma \neq \frac{1}{2}$, the reverse of the implication holds.

We shed more light on these path properties when establishing that Property γ-(RIE) is actually equivalent to imposing a certain regularity condition of the path and along the sequence of partitions, and assuming the existence of the Lévy area of the path, i.e., that

$$\lim_{n\to\infty} \sum_{k=0}^{N_n-1} (X^i_{t^n_k}(X^j_{t^n_{k+1}\wedge t} - X^j_{t^n_k\wedge t}) - X^j_{t^n_k}(X^i_{t^n_{k+1}\wedge t} - X^i_{t^n_k\wedge t})), \qquad i,j = 1,\dots,d,$$

exists, and for $\gamma \neq \frac{1}{2}$, to additionally assuming the existence of the quadratic variation of the path, i.e., that

$$\lim_{n\to\infty} \sum_{k=0}^{N_n-1} (X^i_{t^n_{k+1}\wedge t} - X^i_{t^n_k\wedge t})(X^j_{t^n_{k+1}\wedge t} - X^j_{t^n_k\wedge t}), \qquad i,j = 1,\dots,d,$$

exists. The former seems fitting in the regime of rough path theory, the latter formally establishes the link to the Föllmer integral (3). Also from a practical perspective, this seems rather natural considering that, e.g., almost all sample paths of Brownian motion satisfy these assumptions relative to sequences of partitions fulfilling mild conditions on the mesh size.

We recall that Property (RIE) is satisfied by almost all sample paths of further (continuous) stochastic processes, see in Chapter 3, which makes this notion of pathwise integration applicable to the stochastic setting, especially since the rough and stochastic Stratonovich integral coincide almost surely under Property γ-(RIE) for $\gamma = \frac{1}{2}$.

Another example of a stochastic process that does not fit into the Itô-type but notably into the Stratonovich-type setting is the fractional Brownian motion for Hurst parameter $H \in (\frac{1}{3}, \frac{1}{2})$ as almost all sample paths do not possess quadratic variation but Lévy area, e.g., relative to the sequence of dyadic partitions.

Chapter 6: Universal approximation with Itô-type signatures

This chapter is joint work with Mihriban Ceylan and David J. Prömel.

So far we have focused on the study of rough analysis in the narrower sense of rough path theory, namely enhancing a path with a suitable "second order" process. The signature of a path refers to a particular characteristic of a path that is formally defined as the enhancement of a path by not only the second order but all iterated integrals of the path against itself:

$$\int_{0<t_1<\ldots<t_n<T} \mathrm{d}X^{i_1}_{t_1} \cdots \mathrm{d}X^{i_n}_{t_n},$$

for $i_1, \ldots, i_n \in \{1, \ldots, d\}$, $n \in \mathbb{N}$, see the early works of Chen [28, 29].

In his seminal work [129], Lyons identified this collection of integrals to be, appropriately truncated, the precise information of a path that is required to guarantee continuity of the integral map and the solution map (5) for a nonlinear differential equation driven by a path of very low regularity, i.e., also for paths of finite p-variation for $p \geq 3$.

Furthermore, due to its many rich properties the signature of a path offers a way to faithfully and tractably represent the key features from highly oscillatory data. Thus recently, a plethora of data-driven methods based on the signature of a path is being developed for applications to mathematical finance, see e.g. [127, 11, 19, 100, 12, 45, 44]. These methods are fundamentally based on universal approximation theorems, which state that continuous functionals on the path space can be approximated arbitrarily well by linear functionals on the signature. For financial applications, this requires computing the signature by adopting Stratonovich integration. However, from a modeling perspective, Itô integration is typically the preferred choice of stochastic integration.

In this chapter, we introduce a notion of the signature of the path assuming Property γ-(RIE) that has been introduced in Chapter 5 and yields a unifying framework for pathwise Stratonovich-type and Itô-type integration. Extending the path by suitable quadratic variation terms, we are able to deduce a pathwise universal approximation theorem for linear functionals on the signature. This can be translated into the probabilistic setting for the Itô-signature of continuous semimartingales, making it particularly suitable for financial modeling.

Chapter 1

Universal approximation property of neural stochastic differential equations

Modeling approaches that hybridize the notion of differential equations with neural networks have recently become increasingly of interest, see [63, 30]. In particular, neural stochastic differential equations (neural SDEs) have emerged as a powerful mathematical tool for capturing complex dynamical systems that exhibit randomness, see [157, 99, 106]. Specifically, these are stochastic differential equations in which neural networks are used to parametrize the drift and diffusion coefficient, thus extending the notion of neural ordinary differential equations. Neural SDEs have been successfully applied to develop data-driven methods for modeling, learning, and generating random dynamics due to powerful training technologies. For instance, they serve as continuous-time generative models for irregular time series, see [125, 123, 106, 92], and, notably, as very tractable and universal models for financial markets, thus being of particular interest for financial engineering, see [46, 78, 36, 37, 35, 64]. In other words: neural stochastic differential equations constitute a continuous time counterpart of recurrent neural networks.

What motivates many of these applications is the key insight that neural stochastic differential equations are, at least, expected to approximate general SDEs arbitrarily well, thus providing fairly general and flexible models for stochastic processes and time series, such as recurrent neural networks approximate generic discrete dynamics. In fact, classical universal approximation theorems for neural networks, as proven, e.g., in [50, 90], state that neural networks approximate any continuous function arbitrarily well uniformly on compact subsets of $\mathbb{R}^d$ or in an L^p-sense globally on $\mathbb{R}^d$. Hence, it seems intuitively reasonable that neural SDEs would inherit the universality of neural networks, allowing them to approximate generic SDEs (under mild regularity conditions). However, classical universal approximation theorems do not guarantee any uniform control of the global growth of the involved neural networks and, therefore, do not rigorously imply a universal approximation property for the associated neural SDEs.

In this chapter, we provide a theoretical justification for the universality of neural SDEs: in Section 1.1 we identify various classes of neural networks that have the so-called "universal approximation property under a linear growth constraint", that is, are able to approximate continuous functions locally uniformly subject to a given global linear growth constraint. Exemplary classes of neural networks with this universal approximation property include single hidden layer feed-forward neural networks with linearly activating activation functions, such as logistic sigmoid and hyperbolic tangent, and deep feed-forward neural networks combining rather general activation functions with rectified linear unit (ReLU) activation functions. For the proof of these universal approximation theorems with global constraints, we rely on universal approximation theorems on weight spaces, as proven in [49], as well as on L^p-spaces, as proven in [107], and extend some of the methods of both works.

In Section 1.2 we demonstrate that the "universal approximation property under a linear growth constraint" of neural networks guarantees the universality of the associated neural SDEs. Indeed, assuming that an SDE possesses a unique solution, this solution can be approximated arbitrarily well by solutions of neural SDEs in a standard L^2-norm for stochastic processes if the involved neural networks do satisfy the "universal approximation property under a linear growth constraint". Moreover, we derive quantitative error estimates for the approximation of stochastic differential equations with coefficients that fulfill standard conditions such as Lipschitz and Hölder continuity.

This chapter is structured as follows. In Section 1.1 we derive the "universal approximation property under a linear growth constraint" for various classes of neural networks. For these, we prove in Section 1.2 that the associated neural SDEs can approximate general SDEs.

1.1 Universal approximation property under a linear growth constraint

In this section, we identify various classes of neural networks allowing for the approximation of continuous functions locally uniformly subject to a given linear growth constraint.

We start by precisely formulating the aforementioned approximation property.

The spaces $\mathbb{R}^k$ and $\mathbb{R}^{n_1\times n_2}$ are equipped with the Euclidean norm $|\cdot|$. Let $C([0,T]\times\mathbb{R}^k;\mathbb{R}^{n_1\times n_2})$ be the set of continuous functions $f\colon[0,T]\times\mathbb{R}^k\to\mathbb{R}^{n_1\times n_2}$. Given a set $K\subset[0,T]\times\mathbb{R}^k$ and $f\in C([0,T]\times\mathbb{R}^k;\mathbb{R}^{n_1\times n_2})$, we define

$$\|f\|_{\infty,K}:=\sup_{x\in K}|f(x)|.$$

Moreover, we write $C(\mathbb{R}^d;\mathbb{R}^e)$ for the space of continuous maps $f\colon\mathbb{R}^d\to\mathbb{R}^e$, $C_b^0(\mathbb{R}^d;\mathbb{R}^e)$ for the space of bounded and continuous functions $f\colon\mathbb{R}^d\to\mathbb{R}^e$, and $C^\infty(\mathbb{R}^d;\mathbb{R}^e)$ for the space of smooth functions $f\colon \mathbb{R}^d \to \mathbb{R}^e$, i.e. functions with all its derivatives up to arbitrary order being continuous.

Definition 1.1.1. *A set $\mathcal{NN} \subset C([0,T] \times \mathbb{R}^k; \mathbb{R}^{n_1 \times n_2})$ is said to have the* universal approximation property under a linear growth constraint *if the following property holds:*

For every function $f \in C([0,T] \times \mathbb{R}^k; \mathbb{R}^{n_1 \times n_2})$ with at most linear growth, i.e., there exists a constant $C_f > 0$ such that

$$|f(t,x)| \leq C_f(1+|x|), \qquad t \in [0,T],\ x \in \mathbb{R}^k,$$

for every $\varepsilon \in (0,1)$ and every compact set $K \subset \mathbb{R}^k$, there exists a function $\varphi \in \mathcal{NN}$ such that

$$\|\varphi - f\|_{\infty,[0,T]\times K} \leq \varepsilon,$$

and there exists a constant $\widetilde{C}_f > 0$, not depending on ε and K, such that

$$|\varphi(t,x)| \leq \widetilde{C}_f(1+|x|), \qquad t \in [0,T],\ x \in \mathbb{R}^k.$$

In the following four subsections we provide various classes of neural networks satisfying the universal approximation property under a linear growth constraint.

1.1.1 Linearly activating activation functions

To introduce the first class of neural networks that have the universal approximation property under a linear growth constraint, we rely on the notion of weighted spaces as introduced in [49] in the context of neural networks. To that end, we fix the weight function

$$\psi\colon \mathbb{R}^{k+1} \to (0,\infty), \qquad \psi(x) := 1 + |x|, \qquad x \in \mathbb{R}^{k+1}.$$

The pre-image $\psi^{-1}((0,r])$ is compact in $\mathbb{R}^{k+1}$, for any $r > 0$, and hence, ψ is an admissible weight function and $(\mathbb{R}^{k+1}, \psi)$ is a weighted space in the sense of [49, Section 2.1]. We further introduce the weighted norm $\|\cdot\|_{\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})}$ as

$$\|f\|_{\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})} := \sup_{x\in\mathbb{R}^{k+1}} \frac{|f(x)|}{\psi(x)},$$

for $f\colon \mathbb{R}^{k+1} \to \mathbb{R}^{n_1\times n_2}$ such that $\sup_{x\in\mathbb{R}^{k+1}} \frac{|f(x)|}{\psi(x)} < \infty$. The space $\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})$ is the weighted function space defined as the $\|\cdot\|_{\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})}$-closure of $C_b^0(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})$. Note that $\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})$ is a separable Banach space when equipped with the norm $\|\cdot\|_{\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})}$, which contains $C_b^0(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})$, whereas $C_b^0(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})$ is of course not separable with respect to the uniform norm.

Given an activation function $\rho \in C(\mathbb{R};\mathbb{R})$, a *single hidden layer (feed-forward) neural network* $\varphi\colon \mathbb{R}^{n_0} \to \mathbb{R}^{n_1 \times n_2}$ is defined by

$$\varphi(x) = \sum_{n=1}^{N} w_n \rho(a_n^\top x + b_n), \tag{1.1}$$

for $x \in \mathbb{R}^{n_0}$, where $N \in \mathbb{N}$ denotes the number of neurons, where $w_1, \ldots, w_N \in \mathbb{R}^{n_1 \times n_2}$, $a_1, \ldots, a_N \in \mathbb{R}^{n_0}$ and $b_1, \ldots, b_N \in \mathbb{R}$ denote the linear readouts, weight vectors and biases, respectively. For $\rho \in C(\mathbb{R};\mathbb{R})$, we denote by $\mathcal{NN}^{\rho}_{n_0;n_1 \times n_2}$ the set of neural networks of the form (1.1) with activation function ρ.

Following [49, Definition 4.3], an activation function $\rho \in C(\mathbb{R};\mathbb{R})$ is called *linearly activating* if $\mathcal{NN}^{\rho}_{1;1\times 1} \subseteq \mathcal{B}_\psi(\mathbb{R};\mathbb{R})$ and $\mathcal{NN}^{\rho}_{1;1\times 1}$ is dense in $\mathcal{B}_\psi(\mathbb{R};\mathbb{R})$.

Remark 1.1.2. *An activation function $\rho \in C(\mathbb{R};\mathbb{R})$ is linearly activating if it holds that $\lim_{x\to\pm\infty} \frac{|\rho(ax+b)|}{\psi(x)} = 0$ for any $a \in \mathbb{N}_0$, $b \in \mathbb{R}$, and ρ is sigmoidal, i.e., $\lim_{x\to-\infty} \rho(x) = 0$ and $\lim_{x\to\infty} \rho(x) = 1$, see [49, Proposition 4.4]. Examples include the logistic sigmoid $\rho(x) = \frac{1}{1+\exp(-x)}$ and $\rho(x) = \tanh(x)$. Other conditions for activation functions to be linearly activating are the discriminatory property or conditions on its Fourier transform, which can be found in [49, Proposition 4.4].*

For the single hidden layer neural networks $\mathcal{NN}^{\rho}_{k+1;n_1\times n_2}$ with linearly activating activation function ρ, we obtain the following universal approximation theorem allowing for given a linear growth constraint.

Theorem 1.1.3. *If the activation function $\rho \in C(\mathbb{R};\mathbb{R})$ is linearly activating, then $\mathcal{NN}^{\rho}_{k+1;n_1\times n_2}$ has the universal approximation property under a linear growth constraint in the sense of Definition 1.1.1. Moreover, the constant $\widetilde{C}_f$ in Definition 1.1.1 can be chosen to be $\widetilde{C}_f = (1+T)(1+C_f)$.*

Proof. Let $f \in C([0,T] \times \mathbb{R}^k; \mathbb{R}^{n_1\times n_2})$ be such that there exists a constant $C_f > 0$ satisfying

$$|f(t,x)| \leq C_f(1+|x|), \qquad t \in [0,T],\ x \in \mathbb{R}^k. \tag{1.2}$$

Step 1. We extend f to $\mathbb{R}^{k+1}$ by setting $f(t,x) := f(0,x)$, $t \leq 0$, and $f(t,x) := f(T,x)$, $t \geq T$. Given some function $g \in C^\infty(\mathbb{R};\mathbb{R})$ with compact support, $g\colon \mathbb{R} \to [0,1]$, $g(t) = 1$ for $t \in [0,T]$, we now consider $\tilde{f}(t,x) := f(t,x)g(t)$. Note that (1.2) holds for $\tilde{f}$, which implies that $\|\tilde{f}\|_{\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})} \leq C_f$.

Step 2. Suppose that $\varepsilon \in (0,1)$ and $K \subset \mathbb{R}^k$ is a compact set. Now there exists $\tilde{f}_{\varepsilon,K} \in \mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})$ satisfying (1.2),

$$f(t,x) = \tilde{f}_{\varepsilon,K}(t,x), \qquad t \in [0,T],\ x \in K,$$

and

$$\|\tilde{f}_{\varepsilon,K}\|_{\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})} \leq \|\tilde{f}\|_{\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})} \leq C_f.$$

More precisely, take $\tilde{g} \in C^\infty(\mathbb{R}^{k+1};\mathbb{R})$ with compact support, $\tilde{g}\colon \mathbb{R}^{k+1} \to [0,1]$, $\tilde{g}(t,x) = 1$ for $t \in [0,T]$, $x \in K$, and set $\tilde{f}_{\varepsilon,K} := \tilde{f}\tilde{g}$. Then $\tilde{f}_{\varepsilon,K} \in \mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})$.

Step 3. Let $\mathcal{H} \subseteq \mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R})$ be the additive family given by

$$\mathcal{H} = \{x \mapsto a^\top x + b \,:\, a \in \mathbb{R}^{k+1}, b \in \mathbb{R}\},$$

see [49, Definition 4.1, Example 4.2]. We note that any $\varphi \in \mathcal{NN}^\rho_{k+1;n_1\times n_2}$ is of the form

$$\varphi(x) = \sum_{n=1}^{N} w_n \rho(h_n(x)),$$

where $h_1, \ldots, h_N \in \mathcal{H}$, and $\sup_{x\in\mathbb{R}^{k+1}} \frac{\psi(h(x))}{\psi(x)} < \infty$, for all $h \in \mathcal{H}$. Then, [49, Theorem 4.13] gives that $\mathcal{NN}^\rho_{k+1;n_1\times n_2}$ is dense in $\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})$, i.e., there exists $\varphi \in \mathcal{NN}^\rho_{k+1;n_1\times n_2}$ with

$$\|\varphi - \tilde{f}_{\varepsilon,K}\|_{\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})} \leq \varepsilon \Big(\sup_{(t,x)\in[0,T]\times K} \psi((t,x)) \Big)^{-1}.$$

This implies that

$$\|\varphi - f\|_{\infty,[0,T]\times K} = \|\varphi - \tilde{f}_{\varepsilon,K}\|_{\infty,[0,T]\times K} \leq \frac{\varepsilon}{2},$$

and

$$\begin{aligned} |\varphi(t,x)| &\leq (\|\varphi - \tilde{f}_{\varepsilon,K}\|_{\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})} + \|\tilde{f}_{\varepsilon,K}\|_{\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})})\psi((t,x)) \\ &\leq (1 + C_f)(1 + T)(1 + |x|), \end{aligned}$$

for $t \in [0,T]$, $x \in \mathbb{R}^k$.

Therefore the universal approximation result on $\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})$ implies the universal approximation property under a linear growth constraint in the sense of Definition 1.1.1. □

1.1.2 Combining the ReLU activation function and a general activation function

To allow for an activation function that is not linearly activating, such as the widely used rectified linear unit (ReLU) activation function $\rho(x) := \max(x,0)$, we consider a different neural network architecture.

Let $L, N_0, \ldots, N_L \in \mathbb{N}$, and for any $l \in \{1, \ldots, L\}$, let $w_l\colon \mathbb{R}^{N_{l-1}} \to \mathbb{R}^{N_l}$, $x \mapsto A_l x + b_l$, be an affine function with $A_l \in \mathbb{R}^{N_l \times N_{l-1}}$ and $b_l \in \mathbb{R}^{N_l}$. Given an activation function $\rho \in C(\mathbb{R};\mathbb{R})$, a *deep (feed-forward) neural network* $\varphi\colon \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}$ is defined by

$$\varphi = w_L \circ \rho \circ w_{L-1} \circ \ldots \circ \rho \circ w_1,$$

where $\circ$ denotes the usual composition of functions. Here, ρ is applied componentwise, $L-1$ denotes the number of hidden layers (L is the depth of φ), and $N_1, \ldots, N_{L-1}$ denote the dimensions (widths) of the hidden layers and N_0 and N_L the dimension of the input and the output layer, respectively.

We write $\mathcal{NN}^{\rho}_{N_0;N_L}$ for the set of deep feed-forward neural networks $\varphi\colon \mathbb{R} \to \mathbb{R}$ with activation function ρ, input dimension N_0 and output dimension N_L and an arbitrary number of hidden layers L, see e.g. [46, Appendix B.1]. We then write $\mathcal{NN}^{\rho}_{n_0;n_1,n_2}$ for the set of functions $\varphi\colon \mathbb{R}^{n_0} \to \mathbb{R}^{n_1\times n_2}$ of the form $\varphi = (\varphi^{ij})_{i=1,\ldots,n_1,\, j=1,\ldots,n_2}$, where $\varphi^{ij} \in \mathcal{NN}^{\rho}_{n_0;1}$.

When allowing for two activation functions $\rho_1, \rho_2 \in C(\mathbb{R};\mathbb{R})$, we write $\mathcal{NN}^{\rho_1,\rho_2}_{N_0;N_L}$ and $\mathcal{NN}^{\rho_1,\rho_2}_{n_0;n_1,n_2}$, respectively.

Proposition 1.1.4. *If $\rho_1\colon \mathbb{R} \to \mathbb{R}$ is non-affine continuous and continuously differentiable at at least one point, with non-zero derivative at that point, and ρ_2 is the ReLU activation function, then $\mathcal{NN}^{\rho_1,\rho_2}_{k+1;n_1,n_2}$ has the universal approximation property under a linear growth constraint. Moreover, the constant $\widetilde{C}_f$ in Definition 1.1.1 can be chosen to be $\widetilde{C}_f = \sqrt{n_1 n_2}(1+T)(1+C_f)$.*

Remark 1.1.5. *The condition on ρ_1 in Proposition 1.1.4 is rather mild. For instance, it is satisfied by the frequently used activation functions, and it even includes polynomials. Furthermore, one may also consider both ρ_1 and ρ_2 to be the ReLU activation function.*

Proof of Proposition 1.1.4. We first shall prove that $\mathcal{NN}^{\rho_1,\rho_2}_{n_0;1}$ is dense in $\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})$, $n_0 \in \mathbb{N}$, i.e., for every $f \in \mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})$ and $\varepsilon > 0$ there exists some $\varphi \in \mathcal{NN}^{\rho_1,\rho_2}_{n_0;1}$ such that

$$\|\varphi - f\|_{\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})} = \sup_{x\in\mathbb{R}^{n_0}} \frac{|f(x)-\varphi(x)|}{\psi(x)} < \varepsilon. \tag{1.3}$$

In this proof, we adapt the methods of [49].

Step 1. A vector space $\mathcal{A}$ of maps $a\colon \mathbb{R}^{n_0} \to \mathbb{R}$ is called a *subalgebra* if $\mathcal{A}$ is closed under multiplication, i.e., for every $a_1, a_2 \in \mathcal{A}$, it holds that $a_1 \cdot a_2 \in \mathcal{A}$. Moreover, $\mathcal{A}$ is called *point separating* if for every distinct $x_1, x_2 \in \mathbb{R}^{n_0}$, there exists some $a \in \mathcal{A}$ with $a(x_1) \neq a(x_2)$. $\mathcal{A}$ *vanishes nowhere* if for every $x \in \mathbb{R}^{n_0}$, there exists some $a \in \mathcal{A}$ with $a(x) \neq 0$.

For a given subalgebra $\mathcal{A} \subseteq C(\mathbb{R}^{n_0};\mathbb{R})$, a vector subspace $\mathcal{W} \subseteq C(\mathbb{R}^{n_0};\mathbb{R})$ is called an $\mathcal{A}$*-submodule* if $a \cdot w \in \mathcal{W}$, for all $a \in \mathcal{A}$ and $w \in \mathcal{W}$, where $x \mapsto (a\cdot w)(x) := a(x)w(x)$.

We consider the additive family $\mathcal{H} \subseteq \mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})$ given by

$$\mathcal{H} = \{x \mapsto a^\top x + b \,:\, a \in \mathbb{R}^{n_0}, b \in \mathbb{R}\},$$

see [49, Definition 4.1, Example 4.2], and define $\mathcal{A} := \operatorname{span}(\{\cos\circ h : h \in \mathcal{H}\} \cup \{\sin\circ h : h \in \mathcal{H}\})$. It follows from [49, part (ii) of Lemma 2.7] that $\mathcal{A} \subseteq \mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})$ since $(\cos\circ h)|_K$,

$(\sin\circ h)|_K\in C(K;\mathbb{R})$, for all $h\in\mathcal{H}$ and compact subsets $K\subset\mathbb{R}$, and $\cos\circ h,\sin\circ h\in C_b^0(\mathbb{R}^{n_0};\mathbb{R})$. Moreover, we note that $\mathcal{A}$ is a subalgebra of $\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})$. Further, we define the subset $\mathcal{W}:=\{\mathbb{R}^{n_0}\ni x\mapsto a(x)y\in\mathbb{R}:a\in\mathcal{A},y\in\mathbb{R}\}\subseteq\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})$, which is a vector subspace (as $\mathcal{A}\subseteq\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})$ and $\mathbb{R}$ are both vector (sub)spaces) and an $\mathcal{A}$-submodule by definition.

Step 2. We observe that $\mathcal{A}\subseteq\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})$ vanishes nowhere as $(x\mapsto a(x):=\cos(0)=1)\in\mathcal{A}$. Moreover, $\mathcal{A}\subseteq\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})$ is point separating and consists only of bounded maps.

Hence, $\mathcal{W}$ is dense in $\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})$ by the weighted vector valued Stone–Weierstrass theorem [49, Theorem 3.8].

Step 3. In this step we show that for every $f\in C_b^0(\mathbb{R};\mathbb{R})$ and $\varepsilon>0$, there exists $\varphi\in\mathcal{NN}_{1;1}^{\rho_1,\rho_2}$ such that

$$\sup_{z\in\mathbb{R}}\frac{|\varphi(z)-f(z)|}{\psi(z)}<\varepsilon.$$

We use this result in Step 4 to show that $\mathcal{W}$ is contained in the $\|\cdot\|_{\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})}$-closure of $\mathcal{NN}_{n_0;1}^{\rho_1,\rho_2}$, which then gives (1.3).

Suppose that $f\in C_b^0(\mathbb{R};\mathbb{R})$ and $\varepsilon>0$, and define the constant $C:=\frac{\varepsilon}{3}+\sup_{z\in\mathbb{R}}|f(z)|$. Choose $r>0$ large enough such that $r\geq 3C\varepsilon^{-1}$, and set $K_r:=\psi^{-1}((0,r])$, which is a compact subset in $\mathbb{R}$. Since $\mathcal{NN}_{1;1}^{\rho_1}$ is dense in $C(\mathbb{R};\mathbb{R})$ with respect to the locally uniform norm, see [107, Proposition 3.12], there exists some $\varphi\in\mathcal{NN}_{1;1}^{\rho_1}$ such that

$$\sup_{z\in K_r}|\varphi_1(z)-f(z)|<\frac{\varepsilon}{3},$$

which implies that $|\varphi_1(z)|\leq C$ for all $z\in K_r$.

Let $g\in C_b^0(\mathbb{R};\mathbb{R})$ be the function defined by $g(z)=\min(\max(z,-C),C)$, for $z\in\mathbb{R}$. Thus $g(\varphi_1(z))=\varphi_1(z)$, for all $z\in K_r$. Then we get that

$$\sup_{z\in\mathbb{R}}\frac{|g(\varphi_1(z))-f(z)|}{\psi(z)}\leq\sup_{z\in K_r}|\varphi_1(z)-f(z)|+\sup_{z\in\mathbb{R}\setminus K_r}\frac{|g(\varphi_1(z))|}{\psi(z)}+\sup_{z\in\mathbb{R}\setminus K_r}\frac{|f(z)|}{\psi(z)}<\frac{\varepsilon}{3}+\frac{2C}{R}\leq\varepsilon.$$

We now note that $\mathbb{R}^2\ni(x,y)\mapsto\max(x,y)=\rho_2(x-y)+y$ and $\mathbb{R}^2\ni(x,y)\mapsto\min(x,y)=x-\rho_2(x-y)$. This gives that there exists $\varphi\in\mathcal{NN}_{1;1}^{\rho_1,\rho_2}$, by adding two more hidden layers, calculating

$$\varphi(z):=-\rho_2(-\rho_2(\varphi_1(z)+C)+2C)+C=\min(\max(\varphi_1(z),-C),C)=g(\varphi_1(z)).$$

Step 4. In this step we verify that $\mathcal{W}$ is contained in the $\|\cdot\|_{\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})}$-closure of $\mathcal{NN}_{n_0;1}^{\rho_1,\rho_2}$.

Suppose that $\varepsilon>0$, $h\in\mathcal{H}$, and $y\in\mathbb{R}$. We can assume without loss of generality that $y\neq 0$. Moreover, we consider the finite constant $C_h:=\sup_{x\in\mathbb{R}^{n_0}}\frac{\psi(h(x))}{\psi(x)}+1>0$. By Step 3, there exists some $\varphi\in\mathcal{NN}_{n_0;1}^{\rho_1,\rho_2}$ such that

$$\sup_{z\in\mathbb{R}}\frac{|\varphi(z)-\cos(z)|}{\psi(z)}<\frac{\varepsilon}{C_h|y|}.$$

Now, for the function $(x \mapsto w(x) := \cos(h(x))y) \in \mathcal{W}$, we define $(x \mapsto \varphi(x) := y\varphi(h(x)))$, which is an element of $\mathcal{NN}^{\rho_1,\rho_2}_{n_0;1}$.

Then we have that

$$
\begin{aligned}
\|\varphi - w\|_{\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})} &= \sup_{x\in\mathbb{R}^{N_0}} \frac{|y\varphi(h(x)) - y\cos(h(x))|}{\psi(x)} \\
&\leq |y| \sup_{x\in\mathbb{R}^{n_0}} \frac{|\varphi(h(x)) - \cos(h(x))|}{\psi(x)} \\
&\leq |y| \sup_{x\in\mathbb{R}^{n_0}} \frac{\psi(h(x))}{\psi(x)} \sup_{x\in\mathbb{R}^{n_0}} \frac{|\varphi(h(x)) - \cos(h(x))|}{\psi(h(x))} \\
&\leq C_h|y| \sup_{z\in\mathbb{R}} \frac{|\varphi(z) - \cos(z)|}{\psi(z)} \\
&< \varepsilon.
\end{aligned}
$$

Since ε was chosen arbitrarily, the map $(x \mapsto w(x) = \cos(h(x))y) \in \mathcal{W}$ belongs to the $\|\cdot\|_{\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})}$-closure of $\mathcal{NN}^{\rho_1,\rho_2}_{n_0;1}$, which holds analogously true for $(x \mapsto \sin(h(x))y) \in \mathcal{W}$. Hence, due to the trigonometric identities for the product of cosine and sine, the entire $\mathcal{A}$-submodule $\mathcal{W}$ is contained in the $\|\cdot\|_{\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})}$-closure of $\mathcal{NN}^{\rho_1,\rho_2}_{n_0;1}$.

Since $\mathcal{W}$ is dense in $\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})$ by Step 2, we obtain that $\mathcal{NN}^{\rho_1,\rho_2}_{n_0;1}$ is dense in $\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})$, that is, (1.3) does hold.

Step 5. It remains to show that (1.3) implies the universal approximation property under a linear growth constraint in the sense of Definition 1.1.1.

Let $f \in C([0,T] \times \mathbb{R}^k;\mathbb{R}^{n_1\times n_2})$ be such that there exists a constant $C_f > 0$ satisfying

$$
|f(t,x)| \leq C_f(1 + |x|), \qquad t \in [0,T],\ x \in \mathbb{R}^k. \tag{1.4}
$$

We extend f to $\mathbb{R}^{k+1}$ by setting $f(t,x) := f(0,x)$, $t \leq 0$, and $f(t,x) := f(T,x)$, $t \geq T$. Given some function $g \in C^\infty(\mathbb{R};\mathbb{R})$ with compact support, $g\colon \mathbb{R} \to [0,1]$, $g(t) = 1$ for $t \in [0,T]$, we now consider $\tilde{f}(t,x) := f(t,x)g(t)$. Note that (1.4) holds for $\tilde{f}$, which implies that $\|\tilde{f}\|_{\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})} \leq C_f$.

Step 6. Suppose that $\varepsilon \in (0,1)$ and $K \subset \mathbb{R}^k$ is a compact set. Now there exists $\tilde{f}_{\varepsilon,K} \in \mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})$ satisfying (1.4),

$$
f(t,x) = \tilde{f}_{\varepsilon,K}(t,x), \qquad t \in [0,T],\ x \in K,
$$

and

$$
\|\tilde{f}_{\varepsilon,K}\|_{\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})} \leq \|\tilde{f}\|_{\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})} \leq C_f.
$$

More precisely, take $\tilde{g} \in C^\infty(\mathbb{R}^{k+1};\mathbb{R})$ with compact support, $\tilde{g}\colon \mathbb{R}^{k+1} \to [0,1]$, $\tilde{g}(t,x) = 1$ for $t \in [0,T]$, $x \in K$, and set $\tilde{f}_{\varepsilon,K} := \tilde{f}\tilde{g}$. Then $\tilde{f}_{\varepsilon,K} \in \mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R}^{n_1\times n_2})$.

Step 7. We write $f = (f^{ij})_{i=1,\dots,n_1,\, j=1,\dots,n_2}$, similarly for $\tilde{f}_{\varepsilon,K}$, and let $\delta = \frac{\varepsilon}{\sqrt{n_1 n_2}}$. Then we infer from (1.3) that there exist $\varphi^{ij} \in \mathcal{NN}^{\rho_1,\rho_2}_{k+1;1}$, $i = 1,\dots,n_1$, $j = 1,\dots,n_2$, such that

$$\|\varphi^{ij} - \tilde{f}^{ij}_{\varepsilon,K}\|_{\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R})} \leq \delta \Big(\sup_{(t,x)\in[0,T]\times K} \psi((t,x)) \Big)^{-1}.$$

This implies that

$$\|\varphi^{ij} - f^{ij}\|_{\infty,[0,T]\times K} = \|\varphi^{ij} - \tilde{f}^{ij}_{\varepsilon,K}\|_{\infty,[0,T]\times K} \leq \delta,$$

and

$$\begin{aligned} |\varphi^{ij}(t,x)| &\leq (\|\varphi^{ij} - \tilde{f}^{ij}_{\varepsilon,K}\|_{\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R})} + \|\tilde{f}^{ij}_{\varepsilon,K}\|_{\mathcal{B}_\psi(\mathbb{R}^{k+1};\mathbb{R})})\psi((t,x)) \\ &\leq (1 + C_f)(1+T)(1+|x|), \end{aligned}$$

for $t \in [0,T]$, $x \in \mathbb{R}^k$. Therefore there exists $\varphi = (\varphi^{ij})_{i=1,\dots,n_1,\, j=1,\dots,n_2} \in \mathcal{NN}^{\rho_1,\rho_2}_{k+1;n_1,n_2}$ satisfying

$$\|\varphi - f\|_{\infty,[0,T]\times K} \leq \varepsilon, \qquad |\varphi(t,x)| \leq \sqrt{n_1 n_2}(1+C_f)(1+T)(1+|x|), \qquad t \in [0,T],\ x \in \mathbb{R}^k,$$

which concludes the proof. □

In the course of the proof of Proposition 1.1.4, we have shown a universal approximation property on the weighted space $\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})$.

Corollary 1.1.6. *If $\rho_1 \colon \mathbb{R} \to \mathbb{R}$ be non-affine continuous and continuously differentiable at at least one point, with non-zero derivative at that point, and ρ_2 be the ReLU activation function, then $\mathcal{NN}^{\rho_1,\rho_2}_{n_0;1}$ is dense in $\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})$, i.e., for every $f \in \mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})$ and $\varepsilon > 0$ there exists some $\varphi \in \mathcal{NN}^{\rho_1,\rho_2}_{n_0;1}$ such that*

$$\|f - \varphi\|_{\mathcal{B}_\psi(\mathbb{R}^{n_0};\mathbb{R})} = \sup_{x\in\mathbb{R}^{n_0}} \frac{|f(x) - \varphi(x)|}{\psi(x)} < \varepsilon.$$

Remark 1.1.7. *A universal approximation property on general weighted spaces has been proven in [49, Theorem 4.13], by lifting a universal approximation property of one-dimensional neural networks to an infinite dimensional setting. In our setting, we notice that it suffices to have an approximation property on $C^0_b(\mathbb{R};\mathbb{R})$ with respect to the weighted norm, and it is a sufficient but not necessary condition that the one-dimensional neural networks be a subset of and dense in $\mathcal{B}_\psi(\mathbb{R};\mathbb{R})$. This allows us to handle activation functions that are not linearly activating, but requires considering deep neural networks and the ReLU activation function instead of single hidden layer neural networks.*

Remark 1.1.8. *In Proposition 1.1.4 and Corollary 1.1.6, we consider* $\mathcal{NN}^{\rho_1,\rho_2}_{n_0;1}$ *to be the generally defined class of neural networks. We note however that the functions that do appear here are more precisely linear combinations of neural networks of the form*

$$\mathbb{R}^{n_0} \ni x \mapsto -\rho_2(-\rho_2(\varphi_1(h(x)) + C) + 2C) + C,$$

where $C > 0$*,* $h \in \mathcal{H} = \{\mathbb{R}^{n_0} \ni x \mapsto a^\top x + b \,:\, a \in \mathbb{R}^{n_0}, b \in \mathbb{R}\}$*, and* $\varphi_1 \in \mathcal{NN}^{\rho_1}_{1;1}$ *is a deep feed-forward neural network with activation function* ρ_1 *and fixed width.*

The assumption on ρ_1 *ensures that* $\mathcal{NN}^{\rho_1}_{1;1}$ *is dense in* $C(\mathbb{R};\mathbb{R})$ *with respect to the locally uniform norm. One may therefore relax this assumption and consider* ρ_1 *to be of the form* $\rho(x) = \sin(x) + v(x)\exp(-x)$*, for some* $v\colon \mathbb{R} \to \mathbb{R}$ *that is bounded, continuous and nowhere differentiable, so* ρ_1 *is also nowhere differentiable, see [107, Proposition 4.15].*

It is also possible to assume $\rho_1\colon \mathbb{R} \to \mathbb{R}$ *to be continuous and non-polynomial, and to consider* $\varphi_1\colon \mathbb{R} \to \mathbb{R}$ *to be a deep neural network, where each hidden layer has two neurons with the identity activation function and one neuron with activation function* ρ_1*. These, again, are dense in* $C(\mathbb{R},\mathbb{R})$ *with respect to the locally uniform norm, see [107, Proposition 4.2].*

1.1.3 The ReLU activation function

We want to further examine the universal approximation property under a linear growth constraint for deep neural networks with the ReLU activation function. We present a constructive proof leading to a slightly stronger result compared to Corollary 1.1.10 in the sense that it shows that the constant $\widetilde{C}_f$ does not depend on T, and thus allows for approximation results uniformly in time.

Proposition 1.1.9. *If* ρ *be the ReLU activation function, then* $\mathcal{NN}^{\rho}_{k+1;n_1,n_2}$ *has the universal approximation property under a linear growth constraint. Moreover, the constant* $\widetilde{C}_f$ *in Definition 1.1.1 can be chosen to be* $\widetilde{C}_f = \sqrt{n_1 n_2}(1 + C_f)$*.*

Proof. We shall prove that for any $f \in C(\mathbb{R}^{n_0};\mathbb{R})$, $n_0 \in \mathbb{N}$, for any $\delta \in (0,1)$ and $K \subset \mathbb{R}^{n_0}$ compact, there exists a neural network $\varphi \in \mathcal{NN}^{\rho}_{n_0;1}$ such that

$$\|\varphi - f\|_{\infty,K} \leq \delta \qquad \text{and} \qquad |\varphi(x)| \leq |f(x)| + \delta, \qquad x \in \mathbb{R}^{n_0}. \tag{1.5}$$

Suppose $K \subset \mathbb{R}^{n_0}$ is a compact set and $\delta \in (0,1)$. Without loss of generality, we assume that $K = \prod_{i=1}^{n_0}[a_i, b_i]$, for some $a_i, b_i \in \mathbb{R}$, $i = 1, \ldots, n_0$. Set $c > 0$ and consider $J = \prod_{i=1}^{N_0}[a_i - c, b_i + c]$.

The proof is similar in spirit to the proof of [107, Theorem 4.16]. Since $\mathcal{NN}^{\rho}_{n_0;1}$ is dense in $C(\mathbb{R}^{n_0};\mathbb{R})$ with respect to the locally uniform norm, see [107, Proposition 4.9], there exists $\varphi_1 \in \mathcal{NN}^{\rho}_{n_0;1}$ with fixed width $n_0 + 2$ such that

$$\|\varphi_1 - f\|_{\infty,J} \leq \delta. \tag{1.6}$$

We begin by extending the definition of a neuron, for sake of notation: an enhanced neuron means the composition of an affine map with the activation function ρ with another affine map, and we allow for affine combinations of enhanced neurons. In the proof of [107, Proposition 4.9] and in the following, one may use that $x \mapsto \rho(x+N)-N$ equals the identity function for N suitably large, that is, one enhanced neuron may exactly represent the identity function. This allows us, first, to record the inputs in every hidden layer (called in-register neurons) and, second, to preserve the values of the corresponding neurons in the preceding layer.

In each layer of φ_1, the first n_0 neurons are the in-register neurons, then we have the neuron which bases its computations on the in-register neurons applying ρ, and finally, the out-register neuron, which we associate the output with.

We now modify φ_1 and construct $\varphi \in \mathcal{NN}^{\rho}_{n_0;1}$, by removing the output layer and adding some more hidden layers ($3n_0+1$ to be precise) such that φ equals φ_1 on K and vanishes on $\mathbb{R}^{n_0} \setminus J$, thus (1.5) holds.

To that end, we use that two layers of two enhanced neurons each may represent the continuous piecewise affine function $U_i\colon \mathbb{R} \to \mathbb{R}$, where $U_i(x)=1$, $x \in [a_i, b_i]$, and $U_i(x)=0$, $x \in (-\infty, a_i - c] \cup [b_i + c, \infty)$, $i = 1, \ldots, n_0$, see [107, Lemma B.1].

Similarly, one layer of two enhanced neurons may represent $[0,\infty)^2 \ni (x,y) \mapsto \min(x,y)$, see [107, Lemma B.2].

By adding $2n_0$ hidden layers, we are therefore able to store the values of $U_i(x_i)$, $i = 1, \ldots, n_0$, in the in-register neurons. By adding n_0-1 hidden layers, we are able to compute and store the value of U in one of the in-register neurons, where

$$U(x) := \min_{i=1,\ldots,n_0} U_i(x_i),$$

which approximates the indicator function $\mathbf{1}_K$, mapping into $[0,1]$, with support in J, taking value 1 on K, and value 0 on $\mathbb{R}^{n_0} \setminus J$.

It further holds that $\mathbb{R}^2 \ni (x,y) \mapsto \max(x,y) = \rho(x-y)+y$ and $\mathbb{R}^2 \ni (x,y) \mapsto \min(x,y) = x - \rho(x-y)$. Therefore there exists $\varphi \in \mathcal{NN}^{\rho}_{n_0;1}$, by adding two more hidden layers and the output layer, calculating

$$\varphi := -\rho(-\rho(\varphi_1 + CU) + 2CU) + CU = \min(\max(\varphi_1, -CU), CU),$$

for some suitable constant $C>0$ depending only on f and J such that $|\varphi_1(x)| \leq C$ for any $x \in J$, see (1.6).

By definition, it holds that $U(x)=1$, $x \in K$, and $U(x)=0$, $x \in \mathbb{R}^{n_0} \setminus J$, thus we deduce that

$$\varphi(x) = \varphi_1(x), \qquad x \in K, \qquad \text{and} \qquad \varphi(x) = 0, \qquad x \in \mathbb{R}^{n_0} \setminus J. \tag{1.7}$$

It then immediately follows from (1.6) that

$$\|\varphi - f\|_{\infty,K} \leq \delta.$$

One can further verify that $|\varphi(x)| \leq |\varphi_1(x)|$, $x \in J$. Combining (1.6) and (1.7), we obtain that

$$|\varphi(x)| \leq |f(x)| + \delta, \qquad x \in \mathbb{R}^{n_0}.$$

This proves (1.5).

We now show that this implies the universal approximation property with given linear growth constraint.

Let $f \in C([0,T] \times \mathbb{R}^k; \mathbb{R}^{n_1 \times n_2})$ be such that there exists a constant $C_f > 0$ satisfying

$$|f(t,x)| \leq C_f(1 + |x|), \qquad t \in [0,T],\ x \in \mathbb{R}^k.$$

We extend f to $\mathbb{R}^{k+1}$ by setting $f(t,x) := f(0,x)$, $t \leq 0$, and $f(t,x) := f(T,x)$, $t \geq T$, and write $f = (f^{ij})_{i=1,\dots,n_1,\, j=1,\dots,n_2}$. Suppose that $K \subset \mathbb{R}^k$ is a compact set and $\varepsilon \in (0,1)$, and let $\delta = \frac{\varepsilon}{\sqrt{n_1 n_2}}$. Then we have shown that there exist $\varphi^{ij} \in \mathcal{NN}^{\rho}_{k+1;1}$, $i = 1, \dots, n_1$, $j = 1, \dots, n_2$, such that

$$\|\varphi^{ij} - f^{ij}\|_{\infty,[0,T]\times K} \leq \delta \qquad \text{and} \qquad |\varphi^{ij}(t,x)| \leq (1 + C_f)(1 + |x|), \qquad t \in [0,T],\ x \in \mathbb{R}^k.$$

This implies that there exists $\varphi = (\varphi^{ij})_{i=1,\dots,n_1,\, j=1,\dots,n_2} \in \mathcal{NN}^{\rho}_{k+1;n_1,n_2}$ satisfying

$$\|\varphi - f\|_{\infty,[0,T]\times K} \leq \varepsilon, \qquad |\varphi(t,x)| \leq \sqrt{n_1 n_2}(1 + C_f)(1 + |x|), \qquad t \in [0,T],\ x \in \mathbb{R}^k,$$

which concludes the proof. □

In the course of the proof, we have shown the following corollary, which implies the universal approximation property under a linear growth constraint.

Corollary 1.1.10. *If ρ be the ReLU activation function, then for any $f \in C(\mathbb{R}^{n_0}; \mathbb{R})$, for any $\varepsilon \in (0,1)$ and $K \subset \mathbb{R}^{n_0}$ compact, there exists a neural network $\varphi \in \mathcal{NN}^{\rho}_{n_0;1}$ such that*

$$\|\varphi - f\|_{\infty,K} \leq \varepsilon \qquad \text{and} \qquad |\varphi(x)| \leq |f(x)| + \varepsilon, \qquad x \in \mathbb{R}^{n_0}.$$

1.1.4 Two activation functions: the ReLU activation function and a squashing activation function

When assuming two activation functions in the neural network architecture, a result analogous to Proposition 1.1.9 and Corollary 1.1.10 can be achieved. For this purpose, we introduce the notion of *squashing* activation functions, i.e., monotone and sigmoidal functions, see [90]. More precisely, $\rho \in C(\mathbb{R}; \mathbb{R})$ is squashing, if ρ is monotone, $\rho\colon \mathbb{R} \to [a,b]$, for some $a, b \in \mathbb{R}$, and $\lim_{x\to-\infty} \rho(x) = a$, $\lim_{x\to\infty} \rho(x) = b$. We assume without loss of generality that $a = 0$, $b = 1$.

Proposition 1.1.11. *If $\rho_1 \in C(\mathbb{R};\mathbb{R})$ be squashing and continuous non-polynomial and continuously differentiable at at least one point, with non-zero derivative at that point, and ρ_2 be the ReLU activation function, then $\mathcal{NN}^{\rho_1,\rho_2}_{k+1;n_1,n_2}$ has the universal approximation property under a linear growth constraint. Moreover, the constant $\widetilde{C}_f$ in Definition 1.1.1 can be chosen to be $\widetilde{C}_f = \sqrt{n_1 n_2}(1 + C_f)$.*

Remark 1.1.12. *Examples for activation functions satisfying the assumptions of Proposition 1.1.11 are $\rho_1(x) = \frac{1}{1+\exp(-x)}$, $\rho_1(x) = \tanh(x)$, and $\rho_1(x) = \frac{x}{1+|x|}$.*

Remark 1.1.13. *One may relax the assumption that ρ_1 is squashing and assume that $\rho_1 \in C(\mathbb{R};\mathbb{R})$ be monotone and have one limit, either left or right. Then there exists $\tilde{\rho}_1 \in C(\mathbb{R};\mathbb{R})$ that is squashing, given as a composition of an affine map with ρ_1 with another affine map and ρ_1. This would allow to consider, e.g., $\rho_1(x) = \ln(1 + \exp(x))$.*

Proof. We shall prove that for any $f \in C(\mathbb{R}^{n_0};\mathbb{R})$, $n_0 \in \mathbb{N}$, for any $\delta \in (0,1)$ and $K \subset \mathbb{R}^{n_0}$ compact, there exists a neural network $\varphi \in \mathcal{NN}^{\rho_1,\rho_2}_{n_0;1}$ such that

$$\|\varphi - f\|_{\infty,K} \leq \delta \qquad \text{and} \qquad |\varphi(x)| \leq |f(x)| + \delta, \qquad x \in \mathbb{R}^{n_0}. \tag{1.8}$$

Suppose $K \subset \mathbb{R}^{n_0}$ is a compact set and $\delta \in (0,1)$. Without loss of generality, we assume that $K = \prod_{i=1}^{n_0}[a_i, b_i]$, for some $a_i, b_i \in \mathbb{R}$, $i = 1, \ldots, n_0$. Set $c > 0$ and consider $J = \prod_{i=1}^{N_0}[a_i - c, b_i + c]$.

We follow the constructive proof of Proposition 1.1.9. Since ρ_1 is assumed to be continuous non-polynomial and continuously differentiable at at least one point, with non-zero derivative at that point, $\mathcal{NN}^{\rho_1}_{n_0;1}$ is dense in $C(\mathbb{R}^{n_0};\mathbb{R})$ with respect to the locally uniform norm, see [107, Proposition 4.9]. That is, there exists $\varphi_1 \in \mathcal{NN}^{\rho_1}_{n_0;1}$ (allowing the identity function in the output layer) with fixed width $n_0 + 2$ such that

$$\|\varphi_1 - f\|_{\infty,J} \leq \delta. \tag{1.9}$$

(We note that ρ_1 may be replaced with ρ_2.) We begin by extending the definition of a neuron, for sake of notation: an enhanced neuron means the composition of an affine map with the activation function (here, ρ_2) with another affine map, and we allow for affine combinations of enhanced neurons. In the proof of [107, Proposition 4.9] and in the following, one uses that $x \mapsto \rho_2(x + N) - N$ equals the identity function for N suitably large, that is, one enhanced neuron may exactly represent the identity function. This allows us , first, to record the inputs in every hidden layer (called in-register neurons) and, second, to preserve the values of the corresponding neurons in the preceding layer.

In each layer of φ, the first n_0 neurons are the in-register neurons, then we have the neuron which bases its computations on the in-register neurons applying the activation function, and finally, we have the out-register neuron, which we associate the output with.

We now modify φ_1 and construct $\varphi \in \mathcal{NN}^{\rho_1,\rho_2}_{n_0;1}$, by removing the output layer and adding some more hidden layers such that φ equals φ_1 on K and vanishes on $\mathbb{R}^{n_0} \setminus J$, thus (1.8) holds.

We consider $\zeta > 0$ and set $\eta = \frac{1-\zeta}{2(n_0-1)+3}$. Then there exists some threshold $C_\eta > 0$ such that

$$\rho_1(x) \in [0, \eta), \qquad x \leq -C_\eta, \qquad \text{and} \qquad \rho_1(x) \in (1-\eta, 1], \qquad x \geq C_\eta.$$

We aim to find a neural representation of $\varphi_0 \colon \mathbb{R} \to [0,1]$ which takes values

$$\varphi_0(x) \in (1-\eta, 1], \qquad x \in K, \qquad \text{and} \qquad \varphi_0(x) \in [0, \eta), \qquad x \in \mathbb{R}^{n_0} \setminus J, \tag{1.10}$$

using activation function ρ_1, and store the value of $\varphi_1(x)$ in one of the in-register neurons. Then we use that two layers of two enhanced neurons each, now using activation function ReLU, ρ_2, may represent the continuous piecewise affine function $U \colon \mathbb{R} \to \mathbb{R}$, where

$$U(x) = 1, \qquad x \in [1-\eta, 1], \qquad \text{and} \qquad U(x) = 0, \qquad x \in (-\infty, \eta] \cup [2(1-\eta), \infty),$$

see [107, Lemma B.1], noting that $\eta < 1 - \eta < 1 < 2(1-\eta)$.

We are therefore able to compute and store the value of $U_1(x)$ in one of the in-register neurons, where

$$U_1(x) := 1, \qquad x \in K, \qquad \text{and} \qquad U_1(x) := 0, \qquad x \in \mathbb{R}^{n_0} \setminus J,$$

which approximates the indicator function $\mathbf{1}_K$.

We proceed as in the proof of Proposition 1.1.9: we add two more hidden layers and the output layer, with ρ_2, calculating

$$\varphi(x) = \min(\max(\varphi_1, -CU_1), CU_1),$$

for some suitable constant $C > 0$ depending only on f and J such that $|\varphi_1(x)| \leq C$ for any $x \in J$, see (1.9). It then follows that there exists $\varphi \in \mathcal{NN}^{\rho_1,\rho_2}_{n_0;1}$ which satisfies (1.8) for $\frac{\delta}{2}$.

The rest can be proven following the last paragraph in the proof of Proposition 1.1.9 verbatim.

It remains to show (1.10). We make use of the squashing property of ρ_1 and get that one layer of two enhanced neurons may represent the function $h_i \colon \mathbb{R} \to [-1, 1]$ that satisfies

$$h_i(x) \in (1-2\eta, 1], \quad x \in [a_i, b_i], \quad \text{and} \quad h(x) \in (-\eta, \eta), \quad x \in (-\infty, a_i - c] \cup [b_i + c, \infty),$$

namely

$$h_i(x) = \rho_1(c_1(2x + c - 2a_i)) - \rho_1(c_1(2x - c - 2b_i)),$$

where $c_1 = \frac{C_\eta}{c}$.

We modify h_i by $\tilde{h}_i \colon \mathbb{R} \to [2\eta - 2, 2\eta]$, $x \mapsto h_i(x) - (1 - 2\eta)$, and it holds that

$$\tilde{h}_i(x) \in (0, 2\eta], \quad x \in [a_i, b_i], \qquad \tilde{h}_i(x) \in (-(1-\eta), -(1-3\eta)), \quad x \in (-\infty, a_i - c] \cup [b_i + c, \infty).$$

This implies that

$$\sum_{i=1}^{n_0} \tilde{h}_i(x_i) \in (0, \infty), \qquad x \in K, \qquad \text{and} \qquad \sum_{i=1}^{n_0} \tilde{h}_i(x_i) \in (-\infty, -\zeta), \qquad x \in \mathbb{R}^{n_0} \setminus J,$$

because if $x \in \mathbb{R}^{n_0} \setminus J$, there exists i such that $x_i \in (-\infty, a_i - c) \cup (b_i + c, \infty)$, that is, $\sum_{i=1}^{n_0} \tilde{h}_i(x_i) \leq 2\eta(n_0 - 1) - (1 - 3\eta) = -\zeta$. Lastly, since

$$\rho_1(c_2(2x + \zeta)) \in [0, \eta), \qquad x \leq -\zeta, \qquad \text{and} \qquad \rho_1(c_2(2x + \zeta)) \in (1 - \eta, 1], \qquad x \geq 0,$$

for $c_2 = \frac{C_\eta}{\zeta}$, we consider

$$\varphi_0(x) = \rho_1\Big(c_2\Big(2\sum_{i=1}^{n_0} \tilde{h}_i(x_i) + \zeta\Big)\Big),$$

which gives (1.10). □

1.2 Universal approximation property of neural SDEs

In this section, we derive a universal approximation property of neural stochastic differential equations (neural SDEs) assuming that the involved neural networks satisfy the universal approximation property under a linear growth constraint in the sense of Definition 1.1.1. We start by introducing the probabilistic framework.

Let $T > 0$ be a fixed finite time horizon and let W be a d-dimensional Brownian motion, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)_{t \in [0,T]}$ satisfying the usual conditions, i.e., completeness and right-continuity. Throughout this section, we consider the stochastic differential equation

$$X_t = x_0 + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}W_s, \qquad t \in [0, T], \tag{1.11}$$

where $x_0 \in \mathbb{R}^k$, $b \colon [0, T] \times \mathbb{R}^k \to \mathbb{R}^k$, $\sigma \colon [0, T] \times \mathbb{R}^k \to \mathbb{R}^{k \times d}$ are continuous functions, and $\int_0^t \sigma(s, X_s)\,\mathrm{d}W_s$ is defined as an Itô integral. For a comprehensive introduction to stochastic Itô integration and stochastic differential equations we refer, e.g., to the textbook [104]. Moreover, we make the following assumption.

Assumption 1.2.1. *Let* $b\colon[0,T]\times\mathbb{R}^k\to\mathbb{R}^k$ *and* $\sigma\colon[0,T]\times\mathbb{R}^k\to\mathbb{R}^{k\times d}$ *be continuous functions such that*

$$|b(t,x)|+|\sigma(t,x)|\leq C_{b,\sigma}(1+|x|),\qquad t\in[0,T],\ x\in\mathbb{R}^k,$$

for some constant $C_{b,\sigma}>0$*.*

In order to approximate the general SDE (1.11), we consider sets $\mathcal{NN}_1\subset C([0,T]\times\mathbb{R}^k;\mathbb{R}^k)$ and $\mathcal{NN}_2\subset C([0,T]\times\mathbb{R}^k;\mathbb{R}^{k\times d})$ having the universal approximation property under a linear growth constraint. For $b_\varepsilon\in\mathcal{NN}_1$ and $\sigma_\varepsilon\in\mathcal{NN}_2$, the associated neural SDE is defined as

$$X_t^\varepsilon=x_0+\int_0^t b_\varepsilon(s,X_s^\varepsilon)\,\mathrm{d}s+\int_0^t\sigma_\varepsilon(s,X_s^\varepsilon)\,\mathrm{d}W_s,\qquad t\in[0,T].\tag{1.12}$$

To ensure the existence of a unique solution X^ε to the neural SDE (1.12), it is sufficient that b_ε and σ_ε are Lipschitz continuous with at most linear growth. Let $\mathrm{Lip}([0,T]\times\mathbb{R}^k;\mathbb{R}^{n_1\times n_2})$ be the set of Lipschitz continuous functions $f\colon[0,T]\times\mathbb{R}^k\to\mathbb{R}^{n_1\times n_2}$.

Remark 1.2.2. *The Lipschitz assumption on the neural networks is immediately satisfied if the underlying activation functions are Lipschitz continuous. Many frequently used activation functions are, indeed, Lipschitz continuous functions, including ReLU, hyperbolic tangent, softsign, softplus and sigmoidal activation functions.*

Combining the universal approximation property under a linear growth constraint and [101, Theorem A], we obtain the following universal approximation property of neural SDEs.

Lemma 1.2.3. *Suppose Assumption 1.2.1 and that pathwise uniqueness holds for the SDE* (1.11)*. Moreover, suppose that* $\mathcal{NN}_1\subset\mathrm{Lip}([0,T]\times\mathbb{R}^k;\mathbb{R}^k)$ *and* $\mathcal{NN}_2\subset\mathrm{Lip}([0,T]\times\mathbb{R}^d;\mathbb{R}^{k\times d})$ *have the universal approximation property under a linear growth constraint in the sense of Definition 1.1.1. Let* $K\subset\mathbb{R}^k$ *be a compact set. Then for every* $\varepsilon>0$*, there exist* $b_\varepsilon\in\mathcal{NN}_1$ *and* $\sigma_\varepsilon\in\mathcal{NN}_2$ *such that*

$$\sup_{x_0\in K}\mathbb{E}\bigg[\sup_{t\in[0,T]}|X_t^{\varepsilon,x_0}-X_t^{x_0}|^2\bigg]\leq\varepsilon,$$

where X^{x_0} *and* X^{ε,x_0} *are the solutions to the SDE* (1.11) *and the neural SDE* (1.12)*, with initial value* x_0*, respectively.*

Remark 1.2.4. *The uniform linear growth condition, as required in the definition of the universal approximation property under a linear growth constraint (Definition 1.1.1), is a necessary condition for most approximation and stability results for stochastic differential equations, cf. [154, 74, 131]. For instance, assuming that the involved neural networks are real analytic, the flow of the associated neural stochastic differential equations is real analytic as well and can be used to approximate the flow of fairly general SDEs, see [91, 74].*

Even though we do not apply the full strength of approximation in weighted spaces in this section, we still want to point out that, in contrast to Lemma 1.2.3, we can actually obtain a global approximation result of quantitative nature for the solutions of differential equations and their flows using weighted norms for the involved coefficients. We do only show it in the case of ordinary differential equations here and leave further investigations on weighted spaces and stochastic differential equations to future research.

Lemma 1.2.5. *Let $V_i\colon \mathbb{R}^k \to \mathbb{R}^k$, $i = 1, 2$, be two L-Lipschitz continuous vector fields of at most linear growth, i.e., there exists a constant $L > 0$ such that*

$$|V_i(x) - V_i(y)| \leq L|x - y| \quad \textit{and} \quad |V_i(x)| \leq L(1 + |x|),$$

for $x, y \in \mathbb{R}^k$. Let $\varepsilon > 0$ and suppose that $\|V_1 - V_2\|_{\mathcal{B}_\psi(\mathbb{R}^k;\mathbb{R}^k)} \leq \varepsilon$ with $\psi(x) := 1 + |x|$. Denote by $X^i(x)$ the solution of

$$X^i_t(x) = x + \int_0^t V_i(X^i_s(x))\,\mathrm{d}s, \qquad t \in [0, T],$$

with $X^i_0(x) = x$ for $i = 1, 2$ and $x \in \mathbb{R}^k$. Then, for every $T > 0$ there is a constant $C > 0$ such that

$$\sup_{t \in [0,T]} |X^1_t(x) - X^2_t(x)| \leq 2\varepsilon \max(1, LT) \exp(2LT) T \psi(x)$$

for all $x \in \mathbb{R}^k$.

Proof. We can write

$$\begin{aligned}
X^1_t(x) - X^2_t(x) &= \int_0^t V_1(X^1_s(x))\,\mathrm{d}s - \int_0^t V_2(X^2_s(x))\,\mathrm{d}s \\
&= \int_0^t (V_1(X^1_s(x)) - V_1(X^2_s(x)))\,\mathrm{d}s + \int_0^t (V_1(X^2_s(x)) - V_2(X^2_s(x)))\,\mathrm{d}s \\
&= \int_0^t \int_0^1 \nabla V_1(X^1_s(x) + \theta(X^1_s(x) - X^2_s(x)))\,\mathrm{d}\theta \cdot (X^1_s(x) - X^2_s(x))\,\mathrm{d}s \\
&\quad + \int_0^t \frac{V_1(X^2_s(x)) - V_2(X^2_s(x))}{\psi(X^2_s(x))} \psi(X^2_s(x))\,\mathrm{d}s
\end{aligned}$$

for all $x \in \mathbb{R}^k$ and $t \in [0, T]$. Recall that, since X^2 is the solution of an ordinary differential equation with coefficient of at most linear growth, a straightforward application of Gronwall's inequality yields

$$|X^2_t(x)| \leq \max(1, Lt)(1 + |x|) \exp(Lt), \qquad t \in [0, T],\ x \in \mathbb{R}^k.$$

Hence, we obtain

$$|X^1_t(x) - X^2_t(x)| \leq L \int_0^t |X^1_s(x) - X^2_s(x)|\,\mathrm{d}s + \varepsilon 2 \max(1, Lt) \exp(Lt) \int_0^t \psi(x)\,\mathrm{d}s$$

for all $x \in \mathbb{R}^k$ and $t \in [0, T]$, which allows to conclude the claimed lemma by Gronwall's inequality. $\square$

The universal approximation property provided in Lemma 1.2.3 ensures that general SDEs can be approximated arbitrary well by neural SDEs, assuming the corresponding neural networks satisfy the universal approximation property under a linear growth constraint. In the following subsection we deduce quantitative versions of these approximation results.

1.2.1 Quantitative approximation results for SDEs with Lipschitz continuous coefficients

For SDEs with Lipschitz continuous coefficients, we obtain the following quantitative approximation property of neural SDEs.

Proposition 1.2.6. *Let $p \geq 2$, suppose that Assumption 1.2.1 holds and that the coefficients b, σ of the SDE* (1.11) *satisfy*

$$|b(t,x)-b(t,y)|+|\sigma(t,x)-\sigma(t,y)|\leq L_{b,\sigma}|x-y|, \qquad t \in [0,T],\ x,y \in \mathbb{R}^k,$$

for some constant $L_{b,\sigma} > 0$. Moreover, assume that $\mathcal{NN}_1 \subset \mathrm{Lip}([0,T] \times \mathbb{R}^k; \mathbb{R}^k)$ and $\mathcal{NN}_2 \subset \mathrm{Lip}([0,T] \times \mathbb{R}^k; \mathbb{R}^{k\times d})$ have the universal approximation property under a linear growth constraint in the sense of Definition 1.1.1. Then for every $\varepsilon > 0$, there exist $b_\varepsilon \in \mathcal{NN}_1$ and $\sigma_\varepsilon \in \mathcal{NN}_2$ satisfying

$$\|b^\varepsilon - b\|_{\infty,[0,T]\times K}+\|\sigma^\varepsilon - \sigma\|_{\infty,[0,T]\times K}\leq \delta,$$

where

$$\delta^p := \frac{\varepsilon}{2C}\exp(-CL_{b,\sigma}^2) \qquad \textit{with} \qquad C := 2^{2(p-1)}T^{\frac{p}{2}}\Big(T^{\frac{p}{2}} + \Big(\frac{p^3}{2(p-1)}\Big)^{\frac{p}{2}}\Big),$$

and

$$K := \{x \in \mathbb{R}^k \ :\ |x|^p\leq r\} \qquad \textit{with} \qquad r := \frac{2^{2p}}{\varepsilon}(1+3^{2p-1}|x_0|^{2p})(\exp(\tilde{a})+\exp(a)),$$

where

$$\tilde{a} := 6^{2p-1}\widetilde{C}_{b,\sigma}^{2p}T^p\Big(T^p + \frac{(2p)^{3p}}{2^p(2p-1)^p}\Big) \qquad \textit{and} \qquad a := 6^{2p-1}C_{b,\sigma}^{2p}T^p\Big(T^p + \frac{(2p)^{3p}}{2^p(2p-1)^p}\Big),$$

where $\widetilde{C}_{b,\sigma} = \max(\widetilde{C}_b, \widetilde{C}_\sigma)$, and $\widetilde{C}_b$ and $\widetilde{C}_\sigma$ are given in Definition 1.1.1, such that

$$\mathbb{E}\Big[\sup_{t\in[0,T]}|X_t^\varepsilon - X_t|^p\Big] \leq \varepsilon,$$

where X and X^ε are the solutions to the SDE (1.11) *and the neural SDE* (1.12)*, respectively.*

Proof. First note that for any stochastic process $(Z_t)_{t\in[0,T]}$ and any stopping time $\tau \le T$, we have that

$$\mathbb{E}\Big[\sup_{t\in[0,T]}|Z_t|^p\Big] \le \bigg(\mathbb{E}\Big[\Big(\sup_{t\in[0,T]}|Z_t|^p\Big)^2\Big]\mathbb{P}(\tau<T)\bigg)^{\frac{1}{2}} + \mathbb{E}\Big[\sup_{t\in[0,T]}|Z_{t\wedge\tau}|^p\Big].$$

Fixing $\varepsilon > 0$, and setting $Z := X^\varepsilon - X$, we aim to bound each of the summands on the right-hand side by $\frac{\varepsilon}{2}$.

By the estimate given in [131, Chapter II, Theorem 4.4], we can bound

$$\mathbb{E}\Big[\Big(\sup_{t\in[0,T]}|X_t^\varepsilon - X_t|^p\Big)^2\Big] \le 2^{2p-1}\Big(\mathbb{E}\Big[\sup_{t\in[0,T]}|X_t^\varepsilon|^{2p}\Big] + \mathbb{E}\Big[\sup_{t\in[0,T]}|X_t|^{2p}\Big]\Big) \le \frac{\varepsilon}{2}r.$$

Markov's inequality then implies that

$$\mathbb{P}(\tau<T) \le \frac{\varepsilon}{2}r^{-1},$$

for the stopping time $\tau := \inf\{t\ge 0 : X_t^\varepsilon \notin K\}\wedge T$.

Moreover, by Jensen's inequality and [131, Chapter I, Theorem 7.2], we obtain that

$$\begin{aligned}
&\mathbb{E}[\sup_{t\in[0,u]}|X^\varepsilon_{t\wedge\tau} - X_{t\wedge\tau}|^p]\\
&\quad\le 2^{p-1}\mathbb{E}\Big[\Big(\int_0^{u\wedge\tau}|b_\varepsilon(s,X_s^\varepsilon) - b(s,X_s)|\,\mathrm{d}s\Big)^p\Big]\\
&\qquad + 2^{p-1}\mathbb{E}\Big[\Big(\sup_{t\in[0,u]}\Big|\int_0^{u\wedge\tau}(\sigma_\varepsilon(s,X_s^\varepsilon) - \sigma(s,X_s))\,\mathrm{d}W_s\Big|\Big)^p\Big]\\
&\quad\le (2T)^{p-1}\mathbb{E}\Big[\int_0^{u\wedge\tau}|b_\varepsilon(s,X_s^\varepsilon) - b(s,X_s)|^p\,\mathrm{d}s\Big]\\
&\qquad + 2^{p-1}T^{\frac{p-2}{2}}\Big(\frac{p^3}{2(p-1)}\Big)^{\frac{p}{2}}\mathbb{E}\Big[\int_0^{u\wedge\tau}|\sigma_\varepsilon(s,X_s^\varepsilon) - \sigma(s,X_s)|^p\,\mathrm{d}s\Big]\\
&\quad\le 2^{2(p-1)}T^{p-1}\bigg(\mathbb{E}\Big[\int_0^{u\wedge\tau}|b_\varepsilon(s,X_s^\varepsilon) - b(s,X_s^\varepsilon)|^p\,\mathrm{d}s\Big] + \mathbb{E}\Big[\int_0^{r\wedge\tau}|b(s,X_s^\varepsilon) - b(s,X_s)|^p\,\mathrm{d}s\Big]\bigg)\\
&\qquad + 2^{2(p-1)}T^{\frac{p-2}{2}}\Big(\frac{p^3}{2(p-1)}\Big)^{\frac{p}{2}}\Big(\mathbb{E}\Big[\int_0^{u\wedge\tau}|\sigma_\varepsilon(s,X_s^\varepsilon) - \sigma(s,X_s^\varepsilon)|^p\,\mathrm{d}s\Big]\\
&\qquad + \mathbb{E}\Big[\int_0^{u\wedge\tau}|\sigma(s,X_s^\varepsilon) - \sigma(s,X_s)|^p\,\mathrm{d}s\Big]\Big)\\
&\quad\le \delta^p 2^{2(p-1)}T^{\frac{p}{2}}\Big(T^{\frac{p}{2}} + \Big(\frac{p^3}{2(p-1)}\Big)^{\frac{p}{2}}\Big)\\
&\qquad + 2^{2(p-1)}T^{\frac{p-2}{2}}\Big(T^{\frac{p}{2}} + \Big(\frac{p^3}{2(p-1)}\Big)^{\frac{p}{2}}\Big)L^p_{b,\sigma}\int_0^u \mathbb{E}[\sup_{v\in[0,s]}|X_v^\varepsilon - X_v|^p]\,\mathrm{d}s,
\end{aligned}$$

for any $u\in[0,T]$. By Grönwall's inequality, it then holds that

$$\mathbb{E}\Big[\sup_{t\in[0,T]}|X^\varepsilon_{t\wedge\tau} - X_{t\wedge\tau}|^p\Big] \le \frac{\varepsilon}{2}$$

since we have chosen δ accordingly. Combining the estimates thus concludes the proof. □

1.2.2 Quantitative approximation results for SDEs with Hölder continuous diffusion coefficient

In the one-dimensional case, the Lipschitz assumption on the diffusion coefficient σ in the SDE (1.11) can be relaxed to Hölder continuity, leading to the following quantitative approximation property of neural SDEs.

Proposition 1.2.7. *Let $k = d = 1$, suppose that Assumption 1.2.1 holds and that the coefficients b, σ of the SDE* (1.11) *satisfy*

$$|b(t,x) - b(t,y)| \leq L_{b,\sigma}|x-y| \qquad \text{and} \qquad |\sigma(t,x) - \sigma(t,y)| \leq L_{b,\sigma}|x-y|^{\gamma},$$

for all $x, y \in \mathbb{R}^k$, $t \in [0,T]$ and for some constant $L_{b,\sigma} > 0$ with $\gamma \in [\frac{1}{2}, 1]$. Moreover, assume that $\mathcal{NN} \subset \mathrm{Lip}([0,T] \times \mathbb{R}; \mathbb{R})$ has the universal approximation property under a linear growth constraint in the sense of Definition 1.1.1. Then for every $\varepsilon > 0$, there exist $b_\varepsilon, \sigma_\varepsilon \in \mathcal{NN}$ satisfying

$$\|b^\varepsilon - b\|_{\infty,[0,T]\times K} + \|\sigma^\varepsilon - \sigma\|_{\infty,[0,T]\times K} \leq \delta,$$

where $\alpha > 1$, $\beta > 0$ and $\delta \in (0,1)$ with

$$\Big(\beta + \delta T + \frac{2\alpha}{\beta \log(\alpha)}\delta^2 T + \frac{2\alpha}{\log(\alpha)}\beta^{2\gamma-1}L_{b,\sigma}^2 T\Big)\exp(L_{b,\sigma}T) \leq \frac{\varepsilon}{2},$$

and

$$K := [-r, r] \qquad \text{with} \qquad r := \frac{4}{\varepsilon}(1 + 3|x_0|^2)(\exp((24T + 6T^2)\widetilde{C}_{b,\sigma}^2) + \exp((24T + 6T^2)C_{b,\sigma}^2)),$$

where $\widetilde{C}_{b,\sigma}$ is given in Definition 1.1.1, such that

$$\sup_{t\in[0,T]} \mathbb{E}[|X_t^\varepsilon - X_t|] \leq \varepsilon,$$

where X and X^ε are the solutions to the SDE (1.11) *and the neural SDE* (1.12)*, respectively.*

Proof. First note that for any stochastic process $(Z_t)_{t\in[0,T]}$ and any stopping time $\tau \leq T$, we have that

$$\mathbb{E}[|Z_t|] \leq (\mathbb{E}[|Z_t|^2]\mathbb{P}(\tau < T))^{\frac{1}{2}} + \mathbb{E}[|Z_{t\wedge\tau}|].$$

Fixing $\varepsilon > 0$, $t \in [0,T]$, and setting $Z = X^\varepsilon - X$, we aim to bound each of the summands on the right-hand side by $\frac{\varepsilon}{2}$.

By the estimate given in [131, Chapter II, Theorem 4.4], we can bound

$$\mathbb{E}[|X_t^\varepsilon - X_t|^2] \leq 2\Big(\mathbb{E}\Big[\sup_{t\in[0,T]}|X_t^\varepsilon|^2\Big] + \mathbb{E}\Big[\sup_{t\in[0,T]}|X_t|^2\Big]\Big) \leq \frac{\varepsilon}{2}r,$$

Markov's inequality then implies that

$$\mathbb{P}(\tau < T) \leq \frac{\varepsilon}{2} r^{-1},$$

for the stopping time $\tau := \inf\{t \geq 0 : X_t^\varepsilon \notin K\} \wedge T$.

We apply the idea of [160] to approximate $x \mapsto |x|$, see also [85]. There exists $h \in C^2(\mathbb{R})$ such that $|x| \leq \beta + h(x)$, $|h'(x)| \leq 1$, and $h''(x) \leq \frac{2}{|x|\log(\alpha)} \mathbf{1}_{[\frac{\beta}{\alpha},\beta]}(x)$. By Itô's formula, we then obtain that

$$\begin{aligned}
|X_{t\wedge\tau}^\varepsilon - X_{t\wedge\tau}| &\leq \beta + \int_0^{t\wedge\tau} h'(X_s^\varepsilon - X_s)(b^\varepsilon(s, X_s^\varepsilon) - b(s, X_s))\,\mathrm{d}s \\
&\quad + \frac{1}{2}\int_0^{t\wedge\tau} h''(X_s^\varepsilon - X_s)(\sigma^\varepsilon(s, X_s^\varepsilon) - \sigma_s(s, X_s))^2\,\mathrm{d}s \\
&\quad + \int_0^{t\wedge\tau} h'(X_s^\varepsilon - X_s)(\sigma^\varepsilon(s, X_s^\varepsilon) - \sigma(s, X_s))\,\mathrm{d}W_s \\
&\leq \beta + \delta T + L_{b,\sigma}\int_0^{t\wedge\tau} |X_s^\varepsilon - X_s|\,\mathrm{d}s \\
&\quad + \frac{2\alpha}{\beta\log(\alpha)}\delta^2 T + \frac{2\alpha}{\beta\log(\alpha)}\beta^{2\gamma-1}L_{b,\sigma}^2 T \\
&\quad + \int_0^t h'(X_s^\varepsilon - X_s)(\sigma^\varepsilon(s, X_s^\varepsilon) - \sigma(s, X_s))\,\mathrm{d}W_s,
\end{aligned}$$

for any $t \in [0, T]$. Let $M_t := \int_0^{t\wedge\tau} h'(X_s^\varepsilon - X_s)(\sigma^\varepsilon(s, X_s^\varepsilon) - \sigma(s, X_s))\,\mathrm{d}W_s$, $t \in [0, T]$. Since σ and σ_ε are of linear growth and there exists some constant $C_0 > 0$ depending only on C_σ, $\widetilde{C}_\sigma$, x_0 and T such that

$$\mathbb{E}[|X_t|^2] + \mathbb{E}[|X_t^\varepsilon|^2] \leq C_0^2, \qquad t \in [0, T],$$

see e.g. [131, Chapter II, Corollary 4.6], it holds that $\mathbb{E}[[M]_t] < \infty$ for any $t \in [0, T]$, where $[M]$ denotes the quadratic variation of M. Hence, by [147, Chapter II.6, Corollary 3], M is a martingale. It follows that

$$\begin{aligned}
&\mathbb{E}[|X_{t\wedge\tau}^\varepsilon - X_{t\wedge\tau}|] \\
&\quad \leq \beta + \delta T + \frac{2\alpha}{\beta\log(\alpha)}\delta^2 T + \frac{2\alpha}{\log(\alpha)}\beta^{2\gamma-1}L_{b,\sigma}^2 T + L_{b,\sigma}\int_0^{t\wedge\tau} |X_s^\varepsilon - X_s|\,\mathrm{d}s,
\end{aligned}$$

and thus Grönwall's inequality yields that

$$\mathbb{E}[|X_{t\wedge\tau}^\varepsilon - X_{t\wedge\tau}|] \leq \left(\beta + \delta T + \frac{2\alpha}{\beta\log(\alpha)}\delta^2 T + \frac{2\alpha}{\log(\alpha)}\beta^{2\gamma-1}L_{b,\sigma}^2 T\right)\exp(L_{b,\sigma}T).$$

Choosing α, β and δ such that

$$\left(\beta + \delta T + \frac{2\alpha}{\beta\log(\alpha)}\delta^2 T + \frac{2\alpha}{\log(\alpha)}\beta^{2\gamma-1}L_{b,\sigma}^2 T\right)\exp(L_{b,\sigma}T) \leq \frac{\varepsilon}{2},$$

and combining the above estimates we conclude the proof. $\square$

Chapter 2

Functional differential equations driven by càdlàg rough paths

Stochastic functional differential equations, also known as stochastic delay differential equations, are a natural generalization of stochastic ordinary differential equations, allowing for path-dependent coefficients which may depend on past values of the generated random dynamics. Since numerous real-world phenomena show evidence of a dependence on the past as well as a stochastic behaviour, stochastic functional differential equations serve as mathematical models in many areas ranging from biology to finance. For classical introductions to stochastic functional differential equations we refer, e.g., to [137, 138].

A deterministic approach to stochastic differential equations is provided by rough path theory, initiated by Lyons [129]. Originally designed to treat stochastic ordinary differential equations, it has been extended in various directions, for instance, allowing to deal with stochastic Volterra equations [56], reflected stochastic differential equations [1], stochastic inclusion equations [15], and different classes of stochastic partial differential equations [86, 27]. These rough path approaches contributed many novel insights to the study of the aforementioned equations, such as, but not limited to, new well-posedness and stability results. Comprehensive introductions to rough path theory can be found, e.g., in [130, 71].

In this chapter, we study *rough functional differential equations* (RFDEs)

$$Y_t = y_t + \int_0^t b_s(Y)\,\mathrm{d}s + \int_0^t \sigma_s(Y)\,\mathrm{d}\mathbf{X}_s, \quad t \in [0,T], \tag{2.1}$$

where the driving signal $\mathbf{X}$ is a càdlàg p-rough path for $p \in (2,3)$, the initial condition y is a given controlled path, and the coefficients b, σ are non-anticipative functionals, mapping a controlled path to a controlled path. Assuming a quadratic growth and a Lipschitz-type condition on the path-dependent coefficients b, σ, which both are formulated on the space of controlled paths, we establish the existence of a unique solution to the RFDE (2.3). To that end, we rely on the theory of càdlàg rough paths, as introduced by Friz, Shekhar and Zhang [73, 75], as well as Banach's fixed point theorem. Moreover, we show that the solution map, also known as Itô–Lyons map, mapping the input (initial condition, coefficients, driving signal) of an RFDE to its solution, is locally Lipschitz continuous with respect to suitable distances on the associated spaces of coefficients, controlled paths and rough paths. Let us remark that the continuity of the Itô–Lyons map is one of the most fundamental insights of rough path theory, with many applications to stochastic differential equations, cf. e.g. [71].

The presented results on rough functional differential equations provide a unifying theory, recovering and extending various previous results on different classes of rough differential equations with path-dependent coefficients. Indeed, we deduce the existence of unique solutions as well as the local Lipschitz continuity of the Itô–Lyons map for classical rough differential equations (RDEs), controlled RDEs, RDEs with discrete time dependence and RDEs with constant/variable delay, that are all driven by càdlàg p-rough paths for $p \in (2, 3)$.

In the existing literature, there are several different approaches to deal with rough functional differential equations driven by continuous rough paths. Since the theory of (continuous) rough paths works nicely for infinite dimensional Banach spaces, RDEs with path-dependent coefficients can be treated as Banach space-valued RDEs, see e.g. [14], which requires the coefficients to be Fréchet differentiable and, thus, excludes some interesting examples. Existence, uniqueness and stability results are established by Neuenkirch, Nourdin and Tindel [140] for RDEs with constant delay. The existence of a solution is proven by Ananova [8] for RDEs with path-dependent coefficients, which are assumed to be Dupire differentiable [62], and by Aida [2] for RDEs with coefficients containing path-dependent bounded variation terms. The latter two approaches rely on Schauder's fixed point theorem. Another exemplary class of RFDEs are reflected rough differential equations, see e.g. [1, 55], which, in general, do not possess a unique solution, see [77].

Most applications of rough path theory to stochastic differential equations (SDEs) crucially rely on the construction of suitable (random) rough paths. To apply the developed theory on RFDEs to Itô SDEs with constant delay, we show that a càdlàg martingale together with its delayed version can be lifted to a random rough path in the spirit of stochastic Itô integration. The key challenge to obtain the "delayed" rough path is that a martingale together with its delayed version is, in general, not a martingale itself, thus preventing the direct use of stochastic Itô integration. For related constructions of random rough paths above fractional Brownian motions we refer to [140, 156, 21, 32]. Consequently, one can apply the continuity of the Itô–Lyons map to derive pathwise stability results for stochastic differential equations with constant delay, which plays an important role in many applications, see e.g. [16]. In particular, the map $y \mapsto Y$, mapping the initial condition y to the associated solution Y of an SDE with constant delay, is continuous on the space of controlled paths. This resolves an old observation, pointed out by Mohammed [136], about the non-continuity of the flow of stochastic differential equations with delay, for which the initial condition is in fact an initial path.

This chapter is structured as follows. In Section 2.1 we provide existence, uniqueness and continuity results for rough functional differential equations. In Section 2.2 we prove that various classes of rough differential equations are covered by the presented results on rough functional differential equations. In Section 2.3 we establish the existence of the Itô rough path lift of delayed martingales and discuss applications to stochastic differential equations with delay. Appendix A.1 contains some auxiliary estimates for rough integrals.

2.1 Existence, uniqueness and continuity

Before treating rough functional differential equations (RFDEs), we recall the necessary definitions and some essentials from the theory of càdlàg rough paths, as introduced by Friz and Shekhar [73] and Friz and Zhang [75]. The theory of càdlàg rough paths extends the classical rough path theory, allowing to deal with many stochastic processes with jumps [31], and has numerous applications, e.g., in probability theory [76], numerical analysis [75] and mathematical finance [7].

2.1.1 Essentials on rough path theory

Throughout, let $T > 0$ be a fixed finite time horizon. Let $\Delta_T := \{(s,t) \in [0,T]^2 : s \leq t\}$ be the standard 2-simplex. A function $w\colon \Delta_T \to [0,\infty)$ is called a *control function* if it is superadditive, in the sense that $w(s,u) + w(u,t) \leq w(s,t)$ for all $0 \leq s \leq u \leq t \leq T$. We write $w(s,t-) := \lim_{u \uparrow t} w(s,u)$ if $s < t$, and $w(s,t-) := 0$ if $s = t$.

Whenever $(B, \|\cdot\|)$ is a normed space and $f, g\colon B \to \mathbb{R}$ are two functions on B, we shall write $f \lesssim g$ or $f \leq Cg$ to mean that there exists a constant $C > 0$ such that $f(x) \leq Cg(x)$ for all $x \in B$. The constant C may depend on the normed space, e.g. through its dimension or regularity parameters, and, if we want to emphasize the dependence of the constant C on some particular variables, say $\alpha_1, \dots, \alpha_n$, then we will write $\lesssim_{\alpha_1,\dots,\alpha_n}$ or $C = C_{\alpha_1,\dots,\alpha_n}$. Unless otherwise stated, the dependence of the implicit constants on the variables is locally bounded; that is, if $\alpha_1 \in A_1, \dots, \alpha_n \in A_n$, where $A_1, \dots, A_n$ are compact subsets of the range of $\alpha_1, \dots, \alpha_n$ respectively, then we have that $\sup_{\alpha_1 \in A_1, \dots, \alpha_n \in A_n} C_{\alpha_1,\dots,\alpha_n} < \infty$.

For two vector spaces, the space of linear maps from $E_1 \to E_2$ is denoted by $\mathcal{L}(E_1; E_2)$; and we write $C_b^l = C_b^l(\mathbb{R}^m; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k))$ for the space of l-times differentiable (in the Fréchet sense) functions $f\colon \mathbb{R}^m \to \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k)$ such that f and all its derivatives up to order l are continuous and bounded. We equip this space with the norm

$$\|f\|_{C_b^l} := \|f\|_\infty + \|\mathrm{D}f\|_\infty + \cdots + \|\mathrm{D}^l f\|_\infty,$$

where $\mathrm{D}^n f$ denotes the n-th order derivative of f, and $\|\cdot\|_\infty$ denotes the supremum norm on the corresponding spaces of operators.

For a normed space $(E, |\cdot|)$, let $D([0,T];E)$ be the set of càdlàg (right-continuous with left-limits) paths from $[0,T]$ to E. For $p \geq 1$, the p-variation of a path $X \in D([0,T];E)$ is given by

$$\|X\|_p := \|X\|_{p,[0,T]} \qquad \text{with} \qquad \|X\|_{p,[s,t]} := \Bigg(\sup_{\mathcal{P} \subset [s,t]} \sum_{[u,v] \in \mathcal{P}} |X_v - X_u|^p \Bigg)^{\frac{1}{p}}, \quad (s,t) \in \Delta_T,$$

where the supremum is taken over all possible partitions $\mathcal{P}$ of the interval $[s,t]$. We recall that, given a path X, we have that $\|X\|_p < \infty$ if and only if there exists a control function w such that[1]

$$\sup_{(u,v) \in \Delta_T} \frac{|X_v - X_u|^p}{w(u,v)} < \infty.$$

We write $D^p = D^p([0,T];E)$ for the space of paths $X \in D([0,T];E)$ which satisfy $\|X\|_p < \infty$. Moreover, for a path $X \in D([0,T];\mathbb{R}^d)$, we will often use the shorthand notation:

$$X_{s,t} := X_t - X_s \quad \text{and} \quad X_{t-} := \lim_{u \uparrow t} X_u, \qquad \text{for} \quad (s,t) \in \Delta_T.$$

For $p > 2$ and a two-parameter function $\mathbb{X} \colon \Delta_T \to E$, we similarly define

$$\|\mathbb{X}\|_{\frac{p}{2}} := \|\mathbb{X}\|_{\frac{p}{2},[0,T]} \qquad \text{with} \qquad \|\mathbb{X}\|_{\frac{p}{2},[s,t]} := \Bigg(\sup_{\mathcal{P} \subset [s,t]} \sum_{[u,v] \in \mathcal{P}} |\mathbb{X}_{u,v}|^{\frac{p}{2}} \Bigg)^{\frac{2}{p}}, \quad (s,t) \in \Delta_T.$$

We write $D_2^{\frac{p}{2}} = D_2^{\frac{p}{2}}(\Delta_T;E)$ for the space of all functions $\mathbb{X} \colon \Delta_T \to E$ which satisfy $\|\mathbb{X}\|_{\frac{p}{2}} < \infty$, and are such that the maps $s \mapsto \mathbb{X}_{s,t}$ for fixed t, and $t \mapsto \mathbb{X}_{s,t}$ for fixed s, are both càdlàg.

For $p \in (2,3)$, a pair $\mathbf{X} = (X, \mathbb{X})$ is called a *càdlàg p-rough path* over $\mathbb{R}^d$ if

(i) $X \in D^p([0,T];\mathbb{R}^d)$ and $\mathbb{X} \in D_2^{\frac{p}{2}}(\Delta_T;\mathbb{R}^{d \times d})$, and

(ii) Chen's relation: $\mathbb{X}_{s,t} = \mathbb{X}_{s,u} + \mathbb{X}_{u,t} + X_{s,u} \otimes X_{u,t}$ holds for all $0 \leq s \leq u \leq t \leq T$,

where $\otimes$ denotes the usual tensor product. In component form then, condition (ii) states that $\mathbb{X}^{ij}_{s,t} = \mathbb{X}^{ij}_{s,u} + \mathbb{X}^{ij}_{u,t} + X^i_{s,u} X^j_{u,t}$ for every i and j. We will denote the space of càdlàg p-rough paths by $\mathcal{D}^p = \mathcal{D}^p([0,T];\mathbb{R}^d)$. On the space $\mathcal{D}^p([0,T];\mathbb{R}^d)$, we use the natural seminorm

$$\|\mathbf{X}\|_p := \|\mathbf{X}\|_{p,[0,T]} \qquad \text{with} \qquad \|\mathbf{X}\|_{p,[s,t]} := \|X\|_{p,[s,t]} + \|\mathbb{X}\|_{\frac{p}{2},[s,t]}$$

[1] Here and throughout, we adopt the convention that $\frac{0}{0} := 0$.

for $(s,t) \in \Delta_T$, and the induced distance

$$\|\mathbf{X};\widetilde{\mathbf{X}}\|_p := \|\mathbf{X};\widetilde{\mathbf{X}}\|_{p,[0,T]} \qquad \text{with} \qquad \|\mathbf{X};\widetilde{\mathbf{X}}\|_{p,[s,t]} := \|X-\widetilde{X}\|_{p,[s,t]} + \|\mathbb{X}-\widetilde{\mathbb{X}}\|_{\frac{p}{2},[s,t]},$$

for $(s,t) \in \Delta_T$.

Let $p \in (2,3)$, and $X \in D^p([0,T];\mathbb{R}^d)$. We say that a pair (Y,Y') is a *controlled path* (with respect to X), if

$$Y \in D^p([0,T];E), \quad Y' \in D^p([0,T];\mathcal{L}(\mathbb{R}^d;E)), \quad \text{and} \quad R^Y \in D_2^{\frac{p}{2}}(\Delta_T;E),$$

where R^Y is defined by

$$Y_{s,t} = Y'_s X_{s,t} + R^Y_{s,t} \qquad \text{for all} \quad (s,t) \in \Delta_T.$$

We write $\mathcal{V}^p_X = \mathcal{V}^p_X([0,T];E)$ for the space of E-valued controlled paths, which becomes a Banach space when equipped with the norm $(Y,Y) \mapsto |Y_0| + \|Y,Y'\|_{X,p}$, where

$$\|Y,Y'\|_{X,p} := \|Y,Y'\|_{X,p,[0,T]} \quad \text{with} \quad \|Y,Y'\|_{X,p,[s,t]} := |Y'_s| + \|Y'\|_{p,[s,t]} + \|R^Y\|_{\frac{p}{2},[s,t]}$$

for $(s,t) \in \Delta_T$. We point out that, by definition, for $(s,t) \in \Delta_T$,

$$|Y_{s,t}| \leq |Y'_s||X_{s,t}| + |R^Y_{s,t}| \qquad \text{and} \qquad |Y'_t| \leq |Y'_0| + |Y'_{0,t}|,$$

which implies that

$$\|Y\|_p \leq C_p(\|Y'\|_\infty \|X\|_p + \|R^Y\|_{\frac{p}{2}}) \qquad \text{and} \qquad \|Y'\|_\infty \leq |Y'_0| + \|Y'\|_p,$$

where $\|Y'\|_\infty := \sup_{t\in[0,T]} |Y'_t|$ denotes the supremum seminorm of the path Y'. Given $X, \widetilde{X} \in D^p$, we further introduce the standard "distance"

$$\|Y,Y';\widetilde{Y},\widetilde{Y}'\|_{X,\widetilde{X},p} := \|Y,Y';\widetilde{Y},\widetilde{Y}'\|_{X,\widetilde{X},p,[0,T]}$$

with

$$\|Y,Y';\widetilde{Y},\widetilde{Y}'\|_{X,\widetilde{X},p,[s,t]} := |Y'_s - \widetilde{Y}'_s| + \|Y' - \widetilde{Y}'\|_{p,[s,t]} + \|R^Y - R^{\widetilde{Y}}\|_{\frac{p}{2},[s,t]}$$

for $(s,t) \in \Delta_T$, whenever $(Y,Y') \in \mathcal{V}^p_X$, $(\widetilde{Y},\widetilde{Y}') \in \mathcal{V}^p_{\widetilde{X}}$. Note that $\mathcal{V}^p_X$ and $\mathcal{V}^p_{\widetilde{X}}$ are, in general, different Banach spaces; if $X = \widetilde{X}$, we write $\|\cdot;\cdot\|_{X,p,[s,t]}$.

Given $p \in (2,3)$, $\mathbf{X} = (X,\mathbb{X}) \in \mathcal{D}^p([0,T];\mathbb{R}^d)$ and $(Y,Y') \in \mathcal{V}^p_X([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$, the (forward) rough integral

$$\int_s^t Y_r \,\mathrm{d}\mathbf{X}_r := \lim_{|\mathcal{P}|\to 0} \sum_{[u,v]\in\mathcal{P}} (Y_u X_{u,v} + Y'_u \mathbb{X}_{u,v}), \qquad (s,t) \in \Delta_T, \tag{2.2}$$

exists (in the classical mesh Riemann–Stieltjes sense), where the limit is taken along any sequence of partitions $(\mathcal{P}^n)_{n\in\mathbb{N}}$ of the interval $[s,t]$ such that $|\mathcal{P}^n|\to 0$ as $n\to\infty$. More precisely, in writing the product $Y_u X_{u,v}$, we apply the operator $Y_u \in \mathcal{L}(\mathbb{R}^d;\mathbb{R}^k)$ onto $X_{u,v}\in\mathbb{R}^d$; and in writing the product $Y'_u\mathbb{X}_{u,v}$, we use the natural identification of $\mathcal{L}(\mathbb{R}^d;\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$ with $\mathcal{L}(\mathbb{R}^d\otimes\mathbb{R}^d;\mathbb{R}^k)$. The rough integral comes with the estimate

$$\left|\int_s^t Y_r\,\mathrm{d}\mathbf{X}_r - Y_s X_{s,t} - Y'_s\mathbb{X}_{s,t}\right| \le C\Big(\|R^Y\|_{\frac{p}{2},[s,t)}\|X\|_{p,[s,t]}+\|Y'\|_{p,[s,t)}\|\mathbb{X}\|_{\frac{p}{2},[s,t]}\Big)$$

for some constant C depending only on p; see [75, Proposition 2.6], where

$$\|Y'\|_{p,[s,t)}:=\sup_{u<t}\|Y'\|_{p,[s,u]} \quad \text{and} \quad \|R^Y\|_{\frac{p}{2},[s,t)}:=\sup_{u<t}\|R^Y\|_{\frac{p}{2},[s,u]}.$$

The estimate implies that $(\int_0^\cdot Y_r\,\mathrm{d}\mathbf{X}_r, Y)\in\mathcal{V}_X^p([0,T];\mathbb{R}^k)$ is a controlled path with respect to X, see also [75, Remark 2.8].

For details on the construction of the rough integral with respect to càdlàg p-rough paths and its properties, we refer to [73, 75], and we provide some auxiliary estimates for the rough integral in Appendix A.1.

Let us now consider the rough functional differential equation (RFDE)

$$Y_t = y_t + \int_0^t F_s(Y)\,\mathrm{d}\mathbf{X}_s, \quad t\in[0,T], \tag{2.3}$$

where $\mathbf{X}\in\mathcal{D}^p([0,T];\mathbb{R}^d)$ is a càdlàg p-rough path for $p\in(2,3)$, $(y,y')\in\mathcal{V}_X^p([0,T];\mathbb{R}^k)$ is a given controlled path with respect to X and further, where $(F,F')\colon\mathcal{V}_X^p([0,T];\mathbb{R}^k)\to\mathcal{V}_X^p([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$ is a non-anticipative functional, i.e.

(i) $(F_\cdot(Y),F'_\cdot(Y,Y'))\in\mathcal{V}_X^p([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$,

(ii) $(F_t(Y),F'_t(Y,Y'))=(F_t(Y_{\cdot\wedge t}),F'_t(Y_{\cdot\wedge t},Y'_{\cdot\wedge t}))$ for all $t\in[0,T]$,

for every $(Y,Y')\in\mathcal{V}_X^p([0,T];\mathbb{R}^k)$. The integral in (2.3) is defined as a (forward) rough integral, see (2.2) for its definition. Note that the RFDE (2.1) can be re-written in the form of (2.3), using a standard time-extension of the driving rough path.

2.1.2 Existence and uniqueness

To prove the existence of a unique solution to the rough functional differential equation (2.3), we postulate a quadratic growth and a Lipschitz-type condition on the path-dependent coefficient (F,F'), formulated on the associated path spaces. While a Lipschitz-type condition is expected, the quadratic growth condition appears to be natural in the presented context of (second order) controlled paths, which corresponds to a Taylor expansion with quadratic remainder term.

Assumption 2.1.1. *Let $X \in D^p([0,T];\mathbb{R}^d)$ be given. For every $K > 0$, there exist constants $C_F > 0$, which depends on p and the functional F, and $C_{F,K,X} > 0$, which additionally depends on K and X, such that the map*

$$(F, F'): \mathcal{V}_X^p([0,T];\mathbb{R}^k) \to \mathcal{V}_X^p([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$$

satisfy, for all $(Y,Y'), (\widetilde{Y}, \widetilde{Y}') \in \mathcal{V}_X^p$, and every $0 \leq s < t \leq T$,
(i) *the growth conditions:*

$$\begin{aligned}
&|F_t(Y)| \leq C_F, \\
&|F_{t-,t}(Y)| \leq C_F(1 + \|Y\|_{p,[s,t)} + |Y_{t-,t}|), \\
&\|F(Y)\|_{p,[s,t]} \leq C_F(1 + (|Y'_s| + \|Y'\|_{p,[s,t]})\|X\|_{p,[s,t]} + \|R^Y\|_{\frac{p}{2},[s,t]}), \quad \textit{and} \\
&\|F(Y), F'(Y,Y')\|_{X,p,[s,t]} \leq C_F(1 + \|Y,Y'\|_{X,p,[s,t]})^2(1 + \|X\|_{p,[s,t]})^2;
\end{aligned}$$

(ii) *the Lipschitz conditions:*

$$\begin{aligned}
&\|F(Y) - F(\widetilde{Y})\|_{p,[s,t]} \leq C_{F,K,X}(|Y_s - \widetilde{Y}_s| + \|Y - \widetilde{Y}\|_{p,[s,t]}), \quad \textit{and} \\
&\|F(Y), F'(Y,Y'); F(\widetilde{Y}), F'(\widetilde{Y}, \widetilde{Y}')\|_{X,p,[s,t]} \\
&\quad \leq C_{F,K,X}(|Y_s - \widetilde{Y}_s| + \|Y, Y'; \widetilde{Y}, \widetilde{Y}'\|_{X,p,[s,t]}),
\end{aligned}$$

if $\|Y,Y'\|_{X,p,[s,t]}, \|\widetilde{Y}, \widetilde{Y}'\|_{X,p,[s,t]} \leq K$.

Remark 2.1.2. *The growth and Lipschitz conditions in Assumption 2.1.1 are formulated in terms of both the p-variation of Y and the controlled path norm $(Y,Y') \mapsto |Y_0| + \|Y,Y'\|_{X,p}$ on the space $\mathcal{V}_X^p$ of controlled paths (Y,Y'). To deduce the existence of a unique solution to the RFDE (2.3) under a growth and Lipschitz conditions formulated only in terms of the controlled path norm seems to be far from being obvious. Moreover, notice that the common examples of RDEs with path-dependent coefficients do satisfy Assumption 2.1.1, see Section 2.2 below, demonstrating that Assumption 2.1.1 is, indeed, a natural generalization of the conditions on the coefficients postulated in the existing literature.*

Based on Assumption 2.1.1, we obtain the following global existence and uniqueness result for rough functional differential equations.

Theorem 2.1.3. *Let $\mathbf{X} \in \mathcal{D}^p([0,T];\mathbb{R}^d)$ be a càdlàg p-rough path for $p \in (2,3)$, and $(y,y') \in \mathcal{V}_X^p([0,T];\mathbb{R}^k)$ be a given controlled path with respect to X. Suppose that the non-anticipative functional $(F,F'): \mathcal{V}_X^p([0,T];\mathbb{R}^k) \to \mathcal{V}_X^p([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$ satisfies Assumption 2.1.1 given X. Then, there exists a unique solution to the rough functional differential equation (2.3), i.e. there exists a unique controlled path $(Y,Y') \in \mathcal{V}_X^p([0,T];\mathbb{R}^k)$, with $Y' = y' + F(Y)$, such that*

$$Y_t = y_t + \int_0^t F_s(Y)\,\mathrm{d}\mathbf{X}_s, \qquad t \in [0,T].$$

Moreover, there exists a componentwise non-decreasing function $K_p\colon[0,\infty)^3\to[0,\infty)$ *such that*

$$\|Y,Y'\|_{X,p}\leq K_p(\|y,y'\|_{X,p},C_F,\|\mathbf{X}\|_p).$$

The proof relies on a fixed point approach using Banach's fixed point theorem.

Proof. Step 1: Local solution. We may assume that

$$\|\mathbf{X}\|_p\leq 1 \quad \text{and} \quad \|y'\|_p+\|R^y\|_{\frac{p}{2}}\leq 1.$$

For $t\in(0,T]$, we define the map $\mathcal{M}_t\colon\mathcal{V}^p_X([0,t];\mathbb{R}^k)\to\mathcal{V}^p_X([0,t];\mathbb{R}^k)$ by

$$(Y,Y')\mapsto(Z,Z'):=\mathcal{M}_t(Y,Y'):=\left(y_\cdot+\int_0^\cdot F_s(Y)\,\mathrm{d}\mathbf{X}_s,y'_\cdot+F_\cdot(Y)\right),$$

noting that $\mathcal{M}_t(Y,Y')$ is a controlled path with respect to X as $\mathcal{V}^p_X$ is a Banach space, and introduce the subset of controlled paths

$$\mathcal{B}_t:=\left\{(Y,Y')\in\mathcal{V}^p_X([0,t];\mathbb{R}^k)\,:\,\begin{array}{l}(Y_0,Y_0')=(y_0,y_0'+F_0(y)),\\ \|(Y-y)'\|_{p,[0,t]}\leq 4C_F,\,\|R^Y-R^y\|_{\frac{p}{2},[0,t]}\leq 1\end{array}\right\},$$

which is a complete set as a closed subset of $\mathcal{V}^p_X([0,t];\mathbb{R}^k)$, see [75, Section 3.2].

Invariance. For any $(Y,Y')\in\mathcal{B}_t$, we have that

$$\begin{aligned}\|(Z-y)'\|_{p,[0,t]}&=\|F(Y)\|_{p,[0,t]}\\ &\leq C_F(1+(|Y_0'|+\|Y'\|_{p,[0,t]})\|X\|_{p,[0,t]}+\|R^Y\|_{\frac{p}{2},[0,t]})\\ &\leq C_F+C_F(|Y_0'|+\|Y'\|_{p,[0,t]})\|X\|_{p,[0,t]}+C_F\|R^{Y-y}\|_{\frac{p}{2},[0,t]}+C_F\|R^y\|_{\frac{p}{2},[0,t]}\\ &\leq C_F(1+|Y_0'|+\|Y'\|_{p,[0,t]})\|X\|_{p,[0,t]}+3C_F,\end{aligned}$$

since (F,F') satisfies Assumption 2.1.1 (i), and by the local estimate for rough integration, see Lemma A.1.1,

$$\begin{aligned}\|R^Z-R^y\|_{\frac{p}{2},[0,t]}&=\|R^{\int_0^\cdot F(Y)\mathrm{d}\mathbf{X}}\|_{\frac{p}{2},[0,t]}\\ &\lesssim C_F(1+\|Y,Y'\|_{X,p,[0,t]})^2(1+\|X\|_{p,[0,t]})^2\|\mathbf{X}\|_{p,[0,t]},\end{aligned}$$

where the implicit multiplicative constant depends only on p. Hence, for $t=t_1$ sufficiently small we obtain that $\mathcal{B}_{t_1}$ is invariant under $\mathcal{M}_{t_1}$. Note that t_1 depends on p, $|y_0'|$, C_F and $\|\mathbf{X}\|_p$.

Contraction. Let $(Y,Y'),(\widetilde{Y},\widetilde{Y}')\in\mathcal{B}_t$ for some $t\in(0,t_1]$, that is, setting $K:=5(1+\|y,y'\|_{X,p}+C_F)$, it holds that $\|Y,Y'\|_{X,p,[0,t]},\|\widetilde{Y},\widetilde{Y}'\|_{X,p,[0,t]}\leq K$. We have that

$$\begin{aligned}\|Z'-\widetilde{Z}'\|_{p,[0,t]}&=\|F(Y)-F(\widetilde{Y})\|_{p,[0,t]}\\ &\leq C_{F,K,X}\|Y-\widetilde{Y}\|_{p,[0,t]}\\ &\lesssim_p C_{F,K,X}(\|Y'-\widetilde{Y}'\|_{p,[0,t]}\|X\|_{p,[0,t]}+\|R^Y-R^{\widetilde{Y}}\|_{\frac{p}{2},[0,t]}).\end{aligned}$$

Further, due to Assumption 2.1.1 (ii) and Lemma A.1.2, it holds that

$$
\begin{aligned}
\|R^Z - R^{\widetilde{Z}}\|_{\frac{p}{2},[0,t]} &= \|R^{\int_0^\cdot F(Y)\mathrm{d}\mathbf{X} - \int_0^\cdot F(\widetilde{Y})\mathrm{d}\mathbf{X}}\|_{\frac{p}{2},[0,t]} \\
&\lesssim C_{F,K,X}\|Y,Y';\widetilde{Y},\widetilde{Y}'\|_{X,p,[0,t]}\|\mathbf{X}\|_{p,[0,t]},
\end{aligned}
$$

where the implicit multiplicative constant depends on p and $\|\mathbf{X}\|_p$. Defining an equivalent norm on $\mathcal{V}_X^p$ by

$$
\|Y,Y'\|_{X,p,[0,t]}^{(\delta)} := |Y_0'| + \|Y'\|_{p,[0,t]} + \delta\|R^Y\|_{\frac{p}{2},[0,t]}, \quad \text{for } \delta \geq 1,
$$

we then deduce that

$$
\begin{aligned}
&\|Z - \widetilde{Z}, Z' - \widetilde{Z}'\|_{X,p,[0,t]}^{(\delta)} \\
&\quad\lesssim C_{F,K,X}(\|Y' - \widetilde{Y}'\|_{p,[0,t]}\|X\|_{p,[0,t]} + \|R^Y - R^{\widetilde{Y}}\|_{\frac{p}{2},[0,t]}) \\
&\qquad + \delta C_{F,K,X}(\|Y' - \widetilde{Y}'\|_{p,[0,t]} + \|R^Y - R^{\widetilde{Y}}\|_{\frac{p}{2},[0,t]})\|\mathbf{X}\|_{p,[0,t]} \\
&\quad\lesssim C_{F,K,X}(1+\delta)\|\mathbf{X}\|_{p,[0,t]}\|Y' - \widetilde{Y}'\|_{p,[0,t]} + C_{F,K,X}(1 + \delta\|\mathbf{X}\|_{p,[0,t]})\|R^Y - R^{\widetilde{Y}}\|_{\frac{p}{2},[0,t]} \\
&\quad\lesssim C_{F,K,X}\left((1+\delta)\|\mathbf{X}\|_{p,[0,t]} \vee \frac{1+\delta\|\mathbf{X}\|_{p,[0,t]}}{\delta}\right)\|Y - \widetilde{Y}, Y' - \widetilde{Y}'\|_{X,p,[0,t]}^{(\delta)},
\end{aligned}
$$

where the implicit multiplicative constant depends on p and $\|\mathbf{X}\|_p$. Hence, we can choose δ sufficiently large and $t = t_2 \leq t_1$ sufficiently small such that

$$
C_{F,K,X}\left((1+\delta)\|\mathbf{X}\|_{p,[0,t_2]} \vee \frac{1+\delta\|\mathbf{X}\|_{p,[0,t_2]}}{\delta}\right) \leq 1,
$$

where the left-hand side is up to a multiplicative constant which depends on p and $\|\mathbf{X}\|_p$. It follows that $\mathcal{M}_{t_2}$ is a contraction on the subset of controlled paths $(\mathcal{B}_{t_2}, \|\cdot\|_{X,p,[0,t_2]}^{(\delta)})$. Hence, by Banach's fixed point theorem, there exists a unique fixed point of the map $\mathcal{M}_{t_2}$, which is the unique solution of the RFDE (2.3) over the time interval $[0,t_2]$.

Step 2: Global solution. Let $w\colon \Delta_T \to [0,\infty)$ be the right-continuous control function given by

$$
w(s,t) := \|X\|_{p,[s,t]}^p + \|\mathbb{X}\|_{\frac{p}{2},[s,t]}^{\frac{p}{2}}, \quad (s,t) \in \Delta_T.
$$

We infer from Step 1 that there exists a constant $\gamma > 0$, which depends on p, $\|y,y'\|_{X,p}$, C_F, $C_{F,K,X}$ and $\|\mathbf{X}\|_p$, such that the local solution (Y,Y') established above exists on any interval $[s,t]$ such that $w(s,t) \leq \gamma$, given any initial condition $\xi \in \mathcal{V}_X^p$ with $|\xi_s'| \leq \|y,y'\|_{X,p}$.

By [75, Lemma 1.5], there exists a partition $\mathcal{P} = \{0 = \tau_0 < \tau_1 < \cdots < \tau_N = T\}$ of $[0,T]$, such that $w(\tau_i,\tau_{i+1}-) < \gamma$ for every $i = 0,1,\ldots,N-1$. We can then define the solution (Y,Y') on each of the half-intervals $[\tau_i,\tau_{i+1})$. Given the solutions on $[\tau_i,\tau_{i+1})$, the values $Y_{\tau_{i+1}}$ at the right end-point of the interval are uniquely determined by the jumps of $\mathbf{X}$ at time τ_{i+1}. More precisely, let $y_{0;\cdot} = y_\cdot$, and define y_i, $i = 1,\ldots,N-1$, by

$$
y_{i;t} = y_t + Y_{\tau_i-} - y_{\tau_i-} + F_{\tau_i-}(Y)X_{\tau_i-,\tau_i} + F_{\tau_i-}'(Y,Y')\mathbb{X}_{\tau_i-,\tau_i}, \quad t \in [\tau_i,\tau_{i+1}).
$$

We note that $|y'_{i,\tau_i}|= |y'_i|\leq \|y, y'\|_{X,p}$. Given the initial condition $(y_i, y'_i) \in \mathcal{V}^p_X$, we obtain the solution (Y, Y') on $[\tau_i, \tau_{i+1})$, $i = 0, 1, \ldots, N-1$. By pasting the solutions on each of these subintervals together, with $Y_T = y_{N;T}$, we obtain a unique global solution (Y, Y') of the RFDE (2.3) on the interval $[0, T]$.

Step 3: A priori estimate. It remains to show the existence of a componentwise non-decreasing function $K_p\colon [0,\infty)^3 \to [0,\infty)$ such that

$$\|Y, Y'\|_{X,p}\leq K_p(\|y, y'\|_{X,p}, C_F, \|\mathbf{X}\|_p).$$

Analogously to Step 2, we can obtain a partition $\mathcal{P} = \{0 = \tau_0 < \tau_1 < \cdots < \tau_N = T\}$ and, recalling the definition of $\mathcal{B}_t$, define the solution (Y, Y') on each of the half-intervals $[t_i, t_{i+1})$ such that

$$\|Y'\|_{p,[t_i,t_{i+1})}\leq 4C_F + \|y'\|_{p,[t_i,t_{i+1})} \tag{2.4}$$

as well as

$$\|R^Y\|_{\frac{p}{2},[\tau_i,\tau_{i+1})}\leq 1 + \|R^y\|_{\frac{p}{2},[\tau_i,\tau_{i+1})} \tag{2.5}$$

for all $i = 0, \ldots, N-1$. Here, N depends on p, $\|y, y'\|_{X,p}$, C_F, $C_{F,K,X}$, $\|\mathbf{X}\|_p$ and is, for p fixed, non-decreasing in the other variables. Observe that

$$Y_{t-,t} = y_{t-,t} + \left(\int_0^{\cdot} F_s(Y)\,\mathrm{d}\mathbf{X}_s\right)_{t-,t} = y_{t-,t} + F_{t-}(Y)X_{t-,t} + F'_{t-}(Y, Y')\mathbb{X}_{t-,t},$$

for any $t \in (0, T]$, so we have

$$R^Y_{t-,t} = R^y_{t-,t} + F'_{t-}(Y, Y')\mathbb{X}_{t-,t}.$$

This yields

$$|R^Y_{\tau_{i+1}-,\tau_{i+1}}|\leq \|R^y\|_{p,[\tau_i,\tau_{i+1}]}+(|F'_{\tau_i}(Y, Y')|+\|F'(Y, Y')\|_{p,[\tau_i,\tau_{i+1})})|\mathbb{X}_{\tau_{i+1}-,\tau_{i+1}}|.$$

Now, we use Assumption 2.1.1 (i), i.e.

$$|F'_{\tau_i}(Y, Y')|+\|F'(Y, Y')\|_{p,[\tau_i,\tau_{i+1})}\leq C_F(1 + \|Y, Y'\|_{X,p,[\tau_i,\tau_{i+1})})^2(1 + \|X\|_{p,[\tau_i,\tau_{i+1})})^2.$$

Since

$$\|Y, Y'\|_{X,p,[\tau_i,\tau_{i+1})}\leq |y'_{\tau_i}|+|F_{\tau_i}(Y)|+\|Y'\|_{p,[\tau_i,\tau_{i+1})}+\|R^Y\|_{\frac{p}{2},[\tau_i,\tau_{i+1})},$$

It follows from Assumption 2.1.1 (i), (2.4) and (2.5) that

$$\|Y, Y'\|_{X,p,[\tau_i,\tau_{i+1})}\leq |y'_{\tau_i}|+5(1 + C_F) + \|y'\|_{p,[\tau_i,\tau_{i+1})}+\|R^y\|_{\frac{p}{2},[\tau_i,\tau_{i+1})}.$$

Consequently, there exists a componentwise non-decreasing polynomial $Q^{(R)}_p\colon [0,\infty)^3 \to [0,\infty)$ such that

$$|R^Y_{\tau_{i+1}-,\tau_{i+1}}|\leq Q^{(R)}_p(\|y, y'\|_{X,p,[\tau_i,\tau_{i+1}]}, C_F, \|\mathbf{X}\|_{p,[\tau_i,\tau_{i+1}]})$$

as well as

$$\|R^Y\|_{\frac{p}{2},[\tau_i,\tau_{i+1}]} \leq 1 + \|R^y\|_{\frac{p}{2},[\tau_i,\tau_{i+1}]} + Q_p^{(R)}(\|y,y'\|_{X,p,[\tau_i,\tau_{i+1}]}, C_F, \|\mathbf{X}\|_{p,[\tau_i,\tau_{i+1}]})$$

for all $i = 0, \ldots, N-1$. Moreover, since

$$Y'_{t-,t} = y'_{t-,t} + F_{t-,t}(Y),$$

for any $t \in (0,T]$, we have

$$|Y'_{\tau_{i+1}-,\tau_{i+1}}| \leq \|y'\|_{p,[\tau_i,\tau_{i+1}]} + |F_{\tau_{i+1}-,\tau_{i+1}}(Y)|.$$

By Assumption 2.1.1 (i), it holds that

$$|F_{\tau_{i+1}-,\tau_{i+1}}(Y)| \leq C_F(1 + \|Y\|_{p,[\tau_i,\tau_{i+1})} + |Y_{\tau_{i+1}-,\tau_{i+1}}|),$$

thus, we need to control the jump of Y at τ_{i+1}. For this, note that

$$\begin{aligned}
&|Y_{\tau_{i+1}-,\tau_{i+1}}| \\
&\quad \leq |y_{\tau_{i+1}-,\tau_{i+1}}| + |F_{\tau_{i+1}-}(Y)||X_{\tau_{i+1}-,\tau_{i+1}}| + |F'_{\tau_{i+1}-}(Y)||\mathbb{X}_{\tau_{i+1}-,\tau_{i+1}}| \\
&\quad \leq |y_{\tau_{i+1}-,\tau_{i+1}}| + (|F_{\tau_i}(Y)| + \|F(Y)\|_{p,[\tau_i,\tau_{i+1})})|X_{\tau_{i+1}-,\tau_{i+1}}| \\
&\qquad + (|F'_{\tau_i}(Y,Y')| + \|F'(Y,Y')\|_{p,[\tau_i,\tau_{i+1})})|\mathbb{X}_{\tau_{i+1}-,\tau_{i+1}}| \\
&\quad \leq \|y,y'\|_{p,[\tau_i,\tau_{i+1}]} + (C_F + C_F(1 + (|Y'_{\tau_i}| + \|Y'\|_{\frac{p}{2},[\tau_i,\tau_{i+1})})\|X\|_{p,[\tau_i,\tau_{i+1}]} \\
&\qquad + \|R^Y\|_{\frac{p}{2},[\tau_i,\tau_{i+1})})\|X\|_{p,[\tau_i,\tau_{i+1}]} \\
&\qquad + C_F(1 + \|Y,Y'\|_{p,[\tau_i,\tau_{i+1})})^2(1 + \|X\|_{p,[\tau_i,\tau_{i+1}]})^2\|\mathbb{X}\|_{\frac{p}{2},[\tau_i,\tau_{i+1}]}.
\end{aligned}$$

Using (2.4) and (2.5), we can now conclude that there exist componentwise non-decreasing polynomials $Q_p^{(Y,J)}, Q_p^{(Y')} \colon [0,\infty)^3 \to [0,\infty)$ such that

$$|F_{\tau_{i+1}-,\tau_{i+1}}(Y)| \leq Q_p^{(Y,J)}(\|y,y'\|_{X,p,[\tau_i,\tau_{i+1}]}, C_F, \|\mathbf{X}\|_{p,[\tau_i,\tau_{i+1}]})$$

as well as

$$\|Y'\|_{p,[\tau_i,\tau_{i+1}]} \leq Q_p^{(Y')}(\|y,y'\|_{X,p,[\tau_i,\tau_{i+1}]}, C_F, \|\mathbf{X}\|_{p,[\tau_i,\tau_{i+1}]}).$$

Combining these estimates, we obtain that

$$|Y'_0| + \|Y'\|_p + \|R^Y\|_{\frac{p}{2}} \leq K_p(\|y,y'\|_{X,p}, C_F, \|\mathbf{X}\|_p),$$

which is the assertion. $\square$

2.1.3 Continuity of the Itô–Lyons map

A fundamental contribution of the theory of rough paths is the insight that the solution map, mapping the input (initial condition, coefficients, driving rough path, ...) of a rough differential equation to its solution, is continuous with respect to suitable distances on the spaces of controlled paths as well as of rough paths, see e.g. [71]. In the context of rough differential equations, this solution map is also known as Itô–Lyons map. In the next theorem we present the local Lipschitz continuity of the Itô–Lyons map for rough functional differential equations, based on the following assumption.

Assumption 2.1.4. *Let $X, \widetilde{X} \in D^p([0,T];\mathbb{R}^d)$ be given. For $(G,G') \in \{(F,F'),(\widetilde{F},\widetilde{F}')\}$ and $Z \in \{X,\widetilde{X}\}$ we have: For every $K > 0$, there exist constants $C_G > 0$, which depends on p and the functional G, and $C_{G,K,X,\widetilde{X}} > 0$, which additionally depends on K, X, $\widetilde{X}$ such that the maps*

$$(G,G')\colon \mathcal{V}^p_Z([0,T];\mathbb{R}^k) \to \mathcal{V}^p_Z([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$$

satisfy, for all $(Y,Y') \in \mathcal{V}^p_X$, $(\widetilde{Y},\widetilde{Y}') \in \mathcal{V}^p_{\widetilde{X}}$, and every $0 \le s < t \le T$,
(i) *the growth conditions:*

$$\begin{aligned}
&|G_t(Y)|\le C_G,\\
&|G_{t-,t}(Y)|\le C_G(1+\|Y\|_{p,[s,t)}+|Y_{t-,t}|),\\
&\|G(Y)\|_{p,[s,t]}\le C_G(1+(|Y'_s|+\|Y'\|_{p,[s,t]})\|Z\|_{p,[s,t]}+\|R^Y\|_{\frac{p}{2},[s,t]}), \quad \text{and}\\
&\|G(Y),G'(Y,Y')\|_{Z,p,[s,t]}\le C_G(1+\|Y,Y'\|_{Z,p,[s,t]})^2(1+\|Z\|_{p,[s,t]})^2;
\end{aligned}$$

(ii) *the Lipschitz conditions:*

$$\begin{aligned}
&\|G(Y)-G(\widetilde{Y})\|_{p,[s,t]}\le C_{F,K,X,\widetilde{X}}(|Y_s-\widetilde{Y}_s|+\|Y-\widetilde{Y}\|_{p,[s,t]}), \quad \text{and}\\
&\|G(Y),G'(Y,Y');G(\widetilde{Y}),G'(\widetilde{Y},\widetilde{Y}')\|_{X,\widetilde{X},p,[s,t]}\\
&\quad \le C_{G,K,X,\widetilde{X}}(|Y_s-\widetilde{Y}_s|+\|Y,Y';\widetilde{Y},\widetilde{Y}'\|_{X,\widetilde{X},p,[s,t]}+\|X-\widetilde{X}\|_{p,[s,t]}),
\end{aligned}$$

if $\|Y,Y'\|_{X,p,[s,t]}, \|\widetilde{Y},\widetilde{Y}'\|_{\widetilde{X},p,[s,t]}\le K$.

Moreover, there exists a constant $C_{F-\widetilde{F}} > 0$, which depends on p and the functionals $F-\widetilde{F}$, such that

$$\begin{aligned}
&|(F-\widetilde{F})_t(Y)|\le C_{F-\widetilde{F}},\\
&|(F-\widetilde{F})_{t-,t}(Y)|\le C_{F-\widetilde{F}}(1+\|Y\|_{p,[s,t)}+|Y_{t-,t}|),\\
&\|(F-\widetilde{F})(Y)\|_{p,[s,t]}\le C_{F-\widetilde{F}}(1+(|Y'_s|+\|Y'\|_{p,[s,t]})\|X\|_{p,[s,t]}+\|R^Y\|_{\frac{p}{2},[s,t]}), \quad \text{and}\\
&\|(F-\widetilde{F})(Y),(F'-\widetilde{F}')(Y,Y')\|_{X,p,[s,t]}\le C_{F-\widetilde{F}}(1+\|Y,Y'\|_{X,p,[s,t]})^2(1+\|X\|_{p,[s,t]})^2.
\end{aligned}$$

Theorem 2.1.5. *Let $\mathbf{X}, \widetilde{\mathbf{X}} \in \mathcal{D}^p([0,T];\mathbb{R}^d)$ be càdlàg p-rough paths for $p \in (2,3)$, $(y,y') \in \mathcal{V}^p_X([0,T];\mathbb{R}^k)$, $(\widetilde{y},\widetilde{y}') \in \mathcal{V}^p_{\widetilde{X}}([0,T];\mathbb{R}^k)$ be given controlled paths with respect to X and $\widetilde{X}$, respectively. Suppose that the non-anticipative functionals (F,F'), $(\widetilde{F},\widetilde{F}')$ satisfy Assumption 2.1.4 given $X, \widetilde{X}$.*

Let $(Y,Y) \in \mathcal{V}^p_X([0,T];\mathbb{R}^k)$ be the solution given by Theorem 2.1.3 to the rough functional differential equation (2.3), and $(\widetilde{Y},\widetilde{Y}') \in \mathcal{V}^p_{\widetilde{X}}([0,T];\mathbb{R}^k)$ be the solution to the rough functional differential equation (2.3) driven by $\widetilde{\mathbf{X}}$ with initial condition $(\widetilde{y},\widetilde{y}')$ and functional $(\widetilde{F},\widetilde{F}')$, and suppose that $\|Y,Y'\|_{X,p}, \|\widetilde{Y},\widetilde{Y}'\|_{\widetilde{X},p} \leq K$, for some $K > 0$. Then, we have the estimate

$$\begin{aligned}|Y_0 - \widetilde{Y}_0| &+ \|Y,Y';\widetilde{Y},\widetilde{Y}'\|_{X,\widetilde{X},p}\\ &\lesssim |y_0 - \widetilde{y}_0| + |F_0(y) - \widetilde{F}_0(\widetilde{y})| + \|y,y';\widetilde{y},\widetilde{y}'\|_{X,\widetilde{X},p} + C_{F-\widetilde{F}} + \|\mathbf{X};\widetilde{\mathbf{X}}\|_p,\end{aligned}$$

where the implicit multiplicative constant depends on p, $C_F \vee C_{\widetilde{F}}$, $C_{F,K,X,\widetilde{X}} \vee C_{\widetilde{F},K,X,\widetilde{X}}$, K, $\|\mathbf{X}\|_p$ and $\|\widetilde{\mathbf{X}}\|_p$.

Proof. Step 1: Local estimate. Let $(Y,Y') \in \mathcal{V}^p_X([0,T];\mathbb{R}^k)$, $(\widetilde{Y},\widetilde{Y}') \in \mathcal{V}^p_{\widetilde{X}}([0,T];\mathbb{R}^k)$ be the global solutions to the RFDE (2.3), with data $((y,y'),(F,F'),\mathbf{X})$, $((\widetilde{y},\widetilde{y}'),(\widetilde{F},\widetilde{F}'),\widetilde{\mathbf{X}})$, respectively, see Theorem 2.1.3. Let $t \in (0,T]$. Without loss of generality assume that $C_{\widetilde{F}} \leq C_F$, $C_{\widetilde{F},K,X,\widetilde{X}} \leq C_{F,K,X,\widetilde{X}}$. As

$$\|Y - \widetilde{Y}\|_{p,[0,t]} \leq (|Y_0' - \widetilde{Y}_0'| + \|Y' - \widetilde{Y}'\|_{p,[0,t]})\|X\|_{p,[0,t]} + (|\widetilde{Y}_0'| + \|\widetilde{Y}'\|_{p,[0,t]})\|X - \widetilde{X}\|_{p,[0,t]},$$

Assumption 2.1.4 gives that

$$\begin{aligned}&\|F(Y) - \widetilde{F}(\widetilde{Y})\|_{p,[0,t]}\\ &\quad\leq \|F(Y) - F(\widetilde{Y})\|_{p,[0,t]} + \|(F - \widetilde{F})(\widetilde{Y})\|_{p,[0,t]}\\ &\quad\leq C_{F,K,X,\widetilde{X}}(|Y_0 - \widetilde{Y}_0| + \|Y - \widetilde{Y}\|_{p,[0,t]})\\ &\qquad + C_{F-\widetilde{F}}(1 + (|\widetilde{Y}_0'| + \|\widetilde{Y}'\|_{p,[0,t]})\|X\|_{p,[0,t]} + \|R^{\widetilde{Y}}\|_{\frac{p}{2},[0,t]})\\ &\quad\lesssim_p C_{F,K,X,\widetilde{X}}(|Y_0 - \widetilde{Y}_0| + (|Y_0' - \widetilde{Y}_0'| + \|Y' - \widetilde{Y}'\|_{p,[0,t]})(\|X\|_{p,[0,t]} \vee \|\widetilde{X}\|_{p,[0,t]}))\\ &\qquad + C_{F,K,X,\widetilde{X}}\|R^y - R^{\widetilde{y}}\|_{\frac{p}{2},[0,t]} + C_{F,K,X,\widetilde{X}}\|R^{\int_0^\cdot F(Y)\mathrm{d}\mathbf{X} - \int_0^\cdot \widetilde{F}(\widetilde{Y})\mathrm{d}\widetilde{\mathbf{X}}}\|_{\frac{p}{2},[0,t]}\\ &\qquad + C_{F,K,X,\widetilde{X}}\|X - \widetilde{X}\|_{p,[0,t]} + C_{F-\widetilde{F}}(1+K)(1 + \|X\|_{p,[0,t]} \vee \|\widetilde{X}\|_{p,[0,t]}).\end{aligned}$$

Further, by Lemma A.1.1 and Lemma A.1.2 we have that

$$\begin{aligned}&\|R^{\int_0^\cdot F(Y)\mathrm{d}\mathbf{X} - \int_0^\cdot \widetilde{F}(\widetilde{Y})\mathrm{d}\widetilde{\mathbf{X}}}\|_{\frac{p}{2},[0,t]}\\ &\quad\leq \|R^{\int_0^\cdot F(Y)\mathrm{d}\mathbf{X} - \int_0^\cdot F(\widetilde{Y})\mathrm{d}\widetilde{\mathbf{X}}}\|_{\frac{p}{2},[0,t]} + \|R^{\int_0^\cdot (F-\widetilde{F})(\widetilde{Y})\mathrm{d}\widetilde{\mathbf{X}}}\|_{\frac{p}{2},[0,t]}\\ &\quad\lesssim C_{F,K,X,\widetilde{X}}(|Y_0 - \widetilde{Y}_0| + \|Y,Y';\widetilde{Y},\widetilde{Y}'\|_{X,\widetilde{X},p,[0,t]} + \|X - \widetilde{X}\|_{p,[0,t]})(\|\mathbf{X}\|_{p,[0,t]} \vee \|\widetilde{\mathbf{X}}\|_{p,[0,t]})\\ &\qquad + C_F(1+K)^2(1 + \|X\|_{p,[0,t]} \vee \|\widetilde{X}\|_{p,[0,t]})^2\|\mathbf{X};\widetilde{\mathbf{X}}\|_{p,[0,t]}\\ &\qquad + C_{F-\widetilde{F}}(1+K)^2(1 + \|X\|_{p,[0,t]} \vee \|\widetilde{X}\|_{p,[0,t]})^2(\|\mathbf{X}\|_{p,[0,t]} \vee \|\widetilde{\mathbf{X}}\|_{p,[0,t]}),\end{aligned}$$

where the implicit multiplicative constant depends on p, $\|\mathbf{X}\|_p$ and $\|\widetilde{\mathbf{X}}\|_p$. Combining the results, we get that

$$\begin{aligned}
&\|Y' - \widetilde{Y}'\|_{p,[0,t]} + \|R^Y - R^{\widetilde{Y}}\|_{\frac{p}{2},[0,t]} \\
&\quad \leq \|y' - \widetilde{y}'\|_{p,[0,t]} + \|F(Y) - \widetilde{F}(\widetilde{Y})\|_{p,[0,t]} \\
&\qquad + \|R^y - R^{\widetilde{y}}\|_{\frac{p}{2},[0,t]} + \|R^{\int_0^\cdot F(Y)\,\mathrm{d}\mathbf{X} - \int_0^\cdot \widetilde{F}(\widetilde{Y})\,\mathrm{d}\widetilde{\mathbf{X}}}\|_{\frac{p}{2},[0,t]} \\
&\quad \leq C_1(\|\mathbf{X}\|_{p,[0,t]} \vee \|\widetilde{\mathbf{X}}\|_{p,[0,t]})(\|Y' - \widetilde{Y}'\|_{p,[0,t]} + \|R^Y - R^{\widetilde{Y}}\|_{\frac{p}{2},[0,t]}) \\
&\qquad + C_2(|Y_0 - \widetilde{Y}_0| + |F_0(Y) - \widetilde{F}_0(\widetilde{Y})| + \|y, y'; \widetilde{y}, \widetilde{y}'\|_{X,\widetilde{X},p,[0,t]} + C_{F-\widetilde{F}} + \|\mathbf{X}; \widetilde{\mathbf{X}}\|_{p,[0,t]})
\end{aligned}$$

for some constants $C_1 > 0$, which depends on p, $C_{F,K,X,\widetilde{X}}$, $\|\mathbf{X}\|_p$ and $\|\widetilde{\mathbf{X}}\|_p$, and $C_2 > 1$, which depends additionally on C_F and K. Hence, we can choose t sufficiently small such that $C_1(\|\mathbf{X}\|_{p,[0,t]} \vee \|\widetilde{\mathbf{X}}\|_{p,[0,t]}) \leq \frac{1}{2}$, which implies that

$$\begin{aligned}
&\|Y' - \widetilde{Y}'\|_{p,[0,t]} + \|R^Y - R^{\widetilde{Y}}\|_{\frac{p}{2},[0,t]} \\
&\quad \lesssim |Y_0 - \widetilde{Y}_0| + |F_0(Y) - \widetilde{F}_0(\widetilde{Y})| + \|y, y'; \widetilde{y}, \widetilde{y}'\|_{X,\widetilde{X},p,[0,t]} + C_{F-\widetilde{F}} + \|\mathbf{X}; \widetilde{\mathbf{X}}\|_{p,[0,t]}.
\end{aligned} \tag{2.6}$$

Step 2: Global estimate. Recall the right-continuous control function $w\colon \Delta_T \to [0,\infty)$ given by

$$w(s,t) := \|X\|_{p,[s,t]}^p + \|\mathbb{X}\|_{\frac{p}{2},[s,t]}^{\frac{p}{2}}, \quad (s,t) \in \Delta_T,$$

as introduced in the proof of Theorem 2.1.3, and let $\tilde{w}(s,t) := \|\widetilde{X}\|_{p,[s,t]}^p + \|\widetilde{\mathbb{X}}\|_{\frac{p}{2},[s,t]}^{\frac{p}{2}}$, $(s,t) \in \Delta_T$. We infer from Step 1 that there exists a constant $\gamma > 0$, which depends on p, $C_{F,K,X,\widetilde{X}}$, $\|\mathbf{X}\|_p$ and $\|\widetilde{\mathbf{X}}\|_p$, such that on any interval $[s,t]$ with $w(s,t) \vee \tilde{w}(s,t) \leq \gamma$ the local solutions satisfy an estimate of the form (2.6).

Let $c(s,t) := w(s,t) + \tilde{w}(s,t)$, $(s,t) \in \Delta_T$. Since c is right-continuous, there exists a partition $\mathcal{P} = \{0 = t_0 < \cdots < t_N = T\}$ of $[0,T]$, such that

$$c(t_i, t_{i+1}-) = \gamma, \quad \text{or} \quad c(t_i, t_{i+1}-) < \gamma \quad \text{and} \quad c(t_i, t_{i+1}-) + c(t_{i+1}-, t_{i+1}) \geq \gamma,$$

for every $i = 0, 1, \ldots, N-1$. Since w and $\tilde{w}$ and, thus, c is superadditive, we have that

$$N\gamma \leq \sum_{i=0}^{N-1} c(t_i, t_{i+1}-) + c(t_{i+1}-, t_{i+1}) \leq c(0,T).$$

Therefore, the number of partition points N may be bounded by a constant depending only on γ, $w(0,T)$ and $\tilde{w}(0,T)$. Thus, in this step, we may combine the local estimates on each of the subintervals $[t_i, t_{i+1})$, together with simple estimates on the jumps at the end-points of these subintervals, which we aim to derive, to obtain the global estimate. More precisely, by Step 1, we have the local estimate

$$
\begin{aligned}
&\|Y' - \widetilde{Y}'\|_{p,[t_i,t_{i+1})} + \|R^Y - R^{\widetilde{Y}}\|_{\frac{p}{2},[t_i,t_{i+1})} \\
&\quad \lesssim |Y_{t_i} - \widetilde{Y}_{t_i}| + |F_{t_i}(Y) - \widetilde{F}_{t_i}(\widetilde{Y})| + |y'_{t_i} - \widetilde{y}'_{t_i}| \\
&\qquad + \|y' - \widetilde{y}'\|_{p,[t_i,t_{i+1})} + \|R^y - R^{\widetilde{y}}\|_{\frac{p}{2},[t_i,t_{i+1})} + C_{F-\widetilde{F}} + \|\mathbf{X}; \widetilde{\mathbf{X}}\|_{p,[t_i,t_{i+1})},
\end{aligned} \tag{2.7}
$$

for $i = 0, \ldots, N-1$, where the implicit multiplicative constant depends on p and $C_{F,K,X,\widetilde{X}}$, C_F, K, $\|\mathbf{X}\|_p$, $\|\widetilde{\mathbf{X}}\|_p$, but not on the index i. So, it remains to bound

$$
|Y'_{t_{i+1}-,t_{i+1}} - \widetilde{Y}'_{t_{i+1}-,t_{i+1}}| + |R^Y_{t_{i+1}-,t_{i+1}} - R^{\widetilde{Y}}_{t_{i+1}-,t_{i+1}}|
$$

to extend the previous estimate to $[t_i, t_{i+1}]$.

We note that $(\int_0^{\cdot} F_s(Y)\,\mathrm{d}\mathbf{X}_s)_{t-,t} = F_{t-}(Y)X_{t-,t} + F'_{t-}(Y,Y')\mathbb{X}_{t-,t}$, that is, with $Y_{t-,t} = y_{t-,t} + (\int_0^{\cdot} F_s(Y)\,\mathrm{d}\mathbf{X}_s)_{t-,t}$, it follows that $Y'_{t-,t} = y'_{t-,t} + F_{t-,t}(Y)$ and $R^Y_{t-,t} = R^y_{t-,t} + F'_{t-}(Y,Y')\mathbb{X}_{t-,t}$, for $t \in (0,T]$. Given the assumptions, we then have

$$
\begin{aligned}
&\|Y' - \widetilde{Y}'\|_{p,[t_i,t_{i+1}]} + \|R^Y - R^{\widetilde{Y}}\|_{\frac{p}{2},[t_i,t_{i+1}]} \\
&\quad \leq \|Y' - \widetilde{Y}'\|_{p,[t_i,t_{i+1})} + \|R^Y - R^{\widetilde{Y}}\|_{\frac{p}{2},[t_i,t_{i+1})} \\
&\qquad + |Y'_{t_{i+1}-,t_{i+1}} - \widetilde{Y}'_{t_{i+1}-,t_{i+1}}| + |R^Y_{t_{i+1}-,t_{i+1}} - R^{\widetilde{Y}}_{t_{i+1}-,t_{i+1}}| \\
&\quad \lesssim |Y_{t_i} - \widetilde{Y}_{t_i}| + |F_{t_i}(Y) - \widetilde{F}_{t_i}(\widetilde{Y})| + \|y, y'; \widetilde{y}, \widetilde{y}'\|_{X,\widetilde{X},p,[t_i,t_{i+1}]} + C_{F-\widetilde{F}} + \|\mathbf{X}; \widetilde{\mathbf{X}}\|_{p,[t_i,t_{i+1}]},
\end{aligned} \tag{2.8}
$$

where the implicit multiplicative constant depends on p, C_F, $C_{F,K,X,\widetilde{X}}$, K, $\|\mathbf{X}\|_p$ and $\|\widetilde{\mathbf{X}}\|_p$, and not on the index i. Here, we used Assumption 2.1.4 and the estimate (2.7) to derive that

$$
\begin{aligned}
&|(R^Y_{t_{i+1}-,t_{i+1}} - R^{\widetilde{Y}}_{t_{i+1}-,t_{i+1}}) - (R^y_{t_{i+1}-,t_{i+1}} - R^{\widetilde{y}}_{t_{i+1}-,t_{i+1}})| \\
&\quad \leq |F'_{t_{i+1}-}(Y,Y')\mathbb{X}_{t_{i+1}-,t_{i+1}} - \widetilde{F}'_{t_{i+1}-}(\widetilde{Y},\widetilde{Y}')\widetilde{\mathbb{X}}_{t_{i+1}-,t_{i+1}}| \\
&\quad \leq |F'_{t_{i+1}-}(Y,Y') - \widetilde{F}'_{t_{i+1}-}(\widetilde{Y},\widetilde{Y}')||\mathbb{X}_{t_{i+1}-,t_{i+1}}| + |\widetilde{F}'_{t_{i+1}-}(\widetilde{Y},\widetilde{Y}')||\mathbb{X}_{t_{i+1}-,t_{i+1}} - \widetilde{\mathbb{X}}_{t_{i+1}-,t_{i+1}}| \\
&\quad \leq (|F'_{t_i}(Y,Y') - \widetilde{F}'_{t_i}(\widetilde{Y},\widetilde{Y})| + \|F'(Y,Y') - \widetilde{F}'(\widetilde{Y},\widetilde{Y}')\|_{p,[t_i,t_{i+1})})\|\mathbb{X}\|_{\frac{p}{2},[t_i,t_{i+1}]} \\
&\qquad + (|\widetilde{F}'_{t_i}(\widetilde{Y},\widetilde{Y}')| + \|\widetilde{F}'(\widetilde{Y},\widetilde{Y}')\|_{p,[t_i,t_{i+1})})\|\mathbb{X} - \widetilde{\mathbb{X}}\|_{\frac{p}{2},[t_i,t_{i+1}]} \\
&\quad \lesssim |Y_{t_i} - \widetilde{Y}_{t_i}| + |F_{t_i}(Y) - \widetilde{F}_{t_i}(\widetilde{Y})| + \|y, y'; \widetilde{y}, \widetilde{y}'\|_{X,p,[t_i,t_{i+1})} + C_{F-\widetilde{F}} + \|\mathbf{X}; \widetilde{\mathbf{X}}\|_{p,[t_i,t_{i+1}]},
\end{aligned}
$$

and

$$
\begin{aligned}
&|(Y'_{t_{i+1}-,t_{i+1}} - \widetilde{Y}'_{t_{i+1}-,t_{i+1}}) - (y'_{t_{i+1}-,t_{i+1}} - \widetilde{y}'_{t_{i+1}-,t_{i+1}})| \\
&\quad = |F_{t_{i+1}-,t_{i+1}}(Y) - \widetilde{F}_{t_{i+1}-,t_{i+1}}(\widetilde{Y})| \\
&\quad \leq |(F-\widetilde{F})_{t_{i+1}-,t_{i+1}}(Y)| + |\widetilde{F}_{t_{i+1}-,t_{i+1}}(Y) - \widetilde{F}_{t_{i+1}-,t_{i+1}}(\widetilde{Y})| \\
&\quad \leq C_{F-\widetilde{F}}(1 + (|Y'_{t_i}| + \|Y'\|_{p,[t_i,t_{i+1}]})\|X\|_{p,[t_i,t_{i+1}]} + \|R^Y\|_{\frac{p}{2},[t_i,t_{i+1}]}) \\
&\qquad + C_{F,K,X,\widetilde{X}}(|Y_{t_i} - \widetilde{Y}_{t_i}| + \|Y - \widetilde{Y}\|_{p,[t_i,t_{i+1})} + |Y_{t_{i+1}-,t_{i+1}} - \widetilde{Y}_{t_{i+1}-,t_{i+1}}|) \\
&\quad \lesssim |Y_{t_i} - \widetilde{Y}_{t_i}| + |F_{t_i}(Y) - \widetilde{F}_{t_i}(\widetilde{Y})| + \|y, y'; \widetilde{y}, \widetilde{y}'\|_{X,p,[t_i,t_{i+1})} + C_{F-\widetilde{F}} + \|\mathbf{X}; \widetilde{\mathbf{X}}\|_{p,[t_i,t_{i+1}]},
\end{aligned}
$$

where the implicit multiplicative constant depends on p, C_F, $C_{F,K,X,\widetilde{X}}$, K and $\|\mathbf{X}\|_p \vee \|\widetilde{\mathbf{X}}\|_p$, and not on the index i.

Now, we need to control the term $|Y_{t_i} - \widetilde{Y}_{t_i}| + |F_{t_i}(Y) - \widetilde{F}_{t_i}(\widetilde{Y})| + |y'_{t_i} - \widetilde{y}'_{t_i}|$. For this, we note that

$$
\begin{aligned}
&|Y_{t_i} - \widetilde{Y}_{t_i}| + |F_{t_i}(Y) - \widetilde{F}_{t_i}(\widetilde{Y})| + |y'_{t_i} - \widetilde{y}'_{t_i}| \\
&\quad \leq |Y_{t_{i-1}} - \widetilde{Y}_{t_{i-1}}| + |F_{t_{i-1}}(Y) - \widetilde{F}_{t_{i-1}}(\widetilde{Y})| + |y'_{t_{i-1}} - \widetilde{y}'_{t_{i-1}}| \\
&\qquad + \|Y - \widetilde{Y}\|_{p,[t_{i-1},t_i]} + \|F(Y) - \widetilde{F}(\widetilde{Y})\|_{p,[t_{i-1},t_i]} + \|y' - \widetilde{y}'\|_{p,[t_{i-1},t_i]} \\
&\quad \lesssim |Y_{t_{i-1}} - \widetilde{Y}_{t_{i-1}}| + |F_{t_{i-1}}(Y) - \widetilde{F}_{t_{i-1}}(\widetilde{Y})| + \|y, y'; \widetilde{y}, \widetilde{y}'\|_{X,\widetilde{X},p,[t_{i-1},t_i]} \\
&\qquad + C_{F-\widetilde{F}} + \|\mathbf{X}; \widetilde{\mathbf{X}}\|_{p,[t_{i-1},t_i]} \\
&\qquad + \|Y' - \widetilde{Y}'\|_{p,[t_{i-1},t_i]} + \|R^Y - R^{\widetilde{Y}}\|_{\frac{p}{2},[t_{i-1},t_i]},
\end{aligned}
$$

where the implicit multiplicative constant depends on p, C_F, $C_{F,K,X,\widetilde{X}}$, K, $\|X\|_p$ and $\|\widetilde{X}\|_p$, and not on the index i; thus applying the estimate (2.8) for $i-1$ gives that

$$
\begin{aligned}
&|Y_{t_i} - \widetilde{Y}_{t_i}| + |F_{t_i}(Y) - \widetilde{F}_{t_i}(\widetilde{Y})| + |y'_{t_i} - \widetilde{y}'_{t_i}| \\
&\quad \lesssim |Y_{t_{i-1}} - \widetilde{Y}_{t_{i-1}}| + |F_{t_{i-1}}(Y) - \widetilde{F}_{t_{i-1}}(\widetilde{Y})| + \|y, y'; \widetilde{y}, \widetilde{y}'\|_{X,\widetilde{X},p,[t_{i-1},t_i]} \\
&\qquad + C_{F-\widetilde{F}} + \|\mathbf{X}; \widetilde{\mathbf{X}}\|_{p,[t_{i-1},t_i]}.
\end{aligned}
$$

Iteratively, we obtain for any $i = 1, \ldots, N$ that

$$
\begin{aligned}
&|Y_{t_i} - \widetilde{Y}_{t_i}| + |F_{t_i}(Y) - \widetilde{F}_{t_i}(\widetilde{Y})| + |y'_{t_i} - \widetilde{y}'_{t_i}| \\
&\quad \lesssim |y_0 - \widetilde{y}_0| + |F_0(y) - \widetilde{F}_0(\widetilde{y})| + |y'_0 - \widetilde{y}'_0| \\
&\qquad + \sum_{j=0}^{i-1} \Big(\|y' - \widetilde{y}'\|_{p,[t_j,t_{j+1}]} + \|R^y - R^{\widetilde{y}}\|_{\frac{p}{2},[t_j,t_{j+1}]} + C_{F-\widetilde{F}} + \|\mathbf{X}; \widetilde{\mathbf{X}}\|_{p,[t_j,t_{j+1}]} \Big),
\end{aligned}
$$

that is,

$$
\begin{aligned}
&(|Y_{t_i} - \widetilde{Y}_{t_i}| + |F_{t_i}(Y) - \widetilde{F}_{t_i}(\widetilde{Y})| + |y'_{t_i} - \widetilde{y}'_{t_i}|)^p \\
&\quad \lesssim (|y_0 - \widetilde{y}_0| + |F_0(y) - \widetilde{F}_0(\widetilde{y})| + |y'_0 - \widetilde{y}'_0| + NC_{F-\widetilde{F}})^p \\
&\qquad + \sum_{j=0}^{i-1} \Big(\|y' - \widetilde{y}'\|^p_{p,[t_j,t_{j+1}]} + \|R^y - R^{\widetilde{y}}\|^p_{\frac{p}{2},[t_j,t_{j+1}]} + \|\mathbf{X}; \widetilde{\mathbf{X}}\|^p_{p,[t_j,t_{j+1}]} \Big).
\end{aligned}
$$

This implies that

$$
\begin{aligned}
&|Y_{t_i} - \widetilde{Y}_{t_i}| + |F_{t_i}(Y) - \widetilde{F}_{t_i}(\widetilde{Y})| + |y'_{t_i} - \widetilde{y}'_{t_i}| \\
&\quad \lesssim |y_0 - \widetilde{y}_0| + |F_0(y) - \widetilde{F}_0(\widetilde{y})| + \|y, y'; \widetilde{y}, \widetilde{y}'\|_{X,\widetilde{X},p} + C_{F-\widetilde{F}} + \|\mathbf{X}; \widetilde{\mathbf{X}}\|_p,
\end{aligned}
$$

which is the desired control.

If we plug this into (2.8), it follows that

$$
\begin{aligned}
&\|Y' - \widetilde{Y}'\|_{p,[t_i,t_{i+1}]} + \|R^Y - R^{\widetilde{Y}}\|_{\frac{p}{2},[t_i,t_{i+1}]} \\
&\quad \lesssim |y_0 - \widetilde{y}_0| + |F_0(y) - \widetilde{F}_0(\widetilde{y})| + \|y, y'; \widetilde{y}, \widetilde{y}'\|_{X,\widetilde{X},p} + C_{F-\widetilde{F}} + \|\mathbf{X}; \widetilde{\mathbf{X}}\|_p.
\end{aligned}
$$

Since $\|\cdot\|_{p,[0,T]} \leq N \sum_{i=0}^{N-1} \|\cdot\|_{p,[t_i,t_{i+1}]}$ for any $p \geq 1$, see e.g. [3, Lemma A.1], the estimate finally follows. □

2.2 Examples of RFDEs

The general framework of rough functional differential equations, presented in Section 2.1, allows to treat various classes of rough differential equations. In this section, some exemplary rough functional differential equations are discussed, aiming to develop the main conceptional ideas and demonstrating the scope of RFDEs rather than pushing for the most general results.

2.2.1 Classical RDEs

Let us start with the classical rough differential equation (RDE)

$$
Y_t = y_0 + \int_0^t f(Y_s)\,\mathrm{d}\mathbf{X}_s, \quad t \in [0,T], \tag{2.9}
$$

where $y_0 \in \mathbb{R}^k$, $f \in C_b^3(\mathbb{R}^k; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k))$ and $\mathbf{X} \in \mathcal{D}^p([0,T]; \mathbb{R}^d)$ for $p \in (2,3)$. While the existence and uniqueness of solutions to the RDE (2.9) driven by a continuous rough path and the continuity of the solution map were first proven by Lyons [129], the analogous results for RDEs driven by càdlàg rough paths were more recently obtained by Friz and Zhang [75]. As an application of Theorem 2.1.3 and Theorem 2.1.5, one can recover these results, demonstrating that Assumption 2.1.1 and Assumption 2.1.4 are, indeed, natural generalizations of the classical assumptions of the coefficients of a rough differential equation. Furthermore, note that Corollary 2.2.1 presents the continuity of the solution map with respect to the controlled path norm, which slightly generalizes [75, Theorem 3.8].

Corollary 2.2.1.

(i) If $f \in C_b^3(\mathbb{R}^k; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k))$, there exists a unique solution to the RDE (2.9). *Moreover, there exists a non-decreasing function $K_p \colon [0,\infty)^2 \to [0,\infty)$ such that*

$$\|Y, Y'\|_{X,p} \leq K_p(\|f\|_{C_b^2}, \|\mathbf{X}\|_p).$$

(ii) Let $(Y, Y') \in \mathcal{V}_X^p([0,T]; \mathbb{R}^k)$ be the unique solution to the RDE (2.9). *Moreover, let $\widetilde{y}_0 \in \mathbb{R}^k$, $\widetilde{f} \in C_b^3(\mathbb{R}^k; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k))$, $\widetilde{\mathbf{X}} \in \mathcal{D}^p([0,T]; \mathbb{R}^d)$ with corresponding solution $(\widetilde{Y}, \widetilde{Y}') \in \mathcal{V}_{\widetilde{X}}^p([0,T]; \mathbb{R}^d)$, and suppose that $\|Y, Y'\|_{X,p}, \|\widetilde{Y}, \widetilde{Y}'\|_{\widetilde{X},p} \leq K$, for some $K > 0$. Then, we have the estimate*

$$|Y_0 - \widetilde{Y}_0| + \|Y, Y'; \widetilde{Y}, \widetilde{Y}'\|_{X,\widetilde{X},p} \lesssim |y_0 - \widetilde{y}_0| + \|f - \widetilde{f}\|_{C_b^2} + \|\mathbf{X}; \widetilde{\mathbf{X}}\|_p,$$

where the implicit multiplicative constant depends on p, $\|f\|_{C_b^3} \vee \|\widetilde{f}\|_{C_b^3}$, K, $\|\mathbf{X}\|_p$ and $\|\widetilde{\mathbf{X}}\|_p$.

In order to apply the existence and uniqueness result presented in Theorem 2.1.3, and also the continuity result presented in Theorem 2.1.5, we need to check that the vector field f in the RDE (2.9) satisfies Assumption 2.1.1 and Assumption 2.1.4. This is the content of the next lemma, which we formulate slightly more general, with regard to the dimensions of the underlying spaces, for later use.

Lemma 2.2.2. *Let $f \in C_b^3(\mathbb{R}^m; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k))$ and $X, \widetilde{X} \in D^p([0,T]; \mathbb{R}^d)$. The non-anticipative functional*

$$(F, F') \colon \mathcal{V}_X^p([0,T]; \mathbb{R}^m) \to \mathcal{V}_X^p([0,T]; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k)), \quad (F(Y), F'(Y, Y')) := (f(Y), \mathrm{D}f(Y)Y'),$$

satisfies Assumption 2.1.4 (i) and (ii), and, in particular, Assumption 2.1.1, given $X, \widetilde{X}$.

Proof. Since the proof is fairly standard, we provide only a sketch of a proof, following, e.g., the proofs of [75, Lemma 3.5, Lemma 3.6, Lemma 3.7].

Fix $(s,t) \in \Delta_T$ and let $(Y, Y') \in \mathcal{V}_X^p$, $(\widetilde{Y}, \widetilde{Y}') \in \mathcal{V}_{\widetilde{X}}^p$.

Growth conditions. It is clear that $|F_t(Y)| \leq \|f\|_{C_b^2}$, and it follows from the Lipschitz continuity of f that

$$|F_{t-,t}(Y)| \leq \|F(Y)\|_{p,[s,t]} \leq \|f\|_{C_b^2} \|Y\|_{p,[s,t]}.$$

We now note that $\|Y\|_{p,[s,t]} \leq \|Y\|_{p,[s,t)} + |Y_{t-,t}|$ as well as $\|Y\|_{p,[s,t]} \leq C_p(1+(|Y'_s|+\|Y'\|_{p,[s,t]}) \cdot \|X\|_{p,[s,t]} + \|R^Y\|_{\frac{p}{2},[s,t]})$. Further, it holds that

$$\begin{aligned}
&|F'_s(Y,Y')| + \|F'(Y,Y')\|_{p,[s,t]} = |\mathrm{D}f(Y_s)Y'_s| + \|\mathrm{D}f(Y)Y'\|_{p,[s,t]} \\
&\quad \leq \|f\|_{C_b^2}(|Y'_s| + \|Y'\|_{p,[s,t]})(1 + \|Y\|_{p,[s,t]}) \\
&\quad \lesssim_p \|f\|_{C_b^2}(|Y'_s| + \|Y'\|_{p,[s,t]})(1 + (|Y'_s| + \|Y'\|_{p,[s,t]})\|X\|_{p,[s,t]} + \|R^Y\|_{\frac{p}{2},[s,t]}) \\
&\quad \lesssim_p \|f\|_{C_b^2}(1 + \|Y,Y'\|_{X,p,[s,t]})(1 + \|X\|_{p,[s,t]})
\end{aligned}$$

and by Taylor's expansion,

$$\begin{aligned}
R^{F(Y)}_{s,t} = R^{f(Y)}_{s,t} &= f(Y_t) - f(Y_s) - \mathrm{D}f(Y_s)Y_{s,t} + \mathrm{D}f(Y_s)R^Y_{s,t} \\
&= \frac{1}{2}\mathrm{D}^2 f(Y_s + \lambda Y_{s,t})Y_{s,t}^2 + \mathrm{D}f(Y_s)R^Y_{s,t},
\end{aligned}$$

with $\lambda \in [0,1]$, which implies that

$$\begin{aligned}
&\|R^{F(Y)}\|_{\frac{p}{2},[s,t]} = \|R^{f(Y)}\|_{\frac{p}{2},[s,t]} \leq \|f\|_{C_b^2}(\|Y\|^2_{p,[s,t]} + \|R^Y\|_{\frac{p}{2},[s,t]}) \\
&\quad \lesssim_p \|f\|_{C_b^2}(((|Y'_s| + \|Y'\|_{p,[s,t]})\|X\|_{p,[s,t]} + \|R^Y\|_{\frac{p}{2},[s,t]})^2 + \|R^Y\|_{\frac{p}{2},[s,t]}) \\
&\quad \lesssim_p \|f\|_{C_b^2}(1 + \|Y,Y'\|_{X,p,[s,t]})^2(1 + \|X\|_{p,[s,t]})^2.
\end{aligned}$$

Assumption 2.1.4 (i) therefore holds with some constant $C_F = \|f\|_{C_b^2}$ up to a multiplicative constant which depends on p.

Lipschitz conditions. Fix $K > 0$ and assume that $\|Y,Y'\|_{X,p,[s,t]}, \|\widetilde{Y},\widetilde{Y}'\|_{\widetilde{X},p,[s,t]} \leq K$. The proofs work verbatim as the proofs of [75, Lemma 3.1 and Lemma 3.7]. The constant $C_{F,K,X,\widetilde{X}}$ depends on p and $\|f\|_{C_b^3}$, K, $\|X\|_p$ and $\|\widetilde{X}\|_p$. □

Proof of Corollary 2.2.1. (i) The existence and uniqueness of the solution follows immediately from Lemma 2.2.2 and Theorem 2.1.3. For the a priori estimate, note that $\|y,y'\|_{X,p} = 0$ and $C_F \lesssim_p \|f\|_{C_b^2}$.

(ii) To apply the continuity result presented in Theorem 2.1.5, we need to ensure that the functionals satisfy the required assumptions. For $(F,F'), (\widetilde{F},\widetilde{F}')$, this is given in Lemma 2.2.2, and further,

$$\begin{aligned}
&(F-\widetilde{F},(F-\widetilde{F})')\colon \mathcal{V}^p_X([0,T];\mathbb{R}^k) \to \mathcal{V}^p_X([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k)), \\
&((F-\widetilde{F})(Y),(F-\widetilde{F})'(Y,Y')) := (f(Y) - \widetilde{f}(Y), \mathrm{D}f(Y)Y' - \mathrm{D}\widetilde{f}(Y)Y'),
\end{aligned}$$

satisfies the corresponding estimates in Assumption 2.1.4, since C_b^3 is a vector space. Thus we have, $C_{F-\widetilde{F}} \lesssim_p \|f - \widetilde{f}\|_{C_b^2}$. □

2.2.2 Controlled RDEs

Motivated by pathwise stochastic control, see e.g. [57, 4], and robust stochastic filtering, see e.g. [3], as well as analogously to controlled stochastic differential equations, see e.g. [144], we consider the controlled rough differential equation

$$Y_t = y_t + \int_0^t f(\alpha_s, Y_s)\,\mathrm{d}\mathbf{X}_s, \quad t \in [0,T], \tag{2.10}$$

where $\mathbf{X} \in \mathcal{D}^p([0,T];\mathbb{R}^d)$ for $p \in (2,3)$, $(y,y') \in \mathcal{V}^p_X([0,T];\mathbb{R}^k)$, $f \in C^3_b(\mathbb{R}^{k+e};\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$, and $(\alpha,\alpha') \in \mathcal{V}^p_X([0,T];\mathbb{R}^e)$ is a fixed controlled path, with $e \in \mathbb{N}$. In case of continuous rough paths and controls α of finite $\frac{p}{2}$-variation, controlled RDEs were treated in [3, Theorem 2.3]. The following corollary provides an existence, uniqueness and continuity result for controlled RDEs driven by càdlàg p-rough paths and with controls α, which are only required to be controlled paths.

Corollary 2.2.3.

(i) If $f \in C^3_b(\mathbb{R}^{k+e};\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$ and $(\alpha,\alpha') \in \mathcal{V}^p_X([0,T];\mathbb{R}^e)$, then there exists a unique solution to the controlled rough differential equation (2.10)*. Moreover, there exists a componentwise non-decreasing function $K_p\colon [0,\infty)^5 \to [0,\infty)$ such that*

$$\|Y,Y'\|_{X,p} \leq K_p(\|f\|_{C^2_b}, \|y,y'\|_{X,p}, \|\alpha\|_p, \|\alpha,\alpha'\|_{X,p}, \|\mathbf{X}\|_p).$$

(ii) Let $(Y,Y') \in \mathcal{V}^p_X([0,T];\mathbb{R}^k)$ be the unique solution to the controlled rough differential equation (2.10)*. Moreover, let $(\widetilde{y},\widetilde{y}') \in \mathcal{V}^p_X([0,T];\mathbb{R}^k)$, $(\tilde{\alpha},\tilde{\alpha}') \in \mathcal{V}^p_X([0,T];\mathbb{R}^e)$, $\widetilde{f} \in C^3_b(\mathbb{R}^{k+e};\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$, with corresponding solution $(\widetilde{Y},\widetilde{Y}') \in \mathcal{V}^p_X([0,T];\mathbb{R}^k)$, and suppose that $\|Y,Y'\|_{X,p}, \|\widetilde{Y},\widetilde{Y}'\|_{X,p} \leq K$, for some $K > 0$. Then, we have the estimate*

$$\begin{aligned}&|Y_0 - \widetilde{Y}_0| + \|Y,Y';\widetilde{Y},\widetilde{Y}'\|_{X,p} \\ &\qquad \lesssim |y_0 - \widetilde{y}_0| + \|y,y';\widetilde{y},\widetilde{y}'\|_{X,p} + \|f - \widetilde{f}\|_{C^2_b} + |\alpha_0 - \tilde{\alpha}_0| + \|\alpha,\alpha';\tilde{\alpha},\tilde{\alpha}\|_{X,p},\end{aligned}$$

where the implicit multiplicative constant depends on p, $\|f\|_{C^3_b} \vee \|\widetilde{f}\|_{C^3_b}$, $\|\alpha,\alpha'\|_{X,p}$, $\|\tilde{\alpha},\tilde{\alpha}'\|_{X,p}$, K, and $\|\mathbf{X}\|_p$.

In order to apply the existence and uniqueness result presented in Theorem 2.1.3, and also the continuity result in Theorem 2.1.5, we need to check that the vector field in the RDE (2.10) satisfies Assumption 2.1.4. This is the content of the next lemma. Note that it will be sufficient to check Assumption 2.1.4 (ii) for $X = \widetilde{X}$, since we do not establish stability results with respect to the driving rough path in this subsection. (This also applies to the following subsections.)

Lemma 2.2.4. *Let $f \in C_b^3(\mathbb{R}^{k+e}; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k))$ and $X \in D^p([0,T]; \mathbb{R}^d)$ for $p \in (2,3)$. Further, let $(\alpha, \alpha') \in \mathcal{V}_X^p([0,T]; \mathbb{R}^e)$. The non-anticipative functional*

$$\begin{aligned}&(F, F') \colon \mathcal{V}_X^p([0,T]; \mathbb{R}^k) \to \mathcal{V}_X^p([0,T]; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k)),\\ &(F(Y), F'(Y, Y')) := (f((\alpha, Y)), \mathrm{D}f((\alpha, Y))(\alpha', Y')),\end{aligned}$$

satisfies Assumption 2.1.4 given $X = \widetilde{X}$.

Proof. Fix $(s,t) \in \Delta_T$ and let $(Y, Y') \in \mathcal{V}_X^p$. It is clear that $|F_t(Y)| \leq \|f\|_{C_b^2}$, and we note that

$$\begin{aligned}|F_{t-,t}(Y)| &\leq \|F(Y)\|_{p,[s,t]} = \|f((\alpha, Y))\|_{p,[s,t]}\\ &\leq \|f\|_{C_b^2} \|(\alpha, Y)\|_{p,[s,t]}\\ &\leq \|f\|_{C_b^2}(1 + \|\alpha\|_{p,[s,t]})(1 + \|Y\|_{p,[s,t]}),\end{aligned}$$

and it holds that $\|Y\|_{p,[s,t]} \leq \|Y\|_{p,[s,t)} + |Y_{t-,t}|$ as well as $\|Y\|_{p,[s,t]} \leq C_p(1 + (|Y'_s| + \|Y'\|_{p,[s,t]}) \cdot \|X\|_{p,[s,t]} + \|R^Y\|_{\frac{p}{2},[s,t]})$.

Applying Lemma 2.2.2 to the enlarged controlled path $((\alpha, Y), (\alpha', Y'))$, it follows that

$$\begin{aligned}&\|F(Y), F'(Y, Y')\|_{X,p,[s,t]}\\ &\quad\lesssim \|f\|_{C_b^2}(1 + |(\alpha'_s, Y'_s)| + \|(\alpha', Y')\|_{p,[s,t]} + \|R^{(\alpha,Y)}\|_{\frac{p}{2},[s,t]})^2 (1 + \|X\|_{p,[s,t]})^2\\ &\quad\lesssim \|f\|_{C_b^2}(1 + \|\alpha, \alpha'\|_{X,p})^2 (1 + \|Y, Y'\|_{X,p,[s,t]})^2 (1 + \|X\|_{p,[s,t]})^2.\end{aligned}$$

The growth conditions thus hold with constant $C_F = \|f\|_{C_b^2}$ up to a multiplicative constant, which depends on p, $\|\alpha\|_p$ and $\|\alpha, \alpha'\|_{X,p}$.

Proceeding as in the proof of Lemma 2.2.2, we can show the Lipschitz conditions, observing that

$$|(\alpha, Y)_s - (\alpha, \widetilde{Y})_s| + \|(\alpha, Y) - (\alpha, \widetilde{Y})\|_{p,[s,t]} = |Y_s - \widetilde{Y}_s| + \|Y - \widetilde{Y}\|_{p,[s,t]}$$

and

$$\|(\alpha, Y), (\alpha', Y'); (\alpha, \widetilde{Y}), (\alpha', \widetilde{Y}')\|_{X,p} = \|Y, Y'; \widetilde{Y}, \widetilde{Y}'\|_{X,p},$$

and similarly for each summand of the norm, so the Lipschitz conditions hold with constant $C_{F,K,X,X}$, which depends on p, $\|f\|_{C_b^3}$, K, for $K > 0$, $\|\alpha, \alpha'\|_{X,p}$, and $\|X\|_p$. $\square$

Proof of Corollary 2.2.3. (i) The existence and uniqueness of the solution follows immediately from Lemma 2.2.4 and Theorem 2.1.3. For the a priori estimate, note that $C_F \lesssim_p \|f\|_{C_b^2}(1 + \|\alpha\|_p + \|\alpha, \alpha'\|_{X,p})^2$.

(ii) To apply the continuity result presented in Theorem 2.1.5, we need to ensure that the functionals satisfy the required assumptions. For $(F, F'), (\widetilde{F}, \widetilde{F}')$, this is given in Lemma 2.2.4. Analogously, since $f - \widetilde{f} \in C_b^3$, we note that for

$$(Y, Y') \mapsto ((f - \widetilde{f})((\alpha, Y)), \mathrm{D}(f - \widetilde{f})((\alpha, Y))(\alpha', Y'))$$

the growth conditions hold with constant equal to $\|f - \widetilde{f}\|_{C_b^2}$ up to a multiplicative constant which depends on p, $\|\alpha\|_p$ and $\|\alpha, \alpha'\|_{X,p}$. Further, it follows from the proofs of [75, Lemma 3.1 and Lemma 3.5] that the growth conditions hold for

$$(Y, Y') \mapsto (f((\alpha, Y)) - f((\tilde{\alpha}, Y)), \mathrm{D}f((\alpha, Y))(\alpha', Y') - \mathrm{D}f((\tilde{\alpha}, Y))(\tilde{\alpha}', Y'))$$

with constant equal to $|\alpha_0 - \tilde{\alpha}_0| + \|\alpha - \tilde{\alpha}\|_p + \|\alpha, \alpha'; \tilde{\alpha}, \tilde{\alpha}'\|_{X,p}$ up to a multiplicative constant which depends on p, $\|f\|_{C_b^3}$, $\|\alpha\|_p$, $\|\tilde{\alpha}\|_p$, $\|\alpha, \alpha'\|_{X,p}$, $\|\tilde{\alpha}, \tilde{\alpha}'\|_{X,p}$.

This implies that

$$\begin{aligned}
&(F - \widetilde{F}, F' - \widetilde{F}') \colon \mathcal{V}_X^p([0,T]; \mathbb{R}^k) \to \mathcal{V}_X^p([0,T]; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k)), \\
&((F - \widetilde{F})(Y), (F - \widetilde{F})'(Y, Y')) \\
&\quad := (f((\alpha, Y)) - \widetilde{f}((\tilde{\alpha}, Y)), \mathrm{D}f((\alpha, Y))(\alpha', Y') - \mathrm{D}\widetilde{f}((\tilde{\alpha}, Y))(\tilde{\alpha}', Y')),
\end{aligned}$$

satisfies the corresponding estimates in Assumption 2.1.4 with

$$C_{F - \widetilde{F}} = \|f - \widetilde{f}\|_{C_b^2} + |\alpha_0 - \tilde{\alpha}_0| + \|\alpha - \tilde{\alpha}\|_p + \|\alpha, \alpha'; \tilde{\alpha}, \tilde{\alpha}'\|_{X,p}$$

up to a multiplicative constant, which depends on p, $\|f\|_{C_b^3} \vee \|\widetilde{f}\|_{C_b^3}$, $\|\alpha\|_p$, $\|\tilde{\alpha}\|_p$, $\|\alpha, \alpha'\|_{X,p}$, $\|\tilde{\alpha}, \tilde{\alpha}'\|_{X,p}$. $\square$

2.2.3 RDEs with discrete time dependence

Let us consider the rough differential equation with discrete time dependence

$$Y_t = y_t + \int_0^t f(Y_s, Y_{s \wedge r_1}, \dots Y_{s \wedge r_\ell}) \, \mathrm{d}\mathbf{X}_s, \quad t \in [0, T], \tag{2.11}$$

where $\mathbf{X} \in \mathcal{D}^p([0,T]; \mathbb{R}^d)$ for $p \in (2,3)$, $(y, y') \in \mathcal{V}_X^p([0,T]; \mathbb{R}^k)$, $f \in C_b^3(\mathbb{R}^{k(\ell+1)}; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k))$, and $r_1 < \dots < r_\ell$ be given time points in $[0, T]$, with $\ell \in \mathbb{N}$. For continuous rough paths as driving signals, the existence (without uniqueness) of a solution to the RDE (2.11) was proven in [8, Example 4.2 and Theorem 4.4]. The next proposition provides an existence, uniqueness and continuity result for RDEs with discrete time dependence driven by càdlàg p-rough paths.

Proposition 2.2.5.

(i) In the above setting, there exists a unique solution to the RDE with discrete time dependence (2.11)*. Moreover, there exists a componentwise non-decreasing function* $K_p: \mathbb{N} \times [0, \infty)^3 \to [0, \infty)$ *such that*

$$\|Y, Y'\|_{X,p} \leq K_p(\ell, \|f\|_{C_b^2}, \|y, y'\|_{X,p}, \|\mathbf{X}\|_p).$$

(ii) Let $(Y, Y') \in \mathcal{V}_X^p([0, T]; \mathbb{R}^k)$ *be the unique solution to the RDE with time discrete dependence* (2.11)*. Moreover, let* $(\widetilde{y}, \widetilde{y}') \in \mathcal{V}_X^p([0, T]; \mathbb{R}^k)$ *with corresponding solution* $(\widetilde{Y}, \widetilde{Y}') \in \mathcal{V}_X^p([0, T]; \mathbb{R}^k)$*, and suppose that* $\|Y, Y'\|_{X,p}, \|\widetilde{Y}, \widetilde{Y}'\|_{X,p} \leq K$*, for some* $K > 0$*. Then, we have the estimate*

$$|Y_0 - \widetilde{Y}_0| + \|Y, Y'; \widetilde{Y}, \widetilde{Y}'\|_{X,p} \lesssim |y_0 - \widetilde{y}_0| + \|y, y'; \widetilde{y}, \widetilde{y}'\|_{X,p},$$

where the implicit multiplicative constant depends on p*,* ℓ*,* $\|f\|_{C_b^3}$*,* K*, and* $\|\mathbf{X}\|_p$*.*

Proof. (i) On the interval $[0, r_1]$, we extend the vector field f to map into the space of controlled paths by setting

$$(F, F'): \mathcal{V}_X^p([0, r_1]; \mathbb{R}^k) \to \mathcal{V}_X^p([0, r_1]; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k)), \quad (F(Y), F'(Y, Y')) := (f(\bar{Y}), \mathrm{D}f(\bar{Y})\bar{Y}'),$$

with $(\bar{Y}, \bar{Y}') = ((Y, Y, \dots, Y), (Y', Y', \dots, Y')) \in \mathcal{V}_X^p([0, T]; \mathbb{R}^{k(\ell+1)})$. It follows analogously to Lemma 2.2.2 that the functional satisfies Assumption 2.1.4 (i) and (ii) with constants depending additionally on ℓ, that is, $C_F = \|f\|_{C_b^2}$ up to a multiplicative constant, which depends on p and ℓ, and $C_{F,K,X,X}$ depends on p, ℓ, $\|f\|_{C_b^3}$, K, for $K > 0$, and $\|X\|_p$. Note that it is sufficient to check Assumption 2.1.4 (ii) for $X = \widetilde{X}$. We can thus apply Theorem 2.1.3 to show that there exists a unique solution to the RDE (2.11) on the interval $[0, r_1)$. We now aim to solve the RDE (2.11) iteratively on the subintervals $[r_i, r_{i+1})$, $i = 1, \dots, \ell$, with $r_{\ell+1} = T$. Given the solution on $[r_{i-1}, r_i)$, with $r_0 = 0$, the value Y_{r_i} is determined by the jump of $\mathbf{X}$ at time r_i. We therefore consider $(y_i, y_i') \in \mathcal{V}_X^p([r_i, r_{i+1}]; \mathbb{R}^k)$, where

$$y_{i;t} = y_t + Y_{r_i-} - y_{r_i-} + F_{r_i-}(Y) X_{r_i-, r_i} + F'_{r_i-}(Y, Y') \mathbb{X}_{r_i-, r_i}, \quad t \in [r_i, r_{i+1}],$$

for every $i = 1, \dots, \ell$, and $(\alpha_i, \alpha_i') \in \mathcal{V}_X^p([r_i, r_{i+1}]; \mathbb{R}^{ik})$ be a fixed controlled path, with $\alpha_{i,t} = (Y_{r_1}, \dots, Y_{r_i})$, $t \in [r_i, r_{i+1})$. We set

$$\begin{aligned}
&(F, F'): \mathcal{V}_X^p([r_i, r_{i+1}]; \mathbb{R}^k) \to \mathcal{V}_X^p([r_i, r_{i+1}]; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k)), \\
&(F(Y), F'(Y, Y')) = (f((Y, \alpha_i, Y, \dots, Y)), \mathrm{D}f((Y, \alpha_i, Y, \dots, Y))(Y', \alpha_i', Y', \dots, Y'))
\end{aligned}$$

for $i = 1, \dots, \ell - 1$, and

$$(F(Y), F'(Y, Y')) = (f((Y, \alpha_\ell)), \mathrm{D}f((Y, \alpha_\ell))(Y', \alpha_\ell')),$$

for $i = \ell$. Analogously to Lemma 2.2.4, we can show that the functional satisfies Assumption 2.1.4 (i) and (ii) with constants depending additionally on ℓ, that is $C_F = \|f\|_{C_b^2}$ up to a multiplicative constant, which depends on p and ℓ, see the definition of $(\alpha_i, \alpha_i') \in \mathcal{V}_X^p([r_i, r_{i+1}])$, and $C_{F,K,X,X}$ depends on p, ℓ, $\|f\|_{C_b^3}$, K, for $K > 0$, and $\|X\|_p$. Note that it is again sufficient to check Assumption 2.1.4 (ii) for $X = \widetilde{X}$. We can thus again apply Theorem 2.1.3 to show that there exists a unique solution to the RDE (2.11) on the interval $[r_i, r_{i+1})$, that is

$$Y_t = y_{i;t} + \int_{r_i}^t F_s(Y)\,\mathrm{d}\mathbf{X}_s, \quad t \in [r_i, r_{i+1}),$$

for every $i = 1, \ldots, \ell$. Then, by pasting the solutions on each of these subintervals together, we obtain a unique global solution Y, which holds over the entire interval $[0, T]$.

The a priori estimate follows by iteratively combining the a priori estimate of Corollary 2.2.3, noting that $\alpha_{i,t} = (Y_{r_1}, \ldots, Y_{r_i})$, $t \in [r_i, r_{i+1})$, for $i = 1, \ldots, \ell$.

(ii) Local estimate on $[0, r_1]$*.* To apply the continuity result presented in Theorem 2.1.5 on the subinterval $[0, r_1]$, we need to ensure that the functionals satisfy the required assumptions.

For (F, F'), this is shown in the proof of (i), and as we aim to obtain continuity of the solution map as a function of the initial condition (y, y'), not the vector field f, on the interval $[0, r_1]$ we may consider $(F, F') = (\widetilde{F}, \widetilde{F}')$, so, $(F - \widetilde{F}, F' - \widetilde{F}') = 0$. Theorem 2.1.5 now gives that

$$\|Y' - \widetilde{Y}'\|_{p,[0,r_1]} + \|R^Y - R^{\widetilde{Y}}\|_{p,[0,r_1]} \lesssim |y_0 - \widetilde{y}_0| + \|y, y'; \widetilde{y}, \widetilde{y}'\|_{X,p,[0,r_1]},$$

where the implicit multiplicative constant depends on p, ℓ, $\|f\|_{C_b^3}$, K, and $\|\mathbf{X}\|_p$.

Local estimate on $[r_i, r_{i+1}]$*,* $i = 1, \ldots, \ell$*.* To apply the continuity result presented in Theorem 2.1.5, we need to ensure that the functionals satisfy the required assumptions. For $(F, F'), (\widetilde{F}, \widetilde{F}')$, this is shown in the proof of (i), and for $(F - \widetilde{F}, F' - \widetilde{F}')$, in the proof of part (ii) of Corollary 2.2.3, where the constant $C_{F-\widetilde{F}}$ depends additionally on ℓ. Theorem 2.1.5 then implies that

$$\|Y' - \widetilde{Y}'\|_{p,[r_i,r_{i+1}]} + \|R^Y - R^{\widetilde{Y}}\|_{p,[r_i,r_{i+1}]} \lesssim |Y_{r_i} - \widetilde{Y}_{r_i}| + \|y, y'; \widetilde{y}, \widetilde{y}'\|_{X,p,[r_i,r_{i+1}]},$$

where the implicit multiplicative depends on p, $\|f\|_{C_b^3}$, K and $\|\mathbf{X}\|_p$, see the definition of $(\alpha_i, \alpha_i'), (\tilde{\alpha}_i, \tilde{\alpha}_i') \in \mathcal{V}_X^p([r_i, r_{i+1}]; \mathbb{R}^{ik})$, $(y_i, y_i'), (\widetilde{y}_i, \widetilde{y}_i') \in \mathcal{V}_X^p([r_i, r_{i+1}]; \mathbb{R}^k)$.

Global estimate. Using the methods of the proof of Theorem 2.1.5, and applying the local estimates on the subintervals $[r_i, r_{i+1}]$, $i = 0, 1, \ldots, \ell$, one can then derive that

$$\|Y' - \widetilde{Y}'\|_p + \|R^Y - R^{\widetilde{Y}}\|_{\frac{p}{2}} \lesssim |y_0 - \widetilde{y}_0| + \|y, y'; \widetilde{y}, \widetilde{y}'\|_{X,p},$$

which implies the estimate. $\square$

2.2.4 RDEs with constant delay

Maybe the most prominent example of rough functional differential equations are RDEs with constant delay, cf. e.g. [65, 140, 21, 32, 20]. In the present subsection we consider the delayed rough differential equation

$$Y_t = y_t + \int_0^t f(Y_s, Y_{s-r_1}, \dots, Y_{s-r_\ell})\, \mathrm{d}\mathbf{X}_s, \quad t \in [0,T], \tag{2.12}$$

where $\mathbf{X} \in \mathcal{D}^p([0,T];\mathbb{R}^d)$ for $p \in (2,3)$, $(y, y') \in \mathcal{V}_X^p([0,T];\mathbb{R}^k)$, $f \in C_b^3(\mathbb{R}^{k(\ell+1)}; \mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$, and constant delays $0 < r_1 < \dots < r_\ell$ with $\ell \in \mathbb{N}$. To give a rigorous mathematical meaning to the RDE (2.12), we follow the approach of Neuenkirch, Nourdin and Tindel [140]: we assume that the driving rough path $\mathbf{X}$ is of the form

$$X_t = (Z_t, Z_{t-r_1}, \dots, Z_{t-r_\ell}), \quad t \in [0,T],$$

for a path $Z \in D^p([-r_\ell, T];\mathbb{R}^e)$ with $d = e(\ell+1)$. We extend the vector field f to map into the space of controlled paths by setting

$$(F(Y), F'(Y,Y')) := (f((Y,\alpha)), \mathrm{D}f((Y,\alpha))(Y',\alpha')),$$

for $(\alpha, \alpha') \in \mathcal{V}_X^p([0,T];\mathbb{R}^{k\ell})$, where

$$\alpha = (\alpha_1, \dots, \alpha_\ell) \quad \text{with} \quad \alpha_{j,t} := \begin{cases} Y_{t-r_j}, & t \in [r_j, T] \\ Y_{j;t}, & t \in [0, r_j) \end{cases}, \tag{2.13}$$

for fixed controlled paths $(Y_j, Y_j') \in \mathcal{V}^p_{Z_{\cdot-r_j}}([0,T];\mathbb{R}^k)$ and every $j = 1, \dots, \ell$. This includes the natural case $Y_j = \xi_{\cdot-r_j}$ for an initial path $\xi \in \mathcal{V}_Z^p([-r_\ell, T];\mathbb{R}^k)$.

Note that the postulated form of the rough path $\mathbf{X}$ is essential to ensure the well-posedness of the rough integral appearing in (2.12) and the extension of the solution Y to the interval $[-r_\ell, 0]$ is a standard and necessary way to give a meaning to $f(Y_s, Y_{s-r_1}, \dots, Y_{s-r_\ell})$ on the entire interval $[0,T]$.

For delayed RDEs of the form (2.12) driven by α-Hölder continuous rough paths, existence, uniqueness and continuity of the Itô–Lyons map were first proven in [140] for $\alpha \in (\frac{1}{3}, \frac{1}{2})$. These results were extended in [156] to α-Hölder continuous rough paths for $\alpha \in (\frac{1}{4}, \frac{1}{3})$. A paracontrolled distribution approach to RDEs with constant delay can be found in [146]. Based on the general results of Section 2.1, we can derive the follow proposition.

Proposition 2.2.6.

(i) In the above setting, there exists a unique solution to the delayed RDE (2.12)*. Moreover, there exists a componentwise non-decreasing function* $K_p : \mathbb{N} \times [0,\infty)^4 \to [0,\infty)$ *such that*

$$\|Y,Y'\|_{X,p} \leq K_p\Big(\ell, \|f\|_{C_b^2}, \|y,y'\|_{X,p}, \sum_{j=1}^{\ell} \|Y_j, Y_j'\|_{Z_{\cdot - r_j},p}, \|\mathbf{X}\|_p\Big).$$

(ii) Let $(Y,Y') \in \mathcal{V}_X^p([0,T];\mathbb{R}^k)$ *be the unique solution to the rough differential equation with constant delay* (2.12)*. Moreover, consider* $(\widetilde{y},\widetilde{y}') \in \mathcal{V}_X^p([0,T];\mathbb{R}^k)$*, and fixed controlled paths* $(\widetilde{Y}_j,\widetilde{Y}_j') \in \mathcal{V}_{Z_{\cdot - r_j}}^p([0,T];\mathbb{R}^k)$*,* $j = 1,\dots,\ell$*, with corresponding solution* $(\widetilde{Y},\widetilde{Y}') \in \mathcal{V}_X^p([0,T];\mathbb{R}^k)$*.*

Suppose that $\|Y,Y'\|_{X,p}$*,* $\|\widetilde{Y},\widetilde{Y}'\|_{X,p} \leq K$*, for some* $K > 0$*, and that* $\|Y_j\|_p$*,* $\|\widetilde{Y}_j\|_p$*,* $\|Y_j, Y_j\|_{X,p}$*,* $\|\widetilde{Y}_j,\widetilde{Y}_j'\|_{X,p} \leq L$*, for some* $L > 0$*,* $j = 1,\dots,\ell$*. Then, we have the estimate*

$$\begin{aligned} &|Y_0 - \widetilde{Y}_0| + \|Y,Y';\widetilde{Y},\widetilde{Y}'\|_{X,p} \\ &\quad \lesssim |y_0 - \widetilde{y}_0| + \|y,y';\widetilde{y},\widetilde{y}'\|_{X,p} + \sum_{j=1}^{\ell} |Y_{j;0} - \widetilde{Y}_{j;0}| + \sum_{j=1}^{\ell} \|Y_j,Y_j';\widetilde{Y}_j,\widetilde{Y}_j'\|_{Z_{\cdot - r_j},p}, \end{aligned}$$

where the implicit multiplicative constant depends on p*,* ℓ*,* r_1*,* T*,* $\|f\|_{C_b^3}$*,* K*,* L*, and* $\|\mathbf{X}\|_p$*.*

Proof. (i) The existence and uniqueness of the solution follows by iteratively applying part (i) of Corollary 2.2.3 to intervals of the length r_1.
More precisely, we consider the functional $(F,F') : \mathcal{V}_X^p([0,r_1];\mathbb{R}^k) \to \mathcal{V}_X^p([0,r_1];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$, for $(\alpha,\alpha') \in \mathcal{V}_X^p([0,r_1];\mathbb{R}^{k\ell})$ given by (2.13), and apply part (i) of Corollary 2.2.3 to show that there exists a unique solution to the RDE (2.12) on the interval $[0,r_1)$.

We now aim to solve the RDE (2.12) iteratively on the subintervals $[ir_1,(i+1)r_1]$, $i = 1,\dots,N-1$, assuming that $T = Nr_1$ for some $N \in \mathbb{N}$. Given the solution on $[(i-1)r_1, ir_1)$, the value Y_{ir_1} is determined by the jump of $\mathbf{X}$ on ir_1. We therefore consider $(y_i,y_i') \in \mathcal{V}_X^p([ir_1,(i+1)r_1];\mathbb{R}^k)$, where

$$y_{i;t} = y_t + Y_{ir_1-} - y_{ir_1-} + F_{ir_1-}(Y) X_{ir_1-,ir_1} + F'_{ir_1-}(Y,Y') \mathbb{X}_{ir_1-,ir_1}, \quad t \in [ir_1,(i+1)r_1],$$

for every $i = 1,\dots,N-1$, and

$$(F,F') : \mathcal{V}_X^p([ir_1,(i+1)r_1];\mathbb{R}^k) \to \mathcal{V}_X^p([ir_1,(i+1)r_1];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k)),$$

for $(\alpha,\alpha') \in \mathcal{V}_X^p([ir_1,(i+1)r_1];\mathbb{R}^{k\ell})$ given by (2.13). We again apply part (i) of Corollary 2.2.3 to show that there exists a unique solution to the RDE (2.12) on the interval $[ir_1,(i+1)r_1)$, that is

$$Y_t = y_{i;t} + \int_{ir_1}^t F_s(Y)\,\mathrm{d}\mathbf{X}_s, \quad t \in [ir_1,(i+1)r_1)$$

for every $i = 1, \dots, N-1$. Then, by pasting solutions on each of these subintervals together, we obtain a unique global solution Y to the RDE (2.12), which holds over the interval $[0,T]$.

The a priori estimate follows by iteratively combining the a priori estimate of Corollary 2.2.3, and by the definition of α in (2.13).

(ii) Local estimate on $[ir_1, (i+1)r_1]$, $i = 0, \dots, N-1$. To apply the continuity result presented in Theorem 2.1.5 on the subintervals $[ir_1, (i+1)r_1]$, we need to ensure that the functionals satisfy the required assumptions. This is given for $(F, F'), (\widetilde{F}, \widetilde{F}')$ in Lemma 2.2.4, and for $(F - \widetilde{F}, F' - \widetilde{F}')$ we refer to the proof of part (ii) of Corollary 2.2.3, and write $C_{F-\widetilde{F},i}$ for the corresponding constant. By Theorem 2.1.5, it then holds the estimate

$$\begin{aligned}
&\|Y' - \widetilde{Y}'\|_{p,[ir_1,(i+1)r_1]} + \|R^Y - R^{\widetilde{Y}}\|_{\frac{p}{2},[ir_1,(i+1)r_1]} \\
&\quad\lesssim |Y_{ir_1} - \widetilde{Y}_{ir_1}| + |F_{ir_1}(Y) - \widetilde{F}_{ir_1}(\widetilde{Y})| + \|y, y'; \widetilde{y}, \widetilde{y}'\|_{p,[ir_1,(i+1)r_1]} + C_{F-\widetilde{F},i} \\
&\quad\lesssim |y_0 - \widetilde{y}_0| + \|y, y'; \widetilde{y}, \widetilde{y}'\|_{X,p} + \sum_{j=1}^{\ell} |Y_{j;0} - \widetilde{Y}_{j;0}| + \|Y_j, Y_j'; \widetilde{Y}_j, \widetilde{Y}_j'\|_{Z_{\cdot - r_j},p} \\
&\qquad + \|Y' - \widetilde{Y}'\|_{p,[0,ir_1]} + \|R^Y - R^{\widetilde{Y}}\|_{\frac{p}{2},[0,ir_1]},
\end{aligned}$$

where the implicit multiplicative constant depends on p, ℓ, $\|f\|_{C_b^2}$, K, L, and $\|\mathbf{X}\|_p$, see the definition of $(\alpha, \alpha'), (\tilde{\alpha}, \tilde{\alpha}') \in \mathcal{V}_X^p([0,T]; \mathbb{R}^{k\ell})$.

Global estimate. Iteratively applying the local estimates and, as before, using that $\|\cdot\|_p \leq N \sum_{i=0}^{N-1} \|\cdot\|_{p,[ir_1,(i+1)r_1]}$, one can then derive the estimate. $\square$

2.2.5 RDEs with variable delay

Rough differential equations with variable delay represent a slight generalization of RDEs with constant delay. More precisely, let us consider the rough differential equation with variable delay

$$Y_t = y_t + \int_0^t f(Y_s, Y_{s-\eta(s)}) \,\mathrm{d}\mathbf{X}_s, \quad t \in [0,T], \tag{2.14}$$

where $\mathbf{X} \in \mathcal{D}^p([0,T]; \mathbb{R}^d)$ for $p \in (2,3)$, $(y, y') \in \mathcal{V}_X^p([0,T]; \mathbb{R}^k)$, $f \in C_b^3(\mathbb{R}^{2k}; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k))$, and $\eta(\cdot)$ be a bounded continuous function with $\eta(t) \geq \varepsilon$, $t \in [0,T]$, for some $\varepsilon > 0$, and $\bar{\eta} = \sup\{\eta(t) - t : t \in [0,T]\}$. We assume that the driving rough path $\mathbf{X}$ is of the form

$$X_t = (Z_t, Z_{t-\eta(t)}), \quad t \in [0,T],$$

for a path $Z \in D^p([-\bar{\eta}, T]; \mathbb{R}^e)$ with $d = 2e$. We extend the vector field f into the space of controlled paths by setting

$$\begin{aligned}
&(F, F') \colon \mathcal{V}_X^p([0,T]; \mathbb{R}^k) \to \mathcal{V}_X^p([0,T]; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k)), \\
&(F(Y), F'(Y, Y')) = (f((Y, \alpha)), \mathrm{D}f((Y, \alpha))(Y', \alpha')),
\end{aligned}$$

for $(\alpha, \alpha') \in \mathcal{V}_X^p([0,T];\mathbb{R}^k)$, where

$$\alpha_t := \begin{cases} Y_{t-\eta(t)}, & t \geq \eta(t) \\ Y_{\eta;t}, & t < \eta(t) \end{cases},$$

for some fixed controlled path $(Y_\eta, Y_\eta') \in \mathcal{V}^p_{Z_{\cdot-\eta(\cdot)}}([0,T];\mathbb{R}^k)$.

Corollary 2.2.7.

(i) In the above setting, there exists a unique solution to the delayed RDE (2.14)*. Moreover, there exists a componentwise non-decreasing function* $K_p\colon (0,\infty) \times [0,\infty)^4 \to [0,\infty)$ *such that*

$$\|Y, Y'\|_{X,p} \leq K_p(\varepsilon^{-1}, \|f\|_{C_b^2}, \|y, y'\|_{X,p}, \|Y_\eta, Y_\eta'\|_{X,p}, \|\mathbf{X}\|_p).$$

(ii) Let $(Y, Y') \in \mathcal{V}_X^p([0,T];\mathbb{R}^k)$ *be the unique solution to the RDE with variable delay* (2.14)*. Moreover, consider* $(\widetilde{y}, \widetilde{y}') \in \mathcal{V}_X^p([0,T];\mathbb{R}^k)$*, and a fixed controlled path* $(\widetilde{Y}_\eta, \widetilde{Y}_\eta') \in \mathcal{V}^p_{Z_{\cdot-\eta(\cdot)}}([0,T];\mathbb{R}^k)$ *with corresponding solution* $(\widetilde{Y}, \widetilde{Y}') \in \mathcal{V}_X^p([0,T];\mathbb{R}^k)$*. Suppose that* $\|Y, Y'\|_{X,p}$*,* $\|\widetilde{Y}, \widetilde{Y}'\|_{X,p} \leq K$*, for some* $K > 0$*, and* $\|Y_\eta, Y_\eta\|_{X,p}$*,* $\|\widetilde{Y}_\eta, \widetilde{Y}_\eta'\|_{X,p} \leq L$*, for some* $L > 0$*. Then, we have the estimate*

$$\begin{aligned}&|Y_0 - \widetilde{Y}_0| + \|Y, Y'; \widetilde{Y}, \widetilde{Y}'\|_{X,p} \\ &\quad \lesssim |y_0 - \widetilde{y}_0| + \|y, y'; \widetilde{y}, \widetilde{y}'\|_{X,p} + |Y_{\eta;0} - \widetilde{Y}_{\eta;0}| + \|Y_\eta, Y_\eta'; \widetilde{Y}_\eta, \widetilde{Y}_\eta'\|_{Z_{\cdot-\eta(\cdot)},p},\end{aligned}$$

where the implicit multiplicative constant depends on p*,* ε*,* T*,* η*,* $\|f\|_{C_b^3}$*,* K*,* L*, and* $\|\mathbf{X}\|_p$*.*

Proof. (i) The existence and uniqueness of the solution follows by iteratively applying part (i) of Corollary 2.2.3 to intervals of the length ε, see the proof of part (i) of Proposition 2.2.6 The a priori estimate follows analogously.

(ii) The continuity of the solution map follows analogously to Proposition 2.2.6 (ii). □

2.3 Application to stochastic differential equations with delay

One main application of rough path theory is a pathwise and robust approach to stochastic differential equations, see e.g. [71]. In this section we show how a càdlàg martingale and its delayed version can be lifted to a joint random rough path in the spirit of stochastic Itô integration. Consequently, this allows to apply the results on rough functional differential equations, provided in Section 2.1, to Itô stochastic differential equations (SDEs) with constant delay.

Throughout the entire section, let us consider constant delays $0 < r_1 < \cdots < r_\ell$ with $\ell \in \mathbb{N}$, and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a complete and right-continuous filtration $(\mathcal{F}_t)_{t\in[-r_\ell,T]}$. Let $Z = (Z_t)_{t\in[-r_\ell,T]}$ be an e-dimensional square-integrable càdlàg martingale that is defined on $(\Omega, \mathcal{F}, \mathbb{P})$, with $Z_t = 0$ for $t < 0$. The space of all square-integrable random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ is denoted by L^2 and equipped with the standard L^2-norm.

2.3.1 Delayed martingales as rough paths

The aim of this subsection is to construct a random rough path above the stochastic process $X = (X_t)_{t\in[0,T]}$, defined as

$$X_t := (Z_t, Z_{t-r_1}, \ldots, Z_{t-r_\ell}), \qquad t \in [0,T],$$

in the spirit of stochastic Itô integration. Recall, for a martingale $(S_t)_{t\in[0,T]}$, stochastic Itô integration allows to define the integral $\int_0^t \varphi_s \,\mathrm{d}S_s$ if $(\varphi_t)_{t\in[0,T]}$ is a stochastic process with left-continuous sample paths with right-limits which is adapted to the augmented filtration generated by $(S_t)_{t\in[0,T]}$. For a comprehensive introduction to stochastic integration see, e.g., [147]. In the following, when writing a stochastic integral, like $\int_0^t \varphi_s \,\mathrm{d}S_s$, we will always implicitly refer to the augmented filtration generated by $(S_t)_{t\in[0,T]}$ if not explicitly stated otherwise.

To construct a random rough path above the stochastic process $X = (X_t)_{t\in[0,T]}$, the main challenge is to establish the existence of the random integral $\int_0^t Z_{t-r_{j_1}} \,\mathrm{d}Z_{t-r_{j_2}}$ for $j_1 < j_2$ since $(Z_{t-r_{j_1}})_{t\in[0,T]}$ is, in general, not adapted to the augmented filtration generated by $(Z_{t-r_{j_2}})_{t\in[0,T]}$.

As a first step to construct a random rough path above the stochastic process X, in the next lemma, we derive the existence of an auxiliary process, inspired by the quadratic co-variation of martingales.

Lemma 2.3.1. *Let $Z = (Z_t)_{t\in[-r_\ell,T]}$ be an e-dimensional square-integrable càdlàg martingale that is defined on $(\Omega, \mathcal{F}, \mathbb{P})$, with $Z_t = 0$ for $t < 0$. Then, for $i_1, i_2 = 1, \ldots, e$, $j_1, j_2 = 0, \ldots, \ell$, $j_1 \neq j_2$, we have*

$$\mathbb{E}\Bigg[\sup_{t\in[0,T]}\Big|\sum_{k=0}^{N_n-1} Z^{i_1}_{t^n_k\wedge t-r_{j_1},t^n_{k+1}\wedge t-r_{j_1}} Z^{i_2}_{t^n_k\wedge t-r_{j_2},t^n_{k+1}\wedge t-r_{j_2}} - \sum_{s\le t}\Delta_s Z^{i_1}_{\cdot-r_{j_1}}\Delta_s Z^{i_2}_{\cdot-r_{j_2}}\Big|^2\Bigg] \longrightarrow 0$$

along any sequence of partitions $\mathcal{P}^n = \{0 = t^n_0 < t^n_1 < \cdots < t^n_{N_n} = T\}$, $n \in \mathbb{N}$, of the interval $[0,T]$ with vanishing mesh size, so that $|\mathcal{P}^n| \to 0$ as $n \to \infty$. Here, we write $\Delta_t H = H_{t-,t}$, with $H_{t-} = \lim_{s\uparrow t} H_s$, for the jump of a stochastic process H at time t.

We define the stochastic process

$$[Z^{i_1}_{\cdot-r_{j_1}}, Z^{i_2}_{\cdot-r_{j_2}}]_t := \sum_{s\le t}\Delta_s Z^{i_1}_{\cdot-r_{j_1}}\Delta_s Z^{i_2}_{\cdot-r_{j_2}}, \quad t \in [0,T].$$

This process is càdlàg and has $\mathbb{P}$-almost surely finite $\frac{p}{2}$-variation, that is, $[Z^{i_1}_{\cdot - r_{j_1}}, Z^{i_2}_{\cdot - r_{j_2}}] \in D^{\frac{p}{2}}([0,T];\mathbb{R})$ $\mathbb{P}$-almost surely.

Proof. We assume w.l.o.g. that $j_1 = 0$, and write $r = r_{j_2}$, and $M = Z^{i_1}$, $\widetilde{M} = Z^{i_2}$. For $n \in \mathbb{N}$, we define

$$H^n_t := \sum_{k=0}^{N_n - 1} \widetilde{M}_{t^n_k - r, t^n_{k+1} - r} \mathbf{1}_{(t^n_k, t^n_{k+1}]}(t), \quad t \in [0,T],$$

and note that for $|\mathcal{P}^n| < r$, H^n is indeed a simple predictable process, see [147, Chapter II]. The Itô integral is then given by

$$\int_0^t H^n_s \,\mathrm{d}M_s = \sum_{k=0}^{N_n - 1} \widetilde{M}_{t^n_k \wedge t - r, t^n_{k+1} \wedge t - r} M_{t^n_k \wedge t, t^n_{k+1} \wedge t}.$$

We now aim to show that

$$\mathbb{E}\bigg[\int_0^T (H^n_s - H_s)^2 \,\mathrm{d}[M]_s\bigg] \to 0, \quad \text{as} \quad n \to \infty, \tag{2.15}$$

where $H := \Delta_\cdot \widetilde{M}_{\cdot - r}$, and $[\cdot]$ denotes the quadratic variation. Using the localizing sequence $\tau_m = T \wedge \inf\{t : |\widetilde{M}_t| \geq m\}$, $m \in \mathbb{N}$, and replacing H^n by $H^n_{\cdot \wedge \tau_m}$ and H by $H_{\cdot \wedge \tau_m}$, we may assume that the integrand is uniformly bounded, so that we can apply the dominated convergence theorem. Since $H^n \to H$ converges pointwise as $n \to \infty$, this shows (2.15).

By [96, Chapter I.4], it follows that

$$\mathbb{E}\bigg[\sup_{t \in [0,T]} \Big| \int_0^t H^n_s \,\mathrm{d}M_s - \int_0^t H_s \,\mathrm{d}M_s \Big|^2\bigg] \to 0$$

as $n \to \infty$, thus uniformly in L^2, and as

$$\int_0^t H_s \,\mathrm{d}M_s = \sum_{s \leq t} \Delta_s \widetilde{M}_{\cdot - r} \Delta_s M,$$

this implies the convergence result. Further,

$$4[M, \widetilde{M}_{\cdot - r}] = \sum_{s \leq \cdot} (\Delta_s M + \Delta_s \widetilde{M}_{\cdot - r})^2 - \sum_{s \leq \cdot} (\Delta_s M - \Delta_s \widetilde{M}_{\cdot - r})^2$$

has càdlàg sample paths of finite 1-variation, as both terms on the right hand side are monotonically increasing, which implies that $[M, \widetilde{M}_{\cdot - r}]$ has $\mathbb{P}$-almost surely finite $\frac{p}{2}$-variation, and concludes the proof. □

Proposition 2.3.2. *Let $p \in (2,3)$, and let $Z = (Z_t)_{t \in [-r,T]}$ be an e-dimensional square-integrable càdlàg martingale that is defined on $(\Omega, \mathcal{F}, \mathbb{P})$, with $Z_t = 0$ for $t < 0$. We set*

$X = (Z, Z_{\cdot - r_1}, \ldots, Z_{\cdot - r_\ell})$. *Then, X can be lifted to a random rough path, by defining* $\mathbf{X} = (X, \mathbb{X}) \in \mathcal{D}^p([0,T];\mathbb{R}^d)$, *$\mathbb{P}$-almost surely, with $d = e(\ell+1)$, where*

$$\mathbb{X}^{ij}_{s,t} := \int_s^t X^i_{u-} \,\mathrm{d}X^j_u - X^i_s X^j_{s,t} := \int_0^t X^i_{u-} \,\mathrm{d}X^j_u - \int_0^s X^i_{u-} \,\mathrm{d}X^j_u - X^i_s X^j_{s,t},$$

for $i, j = 1, \ldots, d$ with $i = j$ and $i > j$ such that $X^i = Z^{i_1}_{\cdot - r_{j_1}}$, $X^j = Z^{i_2}_{\cdot - r_{j_2}}$ with $i_1, i_2 = 0, \ldots, e$, $j_1, j_2 = 0, \ldots, \ell$, $j_1 > j_2$, and else

$$\mathbb{X}^{ji}_{s,t} = -\mathbb{X}^{ij}_{s,t} + X^i_{s,t} X^j_{s,t} - [X^i, X^j]_{s,t}$$

for any $(s,t) \in \Delta_T$, and where the integration is defined as a stochastic Itô integral, and $[X^i, X^j]$ is defined in Lemma 2.3.1.

Remark 2.3.3. *The stochastic integral $\int_0^t Z^{i_1}_{u - r_{j_1}-} \,\mathrm{d}Z^{i_2}_{u - r_{j_2}}$ can be defined using classical stochastic Itô integration if $j_1 > j_2$. Indeed, the stochastic process $(Z^{i_2}_{t - r_{j_2}})_{t \in [0,T]}$ is a martingale and the stochastic process $(Z^{i_1}_{t - r_{j_1}})_{t \in [0,T]}$ is predictable, both with respect to the filtration $(\mathcal{F}_{t - r_{j_2}})_{t \in [0,T]}$ with $\mathcal{F}_t := \{\Omega, \emptyset\}$ for $t < 0$.*
Further, for $i \geq j$, we have that $\mathbb{X}^{ij}_{s,t} := \int_0^t X^i_{u-} \,\mathrm{d}X^j_u - \int_0^s X^i_{u-} \,\mathrm{d}X^j_u - X^i_s X^j_{s,t} = \int_s^t X^i_{s,u-} \,\mathrm{d}X^j_u$ $= \int_s^t \int_s^{u-} \mathrm{d}X^i_r \,\mathrm{d}X^j_u$, that is, $\mathbb{X}^{ij}_{s,t}$ coincides with the 2-fold iterated integral, for $(s,t) \in \Delta_T$.

Proof of Proposition 2.3.2. First, by definition Chen's relation does hold: Let $0 \leq s \leq v \leq t \leq T$. Then, we have that

$$\begin{aligned}
&\mathbb{X}^{ii}_{s,v} + \mathbb{X}^{ii}_{v,t} + X^i_{s,v} X^i_{v,t} \\
&\quad = \int_s^t X^i_{u-} \,\mathrm{d}X^i_u - X^i_s X^i_{s,v} - X^i_v X^i_{v,t} + X^i_{s,v} X^i_{v,t} \\
&\quad = \int_s^t X^i_{u-} \,\mathrm{d}X^i_u - X^i_s X^i_{s,t} \\
&\quad = \mathbb{X}^{ii}_{s,t},
\end{aligned}$$

similarly for $\mathbb{X}^{ij}$, and

$$\begin{aligned}
&\mathbb{X}^{ji}_{s,v} + \mathbb{X}^{ji}_{v,t} + X^j_{s,v} X^i_{v,t} \\
&\quad = -\mathbb{X}^{ij}_{s,v} + X^i_{s,v} X^j_{s,v} - [X^i, X^j]_{s,v} - \mathbb{X}^{ij}_{v,t} + X^i_{v,t} X^j_{v,t} - [X^i, X^j]_{v,t} + X^j_{s,v} X^i_{v,t} \\
&\quad = -\mathbb{X}^{ij}_{s,t} + X^i_{s,v} X^j_{v,t} + X^i_{s,v} X^j_{s,v} + X^i_{v,t} X^j_{v,t} + X^i_{v,t} X^j_{s,v} - [X^i, X^j]_{s,t} \\
&\quad = -\mathbb{X}^{ij}_{s,t} + X^i_{s,t} X^j_{s,t} - [X^i, X^j]_{s,t} \\
&\quad = \mathbb{X}^{ji}_{s,t}.
\end{aligned}$$

Further, Z has $\mathbb{P}$-almost surely finite p-variation, see e.g. [120], therefore $X \in D^p([0,T];\mathbb{R}^d)$ $\mathbb{P}$-almost surely. Since the maps $s \mapsto \mathbb{X}_{s,t}$ for fixed t, and $t \mapsto \mathbb{X}_{s,t}$ for fixed s are both càdlàg, it thus remains to show that $\|\mathbb{X}\|_{\frac{p}{2}} < \infty$ $\mathbb{P}$-almost surely.

We define the dyadic stopping times $(\tau_k^n)_{n,k\in\mathbb{N}}$ by

$$t_0^n := 0, \quad \tau_{k+1}^n := \inf\{t \geq \tau_k^n \, : \, |X_t - X_{\tau_k^n}| \geq 2^{-n}\} \wedge T.$$

For $t \in [0,T]$ and $n \in \mathbb{N}$ we introduce the dyadic approximation

$$X_t^n := \sum_{k=0}^{\infty} X_{\tau_k^n} \mathbf{1}_{(\tau_k^n, \tau_{k+1}^n]}(t) \quad \text{and} \quad \int_0^t X_r^{i,n} \, \mathrm{d}X_r^j := \sum_{k=0}^{\infty} X_{\tau_k^n}^i X_{\tau_k^n \wedge t, \tau_{k+1}^n \wedge t}^j,$$

for $i = j$ or $i \neq j$ such that $X^i = Z^{i_1}_{\cdot - r_{j_1}}$, $X^j = Z^{i_2}_{\cdot - r_{j_2}}$ for $i_1, i_2 = 1, \dots, e$, $j_1, j_2 = 0, \dots, \ell$, $j_1 > j_2$.

We now show that for almost every $\omega \in \Omega$, for every $t \in [0,T]$ and for every $\varepsilon \in (0,1)$, there exists a constant $C = C(\omega, \varepsilon)$ such that for all $n \in \mathbb{N}$, we have

$$\left| \left(\int_0^t X_u^{i,n} \, \mathrm{d}X_u^j - \int_0^t X_u^i \, \mathrm{d}X_u^j \right)(\omega) \right| \leq C 2^{-n(1-\varepsilon)}. \tag{2.16}$$

Applying the Burkholder–Davis–Gundy inequality, we have that

$$\mathbb{E}\left[\left(\sup_{t\in[0,T]} \int_0^t (X_u^{i,n} - X_u^i) \, \mathrm{d}X_u^j \right)^2 \right] \lesssim \mathbb{E}\left[\int_0^T (X_u^{i,n} - X_u^i)^2 \, \mathrm{d}[X^j]_u \right] \lesssim 2^{-2n}, \quad n \in \mathbb{N},$$

where the implicit multiplicative constant depends on the quadratic variation $[X^j]$ of X^j. Combining this with Chebyshev's inequality, we obtain for any $\varepsilon \in (0,1)$ that

$$\mathbb{P}\left(\left| \int_0^t (X_u^{i,n} - X_u^i) \, \mathrm{d}X_u^j \right| \geq 2^{-n(1-\varepsilon)} \right) \lesssim 2^{2n(1-\varepsilon)} 2^{-2n} = 2^{-n\varepsilon}.$$

So by the Borel–Cantelli lemma, we have that

$$\sup_{t\in[0,T]} \left(\int_0^t X_u^{i,n} \, \mathrm{d}X_u^j - \int_0^t X_u^i \, \mathrm{d}X_u^j \right) \lesssim 2^{-n(1-\varepsilon)},$$

where the implicit multiplicative constant is a random variable which does not depend on n, which shows (2.16). Proceeding as in the proof of [124, Theorem 3.1], we can show that $\|\mathbb{X}^{ij}\|_{\frac{p}{2}} < \infty$ $\mathbb{P}$-almost surely. Further, let $i \neq j$ as above, then we have for any $(s,t) \in \Delta_T$ that

$$|\mathbb{X}_{s,t}^{ji}|^{\frac{p}{2}} \lesssim \|\mathbb{X}^{ij}\|_{\frac{p}{2},[s,t]}^{\frac{p}{2}} + \|X\|_{p,[s,t]}^p + \|[X^i, X^j]\|_{\frac{p}{2},[s,t]}^{\frac{p}{2}}.$$

Lemma 2.3.1 then ensures that $\|\mathbb{X}^{ji}\|_{\frac{p}{2}} < \infty$ $\mathbb{P}$-almost surely. □

Remark 2.3.4. *For a fractional Brownian motion with Hurst index H and its delayed version a joint rough path was constructed in [140] based on the Russo–Vallois integral [151], assuming that $H > \frac{1}{3}$. This construction was generalized in [156] allowing for $H > \frac{1}{4}$. A related construction of a delayed rough path above a fractional Brownian motion can* *be found in [21]. For a standard Brownian motion a delayed rough path was also defined in [32] based on stochastic Itô integration. While the delayed rough path provided in Proposition 2.3.2 corresponds to stochastic Itô integration, see Proposition 2.3.5 below, the aforementioned constructions of delayed rough paths above (fractional) Brownian motion correspond to Stratonovich integration.*

2.3.2 SDEs with delay as random RDEs

Let us consider the SDE with constant delay

$$\begin{aligned} &Y_t = y_0 + \int_0^t f(Y_{s-}, Y_{s-r_1-}, \dots, Y_{s-r_\ell-})\,\mathrm{d}Z_s, \quad t \in [0,T], \\ &Y_t = y_t, \quad t \in [-r_\ell, 0), \end{aligned} \tag{2.17}$$

where $y \in D^{\frac{p}{2}}([-r_\ell, T]; \mathbb{R}^k)$, $f \in C_b^3(\mathbb{R}^{k(\ell+1)}; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k))$ and the integral is defined as a stochastic Itô integral. For a comprehensive introduction to stochastic Itô integration and SDEs we refer, e.g., to the textbook [147]. It is known that the SDE (2.17) possesses a unique (strong) solution, see e.g. [147, Chapter V, Theorem 7]. It turns out that the solutions to the SDE (2.17) and to the RDE (2.12) driven by the random rough path $\mathbf{X} = (X, \mathbb{X})$, with $\mathbb{X}$ as defined in Proposition 2.3.2, coincide $\mathbb{P}$-almost surely.

Proposition 2.3.5. *Let $p \in (2,3)$, and let $Z = (Z_t)_{t\in[-r,T]}$ be an e-dimensional square-integrable càdlàg martingale, that is defined on $(\Omega, \mathcal{F}, \mathbb{P})$, with $Z_t = 0$ for $t < 0$. We set $X := (Z, Z_{\cdot-r_1}, \dots, Z_{\cdot-r_\ell})$, and let $\mathbf{X} = (X, \mathbb{X})$ be the random rough path, with $\mathbb{X}$ defined as in Proposition 2.3.2.*

(i) Let (V, V') be an adapted stochastic process such that $(V(\omega), V'(\omega)) \in \mathcal{V}^p_{X(\omega)}$ for almost every $\omega \in \Omega$. Then, the rough integral exists and coincides $\mathbb{P}$-almost surely with the stochastic Itô integral, that is

$$\int_0^T V_u\,\mathrm{d}\mathbf{X}_u = \int_0^T V_{u-}\,\mathrm{d}X_u, \quad \mathbb{P}\text{-almost surely.}$$

(ii) The solution of the SDE (2.17) driven by X, and the solution of the RDE (2.12) driven by $\mathbf{X}$, coincide $\mathbb{P}$-almost surely.

Proof. (i) Step 1. [73, Theorem 31] gives that

$$\int_0^T V_u\,\mathrm{d}\mathbf{X}_u = \lim_{|\mathcal{P}|\to 0} \sum_{(s,t)\in\mathcal{P}} (V_s X_{s,t} + V_s' \mathbb{X}_{s,t}),$$

where the limit is taken over any sequence of partitions $\mathcal{P}$ of the interval $[0,T]$ with mesh size $|\mathcal{P}| \to 0$, and it is known that

$$\sum_{(s,t)\in\mathcal{P}} V_s X_{s,t} \to \int_0^T V_{u-}\,\mathrm{d}X_u,$$

in probability as $|\mathcal{P}|\to 0$, see e.g. [147, Chapter II, Theorem 21], therefore the convergence holds $\mathbb{P}$-almost surely, possibly along some subsequence.

Step 2. We are left to show that

$$\lim_{|\mathcal{P}|\to 0} \sum_{(s,t)\in\mathcal{P}} V'_s \mathbb{X}_{s,t} = 0, \tag{2.18}$$

$\mathbb{P}$-almost surely, along some subsequence. It suffices to show that for $i, j = 1, \dots, d$,

$$\sup_{\tau\in[0,T]} \Big| \sum_{(s,t)\in\mathcal{P}\cap[0,\tau]} \mathbb{X}^{ij}_{s,t} \Big| \to 0, \qquad \text{as} \qquad |\mathcal{P}|\to 0, \tag{2.19}$$

in probability, which then implies $\mathbb{P}$-almost sure convergence, along some subsequence: if V' is $\mathbb{P}$-almost surely piecewise constant, then (2.19) implies (2.18). Otherwise, for any $\varepsilon > 0$, there exists a suitable piecewise constant approximation $V'^{,\varepsilon}$ of V' such that

$$\|V' - V'^{,\varepsilon}\|_\infty \le \varepsilon,$$

$\mathbb{P}$-almost surely, see [6, Proposition B.1]. By a standard interpolation argument (e.g. [74, Proposition 5.5]), it follows, for any $q > p$, that

$$\|V' - V'^{,\varepsilon}\|_q \le \|V' - V'^{,\varepsilon}\|_p^{\frac{p}{q}} \|V' - V'^{,\varepsilon}\|_\infty^{1-\frac{p}{q}} \lesssim C\varepsilon^{1-\frac{p}{q}},$$

$\mathbb{P}$-almost surely, where the implicit multiplicative constant C is a random variable which does depend only on p, q and $\|V'\|_p$. Using [161, (5.1)], we obtain that

$$\Big| \sum_{(s,t)\in\mathcal{P}} V'_s \mathbb{X}_{s,t} - \sum_{(s,t)\in\mathcal{P}} V'^{,\varepsilon}_s \mathbb{X}_{s,t} \Big| \le \Big(1 + \zeta\Big(\frac{1}{q} + \frac{2}{p}\Big)\Big) \|V' - V'^{,\varepsilon}\|_q \|\mathbb{X}\|_{\frac{p}{2}} \lesssim C\varepsilon^{1-\frac{p}{q}},$$

$\mathbb{P}$-almost surely for any partition $\mathcal{P}$, where the implicit multiplicative constant C is a random variable which does depend only on p, q, $\|V'\|_p$ and $\|\mathbb{X}\|_{\frac{p}{2}}$. Consequently, if

$$\lim_{|\mathcal{P}|\to 0} \sum_{(s,t)\in\mathcal{P}} V'^{,\varepsilon}_s \mathbb{X}_{s,t} = 0,$$

holds $\mathbb{P}$-almost surely, then so does (2.18), and it suffices to show (2.19).

Step 3. From here on, for the proof of (2.19), we consider the sequence of partitions $\mathcal{P}^n = \{0 = t_0^n < t_1^n < \ldots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$, of the interval $[0, T]$ with vanishing mesh size, so that $|\mathcal{P}^n|\to 0$ as $n \to \infty$. Moreover, let $i \ge j$.

Recall that

$$\mathbb{X}^{ij}_{t_k^n, t_{k+1}^n} = \int_{t_k^n}^{t_{k+1}^n} X^i_{t_k^n, u-} \,\mathrm{d}X^j_u, \qquad k = 0, \dots, N_n - 1.$$

Thus, the Burkholder–Davis–Gundy inequality gives that

$$\mathbb{E}\Big[\sup_{\tau\in[0,T]} \Big| \sum_{(s,t)\in\mathcal{P}^n\cap[0,\tau]} \mathbb{X}^{ij}_{s,t} \Big|^2\Big] \lesssim \mathbb{E}\Big[\int_0^T |X^i_{u_{|\mathcal{P}^n}, u-}|^2 \,\mathrm{d}[X^j]_u\Big],$$

where we write $t_{|\mathcal{P}^n} := \max\{t_k^n \in \mathcal{P}^n : t_k^n \le t\}$, $t \in [0,T]$. Proceeding as in the proof of [76, Lemma 6.1], one can then show that

$$\mathbb{E}\bigg[\int_0^T |X^i_{u_{|\mathcal{P}^n},u-}|^2 \,\mathrm{d}[X^j]_u\bigg] \to 0, \qquad \text{as} \qquad n \to \infty,$$

which gives (2.19).

Therefore, by definition and Lemma 2.3.1 it holds that

$$\begin{aligned}\sum_{k=0}^{N_n-1} \mathbb{X}^{ji}_{t_k^n,t_{k+1}^n} &= -\sum_{k=0}^{N_n-1} \mathbb{X}^{ij}_{t_k^n,t_{k+1}^n} + X^i_{t_k^n,t_{k+1}^n} X^j_{t_k^n,t_{k+1}^n} - [X^i,X^j]_{t_k^n,t_{k+1}^n} \\ &= -\sum_{k=0}^{N_n-1} \mathbb{X}^{ij}_{t_k^n,t_{k+1}^n} + \sum_{k=0}^{N_n-1} X^i_{t_k^n,t_{k+1}^n} X^j_{t_k^n,t_{k+1}^n} - \sum_{s\le T} \Delta_s X^i \Delta_s X^j \\ &\to 0,\end{aligned}$$

as $n \to \infty$, where the convergence holds uniformly in probability, which then concludes the proof.

(ii) Let Y be the solution of the rough differential equation (2.12) driven by the random rough path $\mathbf{X} = (X, \mathbb{X})$, see part (i) of Proposition 2.2.6. We note that the assumption on (V, V') in (i) does fit into this setting, where $(V(\omega), V'(\omega)) = (F(Y(\omega)), F'(Y(\omega), Y'(\omega))$ for some functional F, see Section 2.2.4. As the rough and Itô integral do coincide $\mathbb{P}$-almost surely by (i), we infer that Y is also a solution of the SDE (2.17), which has a unique solution (by e.g. [147, Chapter V, Theorem 7]). □

Remark 2.3.6. *As a consequence of Proposition 2.2.6 and part (ii) of Proposition 2.3.5, one can apply the continuity of the Itô–Lyons map (Theorem 2.1.5) to derive pathwise stability results for stochastic differential equations with delay like* (2.17)*. In particular, the map $y \mapsto Y$, mapping the initial path y to the associated solution Y of the SDE* (2.17)*, is continuous on the space of controlled paths, which resolves an old observation, pointed out by Mohammed [136], about the non-continuity of the flow of stochastic differential equations with delay. The latter is a consequence of the discontinuity of stochastic integration when using an unsuitable topology for the integrands.*

Remark 2.3.7. *While we considered square-integrable martingales and the associated stochastic differential equations with constant delay in this section, the presented results can be generalized in a fairly straightforward manner to:*

(i) càdlàg local martingales using standard localization arguments;

(ii) càdlàg semimartingales using the classical estimates for Young integrals, see e.g. [75, Proposition 2.4] and [124, Theorem 3.1], to show that one can suitably lift X to a random rough path with additional Young integrals;

(iii) Young semimartingales (also known as semimartingales in the sense of Norvaiša [142]), i.e. $Z = M + \varphi$*, for some martingale* M *and some càdlàg adapted process* φ *with* $\varphi(\omega) \in D^q([0,T];\mathbb{R}^e)$ *for almost every* $\omega \in \Omega$*, for some* $q \in [1,2)$*;*

(iv) SDEs/RDEs with variable delay of the form (2.14)*, as long as* η *is assumed to be bounded with* $\eta(t) \geq \varepsilon$*,* $t \in [0,T]$*, for some* $\varepsilon > 0$*.*

Chapter 3

Pathwise convergence of the Euler scheme for rough and stochastic differential equations

Stochastic differential equations serve as models for dynamical systems which evolve randomly in time, and are fundamental mathematical objects, essential to numerous applications in finance, engineering, biology and beyond. In a fairly general form, a stochastic differential equation (SDE) is given by

$$Y_t = y_0 + \int_0^t b(s, Y_s)\,\mathrm{d}s + \int_0^t \sigma(s, Y_s)\,\mathrm{d}X_s, \qquad t \in [0,T], \tag{3.1}$$

where $y_0 \in \mathbb{R}^k$ is the initial condition, $b\colon [0,T] \times \mathbb{R}^k \to \mathbb{R}^k$ and $\sigma\colon [0,T] \times \mathbb{R}^k \to \mathbb{R}^{k \times d}$ are coefficients, and the driving signal $X = (X_t)_{t \in [0,T]}$ is a d-dimensional stochastic process which models the random noise affecting the system.

Assuming that X is a càdlàg semimartingale, such as a Brownian motion or a Lévy process, and the coefficients b, σ are suitably regular, it is well-known that (3.1) is well-posed as an Itô SDE. That is, $\int_0^t \sigma(s, Y_s)\,\mathrm{d}X_s$ can be defined as a stochastic Itô integral, and the equation admits a unique adapted solution $Y = (Y_t)_{t \in [0,T]}$; see e.g. [147]. Unfortunately, such SDEs, including many of those which appear in practical applications, can rarely be solved explicitly, which has led to a vast literature on various numerical approximations of the solutions to SDEs; see e.g. [110].

One of the most common approaches to numerically approximate the solution of a stochastic differential equation is to rely on a time-discretized modification of the equation. This type of discretization is implemented in particular by the Euler scheme (also called the Euler–Maruyama scheme) and its higher order variants. For the SDE (3.1), the (first order) Euler approximation is defined by

$$Y_t^n = y_0 + \sum_{i\,:\,t_{i+1}^n \leq t} b(t_i^n, Y_{t_i^n}^n)(t_{i+1}^n - t_i^n) + \sum_{i\,:\,t_{i+1}^n \leq t} \sigma(t_i^n, Y_{t_i^n}^n)(X_{t_{i+1}^n} - X_{t_i^n}), \tag{3.2}$$

for $t \in [0,T]$, along a sequence of partitions $\mathcal{P}^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$. Higher order Euler approximations, such as the Milstein scheme, introduce additional higher order correction terms in the approximation (3.2), which often involve iterated integrals of the driving signal X. In general, the numerical calculation of the approximation Y^n is carried out path by path, motivating a pathwise convergence analysis of the Euler scheme and its higher order variants. Indeed, it is well-known that, for SDEs driven by Brownian motion, the (higher-order) Euler approximations converge pathwise; see e.g. [23, 102, 84, 109].

A fully pathwise solution theory for SDEs like (3.1) is provided by the theory of rough paths; see e.g. [71, 74]. Loosely speaking, in our context, a rough path is a pair $\mathbf{X} = (X, \mathbb{X})$, consisting of a deterministic càdlàg $\mathbb{R}^d$-valued path X, and a two-parameter càdlàg $\mathbb{R}^{d\times d}$-valued function $\mathbb{X}$, which satisfy certain analytic and algebraic conditions. We will work with càdlàg rough paths with finite p-variation, in the regime with $p \in (2, 3)$, which includes in particular almost any sample path of a general semimartingale X, in which case the corresponding rough path $\mathbf{X} = (X, \mathbb{X})$, is given by $\mathbb{X}_{s,t} = \int_s^t (X_{r-} - X_s) \otimes \mathrm{d}X_r$ via stochastic integration.

Replacing the stochastic driving signal X in (3.1) by a (deterministic) rough path $\mathbf{X} = (X, \mathbb{X})$, we obtain a so-called rough differential equation (RDE). Assuming sufficient regularity of the coefficients b, σ, the RDE (3.1) driven by a given càdlàg rough path $\mathbf{X} = (X, \mathbb{X})$ is well-posed, in the sense that $\int_0^t \sigma(s, Y_s)\,\mathrm{d}\mathbf{X}_s$ is defined as a rough integral, and the equation admits a unique solution $Y = (Y_t)_{t\in[0,T]}$; see [75]. Moreover, if the rough path is the, say, Itô lift of a semimartingale X, then the solution of the resulting random RDE is consistent with the solution of the corresponding SDE driven by X. Both interpretations of the equation are thus essentially equivalent. Furthermore, in contrast to classical SDE theory, rough path theory is not limited to the semimartingale setting, and it comes with powerful pathwise stability estimates.

Rough path theory is intrinsically linked to the numerical approximation of stochastic differential equations, and provides a transparent explanation for the pathwise convergence of higher order Euler approximations and their modifications; see e.g. [70, 74, 54, 75, 126]. More precisely, the existence of a rough path lift of the driving signal is a sufficient condition for the pathwise convergence of higher order Euler schemes for RDEs, thus implying pathwise convergence for the corresponding SDEs driven by, e.g., semimartingales. However, the pathwise convergence of the first order Euler scheme—the most prominent numerical scheme for differential equations—cannot be explained by the rough path lift of the driving signal. Moreover, in general, an Euler approximation cannot converge to the solution of an RDE driven by an arbitrary rough path, for at least two reasons: First, the Euler approximation for an SDE driven by a fractional Brownian motion with Hurst index $H < \frac{1}{2}$ fails to converge (see e.g. [54]), and second, while the rough path lift $\mathbf{X} = (X, \mathbb{X})$ of a path X is not unique, leading to potentially multiple solutions of the RDE, the Euler approximation

Y^n defined in (3.2) is independent of the choice of rough path, and can thus only converge to at most one such solution.

In the present chapter we clarify the gap between rough and stochastic differential equations from the perspective of numerical approximation, by establishing the convergence of the first order Euler scheme for RDEs driven by Itô-type rough path lifts. More precisely, in Theorem 3.1.2 we obtain convergence in p-variation of the Euler scheme for rough differential equations driven by càdlàg paths satisfying a suitable criterion—namely the so-called Property (RIE)—relative to a sequence of partitions with vanishing mesh size.

Property (RIE) was first introduced in [143] and [7], motivated by applications in mathematical finance under model uncertainty. While, strictly speaking, it is a condition on a càdlàg path $X\colon [0,T] \to \mathbb{R}^d$, it always ensures the existence of an Itô-type rough path lift $\mathbf{X} = (X, \mathbb{X})$, allowing one to treat (3.1) as an RDE. Using this fact, we will show that Property (RIE) is a sufficient condition on the sample paths of a stochastic driving signal to guarantee the convergence of the first order Euler scheme for the corresponding SDE. We note in particular that the Euler scheme converges surely on the set where the stochastic driving signal satisfies Property (RIE), which is a stronger statement compared to the earlier results in [23, 102, 84, 109], in which the set on which the Euler scheme converges can depend on the coefficients b, σ. A criterion for Hölder continuous rough paths, related to Property (RIE), was previously introduced by Davie [51], which also allows one to obtain convergence of the Euler scheme for RDEs, and will be discussed in more detail in Remark 3.1.3.

Exploiting the continuity results of rough path theory, in Theorem 3.1.2 we derive a precise error estimate in p-variation for the Euler approximation of RDEs with respect to the discretization error of the driving signal. The convergence rate is expressed transparently, in terms of the mesh size of the approximating partition, and the approximation error of the discretized signal and of its rough path lift. We also obtain an error estimate for the Euler approximation with respect to pathwise perturbations of the driving signal; see Proposition 3.1.11. This latter perturbation is motivated by so-called approximate Euler schemes for SDEs driven by jump processes, see e.g. [95, 150, 53]. For instance, approximate Euler schemes are used for Lévy-driven SDEs, since the increments of Lévy processes cannot always be simulated, and thus the increments of the driving Lévy process need to be approximated by random variables with known distributions.

To obtain pathwise convergence of the Euler scheme in p-variation for a stochastic differential equation, it is then sufficient to verify that the associated stochastic driving signal of the equation satisfies Property (RIE), almost surely, relative to a sequence of partitions; see Sections 3.2 and 3.3. Unsurprisingly, we find that the more regular the driving signal is, the more general the sequence of partitions may be chosen. Indeed, while the sample paths of a Brownian motion satisfy Property (RIE), almost surely, relative to sequences of partitions whose mesh size can converge to zero very slowly, the sample paths of more general Itô processes satisfy Property (RIE), almost surely, relative to sequences of partitions whose mesh size is of order 2^{-n}. For stochastic processes with jumps, such as Lévy processes or general càdlàg semimartingales, one needs to ensure that the jump times are exhausted by the sequence of partitions, which is a necessary condition, for both the Euler scheme to converge pathwise, and for Property (RIE) to be satisfied by the driving signal.

The presented pathwise analysis of the first order Euler approximation is not limited to stochastic differential equations in a semimartingale setting. As examples, we consider mixed SDEs driven by both Brownian motion and fractional Brownian motion with Hurst index $H > \frac{1}{2}$, as in e.g. [162, 135], as well as rough stochastic differential equations, which are differential equations driven by both a rough path and a Brownian motion; see [72]. The latter equations are of interest, e.g., in the context of robust stochastic filtering; see [43, 58].

This chapter is structured as follows. In Section 3.1 we prove the convergence of the Euler scheme for rough differential equations assuming that the driving paths satisfy Property (RIE). In Sections 3.2 and 3.3 we provide various examples of stochastic processes which satisfy Property (RIE) along suitable sequences of partitions, making the established convergence analysis applicable to the corresponding SDEs, and derive associated convergence rates.

3.1 The Euler scheme for rough differential equations

In this section we study convergence of the (first order) Euler scheme for rough differential equations, which does not rely on the Lévy area of the path, and is known to converge pathwise for certain classes of stochastic differential equations. Before treating the Euler scheme, we will first recall some essentials from the theory of càdlàg rough paths, as introduced in [73, 75].

3.1.1 Essentials on rough path theory

A *partition* $\mathcal{P}$ of an interval $[s,t]$ is a finite set of points between and including the points s and t, i.e., $\mathcal{P} = \{s = u_0 < u_1 < \cdots < u_N = t\}$ for some $N \in \mathbb{N}$, and its mesh size is denoted by $|\mathcal{P}| := \max\{|u_{i+1} - u_i| : i = 0, \ldots, N-1\}$. A sequence $(\mathcal{P}^n)_{n\in\mathbb{N}}$ of partitions is said to be *nested*, if $\mathcal{P}^n \subset \mathcal{P}^{n+1}$ for all $n \in \mathbb{N}$.

Throughout, we let $T > 0$ be a fixed finite time horizon. We let $\Delta_T := \{(s,t) \in [0,T]^2 : s \le t\}$ denote the standard 2-simplex. A function $w\colon \Delta_T \to [0,\infty)$ is called a *control function* if it is superadditive, in the sense that $w(s,u) + w(u,t) \le w(s,t)$ for all $0 \le s \le u \le t \le T$. For two vectors $x = (x^1,\dots,x^d)^\top, y = (y^1,\dots,y^d)^\top \in \mathbb{R}^d$ we use the usual tensor product

$$x \otimes y := (x^i y^j)_{i,j=1,\dots,d} \in \mathbb{R}^{d\times d}.$$

Whenever $(B, \|\cdot\|)$ is a normed space and $f, g\colon B \to \mathbb{R}$ are two functions on B, we shall write $f \lesssim g$ or $f \le Cg$ to mean that there exists a constant $C > 0$ such that $f(x) \le Cg(x)$ for all $x \in B$. The constant C may depend on the normed space, e.g. through its dimension or regularity parameters, and, if we want to emphasize the dependence of the constant C on some particular variables, $\alpha_1,\dots,\alpha_n$ say, then we will write $C = C_{\alpha_1,\dots,\alpha_n}$.

For two vector spaces, the space of linear maps from $E_1 \to E_2$ is denoted by $\mathcal{L}(E_1,E_2)$; and we write $C_b^l = C_b^l(\mathbb{R}^m;\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$ for the space of l-times differentiable (in the Fréchet sense) functions $f\colon \mathbb{R}^m \to \mathcal{L}(\mathbb{R}^d;\mathbb{R}^k)$ such that f and all its derivatives up to order l are continuous and bounded. We equip this space with the norm

$$\|f\|_{C^l} := \|f\|_\infty + \|\mathrm{D}f\|_\infty + \cdots + \|\mathrm{D}^l f\|_\infty,$$

where $\mathrm{D}^n f$ denotes the n-th order derivative of f, and $\|\cdot\|_\infty$ denotes the supremum norm on the corresponding spaces of operators.

For a normed space $(E, |\cdot|)$, we let $D([0,T];E)$ denote the set of càdlàg (right-continuous with left-limits) paths from $[0,T]$ to E. For $X \in D([0,T];E)$, the supremum seminorm of the path X is given by

$$\|X\|_\infty := \sup_{t\in[0,T]} |X_t|,$$

and for $p \ge 1$, the p-variation of the path X is given by

$$\|X\|_p := \|X\|_{p,[0,T]} \qquad \text{with} \qquad \|X\|_{p,[s,t]} := \Bigg(\sup_{\mathcal{P}\subset[s,t]} \sum_{[u,v]\in\mathcal{P}} |X_v - X_u|^p\Bigg)^{\frac{1}{p}}, \quad (s,t) \in \Delta_T,$$

where the supremum is taken over all possible partitions $\mathcal{P}$ of the interval $[s,t]$. We recall that, given a path X, we have that $\|X\|_p < \infty$ if and only if there exists a control function w such that[1]

$$\sup_{(u,v)\in\Delta_T} \frac{|X_v - X_u|^p}{w(u,v)} < \infty.$$

We write $D^p = D^p([0,T];E)$ for the space of paths $X \in D([0,T];E)$ which satisfy $\|X\|_p < \infty$. Moreover, for a path $X \in D([0,T];\mathbb{R}^d)$, we will often use the shorthand notation:

$$X_{s,t} := X_t - X_s \quad \text{and} \quad X_{t-} := \lim_{u\uparrow t} X_u, \qquad \text{for} \quad (s,t) \in \Delta_T.$$

[1] Here and throughout, we adopt the convention that $\frac{0}{0} := 0$.

For $r \geq 1$ and a two-parameter function $\mathbb{X}\colon \Delta_T \to E$, we similarly define

$$\|\mathbb{X}\|_r := \|\mathbb{X}\|_{r,[0,T]} \qquad \text{with} \qquad \|\mathbb{X}\|_{r,[s,t]} := \bigg(\sup_{\mathcal{P}\subset[s,t]} \sum_{[u,v]\in\mathcal{P}} |\mathbb{X}_{u,v}|^r\bigg)^{\frac{1}{r}}, \quad (s,t) \in \Delta_T.$$

We write $D_2^r = D_2^r(\Delta_T; E)$ for the space of all functions $\mathbb{X}\colon \Delta_T \to E$ which satisfy $\|\mathbb{X}\|_r < \infty$, and are such that the maps $s \mapsto \mathbb{X}_{s,t}$ for fixed t, and $t \mapsto \mathbb{X}_{s,t}$ for fixed s, are both càdlàg.

For $p \in [2,3)$, a pair $\mathbf{X} = (X, \mathbb{X})$ is called a *càdlàg p-rough path* over $\mathbb{R}^d$ if

(i) $X \in D^p([0,T]; \mathbb{R}^d)$ and $\mathbb{X} \in D_2^{\frac{p}{2}}(\Delta_T; \mathbb{R}^{d\times d})$, and

(ii) Chen's relation: $\mathbb{X}_{s,t} = \mathbb{X}_{s,u} + \mathbb{X}_{u,t} + X_{s,u} \otimes X_{u,t}$ holds for all $0 \leq s \leq u \leq t \leq T$.

In component form, condition (ii) states that $\mathbb{X}^{ij}_{s,t} = \mathbb{X}^{ij}_{s,u} + \mathbb{X}^{ij}_{u,t} + X^i_{s,u} X^j_{u,t}$ for every i and j. We will denote the space of càdlàg p-rough paths by $\mathcal{D}^p = \mathcal{D}^p([0,T]; \mathbb{R}^d)$. On the space $\mathcal{D}^p([0,T]; \mathbb{R}^d)$, we use the natural seminorm

$$\|\mathbf{X}\|_p := \|\mathbf{X}\|_{p,[0,T]} \qquad \text{with} \qquad \|\mathbf{X}\|_{p,[s,t]} := \|X\|_{p,[s,t]} + \|\mathbb{X}\|_{\frac{p}{2},[s,t]}$$

for $(s,t) \in \Delta_T$, and the induced distance

$$\|\mathbf{X}; \widetilde{\mathbf{X}}\|_p := \|\mathbf{X}; \widetilde{\mathbf{X}}\|_{p,[0,T]} \qquad \text{with} \qquad \|\mathbf{X}; \widetilde{\mathbf{X}}\|_{p,[s,t]} := \|X - \widetilde{X}\|_{p,[s,t]} + \|\mathbb{X} - \widetilde{\mathbb{X}}\|_{\frac{p}{2},[s,t]}, \tag{3.3}$$

whenever $\mathbf{X} = (X, \mathbb{X}), \widetilde{\mathbf{X}} = (\widetilde{X}, \widetilde{\mathbb{X}}) \in \mathcal{D}^p([0,T]; \mathbb{R}^d)$.

Let $p \in [2,3)$, $q \in [p, \infty)$ and $r \in [\frac{p}{2}, 2)$ such that $\frac{1}{p} + \frac{1}{r} > 1$ and $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. Let $X \in D^p([0,T]; \mathbb{R}^d)$. We say that a pair (Y, Y') is a *controlled path* (with respect to X), if

$$Y \in D^p([0,T]; E), \quad Y' \in D^q([0,T]; \mathcal{L}(\mathbb{R}^d; E)), \quad \text{and} \quad R^Y \in D_2^r(\Delta_T; E),$$

where R^Y is defined by

$$Y_{s,t} = Y'_s X_{s,t} + R^Y_{s,t} \qquad \text{for all} \quad (s,t) \in \Delta_T.$$

We write $\mathcal{V}^{q,r}_X = \mathcal{V}^{q,r}_X([0,T]; E)$ for the space of E-valued controlled paths, which becomes a Banach space when equipped with the norm

$$(Y, Y') \mapsto |Y_0| + |Y'_0| + \|Y'\|_{q,[0,T]} + \|R^Y\|_{r,[0,T]}.$$

For paths $A \in D^{q_1}$, $H \in D^{q_2}$ for $q_1, q_2 \in [1,2)$, and a rough path $\mathbf{X} \in \mathcal{D}^p$ for $p \in [2,3)$, we consider the rough differential equation (RDE):

$$Y_t = y_0 + \int_0^t b(H_s, Y_s) \,\mathrm{d}A_s + \int_0^t \sigma(H_s, Y_s) \,\mathrm{d}\mathbf{X}_s, \qquad t \in [0,T]. \tag{3.4}$$

Provided that $\frac{1}{p} + \frac{1}{q_1} > 1$ and $\frac{1}{p} + \frac{1}{q_2} > 1$, the first integral in this equation can be defined as a Young integral, whilst the second integral is defined as a rough integral. For precise definitions, constructions and properties of these integrals, we refer to the comprehensive exposition in [75].

Theorem 3.1.1. *Let $p \in [2,3)$ and $q_1, q_2 \in [1,2)$ such that $\frac{1}{p} + \frac{1}{q_1} > 1$ and $\frac{1}{p} + \frac{1}{q_2} > 1$. Let $b \in C_b^2$, $\sigma \in C_b^3$, $y_0 \in \mathbb{R}^k$, $A \in D^{q_1}$, $H \in D^{q_2}$ and $\mathbf{X} = (X, \mathbb{X}) \in \mathcal{D}^p$. Let $r \in [\frac{p}{2} \vee q_1 \vee q_2, 2)$ such that $\frac{1}{p} + \frac{1}{r} > 1$, and let $q \in [p, \infty)$ such that $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. Then there exists a unique path $Y \in D^p$ such that the controlled path $(Y, \sigma(H, Y)) \in \mathcal{V}_X^{q,r}$ satisfies the RDE (3.4).*

Moreover, if $\widetilde{y}_0 \in \mathbb{R}^k$, $\widetilde{A} \in D^{q_1}$, $\widetilde{H} \in D^{q_2}$ and $\widetilde{\mathbf{X}} = (\widetilde{X}, \widetilde{\mathbb{X}}) \in \mathcal{D}^p$ with corresponding solution $\widetilde{Y}$, and if $\|A\|_r, \|\widetilde{A}\|_r, \|H\|_r, \|\widetilde{H}\|_r, \|\mathbf{X}\|_p, \|\widetilde{\mathbf{X}}\|_p \leq L$ for some $L > 0$, then

$$\begin{aligned} &\|Y - \widetilde{Y}\|_p + \|Y' - \widetilde{Y}'\|_q + \|R^Y - R^{\widetilde{Y}}\|_r \\ &\quad \lesssim |y_0 - \widetilde{y}_0| + |H_0 - \widetilde{H}_0| + \|H - \widetilde{H}\|_r + \|A - \widetilde{A}\|_r + \|\mathbf{X}; \widetilde{\mathbf{X}}\|_p, \end{aligned} \tag{3.5}$$

where the implicit multiplicative constant depends only on $p, q, r, \|b\|_{C_b^2}, \|\sigma\|_{C_b^3}$ and L.

The result of Theorem 3.1.1 may be considered classical, and will be unsurprising to readers familiar with RDEs. However, to the best of our knowledge, a proof of the precise statement of the theorem does not appear in the existing literature. A sketch of the proof, based on the proof of [3, Theorem 2.3], is therefore given in Appendix A.2.

3.1.2 Convergence of the Euler scheme

Let us consider the rough differential equation

$$Y_t = y_0 + \int_0^t b(s, Y_s) \, \mathrm{d}s + \int_0^t \sigma(s, Y_s) \, \mathrm{d}\mathbf{X}_s, \qquad t \in [0, T], \tag{3.6}$$

where $y_0 \in \mathbb{R}^k$, $b \in C_b^2(\mathbb{R}^{k+1}; \mathcal{L}(\mathbb{R}; \mathbb{R}^k))$, $\sigma \in C_b^3(\mathbb{R}^{k+1}; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k))$ and $\mathbf{X} = (X, \mathbb{X}) \in \mathcal{D}^p([0,T]; \mathbb{R}^d)$ is the driving càdlàg p-rough path for $p \in [2,3)$. Given a sequence of partitions $\mathcal{P}^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$, the Euler approximation Y^n corresponding to the RDE (3.6) along the partition $\mathcal{P}^n$ is given by

$$Y_t^n = y_0 + \sum_{i \colon t_{i+1}^n \leq t} b(t_i^n, Y_{t_i^n}^n)(t_{i+1}^n - t_i^n) + \sum_{i \colon t_{i+1}^n \leq t} \sigma(t_i^n, Y_{t_i^n}^n)(X_{t_{i+1}^n} - X_{t_i^n}), \tag{3.7}$$

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

It is a classical result in the numerical analysis of stochastic differential equations that, if the driving signal is, e.g., a Brownian motion, then the Euler scheme (often also called the Euler–Maruyama scheme) converges pathwise; see e.g. [109]. On the other hand, it is known that in general the Euler scheme cannot converge if the driving signal is an arbitrary rough path, since the corresponding Euler scheme for stochastic differential equations driven by fractional Brownian motion fails to converge; see [54] for a more detailed discussion on this observation.

Moreover, since the extension of a path X to a rough path $\mathbf{X} = (X, \mathbb{X})$ is not unique, and the Euler approximation Y^n defined in (3.7) is independent of $\mathbb{X}$, the sequence $(Y^n)_{n \in \mathbb{N}}$

cannot converge to the solution of a general rough differential equation. Thus, in order to ensure the convergence of the Euler scheme, it is necessary to identify the "correct" rough path lift $\mathbf{X}$ as the driving signal for the RDE (3.6). A suitable resolution to this is provided by the so-called Property (RIE), as introduced in [143] and [7].

Property (RIE). *Let $p \in (2,3)$ and let $\mathcal{P}^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$, be a sequence of partitions of the interval $[0,T]$ such that $|\mathcal{P}^n| \to 0$ as $n \to \infty$. For $X \in D([0,T];\mathbb{R}^d)$, and each $n \in \mathbb{N}$, we define $X^n\colon [0,T] \to \mathbb{R}^d$ by*

$$X_t^n = X_T \mathbf{1}_{\{T\}}(t) + \sum_{k=0}^{N_n-1} X_{t_k^n} \mathbf{1}_{[t_k^n, t_{k+1}^n)}(t), \qquad t \in [0,T].$$

We assume that:

(i) the sequence of paths $(X^n)_{n\in\mathbb{N}}$ converges uniformly to X as $n \to \infty$,

(ii) the Riemann sums $\int_0^t X_u^n \otimes \mathrm{d}X_u := \sum_{k=0}^{N_n-1} X_{t_k^n} \otimes X_{t_k^n \wedge t, t_{k+1}^n \wedge t}$ converge uniformly as $n \to \infty$ to a limit, which we denote by $\int_0^t X_u \otimes \mathrm{d}X_u$, $t \in [0,T]$,

(iii) and there exists a control function w such that

$$\sup_{(s,t)\in\Delta_T} \frac{|X_{s,t}|^p}{w(s,t)} + \sup_{n\in\mathbb{N}} \sup_{0\le k<\ell\le N_n} \frac{|\int_{t_k^n}^{t_\ell^n} X_u^n \otimes \mathrm{d}X_u - X_{t_k^n} \otimes X_{t_k^n, t_\ell^n}|^{\frac{p}{2}}}{w(t_k^n, t_\ell^n)} \le 1. \tag{3.8}$$

We say that a path $X \in D([0,T];\mathbb{R}^d)$ satisfies Property (RIE) relative to p and $(\mathcal{P}^n)_{n\in\mathbb{N}}$, if p, $(\mathcal{P}^n)_{n\in\mathbb{N}}$ and X together satisfy Property (RIE).

It is known that, if a path $X \in D([0,T];\mathbb{R}^d)$ satisfies Property (RIE), then X extends canonically to a rough path $\mathbf{X} = (X, \mathbb{X}) \in \mathcal{D}^p([0,T];\mathbb{R}^d)$, where the lift $\mathbb{X}$ is defined by

$$\mathbb{X}_{s,t} := \int_s^t X_u \otimes \mathrm{d}X_u - X_s \otimes (X_t - X_s), \qquad (s,t) \in \Delta_T, \tag{3.9}$$

with $\int_s^t X_u \otimes \mathrm{d}X_u := \int_0^t X_u \otimes \mathrm{d}X_u - \int_0^s X_u \otimes \mathrm{d}X_u$, and the existence of the integral $\int_0^t X_u \otimes \mathrm{d}X_u$ is ensured by condition (ii) of Property (RIE); see [7, Lemma 2.13]. When assuming Property (RIE) for a path X, we will always work with the rough path $\mathbf{X} = (X, \mathbb{X})$ defined via (3.9), and note that $\mathbf{X} = (X, \mathbb{X})$ corresponds to the Itô rough path lift of a stochastic process, since the "iterated integral" $\mathbb{X}$ is given as a limit of left-point Riemann sums, analogously to the stochastic Itô integral.

Postulating Property (RIE) for the driving signal of a rough differential equation ensures that the (first order) Euler approximation converges to the solution of the equation, as stated precisely in the next theorem.

Theorem 3.1.2. *Suppose that $X\colon[0,T]\to\mathbb{R}^d$ satisfies Property* (RIE) *relative to some $p\in(2,3)$ and a sequence of partitions $(\mathcal{P}^n)_{n\in\mathbb{N}}$, and let $\mathbf{X}$ be the canonical rough path lift of X, as defined in* (3.9). *Let Y be the solution of the RDE* (3.6) *driven by $\mathbf{X}$, and let Y^n be the Euler approximation defined in* (3.7). *Then,*

$$\|Y^n-Y\|_{p'}\longrightarrow 0 \qquad \text{as} \quad n\longrightarrow\infty,$$

for any $p'\in(p,3)$, and the rate of convergence is determined by the estimate

$$\|Y^n-Y\|_{p'}\lesssim|\mathcal{P}^n|^{1-\frac{1}{q}}+\|X^n-X\|_\infty^{1-\frac{p}{p'}}+\bigg\|\int_0^\cdot X_u^n\otimes\mathrm{d}X_u-\int_0^\cdot X_u\otimes\mathrm{d}X_u\bigg\|_\infty^{1-\frac{p}{p'}},\tag{3.10}$$

which holds for any $q\in(1,2)$ such that $\frac{1}{p'}+\frac{1}{q}>1$, where the implicit multiplicative constant depends only on $p,p',q,\|b\|_{C_b^2},\|\sigma\|_{C_b^3},T,|X_0|$ and $w(0,T)$, where w is the control function for which (3.8) *holds.*

Note that Property (RIE) implies that each of the terms on the right-hand side of (3.10) tends to zero as $n\to\infty$.

Remark 3.1.3. *In [51], A. M. Davie observed that, under suitable conditions, the first order Euler scheme along equidistant partitions converges to the solution of a given rough differential equation. More precisely, for $p\in(2,3)$ and $\alpha:=\frac{1}{p}$, let $\mathbf{X}=(X,\mathbb{X})$ be an α-Hölder continuous rough path, so that $|X_{s,t}|\lesssim|t-s|^\alpha$ and $|\mathbb{X}_{s,t}|\lesssim|t-s|^{2\alpha}$ for $(s,t)\in\Delta_T$, such that, for some $\beta\in(1-\alpha,2\alpha)$, there exists a constant $C>0$ such that*

$$\bigg|\sum_{j=k}^{\ell-1}\mathbb{X}_{jh,(j+1)h}\bigg|\le C(\ell-k)^\beta h^{2\alpha}$$

whenever $h>0$ and $0\le k<\ell$ are integers such that $\ell h\le T$. Under this condition on the driving signal $\mathbf{X}$, [51, Theorem 7.1] states that the Euler approximations Y^n, as defined in (3.7)*, converge uniformly to the solution Y of the RDE* (3.6) *along the equidistant partitions $(\mathcal{P}_U^n)_{n\in\mathbb{N}}$, where $\mathcal{P}_U^n=\{\frac{iT}{n}:i=0,1,\dots,n\}$. Note that Davie's condition implies Property* (RIE)*—see [143, Appendix B]—and is thus less general, even in the case of Hölder continuous rough paths.*

The rest of this subsection is devoted to the proof of Theorem 3.1.2, which first requires us to establish some auxiliary results.

In the following, we will always assume that $X\colon[0,T]\to\mathbb{R}^d$ satisfies Property (RIE) relative to some $p\in(2,3)$ and a sequence of partitions $(\mathcal{P}^n)_{n\in\mathbb{N}}$. As the piecewise constant approximation X^n (as defined in Property (RIE)) has finite 1-variation, it possesses a canonical rough path lift $\mathbf{X}^n=(X^n,\mathbb{X}^n)\in\mathcal{D}^p([0,T];\mathbb{R}^d)$, with $\mathbb{X}^n$ given by

$$\mathbb{X}_{s,t}^n:=\int_s^t X_{s,u}^n\otimes\mathrm{d}X_u^n,\qquad(s,t)\in\Delta_T,\tag{3.11}$$

where the integral is defined as a classical limit of left-point Riemann sums. Note that, while [75, Section 5.3] discretizes the rough path $\mathbf{X} = (X, \mathbb{X})$ in a piecewise constant manner, here we instead discretize the path X and then extend it to a rough path $\mathbf{X}^n = (X^n, \mathbb{X}^n)$ via (3.11).

As a first step towards the proof of Theorem 3.1.2, we establish the convergence of the rough paths $(\mathbf{X}^n)_{n\in\mathbb{N}}$ to the rough path $\mathbf{X}$ in a suitable rough path distance. For this purpose, we need two auxiliary lemmas.

Lemma 3.1.4. *Suppose that $X\colon [0,T] \to \mathbb{R}^d$ satisfies Property* (RIE) *relative to some $p \in (2,3)$ and a sequence of partitions $(\mathcal{P}^n)_{n\in\mathbb{N}}$. Then, we have the estimate*

$$\sup_{(s,t)\in\Delta_T} |\mathbb{X}^n_{s,t} - \mathbb{X}_{s,t}| \leq 2\|X\|_\infty \|X^n - X\|_\infty + \sup_{(s,t)\in\Delta_T} \left| \int_s^t X^n_{s,u} \otimes \mathrm{d}X_u - \mathbb{X}_{s,t} \right|,$$

where $\mathbb{X}^n$ and $\mathbb{X}$ were defined in (3.11) *and* (3.9)*, respectively. In particular, we have that*

$$\mathbb{X}^n \longrightarrow \mathbb{X} \quad \textit{uniformly as} \quad n \longrightarrow \infty.$$

Proof. Since

$$|\mathbb{X}^n_{s,t} - \mathbb{X}_{s,t}| \leq \left| \mathbb{X}^n_{s,t} - \int_s^t X^n_{s,u} \otimes \mathrm{d}X_u \right| + \left| \int_s^t X^n_{s,u} \otimes \mathrm{d}X_u - \mathbb{X}_{s,t} \right|,$$

and the limit in condition (ii) of Property (RIE) holds uniformly, it is enough to prove that the function given by

$$\Lambda^n_{s,t} := \mathbb{X}^n_{s,t} - \int_s^t X^n_{s,u} \otimes \mathrm{d}X_u = \int_s^t X^n_{s,u} \otimes \mathrm{d}(X^n - X)_u$$

satisfies

$$\sup_{(s,t)\in\Delta_T} |\Lambda^n_{s,t}| \leq 2\|X\|_\infty \|X^n - X\|_\infty. \tag{3.12}$$

If $t^n_k \leq s < t \leq t^n_{k+1}$ for some k, then $X^n_{s,u} = X_{t^n_k, t^n_k} = 0$ for every $u \in [s,t)$, so that $\Lambda^n_{s,t} = 0$. Otherwise, let k_0 be the smallest k such that $t^n_k \in (s,t)$, and let k_1 be the largest such k. It is straightforward to see that the triplet $(X^n - X, X^n, \Lambda^n)$ satisfies Chen's relation:

$$\Lambda^n_{s,t} = \Lambda^n_{s,u} + \Lambda^n_{u,t} + X^n_{s,u} \otimes (X^n - X)_{u,t}$$

for all $s \leq u \leq t$, from which it follows that

$$\Lambda^n_{s,t} = \Lambda^n_{s,t^n_{k_0}} + \Lambda^n_{t^n_{k_0}, t^n_{k_1}} + \Lambda^n_{t^n_{k_1}, t} + X^n_{s,t^n_{k_0}} \otimes (X^n - X)_{t^n_{k_0}, t^n_{k_1}} + X^n_{s,t^n_{k_1}} \otimes (X^n - X)_{t^n_{k_1}, t}.$$

As we already observed, we have that $\Lambda^n_{s,t^n_{k_0}} = \Lambda^n_{t^n_{k_1},t} = 0$. In fact, we also have that

$$\begin{aligned}\Lambda^n_{t^n_{k_0},t^n_{k_1}} &= \int_{t^n_{k_0}}^{t^n_{k_1}} X^n_{t^n_{k_0},u} \otimes \mathrm{d}(X^n - X)_u = \sum_{i=k_0}^{k_1-1} \int_{t^n_i}^{t^n_{i+1}} X^n_{t^n_{k_0},u} \otimes \mathrm{d}(X^n - X)_u \\ &= \sum_{i=k_0}^{k_1-1} \int_{t^n_i}^{t^n_{i+1}} X_{t^n_{k_0},t^n_i} \otimes \mathrm{d}(X^n - X)_u = \sum_{i=k_0}^{k_1-1} X_{t^n_{k_0},t^n_i} \otimes (X^n - X)_{t^n_i,t^n_{i+1}} = 0.\end{aligned} \tag{3.13}$$

Since $(X^n - X)_{t^n_{k_0}} = (X^n - X)_{t^n_{k_1}} = 0$, we simply obtain $\Lambda^n_{s,t} = X^n_{s,t^n_{k_1}} \otimes (X^n_t - X_t)$, from which (3.12) follows. □

Lemma 3.1.5. *Suppose that $X\colon [0,T] \to \mathbb{R}^d$ satisfies Property* (RIE) *relative to some $p \in (2,3)$ and a sequence of partitions $(\mathcal{P}^n)_{n\in\mathbb{N}}$. Let w be the control function with respect to which X satisfies the inequality* (3.8)*. Then, there exists a constant C, which depends only on p, such that*

$$\|\mathbb{X}^n\|_{\frac{p}{2}} \leq C w(0,T)^{\frac{2}{p}} \tag{3.14}$$

for every $n \in \mathbb{N}$, where $\mathbb{X}^n$ was defined in (3.11)*.*

Proof. Let $n \in \mathbb{N}$, and let $(s,t) \in \Delta_T$. If $t^n_k \leq s < t \leq t^n_{k+1}$ for some k, then $X^n_{s,u} = X_{t^n_k,t^n_k} = 0$ for every $u \in [s,t)$, so that $\mathbb{X}^n_{s,t} = 0$. Otherwise, let k_0 be the smallest k such that $t^n_k \in (s,t)$, and let k_1 be the largest such k. It is straightforward to see that $(X^n, \mathbb{X}^n)$ satisfies Chen's relation:

$$\mathbb{X}^n_{s,t} = \mathbb{X}^n_{s,u} + \mathbb{X}^n_{u,t} + X^n_{s,u} \otimes X^n_{u,t}$$

for all $s \leq u \leq t$, from which it follows that

$$\mathbb{X}^n_{s,t} = \mathbb{X}^n_{s,t^n_{k_0}} + \mathbb{X}^n_{t^n_{k_0},t^n_{k_1}} + \mathbb{X}^n_{t^n_{k_1},t} + X^n_{s,t^n_{k_0}} \otimes X^n_{t^n_{k_0},t^n_{k_1}} + X^n_{s,t^n_{k_1}} \otimes X^n_{t^n_{k_1},t}.$$

As we have already seen, we have that $\mathbb{X}^n_{s,t^n_{k_0}} = \mathbb{X}^n_{t^n_{k_1},t} = 0$. Recalling the calculation in (3.13), we note that

$$\mathbb{X}^n_{t^n_{k_0},t^n_{k_1}} = \int_{t^n_{k_0}}^{t^n_{k_1}} X^n_{t^n_{k_0},u} \otimes \mathrm{d}X^n_u = \int_{t^n_{k_0}}^{t^n_{k_1}} X^n_{t^n_{k_0},u} \otimes \mathrm{d}X_u,$$

and hence, by the inequality in (3.8), that

$$|\mathbb{X}^n_{t^n_{k_0},t^n_{k_1}}|^{\frac{p}{2}} = \left| \int_{t^n_{k_0}}^{t^n_{k_1}} X^n_{t^n_{k_0},u} \otimes \mathrm{d}X_u \right|^{\frac{p}{2}} \leq w(t^n_{k_0}, t^n_{k_1}) \leq w(t^n_{k_0-1}, t^n_{k_1+1}).$$

We estimate the remaining terms as

$$\begin{aligned}&|X^n_{s,t^n_{k_0}} \otimes X^n_{t^n_{k_0},t^n_{k_1}}|^{\frac{p}{2}} + |X^n_{s,t^n_{k_1}} \otimes X^n_{t^n_{k_1},t}|^{\frac{p}{2}} \lesssim |X^n_{s,t^n_{k_0}}|^p + |X^n_{t^n_{k_0},t^n_{k_1}}|^p + |X^n_{s,t^n_{k_1}}|^p + |X^n_{t^n_{k_1},t}|^p \\ &\leq |X_{t^n_{k_0-1},t^n_{k_0}}|^p + |X_{t^n_{k_0},t^n_{k_1}}|^p + |X_{t^n_{k_0-1},t^n_{k_1}}|^p + |X_{t^n_{k_1},t^n_{k_1+1}}|^p \\ &\leq w(t^n_{k_0-1}, t^n_{k_0}) + w(t^n_{k_0}, t^n_{k_1}) + w(t^n_{k_0-1}, t^n_{k_1}) + w(t^n_{k_1}, t^n_{k_1+1}) \\ &\leq 2w(t^n_{k_0-1}, t^n_{k_1+1}).\end{aligned}$$

Putting this together, we have that

$$|\mathbb{X}^n_{s,t}|^{\frac{p}{2}} \leq \widetilde{C} w(t^n_{k_0-1}, t^n_{k_1+1})$$

for some constant $\widetilde{C}$. It follows that, for an arbitrary partition $\mathcal{P}$ of the interval $[0,T]$, we have the bound

$$\sum_{[s,t]\in\mathcal{P}} |\mathbb{X}^n_{s,t}|^{\frac{p}{2}} \leq 3\widetilde{C} w(0,T),$$

and hence that (3.14) holds with $C = (3\widetilde{C})^{\frac{2}{p}}$. □

Using the previous two lemmas, we can now infer the convergence of the rough paths $(\mathbf{X}^n)_{n\in\mathbb{N}}$ to the rough path $\mathbf{X}$.

Lemma 3.1.6. *Suppose that $X\colon[0,T]\to\mathbb{R}^d$ satisfies Property* (RIE) *relative to some $p\in(2,3)$ and a sequence of partitions $(\mathcal{P}^n)_{n\in\mathbb{N}}$. Let $\mathbf{X}=(X,\mathbb{X})$ and $\mathbf{X}^n=(X^n,\mathbb{X}^n)$ be the càdlàg rough paths defined via* (3.9) *and* (3.11)*, respectively. Then, for any $p'>p$, we have that*

$$\|\mathbf{X}^n;\mathbf{X}\|_{p'} \longrightarrow 0 \qquad \text{as} \quad n \longrightarrow \infty, \tag{3.15}$$

with a rate of convergence given by

$$\|\mathbf{X}^n;\mathbf{X}\|_{p'} \lesssim \|X^n - X\|_\infty^{1-\frac{p}{p'}} + \sup_{(s,t)\in\Delta_T} \left| \int_s^t X^n_{s,u}\otimes \mathrm{d}X_u - \mathbb{X}_{s,t}\right|^{1-\frac{p}{p'}}, \tag{3.16}$$

where the implicit multiplicative constant depends only on $p,p',|X_0|$ and $w(0,T)$, where w is the control function for which (3.8) *holds.*

Proof. By a standard interpolation estimate (e.g. [74, Proposition 5.5]), it follows, for any $p'>p$, that

$$\|X^n - X\|_{p'} \leq \|X^n - X\|_p^{\frac{p}{p'}} \|X^n - X\|_\infty^{1-\frac{p}{p'}}.$$

We similarly have that

$$\|\mathbb{X}^n - \mathbb{X}\|_{\frac{p'}{2}} \leq \|\mathbb{X}^n - \mathbb{X}\|_{\frac{p}{2}}^{\frac{p}{p'}} \sup_{(s,t)\in\Delta_T} |\mathbb{X}^n_{s,t} - \mathbb{X}_{s,t}|^{1-\frac{p}{p'}}.$$

We recall from Lemma 3.1.4 that

$$\sup_{(s,t)\in\Delta_T} |\mathbb{X}^n_{s,t} - \mathbb{X}_{s,t}| \leq 2\|X\|_\infty \|X^n - X\|_\infty + \sup_{(s,t)\in\Delta_T} \left|\int_s^t X^n_{s,u}\otimes\mathrm{d}X_u - \mathbb{X}_{s,t}\right|.$$

We have that $\sup_{n\in\mathbb{N}}\|X^n\|_p \leq \|X\|_p$ and $\|X\|_\infty \leq |X_0| + \|X\|_p \leq |X_0| + w(0,T)^{\frac{1}{p}}$, and, by the lower semi-continuity of the $\frac{p}{2}$-variation norm and Lemma 3.1.5, $\|\mathbb{X}\|_{\frac{p}{2}} \leq \liminf_{n\to\infty}\|\mathbb{X}^n\|_{\frac{p}{2}} \leq \sup_{n\in\mathbb{N}}\|\mathbb{X}^n\|_{\frac{p}{2}} \leq Cw(0,T)^{\frac{2}{p}}$. Putting this together, we conclude that (3.16) holds. By conditions (i) and (ii) in Property (RIE), the convergence in (3.15) then also follows. □

As a next step towards the proof of Theorem 3.1.2, we introduce a discretized version of the RDE (3.6). For this purpose, we define a time discretization path along $\mathcal{P}^n$ by

$$\gamma_t^n := T\mathbf{1}_{\{T\}}(t) + \sum_{k=0}^{N_n-1} t_k^n \mathbf{1}_{[t_k^n, t_{k+1}^n)}(t), \qquad t \in [0,T], \tag{3.17}$$

and consider the rough differential equation

$$\widetilde{Y}_t^n = y_0 + \int_0^t b(\gamma_s^n, \widetilde{Y}_s^n) \,\mathrm{d}\gamma_s^n + \int_0^t \sigma(\gamma_s^n, \widetilde{Y}_s^n) \,\mathrm{d}\mathbf{X}_s^n, \qquad t \in [0,T]. \tag{3.18}$$

Thanks to Lemma 3.1.6 and the local Lipschitz continuity of the Itô–Lyons map, we obtain the following proposition.

Proposition 3.1.7. *Suppose that $X\colon [0,T] \to \mathbb{R}^d$ satisfies Property* (RIE) *relative to some $p \in (2,3)$ and a sequence of partitions $(\mathcal{P}^n)_{n\in\mathbb{N}}$. Let Y be the solution of the RDE* (3.6)*, and let $\widetilde{Y}^n$ be the solution of the RDE* (3.18)*. Then,*

$$\|\widetilde{Y}^n - Y\|_{p'} \longrightarrow 0 \qquad \text{as} \quad n \longrightarrow \infty, \tag{3.19}$$

for any $p' \in (p,3)$, with a rate of convergence given by

$$\|\widetilde{Y}^n - Y\|_{p'} \lesssim |\mathcal{P}^n|^{1-\frac{1}{q}} + \|X^n - X\|_\infty^{1-\frac{p}{p'}} + \left\| \int_0^\cdot X_u^n \otimes \mathrm{d}X_u - \int_0^\cdot X_u \otimes \mathrm{d}X_u \right\|_\infty^{1-\frac{p}{p'}},$$

for any $q \in (1,2)$ such that $\frac{1}{p'} + \frac{1}{q} > 1$, where the implicit multiplicative constant depends only on $p, p', q, \|b\|_{C_b^2}, \|\sigma\|_{C_b^3}, T, |X_0|$ and $w(0,T)$, where w is the control function for which (3.8) *holds.*

Proof. Setting $\gamma_t := t$ for $t \in [0,T]$, the RDE (3.6) may be rewritten as

$$Y_t = y_0 + \int_0^t b(\gamma_s, Y_s) \,\mathrm{d}\gamma_s + \int_0^t \sigma(\gamma_s, Y_s) \,\mathrm{d}\mathbf{X}_s, \qquad t \in [0,T].$$

Hence, by Theorem 3.1.1, we know that

$$\|\widetilde{Y}^n - Y\|_{p'} \lesssim \|\gamma^n - \gamma\|_q + \|\mathbf{X}^n; \mathbf{X}\|_{p'} \tag{3.20}$$

for any $p' \in (p,3)$ and any $q \in [1,2)$ such that $\frac{1}{p'} + \frac{1}{q} > 1$.

Note that γ^n and γ have finite 1-variation, with $\|\gamma^n\|_1 = \|\gamma\|_1 = T$, and $\|\gamma^n - \gamma\|_1 = 2T$. Although γ^n does not converge to γ in 1-variation, it is straightforward to see by interpolation that

$$\|\gamma^n - \gamma\|_q \leq \|\gamma^n - \gamma\|_1^{\frac{1}{q}} \|\gamma^n - \gamma\|_\infty^{1-\frac{1}{q}} = (2T)^{\frac{1}{q}} |\mathcal{P}^n|^{1-\frac{1}{q}}$$

for any $q > 1$. Combining this with the estimate in (3.20) and the result of Lemma 3.1.6, we infer the convergence in (3.19), and the estimate

$$\begin{aligned}\|\widetilde{Y}^n - Y\|_{p'} &\lesssim \|\gamma^n - \gamma\|_q + \|X^n - X\|_\infty^{1-\frac{p}{p'}} + \sup_{(s,t)\in\Delta_T} \left| \int_s^t X^n_{s,u} \otimes \mathrm{d}X_u - \mathbb{X}_{s,t} \right|^{1-\frac{p}{p'}} \\ &\lesssim |\mathcal{P}^n|^{1-\frac{1}{q}} + \|X^n - X\|_\infty^{1-\frac{p}{p'}} + \left\| \int_0^\cdot X^n_u \otimes \mathrm{d}X_u - \int_0^\cdot X_u \otimes \mathrm{d}X_u \right\|_\infty^{1-\frac{p}{p'}}.\end{aligned}$$

□

Remark 3.1.8. *For a path $A \in D^1([0,T];\mathbb{R}^d)$ of finite 1-variation, let us consider the controlled ordinary differential equation (ODE)*

$$Z_t = z_0 + \int_0^t \sigma(Z_s)\,\mathrm{d}A_s, \qquad t \in [0,T], \tag{3.21}$$

where the integral is interpreted in the Riemann–Stieltjes sense. It is a classical result that, provided σ is sufficiently regular, the ODE in (3.21) is well-posed, and that the solution map $\Phi\colon A \mapsto Z$ is continuous with respect to the 1-variation norm $\|\cdot\|_1$. A major insight of the theory of rough paths is that the solution map Φ can be extended from the space of smooth paths to the space $\mathscr{C}^{0,p\text{-var}}([0,T];\mathbb{R}^d)$ of continuous geometric rough paths for $p \in (2,3)$; see e.g. [74]. Of course, the closure of a set containing only continuous paths with respect to p-variation norms will again only contain continuous paths.

In the current framework of càdlàg rough paths, Lemma 3.1.6 and Proposition 3.1.7 motivate us to consider instead the closure of càdlàg paths of finite 1-variation. For $p \in (2,3)$, let $\mathcal{D}^{0,p}([0,T];\mathbb{R}^d)$ denote the closure of the set

$$\left\{ \mathbf{A} = (A, \mathbb{A}) \,:\, A \in D^1([0,T];\mathbb{R}^d) \;\text{ and }\; \mathbb{A}_{s,t} := \int_s^t A_{s,u} \otimes \mathrm{d}A_u \;\text{ for all }\; (s,t) \in \Delta_T \right\}$$

with respect to the rough path distance $\|\cdot\,;\cdot\|_p$ (as defined in (3.3)), where $\int_s^t A_{s,u} \otimes \mathrm{d}A_u$ is defined as a left-point Riemann–Stieltjes integral. Then, the solution map $\Phi\colon A \mapsto Z$ extends continuously to the space $\mathcal{D}^{0,p}([0,T];\mathbb{R}^d)$ by Theorem 3.1.1, and every path satisfying Property (RIE) is in $\mathcal{D}^{0,p'}([0,T];\mathbb{R}^d)$ for $p' \in (p,3)$ by Lemma 3.1.6.

Next, we shall verify that the piecewise constant approximation X^n of X, as defined in Property (RIE), itself satisfies Property (RIE) relative to any sequence of partitions $(\widetilde{\mathcal{P}}^m)_{m\in\mathbb{N}}$ which are coarser than $\mathcal{P}^n$ and have vanishing mesh size.

Lemma 3.1.9. *Suppose that a path X satisfies Property (RIE) relative to $p \in (2,3)$ and a sequence of partitions $(\mathcal{P}^n)_{n\in\mathbb{N}}$, and let X^n be the usual piecewise constant approximation of X along $\mathcal{P}^n$. Then the path X^n satisfies Property (RIE) relative to p and any sequence of partitions $(\widetilde{\mathcal{P}}^m)_{m\in\mathbb{N}}$ such that $\mathcal{P}^n \subseteq \widetilde{\mathcal{P}}^m$ for every $m \in \mathbb{N}$, and $|\widetilde{\mathcal{P}}^m| \to 0$ as $m \to \infty$.*

Proof. We need to verify each of the conditions (i)–(iii) of Property (RIE) along the sequence of partitions $(\widetilde{\mathcal{P}}^m)_{m\in\mathbb{N}}$. Since $\mathcal{P}^n \subseteq \widetilde{\mathcal{P}}^m$ for every $m \in \mathbb{N}$, the piecewise constant approximation of X^n along the partition $\widetilde{\mathcal{P}}^m$ is simply the path X^n itself. Conditions (i) and (ii) thus hold trivially.

Let $w_{1,n}$ be the control function given by $w_{1,n}(s,t) := \|X^n\|^p_{p,[s,t]}$, so that $|X^n_{s,t}|^p \leq w_{1,n}(s,t)$ for all $(s,t) \in \Delta_T$, and similarly let $w_{2,n}$ be the control function given by $w_{2,n}(s,t) := \|X^n\|^{\frac{p}{2}}_{\frac{p}{2},[s,t]}$. Let us also write $\widetilde{\mathcal{P}}^m = \{0 = r^m_0 < r^m_1 < \cdots < r^m_{\widetilde{N}_m} = T\}$ for each $m \in \mathbb{N}$. Then, for any $m \in \mathbb{N}$ and any $0 \leq k < \ell \leq \widetilde{N}_m$, using the standard estimate for Young integration (see e.g. [75, Proposition 2.4]) we have that

$$\begin{aligned}\left|\int_{r^m_k}^{r^m_\ell} X^n_u \otimes \mathrm{d}X^n_u - X^n_{r^m_k} \otimes X^n_{r^m_k,r^m_\ell}\right|^{\frac{p}{2}} &\lesssim \|X^n\|^{\frac{p}{2}}_{p,[r^m_k,r^m_\ell]} \|X^n\|^{\frac{p}{2}}_{\frac{p}{2},[r^m_k,r^m_\ell]} \\ &\leq \|X^n\|^{\frac{p}{2}}_p \|X^n\|^{\frac{p}{2}}_{\frac{p}{2},[r^m_k,r^m_\ell]} \leq \|X\|^{\frac{p}{2}}_p w_{2,n}(r^m_k, r^m_\ell).\end{aligned}$$

Thus condition (iii) holds for X^n with the control function $w_{3,n}$, given by

$$w_{3,n}(s,t) := w_{1,n}(s,t) + \|X\|^{\frac{p}{2}}_p w_{2,n}(s,t), \qquad (s,t) \in \Delta_T.$$

□

We are now in a position to complete the proof of Theorem 3.1.2. For this, we will apply in particular the result of Theorem A.3.2, which states that, under Property (RIE), the rough integral can be obtained as a limit of classical left-point Riemann sums.

Proof of Theorem 3.1.2. Note that the Euler scheme in (3.7) may be expressed as the solution of the controlled ODE

$$Y^n_t = y_0 + \int_0^t b(\gamma^n_s, Y^n_s)\,\mathrm{d}\gamma^n_s + \int_0^t \sigma(\gamma^n_s, Y^n_s)\,\mathrm{d}X^n_s, \qquad t \in [0,T], \tag{3.22}$$

where γ^n denotes the time discretization path along $\mathcal{P}^n$ defined in (3.17), and the integrals are defined as limits of left-point Riemann sums. Recall that $\widetilde{Y}^n$ denotes the solution of the RDE in (3.18), that is

$$\widetilde{Y}^n_t = y_0 + \int_0^t b(\gamma^n_s, \widetilde{Y}^n_s)\,\mathrm{d}\gamma^n_s + \int_0^t \sigma(\gamma^n_s, \widetilde{Y}^n_s)\,\mathrm{d}\mathbf{X}^n_s, \qquad t \in [0,T], \tag{3.23}$$

where $\mathbf{X}^n$ is the canonical rough path lift of X^n, as constructed in (3.11).

Since X^n is piecewise constant, it is clear from the definition of $\mathbb{X}^n$ that $\mathbb{X}^n_{s,t} = 0$ for any times $s \leq t$ which lie in the same subinterval $[t^n_k, t^n_{k+1})$ of the partition $\mathcal{P}^n$. Since γ^n is also constant on each such subinterval, it follows from the definitions of Young and rough integrals that the solution $\widetilde{Y}^n$ of (3.23) is itself also piecewise constant along the partition $\mathcal{P}^n$.

Let $\widetilde{\mathcal{P}}^m = \{0 = r_0^m < r_1^m < \cdots < r_{\widetilde{N}_m}^m = T\}$, $m \in \mathbb{N}$, be any sequence of partitions with mesh size converging to 0, such that $\mathcal{P}^n \subseteq \widetilde{\mathcal{P}}^m$ for every $m \in \mathbb{N}$. By Lemma 3.1.9, we have that the path X^n satisfies Property (RIE) relative to p and the sequence $(\widetilde{\mathcal{P}}^m)_{m\in\mathbb{N}}$. Since γ^n and $\widetilde{Y}^n$ are piecewise constant along the partition $\mathcal{P}^n$, it is clear that the jump times of the integrand $s \mapsto \sigma(\gamma_s^n, \widetilde{Y}_s^n)$ all belong to $\mathcal{P}^n$, and thus also belong to the set $\liminf_{m\to\infty} \widetilde{\mathcal{P}}^m$. It thus follows from Theorem A.3.2 that the rough integral $\int_0^t \sigma(\gamma_s^n, \widetilde{Y}_s^n) \,\mathrm{d}\mathbf{X}_s^n$ is equal to a limit of left-point Riemann sums along the sequence $(\widetilde{\mathcal{P}}^m)_{m\in\mathbb{N}}$. That is, for any $t \in [0,T]$, we have that

$$\begin{aligned}\int_0^t \sigma(\gamma_s^n, \widetilde{Y}_s^n) \,\mathrm{d}\mathbf{X}_s^n &= \lim_{m\to\infty} \sum_{k=0}^{\widetilde{N}_m-1} \sigma(\gamma_{r_k^m}^n, \widetilde{Y}_{r_k^m}^n) X_{r_k^m\wedge t, r_{k+1}^m\wedge t}^n \\ &= \sum_{k=0}^{N_n-1} \sigma(\gamma_{t_k^n}^n, \widetilde{Y}_{t_k^n}^n) X_{t_k^n\wedge t, t_{k+1}^n\wedge t}^n = \int_0^t \sigma(\gamma_s^n, \widetilde{Y}_s^n) \,\mathrm{d}X_s^n.\end{aligned}$$

Since these integrals are equal, it follows that the ODE in (3.22) and the RDE in (3.23) are consistent, so that $Y^n = \widetilde{Y}^n$. The result then follows from Proposition 3.1.7. $\square$

3.1.3 Error bound for an approximate Euler scheme

In general, the Euler scheme (3.7) is not applicable to numerically approximate the solution of a stochastic differential equation driven by an arbitrary Lévy process, since the increments of Lévy processes cannot always be simulated. Therefore, to obtain a numerical approximation of the solution of such a Lévy-driven SDE, one needs to consider approximate Euler schemes—see e.g. [95, 150, 53]—where the increments of the driving Lévy process are approximated by random variables with known distributions.

As a pathwise counterpart, we introduce the approximate Euler scheme $\widehat{Y}^n$ of the rough differential equation (3.6) along the partition $\mathcal{P}^n$, given by

$$\widehat{Y}_t^n = y_0 + \sum_{i:t_{i+1}^n\le t} b(t_i^n, \widehat{Y}_{t_i^n}^n)(t_{i+1}^n - t_i^n) + \sum_{i:t_{i+1}^n\le t} \sigma(t_i^n, \widehat{Y}_{t_i^n}^n)(\widehat{X}_{t_{i+1}^n} - \widehat{X}_{t_i^n}), \qquad (3.24)$$

for $t \in [0,T]$, with the modified driving signal

$$\widehat{X} := X + \varphi,$$

where $\varphi \in D^q([0,T];\mathbb{R}^d)$, for some $q \in [1,2)$ such that $\frac{1}{p} + \frac{1}{q} > 1$, and, as usual, we write $\mathcal{P}^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$.

While the approximation error of the Euler scheme (3.7) was only caused by discretizing the time interval $[0,T]$, the approximate Euler scheme (3.24) produces an additional approximation error due to taking the modified driving signal $\widehat{X}$ as an input, instead of the actual driving signal X of the RDE (3.6).

To ensure the convergence of the approximate Euler scheme, we first need to verify that, if the actual driving signal satisfies Property (RIE), then the same is true of the modified driving signal.

Proposition 3.1.10. *Suppose that $X \in D([0,T];\mathbb{R}^d)$ satisfies Property* (RIE) *relative to some $p \in (2,3)$ and a sequence of partitions $\mathcal{P}^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$. Let $\varphi \in D^q([0,T];\mathbb{R}^d)$ for some $q \in [1,2)$ such that $\frac{1}{p} + \frac{1}{q} > 1$. For each $n \in \mathbb{N}$, we define $\varphi^n \colon [0,T] \to \mathbb{R}^d$ by*

$$\varphi_t^n = \varphi_T \mathbf{1}_{\{T\}}(t) + \sum_{k=0}^{N_n-1} \varphi_{t_k^n} \mathbf{1}_{[t_k^n, t_{k+1}^n)}(t), \qquad t \in [0,T], \tag{3.25}$$

as the discretization of φ along $\mathcal{P}^n$. Suppose that $\|\varphi^n - \varphi\|_q \to 0$ as $n \to \infty$. Then the path $\widehat{X} = X + \varphi$ satisfies Property (RIE) *relative to p and $(\mathcal{P}^n)_{n\in\mathbb{N}}$.*

Proof. We need to verify the conditions (i)–(iii) of Property (RIE).

(i): Letting $\widehat{X}^n$ denote the piecewise constant approximation of $\widehat{X}$ along the partition $\mathcal{P}^n$, it is clear that $\widehat{X}^n = X^n + \varphi^n$ for each $n \in \mathbb{N}$. Since X^n converges uniformly to X by Property (RIE), and $\|\varphi^n - \varphi\|_q \to 0$ by assumption, it is clear that $\widehat{X}^n$ converges uniformly to $\widehat{X}$ as $n \to \infty$.

(ii): We need to verify that the integral

$$\int_0^t \widehat{X}_u^n \otimes \mathrm{d}\widehat{X}_u = \int_0^t X_u^n \otimes \mathrm{d}X_u + \int_0^t X_u^n \otimes \mathrm{d}\varphi_u + \int_0^t \varphi_u^n \otimes \mathrm{d}X_u + \int_0^t \varphi_u^n \otimes \mathrm{d}\varphi_u,$$

converges as $n \to \infty$ to the limit

$$\int_0^t \widehat{X}_u \otimes \mathrm{d}\widehat{X}_u := \int_0^t X_u \otimes \mathrm{d}X_u + \int_0^t X_u \otimes \mathrm{d}\varphi_u + \int_0^t \varphi_u \otimes \mathrm{d}X_u + \int_0^t \varphi_u \otimes \mathrm{d}\varphi_u,$$

uniformly in $t \in [0,T]$, where the latter three integrals are defined as Young integrals.

Since X satisfies Property (RIE), we have that

$$\left\| \int_0^\cdot X_u^n \otimes \mathrm{d}X_u - \int_0^\cdot X_u \otimes \mathrm{d}X_u \right\|_\infty \longrightarrow 0 \qquad \text{as} \quad n \longrightarrow \infty.$$

Let $p' > p$ such that $\frac{1}{p'} + \frac{1}{q} > 1$. By the standard estimate for Young integrals—see e.g. [75, Proposition 2.4]—we have, for all $t \in [0,T]$, that

$$\left| \int_0^t X_u^n \otimes \mathrm{d}\varphi_u - \int_0^t X_u \otimes \mathrm{d}\varphi_u \right| \lesssim \|X^n - X\|_{p'} \|\varphi\|_q.$$

It follows by interpolation (see e.g. [74, Proposition 5.5]) that

$$\|X^n - X\|_{p'} \leq \|X^n - X\|_\infty^{1-\frac{p}{p'}} \|X^n - X\|_p^{\frac{p}{p'}}.$$

Since X^n converges uniformly to X as $n \to \infty$, and $\sup_{n\in\mathbb{N}}\|X^n\|_p \leq \|X\|_p < \infty$, we deduce that

$$\left\| \int_0^{\cdot} X_u^n \otimes \mathrm{d}\varphi_u - \int_0^{\cdot} X_u \otimes \mathrm{d}\varphi_u \right\|_\infty \longrightarrow 0 \qquad \text{as} \quad n \longrightarrow \infty.$$

Similarly, for each $t \in [0,T]$, it holds that

$$\left| \int_0^t \varphi_u^n \otimes \mathrm{d}X_u - \int_0^t \varphi_u \otimes \mathrm{d}X_u \right| \lesssim \|\varphi^n - \varphi\|_q \|X\|_p,$$

and

$$\left| \int_0^t \varphi_u^n \otimes \mathrm{d}\varphi_u - \int_0^t \varphi_u \otimes \mathrm{d}\varphi_u \right| \lesssim \|\varphi^n - \varphi\|_q \|\varphi\|_q,$$

and, since $\|\varphi^n - \varphi\|_q \to 0$ as $n \to \infty$, we infer the required convergence.

(iii): We aim to find a control function w such that

$$\sup_{(s,t)\in\Delta_T} \frac{|\widehat{X}_{s,t}|^p}{w(s,t)} + \sup_{n\in\mathbb{N}} \sup_{0\leq k<\ell\leq N_n} \frac{|\int_{t_k^n}^{t_\ell^n} \widehat{X}_{t_k^n,u}^n \otimes \mathrm{d}\widehat{X}_u|^{\frac{p}{2}}}{w(t_k^n, t_\ell^n)} \leq 1, \tag{3.26}$$

where

$$\int_{t_k^n}^{t_\ell^n} \widehat{X}_{t_k^n,u}^n \otimes \mathrm{d}\widehat{X}_u = \int_{t_k^n}^{t_\ell^n} X_{t_k^n,u}^n \otimes \mathrm{d}X_u + \int_{t_k^n}^{t_\ell^n} X_{t_k^n,u}^n \otimes \mathrm{d}\varphi_u + \int_{t_k^n}^{t_\ell^n} \varphi_{t_k^n,u}^n \otimes \mathrm{d}X_u + \int_{t_k^n}^{t_\ell^n} \varphi_{t_k^n,u}^n \otimes \mathrm{d}\varphi_u.$$

Let w_X be the control function with respect to which X satisfies Property (RIE), and define moreover the control function w_φ, given by $w_\varphi(s,t) = \|\varphi\|_{q,[s,t]}^q$ for $(s,t) \in \Delta_T$.

We have from Property (RIE) that

$$\sup_{(s,t)\in\Delta_T} \frac{|\widehat{X}_{s,t}|^p}{w_X(s,t) + w_\varphi(s,t)} \lesssim \sup_{(s,t)\in\Delta_T} \frac{|X_{s,t}|^p}{w_X(s,t)} + \sup_{(s,t)\in\Delta_T} \frac{|\varphi_{s,t}|^p}{w_\varphi(s,t)} \leq 2,$$

and that

$$\sup_{n\in\mathbb{N}} \sup_{0\leq k<\ell\leq N_n} \frac{|\int_{t_k^n}^{t_\ell^n} X_{t_k^n,u}^n \otimes \mathrm{d}X_u|^{\frac{p}{2}}}{w_X(t_k^n, t_\ell^n)} \leq 1.$$

By the standard estimate for Young integrals (see e.g. [75, Proposition 2.4]), for every $n \in \mathbb{N}$ and $0 \leq k < \ell \leq N_n$, we have

$$\begin{aligned} \left| \int_{t_k^n}^{t_\ell^n} X_{t_k^n,u}^n \otimes \mathrm{d}\varphi_u \right|^{\frac{p}{2}} &\lesssim \|X^n\|_{p,[t_k^n,t_\ell^n]}^{\frac{p}{2}} \|\varphi\|_{q,[t_k^n,t_\ell^n]}^{\frac{p}{2}} \\ &\leq \|X\|_{p,[t_k^n,t_\ell^n]}^{\frac{p}{2}} \|\varphi\|_{q,[t_k^n,t_\ell^n]}^{\frac{p}{2}} \leq w_X(t_k^n, t_\ell^n)^{\frac{1}{2}} w_\varphi(t_k^n, t_\ell^n)^{\frac{p}{2q}}, \end{aligned}$$

and we can similarly obtain

$$\left| \int_{t_k^n}^{t_\ell^n} \varphi_{t_k^n,u}^n \otimes \mathrm{d}X_u \right|^{\frac{p}{2}} \lesssim w_X(t_k^n, t_\ell^n)^{\frac{1}{2}} w_\varphi(t_k^n, t_\ell^n)^{\frac{p}{2q}}$$

and

$$\left|\int_{t_k^n}^{t_\ell^n} \varphi^n_{t_k^n,u} \otimes \mathrm{d}\varphi_u\right|^{\frac{p}{2}} \lesssim w_\varphi(t_k^n, t_\ell^n)^{\frac{p}{q}}.$$

Since $p \in (2,3)$ and $q \in [1,2)$, we have that $\frac{1}{2} + \frac{p}{2q} > 1$ and $\frac{p}{q} > 1$, and it follows that the maps $(s,t) \mapsto w_X(s,t)^{\frac{1}{2}} w_\varphi(s,t)^{\frac{p}{2q}}$ and $(s,t) \mapsto w_\varphi(s,t)^{\frac{p}{q}}$ are superadditive and thus control functions. We deduce that (3.26) holds with a control function w of the form

$$w(s,t) = C\Big(w_X(s,t) + w_\varphi(s,t) + w_X(s,t)^{\frac{1}{2}} w_\varphi(s,t)^{\frac{p}{2q}} + w_\varphi(s,t)^{\frac{p}{q}}\Big), \qquad (s,t) \in \Delta_T,$$

where $C > 0$ is a suitable constant which depends only on p and q. $\square$

By Proposition 3.1.10, the modified driving signal $\widehat{X}$ satisfies Property (RIE), and can thus be canonically lifted to a rough path $\widehat{\mathbf{X}} = (\widehat{X}, \widehat{\mathbb{X}}) \in \mathscr{D}^p([0,T];\mathbb{R}^d)$ via (3.9). By Theorem 3.1.1, the rough differential equation (3.6) driven by $\widehat{\mathbf{X}}$ has a unique solution $\widehat{Y}$, and the approximate Euler scheme $\widehat{Y}^n$ in (3.24) converges to $\widehat{Y}$ by Theorem 3.1.2.

The next proposition provides an error and convergence analysis for the approximate Euler scheme (3.24) with respect to the solution Y of the RDE (3.6) driven by the rough path $\mathbf{X} = (X, \mathbb{X})$ under Property (RIE).

Proposition 3.1.11. *Suppose that $X \in D([0,T];\mathbb{R}^d)$ satisfies Property* (RIE) *relative to $p \in (2,3)$ and a sequence of partitions $(\mathcal{P}^n)_{n\in\mathbb{N}}$, and let $\mathbf{X}$ be its canonical rough path lift. Let $\varphi \in D^q([0,T];\mathbb{R}^d)$ for some $q \in (1,2)$ such that $\frac{1}{p} + \frac{1}{q} > 1$, let φ^n be the piecewise constant approximation of φ, as defined in* (3.25)*, and assume that $\|\varphi^n - \varphi\|_q \to 0$ as $n \to \infty$. Let Y be the solution of the RDE* (3.6) *driven by $\mathbf{X}$, and let $\widehat{Y}^n$ be the approximate Euler scheme defined in* (3.24)*. We have the error estimate*

$$\begin{aligned}\|\widehat{Y}^n - Y\|_{p'} \lesssim\ & (1 + \|X\|_p + \|\varphi\|_q)\|\varphi\|_q + |\mathcal{P}^n|^{1-\frac{1}{q}} + (\|X^n - X\|_\infty + \|\varphi^n - \varphi\|_\infty)^{1-\frac{p}{p'}} \\ & + \bigg(\bigg\|\int_0^{\cdot} X_u^n \otimes \mathrm{d}X_u - \int_0^{\cdot} X_u \otimes \mathrm{d}X_u\bigg\|_\infty + \|X^n - X\|_{p'} + \|\varphi^n - \varphi\|_q\bigg)^{1-\frac{p}{p'}}\end{aligned}$$

for any $p' \in (p,3)$ such that $\frac{1}{p'} + \frac{1}{q} > 1$, where the implicit multiplicative constant depends on $p, p', q, \|b\|_{C_b^2}, \|\sigma\|_{C_b^3}, T, \|X\|_\infty, \|\mathbf{X}\|_p, \|\varphi\|_\infty, \|\varphi\|_q$ and $w(0,T)$, where w is the control function for which (3.8) *holds. In particular, we have that*

$$\limsup_{n\to\infty} \|\widehat{Y}^n - Y\|_{p'} \lesssim (1 + \|X\|_p + \|\varphi\|_q)\|\varphi\|_q. \tag{3.27}$$

Proof. By Proposition 3.1.10, we know that the path $\widehat{X} = X + \varphi$ satisfies Property (RIE) relative to p and $(\mathcal{P}^n)_{n\in\mathbb{N}}$. Let $\widehat{\mathbf{X}}$ be the canonical rough path lift of $\widehat{X}$, and let Y and $\widehat{Y}$ be the solutions of the RDE (3.6) driven by $\mathbf{X}$ and $\widehat{\mathbf{X}}$ respectively. It is clear that

$$\|\widehat{Y}^n - Y\|_{p'} \leq \|\widehat{Y}^n - \widehat{Y}\|_{p'} + \|\widehat{Y} - Y\|_{p'}.$$

By Theorem 3.1.1, we have the estimate

$$\|\widehat{Y} - Y\|_{p'} \lesssim \|\widehat{\mathbf{X}}; \mathbf{X}\|_{p'},$$

and, by Theorem 3.1.2, we have that

$$\|\widehat{Y}^n - \widehat{Y}\|_{p'} \lesssim |\mathcal{P}^n|^{1-\frac{1}{q}} + \|\widehat{X}^n - \widehat{X}\|_\infty^{1-\frac{p}{p'}} + \left\| \int_0^\cdot \widehat{X}_u^n \otimes \mathrm{d}\widehat{X}_u - \int_0^\cdot \widehat{X}_u \otimes \mathrm{d}\widehat{X}_u \right\|_\infty^{1-\frac{p}{p'}},$$

where $\widehat{X}^n$ is the piecewise constant approximation of $\widehat{X}$ along $\mathcal{P}^n$. Since $\widehat{X}^n = X^n + \varphi^n$, we can bound

$$\|\widehat{X}^n - \widehat{X}\|_\infty \leq \|X^n - X\|_\infty + \|\varphi^n - \varphi\|_\infty.$$

As shown in the proof of Proposition 3.1.10,

$$\begin{aligned} &\left\| \int_0^\cdot \widehat{X}_u^n \otimes \mathrm{d}\widehat{X}_u - \int_0^\cdot \widehat{X}_u \otimes \mathrm{d}\widehat{X}_u \right\|_\infty \\ &\quad \lesssim \left\| \int_0^\cdot X_u^n \otimes \mathrm{d}X_u - \int_0^\cdot X_u \otimes \mathrm{d}X_u \right\|_\infty + \|X^n - X\|_{p'} \|\varphi\|_q + \|\varphi^n - \varphi\|_q (\|X\|_p + \|\varphi\|_q). \end{aligned}$$

We also note that

$$\widehat{\mathbb{X}}_{s,t} - \mathbb{X}_{s,t} = \int_s^t X_{s,u} \otimes \mathrm{d}\varphi_u + \int_s^t \varphi_{s,u} \otimes \mathrm{d}X_u + \int_s^t \varphi_{s,u} \otimes \mathrm{d}\varphi_u$$

for $(s,t) \in \Delta_T$, so that, by the standard estimate for Young integrals (see e.g. [75, Proposition 2.4]), we obtain

$$|\widehat{\mathbb{X}}_{s,t} - \mathbb{X}_{s,t}| \lesssim \|X\|_{p,[s,t]} \|\varphi\|_{q,[s,t]} + \|\varphi\|_{q,[s,t]}^2.$$

This implies that, for any partition $\mathcal{P}$ of the interval $[0,T]$,

$$\begin{aligned} \sum_{[s,t] \in \mathcal{P}} |\widehat{\mathbb{X}}_{s,t} - \mathbb{X}_{s,t}|^{\frac{p}{2}} &\lesssim \sum_{[s,t] \in \mathcal{P}} \left(\|X\|_{p,[s,t]}^{\frac{p}{2}} \|\varphi\|_{q,[s,t]}^{\frac{p}{2}} + \|\varphi\|_{q,[s,t]}^p \right) \\ &\leq \Bigg(\sum_{[s,t] \in \mathcal{P}} \|X\|_{p,[s,t]}^p \Bigg)^{\frac{1}{2}} \Bigg(\sum_{[s,t] \in \mathcal{P}} \|\varphi\|_{q,[s,t]}^p \Bigg)^{\frac{1}{2}} + \sum_{[s,t] \in \mathcal{P}} \|\varphi\|_{q,[s,t]}^p \\ &\leq \Bigg(\sum_{[s,t] \in \mathcal{P}} \|X\|_{p,[s,t]}^p \Bigg)^{\frac{1}{2}} \Bigg(\sum_{[s,t] \in \mathcal{P}} \|\varphi\|_{q,[s,t]}^q \Bigg)^{\frac{p}{2q}} + \Bigg(\sum_{[s,t] \in \mathcal{P}} \|\varphi\|_{q,[s,t]}^q \Bigg)^{\frac{p}{q}} \leq \|X\|_p^{\frac{p}{2}} \|\varphi\|_q^{\frac{p}{2}} + \|\varphi\|_q^p, \end{aligned}$$

so that $\|\widehat{\mathbb{X}} - \mathbb{X}\|_{\frac{p}{2}} \lesssim \|X\|_p \|\varphi\|_q + \|\varphi\|_q^2$. We thus deduce that

$$\|\widehat{\mathbf{X}}; \mathbf{X}\|_{p'} \leq \|\widehat{X} - X\|_p + \|\widehat{\mathbb{X}} - \mathbb{X}\|_{\frac{p}{2}} \lesssim (1 + \|X\|_p + \|\varphi\|_q) \|\varphi\|_q,$$

and combining the estimates above, we obtain the desired error estimate. □

Remark 3.1.12. *As an immediate consequence of Proposition 3.1.11, if the modified driving signal* $\widehat{X}$ *converges to the driving signal* X*, then the approximate Euler scheme converges to the solution* Y *of the RDE* (3.6)*. More precisely, let us consider the approximate Euler scheme* $\check{Y}^n$ *of the RDE* (3.6) *along the partition* $\mathcal{P}^n$*, given by*

$$\check{Y}_t^n = y_0 + \sum_{i:\, t_{i+1}^n \leq t} b(t_i^n, \check{Y}_{t_i^n}^n)(t_{i+1}^n - t_i^n) + \sum_{i:\, t_{i+1}^n \leq t} \sigma(t_i^n, \check{Y}_{t_i^n}^n)(\check{X}_{t_{i+1}^n}^n - \check{X}_{t_i^n}^n),$$

for $t \in [0,T]$*, with the modified driving signal*

$$\check{X}^n := X + \psi^n,$$

where $\psi^n \in D^q([0,T];\mathbb{R}^d)$ *for some* $q \in (1,2)$ *such that* $\frac{1}{p} + \frac{1}{q} > 1$*. In the setting of Proposition 3.1.11 with* $\varphi = \psi^n$*, if* $\|\psi^n\|_q \to 0$ *as* $n \to \infty$*, then*

$$\|\check{Y}^n - Y\|_{p'} \longrightarrow 0 \qquad \text{as} \quad n \longrightarrow \infty,$$

for any $p' \in (p,3)$ *such that* $\frac{1}{p'} + \frac{1}{q} > 1$*, which follows from* (3.27)*.*

Remark 3.1.13. *In this section we handled the modified driving signal* $X+\varphi$ *by considering the rough path lift* $\widehat{\mathbf{X}}$ *of* $\widehat{X} = X + \varphi$*, and considering the solution* $\widehat{Y}$ *of the RDE* (3.6) *driven by* $\widehat{\mathbf{X}}$*. An alternative, equally valid approach would be to instead absorb* φ *into the drift of the RDE. The resulting equation would not strictly speaking be of the form in* (3.6)*, but it would still fall into the regime of the more general RDE in* (3.4)*, and an error estimate could be obtained using the stability estimate in Theorem 3.1.1.*

3.2 Application to stochastic differential equations

In this section we apply the deterministic theory developed in Section 3.1, regarding the Euler scheme for RDEs, to stochastic differential equations (SDEs). For this purpose, we now let X be a d-dimensional càdlàg semimartingale, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)_{t\in[0,T]}$ satisfying the usual conditions, i.e., completeness and right-continuity. We consider the stochastic differential equation

$$Y_t = y_0 + \int_0^t b(s, Y_{s-})\,\mathrm{d}s + \int_0^t \sigma(s, Y_{s-})\,\mathrm{d}X_s, \qquad t \in [0,T], \tag{3.28}$$

where $y_0 \in \mathbb{R}^k$, $b \in C_b^2(\mathbb{R}^{k+1}; \mathcal{L}(\mathbb{R};\mathbb{R}^k))$ and $\sigma \in C_b^3(\mathbb{R}^{k+1}; \mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$, and $\int_0^t \sigma(s, Y_{s-})\,\mathrm{d}X_s$ is defined as an Itô integral. For a comprehensive introduction to stochastic Itô integration and SDEs we refer, e.g., to the textbook [147]. It is well-known that the SDE (3.28) possesses a unique (strong) solution (see e.g. [147, Chapter V, Theorem 6]), and that the semimartingale X can be lifted to a random rough path via Itô integration, by defining $\mathbf{X} = (X, \mathbb{X}) \in \mathcal{D}^p([0,T];\mathbb{R}^d)$, $\mathbb{P}$-a.s., for any $p \in (2,3)$, where

$$\mathbb{X}_{s,t} := \int_s^t (X_{r-} - X_s) \otimes \mathrm{d}X_r = \int_s^t X_{r-} \otimes \mathrm{d}X_r - X_s \otimes X_{s,t}, \qquad (s,t) \in \Delta_T; \tag{3.29}$$

see [124, Proposition 3.4] or [75, Theorem 6.5]. It turns out that, if the semimartingale X satisfies Property (RIE) relative to $p \in (2,3)$ and a suitable sequence of partitions $(\mathcal{P}^n)_{n\in\mathbb{N}}$, then the solutions to the SDE (3.28) and to the RDE (3.6) driven by the random rough path $\mathbf{X} = (X, \mathbb{X})$ coincide $\mathbb{P}$-almost surely.

Lemma 3.2.1. *Let $p \in (2,3)$ and let $\mathcal{P}^n = \{\tau_k^n\}$, $n \in \mathbb{N}$, be a sequence of adapted partitions (so that each τ_k^n is a stopping time), such that, for almost every $\omega \in \Omega$, $(\mathcal{P}^n(\omega))_{n\in\mathbb{N}}$ is a sequence of (finite) partitions of $[0,T]$ with vanishing mesh size. Let X be a càdlàg semimartingale, and suppose that, for almost every $\omega \in \Omega$, the sample path $X(\omega)$ satisfies Property* (RIE) *relative to p and $(\mathcal{P}^n(\omega))_{n\in\mathbb{N}}$.*

(i) The random rough paths $\mathbf{X} = (X, \mathbb{X})$, with $\mathbb{X}$ defined pathwise via (3.9)*, and with $\mathbb{X}$ defined by stochastic integration as in* (3.29)*, coincide $\mathbb{P}$-almost surely.*

(ii) The solution of the SDE (3.28) *driven by X, and the solution of the RDE* (3.6) *driven by the random rough path $\mathbf{X} = (X, \mathbb{X})$, coincide $\mathbb{P}$-almost surely.*

Proof. (i): By construction, the pathwise rough integral $\int_0^t X_u(\omega) \otimes \mathrm{d}X_u(\omega)$ constructed via Property (RIE) is given by the limit as $n \to \infty$ of left-point Riemann sums:

$$\sum_{k=0}^{N_n-1} X_{\tau_k^n(\omega)}(\omega) \otimes X_{\tau_k^n(\omega)\wedge t, \tau_{k+1}^n(\omega)\wedge t}(\omega). \tag{3.30}$$

It is known that these Riemann sums also converge uniformly in probability to the Itô integral $\int_0^t X_{u-} \otimes \mathrm{d}X_u$ (see e.g. [147, Chapter II, Theorem 21]), and the result thus follows from the (almost sure) uniqueness of limits.

(ii): In the following, we adopt the notation J_F for the set of jump times of a path F, and we write $\liminf_{n\to\infty} \mathcal{P}^n := \bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} \mathcal{P}^n$.

Let Y be the solution of the RDE (3.6) driven by the random rough path $\mathbf{X} = (X, \mathbb{X})$. By the definition of $\mathbb{X}$ in terms of a limit of the Riemann sums in (3.30), it is straightforward to see that $\mathbb{X}_{t-,t} = 0$ unless X has a jump at time t. It follows from the definition of rough integration that the integral $t \mapsto \int_0^t \sigma(s, Y_s)\,\mathrm{d}\mathbf{X}_s$ can only have a jump at the jump times of X, and it follows that the same is true of the solution Y of the RDE (3.6), i.e., $J_Y \subseteq J_X$.

Since the piecewise constant approximation X^n of X along $\mathcal{P}^n$ converges uniformly to X (by condition (i) of Property (RIE)), we have from Proposition A.3.1 that $J_X \subseteq$

$\liminf_{n\to\infty} \mathcal{P}^n$. Since $J_Y \subseteq J_X$, we have that $J_Y \subseteq \liminf_{n\to\infty} \mathcal{P}^n$. It then follows from Theorem A.3.2 that

$$\int_0^t \sigma(s, Y_s) \,\mathrm{d}\mathbf{X}_s = \lim_{n\to\infty} \sum_{k=0}^{N_n-1} \sigma(\tau_k^n, Y_{\tau_k^n}) X_{\tau_k^n\wedge t, \tau_{k+1}^n \wedge t}.$$

Since these Riemann sums also converge in probability to the Itô integral $\int_0^t \sigma(s, Y_{s-}) \,\mathrm{d}X_s$ (see e.g. [147, Chapter II, Theorem 21]), these integrals coincide almost surely. We infer that Y is also a solution of the SDE (3.28), which has a unique solution (by e.g. [147, Chapter V, Theorem 6]). □

As a consequence of Theorem 3.1.2 and Lemma 3.2.1, for semimartingales which satisfy Property (RIE) relative to a sequence of adapted partitions, the Euler scheme (3.7) converges pathwise to the solution of the SDE (3.28). In the following subsections we verify Property (RIE) for various semimartingales relative to suitable sequences of partitions, and derive the pathwise convergence rate of the associated Euler scheme with respect to the p-variation norm.

3.2.1 Brownian motion

We start with the most prominent example of a semimartingale, by taking $X = W$ to be a d-dimensional Brownian motion $W = (W_t)_{t\in[0,T]}$ with respect to the underlying filtration $(\mathcal{F}_t)_{t\in[0,T]}$.

Proposition 3.2.2. *Let $p \in (2,3)$ and let $\mathcal{P}^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$, be a sequence of equidistant partitions of the interval $[0,T]$, so that, for each $n \in \mathbb{N}$, there exists some $\pi_n > 0$ such that $t_{i+1}^n - t_i^n = \pi_n$ for each $0 \leq i < N_n$. If $\pi_n^{2-\frac{4}{p}} \log(n) \to 0$ as $n \to \infty$, then, for almost every $\omega \in \Omega$, the sample path $W(\omega)$ satisfies Property* (RIE) *relative to p and $(\mathcal{P}^n)_{n\in\mathbb{N}}$.*

Proof. As stated in Remark 3.1.3, Davie's condition implies Property (RIE). While [143, Appendix B] show this for the sequence of partitions $(\mathcal{P}_U^n)_{n\in\mathbb{N}}$, where $\mathcal{P}_U^n = \{\frac{iT}{n} : i = 0, 1, \dots, n\}$, i.e. $\pi_n = \frac{T}{n}$, their proof actually holds for any sequence of equidistant partitions of the interval $[0,T]$. We therefore show the necessary condition proposed in [51], under the assumption that $\pi_n^{2-\frac{4}{p}} \log(n) \to 0$ as $n \to \infty$.

More precisely, let $\mathbf{W} = (W, \mathbb{W})$ be the Itô Brownian rough path lift of W. Write $\alpha := \frac{1}{p}$ and let $\beta \in (1-\alpha, 2\alpha)$. We show that, almost surely, there exists a constant $C > 0$ such that

$$\Big| \sum_{m=k}^{\ell-1} \mathbb{W}^{ij}_{t_m^n, t_{m+1}^n} \Big| \leq C(\ell - k)^\beta \pi_n^{2\alpha},$$

for every $i, j = 1, \dots, d$ and $n \in \mathbb{N}$, whenever $0 < k < \ell$ are integers such that $\ell\pi_n \leq T$.

Step 1. We recall that a (zero mean) random variable Z is said to be *sub-Gaussian* if its sub-Gaussian norm $\|Z\|_{\psi_2} := \inf\{z > 0 : \mathbb{E}[\exp(Z^2/z^2)] \leq 2\}$ is finite. It is well-known that the sub-Gaussian property admits an equivalent formulation; namely, Z is sub-Gaussian if and only if $\mathbb{E}[\exp(\lambda^2 Z^2)] \leq \exp(\lambda^2 K^2)$ holds for all λ such that $|\lambda| \leq \frac{1}{K}$, for some $K > 0$. In this case we have $\|Z\|_{\psi_2} = K$ up to a multiplicative constant.

We will prove that $\mathbb{W}^{ij}_{t^n_m, t^n_{m+1}}$, $m = k, \dots, \ell - 1$, are independent sub-Gaussian random variables with sub-Gaussian norm $\|\mathbb{W}^{ij}_{t^n_m, t^n_{m+1}}\|_{\psi_2} = C\pi_n$ for some $C > 0$.

First, we note that, by [74, Proposition 13.4], for all $m \in \mathbb{N}$, the random variables $\frac{\mathbb{W}^{ij}_{t^n_m, t^n_{m+1}}}{t^n_{m+1} - t^n_m}$ are independent and identically distributed, with the same distribution as $\mathbb{W}^{ij}_{0,1}$, and that the latter satisfies $\mathbb{E}[\exp(\eta \mathbb{W}^{ij}_{0,1})] < \infty$ for some sufficiently small $\eta > 0$, which is equivalent to the Gaussian tail property, i.e., that $\|\mathbb{W}^{ij}_{0,1}\|_{L^q} \leq c\sqrt{q}$ for all $q \geq 1$, where the constant c is independent of q; see [74, Lemma A.17]. As a consequence, using the fact that $t^n_{m+1} - t^n_m = \pi_n$ for all m, and setting $q = 2\nu$, we deduce that

$$\mathbb{E}[|\mathbb{W}^{ij}_{t^n_m, t^n_{m+1}}|^{2\nu}] \leq c^\nu \nu^\nu \pi_n^{2\nu}, \qquad \nu \in \mathbb{N}, \tag{3.31}$$

for a new constant $c > 0$ which does not depend on ν.

We now aim to show that there exists a constant $C > 0$ such that

$$\mathbb{E}[\exp(\lambda^2 (\mathbb{W}^{ij}_{t^n_m, t^n_{m+1}})^2)] \leq \exp(C^2 \pi_n^2 \lambda^2), \tag{3.32}$$

for all λ such that $|\lambda| \leq \frac{1}{C\pi_n}$, which then implies that $\mathbb{W}^{ij}_{t^n_m, t^n_{m+1}}$ is sub-Gaussian with norm $\|\mathbb{W}^{ij}_{t^n_m, t^n_{m+1}}\|_{\psi_2} = C\pi_n$, up to a multiplicative constant which we may then absorb into C. Using the Taylor expansion for the exponential function, we get, for $\lambda \in \mathbb{R}$, that

$$\mathbb{E}[\exp(\lambda^2 (\mathbb{W}^{ij}_{t^n_m, t^n_{m+1}})^2)] = \mathbb{E}\left[1 + \sum_{\nu=1}^{\infty} \frac{\lambda^{2\nu} (\mathbb{W}^{ij}_{t^n_m, t^n_{m+1}})^{2\nu}}{\nu!}\right] = 1 + \sum_{\nu=1}^{\infty} \frac{\lambda^{2\nu} \mathbb{E}[(\mathbb{W}^{ij}_{t^n_m, t^n_{m+1}})^{2\nu}]}{\nu!}.$$

By the bound in (3.31) and Stirling's approximation (which implies in particular that $\nu! \geq (\frac{\nu}{e})^\nu$ for all $\nu \geq 1$), we obtain

$$\mathbb{E}[\exp(\lambda^2 (\mathbb{W}^{ij}_{t^n_m, t^n_{m+1}})^2)] \leq 1 + \sum_{\nu=1}^{\infty} (ec\lambda^2 \pi_n^2)^\nu = \frac{1}{1 - ec\lambda^2 \pi_n^2} \leq \exp(2ec\lambda^2 \pi_n^2),$$

which is valid provided that

$$ec\lambda^2 \pi_n^2 \leq \frac{1}{2}, \tag{3.33}$$

since $\frac{1}{1-x} \leq \exp(2x)$ for $x \in [0, \frac{1}{2}]$. We then obtain (3.32) by choosing $C = \sqrt{2ec}$, and note that then (3.33) does indeed hold when $|\lambda| \leq \frac{1}{C\pi_n}$.

Step 2. Let $C > 0$ be the constant found above, so that $\|\mathbb{W}^{ij}_{t^n_m, t^n_{m+1}}\|_{\psi_2} = C\pi_n$. Then Hoeffding's inequality (see e.g. [158, Theorem 2.6.2]) gives

$$\mathbb{P}\Big(\Big|\sum_{m=k}^{\ell-1} \mathbb{W}^{ij}_{t^n_m, t^n_{m+1}}\Big| \geq C(\ell-k)^\beta \pi_n^{2\alpha}\Big) \leq 2\exp\Big(-\frac{C^2(\ell-k)^{2\beta}\pi_n^{4\alpha}}{\sum_{m=k}^{\ell-1}\|\mathbb{W}^{ij}_{t^n_m, t^n_{m+1}}\|^2_{\psi_2}}\Big)$$
$$= 2\exp\Big(-\frac{(\ell-k)^{2\beta-1}}{{\pi_n}^{2-4\alpha}}\Big).$$

Since $\beta > 1 - \alpha > \frac{1}{2}$, we can bound this further by

$$\mathbb{P}\Big(\Big|\sum_{m=k}^{\ell-1} \mathbb{W}^{ij}_{t^n_m, t^n_{m+1}}\Big| \geq C(\ell-k)^\beta \pi_n^{2\alpha}\Big) \leq 2\exp\Big(-\frac{1}{{\pi_n}^{2-4\alpha}}\Big) = 2n^{-\frac{1}{\gamma_n}},$$

where we denote $\gamma_n = {\pi_n}^{2-4\alpha}\log(n)$. Since, by assumption, $\gamma_n \to 0$ as $n \to \infty$, we have that $\frac{1}{\gamma_n} > 1$ for all sufficiently large $n \in \mathbb{N}$, and hence that the series $\sum_{n\in\mathbb{N}} n^{-\frac{1}{\gamma_n}}$ is absolutely convergent. The desired statement then follows from the Borel–Cantelli lemma. □

Remark 3.2.3. *Proposition 3.2.2 can be generalized to any sequence of partitions $(\mathcal{P}^n)_{n\in\mathbb{N}}$, which possibly consists of non-equidistant partitions, such that $|\mathcal{P}^n|^{2-\frac{4}{p}}\log(n) \to 0$ as $n \to \infty$, provided that there exists a positive number $\eta > 0$ such that*

$$\frac{|\mathcal{P}^n|}{\min_{0\leq k<N_n}|t^n_{k+1} - t^n_k|} \leq \eta$$

for every $n \in \mathbb{N}$. This additional condition requires that the sequence $(\mathcal{P}^n)_{n\in\mathbb{N}}$ be a "balanced partition sequence" in the sense of [38].

Remark 3.2.4. *Combining Proposition 3.2.2 with Lemma 3.1.6, we infer that the piecewise constant approximations of a Brownian motion along equidistant partitions converge to its Itô rough path lift, which, as far as we are aware, is a novel construction of this lift. Existing approximations of Brownian rough path are all continuous approximations, such as piecewise linear or mollifier approximations—cf. [74]—which play a crucial role, e.g., in the rough path based proofs of Wong–Zakai results, support theorems and large deviation principles.*

Corollary 3.2.5. *Let $p \in (2,3)$ and let $\mathcal{P}^n_U = \{0 = t^n_0 < t^n_1 < \cdots < t^n_n = T\}$, $n \in \mathbb{N}$, with $t^n_i = \frac{iT}{n}$, be the sequence of equidistant partitions with width $\frac{T}{n}$ of the interval $[0,T]$. Let Y be the solution of the SDE (3.28) driven by a Brownian motion W, and let Y^n be the corresponding Euler approximation along $\mathcal{P}^n_U$, as defined in (3.7). For any $p' \in (p,3)$, $q \in (1,2)$ and $\beta \in (1-\frac{1}{p}, \frac{2}{p})$ such that $\frac{1}{p'} + \frac{1}{q} > 1$, there exists a random variable C, which does not depend on n, such that*

$$\|Y^n - Y\|_{p'} \leq C(n^{-(1-\frac{1}{q})} + n^{-(\frac{2}{p}-\beta)(1-\frac{p}{p'})}), \qquad n \in \mathbb{N}. \tag{3.34}$$

Proof. Since $|\mathcal{P}_U^n| = \frac{T}{n}$, we have that $|\mathcal{P}_U^n|^{2-\frac{4}{p}} \log(n) \to 0$ as $n \to \infty$. Thus, by Proposition 3.2.2, for almost every $\omega \in \Omega$, the sample path $W(\omega)$ satisfies Property (RIE) relative to p and $(\mathcal{P}_U^n)_{n\in\mathbb{N}}$, which allows us to apply the result of Theorem 3.1.2.

Since the sample paths of W are almost surely $\frac{1}{p}$-Hölder continuous, it is easy to see that

$$\|W^n - W\|_\infty \lesssim n^{-\frac{1}{p}}, \qquad n \in \mathbb{N},$$

where the implicit multiplicative constant is a random variable which does not depend on n. Moreover, by [143, Appendix B], the left-point Riemann sums along $(\mathcal{P}_U^n)_{n\in\mathbb{N}}$ converge uniformly as $n \to \infty$, with rate $n^{-(\frac{2}{p}-\beta)}$ for $\beta \in (1-\frac{1}{p}, \frac{2}{p})$, i.e.,

$$\left\| \int_0^\cdot W_u^n \otimes \mathrm{d}W_u - \int_0^\cdot W_u \otimes \mathrm{d}W_u \right\|_\infty \lesssim n^{-(\frac{2}{p}-\beta)}, \qquad n \in \mathbb{N}.$$

Hence, by Theorem 3.1.2, we get that

$$\|Y^n - Y\|_{p'} \lesssim n^{-(1-\frac{1}{q})} + n^{-\frac{1}{p}(1-\frac{p}{p'})} + n^{-(\frac{2}{p}-\beta)(1-\frac{p}{p'})}.$$

Since $\frac{1}{p} < 1 - \frac{1}{p} < \beta$ for $p \in (2,3)$, this gives the rate of convergence in (3.34). $\square$

3.2.2 Itô processes

In this subsection we let X be an Itô process. More precisely, we suppose that

$$X_t = x_0 + \int_0^t b_r \,\mathrm{d}r + \int_0^t H_r \,\mathrm{d}W_r, \qquad t \in [0,T], \tag{3.35}$$

for some $x_0 \in \mathbb{R}^d$, and some locally bounded predictable integrands $b\colon \Omega \times [0,T] \to \mathbb{R}^d$ and $H\colon \Omega \times [0,T] \to \mathcal{L}(\mathbb{R}^m;\mathbb{R}^d)$, where W is an $\mathbb{R}^m$-valued Brownian motion. We consider the sequence of dyadic partitions $(\mathcal{P}_D^n)_{n\in\mathbb{N}}$ of $[0,T]$, given by

$$\mathcal{P}_D^n := \{0 = t_0^n < t_1^n < \cdots < t_{2^n}^n = T\} \qquad \text{with} \quad t_k^n := k2^{-n}T \quad \text{for} \quad k = 0,1,\ldots,2^n. \tag{3.36}$$

In the next proposition we will show that X satisfies Property (RIE) along the sequence of partitions $(\mathcal{P}_D^n)_{n\in\mathbb{N}}$, and establish the rate of convergence of the associated Euler scheme. Note that, in contrast to the proof of Proposition 3.2.2, for general Itô processes we cannot rely on the concentration of measure inequality for sub-Gaussian distributions.

Proposition 3.2.6. *Let $p \in (2,3)$ and let X be an Itô process of the form in* (3.35)*. Let Y be the solution of the SDE* (3.28) *driven by X, and let Y^n denote the corresponding Euler approximation, as defined in* (3.7)*, based on X and the sequence of dyadic partitions $(\mathcal{P}_D^n)_{n\in\mathbb{N}}$.*

(i) For almost every $\omega \in \Omega$, the sample path $X(\omega)$ satisfies Property (RIE) *relative to p and $(\mathcal{P}^n_D)_{n\in\mathbb{N}}$.*

(ii) For any $p' \in (p,3)$ and $q \in (1,2)$ such that $\frac{1}{p'} + \frac{1}{q} > 1$, and any $\varepsilon \in (0,1)$, there exists a random variable C, which does not depend on n, such that

$$\|Y^n - Y\|_{p'} \leq C(2^{-n(1-\frac{1}{q})} + 2^{-n(\frac{1}{p}-\frac{1}{p'})} + 2^{-\frac{n}{2}(1-\varepsilon)(1-\frac{p}{p'})}), \qquad n \in \mathbb{N}, \tag{3.37}$$

and

$$\|Y^n - Y\|_3 \leq C2^{-n(\frac{1}{6}-\varepsilon)}, \qquad n \in \mathbb{N}. \tag{3.38}$$

Proof. (i): By a localization argument, we may assume that b and H are globally bounded. Let

$$A_t := \int_0^t b_r \,\mathrm{d}r \qquad \text{and} \qquad M_t := \int_0^t H_r \,\mathrm{d}W_r$$

for $t \in [0,T]$, so that $X = x_0 + A + M$, and recall that we denote the piecewise constant approximation of X along $\mathcal{P}^n_D$ by

$$X^n_t = X_T \mathbf{1}_T(t) + \sum_{k=0}^{2^n-1} X_{t^n_k} \mathbf{1}_{[t^n_k, t^n_{k+1})}(t), \qquad t \in [0,T],$$

with $t^n_k = k2^{-n}T$ for each $k = 0,1,\dots,2^n$ and $n \in \mathbb{N}$. Note that, by the uniform continuity of the sample paths of X, it is clear that X^n converges uniformly to X almost surely as $n \to \infty$.

Step 1. In this step we verify that the sample paths of X are almost surely $\frac{1}{p}$-Hölder continuous. This is a standard application of the Burkholder–Davis–Gundy inequality. Indeed, for any $q \geq 1$, using the boundedness of H, and writing $[\cdot]$ for quadratic variation, we have that

$$\mathbb{E}[|M_t - M_s|^q] = \mathbb{E}\bigg[\bigg|\int_s^t H_u \,\mathrm{d}W_u\bigg|^q\bigg] \lesssim \mathbb{E}\bigg[\bigg[\int_0^\cdot H \,\mathrm{d}W\bigg]_{s,t}^{\frac{q}{2}}\bigg] \lesssim |t-s|^{\frac{q}{2}},$$

so that $\|M_t - M_s\|_{L^q} \lesssim |t-s|^{\frac{1}{2}}$. By the Kolmogorov continuity theorem (see e.g. [74, Theorem A.10]), it follows that $\mathbb{E}[\|M\|_{\gamma\text{-Höl}}] < \infty$, where $\|\cdot\|_{\gamma\text{-Höl}}$ denotes the γ-Hölder norm, for any $\gamma \in [0, \frac{1}{2} - \frac{1}{q})$, which, taking q sufficiently large, implies that the sample paths of M are almost surely $\frac{1}{p}$-Hölder continuous. Since $A = \int_0^\cdot b_r \,\mathrm{d}r$ with the bounded integrand b, the sample paths of A are Lipschitz continuous, and thus also $\frac{1}{p}$-Hölder continuous.

Step 2. In this step we show that, almost surely, $\int_0^\cdot X^n_u \otimes \mathrm{d}X_u$ converges uniformly to the Itô integral $\int_0^\cdot X_u \otimes \mathrm{d}X_u$ as $n \to \infty$. For this purpose, we write $X^n = x_0 + A^n + M^n$, where

$$A^n_t := A_T \mathbf{1}_{\{T\}}(t) + \sum_{k=0}^{2^n-1} A_{t^n_k} \mathbf{1}_{[t^n_k, t^n_{k+1})}(t) \quad \text{and} \quad M^n_t := M_T \mathbf{1}_{\{T\}}(t) + \sum_{k=0}^{2^n-1} M_{t^n_k} \mathbf{1}_{[t^n_k, t^n_{k+1})}(t),$$

for $t \in [0,T]$. Since $X = x_0 + A + M$, we obtain

$$\begin{aligned}
&\mathbb{E}\bigg[\bigg\| \int_0^{\cdot} X_u^n \otimes \mathrm{d}X_u - \int_0^{\cdot} X_u \otimes \mathrm{d}X_u \bigg\|_\infty^2\bigg] \\
&\quad\lesssim \mathbb{E}\bigg[\bigg\| \int_0^{\cdot} (A_u^n - A_u) \otimes \mathrm{d}A_u \bigg\|_\infty^2\bigg] + \mathbb{E}\bigg[\bigg\| \int_0^{\cdot} (M_u^n - M_u) \otimes \mathrm{d}A_u \bigg\|_\infty^2\bigg] \\
&\quad\quad + \mathbb{E}\bigg[\bigg\| \int_0^{\cdot} (A_u^n - A_u) \otimes \mathrm{d}M_u \bigg\|_\infty^2\bigg] + \mathbb{E}\bigg[\bigg\| \int_0^{\cdot} (M_u^n - M_u) \otimes \mathrm{d}M_u \bigg\|_\infty^2\bigg].
\end{aligned} \tag{3.39}$$

Applying the Burkholder–Davis–Gundy inequality, the fact that $[M] = [\int_0^{\cdot} H_t \,\mathrm{d}W_t] = \int_0^{\cdot} |H_t|^2 \,\mathrm{d}t$, and the boundedness of H, we can bound

$$\begin{aligned}
&\mathbb{E}\bigg[\bigg\| \int_0^{\cdot} (M_u^n - M_u) \otimes \mathrm{d}M_u \bigg\|_\infty^2\bigg] \lesssim \mathbb{E}\bigg[\int_0^T |M_t^n - M_t|^2 \,\mathrm{d}[M]_t\bigg] \\
&\lesssim \int_0^T \mathbb{E}[|M_t^n - M_t|^2] \,\mathrm{d}t = \sum_{k=0}^{2^n-1} \int_{t_k^n}^{t_{k+1}^n} \mathbb{E}[|M_{t_k^n} - M_t|^2] \,\mathrm{d}t \lesssim \sum_{k=0}^{2^n-1} \int_{t_k^n}^{t_{k+1}^n} \mathbb{E}[|[M]_{t_k^n,t}|] \,\mathrm{d}t \\
&= \sum_{k=0}^{2^n-1} \int_{t_k^n}^{t_{k+1}^n} \mathbb{E}\bigg[\int_{t_k^n}^t |H_r|^2 \,\mathrm{d}r\bigg] \,\mathrm{d}t \lesssim \sum_{k=0}^{2^n-1} \int_{t_k^n}^{t_{k+1}^n} (t - t_k^n) \,\mathrm{d}t \le \sum_{k=0}^{2^n-1} (t_{k+1}^n - t_k^n)^2 = 2^{-n}.
\end{aligned}$$

The other terms on the right-hand side of (3.39) can be bounded similarly by 2^{-n}, up to a constant which does not depend on n, and we thus have that

$$\mathbb{E}\bigg[\bigg\| \int_0^{\cdot} X_u^n \otimes \mathrm{d}X_u - \int_0^{\cdot} X_u \otimes \mathrm{d}X_u \bigg\|_\infty^2\bigg] \lesssim 2^{-n},$$

for every $n \in \mathbb{N}$. By Markov's inequality, for any $\varepsilon \in (0,1)$, we then have that

$$\begin{aligned}
&\mathbb{P}\bigg(\bigg\| \int_0^{\cdot} X_u^n \otimes \mathrm{d}X_u - \int_0^{\cdot} X_u \otimes \mathrm{d}X_u \bigg\|_\infty \ge 2^{-\frac{n}{2}(1-\varepsilon)}\bigg) \\
&\quad \le 2^{n(1-\varepsilon)} \mathbb{E}\bigg[\bigg\| \int_0^{\cdot} X_u^n \otimes \mathrm{d}X_u - \int_0^{\cdot} X_u \otimes \mathrm{d}X_u \bigg\|_\infty^2\bigg] \lesssim 2^{n(1-\varepsilon)} 2^{-n} = 2^{-n\varepsilon}.
\end{aligned}$$

It then follows from the Borel–Cantelli lemma that, almost surely,

$$\bigg\| \int_0^{\cdot} X_u^n \otimes \mathrm{d}X_u - \int_0^{\cdot} X_u \otimes \mathrm{d}X_u \bigg\|_\infty < 2^{-\frac{n}{2}(1-\varepsilon)} \tag{3.40}$$

for all sufficiently large n, which implies the desired convergence.

Step 3. Let $\varepsilon \in (0,1)$ and $\rho = 2 + \frac{(1-\varepsilon)(p-2)}{4} \in (2,3)$. We infer from Step 1 above that the sample paths of X are almost surely $\frac{1}{\rho}$-Hölder continuous, from which it follows that

$$|X_{s,t}| \lesssim |t-s|^{\frac{1}{\rho}},$$

where the implicit multiplicative constant is a random variable which does not depend on s or t. Proceeding as in the proof of [124, Lemma 3.2], we can show, for any $0 \le k < \ell \le 2^n$, and writing $N = \ell - k = 2^n |t_\ell^n - t_k^n| T^{-1}$, that

$$\bigg| \int_{t_k^n}^{t_\ell^n} X_u^n \otimes \mathrm{d}X_u - X_{t_k^n} \otimes X_{t_k^n, t_\ell^n} \bigg| \lesssim N^{1-\frac{2}{\rho}} |t_\ell^n - t_k^n|^{\frac{2}{\rho}} \lesssim 2^{n(1-\frac{2}{\rho})} |t_\ell^n - t_k^n| \le 2^{n(\rho-2)} |t_\ell^n - t_k^n|.$$

If $2^{-n} \geq |t^n_\ell - t^n_k|^{\frac{4}{p(1-\varepsilon)}}$, then it follows that

$$\left| \int_{t^n_k}^{t^n_\ell} X^n_u \otimes \mathrm{d}X_u - X_{t^n_k} \otimes X_{t^n_k, t^n_\ell} \right| \lesssim |t^n_\ell - t^n_k|^{1 - \frac{4}{p(1-\varepsilon)}(\rho - 2)} = |t^n_\ell - t^n_k|^{\frac{2}{p}}.$$

We will now aim to obtain the same estimate in the case that $2^{-n} < |t^n_\ell - t^n_k|^{\frac{4}{p(1-\varepsilon)}}$. To this end, let $\mathbb{X}$ denote the second level component of the Itô rough path lift of X, as defined in (3.29). It follows from the Kolmogorov criterion for rough paths (see [71, Theorem 3.1]) that

$$|\mathbb{X}_{s,t}| \lesssim |t - s|^{\frac{2}{p}}, \tag{3.41}$$

where the implicit multiplicative constant is a random variable which does not depend on s or t. Using the bounds in (3.40) and (3.41), we then have, for all sufficiently large n, that

$$\begin{aligned}
&\left| \int_{t^n_k}^{t^n_\ell} X^n_u \otimes \mathrm{d}X_u - X_{t^n_k} \otimes X_{t^n_k, t^n_\ell} \right| \\
&\quad = \left| \int_{t^n_k}^{t^n_\ell} X^n_u \otimes \mathrm{d}X_u - \int_{t^n_k}^{t^n_\ell} X_u \otimes \mathrm{d}X_u + \int_{t^n_k}^{t^n_\ell} X_u \otimes \mathrm{d}X_u - X_{t^n_k} \otimes X_{t^n_k, t^n_\ell} \right| \\
&\quad \leq 2 \left\| \int_0^\cdot X^n_u \otimes \mathrm{d}X_u - \int_0^\cdot X_u \otimes \mathrm{d}X_u \right\|_\infty + |\mathbb{X}_{t^n_k, t^n_\ell}| \\
&\quad \lesssim 2^{-\frac{n}{2}(1-\varepsilon)} + |t^n_\ell - t^n_k|^{\frac{2}{p}} \\
&\quad \lesssim |t^n_\ell - t^n_k|^{\frac{2}{p}}.
\end{aligned}$$

We have thus established that

$$\left| \int_{t^n_k}^{t^n_\ell} X^n_u \otimes \mathrm{d}X_u - X_{t^n_k} \otimes X_{t^n_k, t^n_\ell} \right|^{\frac{p}{2}} \lesssim |t^n_\ell - t^n_k|$$

holds for all $0 \leq k < \ell \leq 2^n$ and all sufficiently large n. It follows that there exists a random control function $w(s,t) := c|t - s|$, for some random variable c, such that

$$\sup_{(s,t) \in \Delta_T} \frac{|X_{s,t}|^p}{w(s,t)} + \sup_{n \in \mathbb{N}} \sup_{0 \leq k < \ell \leq 2^n} \frac{|\int_{t^n_k}^{t^n_\ell} X^n_u \otimes \mathrm{d}X_u - X_{t^n_k} \otimes X_{t^n_k, t^n_\ell}|^{\frac{p}{2}}}{w(t^n_k, t^n_\ell)} \leq 1$$

holds almost surely. This means that, for almost every $\omega \in \Omega$, the sample path $X(\omega)$ satisfies Property (RIE) relative to any $p \in (2,3)$ and the sequence of dyadic partitions $(\mathcal{P}^n_D)_{n \in \mathbb{N}}$.

(ii): Since the sample paths of X are almost surely $\frac{1}{p}$-Hölder continuous (by Step 1 above), it is straightforward to see that

$$\|X^n - X\|_\infty \lesssim 2^{-\frac{n}{p}}, \qquad n \in \mathbb{N},$$

and, recalling (3.40), we have that

$$\left\| \int_0^{\cdot} X_u^n \otimes \mathrm{d}X_u - \int_0^{\cdot} X_u \otimes \mathrm{d}X_u \right\|_{\infty} \lesssim 2^{-\frac{n}{2}(1-\varepsilon)}, \qquad n \in \mathbb{N}.$$

Hence, by Theorem 3.1.2, we deduce that

$$\|Y^n - Y\|_3 \leq \|Y^n - Y\|_{p'} \lesssim 2^{-n(1-\frac{1}{q})} + 2^{-\frac{n}{p}(1-\frac{p}{p'})} + 2^{-\frac{n}{2}(1-\varepsilon)(1-\frac{p}{p'})},$$

for any $p' \in (p,3)$ and $q \in (1,2)$ such that $\frac{1}{p'} + \frac{1}{q} > 1$, which leads to (3.37). Choosing p sufficiently close to 2, p' to 3, and q to $\frac{3}{2}$, and replacing ε by 6ε, then reveals (3.38). □

3.2.3 Lévy processes

Let $L = (L_t)_{t \in [0,T]}$ be a d-dimensional Lévy process with characteristics (λ, Σ, ν). In this section, we shall work under the assumption that $\int_{|x|<1} |x|^q \nu(\mathrm{d}x) < \infty$ for some $q \in [1,2)$.

By the Lévy–Itô decomposition (see e.g. [10, Theorem 2.4.16]), there exists a Brownian motion W with covariance matrix Σ, and an independent Poisson random measure μ on $[0,T] \times (\mathbb{R}^d \setminus \{0\})$ with compensator ν, such that $L = W + \varphi$, where

$$\varphi_t = \lambda t + \int_{|x| \geq 1} x \mu(t, \mathrm{d}x) + \int_{|x|<1} x(\mu(t, \mathrm{d}x) - t\nu(\mathrm{d}x)), \qquad t \in [0,T]. \tag{3.42}$$

Since $\int_{|x|<1} |x|^q \nu(\mathrm{d}x) < \infty$, we have that $\varphi(\omega) \in D^q([0,T]; \mathbb{R}^d)$ for almost every $\omega \in \Omega$; see [10, Theorem 2.4.25] and [26, Théorème IIIb].

Let $(\mathcal{P}_D^n)_{n \in \mathbb{N}}$ be the dyadic partitions of $[0,T]$, as defined in (3.36). For each $n \in \mathbb{N}$, we also let $J^n = \{t \in (0,T] : |\Delta \varphi_t| \geq 2^{-n}\}$, where $\Delta \varphi_t = \varphi_t - \varphi_{t-}$ denotes the jump of φ at time t, and we let

$$\mathcal{P}_L^n = \mathcal{P}_D^n \cup J^n. \tag{3.43}$$

We will consider $(\mathcal{P}_L^n)_{n \in \mathbb{N}}$ as our sequence of adapted partitions, noting in particular that, for almost every $\omega \in \Omega$, $(\mathcal{P}_L^n(\omega))_{n \in \mathbb{N}}$ is a nested sequence of (finite) partitions with vanishing mesh size, and that $\{t \in (0,T] : L_{t-}(\omega) \neq L_t(\omega)\} \subseteq \cup_{n \in \mathbb{N}} \mathcal{P}_L^n(\omega)$.

Remark 3.2.7. *In order to obtain pointwise convergence of an Euler scheme, it is necessary that the jump times of the driving signal belong to the partitions used to construct the discretization, a fact which follows immediately from Proposition A.3.1, necessitating the inclusion of the jump times $(J^n)_{n \in \mathbb{N}}$ above.*

Proposition 3.2.8. *Let L be a d-dimensional Lévy process with characteristics (λ, Σ, ν), and assume that $\int_{|x|<1} |x|^q \nu(\mathrm{d}x) < \infty$ for some $q \in [1,2)$. Let $p \in (2,3)$ such that $\frac{1}{p} + \frac{1}{q} > 1$. Let Y be the solution of the SDE (3.28) driven by L, and let Y^n be the corresponding Euler approximation along $\mathcal{P}_L^n$, as defined in (3.7).*

(i) *For almost every $\omega \in \Omega$, the sample path $L(\omega)$ satisfies Property* (RIE) *relative to p and $(\mathcal{P}^n_L(\omega))_{n\in\mathbb{N}}$.*

(ii) *For any $p' \in (p,3)$ and $q' \in (q,2)$ such that $\frac{1}{p'} + \frac{1}{q'} > 1$, any $\gamma \in (0, \frac{1}{p})$, and any $\delta \in (0, 1 - \frac{q}{2})$, there exists a random variable C, which does not depend on n, such that*

$$\|Y^n - Y\|_{p'} \leq C\Big(2^{-n(1-\frac{1}{q'})} + (2^{-n(\frac{1}{p}-\gamma)} + 2^{-n(\frac{1}{p}-\frac{1}{p'})} + 2^{-n\delta(1-\frac{q}{q'})})^{1-\frac{p}{p'}}\Big), \qquad n \in \mathbb{N}.$$

To prove this statement, we need the following lemma.

Lemma 3.2.9. *Let $p \in (2,3)$, let W be a d-dimensional Brownian motion with covariance matrix Σ, and let $(\mathcal{P}^n_L)_{n\in\mathbb{N}}$ be the sequence of adapted partitions defined in* (3.43)*. For almost every $\omega \in \Omega$, the sample path $W(\omega)$ satisfies Property* (RIE) *relative to p and $(\mathcal{P}^n_L(\omega))_{n\in\mathbb{N}}$.*

Proof. We need to verify each of the conditions (i)–(iii) in Property (RIE).

(i): Since the sample paths of W are uniformly continuous on the compact interval $[0,T]$, it is straightforward to see that $W^n(\omega) \to W(\omega)$ uniformly as $n \to \infty$ for almost every $\omega \in \Omega$, where W^n denotes the piecewise constant approximation of W along $\mathcal{P}^n_L$.

(ii): It follows from the Kolmogorov continuity criterion that the sample paths of Brownian motion are almost surely $\frac{1}{p}$-Hölder continuous, and that the Hölder constant $\|W\|_{\frac{1}{p}\text{-Höl}}$ has finite moments of all orders (see e.g. [18, Theorem A.1]). Applying the Burkholder–Davis–Gundy inequality, we then have that

$$\begin{aligned}\mathbb{E}\bigg[\bigg\|\int_0^\cdot W^n_u \otimes \mathrm{d}W_u - \int_0^\cdot W_u \otimes \mathrm{d}W_u\bigg\|_\infty^2\bigg] &\lesssim \mathbb{E}\bigg[\int_0^T |W^n_t - W_t|^2 \,\mathrm{d}t\bigg] \\ &\leq \mathbb{E}\bigg[\|W\|^2_{\frac{1}{p}\text{-Höl}}\int_0^T |\mathcal{P}^n_L|^{\frac{2}{p}} \,\mathrm{d}t\bigg] \lesssim \mathbb{E}[\|W\|^2_{\frac{1}{p}\text{-Höl}}]2^{-\frac{2n}{p}}.\end{aligned}$$

Let $\gamma \in (0, \frac{1}{p})$ and $\varepsilon = 1 - \frac{2}{p} + 2\gamma \in (1 - \frac{2}{p}, 1)$. By Markov's inequality, we infer that

$$\mathbb{P}\bigg(\bigg\|\int_0^\cdot W^n_u \otimes \mathrm{d}W_u - \int_0^\cdot W_u \otimes \mathrm{d}W_u\bigg\|_\infty \geq 2^{-\frac{n}{2}(1-\varepsilon)}\bigg) \lesssim 2^{-\frac{2n}{p}+n(1-\varepsilon)} = 2^{-2n\gamma}.$$

By the Borel–Cantelli lemma, we then have that, almost surely,

$$\bigg\|\int_0^\cdot W^n_u \otimes \mathrm{d}W_u - \int_0^\cdot W_u \otimes \mathrm{d}W_u\bigg\|_\infty < 2^{-\frac{n}{2}(1-\varepsilon)} \tag{3.44}$$

for all sufficiently large n. It follows that $(\int_0^\cdot W^n_u \otimes \mathrm{d}W_u)(\omega)$ converges uniformly to $(\int_0^\cdot W_u \otimes \mathrm{d}W_u)(\omega)$ as $n \to \infty$ for almost every $\omega \in \Omega$.

(iii): Let $\rho = 2 + \frac{(1-\varepsilon)(p-2)}{4} \in (2,3)$. Since the sample paths of W are almost surely $\frac{1}{\rho}$-Hölder continuous, it follows that

$$|W_{s,t}|^\rho \lesssim |t-s|,$$

where the implicit multiplicative constant is a random variable which does not depend on s or t. Proceeding as in the proof of [124, Lemma 3.2], we can show, for any $0 \le k < \ell$, and writing $N = \ell - k$, we can show that

$$\left| \int_{t_k^n}^{t_\ell^n} W_u^n \otimes \mathrm{d}W_u - W_{t_k^n} \otimes W_{t_k^n, t_\ell^n} \right| \lesssim N^{1-\frac{2}{\rho}} |t_\ell^n - t_k^n|^{\frac{2}{\rho}},$$

where $\{0 = t_0^n < t_1^n < \cdots\}$ are the partition points of $\mathcal{P}_L^n(\omega)$ for some (here fixed) $\omega \in \Omega$. Using $|\cdot|$ here to denote the cardinality of a set, we note that the number N can be bounded by

$$\begin{aligned} N &\le |\mathcal{P}_D^n(\omega) \cap (t_k^n, t_\ell^n]| + |J^n(\omega) \cap (t_k^n, t_\ell^n]| \le 2^n T^{-1} |t_\ell^n - t_k^n| + 2^{nq} \sum_{t \in J^n(\omega) \cap (t_k^n, t_\ell^n]} |\Delta \varphi_t(\omega)|^q \\ &\lesssim 2^n |t_\ell^n - t_k^n| + 2^{nq} \|\varphi(\omega)\|_{q,[t_k^n, t_\ell^n]}^q \le 2^{n\rho} c(t_k^n, t_\ell^n), \end{aligned}$$

where c is the control function defined by $c(s,t) := |t-s| + \|\varphi(\omega)\|_{q,[s,t]}^q$ for $(s,t) \in \Delta_T$. If $2^{-n} \ge c(t_k^n, t_\ell^n)^{\frac{4}{p(1-\varepsilon)}}$, this implies that

$$\left| \int_{t_k^n}^{t_\ell^n} W_u^n \otimes \mathrm{d}W_u - W_{t_k^n} \otimes W_{t_k^n, t_\ell^n} \right| \lesssim 2^{n(\rho-2)} c(t_k^n, t_\ell^n) \le c(t_k^n, t_\ell^n)^{1 - \frac{4}{p(1-\varepsilon)}(\rho-2)} = c(t_k^n, t_\ell^n)^{\frac{2}{p}}.$$

In the case that $2^{-n} < c(t_k^n, t_\ell^n)^{\frac{4}{p(1-\varepsilon)}}$, we can follow the same argument as in Step 3 of the proof of part (i) of Proposition 3.2.6 (using in particular the bound in (3.44)) to obtain again that

$$\left| \int_{t_k^n}^{t_\ell^n} W_u^n \otimes \mathrm{d}W_u - W_{t_k^n} \otimes W_{t_k^n, t_\ell^n} \right| \lesssim c(t_k^n, t_\ell^n)^{\frac{2}{p}},$$

where, as usual, the implicit multiplicative constant depends on ω, but not on n.

It follows that there exists a random control function w such that

$$\sup_{(s,t) \in \Delta_T} \frac{|W_{s,t}|^p}{w(s,t)} + \sup_{n \in \mathbb{N}} \sup_{0 \le k < \ell} \frac{|\int_{t_k^n}^{t_\ell^n} W_u^n \otimes \mathrm{d}W_u - W_{t_k^n} \otimes W_{t_k^n, t_\ell^n}|^{\frac{p}{2}}}{w(s,t)} \le 1$$

holds almost surely. □

Proof of Proposition 3.2.8. Let W be a Brownian motion with covariance matrix Σ, and let φ be the process defined in (3.42), so that $L = W + \varphi$. As usual, we let L^n, W^n and φ^n denote the piecewise constant approximations of L, W and φ respectively, along the adapted partition $\mathcal{P}_L^n$.

Recalling (3.42), we see that we can write $\varphi = \eta + \xi$, where

$$\eta_t := \lambda t + \int_{|x| \ge 2^{-n}} x \mu(t, \mathrm{d}x) - t \int_{2^{-n} \le |x| < 1} x \nu(\mathrm{d}x) \tag{3.45}$$

and

$$\xi_t := \int_{|x|<2^{-n}} x(\mu(t, \mathrm{d}x) - t\nu(\mathrm{d}x)).$$

Let η^n and ξ^n denote the piecewise constant approximations of η and ξ along $\mathcal{P}_L^n$. Recalling how the adapted partition $\mathcal{P}_L^n$ was defined in (3.43), we note that, when estimating the difference $\eta^n - \eta$, we may ignore all jumps of size greater than 2^{-n}, and may thus ignore the first integral on the right-hand side of (3.45). We then have that

$$\begin{aligned}\|\eta^n - \eta\|_\infty &\leq 2^{-n}T|\lambda| + 2^{-n}T\int_{2^{-n}\leq|x|<1}|x|\nu(\mathrm{d}x) \\ &\leq 2^{-n}T|\lambda| + 2^{-n(2-q)}T\int_{2^{-n}\leq|x|<1}|x|^q\nu(\mathrm{d}x) \lesssim 2^{-n(2-q)}.\end{aligned} \tag{3.46}$$

Writing $\langle\cdot\rangle$ for the predictable quadratic variation, we have (see e.g. [96, Chapter 2, Theorem 1.33]) that

$$\mathbb{E}[\langle\xi\rangle_T] \leq T\int_{|x|<2^{-n}}|x|^2\nu(\mathrm{d}x) \leq 2^{-n(2-q)}T\int_{|x|<2^{-n}}|x|^q\nu(\mathrm{d}x).$$

Since this quantity is finite, the process ξ is a square integrable martingale, and in particular $\mathbb{E}[[\xi]_T] = \mathbb{E}[\langle\xi\rangle_T]$, where $[\cdot]$ denotes the usual quadratic variation. By the Burkholder–Davis–Gundy inequality, we then have that

$$\mathbb{E}[\|\xi\|_\infty^2] \lesssim \mathbb{E}[[\xi]_T] = \mathbb{E}[\langle\xi\rangle_T] \lesssim 2^{-n(2-q)}. \tag{3.47}$$

Note that, for any $a > 0$, if $\|\xi\|_\infty < \frac{a}{2}$, then $\|\xi^n - \xi\|_\infty < a$. It follows that, for any $\delta \in (0, 1 - \frac{q}{2})$,

$$\mathbb{P}(\|\xi^n - \xi\|_\infty \geq 2^{-n\delta}) \leq \mathbb{P}(\|\xi\|_\infty \geq 2^{-1-n\delta}).$$

By Markov's inequality and the bound in (3.47), we see that

$$\mathbb{P}(\|\xi^n - \xi\|_\infty \geq 2^{-n\delta}) \lesssim 2^{2-n(2-q-2\delta)},$$

and the Borel–Cantelli lemma then implies that, almost surely,

$$\|\xi^n - \xi\|_\infty \lesssim 2^{-n\delta}, \tag{3.48}$$

where the implicit multiplicative constant is a random variable which does not depend on n. It follows from (3.46) and (3.48) that

$$\|\varphi^n - \varphi\|_\infty \lesssim 2^{-n\delta}. \tag{3.49}$$

Let $p' \in (p, 3)$ and $q' \in (q, 2)$ such that $\frac{1}{p'} + \frac{1}{q'} > 1$. Using interpolation, the fact that $\sup_{n\in\mathbb{N}}\|\varphi^n\|_q \leq \|\varphi\|_q$, and the bound in (3.49), we have that, almost surely,

$$\|\varphi^n - \varphi\|_{q'} \leq \|\varphi^n - \varphi\|_\infty^{1-\frac{q}{q'}}\|\varphi^n - \varphi\|_q^{\frac{q}{q'}} \lesssim \|\varphi^n - \varphi\|_\infty^{1-\frac{q}{q'}} \lesssim 2^{-n\delta(1-\frac{q}{q'})}. \tag{3.50}$$

We also have from Lemma 3.2.9 that, for almost every $\omega \in \Omega$, the sample path $W(\omega)$ satisfies Property (RIE) relative to p and $(\mathcal{P}^n_L(\omega))_{n\in\mathbb{N}}$. Thus, by Proposition 3.1.10, for almost every $\omega \in \Omega$, the sample path $L(\omega)$ satisfies Property (RIE) relative to p and $(\mathcal{P}^n(\omega))_{n\in\mathbb{N}}$, which establishes part (i).

Since the sample paths of W are almost surely $\frac{1}{p}$-Hölder continuous, it is straightforward to see that

$$\|W^n - W\|_\infty \lesssim 2^{-\frac{n}{p}},$$

where the implicit multiplicative constant depends on the (random) Hölder constant of the path. Since $L = W + \varphi$, we have that

$$\|L^n - L\|_\infty \leq \|W^n - W\|_\infty + \|\varphi^n - \varphi\|_\infty \lesssim 2^{-\frac{n}{p}} + 2^{-n\delta}.$$

We recall from (3.44) that

$$\left\| \int_0^\cdot W^n_u \otimes \mathrm{d}W_u - \int_0^\cdot W_u \otimes \mathrm{d}W_u \right\|_\infty \lesssim 2^{-\frac{n}{2}(1-\varepsilon)} = 2^{-n(\frac{1}{p}-\gamma)}$$

for any $\gamma \in (0, \frac{1}{p})$. We obtained a bound for $\|\varphi^n - \varphi\|_{q'}$ in (3.50), and an analogous argument also shows that

$$\|W^n - W\|_{p'} \leq \|W^n - W\|_\infty^{1-\frac{p}{p'}} \|W^n - W\|_p^{\frac{p}{p'}} \lesssim \|W^n - W\|_\infty^{1-\frac{p}{p'}} \lesssim 2^{-n(\frac{1}{p}-\frac{1}{p'})}.$$

Using the standard estimate for Young integrals (see e.g. [75, Proposition 2.4]), similarly to the proof of Proposition 3.1.10, we then obtain

$$\begin{aligned}
&\left\| \int_0^\cdot L^n_u \otimes \mathrm{d}L_u - \int_0^\cdot L_u \otimes \mathrm{d}L_u \right\|_\infty \\
&\lesssim \left\| \int_0^\cdot W^n_u \otimes \mathrm{d}W_u - \int_0^\cdot W_u \otimes \mathrm{d}W_u \right\|_\infty + \|W^n - W\|_{p'} \|\varphi\|_q + \|\varphi^n - \varphi\|_{q'} (\|W\|_p + \|\varphi\|_q) \\
&\lesssim 2^{-n(\frac{1}{p}-\gamma)} + 2^{-n(\frac{1}{p}-\frac{1}{p'})} + 2^{-n\delta(1-\frac{q}{q'})}.
\end{aligned}$$

Hence, by Theorem 3.1.2, we establish the estimate in part (ii). □

3.2.4 Càdlàg semimartingales

In this section, we consider the case when X is a general càdlàg semimartingale. As noted in Remark 3.2.7, to hope for pointwise convergence of the Euler scheme, we need to ensure that the sequence of partitions exhausts all the jump times of X. With this in mind, for each $n \in \mathbb{N}$, we introduce the stopping times $(\tau^n_k)_{k\in\mathbb{N}\cup\{0\}}$, such that $\tau^n_0 = 0$, and

$$\tau^n_k = \inf\{t > \tau^n_{k-1} : |t - \tau^n_{k-1}| + |X_t - X_{\tau^n_{k-1}}| \geq 2^{-n}\} \wedge T, \qquad k \in \mathbb{N}. \tag{3.51}$$

We then define a sequence of adapted partitions $(\mathcal{P}_X^n)_{n\in\mathbb{N}}$ by

$$\mathcal{P}_X^n = \{\tau_k^n : k \in \mathbb{N} \cup \{0\}\}.$$

Note that, for almost every $\omega \in \Omega$, $(\mathcal{P}_X^n(\omega))_{n\in\mathbb{N}}$ is a sequence of (finite) partitions with vanishing mesh size. The next result verifies that X satisfies Property (RIE) relative to any $p \in (2,3)$ and $(\mathcal{P}_X^n)_{n\in\mathbb{N}}$, and establishes the rate of convergence of the associated Euler scheme.

Proposition 3.2.10. *Let $p \in (2,3)$, and let X be a d-dimensional càdlàg semimartingale. Let Y be the solution of the SDE* (3.28) *driven by X, and let Y^n be the corresponding Euler approximation along $\mathcal{P}_X^n$, as defined in* (3.7).

(i) For almost every $\omega \in \Omega$, the sample path $X(\omega)$ satisfies Property (RIE) *relative to p and $(\mathcal{P}_X^n(\omega))_{n\in\mathbb{N}}$.*

(ii) For any $p' \in (p,3)$ and $q \in (1,2)$ such that $\frac{1}{p'} + \frac{1}{q} > 1$, and any $\varepsilon \in (0,1)$, there exists a random variable C, which does not depend on n, such that

$$\|Y^n - Y\|_{p'} \leq C(2^{-n(1-\frac{1}{q})} + 2^{-n(1-\varepsilon)(1-\frac{p}{p'})}), \qquad n \in \mathbb{N}, \tag{3.52}$$

and

$$\|Y^n - Y\|_3 \leq C2^{-n(\frac{1}{3}-\varepsilon)}, \qquad n \in \mathbb{N}. \tag{3.53}$$

Proof. (i): The proof is just a slight modification of the proof of [7, Proposition 4.1], and is therefore omitted here for brevity. It is actually slightly easier, as here we do not require the sequence of partitions to be nested, and the sequence of stopping times in (3.51) is constructed to ensure that the mesh size vanishes, even if X exhibits intervals of constancy.

(ii): By the definition of the partition $\mathcal{P}_X^n$, it is clear that

$$\|X^n - X\|_\infty \leq 2^{-n}.$$

By an application of the Burkholder–Davis–Gundy inequality and the Borel–Cantelli lemma, as in the proof of [124, Proposition 3.4], one can show that

$$\left\| \int_0^\cdot X_{u-}^n \otimes \mathrm{d}X_u - \int_0^\cdot X_{u-} \otimes \mathrm{d}X_u \right\|_\infty \lesssim 2^{-n(1-\varepsilon)}, \qquad n \in \mathbb{N},$$

where the implicit multiplicative constant is a random variable which does not depend on n.

It thus follows from Theorem 3.1.2 that

$$\|Y^n - Y\|_3 \leq \|Y^n - Y\|_{p'} \lesssim 2^{-n(1-\frac{1}{q})} + 2^{-n(1-\frac{p}{p'})} + 2^{-n(1-\varepsilon)(1-\frac{p}{p'})},$$

which leads to (3.52). Choosing p sufficiently close to 2, p' to 3, and q to $\frac{3}{2}$, and replacing ε by 3ε, then reveals (3.53). □

3.3 Application to differential equations driven by non-semimartingales

While in the previous section we considered stochastic differential equations driven by various classes of semimartingales, like the general theory of rough paths, the deterministic theory developed in Section 3.1 is not limited to the semimartingale framework. In this section we investigate Property (RIE) in the context of "mixed" and "rough" stochastic differential equations. The main insight is again that the random driving signals of these equations do, indeed, satisfy Property (RIE) and, thus the pathwise convergence results regarding the Euler scheme, as presented in Theorem 3.1.2 and Proposition 3.1.11, are applicable.

Remark 3.3.1. *Further examples of stochastic processes which fulfill Property* (RIE) *almost surely include* p-semimartingales *(also known as* Young semimartingales*) in the sense of Norvaiša [142], as well as* typical price paths *in the sense of Vovk, relative to suitable sequences of adapted partitions. The pathwise convergence of the Euler scheme is thus immediately applicable to differential equations driven by such p-semimartingales [114] and typical price paths [17].*

3.3.1 Mixed stochastic differential equations

Differential equations driven by both a Brownian motion as well as a fractional Brownian motion with Hurst parameter $H > \frac{1}{2}$ are classical objects in stochastic analysis; see e.g. [162, 135]. More precisely, a "mixed" stochastic differential equation (mixed SDE) is given by

$$Y_t = y_0 + \int_0^t b(s, Y_s)\,\mathrm{d}s + \int_0^t \sigma_1(s, Y_s)\,\mathrm{d}W_s + \int_0^t \sigma_2(s, Y_s)\,\mathrm{d}W_s^H, \qquad t \in [0, T], \tag{3.54}$$

where $b \in C_b^2(\mathbb{R}^{k+1}; \mathcal{L}(\mathbb{R}; \mathbb{R}^k))$, $\sigma_1 \in C_b^3(\mathbb{R}^{k+1}; \mathcal{L}(\mathbb{R}^{d_1}; \mathbb{R}^k))$, $\sigma_2 \in C_b^3(\mathbb{R}^{k+1}; \mathcal{L}(\mathbb{R}^{d_2}; \mathbb{R}^k))$ and $y_0 \in \mathbb{R}^k$. Here, W is a d_1-dimensional standard Brownian motion, and W^H is a d_2-dimensional fractional Brownian motion with Hurst parameter $H > \frac{1}{2}$, which are independent and both defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ satisfying the usual conditions.

The mixed SDE (3.54) lies outside the semimartingale framework, but there are various ways to provide a rigorous meaning to its solution. Here we consider the mixed SDE (3.54) as a random rough differential equation, driven by the Itô rough path lift of (W, W^H), the existence of which follows from Lemma 3.3.2 below. In particular, it then follows from Theorem 3.1.1 that there exists a unique solution Y to (3.54).

Lemma 3.3.2. *Let W be a standard Brownian motion, and let W^H be a fractional Brownian motion with Hurst parameter $H \in (\frac{1}{2}, 1)$. Let $p \in (2, 3)$ such that $\frac{1}{p} + H > 1$, and let*

$\mathcal{P}^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$, be a sequence of equidistant partitions of the interval $[0,T]$, so that, for each $n \in \mathbb{N}$, there exists some $\pi_n > 0$ such that $t_{i+1}^n - t_i^n = \pi_n$ for each $0 \leq i < N_n$. If $\pi_n^{2-\frac{4}{p}} \log(n) \to 0$ as $n \to \infty$, then, for almost every $\omega \in \Omega$, the sample path $(W(\omega), W^H(\omega))$ satisfies Property (RIE) *relative to p and $(\mathcal{P}^n)_{n\in\mathbb{N}}$.*

Proof. We first note that the process $(W, 0)$ satisfies the hypotheses of Theorem 3.2.2, and thus that almost all of its sample paths satisfy Property (RIE) relative to p and $(\mathcal{P}^n)_{n\in\mathbb{N}}$. Let $\frac{1}{H} < q < q' < 2$ such that $\frac{1}{p} + \frac{1}{q'} > 1$. Since $\frac{1}{q} < H$, it is well-known that the sample paths of $(0, W^H)$ are almost surely $\frac{1}{q}$-Hölder continuous, and hence that $\|W^H\|_q < \infty$. Writing $W^{H,n}$ for the usual piecewise constant approximation of W^H along $\mathcal{P}^n$, we have by interpolation that

$$\|W^{H,n} - W^H\|_{q'} \leq \|W^{H,n} - W^H\|_\infty^{1-\frac{q}{q'}} \|W^{H,n} - W^H\|_q^{\frac{q}{q'}} \lesssim \|W^{H,n} - W^H\|_\infty^{1-\frac{q}{q'}} \longrightarrow 0$$

as $n \to \infty$. The result then follows by applying Proposition 3.1.10 to $(W, 0) + (0, W^H)$. □

3.3.2 Rough stochastic differential equations

Rough stochastic differential equations (rough SDEs) are differential equations driven by both a rough path and a semimartingale. These equations first appeared in the context of robust stochastic filtering—see [43, 58]—and were recently studied in a general form in [72]. In this section we will adapt the setting of [58], which allows to treat Hölder continuous rough paths and Brownian motion as driving signals.

We let $\eta\colon [0,T] \to \mathbb{R}^d$ be a deterministic path which is $\frac{1}{p}$-Hölder continuous for some $p \in (2,3)$, and which satisfies Property (RIE) relative to p and the dyadic partitions $(\mathcal{P}_D^n)_{n\in\mathbb{N}}$, as defined in (3.36). We write $\boldsymbol{\eta} = (\boldsymbol{\eta}^1, \boldsymbol{\eta}^2)$ for the canonical rough path lift of η, with $\boldsymbol{\eta}^2$ defined as in (3.9), so that $\boldsymbol{\eta}_{s,t}^2 = \int_s^t \eta_{s,u} \otimes \mathrm{d}\eta_u$ for each $(s,t) \in \Delta_T$. We also let W be an $\mathbb{R}^e$-valued Brownian motion. For vector fields $a \in C_b^2$ and $b, c \in C_b^3$, and an initial value $y_0 \in \mathbb{R}^k$, we then consider the rough SDE

$$Y_t = y_0 + \int_0^t a(Y_s)\,\mathrm{d}s + \int_0^t b(Y_s)\,\mathrm{d}\boldsymbol{\eta}_s + \int_0^t c(Y_s)\,\mathrm{d}W_s, \qquad t \in [0,T]. \tag{3.55}$$

To give a rigorous meaning to the rough SDE (3.55), following the method introduced in [58], we need to construct a suitable joint rough path lift $\boldsymbol{\Lambda}(\omega)$ above the $\mathbb{R}^{d+e}$-valued path $(\eta, W(\omega))$ for almost every $\omega \in \Omega$. Indeed, the (pathwise) unique solution to the random RDE

$$Y_t = y_0 + \int_0^t a(Y_s)\,\mathrm{d}s + \int_0^t (b,c)(Y_s)\,\mathrm{d}\boldsymbol{\Lambda}_s, \qquad t \in [0,T],$$

is then defined to be the solution to the rough SDE (3.55).

To construct the Itô rough path lift of (η, W), we need the existence of the quadratic covariation of η and W along the dyadic partitions. More precisely, writing $\mathcal{P}_D^n = \{0 = t_0^n < t_1^n < \cdots < t_{2^n}^n = T\}$ with $t_k^n = k2^{-n}T$, we need to establish that, for almost every $\omega \in \Omega$, the limit

$$\langle \eta, W(\omega)\rangle_t := \lim_{n\to\infty} \sum_{k=0}^{2^n-1} \eta_{t_k^n\wedge t, t_{k+1}^n\wedge t} \otimes W_{t_k^n\wedge t, t_{k+1}^n\wedge t}(\omega) \tag{3.56}$$

exists and holds uniformly for $t \in [0, T]$.

Lemma 3.3.3. *Let $\alpha \in (0, 1]$, let $\eta\colon [0, T] \to \mathbb{R}$ be an α-Hölder continuous deterministic path, and let W be a one-dimensional Brownian motion. Then, for almost every $\omega \in \Omega$, the quadratic covariation of η and $W(\omega)$ along the dyadic partitions, in the sense of* (3.56)*, exists, and satisfies $\langle \eta, W(\omega)\rangle_t = 0$ for all $t \in [0, T]$.*

Proof. We consider the discrete-time martingale given by $t \mapsto \sum_{k\colon t_{k+1}^n\le t} \eta_{t_k^n,t_{k+1}^n} W_{t_k^n,t_{k+1}^n}$ for $t \in \mathcal{P}_D^n$, for some fixed $n \in \mathbb{N}$. By the Burkholder–Davis–Gundy inequality, we have that

$$\begin{aligned}\mathbb{E}\bigg[\bigg\|\sum_{k\colon t_{k+1}^n\le\cdot} \eta_{t_k^n,t_{k+1}^n} W_{t_k^n,t_{k+1}^n}\bigg\|_\infty^2\bigg] &\lesssim \mathbb{E}\bigg[\sum_{k=0}^{2^n-1} (\eta_{t_k^n,t_{k+1}^n} W_{t_k^n,t_{k+1}^n})^2\bigg] = \sum_{k=0}^{2^n-1} (\eta_{t_k^n,t_{k+1}^n})^2 (t_{k+1}^n - t_k^n)\\ &\lesssim \sum_{k=0}^{2^n-1} (t_{k+1}^n - t_k^n)^{1+2\alpha} \lesssim (2^{-n}T)^{2\alpha} \sum_{k=0}^{2^n-1} (t_{k+1}^n - t_k^n)\\ &\lesssim 2^{-2n\alpha}.\end{aligned}$$

For any $\varepsilon \in (0, 1)$, we then have, by Markov's inequality, that

$$\mathbb{P}\bigg(\bigg\|\sum_{k\colon t_{k+1}^n\le\cdot} \eta_{t_k^n,t_{k+1}^n} W_{t_k^n,t_{k+1}^n}\bigg\|_\infty \ge 2^{-n\alpha(1-\varepsilon)}\bigg) \lesssim 2^{-2n\alpha\varepsilon},$$

and the Borel–Cantelli lemma then implies that

$$\bigg\|\sum_{k\colon t_{k+1}^n\le\cdot} \eta_{t_k^n,t_{k+1}^n} W_{t_k^n,t_{k+1}^n}\bigg\|_\infty \lesssim 2^{-n\alpha(1-\varepsilon)},$$

where the implicit multiplicative constant is a random variable which does not depend on n.

For a given $t \in [0, T]$ and $n \in \mathbb{N}$, let k_0 be such that $t \in [t_{k_0}^n, t_{k_0+1}^n]$. Since η is α-Hölder continuous, and the sample paths of W are almost surely β-Hölder continuous for any $\beta \in (0, \frac{1}{2})$, we have that

$$|\eta_{t_{k_0}^n,t} W_{t_{k_0}^n,t}| \lesssim (t - t_{k_0}^n)^{\alpha+\beta} \lesssim 2^{-n(\alpha+\beta)}.$$

We thus have the bound

$$\Big|\sum_{k=0}^{2^n-1}\eta_{t_k^n\wedge t,t_{k+1}^n\wedge t}W_{t_k^n\wedge t,t_{k+1}^n\wedge t}\Big| \leq \Big|\sum_{k:t_{k+1}^n\leq t}\eta_{t_k^n,t_{k+1}^n}W_{t_k^n,t_{k+1}^n}\Big| + |\eta_{t_{k_0}^n,t}W_{t_{k_0}^n,t}|$$
$$\lesssim 2^{-n\alpha(1-\varepsilon)} + 2^{-n(\alpha+\beta)},$$

where the implicit multiplicative constant is a random variable which does not depend on t or n. It follows that, almost surely,

$$\sum_{k=0}^{2^n-1}\eta_{t_k^n\wedge t,t_{k+1}^n\wedge t}W_{t_k^n\wedge t,t_{k+1}^n\wedge t} \longrightarrow 0 \qquad \text{as} \quad n\longrightarrow\infty,$$

uniformly for $t\in[0,T]$. $\square$

It is shown in [58, Theorem 1], with integrals defined in the Stratonovich sense, that an analogous object to the process $\mathbf{\Lambda}$ described in (3.57) below provides a geometric rough path lift of (η, W). In the next theorem we establish that $\mathbf{\Lambda}$ is the Itô rough path lift of (η, W), and moreover that it may be obtained as the canonical lift via Property (RIE), thus making our convergence analysis of the Euler scheme applicable to the rough SDE (3.55).

Theorem 3.3.4. *Let $p\in(2,3)$. Let η be a $\frac{1}{p}$-Hölder continuous $\mathbb{R}^d$-valued path which satisfies Property* (RIE) *relative to p and the sequence of dyadic partitions $(\mathcal{P}_D^n)_{n\in\mathbb{N}}$, and write $\boldsymbol{\eta}=(\boldsymbol{\eta}^1,\boldsymbol{\eta}^2)$ for the canonical rough path lift of η, so that $\boldsymbol{\eta}^1=\eta$, and $\boldsymbol{\eta}^2_{s,t}=\int_s^t\eta_{s,u}\otimes \mathrm{d}\eta_u$, defined as in* (3.9)*, for every $(s,t)\in\Delta_T$. Let W be an $\mathbb{R}^e$-valued Brownian motion, and write $\mathbf{W}=(W,\mathbb{W})$ for the Itô rough path lift of W, so that $\mathbb{W}_{s,t}=\int_s^t W_{s,u}\otimes \mathrm{d}W_u$, defined as an Itô integral, for every $(s,t)\in\Delta_T$.*

For any $p'\in(p,3)$ and almost every $\omega\in\Omega$, the $\mathbb{R}^{d+e}$-valued path $(\eta, W(\omega))$ satisfies Property (RIE) *relative to p' and $(\mathcal{P}_D^n)_{n\in\mathbb{N}}$.*

Moreover, for almost every $\omega\in\Omega$, the canonical rough path lift $\mathbf{\Lambda}(\omega)=(\mathbf{\Lambda}^1(\omega),\mathbf{\Lambda}^2(\omega))\in\mathbb{R}^{d+e}\oplus\mathbb{R}^{(d+e)\times(d+e)}$ of $(\eta, W(\omega))$ (constructed via Property (RIE) *as in* (3.9)*) is given by $\mathbf{\Lambda}^1(\omega)=(\eta,W(\omega))$, and*

$$\mathbf{\Lambda}^2_{s,t}=\begin{pmatrix}\boldsymbol{\eta}^2_{s,t} & \int_s^t\eta_{s,u}\otimes \mathrm{d}W_u\\ W_{s,t}\otimes\eta_{s,t}-(\int_s^t\eta_{s,u}\otimes \mathrm{d}W_u)^\top & \mathbb{W}_{s,t}\end{pmatrix} \tag{3.57}$$

for every $(s,t)\in\Delta_T$, where $\int_s^t\eta_{s,u}\otimes\mathrm{d}W_u$ is defined as an Itô integral, and $(\cdot)^\top$ denotes matrix transposition.

Proof. Let $p'\in(p,3)$. It follows from the Kolmogorov criterion for rough paths (see [71, Theorem 3.1]) that, for almost every $\omega\in\Omega$,

$$\Big|\Big(\int_s^t\eta_{s,u}\otimes\mathrm{d}W_u\Big)(\omega)\Big|\lesssim|t-s|^{\frac{2}{p'}} \qquad \text{for all} \quad (s,t)\in\Delta_T, \tag{3.58}$$

and moreover that $\mathbf{\Lambda}(\omega) = (\mathbf{\Lambda}^1(\omega), \mathbf{\Lambda}^2(\omega))$ is a $\frac{1}{p'}$-Hölder continuous rough path. We will show that $(\eta, W(\omega))$ satisfies Property (RIE), and that the associated canonical rough path is indeed given by $\mathbf{\Lambda}(\omega)$.

Step 1. As usual, we let η^n and W^n denote the piecewise constant approximations of η and W respectively, along $\mathcal{P}^n_D$. By assumption, η satisfies Property (RIE) relative to p and $(\mathcal{P}^n_D)_{n\in\mathbb{N}}$. By Proposition 3.2.2 (or Proposition 3.2.6), for almost every $\omega \in \Omega$, the sample path $W(\omega)$ also satisfies Property (RIE) relative to p and $(\mathcal{P}^n_D)_{n\in\mathbb{N}}$.

It follows from the first condition in Property (RIE) for η and $W(\omega)$ that, for almost every $\omega \in \Omega$,

$$(\eta^n, W^n(\omega)) \longrightarrow (\eta, W(\omega)) \qquad \text{uniformly as} \quad n \longrightarrow \infty,$$

so that this condition also holds for the pair $(\eta, W(\omega))$. Moreover, it follows from the second condition in Property (RIE) that $\int_0^\cdot \eta^n_u \otimes \mathrm{d}\eta_u$ converges uniformly to $\int_0^\cdot \eta_u \otimes \mathrm{d}\eta_u$, and, for almost every $\omega \in \Omega$, that $(\int_0^\cdot W^n_u \otimes \mathrm{d}W_u)(\omega)$ converges uniformly to $(\int_0^\cdot W_u \otimes \mathrm{d}W_u)(\omega)$.

By the Burkholder–Davis–Gundy inequality, and the observation that $\|\eta^n - \eta\|_\infty \lesssim 2^{-\frac{n}{p}}$, we have that

$$\mathbb{E}\bigg[\bigg\|\int_0^\cdot \eta^n_u \otimes \mathrm{d}W_u - \int_0^\cdot \eta_u \otimes \mathrm{d}W_u\bigg\|_\infty^2\bigg] \lesssim \mathbb{E}\bigg[\int_0^T |\eta^n_u - \eta_u|^2 \,\mathrm{d}u\bigg] \lesssim 2^{-\frac{2n}{p}}.$$

For any $\varepsilon \in (1 - \frac{2}{p}, 1)$, it then follows from Markov's inequality that

$$\mathbb{P}\bigg(\bigg\|\int_0^\cdot \eta^n_u \otimes \mathrm{d}W_u - \int_0^\cdot \eta_u \otimes \mathrm{d}W_u\bigg\|_\infty \geq 2^{-\frac{n}{2}(1-\varepsilon)}\bigg) \lesssim 2^{n(1-\frac{2}{p}-\varepsilon)}.$$

The Borel–Cantelli lemma then implies that, for almost every $\omega \in \Omega$,

$$\bigg\|\bigg(\int_0^\cdot \eta^n_u \otimes \mathrm{d}W_u - \int_0^\cdot \eta_u \otimes \mathrm{d}W_u\bigg)(\omega)\bigg\|_\infty \lesssim 2^{-\frac{n}{2}(1-\varepsilon)} \tag{3.59}$$

for all $n \in \mathbb{N}$, and in particular that $(\int_0^\cdot \eta^n_u \otimes \mathrm{d}W_u)(\omega)$ converges uniformly to $(\int_0^\cdot \eta_u \otimes \mathrm{d}W_u)(\omega)$ as $n \to \infty$.

Let us write $\mathcal{P}^n_D = \{0 = t^n_0 < t^n_1 < \cdots < t^n_{2^n} = T\}$ for $n \in \mathbb{N}$, where $t^n_k = k2^{-n}T$. It is straightforward to verify that, for any $t \in [0, T]$,

$$W_t \otimes \eta_t = \int_0^t W^n_u \otimes \mathrm{d}\eta_u + \bigg(\int_0^t \eta^n_u \otimes \mathrm{d}W_u\bigg)^\top + \langle W, \eta \rangle^n_t,$$

where, by Lemma 3.3.3, the discrete quadratic variation $\langle W, \eta \rangle^n_t := \sum_{k=0}^{2^n-1} W_{t^n_k \wedge t, t^n_{k+1} \wedge t} \otimes \eta_{t^n_k \wedge t, t^n_{k+1} \wedge t}$ almost surely converges uniformly to $\langle W, \eta \rangle_t = 0$ as $n \to \infty$. We then see that, for almost every $\omega \in \Omega$,

$$\int_0^t W^n_u(\omega) \otimes \mathrm{d}\eta_u \longrightarrow W_t(\omega) \otimes \eta_t - \bigg(\int_0^t \eta_u \otimes \mathrm{d}W_u\bigg)^\top(\omega)$$

as $n \to \infty$, uniformly in $t \in [0,T]$. We have thus established that, for almost every $\omega \in \Omega$, the path $(\eta, W(\omega))$ also satisfies the second condition of Property (RIE), and moreover that the resulting canonical rough path is indeed given by (3.57).

Step 2. It remains to show that $(\eta, W(\omega))$ satisfies the third condition of Property (RIE) relative to p' and $(\mathcal{P}^n_D)_{n\in\mathbb{N}}$.

Since η satisfies Property (RIE) relative to p and $(\mathcal{P}^n_D)_{n\in\mathbb{N}}$, there exists a control function w_η such that

$$\sup_{(s,t)\in\Delta_T} \frac{|\eta_{s,t}|^p}{w_\eta(s,t)} + \sup_{n\in\mathbb{N}} \sup_{0\leq k<\ell\leq 2^n} \frac{|\int_{t^n_k}^{t^n_\ell} \eta^n_u \otimes \mathrm{d}\eta_u - \eta_{t^n_k} \otimes \eta_{t^n_k,t^n_\ell}|^{\frac{p}{2}}}{w_\eta(t^n_k,t^n_\ell)} \leq 1, \tag{3.60}$$

which implies that the same inequality also holds with p replaced by p' (possibly with a different control function, but without loss of generality we may assume that w_η remains valid for p'). Similarly, since for almost every $\omega \in \Omega$ the sample path $W(\omega)$ satisfies Property (RIE) relative to p (and therefore also to p') and $(\mathcal{P}^n_D)_{n\in\mathbb{N}}$, there exists a control function c such that

$$\sup_{(s,t)\in\Delta_T} \frac{|W_{s,t}(\omega)|^{p'}}{c(s,t)} + \sup_{n\in\mathbb{N}} \sup_{0\leq k<\ell\leq 2^n} \frac{|(\int_{t^n_k}^{t^n_\ell} W^n_u \otimes \mathrm{d}W_u - W_{t^n_k} \otimes W_{t^n_k,t^n_\ell})(\omega)|^{\frac{p'}{2}}}{c(t^n_k,t^n_\ell)} \leq 1. \tag{3.61}$$

Step 3. Let $\beta \in (0,\frac{1}{2})$. Since η is $\frac{1}{p}$-Hölder continuous, and the sample paths of W are almost surely β-Hölder continuous, we have that

$$|\eta_{t^n_{i-1}} \otimes W_{t^n_{i-1},t^n_i} + \eta_{t^n_i} \otimes W_{t^n_i,t^n_{i+1}} - \eta_{t^n_{i-1}} \otimes W_{t^n_{i-1},t^n_{i+1}}| = |\eta_{t^n_{i-1},t^n_i} \otimes W_{t^n_i,t^n_{i+1}}| \lesssim |t^n_{i+1} - t^n_{i-1}|^{\frac{1}{p}+\beta}$$

for any $i = 1, \ldots, N_n - 1$, where the implicit multiplicative constant is a random variable, and we can follow the proof of [124, Lemma 3.2] to deduce that, for almost any fixed $\omega \in \Omega$, for any $k < \ell$, and writing $N = \ell - k = 2^n|t^n_\ell - t^n_k|T^{-1}$,

$$\left|\left(\int_{t^n_k}^{t^n_\ell} \eta^n_u \otimes \mathrm{d}W_u\right)(\omega) - \eta_{t^n_k} \otimes W_{t^n_k,t^n_\ell}(\omega)\right| \lesssim N^{1-\frac{2}{\rho}}|t^n_\ell - t^n_k|^{\frac{2}{\rho}} \lesssim 2^{n(1-\frac{2}{\rho})}|t^n_\ell - t^n_k|,$$

where $\frac{2}{\rho} = \frac{1}{p} + \beta$.

Let $\varepsilon \in (1 - \frac{2}{p}, 1)$. If $2^{-n} \geq |t^n_\ell - t^n_k|^{\frac{4}{p(1-\varepsilon)}}$, then

$$\left|\left(\int_{t^n_k}^{t^n_\ell} \eta^n_u \otimes \mathrm{d}W_u\right)(\omega) - \eta_{t^n_k} \otimes W_{t^n_k,t^n_\ell}(\omega)\right| \lesssim |t^n_\ell - t^n_k|^{1-\frac{4}{p(1-\varepsilon)}(1-\frac{2}{\rho})}.$$

By choosing ε close to $1 - \frac{2}{p}$, we can make the above exponent $1 - \frac{4}{p(1-\varepsilon)}(1 - \frac{2}{\rho})$ arbitrarily close to $\frac{4}{\rho} - 1 = \frac{2}{p} + 2\beta - 1$. By then choosing β close to $\frac{1}{2}$, we can make this value arbitrarily close to $\frac{2}{p}$ from below. In particular, by making suitable choices of ε and β, we can ensure that $1 - \frac{4}{p(1-\varepsilon)}(1 - \frac{2}{\rho}) = \frac{2}{p'}$, and we obtain

$$\left|\left(\int_{t_k^n}^{t_\ell^n} \eta_u^n \otimes \mathrm{d}W_u\right)(\omega) - \eta_{t_k^n} \otimes W_{t_k^n, t_\ell^n}(\omega)\right| \lesssim |t_\ell^n - t_k^n|^{\frac{2}{p'}}. \tag{3.62}$$

We will now aim to obtain the same estimate in the case that $2^{-n} < |t_\ell^n - t_k^n|^{\frac{4}{p(1-\varepsilon)}}$, with ε chosen as above. Recalling (3.58) and (3.59), we have that

$$\begin{aligned}
&\left|\left(\int_{t_k^n}^{t_\ell^n} \eta_u^n \otimes \mathrm{d}W_u\right)(\omega) - \eta_{t_k^n} \otimes W_{t_k^n, t_\ell^n}(\omega)\right| \\
&= \left|\left(\int_{t_k^n}^{t_\ell^n} \eta_u^n \otimes \mathrm{d}W_u\right)(\omega) - \left(\int_{t_k^n}^{t_\ell^n} \eta_u \otimes \mathrm{d}W_u\right)(\omega)\right. \\
&\quad \left. + \left(\int_{t_k^n}^{t_\ell^n} \eta_u \otimes \mathrm{d}W_u\right)(\omega) - \eta_{t_k^n} \otimes W_{t_k^n, t_\ell^n}(\omega)\right| \\
&\leq 2\left\|\left(\int_0^\cdot \eta_u^n \otimes \mathrm{d}W - \int_0^\cdot \eta_u \otimes \mathrm{d}W\right)(\omega)\right\|_\infty + \left|\left(\int_{t_k^n}^{t_\ell^n} \eta_{t_k^n, u} \otimes \mathrm{d}W_u\right)(\omega)\right| \\
&\lesssim 2^{-\frac{n}{2}(1-\varepsilon)} + |t_\ell^n - t_k^n|^{\frac{2}{p'}} \\
&\lesssim |t_\ell^n - t_k^n|^{\frac{2}{p'}}.
\end{aligned}$$

Combining this with (3.62), we conclude that

$$\sup_{n \in \mathbb{N}} \sup_{0 \leq k < \ell \leq 2^n} \frac{|(\int_{t_k^n}^{t_\ell^n} \eta_u^n \otimes \mathrm{d}W_u)(\omega) - \eta_{t_k^n} \otimes W_{t_k^n, t_\ell^n}(\omega)|^{\frac{p'}{2}}}{C(\omega)|t_\ell^n - t_k^n|} \leq 1, \tag{3.63}$$

for a suitable random variable C.

Step 4. For any $n \in \mathbb{N}$ and $0 \leq k < \ell \leq 2^n$, it is straightforward to verify that

$$|\eta_{t_k^n, t_\ell^n}|^2 = 2\int_{t_k^n}^{t_\ell^n} \eta_{t_k^n, u}^n \cdot \mathrm{d}\eta_u + \sum_{i=k}^{\ell-1} |\eta_{t_i^n, t_{i+1}^n}|^2,$$

where $\cdot$ denotes the Euclidean inner product. It follows from (3.60) that $|\eta_{t_k^n, t_\ell^n}|^2 \lesssim w_\eta(t_k^n, t_\ell^n)^{\frac{2}{p'}}$, and that

$$\sup_{n \in \mathbb{N}} \sup_{0 \leq k < \ell \leq 2^n} \frac{|\int_{t_k^n}^{t_\ell^n} \eta_{t_k^n, u}^n \cdot \mathrm{d}\eta_u|^{\frac{p'}{2}}}{w_\eta(t_k^n, t_\ell^n)} \lesssim 1,$$

from which we then have that

$$\sup_{n \in \mathbb{N}} \sup_{0 \leq k < \ell \leq 2^n} \frac{|\sum_{i=k}^{\ell-1} |\eta_{t_i^n, t_{i+1}^n}|^2|^{\frac{p'}{2}}}{w_\eta(t_k^n, t_\ell^n)} \lesssim 1.$$

The same argument holds for the sample paths of W, and since

$$\left|\sum_{i=k}^{\ell-1} W_{t_i^n, t_{i+1}^n} \otimes \eta_{t_i^n, t_{i+1}^n}\right| \lesssim \sum_{i=k}^{\ell-1} |W_{t_i^n, t_{i+1}^n}|^2 + \sum_{i=k}^{\ell-1} |\eta_{t_i^n, t_{i+1}^n}|^2,$$

we deduce that

$$\sup_{n\in\mathbb{N}} \sup_{0\le k<\ell\le 2^n} \frac{|\sum_{i=k}^{\ell-1} W_{t_i^n,t_{i+1}^n} \otimes \eta_{t_i^n,t_{i+1}^n}|^{\frac{p'}{2}}}{w_\eta(t_k^n,t_\ell^n) + c(t_k^n,t_\ell^n)} \lesssim 1. \tag{3.64}$$

By the Hölder continuity of η and W, it is clear that $|W_{t_k^n,t_\ell^n} \otimes \eta_{t_k^n,t_\ell^n}| \lesssim |t_\ell^n - t_k^n|^{\frac{2}{p'}}$, so that

$$\sup_{n\in\mathbb{N}} \sup_{0\le k<\ell\le 2^n} \frac{|W_{t_k^n,t_\ell^n} \otimes \eta_{t_k^n,t_\ell^n}|^{\frac{p'}{2}}}{|t_\ell^n - t_k^n|} \lesssim 1. \tag{3.65}$$

For any $n \in \mathbb{N}$ and $0 \le k < \ell \le 2^n$, it is straightforward to verify that

$$W_{t_k^n,t_\ell^n} \otimes \eta_{t_k^n,t_\ell^n} = \int_{t_k^n}^{t_\ell^n} W^n_{t_k^n,u} \otimes \mathrm{d}\eta_u + \bigg(\int_{t_k^n}^{t_\ell^n} \eta^n_{t_k^n,u} \otimes \mathrm{d}W_u\bigg)^{\top} + \sum_{i=k}^{\ell-1} W_{t_i^n,t_{i+1}^n} \otimes \eta_{t_i^n,t_{i+1}^n}.$$

Recalling (3.63), (3.64) and (3.65), we thus have that

$$\sup_{n\in\mathbb{N}} \sup_{0\le k<\ell\le 2^n} \frac{|\int_{t_k^n}^{t_\ell^n} W^n_{t_k^n,u} \otimes \mathrm{d}\eta_u|^{\frac{p'}{2}}}{\hat{w}(t_k^n,t_\ell^n)} \le 1$$

for a suitable random control function $\hat{w}$. Combining this with (3.60), (3.61) and (3.63), we conclude that, for almost every $\omega \in \Omega$, the path $(\eta, W(\omega))$ indeed satisfies the third condition of Property (RIE). □

Remark 3.3.5. *A joint rough path lift of (η, W) is constructed in [58, Section 2] which allows* (3.55) *to be treated as a rough Stratonovich SDE. Since the construction of the joint lift $\mathbf{\Lambda}$ above is based on a piecewise constant approximation, as in Property* (RIE)*, rather than on linear interpolations as considered in [58], Theorem 3.3.4 provides a joint Itô-type rough path lift of (η, W) and, thus, an Itô interpretation of the rough SDE* (3.55)*, consistent with that in [72].*

Chapter 4

Pathwise analysis of log-optimal portfolios

A central challenge in mathematical finance, financial economics, and related fields is to understand the decision making of rational agents facing financial markets with their random evolution of asset prices. A major approach to this challenge, initiated by Merton [132], is the study of utility maximization problems in continuous-time financial markets. By now, a vast number of researchers contributed to this approach and investigated various facets of utility maximization problems; see, e.g., [98] and the references therein. For instance, a large body of work is devoted to constructing closed-form solutions, which are of particular interest from a practitioner's perspective; see, e.g., [133, 108, 163, 113, 81].

In classical portfolio theory, the utility maximization problems are considered and solved with the implicit assumption that the underlying model for the asset prices is perfectly specified, that is, the model parameters (trend and volatility) are fully known. Consequently, essentially all "optimal" portfolios in the literature depend on the underlying model parameters. However, in reality, due to the necessity to use statistical estimation to determine the underlying models, there is always a natural uncertainty about the model parameters and, even worse, about the underlying model itself. Especially, estimating the trend of the time-evolution of an asset price on a financial market is known to be a notoriously difficult problem, cf. [149]. Hence, to deal with model uncertainty and to understand its implications is of utmost importance in portfolio theory.

Various approaches have been developed in mathematical finance to treat model uncertainty in the context of portfolio optimization. Let us briefly mention in the following three major research areas which are most related to this chapter. The sensitivity analysis of utility maximization problems is based on classical probabilistic modeling and studies the impact of model perturbations to decision making; see, e.g., [105, 118, 159, 139]. Robust portfolio theory does not fix a fully specified underlying model, instead, it introduces a "worst-case" approach, also called Knightian approach, aiming to solve utility maximization problems simultaneously for a family of models; see, e.g., [155, 22, 141, 145]. In model-free portfolio theory, portfolios are constructed without any underlying probabilistic framework and their performance is analyzed in an entirely pathwise manner; see, e.g., [152, 48, 103, 5].

In this chapter, we develop a methodology that allows for a pathwise analysis of portfolios for individual price trajectories generated by standard models for financial markets.

In doing so, we take a model-free perspective on portfolio theory to conduct a sensitivity analysis in the classical sense. As such, it presents a novel approach to addressing the aforementioned issue of model uncertainty in the context of portfolio optimization. Notably, this one framework additionally allows us to analyze the time discretization error of portfolios explained below.

As a prototypical example of an "optimal" portfolio, we investigate the log-optimal portfolio of a classical investment-consumption optimization problem in a frictionless financial market, modeled by an Itô diffusion process, see [132, 133]. Let us quickly recall this classical utility maximization problem. We assume that the discounted price process $(\bar{S}_t)_{t\in[0,T]}$ is given by

$$\bar{S}_t = s_0 + \int_0^t \bar{b}_s \,\mathrm{d}s + \int_0^t \bar{\sigma}_s \,\mathrm{d}\bar{W}_s, \qquad t \in [0,T],$$

where s_0 is a constant, $\bar{b}$, $\bar{\sigma}$ are suitable, predictable process, and $\bar{W}$ is a Brownian motion. It is well-known that there exists a log-optimal portfolio $(\bar{\varphi}, \bar{\kappa})$ given a consumption clock K, that is,

$$\mathbb{E}\Big[\int_0^T \log(\bar{\kappa}_t) \,\mathrm{d}K_t\Big] = \sup_{(\bar{\phi},\bar{\chi})} \mathbb{E}\Big[\int_0^T \log(\bar{\chi}_t) \,\mathrm{d}K_t\Big],$$

where the supremum is taken over all admissible portfolios $(\bar{\phi}, \bar{\chi})$; see, e.g., [79, 80]. Before presenting our pathwise analysis of the log-optimal portfolio, we would like to emphasize that, based on the developed methodology, an analogous pathwise analysis can be carried out for numerous portfolios which are known to be "optimal" from classical portfolio theory.

As a foundation, we set up a suitable pathwise Itô-type integration, relying on the theory of càdlàg rough paths, see [73, 75], and, more specifically, the so-called Property (RIE) as introduced in [143, 7]. While rough path theory, of course, provides a comprehensive theory of rough integration as well, some fundamental results essential in the specific context of mathematical finance still need to be proven, and some care is required to obtain the natural economic interpretation of all involved integrals and related objects. In particular, assuming that a "noise" path W satisfies Property (RIE), the discounted price path $(S_t)_{t\in[0,T]}$ can be modeled by the (rough) differential equation

$$S_t = s_0 + \int_0^t \hat{b}_s \,\mathrm{d}s + \int_0^t \hat{\sigma}_s \,\mathrm{d}W_s, \qquad t \in [0,T], \tag{4.1}$$

where s_0 is a constant and $\hat{b}$, $\hat{\sigma}$ are suitable paths. The rationale behind the deterministic price dynamics (4.1) is that W corresponds to a fixed realization of the noise, modeling the randomness of price processes, and $\hat{b}, \hat{\sigma}$ are the model parameters specifying the model actually used for the asset prices. Hence, (4.1) provides a transparent distinction of model uncertainty and randomness. In this chapter, we work in one of the following settings for (4.1):

- local volatility models: $\hat{b}_s = b(s, S_s)$ and $\hat{\sigma}_s = \sigma(s, S_s)$ with $b \in C_b^3(\mathbb{R}^{k+1}; \mathcal{L}(\mathbb{R}; \mathbb{R}^k))$ and $\sigma \in C_b^3(\mathbb{R}^{k+1}; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k))$,
- Black–Scholes-type models: $\hat{b}_s = b_s S_s$ and $\hat{\sigma}_s = \sigma_s S_s$, where b and σ are controlled paths with respect to W.

Note that it is necessary to differentiate between these two settings since rough differential equations, like (4.1), with unbounded coefficients are a delicate challenge in rough path theory and can only be treated in specific situations; see, e.g., [119].

Based on the developed pathwise Itô-type integration, we can construct the log-optimal portfolio of Merton's investment-consumption problem entirely pathwise, given the model parameters b, σ and a fixed "noise" path W. Moreover, assuming that the "noise" paths W are realizations of a Brownian motion $\bar{W}$, the pathwise construction of the log-optimal portfolio is, indeed, a solution to Merton's classical investment-consumption problem in a frictionless financial market, modeled by an Itô diffusion process. For that reason, we shall call this pathwise constructed portfolio the pathwise log-optimal portfolio (φ, κ). However, let us remark that for the construction of the pathwise log-optimal portfolio (φ, κ) as well as its pathwise analysis, the "noise" path W can be a rather general deterministic path and does not need to be a sample path of a Brownian motion.

The present pathwise framework and the pathwise construction of the log-optimal portfolio allow us to analyze the dependency of the pathwise log-optimal portfolio on the model parameters for a fixed "noise" path. Relying on continuity estimates for rough integration and rough differential equations, we prove that the pathwise log-optimal portfolio and its associated capital process depend in a locally Lipschitz continuous way on the model parameters b, σ. For instance, in the case of local volatility models, the stability of the pathwise log-optimal portfolio and its associated capital process reads as follows

$$\|(\varphi^{(b,\sigma)}, \kappa^{(b,\sigma)}) - (\varphi^{(\tilde{b},\tilde{\sigma})}, \kappa^{(\tilde{b},\tilde{\sigma})})\|_\infty \lesssim \|b - \tilde{b}\|_{C_b^2} + \|\sigma - \tilde{\sigma}\|_{C_b^2}$$

and

$$\|V^{(b,\sigma)} - V^{(\tilde{b},\tilde{\sigma})}\|_\infty \lesssim \|b - \tilde{b}\|_{C_b^2} + \|\sigma - \tilde{\sigma}\|_{C_b^2},$$

where $(\varphi^{(b,\sigma)}, \kappa^{(b,\sigma)})$, $(\varphi^{(\tilde{b},\tilde{\sigma})}, \kappa^{(\tilde{b},\tilde{\sigma})})$ denote the pathwise log-optimal portfolios and $V^{(b,\sigma)}$, $V^{(\tilde{b},\tilde{\sigma})}$ the associated capital processes, given the model parameters b, σ and $\tilde{b}$, $\tilde{\sigma}$, respectively. The precise statements of the pathwise stability estimates with respect to model parameters can be found in Sections 4.3.2 and 4.4.2.

Model uncertainty is, of course, not the only source of trouble when aiming to implement a theoretically optimal portfolio on a real financial market. An other major challenge is the necessary time-discretization of portfolios and trading strategies in general. Indeed, while trading can be done at very high frequency, there is still some gap between high-frequency trading and continuous-time trading, and there is often a desire for various reasons to rebalance a portfolio with a lower frequency. With this in mind, we derive the convergence of the time-discretized versions of the pathwise log-optimal portfolio to its continuous-time version as well as the convergence of the associated capital processes. Moreover, we obtain quantitative bounds for the discretization error for the pathwise log-optimal portfolio as well as for the associated capital processes. The precise estimates for the discretization errors can be found in Sections 4.3.3 and 4.4.3.

This chapter is structured as follows. Section 4.1 presents the classical investment-consumption optimization problem in a probabilistic setting. In Section 4.2, we recall some essential background from rough path theory and set up the pathwise approach to stochastic Itô integration. In the case of price trajectories generated by local volatility models, the pathwise analysis of the log-optimal portfolio is developed in Section 4.3, and, in the case of price trajectories generated by Black–Scholes-type models, in Section 4.4. Appendix A.4 establishes several elementary results in the theory of càdlàg rough paths.

4.1 Portfolio optimization in a probabilistic setting

Before setting up a pathwise stability analysis of optimal portfolios, let us recall the classical formulation of (and the well-known solution to) an optimal investment-consumption problem in a probabilistic setting à la Merton [133]. For this purpose, we fix an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)_{t\in[0,T]}$ satisfying the usual conditions, i.e., completeness and right-continuity.

Following [79, 80], we consider an optimal investment-consumption problem, where "optimal" refers to the maximization of the expected log-utility from the investor's consumption over a finite time horizon $T > 0$. In the next subsection, we give a precise formulation of the investment-consumption problem.

4.1.1 A classical investment-consumption optimization problem

The underlying frictionless financial market consists of $k+1$ assets, where the discounted price process $\bar{S} = (\bar{S}^0, \bar{S}^1, \ldots, \bar{S}^k) = (\bar{S}_t)_{t\in[0,T]}$ is an $\mathbb{R}^{k+1}$-valued $\mathcal{F}_t$-adapted semimartingale on $(\Omega, \mathcal{F}, \mathbb{P})$, with $\bar{S}^0 \equiv 1$.

Following [79], we adopt the following standard setup:

- A self-financing trading strategy $\bar{\varphi} \in L(\bar{S})$, where $L(\bar{S})$ denotes the space of all $\bar{S}$-integrable predictable processes in the spirit of stochastic Itô integration, is called *admissible*, and denoted by $\bar{\varphi} \in \mathfrak{S}$, if $\int_0^t \bar{\varphi}_s^\top \,\mathrm{d}\bar{S}_s \geq -1$ for all $t \in [0,T]$, $\mathbb{P}$-almost surely, where $(\cdot)^\top$ denotes matrix transposition.

- The *consumption clock* $K\colon [0,T] \to \mathbb{R}$ is an increasing deterministic càdlàg function, and $\mathfrak{K}$ denotes the set of all non-negative optional processes $\bar{\kappa}$, called the *consumption rate*, such that $\int_0^t \bar{\kappa}_s \,\mathrm{d}K_s < \infty$ for all $t \in [0,T]$, $\mathbb{P}$-almost surely. For $\bar{\kappa} \in \mathfrak{K}$, the consumption process is given by $\int_0^\cdot \bar{\kappa}_s \,\mathrm{d}K_s$.

- A pair $(\bar{\varphi}, \bar{\kappa}) \in \mathfrak{S} \times \mathfrak{K}$ belongs to the set $\mathfrak{P}$ of *admissible portfolios* if the discounted wealth process $(\bar{V}_t)_{t\in[0,T]}$, given by

$$\bar{V}_t(\bar{\varphi}, \bar{\kappa}) := 1 + \int_0^t \bar{\varphi}_s^\top \,\mathrm{d}\bar{S}_s - \int_0^t \bar{\kappa}_s \,\mathrm{d}K_s, \qquad t \in [0,T],$$

  is non-negative, $\mathbb{P}$-almost surely.

Typical choices of the consumption clock are $K_t = \mathbf{1}_{[T,\infty)}(t)$, i.e., consumption only at time T, and $K_t = \sum_{s\leq t} \mathbf{1}_{\mathbb{N}}(s)$, i.e., consumption only at integer times.

Occasionally, as will become apparent, we will identify $\bar{S}$ with the $\mathbb{R}^k$-valued process $(\bar{S}^1, \ldots, \bar{S}^k)$, and similarly for $\bar{\varphi}$.

A pair $(\bar{\varphi}, \bar{\kappa}) \in \mathfrak{P}$ is called a *log-optimal portfolio* if $(\bar{\varphi}, \bar{\kappa})$ maximizes the map $\Phi_{\log}\colon \mathfrak{P} \to \mathbb{R}$, given by

$$(\bar{\phi}, \bar{\chi}) \mapsto \mathbb{E}\bigg[\int_0^T \log(\bar{\chi}_t) \,\mathrm{d}K_t\bigg],$$

over all $(\bar{\phi}, \bar{\chi}) \in \mathfrak{P}$.

4.1.2 The log-optimal portfolio for the investment-consumption problem

Finding log-optimal portfolios in the context of expected utility maximization is a well-studied mathematical problem; see, e.g., [112] for a classical introduction. For example, in a general semimartingale framework, the works of Goll and Kallsen [79, 80] provide explicit formulae in terms of semimartingale characteristics. In the following, we recall the result of [79] in the case that the discounted price process is modeled by an Itô process.

Let $\bar{W} = (\bar{W}_t)_{t\in[0,T]}$ be a d-dimensional Brownian motion defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with respect to $(\mathcal{F}_t)_{t\in[0,T]}$, and suppose that the discounted price process $(\bar{S}_t)_{t\in[0,T]}$ is given by

$$\bar{S}_t = s_0 + \int_0^t \bar{b}_s \,\mathrm{d}s + \int_0^t \bar{\sigma}_s \,\mathrm{d}\bar{W}_s, \qquad t \in [0,T],$$

where $s_0 \in \mathbb{R}^k$, $\bar{b}$ is a predictable, locally integrable $\mathbb{R}^k$-valued process, and $\bar{\sigma}$ is a predictable, locally square integrable $\mathbb{R}^{k\times d}$-valued process such that $\bar{\sigma}_t\bar{\sigma}_t^\top$ is a positive definite $k \times k$-matrix for every t, where each coefficient is bounded away from zero.

In the present setting, [79, Theorem 3.1], which formulates the solution to the optimal investment-consumption problem, reads as follows.

Theorem 4.1.1. *Assume that there exists an $\mathbb{R}^k$-valued process $\bar{H} \in L(\bar{S})$ such that*

$$\bar{b}_t - \bar{c}_t\bar{H}_t = 0 \qquad \textit{with} \qquad \bar{c}_t := \bar{\sigma}_t\bar{\sigma}_t^\top, \qquad t \in [0,T],$$

holds $\mathbb{P} \otimes \mathrm{d}t$-almost everywhere, and set

$$\bar{\kappa}_t := \frac{1}{K_T}\bar{\mathcal{E}}\Big(\int_0^\cdot \bar{H}_s^\top \,\mathrm{d}\bar{S}_s\Big)_t, \qquad \bar{V}_t := \bar{\kappa}_t(K_T - K_t),$$

$$\bar{\varphi}_t^i := \bar{H}_t^i\bar{V}_{t-}, \quad i = 1,\dots,k, \qquad \bar{\varphi}_t^0 := \int_0^t \bar{\varphi}_s^\top \,\mathrm{d}\bar{S}_s - \sum_{i=1}^k \bar{\varphi}_t^i\bar{S}_t^i, \qquad t \in [0,T],$$

where we set $\bar{V}_{0-} := 0$, and $\bar{\mathcal{E}}$ denotes the stochastic exponential. Then, $(\bar{\varphi}, \bar{\kappa}) \in \mathfrak{P}$ is a log-optimal portfolio with discounted wealth process $(\bar{V}_t)_{t\in[0,T]}$.

Remark 4.1.2. *If the price process $\bar{S}$ is given by a linear stochastic differential equation, i.e., if $\bar{b}^i = \bar{S}^i\hat{b}^i$ and $\bar{\sigma}^{i,j} = \bar{S}^i\hat{\sigma}^{ij}$, for some predictable $\hat{b}^i$, $\hat{\sigma}^{ij}$, $i = 1,\dots,k$, $j = 1,\dots,d$, then the previous theorem can be rephrased as follows; see also [79, Example 4.2].*

Assume that there exists a predictable, $\mathbb{R}^k$-valued process $\bar{h}$ such that

$$\hat{b}_t - \hat{c}_t\bar{h}_t = 0 \qquad \textit{with} \qquad \hat{c}_t = \hat{\sigma}_t\hat{\sigma}_t^\top, \qquad t \in [0,T],$$

holds $\mathbb{P} \otimes \mathrm{d}t$-almost everywhere, and set

$$\bar{H}_t^i := \frac{\bar{h}_t^i}{\bar{S}_t^i}, \qquad t \in [0,T], \qquad i = 1,\dots,k,$$

and $\bar{\kappa}$, $\bar{V}$, $\bar{\varphi}^i$, $i = 0,\dots,k$, as defined in Theorem 4.1.1. Then, $(\bar{\varphi}, \bar{\kappa}) \in \mathfrak{P}$ is a log-optimal portfolio with discounted wealth process $(\bar{V}_t)_{t\in[0,T]}$.

4.2 Pathwise stochastic analysis

Developing a methodology that allows for a pathwise analysis of optimal portfolios requires, unsurprisingly, an underlying pathwise framework. To that end, we rely on the theory of rough paths—see, e.g., [71] for an introductory textbook—and, more specifically, the so-called Property (RIE) as introduced in [143, 7], which provides a suitable foundation for the use of rough path theory in mathematical finance. We start by recalling some essentials from the theory of càdlàg rough paths. For a more comprehensive introduction we refer to [73, 75].

4.2.1 Essentials of rough path theory

A *partition* $\mathcal{P}$ of an interval $[s,t]$ is a finite set of points between and including the points s and t, i.e., $\mathcal{P} = \{s = u_0 < u_1 < \cdots < u_N = t\}$ for some $N \in \mathbb{N}$, and its mesh size is denoted by $|\mathcal{P}| := \max\{|u_{i+1} - u_i| \colon i = 0, \ldots, N-1\}$.

Throughout, we let $T > 0$ be a fixed finite time horizon. We let $\Delta_T := \{(s,t) \in [0,T]^2 : s \leq t\}$ denote the standard 2-simplex. A function $w \colon \Delta_T \to [0,\infty)$ is called a *control function* if it is superadditive, in the sense that $w(s,u) + w(u,t) \leq w(s,t)$ for all $0 \leq s \leq u \leq t \leq T$. For two vectors $x = (x^1, \ldots, x^d)^\top, y = (y^1, \ldots, y^d)^\top \in \mathbb{R}^d$ we use the usual tensor product

$$x \otimes y := (x^i y^j)_{i,j=1,\ldots,d} \in \mathbb{R}^{d \times d}.$$

Whenever $(B, \|\cdot\|)$ is a normed space and $f, g \colon B \to \mathbb{R}$ are two functions on B, we shall write $f \lesssim g$ or $f \leq Cg$ to mean that there exists a constant $C > 0$ such that $f(x) \leq Cg(x)$ for all $x \in B$. The constant C may depend on the normed space, e.g., through its dimension or regularity parameters.

For two vector spaces, the space of linear maps from $E_1 \to E_2$ is denoted by $\mathcal{L}(E_1; E_2)$, and we write, e.g., $C_b^l = C_b^l(\mathbb{R}^m; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k))$ for the space of l-times differentiable (in the Fréchet sense) functions $f \colon \mathbb{R}^m \to \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k)$ such that f and all its derivatives up to order l are continuous and bounded. We equip this space with the norm

$$\|f\|_{C_b^l} := \|f\|_\infty + \|\mathrm{D}f\|_\infty + \cdots + \|\mathrm{D}^l f\|_\infty,$$

where $\mathrm{D}^n f$ denotes the n-th order derivative of f, and $\|\cdot\|_\infty$ denotes the supremum norm on the corresponding spaces of operators.

For a normed space $(E, |\cdot|)$, we let $D([0,T]; E)$ denote the set of càdlàg (right-continuous with left-limits) paths from $[0,T] \to E$. For $X \in D([0,T]; E)$, the supremum norm of the path X is given by

$$\|X\|_\infty := \sup_{t \in [0,T]} |X_t|,$$

and for $p \geq 1$, the p-variation of the path X is given by

$$\|X\|_p := \|X\|_{p,[0,T]} \qquad \text{with} \qquad \|X\|_{p,[s,t]} := \Big(\sup_{\mathcal{P} \subset [s,t]} \sum_{[u,v] \in \mathcal{P}} |X_v - X_u|^p \Big)^{\frac{1}{p}}, \quad (s,t) \in \Delta_T,$$

where the supremum is taken over all possible partitions $\mathcal{P}$ of the interval $[s,t]$. We recall that, given a path X, we have that $\|X\|_p < \infty$ if and only if there exists a control function w such that[1]

$$\sup_{(u,v) \in \Delta_T} \frac{|X_v - X_u|^p}{w(u,v)} < \infty.$$

We write $D^p = D^p([0,T];E)$ for the space of paths $X \in D([0,T];E)$ which satisfy $\|X\|_p < \infty$.

For a path $X \in D([0,T];E)$, we will use the shorthand notation:

$$X_{s,t} := X_t - X_s \quad \text{and} \quad X_{t-} := \lim_{u \uparrow t} X_u, \qquad \text{for} \quad (s,t) \in \Delta_T.$$

For $r \geq 1$ and a two-parameter function $\mathbb{X} \colon \Delta_T \to E$, we similarly define

$$\|\mathbb{X}\|_r := \|\mathbb{X}\|_{r,[0,T]} \qquad \text{with} \qquad \|\mathbb{X}\|_{r,[s,t]} := \Big(\sup_{\mathcal{P} \subset [s,t]} \sum_{[u,v] \in \mathcal{P}} |\mathbb{X}_{u,v}|^r \Big)^{\frac{1}{r}}, \quad (s,t) \in \Delta_T.$$

We write $D_2^r = D_2^r(\Delta_T;E)$ for the space of all functions $\mathbb{X} \colon \Delta_T \to E$ which satisfy $\|\mathbb{X}\|_r < \infty$, and are such that the maps $s \mapsto \mathbb{X}_{s,t}$ for fixed t, and $t \mapsto \mathbb{X}_{s,t}$ for fixed s, are both càdlàg.

For $p \in [2,3)$, a pair $\mathbf{X} = (X, \mathbb{X})$ is called a *càdlàg rough path* over $\mathbb{R}^d$ if

(i) $X \in D^p([0,T];\mathbb{R}^d)$ and $\mathbb{X} \in D_2^{\frac{p}{2}}(\Delta_T;\mathbb{R}^{d \times d})$, and

(ii) Chen's relation: $\mathbb{X}_{s,t} = \mathbb{X}_{s,u} + \mathbb{X}_{u,t} + X_{s,u} \otimes X_{u,t}$ holds for all $0 \leq s \leq u \leq t \leq T$.

In component form, condition (ii) states that $\mathbb{X}^{ij}_{s,t} = \mathbb{X}^{ij}_{s,u} + \mathbb{X}^{ij}_{u,t} + X^i_{s,u} X^j_{u,t}$ for every i and j. We will denote the space of càdlàg rough paths by $\mathcal{D}^p = \mathcal{D}^p([0,T];\mathbb{R}^d)$. On the space $\mathcal{D}^p([0,T];\mathbb{R}^d)$, we use the natural seminorm

$$\|\mathbf{X}\|_p := \|\mathbf{X}\|_{p,[0,T]} \qquad \text{with} \qquad \|\mathbf{X}\|_{p,[s,t]} := \|X\|_{p,[s,t]} + \|\mathbb{X}\|_{\frac{p}{2},[s,t]}$$

for $(s,t) \in \Delta_T$, and the induced distance

$$\|\mathbf{X}; \widetilde{\mathbf{X}}\|_p := \|\mathbf{X}; \widetilde{\mathbf{X}}\|_{p,[0,T]} \qquad \text{with} \qquad \|\mathbf{X}; \widetilde{\mathbf{X}}\|_{p,[s,t]} := \|X - \widetilde{X}\|_{p,[s,t]} + \|\mathbb{X} - \widetilde{\mathbb{X}}\|_{\frac{p}{2},[s,t]},$$

whenever $\mathbf{X} = (X, \mathbb{X}), \widetilde{\mathbf{X}} = (\widetilde{X}, \widetilde{\mathbb{X}}) \in \mathcal{D}^p([0,T];\mathbb{R}^d)$. Recall that the rough path $\mathbf{X} = (X, \mathbb{X})$ above a path X is not unique.

[1] Here and throughout, we adopt the convention that $\frac{0}{0} := 0$.

Let $p \in [2,3)$, $q \in [p,\infty)$ and $r \in [\frac{p}{2},2)$ such that $\frac{1}{p}+\frac{1}{r} > 1$ and $\frac{1}{p}+\frac{1}{q} = \frac{1}{r}$. Let $X \in D^p([0,T];\mathbb{R}^d)$. We say that a pair (Y,Y') is a *controlled path* (with respect to X), if

$$Y \in D^p([0,T];E), \quad Y' \in D^q([0,T];\mathcal{L}(\mathbb{R}^d;E)), \quad \text{and} \quad R^Y \in D_2^r(\Delta_T;E),$$

where R^Y is defined by

$$Y_{s,t} = Y'_s X_{s,t} + R^Y_{s,t} \qquad \text{for all} \quad (s,t) \in \Delta_T.$$

We write $\mathcal{V}^{q,r}_X = \mathcal{V}^{q,r}_X([0,T];E)$ for the space of E-valued controlled paths, which becomes a Banach space when equipped with the norm $(Y,Y') \mapsto \|Y,Y'\|_{\mathcal{V}^{q,r}_X}$, where

$$\|Y,Y'\|_{\mathcal{V}^{q,r}_X} := \|Y,Y'\|_{\mathcal{V}^{q,r}_X,[0,T]},$$

with

$$\|Y,Y'\|_{\mathcal{V}^{q,r}_X,[s,t]} := |Y_s| + |Y'_s| + \|Y'\|_{q,[s,t]} + \|R^Y\|_{r,[s,t]},$$

for $(s,t) \in \Delta_T$. It is straightforward to see that

$$\|Y\|_p \leq \|Y'\|_\infty \|X\|_p + \|R^Y\|_r \qquad \text{and} \qquad \|Y'\|_\infty \leq |Y'_0| + \|Y'\|_q,$$

so that in particular

$$\|Y\|_\infty \leq (1 + \|X\|_p)\|Y,Y'\|_{\mathcal{V}^{q,r}_X}. \tag{4.2}$$

We further introduce the standard "distance"

$$\|Y;\widetilde{Y}\|_{\mathcal{V}^{q,r}_X,\mathcal{V}^{q,r}_{\widetilde{X}}} := \|Y;\widetilde{Y}\|_{\mathcal{V}^{q,r}_X,\mathcal{V}^{q,r}_{\widetilde{X}},[0,T]}$$

with

$$\|Y;\widetilde{Y}\|_{\mathcal{V}^{q,r}_X,\mathcal{V}^{q,r}_{\widetilde{X}},[s,t]} := |Y_s - \widetilde{Y}_s| + |Y'_s - \widetilde{Y}'_s| + \|Y' - \widetilde{Y}'\|_{q,[s,t]} + \|R^Y - R^{\widetilde{Y}}\|_{r,[s,t]},$$

for $(s,t) \in \Delta_T$, whenever $(Y,Y') \in \mathcal{V}^{q,r}_X$, $(\widetilde{Y},\widetilde{Y}') \in \mathcal{V}^{q,r}_{\widetilde{X}}$. Note that, in general, $\mathcal{V}^{q,r}_X$ and $\mathcal{V}^{q,r}_{\widetilde{X}}$ are different Banach spaces; if $X = \widetilde{X}$, we write $\|Y;\widetilde{Y}\|_{\mathcal{V}^{q,r}_X}$. When $q = p$ and $r = \frac{p}{2}$, we write $\mathcal{V}^p_X = \mathcal{V}^{p,\frac{p}{2}}_X$.

We also note that

$$\|Y - \widetilde{Y}\|_p \leq C(\|Y;\widetilde{Y}\|_{\mathcal{V}^{q,r}_X,\mathcal{V}^{q,r}_{\widetilde{X}}} + \|X - \widetilde{X}\|_p) \tag{4.3}$$

for some constant C which depends only on $\|Y\|_{\mathcal{V}^{q,r}_X}$, $\|\widetilde{Y}\|_{\mathcal{V}^{q,r}_{\widetilde{X}}}$, $\|X\|_p$ and $\|\widetilde{X}\|_p$.

Given $p \in (2,3)$, $\mathbf{X} = (X,\mathbb{X}) \in \mathcal{D}^p([0,T];\mathbb{R}^d)$ and $(Y,Y') \in \mathcal{V}^{q,r}_X([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$, the (forward) rough integral

$$\int_s^t Y_u \,\mathrm{d}\mathbf{X}_u := \lim_{|\mathcal{P}^n| \to 0} \sum_{[u,v] \in \mathcal{P}^n} (Y_u X_{u,v} + Y'_u \mathbb{X}_{u,v}), \qquad (s,t) \in \Delta_T, \tag{4.4}$$

exists (in the classical mesh Riemann–Stieltjes sense), where the limit is taken along any sequence of partitions $(\mathcal{P}^n)_{n\in\mathbb{N}}$ of the interval $[s,t]$ such that $|\mathcal{P}^n|\to 0$ as $n\to\infty$. To be precise, in writing the product $Y_u X_{u,v}$, we apply the operator $Y_u \in \mathcal{L}(\mathbb{R}^d;\mathbb{R}^k)$ onto $X_{u,v}\in\mathbb{R}^d$, and in writing the product $Y'_u\mathbb{X}_{u,v}$, we use the natural identification of $\mathcal{L}(\mathbb{R}^d;\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$ with $\mathcal{L}(\mathbb{R}^d\otimes\mathbb{R}^d;\mathbb{R}^k)$. The rough integral comes with the estimate

$$\left|\int_s^t Y_u\,\mathrm{d}\mathbf{X}_u - Y_s X_{s,t} - Y'_s\mathbb{X}_{s,t}\right| \leq C\Big(\|R^Y\|_{r,[s,t)}\|X\|_{p,[s,t]}+\|Y'\|_{q,[s,t)}\|\mathbb{X}\|_{\frac{p}{2},[s,t]}\Big)$$

for some constant C which depends only on p,q and r (see, e.g., [7, Proposition 2.4 and Remark 2.5]), where

$$\|Y'\|_{q,[s,t)} := \sup_{u<t}\|Y'\|_{q,[s,u]} \qquad \text{and} \qquad \|R^Y\|_{r,[s,t)} := \sup_{u<t}\|R^Y\|_{r,[s,u]}.$$

This implies that $(\int_0^\cdot Y_u\,\mathrm{d}\mathbf{X}_u, Y)\in\mathcal{V}_X^{q,r}$ is a controlled path with respect to X, and satisfies

$$\left\|\int_0^\cdot Y_u\,\mathrm{d}\mathbf{X}_u\right\|_{\mathcal{V}_X^{q,r}} \leq C, \tag{4.5}$$

for some constant C depending only on $p,q,r,\|Y\|_{\mathcal{V}_X^{q,r}}$ and $\|\mathbf{X}\|_p$.

Given a rough path $\mathbf{X}=(X,\mathbb{X})\in\mathcal{D}^p([0,T];\mathbb{R}^d)$ with $p\in[2,3)$, there exists a unique controlled path $(Y,Y')\in\mathcal{V}_X^p([0,T];\mathbb{R}^k)$ satisfying the rough differential equation (RDE)

$$Y_t = y_0 + \int_0^t b(s,Y_s)\,\mathrm{d}s + \int_0^t \sigma(s,Y_s)\,\mathrm{d}\mathbf{X}_s, \qquad t\in[0,T],$$

if $b\in C_b^2(\mathbb{R}^{k+1};\mathcal{L}(\mathbb{R};\mathbb{R}^k))$ and $\sigma\in C_b^3(\mathbb{R}^{k+1};\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$; see, e.g., Theorem 3.1.1.

4.2.2 Pathwise Itô-type integration

Rough path theory provides a pathwise approach to stochastic integration and stochastic differential equations. In particular, it allows to recover the stochastic Itô and Stratonovich integrals by choosing the corresponding rough path lift of a semimartingale. Consequently, from a financial modelling perspective, choosing the rough path above a given path without care could create arbitrage. Moreover, the definition of the rough integral (4.4) lacks a canonical interpretation in mathematical finance. To overcome these issues, we rely on the so-called Property (RIE), as introduced in [143, 7].

Property (RIE). *Let $p\in(2,3)$ and let $\mathcal{P}^n=\{0=t_0^n<t_1^n<\cdots<t_{N_n}^n=T\}$, $n\in\mathbb{N}$, be a sequence of partitions of the interval $[0,T]$ such that $|\mathcal{P}^n|\to 0$ as $n\to\infty$. For $X\in D([0,T];\mathbb{R}^d)$, and each $n\in\mathbb{N}$, we define $X^n\colon[0,T]\to\mathbb{R}^d$ by*

$$X_t^n = X_T\mathbf{1}_{\{T\}}(t) + \sum_{k=0}^{N_n-1} X_{t_k^n}\mathbf{1}_{[t_k^n,t_{k+1}^n)}(t), \qquad t\in[0,T].$$

We assume that:

(i) the sequence of paths $(X^n)_{n\in\mathbb{N}}$ converges uniformly to X as $n\to\infty$,

(ii) the Riemann sums $\int_0^t X^n_u \otimes \mathrm{d}X_u := \sum_{k=0}^{N_n-1} X_{t^n_k} \otimes X_{t^n_k\wedge t, t^n_{k+1}\wedge t}$ converge uniformly as $n\to\infty$ to a limit, which we denote by $\int_0^t X_u\otimes \mathrm{d}X_u$, $t\in[0,T]$,

(iii) and there exists a control function w such that

$$\sup_{(s,t)\in\Delta_T} \frac{|X_{s,t}|^p}{w(s,t)} + \sup_{n\in\mathbb{N}} \sup_{0\leq k<\ell\leq N_n} \frac{|\int_{t^n_k}^{t^n_\ell} X^n_u \otimes \mathrm{d}X_u - X_{t^n_k}\otimes X_{t^n_k,t^n_\ell}|^{\frac{p}{2}}}{w(t^n_k,t^n_\ell)} \leq 1. \tag{4.6}$$

We say that a path $X\in D([0,T];\mathbb{R}^d)$ satisfies Property (RIE) relative to p and $(\mathcal{P}^n)_{n\in\mathbb{N}}$, if p, $(\mathcal{P}^n)_{n\in\mathbb{N}}$ and X together satisfy Property (RIE).

It is known that, if a path $X\in D([0,T];\mathbb{R}^d)$ satisfies Property (RIE), then X extends canonically to a rough path $\mathbf{X}=(X,\mathbb{X})\in\mathcal{D}^p([0,T];\mathbb{R}^d)$, where the lift $\mathbb{X}$ is defined by

$$\mathbb{X}_{s,t} := \int_s^t X_u\otimes \mathrm{d}X_u - X_s\otimes(X_t-X_s), \qquad (s,t)\in\Delta_T, \tag{4.7}$$

with $\int_s^t X_u\otimes\mathrm{d}X_u := \int_0^t X_u\otimes\mathrm{d}X_u - \int_0^s X_u\otimes\mathrm{d}X_u$, and the existence of the integral $\int_0^t X_u\otimes \mathrm{d}X_u$ is ensured by condition (ii) of Property (RIE); see [7, Lemma 2.13]. When assuming Property (RIE) for a path X, we will always work with the rough path $\mathbf{X}=(X,\mathbb{X})$ defined via (4.7), and note that $\mathbf{X}=(X,\mathbb{X})$ corresponds to the Itô rough path lift of a stochastic process, since the "iterated integral" $\mathbb{X}$ is given as a limit of left-point Riemann sums, analogously to the stochastic Itô integral.

Property (RIE) not only ensures the existence of a suitable rough path lift of a path, but also allows the rough integral to be expressed as a classical limit of Riemann sums. Consequently, the rough integral possesses the natural interpretation in a financial context as the capital process of a portfolio. The next theorem is a slight generalization of [7, Theorem 2.15].

Theorem 4.2.1. *Let $p\in(2,3)$, $q\in[p,\infty)$ and $r\in[\frac{p}{2},2)$ such that $\frac{1}{p}+\frac{1}{r}>1$ and $\frac{1}{p}+\frac{1}{q}=\frac{1}{r}$, and let $\mathcal{P}^n=\{0=t^n_0<t^n_1<\cdots<t^n_{N_n}=T\}$, $n\in\mathbb{N}$, be a sequence of partitions such that $|\mathcal{P}^n|\to 0$ as $n\to\infty$. Suppose that $X\in D([0,T];\mathbb{R}^d)$ satisfies Property* (RIE) *relative to p and $(\mathcal{P}^n)_{n\in\mathbb{N}}$, and let $\mathbf{X}$ be the canonical rough path lift of X, as constructed in* (4.7)*. Let $(F,F'),(G,G')\in\mathcal{V}^{q,r}_X$ be controlled paths with respect to X, and suppose that $J_F\subseteq\liminf_{n\to\infty}\mathcal{P}^n := \bigcup_{m\in\mathbb{N}}\bigcap_{n\geq m}\mathcal{P}^n$, where $J_F:=\{t\in(0,T]: F_{t-}\neq F_t\}$ denotes the set of jump times of F. Then, the limit*

$$\int_0^t F_u\,\mathrm{d}G_u := \lim_{n\to\infty}\sum_{k=0}^{N_n-1} F_{t^n_k}\otimes G_{t^n_k\wedge t, t^n_{k+1}\wedge t},$$

exists, where the convergence holds uniformly for $t \in [0,T]$, and it coincides with the rough integral of (F,F') against (G,G'), as defined in (A.3)*.*

The proof of Theorem 4.2.1 follows the proof of [7, Theorem 2.15] almost verbatim. The only difference is that, rather than using [7, Proposition 2.14] to establish the uniform convergence of F^n to F, we can instead use [6, Proposition B.1] (which does not require the sequence of partitions to be nested).

A crucial observation for our pathwise analysis of log-optimal portfolios is that, if a path X satisfies Property (RIE), then suitable controlled paths relative to X do as well. This is made precise in Theorem 4.2.2 below. In particular, a corollary of this result is that if X satisfies Property (RIE), and Y is the solution to an RDE driven by the canonical rough path lift of X, then Y itself satisfies Property (RIE) with respect to the same sequence of partitions.

Theorem 4.2.2. *Suppose that $X \in D([0,T];\mathbb{R}^d)$ satisfies Property* (RIE) *relative to some $p \in (2,3)$ and a sequence of partitions $\mathcal{P}^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$. Let $(Y,Y') \in \mathcal{V}_X^p$ be a controlled path such that $J_Y \subseteq \liminf_{n\to\infty} \mathcal{P}^n$, where $J_Y := \{t \in (0,T] : Y_{t-} \neq Y_t\}$ denotes the set of jump times of Y. Then, Y satisfies Property* (RIE) *relative to p and $(\mathcal{P}^n)_{n\in\mathbb{N}}$.*

Proof. For each $n \in \mathbb{N}$, let

$$Y_t^n = Y_T \mathbf{1}_{\{T\}}(t) + \sum_{k=0}^{N_n-1} Y_{t_k^n} \mathbf{1}_{[t_k^n, t_{k+1}^n)}(t), \qquad t \in [0,T],$$

be the piecewise constant approximation of Y along $\mathcal{P}^n$. Since $J_Y \subseteq \liminf_{n\to\infty} \mathcal{P}^n$, we have from Proposition A.3.1 that $Y^n \to Y$ uniformly as $n \to \infty$, so that part (i) of Property (RIE) holds.

By Lemma A.4.1, we can define the rough integral of the controlled path (Y,Y') against itself as

$$\int_0^t Y_r \,\mathrm{d}Y_r := \lim_{|\mathcal{P}|\to 0} \sum_{[u,v]\in\mathcal{P}} Y_u \otimes Y_{u,v} + (Y'_u \otimes Y'_u)\mathbb{X}_{u,v}, \qquad t \in [0,T],$$

relative to the rough path $\mathbf{X} = (X, \mathbb{X})$, where the limit exists along any sequence of partitions $\mathcal{P}$ of the interval $[0,t]$ with mesh size tending to zero. We have from Theorem 4.2.1 that

$$\int_0^t Y_r^n \otimes \mathrm{d}Y_r = \sum_{k=0}^{N_n-1} Y_{t_k^n} \otimes Y_{t_k^n\wedge t, t_{k+1}^n \wedge t} \longrightarrow \int_0^t Y_r \,\mathrm{d}Y_r \qquad \text{as} \qquad n \longrightarrow \infty, \tag{4.8}$$

where the convergence is uniform in $t \in [0,T]$, which gives part (ii) of Property (RIE).

As the piecewise constant approximation X^n as defined in Property (RIE) has finite 1-variation, we also have that $\mathbf{X}^n = (X, X^n, \mathbb{X}^n)$ is a càdlàg rough path in the sense of [7, Definition 2.1], where

$$\mathbb{X}^n_{s,t} := \int_s^t X^n_u \otimes \mathrm{d}X_u - X^n_s \otimes X_{s,t}, \qquad (s,t) \in \Delta_T. \tag{4.9}$$

We note that (Y^n, Y') is a controlled path with respect to X^n. We can therefore consider the rough integral of (Y^n, Y') against (Y, Y') relative to the rough path $\mathbf{X}^n$ in the sense of [7, Proposition 2.4], which is given by

$$\int_0^t Y^n_r \,\mathrm{d}Y_r = \lim_{|\mathcal{P}|\to 0} \sum_{[u,v]\in\mathcal{P}} Y^n_u \otimes Y_{u,v} + (Y'_u \otimes Y'_u)\mathbb{X}^n_{u,v}.$$

For any refinement $\widetilde{\mathcal{P}}$ of the partition $(\mathcal{P}^n \cup \{t\}) \cap [0,t]$ and any $[u,v] \in \widetilde{\mathcal{P}}$, there exists a k such that $t^n_k \leq u < v \leq t^n_{k+1}$ which, recalling (4.9), implies that $\mathbb{X}^n_{u,v} = 0$. Thus,

$$\int_0^t Y^n_r \,\mathrm{d}Y_r = \lim_{|\widetilde{\mathcal{P}}|\to 0} \sum_{[u,v]\in\widetilde{\mathcal{P}}} Y^n_u \otimes Y_{u,v} = \sum_{k=0}^{N_n-1} Y_{t^n_k} \otimes Y_{t^n_k\wedge t, t^n_{k+1}\wedge t},$$

so that the rough integral $\int_0^t Y^n_r \,\mathrm{d}Y_r$ coincides with the Riemann–Stieltjes integral on the left-hand side of (4.8).

Let us fix $0 \leq k < \ell \leq N_n$. By the estimate in [7, Proposition 2.4], we have that

$$\begin{aligned} &\left| \int_{t^n_k}^{t^n_\ell} Y^n_r \,\mathrm{d}Y_r - Y_{t^n_k} \otimes Y_{t^n_k, t^n_\ell} - (Y'_{t^n_k} \otimes Y'_{t^n_k})\mathbb{X}^n_{t^n_k, t^n_\ell} \right| \\ &\quad \lesssim \|Y'\|_\infty (\|Y'\|^p_{p,[t^n_k,t^n_\ell]} + \|X^n\|^p_{p,[t^n_k,t^n_\ell]})^{\frac{2}{p}} \|X\|_{p,[t^n_k,t^n_\ell]} + \|Y^n\|_{p,[t^n_k,t^n_\ell]} \|R^Y\|_{\frac{p}{2},[t^n_k,t^n_\ell]} \\ &\qquad + \|R^{Y^n}\|_{\frac{p}{2},[t^n_k,t^n_\ell]} \|Y'\|_\infty \|X\|_{p,[t^n_k,t^n_\ell]} + \|Y' \otimes Y'\|_{p,[t^n_k,t^n_\ell]} \|\mathbb{X}^n\|_{\frac{p}{2},[t^n_k,t^n_\ell]}. \end{aligned} \tag{4.10}$$

It is clear that the functions given by $w_1(s,t) := \|Y'\|^p_{p,[s,t]}$, $w_2(s,t) := \|X\|^p_{p,[s,t]}$, and $w_3(s,t) := \|R^Y\|^{\frac{p}{2}}_{\frac{p}{2},[s,t]}$ for $(s,t) \in \Delta_T$ are all controls. Since $t^n_k, t^n_\ell \in \mathcal{P}^n$, we have that

$$\begin{aligned} &\|X^n\|_{p,[t^n_k,t^n_\ell]} \leq \|X\|_{p,[t^n_k,t^n_\ell]} = w_2(t^n_k, t^n_\ell)^{\frac{1}{p}}, \\ \text{and} \qquad &\|Y^n\|_{p,[t^n_k,t^n_\ell]} \leq \|Y\|_{p,[t^n_k,t^n_\ell]} \leq \|Y\|_p. \end{aligned}$$

Let w denote the control with respect to which (4.6) holds for X. Note that X also satisfies Property (RIE) over the subinterval $[t^n_k, t^n_\ell]$, with respect to p, the sequence of partitions $(\mathcal{P}^m \cap [t^n_k, t^n_\ell])_{m\geq n}$, and the same control w. It then follows from [7, Lemma 2.12] that

$$\sup_{m\geq n} \|\mathbb{X}^m\|_{\frac{p}{2},[t^n_k,t^n_\ell]} \lesssim w(t^n_k, t^n_\ell)^{\frac{2}{p}}.$$

We also infer from the proof of [7, Theorem 2.15] that

$$\begin{aligned}\sup_{m\geq n}\|R^{Y^m}\|^{\frac{p}{2}}_{\frac{p}{2},[t^n_k,t^n_\ell]} &\lesssim \|Y'\|^p_{p,[t^n_k,t^n_\ell]}+\|X\|^p_{p,[t^n_k,t^n_\ell]}+\|R^Y\|^{\frac{p}{2}}_{\frac{p}{2},[t^n_k,t^n_\ell]}\\ &= w_1(t^n_k,t^n_\ell)+w_2(t^n_k,t^n_\ell)+w_3(t^n_k,t^n_\ell).\end{aligned}$$

The estimate in (4.10) then implies that

$$\begin{aligned}&\bigg|\int_{t^n_k}^{t^n_\ell} Y^n_u \,\mathrm{d}Y_u - Y_{t^n_k}\otimes Y_{t^n_k,t^n_\ell} - (Y'_{t^n_k}\otimes Y'_{t^n_k})\mathbb{X}^n_{t^n_k,t^n_\ell}\bigg|^{\frac{p}{2}}\\ &\quad\lesssim \|Y'\|^{\frac{p}{2}}_\infty(w_1(t^n_k,t^n_\ell)+w_2(t^n_k,t^n_\ell))\|X\|^{\frac{p}{2}}_p+\|Y\|^{\frac{p}{2}}_p w_3(t^n_k,t^n_\ell)\\ &\qquad+(w_1(t^n_k,t^n_\ell)+w_2(t^n_k,t^n_\ell)+w_3(t^n_k,t^n_\ell))\|Y'\|^{\frac{p}{2}}_\infty\|X\|^{\frac{p}{2}}_p+\|Y'\otimes Y'\|^{\frac{p}{2}}_p w(t^n_k,t^n_\ell).\end{aligned}$$

Since we can also bound

$$|(Y'_{t^n_k}\otimes Y'_{t^n_k})\mathbb{X}^n_{t^n_k,t^n_\ell}|^{\frac{p}{2}}\leq \|Y'\otimes Y'\|^{\frac{p}{2}}_\infty\|\mathbb{X}^n\|^{\frac{p}{2}}_{\frac{p}{2},[t^n_k,t^n_\ell]}\lesssim \|Y'\otimes Y'\|^{\frac{p}{2}}_\infty w(t^n_k,t^n_\ell),$$

it is then clear how to choose a control w_4 such that

$$\bigg|\int_{t^n_k}^{t^n_\ell} Y^n_u \,\mathrm{d}Y_u - Y_{t^n_k}\otimes Y_{t^n_k,t^n_\ell}\bigg|^{\frac{p}{2}} \leq w_4(t^n_k,t^n_\ell).$$

Since w_4 does not depend on the choices of $n\in\mathbb{N}$ or $0\leq k<l\leq N_n$, we have established part (iii) of Property (RIE). □

4.2.3 Consistency of rough and stochastic integration

In this subsection we briefly discuss the relation between the deterministic theory of rough integration, as developed in Sections 4.2.1 and 4.2.2, and stochastic integration. As before, we fix a probability space $(\Omega,\mathcal{F},\mathbb{P})$ with a filtration $(\mathcal{F}_t)_{t\in[0,T]}$ satisfying the usual conditions. As shown in Section 3.2, the sample paths of various stochastic processes, such as Brownian motion, Itô processes and Lévy processes, almost surely satisfy Property (RIE) relative to $p\in(2,3)$ and suitable sequences of partitions $(\mathcal{P}^n)_{n\in\mathbb{N}}$. In the current work, we fundamentally rely on Brownian motion, and thus recall the corresponding result in the following remark, which combines the content of Lemma 3.2.1 and Proposition 3.2.2.

Remark 4.2.3. *Let $\bar{W}=(\bar{W}_t)_{t\in[0,T]}$ be d-dimensional Brownian motion, $p\in(2,3)$ and $\mathcal{P}^n=\{0=t^n_0<t^n_1<\cdots<t^n_{N_n}=T\}$, $n\in\mathbb{N}$, be a sequence of equidistant partitions of the interval $[0,T]$, so that, for each $n\in\mathbb{N}$, there exists some $\pi_n>0$ such that $t^n_{i+1}-t^n_i=\pi_n$ for each $0\leq i<N_n$. If $\pi_n^{2-\frac{4}{p}}\log(n)\to 0$ as $n\to\infty$, then, for almost every $\omega\in\Omega$, the sample path $W(\omega)$ satisfies Property* (RIE) *relative to p and $(\mathcal{P}^n)_{n\in\mathbb{N}}$.*

Moreover, the canonical rough path $\bar{\mathbf{W}}(\omega) = (\bar{W}(\omega), \bar{\mathbb{W}}(\omega))$ *defined via Property* (RIE) *corresponds almost surely to the random rough path defined via Itô integration, namely, where*

$$\bar{\mathbb{W}}_{s,t} := \int_s^t \bar{W}_{s,r} \otimes \mathrm{d}\bar{W}_r = \int_s^t \bar{W}_r \otimes \mathrm{d}\bar{W}_r - \bar{W}_s \otimes \bar{W}_{s,t}, \qquad (s,t) \in \Delta_T.$$

Property (RIE) also ensures that a random rough integral against a semimartingale coincides almost surely with the associated stochastic Itô integral.

Proposition 4.2.4. *Let* $X = (X_t)_{t \in [0,T]}$ *be a* d*-dimensional càdlàg semimartingale and let* (Y, Y') *be a càdlàg stochastic process adapted to* $(\mathcal{F}_t)_{t \in [0,T]}$*. Let* $p \in (2,3)$*. By part (i) of Proposition 3.2.10, there exists an adapted sequence of partitions* $\mathcal{P}^n = \{\tau_k^n\}$*,* $n \in \mathbb{N}$*, (so that each* $\tau_k^n \in \mathcal{P}^n$ *is a stopping time), such that, for almost every* $\omega \in \Omega$*, the path* $X(\omega)$ *satisfies Property* (RIE) *relative to p and* $(\mathcal{P}^n(\omega))_{n \in \mathbb{N}}$*. Suppose that, for almost every* $\omega \in \Omega$*,* $(Y(\omega), Y'(\omega))$ *is a controlled path in* $\mathcal{V}^{q,r}_{X(\omega)}$ *with* $J_{Y(\omega)} \subseteq \liminf_{n \to \infty} \mathcal{P}^n(\omega)$*, where* $J_{Y(\omega)}$ *denotes the set of jump times of* $Y(\omega)$*. Then the rough and Itô integrals of* Y *against* X *coincide* $\mathbb{P}$*-almost surely, that is,*

$$\int_0^t Y_s(\omega) \, \mathrm{d}\mathbf{X}_s(\omega) = \bigg(\int_0^t Y_{s-} \, \mathrm{d}X_s \bigg)(\omega) \qquad \text{for all } t \in [0,T],$$

holds for almost every $\omega \in \Omega$*, where* $\mathbf{X}(\omega)$ *is the canonical rough path lift of* $X(\omega)$ *as defined via Property* (RIE)*.*

Proof. By, e.g., [147, Chapter II, Theorem 21], we have that

$$\sum_{k=0}^{N_n - 1} Y_{\tau_k^n} X_{\tau_k^n \wedge t, \tau_{k+1}^n \wedge t} \longrightarrow \int_0^t Y_{s-} \, \mathrm{d}X_s \qquad \text{as} \quad n \to \infty,$$

where the convergence holds uniformly (in $t \in [0,T]$) in probability. By taking a subsequence if necessary, we can then assume that the (uniform) convergence holds almost surely. On the other hand, by Theorem A.3.2, we know that, for almost every $\omega \in \Omega$,

$$\sum_{k=0}^{N_n - 1} Y_{\tau_k^n(\omega)}(\omega) X_{\tau_k^n(\omega) \wedge t, \tau_{k+1}^n(\omega) \wedge t}(\omega) \longrightarrow \int_0^t Y_s(\omega) \, \mathrm{d}\mathbf{X}_s(\omega) \qquad \text{as} \quad n \to \infty$$

uniformly for $t \in [0,T]$. The result thus follows by the uniqueness of limits. □

4.3 Local volatility models: pathwise analysis of log-optimal portfolios

In this section we shall study log-optimal portfolios for the investment-consumption problem, acting on deterministic price paths generated by local volatility models, defined in a pathwise manner. To this end, we make the following assumption throughout this section.

Assumption 4.3.1. *Let $p \in (2,3)$ and let $\mathcal{P}^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$, be a sequence of equidistant partitions of the interval $[0,T]$, such that,*

- *for each $n \in \mathbb{N}$, there exists some $\pi_n > 0$ such that $t_{i+1}^n - t_i^n = \pi_n$ for each $0 \le i < N_n$,*
- *$\pi_n^{2-\frac{4}{p}} \log(n) \to 0$ as $n \to \infty$,*
- *$J_K \subseteq \liminf_{n\to\infty} \mathcal{P}^n$ with $J_K := \{t \in (0,T] : K_{t-} \neq K_t\}$,*

where the consumption clock $K \colon [0,T] \to \mathbb{R}$ is fixed as in Section 4.1.1. Moreover, the deterministic path $W \colon [0,T] \to \mathbb{R}^d$ satisfies Property (RIE) *relative to p and $(\mathcal{P}^n)_{n\in\mathbb{N}}$.*

We suppose that the discounted price path $(S_t)_{t\in[0,T]}$ satisfies the rough differential equation

$$S_t = s_0 + \int_0^t b(s, S_s)\,\mathrm{d}s + \int_0^t \sigma(s, S_s)\,\mathrm{d}\mathbf{W}_s, \qquad t \in [0,T], \tag{4.11}$$

where $s_0 \in \mathbb{R}^k$, $b \in C_b^3(\mathbb{R}^{k+1}; \mathcal{L}(\mathbb{R};\mathbb{R}^k))$, $\sigma \in C_b^3(\mathbb{R}^{k+1}; \mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$, and $\mathbf{W} = (W, \mathbb{W})$ is the canonical rough path lift of W as defined in (4.7).

Remark 4.3.2. *If W is a realization of a Brownian motion, the dynamics of the RDE* (4.11) *can be seen as a fixed realization of a local volatility model for a financial market.*

Indeed, let us assume that $\bar{W} = (\bar{W}_t)_{t\in[0,T]}$ is a d-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with respect to an underlying filtration $(\mathcal{F}_t)_{t\in[0,T]}$. It is well-known that the stochastic differential equation (SDE)

$$\bar{S}_t = s_0 + \int_0^t b(s, \bar{S}_s)\,\mathrm{d}s + \int_0^t \sigma(s, \bar{S}_s)\,\mathrm{d}\bar{W}_s, \qquad t \in [0,T], \tag{4.12}$$

has a unique strong solution, where $\int_0^t \sigma(s, \bar{S}_s)\,\mathrm{d}\bar{W}_s$ denotes the stochastic Itô integral; see, e.g., [147, Chapter V, Theorem 6]. Note that the Itô diffusion $(\bar{S}_t)_{t\in[0,T]}$ represents many standard models for financial markets, including local volatility models.

Recall that for almost every $\omega \in \Omega$, the sample path $\bar{W}(\omega)$ of a Brownian motion satisfies Property (RIE) *relative to p and $(\mathcal{P}^n)_{n\in\mathbb{N}}$; see Remark 4.2.3. Hence, for almost every $\omega \in \Omega$, the solution $(\bar{S}_t(\omega))_{t\in[0,T]}$ of the SDE* (4.12) *driven by $\bar{W}$, and the solution $(S_t)_{t\in[0,T]}$ of the rough differential equation driven by the rough path $\mathbf{W} = (W, \mathbb{W}) := \bar{\mathbf{W}}(\omega) = (\bar{W}(\omega), \bar{\mathbb{W}}(\omega))$ coincide; see Lemma 3.2.1. In other words, $(S_t)_{t\in[0,T]}$ can be understood as a fixed realization of the probabilistic model $(\bar{S}_t)_{t\in[0,T]}$.*

In the present setting, it will be convenient to equivalently reformulate the RDE (4.11) into the RDE

$$S_t = s_0 + \int_0^t (b, \sigma)(s, S_s)\,\mathrm{d}(\cdot, \mathbf{W})_s, \qquad t \in [0,T], \tag{4.13}$$

where $(\cdot, \mathbf{W})$ denotes the time-extended rough path of $\mathbf{W}$, i.e., the path-level of $(\cdot, \mathbf{W})$ is given by $(t, W_t)_{t\in[0,T]}$ and the missing integrals $\int W^j \, \mathrm{d}t$, $\int t \, \mathrm{d}W^j$, $j = 1, \dots, d$, to define a rough path are canonically defined as Riemann–Stieltjes integrals. By classical rough path theory (e.g., [4, Theorem 2.5]), for any $b \in C_b^3(\mathbb{R}^{k+1}; \mathcal{L}(\mathbb{R}; \mathbb{R}^k))$, $\sigma \in C_b^3(\mathbb{R}^{k+1}; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k))$, there exists a unique solution $(S, S') \in \mathcal{V}^p_{(\cdot,W)}$ to the RDE (4.13), where $S' = (b, \sigma)(\cdot, S)$. Moreover, $(S_t)_{t\in[0,T]}$ satisfies the RDE (4.13) if and only if it satisfies the RDE (4.11). For later reference, we also remark that $(\cdot, W)$ satisfies Property (RIE) relative to p and $(\mathcal{P}^n)_{n\in\mathbb{N}}$ by Proposition 3.1.10.

4.3.1 Pathwise construction of log-optimal portfolios

As a first step to a pathwise analysis of optimal portfolios, we prove a pathwise construction of the log-optimal portfolio, supposing that the underlying price dynamics of the financial market are given by a local volatility model. Recall that in the probabilistic setting the log-optimal portfolio is well-known and was presented in Theorem 4.1.1, which is due to [79].

Let $\mathcal{A} \subset C_b^3([0,T] \times \mathbb{R}^k; \mathcal{L}(\mathbb{R}^d, \mathbb{R}^k))$ be the class of functions σ such that $\sigma(t,x)\sigma(t,x)^\top \in \mathrm{GL}(\mathbb{R}^{k\times k})$ for all (t,x), where each coefficient is uniformly bounded away from zero, endowed with the $\|\cdot\|_{C_b^3}$ norm. Here, $\mathrm{GL}(\mathbb{R}^{k\times k})$ denotes the general linear group of degree k. For a $k \times k$-matrix we write $\det(\cdot)$ for its determinant, and $(\cdot)^\top$ denotes matrix transposition. Given a path W, the time-extended path is denoted by $(\cdot, W) = (t, W_t)_{t\in[0,T]}$.

Lemma 4.3.3. *For $(b, \sigma) \in C_b^3([0,T] \times \mathbb{R}^k; \mathcal{L}(\mathbb{R}; \mathbb{R}^k)) \times \mathcal{A}$, let*

$$H_t := H_t^{(b,\sigma)} := c(t, S_t)^{-1} b(t, S_t) \qquad \text{with} \qquad c(t, S_t) := c_t^{(b,\sigma)} := \sigma(t, S_t)\sigma(t, S_t)^\top,$$

for $t \in [0,T]$, and set $(\varphi, \kappa) := (\varphi^{(b,\sigma)}, \kappa^{(b,\sigma)}) := (\varphi^{(b,\sigma),0}, \dots, \varphi^{(b,\sigma),k}, \kappa^{(b,\sigma)})$, with

$$\kappa_t := \kappa_t^{(b,\sigma)} := \frac{1}{K_T} \mathcal{E}\Big(\sum_{i=1}^k \int_0^\cdot H_s^i \, \mathrm{d}S_s^i\Big)_t, \qquad V_t := V_t^{(b,\sigma)} := \kappa_t (K_T - K_t),$$

$$\varphi_t^i := \varphi_t^{(b,\sigma),i} := H_t^i V_t, \qquad i = 1, \dots, k, \qquad \varphi_t^0 := \varphi_t^{(b,\sigma),0} := \sum_{i=1}^k \int_0^t \varphi_s^i \, \mathrm{d}S_s^i - \varphi_t^i S_t^i,$$

for $t \in [0,T]$, where $\int_0^t \varphi_s^i \, \mathrm{d}S_s^i$ is the rough integral, and $\mathcal{E}$ is the rough exponential as defined in Lemma A.4.7. Then, φ, κ and V are all well-defined and are controlled paths with respect to W and, in particular, with respect to $(\cdot, W)$.

Proof. It is a well-known result in rough path theory (e.g., [75, Lemma 3.5]) that the composition of a controlled path with a regular function remains a controlled path. More precisely, since $S \in \mathcal{V}_W^p$, we have that $\sigma(\cdot, S)$ and $b(\cdot, S)$ are controlled paths in $\mathcal{V}_W^p$. We recall that $\det(\sigma(\cdot,S)\sigma(\cdot,S)^\top)$ is bounded away from zero. We then obtain (componentwise) that H is a controlled path in $\mathcal{V}^p_W$, since the sum and the product of two (real-valued) controlled paths is again a controlled path by Lemma A.4.3, as well as the inverse of a controlled path which is bounded away from zero (as a composition with the smooth function $x \mapsto \frac{1}{x}$).

By Lemma A.4.1, since H^i and S^i are both controlled paths, the rough integral $\int_0^\cdot H^i_s \,\mathrm{d}S^i_s$ is well-defined and is itself a controlled path for each $i = 1,\dots,k$. By Lemma A.4.6, the path $Z_t := \sum_{i=1}^k \int_0^t H^i_s \,\mathrm{d}S^i_s$ may then considered as a rough path, and Lemma A.4.7 then implies that the rough exponential $\mathcal{E}(Z)$, and hence also $\kappa = \frac{1}{K_T}\mathcal{E}(Z)$, are controlled paths.

Since the consumption clock K is a càdlàg (deterministic) and increasing function (and thus of finite 1-variation), by Lemma A.4.3, the wealth process V is a controlled path in $\mathcal{V}^p_W$, as the product of two controlled paths.

By similar arguments, we see that $\varphi^i \in \mathcal{V}^p_W$, $i = 0,1,\dots,k$, are also all controlled paths with respect to W, and hence also with respect to $(\cdot, W)$. □

The portfolio constructed in Lemma 4.3.3 in a pathwise manner agrees, indeed, with the log-optimal portfolio for the investment-consumption problem as considered in Section 4.1, if the underlying frictionless financial market is generated by a local volatility model, such as the stochastic differential equation (4.12). Hence, in the following we shall call the portfolio $(\varphi,\kappa) = (\varphi^{(b,\sigma)}, \kappa^{(b,\sigma)})$ from Lemma 4.3.3 a *pathwise log-optimal portfolio*.

Lemma 4.3.4. *Suppose that the discounted price process* $(\bar{S}_t)_{t\in[0,T]}$ *is modelled by the SDE* (4.12) *driven by a Brownian motion* $\bar{W}$ *on a probability space* $(\Omega,\mathcal{F},\mathbb{P})$ *with respect to an underlying filtration* $(\mathcal{F}_t)_{t\in[0,T]}$. *Then the log-optimal portfolio* $(\bar{\varphi},\bar{\kappa})$*, as provided in Theorem 4.1.1, and the pathwise log-optimal portfolio* (φ,κ)*, as provided in Lemma 4.3.3, coincide* $\mathbb{P}$*-almost surely, where* (φ,κ) *is constructed given the realization* $\mathbf{W} := \bar{\mathbf{W}}(\omega)$ *of the Itô rough path lift of the Brownian motion* $\bar{W}$*, for almost every* $\omega \in \Omega$.

One may note that in Theorem 4.1.1 we had $\bar{\varphi}^i_t = \bar{H}^i_t \bar{V}_{t-}$, but in Lemma 4.3.3 we have $\varphi^i_t = H^i_t V_t$. This is only to be consistent with standard rough analysis in which controlled paths are assumed to be càdlàg, and makes no difference to the value of the rough integral $\int_0^t \varphi^i_s \,\mathrm{d}S^i_s$, as explained in [7, Remark 3.3].

Proof. In this proof we consider S, φ, κ, etc., as random controlled paths, in the sense that S is a stochastic process such that $S(\omega)$ is a controlled path for almost every $\omega \in \Omega$. In particular, S is defined pathwise as the solution to the RDE (4.11), and (φ,κ) is defined pathwise via Lemma 4.3.3, given a realization of the Brownian motion $\bar{W}$; see Remark 4.3.2.

By the associativity property of rough integrals (see Proposition A.4.5), and the consistency of rough and stochastic integrals (as established in Proposition 4.2.4), we have that

$$\begin{aligned}\int_0^t H_s^i \,\mathrm{d}S_s^i &= \int_0^t H_s^i b^i(s, S_s)\,\mathrm{d}s + \int_0^t H_s^i \sigma^{i\cdot}(s, S_s)\,\mathrm{d}\mathbf{W}_s \\ &= \int_0^t \bar{H}_s^i b^i(s, \bar{S}_s)\,\mathrm{d}s + \int_0^t \bar{H}_s^i \sigma^{i\cdot}(s, \bar{S}_s)\,\mathrm{d}\bar{W}_s = \int_0^t \bar{H}_s^i \,\mathrm{d}\bar{S}_s^i,\end{aligned}$$

almost surely for each $i = 1, \ldots, k$, where $\int_0^t \bar{H}_s^i \,\mathrm{d}\bar{S}_s^i$ is the stochastic Itô integral of $\bar{H}_t = c(t, \bar{S}_t)^{-1} b(t, \bar{S}_t)$ against the price process $\bar{S}$ in (4.12).

Let $Z := \sum_{i=1}^k \int_0^\cdot H_s^i \,\mathrm{d}S_s^i$ and $\bar{Z} = \sum_{i=1}^k \int_0^\cdot \bar{H}_s^i \,\mathrm{d}\bar{S}_s^i = \int_0^\cdot \bar{H}_s^\top \,\mathrm{d}\bar{S}_s$. By Lemma A.4.6, Z admits a canonical rough path lift $\mathbf{Z} \in \mathcal{D}^p$. Further, since $(Z, Z') \in \mathcal{V}^p_{(\cdot, W)}$ and W satisfies Property (RIE) relative to p and $(\mathcal{P}^n)_{n \in \mathbb{N}}$, Z also satisfies Property (RIE) by Theorem 4.2.2. By [7, Proposition 2.18], this implies that the rough path bracket $[\mathbf{Z}]$ coincides with the quadratic variation $[Z]$ of Z along $(\mathcal{P}^n)_{n \in \mathbb{N}}$ in the sense of Föllmer, that is

$$[\mathbf{Z}]_t = [Z]_t = \lim_{n \to \infty} \sum_{k=0}^{N_n - 1} (Z_{t_k^n \wedge t, t_{k+1}^n \wedge t})^2, \qquad t \in [0, T].$$

On the other hand, we have that

$$[\bar{Z}]_t = \lim_{n \to \infty} \sum_{k=0}^{N_n - 1} (\bar{Z}_{t_k^n \wedge t, t_{k+1}^n \wedge t})^2, \qquad t \in [0, T],$$

where the convergence holds uniformly (in $t \in [0, T]$) in probability. By taking a subsequence if necessary, we can then assume that the uniform convergence holds almost surely, and it follows that $[\bar{Z}] = [\mathbf{Z}]$ almost surely. In particular, by Lemma A.4.7, we have that

$$\mathcal{E}(Z)_t = \exp\Big(Z_t - \frac{1}{2}[\mathbf{Z}]_t\Big) = \exp\Big(\bar{Z}_t - \frac{1}{2}[\bar{Z}]_t\Big) = \bar{\mathcal{E}}(\bar{Z})_t$$

almost surely, where $\bar{\mathcal{E}}(\bar{Z})$ denotes the stochastic exponential of $\bar{Z}$. Moreover, it holds that $\sum_{i=1}^k \int_0^\cdot \varphi_s^i \,\mathrm{d}S_s^i = \int_0^\cdot \bar{\varphi}_s^\top \,\mathrm{d}\bar{S}_s$ almost surely. Thus, the log-optimal portfolio $(\bar{\varphi}, \bar{\kappa})$, as provided in Theorem 4.1.1, and the pathwise log-optimal portfolio (φ, κ), as provided in Lemma 4.3.3, coincide almost surely. □

Remark 4.3.5. *We take $\bar{W}$ to be a Brownian motion to ensure that the pathwise log-optimal portfolio (φ, κ), as constructed in Lemma 4.3.3, is, indeed, a log-optimal portfolio for the investment-consumption problem in the setting of local volatility models. However, we emphasize that the construction of the pathwise portfolio (φ, κ) as well as its pathwise analysis developed in Sections 4.3.2 and 4.3.3 works for any path W satisfying Assumption 4.3.1.*

4.3.2 Stability of pathwise log-optimal portfolios with respect to drift and volatility

Having at hand a pathwise construction of log-optimal portfolios, we are in a position to study its pathwise stability properties. In this subsection, we analyze the stability of the log-optimal portfolio and the associated capital process with respect to the model parameters b and σ.

In particular, the following result shows that the pathwise log-optimal portfolio $(\varphi, \kappa) = (\varphi^{(b,\sigma)}, \kappa^{(b,\sigma)})$ and its associated capital processes $V = V^{(b,\sigma)}$ are locally Lipschitz continuous with respect to these parameters.

Theorem 4.3.6. *For $(b, \sigma), (\tilde{b}, \tilde{\sigma}) \in C_b^3([0, T] \times \mathbb{R}^k; \mathbb{R}^k) \times \mathcal{A}$, let $(\varphi^{(b,\sigma)}, \kappa^{(b,\sigma)})$ and $(\varphi^{(\tilde{b},\tilde{\sigma})}, \kappa^{(\tilde{b},\tilde{\sigma})})$ be the corresponding pathwise log-optimal portfolios, as constructed in Lemma 4.3.3. Let M be an upper bound for*

$$\|b\|_{C_b^3}, \|\tilde{b}\|_{C_b^3}, \|\sigma\|_{C_b^3}, \|\tilde{\sigma}\|_{C_b^3}, 1/\inf_{(t,x)} |\det(\sigma(t,x)\sigma(t,x)^\top)|, 1/\inf_{(t,x)} |\det(\tilde{\sigma}(t,x)\tilde{\sigma}(t,x)^\top)|$$

and $\|(\cdot, \mathbf{W})\|_p$. We then have that

$$\|(\varphi^{(b,\sigma)}, \kappa^{(b,\sigma)}); (\varphi^{(\tilde{b},\tilde{\sigma})}, \kappa^{(\tilde{b},\tilde{\sigma})})\|_{\mathcal{V}^p_{(\cdot,W)}} \lesssim \|b - \tilde{b}\|_{C_b^2} + \|\sigma - \tilde{\sigma}\|_{C_b^2}$$

and

$$\|V^{(b,\sigma)}; V^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}} \lesssim \|b - \tilde{b}\|_{C_b^2} + \|\sigma - \tilde{\sigma}\|_{C_b^2},$$

and in particular that

$$\|(\varphi^{(b,\sigma)}, \kappa^{(b,\sigma)}) - (\varphi^{(\tilde{b},\tilde{\sigma})}, \kappa^{(\tilde{b},\tilde{\sigma})})\|_\infty \lesssim \|b - \tilde{b}\|_{C_b^2} + \|\sigma - \tilde{\sigma}\|_{C_b^2}$$

and

$$\|V^{(b,\sigma)} - V^{(\tilde{b},\tilde{\sigma})}\|_\infty \lesssim \|b - \tilde{b}\|_{C_b^2} + \|\sigma - \tilde{\sigma}\|_{C_b^2},$$

where the implicit multiplicative constants depend only on p, k, d, M, s_0 and the consumption clock K.

Proof. Step 1. Using the classical result from rough path theory (e.g., [4, Theorem 2.5]), for any $(b, \sigma) \in C_b^3([0, T] \times \mathbb{R}^k; \mathbb{R}^k) \times \mathcal{A}$, we recall that there exists a unique solution $(S^{(b,\sigma)}, (S^{(b,\sigma)})') \in \mathcal{V}^p_{(\cdot,W)}$ to the rough differential equation

$$S_t^{(b,\sigma)} = s_0 + \int_0^t (b, \sigma)(s, S_s^{(b,\sigma)}) \,\mathrm{d}(\cdot, \mathbf{W})_s, \qquad t \in [0, T],$$

where $(S^{(b,\sigma)})' = (b, \sigma)(\cdot, S^{(b,\sigma)})$.

By part (i) of Corollary 2.2.3, we get that

$$\|S^{(b,\sigma)}\|_{\mathcal{V}^p_{(\cdot,W)}}+\|S^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}}\leq C,$$

where $C>0$ depends only on p, M, and s_0. Furthermore, by the continuity of the solution map with respect to the controlled path norm, for any $(b,\sigma),(\tilde{b},\tilde{\sigma})\in C_b^3([0,T]\times\mathbb{R}^k;\mathbb{R}^k)\times\mathcal{A}$, see part (ii) of Corollary 2.2.3, it holds that

$$\|S^{(b,\sigma)};S^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|b-\tilde{b}\|_{C_b^2}+\|\sigma-\tilde{\sigma}\|_{C_b^2}, \tag{4.14}$$

where the implicit multiplicative constant depends only on p, M, and s_0.

Step 2. Another well-known result in rough path theory (e.g., [75, Lemma 3.5]) is that the compositions of controlled paths with regular functions remain a controlled path and that such a composition is locally Lipschitz continuous. More precisely, for (b,σ) and $(\tilde{b},\tilde{\sigma})$ in $C_b^3([0,T]\times\mathbb{R}^d;\mathbb{R}^d)\times\mathcal{A}$, we have that $b(\cdot,S^{(b,\sigma)})$, $\sigma(\cdot,S^{(b,\sigma)})$, $\tilde{b}(\cdot,S^{(\tilde{b},\tilde{\sigma})})$ and $\tilde{\sigma}(\cdot,S^{(\tilde{b},\tilde{\sigma})})$ are controlled paths in $\mathcal{V}^p_{(\cdot,W)}$, and we obtain that

$$\|b(\cdot,S^{(b,\sigma)})\|_{\mathcal{V}^p_{(\cdot,W)}}+\|\sigma(\cdot,S^{(b,\sigma)})\|_{\mathcal{V}^p_{(\cdot,W)}}\leq C, \tag{4.15}$$

where $C>0$ depends only on p, M, and s_0, see, e.g., [75, Lemma 3.5]; the same holds for $\tilde{b}(\cdot,S^{(\tilde{b},\tilde{\sigma})})$ and $\tilde{\sigma}(\cdot,S^{(\tilde{b},\tilde{\sigma})})$. It also holds that

$$\|b(\cdot,S^{(b,\sigma)});\tilde{b}(\cdot,S^{(\tilde{b},\tilde{\sigma})})\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|b-\tilde{b}\|_{C_b^2}+\|S^{(b,\sigma)};S^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}}$$

and

$$\|\sigma(\cdot,S^{(b,\sigma)});\tilde{\sigma}(\cdot,S^{(\tilde{b},\tilde{\sigma})})\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|\sigma-\tilde{\sigma}\|_{C_b^2}+\|S^{(b,\sigma)};S^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}},$$

where the implicit multiplicative constant depends only on p, M, and s_0. Combining the above estimates with (4.14), we get that

$$\|b(\cdot,S^{(b,\sigma)});\tilde{b}(\cdot,S^{(\tilde{b},\tilde{\sigma})})\|_{\mathcal{V}^p_{(\cdot,W)}}+\|\sigma(\cdot,S^{(b,\sigma)});\tilde{\sigma}(\cdot,S^{(\tilde{b},\tilde{\sigma})})\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|b-\tilde{b}\|_{C_b^2}+\|\sigma-\tilde{\sigma}\|_{C_b^2}, \tag{4.16}$$

where the implicit multiplicative constant depends only on p, M, and s_0.

Step 3. Let $c^{(b,\sigma)}:=\sigma(\cdot,S^{(b,\sigma)})\sigma(\cdot,S^{(b,\sigma)})^\top$ and $c^{(\tilde{b},\tilde{\sigma})}:=\tilde{\sigma}(\cdot,S^{(\tilde{b},\tilde{\sigma})})\tilde{\sigma}(\cdot,S^{(\tilde{b},\tilde{\sigma})})^\top$. We recall that $\det(\sigma(\cdot,S^{(b,\sigma)})\sigma(\cdot,S^{(b,\sigma)})^\top)$ and $\det(\tilde{\sigma}(\cdot,S^{(\tilde{b},\tilde{\sigma})})\tilde{\sigma}(\cdot,S^{(\tilde{b},\tilde{\sigma})})^\top)$ are bounded away from zero by assumption. We then obtain (componentwise) that $(c^{(b,\sigma)})^{-1}$ and $(c^{(\tilde{b},\tilde{\sigma})})^{-1}$ are controlled paths in $\mathcal{V}^p_{(\cdot,W)}$ since the sum and the product of (real-valued) controlled paths is again a controlled path (see Lemma A.4.3), as well as the inverse of a controlled path that is bounded away from zero (as a composition with the regular function $x\mapsto\frac{1}{x}$). Applying Lemma A.4.3 and the estimate (4.15), we can derive for each component that

$$\|((c^{(b,\sigma)})^{-1})^{ij}\|_{\mathcal{V}^p_{(\cdot,W)}}\leq C, \tag{4.17}$$

where $C > 0$ depends only on p, k, d, M, and s_0; the same holds for $((c^{(\tilde{b},\tilde{\sigma})})^{-1})^{ij}$.

By Lemma A.4.4 and since the composition of a controlled path with a regular function is locally Lipschitz continuous, we can check with (4.15) and (4.16) that

$$\|((c^{(b,\sigma)})^{-1})^{ij}; ((c^{(\tilde{b},\tilde{\sigma})})^{-1})^{ij}\|_{\mathcal{V}^p_{(\cdot,W)}} \lesssim \|b - \tilde{b}\|_{C_b^2} + \|\sigma - \tilde{\sigma}\|_{C_b^2}, \tag{4.18}$$

where the implicit multiplicative constant depends only on p, k, d, M, and s_0.

Step 4. We now recall the proof of Lemma 4.3.3, particularly, that

$$H^{(b,\sigma)} := (c^{(b,\sigma)})^{-1} b(\cdot, S^{(b,\sigma)}) \qquad \text{and} \qquad H^{(\tilde{b},\tilde{\sigma})} := (c^{(\tilde{b},\tilde{\sigma})})^{-1} \tilde{b}(\cdot, S^{(\tilde{b},\tilde{\sigma})})$$

are controlled paths in $\mathcal{V}^p_{(\cdot,W)}$. Particularly, $H^{(b,\sigma),i} = \sum_{j=1}^k ((c^{(b,\sigma)})^{-1})^{ij} b(\cdot, S^{(b,\sigma)})^j$, $i = 1, \dots, k$. Then, Lemma A.4.3 implies that

$$\|H^{(b,\sigma),i}\|_{\mathcal{V}^p_{(\cdot,W)}} \leq C, \tag{4.19}$$

where $C > 0$ depends only on p, M, $\|((c^{(b,\sigma)})^{-1})^{ij}\|_{\mathcal{V}^p_{(\cdot,W)}}$, $\|b(\cdot, S^{(b,\sigma)})^j\|_{\mathcal{V}^p_{(\cdot,W)}}$, $i, j = 1, \dots, k$, that is, only on p, k, d, M, and s_0, where we have applied the estimates (4.17) and (4.15); the same holds for $H^{(\tilde{b},\tilde{\sigma}),i}$.

Lemma A.4.4 then gives that

$$\begin{aligned} &\|H^{(b,\sigma),i}; H^{(\tilde{b},\tilde{\sigma}),i}\|_{\mathcal{V}^p_{(\cdot,W)}} \\ &\quad \lesssim \sum_{j=1}^k \|((c^{(b,\sigma)})^{-1})^{ij}; ((c^{(\tilde{b},\tilde{\sigma})})^{-1})^{ij}\|_{\mathcal{V}^p_{(\cdot,W)}} + \|b(\cdot, S^{(b,\sigma)})^j; \tilde{b}(\cdot, S^{(\tilde{b},\tilde{\sigma})})^j\|_{\mathcal{V}^p_{(\cdot,W)}}, \end{aligned}$$

where the implicit multiplicative constant depends only on

$$p, M, \|((c^{(b,\sigma)})^{-1})^{ij}\|_{\mathcal{V}^p_{(\cdot,W)}}, \|((c^{(\tilde{b},\tilde{\sigma})})^{-1})^{ij}\|_{\mathcal{V}^p_{(\cdot,W)}}, \|b(\cdot, S^{(b,\sigma)})^j\|_{\mathcal{V}^p_{(\cdot,W)}} \text{ and } \|\tilde{b}(\cdot, S^{(\tilde{b},\tilde{\sigma})})^j\|_{\mathcal{V}^p_{(\cdot,W)}},$$

for $i, j = 1, \dots, k$. Using the estimates (4.17) and (4.15), (4.18) and (4.16), it follows that

$$\|H^{(b,\sigma),i}; H^{(\tilde{b},\tilde{\sigma}),i}\|_{\mathcal{V}^p_{(\cdot,W)}} \lesssim \|b - \tilde{b}\|_{C_b^2} + \|\sigma - \tilde{\sigma}\|_{C_b^2}, \tag{4.20}$$

where the implicit multiplicative constant depends only on p, k, d, M, and s_0.

Step 5. Let $\vartheta^{(b,\sigma)} := \sigma(\cdot, S^{(b,\sigma)})^\top H^{(b,\sigma)}$, $\vartheta^{(\tilde{b},\tilde{\sigma})} = \tilde{\sigma}(\cdot, S^{(\tilde{b},\tilde{\sigma})})^\top H^{(\tilde{b},\tilde{\sigma})}$. Then,

$$\theta^{(b,\sigma)} := (\frac{1}{2}(\vartheta^{(b,\sigma)})^\top \vartheta^{(b,\sigma)}, (\vartheta^{(b,\sigma)})^\top) \quad \text{and} \quad \theta^{(\tilde{b},\tilde{\sigma})} := (\frac{1}{2}(\vartheta^{(\tilde{b},\tilde{\sigma})})^\top \vartheta^{(\tilde{b},\tilde{\sigma})}, (\vartheta^{(\tilde{b},\tilde{\sigma})})^\top)$$

are controlled paths in $\mathcal{V}^p_{(\cdot,W)}$, as, again, the sum and product of controlled paths remains a controlled path. Using the same arguments as above and combining the estimates (4.15) and (4.19), (4.16) and (4.20), we get that

$$\|\theta^{(b,\sigma)}\|_{\mathcal{V}^p_{(\cdot,W)}} \leq C, \tag{4.21}$$

where $C > 0$ depends only on p, k, d, M, and s_0; the same holds for $\theta^{(\tilde{b},\tilde{\sigma})}$; and

$$\|\theta^{(b,\sigma)};\theta^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|b-\tilde{b}\|_{C_b^2}+\|\sigma-\tilde{\sigma}\|_{C_b^2}, \tag{4.22}$$

where the implicit multiplicative constant depends only on p, k, d, M, and s_0.

Define the rough integrals $U^{(b,\sigma)} := \int_0^\cdot \theta_t^{(b,\sigma)}\,\mathrm{d}(\cdot,\mathbf{W})_t$ and $U^{(\tilde{b},\tilde{\sigma})} := \int_0^\cdot \theta_t^{(\tilde{b},\tilde{\sigma})}\,\mathrm{d}(\cdot,\mathbf{W})_t$, which are controlled paths in $\mathcal{V}^p_{(\cdot,W)}$. Using the estimate (4.5) for the rough integral and the estimate (4.21), it holds that

$$\|U^{(b,\sigma)}\|_{\mathcal{V}^p_{(\cdot,W)}}\leq C, \tag{4.23}$$

where $C > 0$ depends only on p, k, d, M, and s_0; the same holds for $U^{(\tilde{b},\tilde{\sigma})}$. Consequently, we get

$$\|U^{(b,\sigma)}\|_\infty\leq C_0, \tag{4.24}$$

where $C_0 > 0$ depends only on p, k, d, M, and s_0. Furthermore, using the stability of rough integrals, e.g., [75, Lemma 3.4], and the estimate (4.22), it immediately follows that

$$\|U^{(b,\sigma)};U^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|b-\tilde{b}\|_{C_b^2}+\|\sigma-\tilde{\sigma}\|_{C_b^2}, \tag{4.25}$$

where the implicit multiplicative constant depends only on p, k, d, M, and s_0.

Step 6. Proceeding as in the proof of Lemma 4.3.4, it follows from Lemma A.4.7 that for $t \in [0,T]$,

$$\begin{aligned}
\mathcal{E}(Z^{(b,\sigma)})_t &= \exp(Z_t^{(b,\sigma)} - \frac{1}{2}[\mathbf{Z}^{(b,\sigma)}]_t) \\
&= \exp\Big(\frac{1}{2}\int_0^t (\vartheta_s^{(b,\sigma)})^\top \vartheta_s^{(b,\sigma)}\,\mathrm{d}s + \int_0^t (\vartheta_s^{(b,\sigma)})^\top\,\mathrm{d}\mathbf{W}_s\Big) \\
&= \exp(U_t^{(b,\sigma)}).
\end{aligned}$$

By Lemma 4.3.3, we have that

$$\kappa_t^{(b,\sigma)} := \frac{1}{K_T}\mathcal{E}(Z^{(b,\sigma)})_t, \qquad \kappa_t^{(\tilde{b},\tilde{\sigma})} := \frac{1}{K_T}\mathcal{E}(Z^{(\tilde{b},\tilde{\sigma})})_t$$

are the pathwise defined optimal consumption rates for the log-utility on the financial market modeled by $S^{(b,\sigma)}$ and $S^{(\tilde{b},\tilde{\sigma})}$, respectively, and are controlled paths in $\mathcal{V}^p_{(\cdot,W)}$. We therefore get with (4.23) and (4.24) that

$$\|\kappa^{(b,\sigma)}\|_{\mathcal{V}^p_{(\cdot,W)}}\leq C, \tag{4.26}$$

where $C > 0$ depends only on p, k, d, M, s_0, $\|\exp\|_{C_b^2(\{y:|y|\leq C_0\};\mathbb{R})}$, and the consumption clock K, as it is a composition of a controlled path with a regular function; see, e.g., [75, Lemma 3.5]. The same holds for $\kappa^{(\tilde{b},\tilde{\sigma})}$.

Because the composition of a controlled path with a regular function is locally Lipschitz continuous (see, e.g., [75, Lemma 3.5]), it follows with (4.25) that

$$\|\kappa^{(b,\sigma)};\kappa^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|b-\tilde{b}\|_{C_b^2}+\|\sigma-\tilde{\sigma}\|_{C_b^2}, \tag{4.27}$$

where the implicit multiplicative constant depends only on p, k, d, M, s_0, and K.

Step 7. Since K_t, $t\in[0,T]$, is a càdlàg (deterministic) and increasing function (so of finite 1-variation), we recall that by Lemma A.4.3, the wealth process $V_t^{(b,\sigma)} := \kappa_t^{(b,\sigma)}(K_T - K_t)$, $t\in[0,T]$, (as the product of two controlled paths) is a controlled path in $\mathcal{V}^p_{(\cdot,W)}$. One can deduce with (4.26) that

$$\|V^{(b,\sigma)}\|_{\mathcal{V}^p_{(\cdot,W)}}\leq C, \tag{4.28}$$

where $C>0$ depends only on p, k, d, M, s_0, and K.

By Lemma A.4.4,

$$\|V^{(b,\sigma)};V^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|\kappa^{(b,\sigma)};\kappa^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}},$$

where the implicit multiplicative constant depends only on p, M, K, $\|\kappa^{(b,\sigma)}\|_{\mathcal{V}^p_{(\cdot,W)}}$ and $\|\kappa^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}}$. Combining this with (4.26) and (4.27), it holds that

$$\|V^{(b,\sigma)};V^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|b-\tilde{b}\|_{C_b^2}+\|\sigma-\tilde{\sigma}\|_{C_b^2}, \tag{4.29}$$

where the implicit multiplicative constant depends only on p, k, d, M, s_0, and K.

Step 8. By Lemma A.4.3, $\varphi_t^{(b,\sigma),i} := H_t^{(b,\sigma),i}V_{t-}^{(b,\sigma)}$, $\varphi_t^{(\tilde{b},\tilde{\sigma}),i} := H_t^{(\tilde{b},\tilde{\sigma}),i}V_{t-}^{(\tilde{b},\tilde{\sigma})}$, $i=1,\dots,k$, are controlled paths in $\mathcal{V}^p_{(\cdot,W)}$, and

$$\|\varphi^{(b,\sigma),i}\|_{\mathcal{V}^p_{(\cdot,W)}}\leq C, \tag{4.30}$$

where $C>0$ depends only on p, M, $\|H^{(b,\sigma),i}\|_{\mathcal{V}^p_{(\cdot,W)}}$, $\|V^{(b,\sigma)}\|_{\mathcal{V}^p_{(\cdot,W)}}$, that is, only on p, k, d, M, s_0, and K, see (4.19) and (4.28); the same holds for $\varphi^{(\tilde{b},\tilde{\sigma}),i}$. By Lemma A.4.4,

$$\|\varphi^{(b,\sigma),i};\varphi^{(\tilde{b},\tilde{\sigma}),i}\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|H^{(b,\sigma),i};H^{(\tilde{b},\tilde{\sigma}),i}\|_{\mathcal{V}^p_{(\cdot,W)}}+\|V^{(b,\sigma),i};V^{(\tilde{b},\tilde{\sigma}),i}\|_{\mathcal{V}^p_{(\cdot,W)}},$$

where the implicit multiplicative constant depends only on p, M, $\|H^{(b,\sigma),i}\|_{\mathcal{V}^p_{(\cdot,W)}}$, $\|H^{(\tilde{b},\tilde{\sigma}),i}\|_{\mathcal{V}^p_{(\cdot,W)}}$, $\|V^{(b,\sigma)}\|_{\mathcal{V}^p_{(\cdot,W)}}$, $\|V^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}}$. This gives with (4.19) and (4.28), (4.20) and (4.29) that

$$\|\varphi^{(b,\sigma),i};\varphi^{(\tilde{b},\tilde{\sigma}),i}\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|b-\tilde{b}\|_{C_b^2}+\|\sigma-\tilde{\sigma}\|_{C_b^2}, \tag{4.31}$$

where the implicit multiplicative constant depends only on p, k, d, M, s_0, and K. Finally, we consider

$$\varphi_t^{(b,\sigma),0} = \sum_{i=1}^k \int_0^t \varphi_s^{(b,\sigma),i}\,\mathrm{d}S_s^{(b,\sigma),i} - \varphi_t^{(b,\sigma),i}S_t^{(b,\sigma),i}$$

and

$$\varphi_t^{(\tilde{b},\tilde{\sigma}),0} = \sum_{i=1}^{k} \int_0^t \varphi_s^{(\tilde{b},\tilde{\sigma}),i} \,\mathrm{d}S_s^{(\tilde{b},\tilde{\sigma}),i} - \varphi_t^{(\tilde{b},\tilde{\sigma}),i} S_t^{(\tilde{b},\tilde{\sigma}),i},$$

for $t \in [0,T]$. By the associativity property of rough integrals, it holds that

$$\begin{aligned}\int_0^{\cdot} \varphi_t^{(b,\sigma),i} \,\mathrm{d}S_t^{(b,\sigma),i} &= \int_0^{\cdot} (\varphi_t^{(b,\sigma),i} b(t, S_t^{(b,\sigma)})^i, \varphi_t^{(b,\sigma),i} \sigma(t, S_t^{(b,\sigma)})^{i\cdot}) \,\mathrm{d}(\cdot, \mathbf{W})_t \\ &=: \int_0^{\cdot} \psi_t^{(b,\sigma),i} \,\mathrm{d}(\cdot, \mathbf{W})_t,\end{aligned}$$

similarly for $\int_0^{\cdot} \varphi_t^{(\tilde{b},\tilde{\sigma}),i} \,\mathrm{d}S_t^{(\tilde{b},\tilde{\sigma}),i}$. Using the same arguments as above, by (4.15) and (4.30), (4.16) and (4.31), it holds that

$$\|\psi^{(b,\sigma),i}; \psi^{(\tilde{b},\tilde{\sigma}),i}\|_{\mathcal{V}^p_{(\cdot,W)}} \lesssim \|b - \tilde{b}\|_{C_b^2} + \|\sigma - \tilde{\sigma}\|_{C_b^2},$$

where the implicit multiplicative constant depends only on p, k, d, M, s_0, and K. Therefore, using the stability of rough integrals(e.g., [75, Lemma 3.4]), Lemma A.4.4, and the estimates (4.30), (4.31) and (4.14), we can derive that

$$\|\varphi^{(b,\sigma),0}; \varphi^{(\tilde{b},\tilde{\sigma}),0}\|_{\mathcal{V}^p_{(\cdot,W)}} \lesssim \|b - \tilde{b}\|_{C_b^2} + \|\sigma - \tilde{\sigma}\|_{C_b^2},$$

where the implicit multiplicative constant depends only on p, k, d, M, s_0, and K.

Hence, since we can bound the supremum norm by the controlled path norm, see (4.2), the (local) Lipschitz continuity for optimal portfolios and wealth processes follows. □

4.3.3 Discretization error of pathwise log-optimal portfolios

To implement the pathwise log-optimal portfolio on a real financial market would require to trade continuously in time. In reality, trading might be done at a very high frequency but still on a discrete time grid and, thus, requires a discretization of any theoretically optimal portfolio.

In this subsection, we introduce a time-discrete version of the pathwise log-optimal portfolio, as constructed in Lemma 4.3.3, and derive quantitative, pathwise error estimates resulting from this discretization for the portfolios as well as for their associated capital processes.

To define the time-discrete version of the pathwise log-optimal portfolio, we start by discretizing the underlying price paths. To that end, we recall that W and the sequence $(\mathcal{P}^n)$ of partitions satisfy Assumption 4.3.1, where $\mathcal{P}^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$. For $n \in \mathbb{N}$, let $W^n \colon [0,T] \to \mathbb{R}^d$ be the piecewise constant approximation of W along $\mathcal{P}^n$, that is,

$$W_t^n := W_T \mathbf{1}_{\{T\}}(t) + \sum_{k=0}^{N_n-1} W_{t_k^n} \mathbf{1}_{[t_k^n, t_{k+1}^n)}(t), \qquad t \in [0,T],$$

and, setting $\gamma_t := t$, we define a time discretization path along $(\mathcal{P}^n)$ by

$$\gamma_t^n := T\mathbf{1}_{\{T\}}(t) + \sum_{k=0}^{N_n-1} t_k^n \mathbf{1}_{[t_k^n, t_{k+1}^n)}(t), \qquad t \in [0,T].$$

To discretize the price path S, we use the classical Euler approximation S^n corresponding to the RDE (4.13) along the partition $\mathcal{P}^n$, which is given by

$$S_t^n = s_0 + \sum_{i: t_{i+1}^n \le t} b(t_i^n, S_{t_i^n}^n)(t_{i+1}^n - t_i^n) + \sum_{i: t_{i+1}^n \le t} \sigma(t_i^n, S_{t_i^n}^n)(W_{t_{i+1}^n} - W_{t_i^n}), \quad t \in [0,T]. \tag{4.32}$$

The time-discrete version (φ^n, κ^n) of the pathwise log-optimal portfolio is defined by

$$\begin{aligned}
&H_t^n := (\sigma(\gamma_t^n, S_t^n)\sigma(\gamma_t^n, S_t^n)^\top)^{-1} b(\gamma_t^n, S_t^n), \qquad \vartheta_t^n := \sigma(\gamma_t^n, S_t^n)^\top H_t^n, \\
&\theta_t^n := (\frac{1}{2}(\vartheta_t^n)^\top \vartheta_t^n, (\vartheta_t^n)^\top), \\
&\kappa_t^n := \frac{1}{K_T} \exp\bigg(\int_0^t \theta_s^n \,\mathrm{d}(\gamma^n, W^n)_s\bigg), \qquad v_t^n := \kappa_t^n (K_T - K_t^n), \\
&\varphi_t^{n,i} := H_t^{n,i} v_t^n, \qquad i = 1, \dots, k, \qquad \varphi_t^{n,0} := \sum_{i=1}^k \int_0^t \varphi_s^{n,i} \,\mathrm{d}S_s^{n,i} - \varphi_t^{n,i} S_t^{n,i}, \\
&V_t^n := 1 + \sum_{i=1}^k \int_0^t \varphi_s^{n,i} \,\mathrm{d}S_s^{n,i} - \int_0^t \kappa_s^n \,\mathrm{d}K_s, \qquad t \in [0,T],
\end{aligned}$$

where all above integrals are just left-point Riemann sums and K^n denotes the piecewise constant approximation of K along $\mathcal{P}^n$. For these time-discrete portfolios and their associated capital processes, we obtain the following convergence result with quantitative error estimates.

Theorem 4.3.7. *For $(b,\sigma) \in C_b^3([0,T] \times \mathbb{R}^k; \mathbb{R}^k) \times \mathcal{A}$, let $(\varphi^{(b,\sigma)}, \kappa^{(b,\sigma)})$ be the pathwise log-optimal portfolio, as constructed in Lemma 4.3.3. Then,*

$$\|(\varphi^n, \kappa^n) - (\varphi, \kappa)\|_{p'} \longrightarrow 0 \qquad \text{as} \qquad n \longrightarrow \infty$$

and

$$\|V^n - V\|_{p'} \longrightarrow 0 \qquad \text{as} \qquad n \longrightarrow \infty,$$

for any $p' \in (p,3)$, with a rate of convergence given by

$$\begin{aligned}
\|(\varphi^n, \kappa^n) - (\varphi, \kappa)\|_{p'} \lesssim |\mathcal{P}^n|^{1-\frac{1}{q}} + (|\mathcal{P}^n| + \|W^n - W\|_\infty)^{1-\frac{p}{p'}} \\
+ \bigg(|\mathcal{P}^n|^{1-\frac{1}{q}} + \bigg\| \int_0^\cdot W_t^n \otimes \mathrm{d}W_t - \int_0^\cdot W_t \otimes \mathrm{d}W_t \bigg\|_\infty \bigg)^{1-\frac{p}{p'}}
\end{aligned}$$

and

$$\|V^n - V\|_{p'} \lesssim |\mathcal{P}^n|^{1-\frac{1}{q}} + (|\mathcal{P}^n| + \|W^n - W\|_\infty)^{1-\frac{p}{p'}} \\ + \left(|\mathcal{P}^n|^{1-\frac{1}{q}} + \left\| \int_0^\cdot W_t^n \otimes \mathrm{d}W_t - \int_0^\cdot W_t \otimes \mathrm{d}W_t \right\|_\infty \right)^{1-\frac{p}{p'}},$$

for any $q \in (1,2)$ *such that* $\frac{1}{p'} + \frac{1}{q} > 1$*, where the implicit multiplicative constant depends only on* p*,* p'*,* q*,* k*,* d*,* $\|b\|_{C_b^3}$*,* $\|\sigma\|_{C_b^3}$*,* $1/\inf_{(t,x)}|\det(\sigma(t,x)\sigma(t,x)^\top)|$*,* T*,* s_0*,* $\|W\|_p$ *and* $w(0,T)$*, where* w *is the control function for which* (4.6) *holds for* $(\cdot, W)$*.*

Remark 4.3.8. *The convergence results and quantitative estimates in Theorem 4.3.7 hold true when replacing the* p'*-variation seminorm* $\|\cdot\|_{p'}$ *by the supremum seminorm* $\|\cdot\|_\infty$*.*

Remark 4.3.9. *The rate of convergence provided in Theorem 4.3.7 does agree with the pathwise rate of convergence of the Euler scheme for the RDE* (4.13)*, as obtained in Theorem 3.1.2. In other words, the rate of convergence for the pathwise log-optimal portfolio appears to be as good as the rate of convergence of the Euler scheme* (4.32)*.*

Before we present the proof, some preliminary steps are necessary. We start by noting that, as W^n has finite 1-variation, W^n possesses a canonical rough path lift $\mathbf{W}^n = (W^n, \mathbb{W}^n) \in \mathcal{D}^p([0,T], \mathbb{R}^d)$, with $\mathbb{W}^n$ given by

$$\mathbb{W}^n_{s,t} := \int_s^t W^n_{s,u} \otimes \mathrm{d}W^n_u, \qquad (s,t) \in \Delta_T,$$

where the integral is defined as a classical limit of left-point Riemann sums. Similarly, we can define a time-space rough path $(\cdot, \mathbf{W})^n$ of $(\cdot, W)^n := (\gamma^n, W^n)$.

Lemma 4.3.10. *There exists a constant* $C > 0$*, which depends only on* p*,* T*,* $\|W\|_p$ *and* $w(0,T)$*, where* w *is the control function for which* (4.6) *holds for* $(\cdot, W)$*, such that*

$$\|(\cdot, \mathbf{W})^n\|_p + \|(\cdot, \mathbf{W})\|_p \leq C,$$

for every $n \in \mathbb{N}$*. For any* $p' \in (p, 3)$*, we have that*

$$\|(\cdot, \mathbf{W})^n; (\cdot, \mathbf{W})\|_{p'} \longrightarrow 0 \qquad as \qquad n \longrightarrow \infty,$$

with a rate of convergence given by

$$\|(\cdot, \mathbf{W})^n; (\cdot, \mathbf{W})\|_{p'} \lesssim (|\mathcal{P}^n| + \|W^n - W\|_\infty)^{1-\frac{p}{p'}} \\ + \left(|\mathcal{P}^n|^{1-\frac{1}{q}} + \left\| \int_0^\cdot W_t^n \otimes \mathrm{d}W_t - \int_0^\cdot W_t \otimes \mathrm{d}W_t \right\|_\infty \right)^{1-\frac{p}{p'}},$$

for any $q \in (1,2)$ *such that* $\frac{1}{p'} + \frac{1}{q} > 1$*, where the implicit multiplicative constant depends only on* p*,* p'*,* q*,* d*,* T *and* $\|W\|_\infty$*,* $\|W\|_p$ *and* $w(0,T)$*.*

Proof. As stated in the proof of [7, Lemma 2.13],

$$\|(\cdot,\mathbf{W})\|_p \leq C,$$

for some $C>0$ depending only on p, T, $\|W\|_p$ and $w(0,T)$, where w is the control function for which (4.6) holds for $(\cdot,W)$. Applying Lemma 3.1.5, we get, for each $n\in\mathbb{N}$, that

$$\|(\cdot,\mathbf{W})^n\|_p \leq C,$$

for some $C>0$ depending on p, T, $\|W\|_p$ and $w(0,T)$, but not on n.

Further, by Lemma 3.1.6, it holds that

$$\begin{aligned}
&\|(\cdot,\mathbf{W})^n;(\cdot,\mathbf{W})\|_{p'}\\
&\quad\lesssim \|(\gamma^n,W^n)-(\gamma,W)\|_\infty^{1-\frac{p}{p'}}\\
&\qquad+\sup_{(s,t)\in\Delta_T}\left|\int_s^t(\gamma^n,W^n)_{s,u}\otimes \mathrm{d}(\gamma,W)_u-\int_s^t(\gamma,W)_{s,u}\otimes \mathrm{d}(\gamma,W)_u\right|^{1-\frac{p}{p'}}\\
&\quad\lesssim \|(\gamma^n,W^n)-(\gamma,W)\|_\infty^{1-\frac{p}{p'}}+\left\|\int_0^\cdot(\gamma^n,W^n)_t\otimes \mathrm{d}(\gamma,W)_t-\int_0^\cdot(\gamma,W)_t\otimes \mathrm{d}(\gamma,W)_t\right\|_\infty^{1-\frac{p}{p'}},
\end{aligned}$$

where the implicit multiplicative constant depends only on p, p', $\|(\cdot,W)\|_\infty$, $\|(\cdot,\mathbf{W})\|_p$ and $w(0,T)$, where w is the control function for which (4.6) holds for $(\cdot,W)$. Therefore, in view of the above bound on $\|(\cdot,\mathbf{W})\|_p$, this constant in fact depends on p, p', T, $\|W\|_\infty$, $\|W\|_p$ and $w(0,T)$.

It is straightforward to see that

$$\|(\gamma^n,W^n)-(\gamma,W)\|_\infty \lesssim |\mathcal{P}^n|+\|W^n-W\|_\infty, \qquad n\in\mathbb{N}.$$

Let $q\in(1,2)$ such that $\frac{1}{p'}+\frac{1}{q}>1$. We note that γ^n and γ have finite 1-variation, with $\|\gamma^n\|_1=\|\gamma\|_1=T$. It follows by interpolation that

$$\|\gamma^n-\gamma\|_q \leq (2T)^{\frac{1}{q}}|\mathcal{P}^n|^{1-\frac{1}{q}}.$$

By the standard estimate for Young integrals—see, e.g., [75, Proposition 2.4]—we then have, for all $t\in[0,T]$, that

$$\left|\int_0^t\gamma_u^n\,\mathrm{d}\gamma_u-\int_0^t\gamma_u\,\mathrm{d}\gamma_u\right|\lesssim\|\gamma^n-\gamma\|_q\|\gamma\|_1\leq(2T)^{\frac{1}{q}}|\mathcal{P}^n|^{1-\frac{1}{q}}T.$$

Similarly, for each $t\in[0,T]$, it holds that

$$\left|\int_0^t\gamma_u^n\,\mathrm{d}W_u^j-\int_0^t\gamma_u\,\mathrm{d}W_u^j\right|\lesssim\|\gamma^n-\gamma\|_q\|W\|_p\lesssim(2T)^{\frac{1}{q}}|\mathcal{P}^n|^{1-\frac{1}{q}}\|W\|_p,$$

and

$$\left|\int_0^t W_u^{n,j}\,\mathrm{d}\gamma_u-\int_0^t W_u^j\,\mathrm{d}\gamma_u\right|\lesssim\|W^n-W\|_p\|\gamma\|_1\lesssim T|\mathcal{P}^n|^{\frac{1}{p}}.$$

Combining the estimates, we obtain the desired rate of convergence. $\square$

It has been established in the proof of Theorem 3.1.2 that the Euler scheme (4.32) and the RDE

$$\widetilde{S}^n_t = s_0 + \int_0^t b(\gamma^n_s, \widetilde{S}^n_s)\,\mathrm{d}\gamma^n_s + \int_0^t \sigma(\gamma^n_s, \widetilde{S}^n_s)\,\mathrm{d}\mathbf{W}^n_s = s_0 + \int_0^t (b,\sigma)(\gamma^n_s, \widetilde{S}^n_s)\,\mathrm{d}(\cdot, \mathbf{W})^n_s, \quad t \in [0,T],$$

coincide. Furthermore, proceeding as in the proof of Lemma 4.3.3, one can show that H^n, θ^n, κ^n, φ^n are controlled paths in $\mathcal{V}^p_{(\cdot,W)^n}$, and $\int_0^t \varphi^{n,i}_s \,\mathrm{d}S^{n,i}_s$ is a rough integral defined as in Lemma A.4.1.

Let $\widetilde{\mathcal{P}}^m = \{0 = r^m_0 < r^m_1 < \cdots < r^m_{\widetilde{N}_m} = T\}$, $m \in \mathbb{N}$, be any sequence of partitions with mesh size converging to 0, such that $\mathcal{P}^n \subseteq \widetilde{\mathcal{P}}^m$ for every $m \in \mathbb{N}$. In much the same way as in the proof of Theorem 3.1.2, one can show that the rough integral $\int_0^t \theta^n_s \,\mathrm{d}(\cdot, \mathbf{W})^n_s$ is equal to a limit of left-point Riemann sums along the sequence $(\widetilde{\mathcal{P}}^m)_{m\in\mathbb{N}}$. That is, for any $t \in [0,T]$, we have

$$\begin{aligned}
\int_0^t \theta^n_s \,\mathrm{d}(\gamma^n, W^n)_s &= \sum_{k=0}^{N_n-1} \theta^n_{t^n_k} (\gamma^n, W^n)_{t^n_k\wedge t, t^n_{k+1}\wedge t} \\
&= \lim_{m\to\infty} \sum_{k=0}^{\widetilde{N}_m-1} \theta^n_{r^m_k} (\gamma^n, W^n)_{r^m_k\wedge t, r^m_{k+1}\wedge t} = \int_0^t \theta^n_s \,\mathrm{d}(\cdot, \mathbf{W})^n_s
\end{aligned}$$

and, consequently, we have

$$\kappa^n_t := \frac{1}{K_T} \exp\bigg(\int_0^t \theta^n_s \,\mathrm{d}(\gamma^n, W^n)_s \bigg) = \frac{1}{K_T} \exp\bigg(\int_0^t \theta^n_s \,\mathrm{d}(\cdot, \mathbf{W})^n_s \bigg).$$

Similarly, applying the associativity property of rough integrals and Theorem 4.2.1, for any $t \in [0,T]$, it holds that the rough integral of the controlled path $\varphi^{n,i}$ against the controlled path $S^{n,i}$ is given by left-point Riemann sums, and so is the integral of κ^n against K since the paths are of finite 1-variation. Now we are finally able to prove Theorem 4.3.7.

Proof of Theorem 4.3.7. Using part (i) of Corollary 2.2.3 and the estimate in Lemma 4.3.10, we deduce that there exists a constant $L > 0$ depending only on p', T, s_0, $\|b\|_{C^2_b}$, $\|\sigma\|_{C^2_b}$, $\|W\|_p$ and $w(0,T)$ such that $\sup_{n\in\mathbb{N}} \|S^n\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}}, \|S\|_{\mathcal{V}^{p'}_{(\cdot,W)}} \leq L$. Further, Theorem 3.1.1 gives that

$$\|S^n; S\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}, \mathcal{V}^{p'}_{(\cdot,W)}} \lesssim \|\gamma^n - \gamma\|_q + \|(\cdot, \mathbf{W})^n; (\cdot, \mathbf{W})\|_{p'} \lesssim |\mathcal{P}^n|^{1-\frac{1}{q}} + \|(\cdot, \mathbf{W})^n; (\cdot, \mathbf{W})\|_{p'},$$

for any $q \in (1,2)$ such that $\frac{1}{p'} + \frac{1}{q} > 1$, where the implicit multiplicative constant depends only on p', q, T, $\|b\|_{C^3_b}$, $\|\sigma\|_{C^3_b}$, $\|W\|_p$ and $w(0,T)$.

We note that γ^n, γ are controlled paths with respect to $(\cdot, W)$, with $\|\gamma^n\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}} = \|\gamma\|_{\mathcal{V}^{p'}_{(\cdot,W)}} = 1$ and $\|\gamma^n; \gamma\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}, \mathcal{V}^{p'}_{(\cdot,W)}} = 0$. Since the composition of controlled paths with regular composition remains a controlled path and such a composition is locally Lipschitz continuous, it therefore holds that

$$\|(b,\sigma)(\gamma^n, S^n)\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}} + \|(b,\sigma)(\gamma, S)\|_{\mathcal{V}^{p'}_{(\cdot,W)}} \leq C,$$

where $C > 0$ depends only on p', T, $\|b\|_{C_b^2}$, $\|\sigma\|_{C_b^2}$, L and $\|W\|_p$, see, e.g. [75, Lemma 3.5], and consequently

$$\begin{aligned}
&\|(b,\sigma)(\gamma, S); (b,\sigma)(\gamma^n, S^n)\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}, \mathcal{V}^{p'}_{(\cdot,W)}} \\
&\quad\lesssim \|(\gamma^n, S^n); (\gamma, S)\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}, \mathcal{V}^{p'}_{(\cdot,W)}} + \|(\cdot, W)^n - (\cdot, W)\|_{p'} \\
&\quad\lesssim \|\gamma^n; \gamma\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}, \mathcal{V}^{p'}_{(\cdot,W)}} + \|S^n; S\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}, \mathcal{V}^{p'}_{(\cdot,W)}} + \|(\cdot, W)^n - (\cdot, W)\|_{p'} \\
&\quad\lesssim |\mathcal{P}^n|^{1-\frac{1}{q}} + \|(\cdot, \mathbf{W})^n; (\cdot, \mathbf{W})\|_{p'},
\end{aligned}$$

where the implicit multiplicative constant depends only on p', q, T, $\|b\|_{C_b^2}$, $\|\sigma\|_{C_b^2}$, L, $\|W\|_p$ and $w(0,T)$.

Following the arguments of the proof of Theorem 4.3.6 from Step 3 on and applying the above estimates, one can derive that

$$\|(\varphi^n, \kappa^n)\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}} + \|(\varphi, \kappa)\|_{\mathcal{V}^{p'}_{(\cdot,W)}} \leq C,$$

and

$$\|V\|_{\mathcal{V}^{p'}_{(\cdot,W)}} \leq C,$$

where $C > 0$ depends only on p', k, d, T, $\|b\|_{C_b^2}$, $\|\sigma\|_{C_b^2}$, L, $\|W\|_p$, $w(0,T)$ and the consumption clock K. Using standard estimates for Young and rough integrals, see, e.g., (4.5) and [75, Proposition 3.4], we also obtain that

$$\|V^n\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}} \leq C.$$

Further, since κ^n and φ^n depend on (the composition of regular functions with, or products of) controlled paths of the form $(\sigma(\gamma_t^n, S_t^n)\sigma(\gamma_t^n, S_t^n)^\top)^{-1} b(\gamma_t^n, S_t^n)$ and $\sigma(\gamma_t^n, S_t^n)$; κ and φ depend on (the composition of regular functions with, or products of) controlled paths of the form $(\sigma(\gamma_t, S_t)\sigma(\gamma_t, S_t)^\top)^{-1} b(\gamma_t, S_t)$ and $\sigma(\gamma_t, S_t)$, using the above bound on $\|(b,\sigma)(\gamma, S); (b,\sigma)(\gamma^n, S^n)\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}, \mathcal{V}^{p'}_{(\cdot,W)}}$, we can obtain that

$$\|(\varphi^n, \kappa^n), (\varphi, \kappa)\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}, \mathcal{V}^{p'}_{(\cdot,W)}} \lesssim |\mathcal{P}^n|^{1-\frac{1}{q}} + \|(\cdot, \mathbf{W}); (\cdot, \mathbf{W})^n\|_{p'},$$

where the implicit multiplicative constant depends only on p', q, k, d, T, $\|b\|_{C_b^3}$, $\|\sigma\|_{C_b^3}$, $1/\inf_{(t,x)}|\det(\sigma(t,x)\sigma(t,x)^\top)|$, s_0, $\|W\|_p$, $w(0,T)$ and the consumption clock K. Applying standard estimates for Young and rough integrals (e.g., [75, Proposition 2.4, Lemma 3.1 and Lemma 3.7], we can also obtain that

$$\|V^n;V\|_{\mathcal{V}^{p'}_{(\cdot,W)^n},\mathcal{V}^{p'}_{(\cdot,W)}}\lesssim|\mathcal{P}^n|^{1-\frac{1}{q}}+\|(\cdot,\mathbf{W});(\cdot,\mathbf{W})^n\|_{p'}.$$

Hence, since we can bound the p-norm by the controlled path norm, see (4.3), combining this with the rate of convergence stated in Lemma 4.3.10, we infer the convergence and the estimate. □

If we assume stronger regularity properties of the "driving noise" path W and the sequence of partitions, we can make the quantitative estimates, provided in Theorem 4.3.7, more explicit. For instance, considering the regularity properties of Brownian sample paths, we can derive the following two corollaries.

Corollary 4.3.11. *Let $p \in (2,3)$ and $(\mathcal{P}^n_U)_{n\in\mathbb{N}}$ be the sequence of equidistant partitions $(\mathcal{P}^n_U)_{n\in\mathbb{N}}$ of $[0,T]$ with width $\frac{T}{n}$. Let W be a $\frac{1}{p}$-Hölder continuous path satisfying Assumption 4.3.1 relative to p and $(\mathcal{P}^n_U)_{n\in\mathbb{N}}$, and*

$$\left\|\int_0^\cdot W^n_t\otimes \mathrm{d}W_t-\int_0^\cdot W_t\otimes \mathrm{d}W_t\right\|_\infty\lesssim n^{-(\frac{2}{p}-\beta)},\qquad n\in\mathbb{N},\tag{4.33}$$

for $\beta\in(1-\frac{1}{p},\frac{2}{p})$. Then for any $p'\in(p,3)$, $q\in(1,2)$ such that $\frac{1}{p'}+\frac{1}{q}>1$, there exists a constant $C>0$, which does not depend on n, such that

$$\|(\varphi^n,\kappa^n)-(\varphi,\kappa)\|_{p'}\leq C(n^{-(1-\frac{1}{q})(1-\frac{p}{p'})}+n^{-(\frac{2}{p}-\beta)(1-\frac{p}{p'})}),\qquad n\in\mathbb{N},$$

and

$$\|V^n-V\|_{p'}\leq C(n^{-(1-\frac{1}{q})(1-\frac{p}{p'})}+n^{-(\frac{2}{p}-\beta)(1-\frac{p}{p'})}),\qquad n\in\mathbb{N}.$$

Proof. Since W is assumed to be $\frac{1}{p}$-Hölder continuous, we have that

$$\|W^n-W\|_\infty\lesssim n^{-\frac{1}{p}},\qquad n\in\mathbb{N}.$$

We may combine this with (4.33) and Theorem 4.3.7. Since $\frac{1}{p}<1-\frac{1}{p}<\beta$ for $p\in(2,3)$, this gives the claimed rate of convergence. □

Remark 4.3.12. *Almost all sample paths of a d-dimensional Brownian motion satisfy Property* (RIE) *relative to p and $(\mathcal{P}^n_U)_{n\in\mathbb{N}}$, as shown in Proposition 3.2.2, and, thus, Assumption 4.3.1 is satisfied if the sequence $(\mathcal{P}^n_U)_{n\in\mathbb{N}}$ of partitions exhausts the jumps of the consumption clock K. Moreover, by [143, Appendix B],* (4.33) *holds true almost surely for sample paths of a Brownian motion. Hence, the sample paths of a Brownian motion fulfill the conditions of Corollary 4.3.11 almost surely.*

Corollary 4.3.13. *Let $p \in (2,3)$ and $(\mathcal{P}_D^n)_{n\in\mathbb{N}}$ be the sequence of dyadic partitions of $[0,T]$, given by*

$$\mathcal{P}_D^n := \{0 = t_0^n < t_1^n < \cdots < t_{2^n}^n = T\} \quad \textit{with} \quad t_k^n := k2^{-n}T \quad \textit{for} \quad k = 0, 1, \ldots, 2^n.$$

Let W be a $\frac{1}{p}$-Hölder continuous path satisfying Assumption 4.3.1 relative to p and $(\mathcal{P}_D^n)_{n\in\mathbb{N}}$, and for any $\varepsilon \in (0,1)$,

$$\left\| \int_0^\cdot W_t^n \otimes \mathrm{d}W_t - \int_0^\cdot W_t \otimes \mathrm{d}W_t \right\|_\infty \lesssim 2^{-\frac{n}{2}(1-\varepsilon)}, \qquad n \in \mathbb{N}. \tag{4.34}$$

Then, for any $p' \in (p,3)$ and $q \in (1,2)$ such that $\frac{1}{p'} + \frac{1}{q} > 1$, and any $\varepsilon \in (0,1)$, there exists a constant $C > 0$, which does not depend on n, such that

$$\|(\varphi^n, \kappa^n) - (\varphi, \kappa)\|_{p'} \leq C(2^{-n(1-\frac{1}{q})(1-\frac{p}{p'})} + 2^{-n(\frac{1}{p}-\frac{1}{p'})} + 2^{-\frac{n}{2}(1-\varepsilon)(1-\frac{p}{p'})}), \qquad n \in \mathbb{N}$$

and

$$\|V^n - V\|_{p'} \leq C(2^{-n(1-\frac{1}{q})(1-\frac{p}{p'})} + 2^{-n(\frac{1}{p}-\frac{1}{p'})} + 2^{-\frac{n}{2}(1-\varepsilon)(1-\frac{p}{p'})}), \qquad n \in \mathbb{N}.$$

Proof. Since W is assumed to be $\frac{1}{p}$-Hölder continuous, we have that

$$\|W^n - W\|_\infty \lesssim 2^{-\frac{n}{p}}, \qquad n \in \mathbb{N}.$$

We may combine this with (4.34) and Theorem 4.3.7, which gives the claimed rate of convergence. □

Remark 4.3.14. *Almost all sample paths of a d-dimensional Brownian motion satisfy Property* (RIE) *relative to p and $(\mathcal{P}_D^n)_{n\in\mathbb{N}}$ almost surely, as shown in Proposition 3.2.6, and, thus, Assumption 4.3.1 is satisfied if the $(\mathcal{P}_U^n)_{n\in\mathbb{N}}$ exhausts the jumps of the consumption clock K. Moreover, as shown in the proof of part (ii) of Proposition 3.2.6, for any $\varepsilon \in (0,1)$, almost all sample paths of a Brownian motion fulfill for any $\varepsilon \in (0,1)$*

$$\left\| \int_0^\cdot W_t^n \otimes \mathrm{d}W_t - \int_0^\cdot W_t \otimes \mathrm{d}W_t \right\|_\infty \lesssim 2^{-\frac{n}{2}(1-\varepsilon)},$$

for all sufficiently large n. Hence, the sample paths of a Brownian motion fulfill the conditions of Corollary 4.3.13 almost surely.

4.4 Black–Scholes-type models: pathwise analysis of log-optimal portfolios

In this section, we shall study log-optimal portfolios for the investment-consumption problem, acting on deterministic price paths generated by Black–Scholes-type models, defined in a pathwise manner.

This section is structured the same as Section 4.3 for readability. We point out that similar arguments apply and the method of proof carries over but due to the unboundedness of the coefficients, we need to treat this case separately.

To this end, we again make the standing Assumption 4.3.1, which is recalled by the following:

Assumption. *Let $p \in (2,3)$ and let $\mathcal{P}^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$, be a sequence of equidistant partitions of the interval $[0,T]$, such that,*

- *for each $n \in \mathbb{N}$, there exists some $\pi_n > 0$ such that $t_{i+1}^n - t_i^n = \pi_n$, for each $0 \le i < N_n$,*
- *$|\mathcal{P}^n|^{2-\frac{4}{p}} \log(n) \to 0$ as $n \to \infty$,*
- *$J_K \subseteq \liminf_{n\to\infty} \mathcal{P}^n$ with $J_K := \{t \in (0,T] : K_{t-} \neq K_t\}$,*

where the consumption clock $K\colon [0,T] \to \mathbb{R}$ is fixed as in Section 4.1.1. Moreover, the deterministic path $W\colon [0,T] \to \mathbb{R}^d$ satisfies Property (RIE) *relative to p and $(\mathcal{P}^n)_{n\in\mathbb{N}}$.*

We suppose that the discounted price path $(S_t)_{t\in[0,T]}$ satisfies the linear rough differential equations

$$S_t^i = s_0^i + \int_0^t S_s^i b_s^i \,\mathrm{d}s + \int_0^t S_s^i \sigma_s^{i\cdot} \,\mathrm{d}\mathbf{W}_s, \qquad t \in [0,T], \qquad i = 1, \dots, k, \tag{4.35}$$

where $s_0 \in \mathbb{R}^k$ and $b^i, \sigma^{i\cdot}$ are controlled paths with respect to W, more precisely, $b^i \in \mathcal{V}_W^p([0,T];\mathbb{R})$, $\sigma^{i\cdot} \in \mathcal{V}_W^p([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}))$, and $\mathbf{W} = (W, \mathbb{W})$ is the canonical rough path lift of W as defined in (4.7).

By Lemma A.4.6 and Proposition A.4.5, this is equivalent to solving the linear rough differential equations

$$S_t^i = s_0^i + \int_0^t S_s^i \,\mathrm{d}\mathbf{\Xi}_s^i, \qquad t \in [0,T], \qquad i = 1, \dots, k, \tag{4.36}$$

where $\Xi^i := \int_0^\cdot b_t^i \,\mathrm{d}t + \int_0^\cdot \sigma_t^{i\cdot} \,\mathrm{d}\mathbf{W}_t$, which is a controlled path in $\mathcal{V}_W^p$ and thus, admits a canonical rough path lift $\mathbf{\Xi}^i$ by Lemma A.4.6. Particularly, by Lemma A.4.7, we obtain that the solution is given by the rough exponential $S^i = \mathcal{E}(\Xi^i)$.

Remark 4.4.1. *If W is a realization of a Brownian motion, the dynamics of the RDE* (4.35) *can be seen as a fixed realization of a Black–Scholes-type model for a financial market.*

Indeed, let us assume that $\bar{W} = (\bar{W}_t)_{t\in[0,T]}$ is a d-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with respect to an underlying filtration $(\mathcal{F}_t)_{t\in[0,T]}$. It is well-known that the linear stochastic differential equations

$$\bar{S}_t^i = s_0^i + \int_0^t \bar{S}_s^i \bar{b}_s^i \,\mathrm{d}s + \int_0^t \bar{S}_s^i \bar{\sigma}_s^{i\cdot} \,\mathrm{d}\bar{W}_s, \qquad t \in [0,T], \qquad i = 1, \dots, k, \tag{4.37}$$

have a unique strong solution, where $\bar{b}^i$, $\bar{\sigma}^{ij}$ *are predictable processes,* $\sigma^{i\cdot} = (\sigma^{i1}, \dots, \sigma^{id})$, $i = 1, \dots, k$, $j = 1, \dots, d$, *and* $\bar{b}, \bar{\sigma}$ *are controlled paths with respect to* $\bar{W}$ *and* $\bar{\sigma}\bar{\sigma}^\top \in \mathrm{GL}(\mathbb{R}^{k \times k})$, *almost surely, and* $\int_0^t \bar{S}_s^i \bar{\sigma}_s^{i\cdot} \,\mathrm{d}\bar{W}_s$ *denotes a stochastic Itô integral; see, e.g., [147, Chapter V, Theorem 6]. Note that this model* $(\bar{S}_t)_{t \in [0,T]}$ *includes the classical Black–Scholes model, and some stochastic volatility models, where the volatility is modeled by an SDE driven by* $\bar{W}$.

The SDE (4.37) *can be solved explicitly, and its solution is given by the stochastic exponential* $\bar{S}^i = s_0^i \mathcal{E}(\bar{\Xi}^i)$, *for* $\bar{\Xi}^i := \int_0^\cdot \bar{b}_t^i \,\mathrm{d}t + \int_0^\cdot \bar{\sigma}_t^{i\cdot} \,\mathrm{d}\bar{W}_t$, *that is,*

$$\bar{S}_t^i = s_0^i \exp(\bar{\Xi}_t^i - \frac{1}{2}[\bar{\Xi}^i]_t), \qquad t \in [0,T], \qquad i = 1, \dots, k,$$

where $[\bar{\Xi}^i]$ *denotes the quadratic variation; see, e.g. [147, Chapter II, Theorem 37].*

Recall that for almost every $\omega \in \Omega$, *the sample path* $\bar{W}(\omega)$ *of a Brownian motion satisfies Property* (RIE) *relative to* p *and* $(\mathcal{P}^n)_{n \in \mathbb{N}}$; *see Remark 4.2.3. Hence, for almost every* $\omega \in \Omega$, $\Xi^i = \bar{\Xi}^i(\omega)$, *for* Ξ^i *defined as in* (4.36) *for the controlled paths* $b = \bar{b}(\omega)$, $\sigma = \bar{\sigma}(\omega)$ *and the rough path* $\mathbf{W} := \bar{\mathbf{W}}(\omega)$, *because the (random) rough integral and the Itô integral coincide; see Proposition 4.2.4. For almost every* $\omega \in \Omega$, Ξ^i *is a controlled path in* $\mathcal{V}_W^p$ *and thus, admits a canonical rough path lift* $\mathbf{\Xi}^i$.

Since for almost every $\omega \in \Omega$, $\Xi^i = \bar{\Xi}^i(\omega)$ *satisfies Property* (RIE) *by Theorem 4.2.2, it follows from [7, Proposition 2.18] that for almost every* $\omega \in \Omega$, *the quadratic variation* $[\bar{\Xi}^i](\omega)$ *and the rough path bracket* $[\mathbf{\Xi}^i]$ *coincide because* $[\mathbf{\Xi}^i]$ *coincides with the quadratic variation of* Ξ^i *along* $(\mathcal{P}^n)_{n \in \mathbb{N}}$ *in the sense of Föllmer.*

Finally, by Lemma A.4.7, for almost every $\omega \in \Omega$, *the stochastic exponential* $\bar{S}^i(\omega)$ *of* $\bar{\Xi}^i$ *and the rough exponential* $S = \bar{S}(\omega)$ *of* $\Xi^i = \bar{\Xi}^i(\omega)$, *which solves the linear rough differential equation* (4.36), *coincide. In other words,* $(S_t)_{t \in [0,T]}$ *can be understood as a fixed realization of the probabilistic model* $(\bar{S}_t)_{t \in [0,T]}$.

In the present setting, it will be convenient to equivalently reformulate the RDEs (4.35) into the RDEs

$$S_t^i = s_0^i + \int_0^t (S_s^i b_s^i, S_s^i \sigma_s^{i\cdot}) \,\mathrm{d}(\cdot, \mathbf{W})_s, \qquad t \in [0,T], \qquad i = 1, \dots, k, \tag{4.38}$$

where $(\cdot, \mathbf{W})$ denotes the time-extended rough path of $\mathbf{W}$, i.e., the path-level of $(\cdot, \mathbf{W})$ is given by $(t, W_t)_{t \in [0,T]}$ and the missing integrals $\int \bar{W}_t^j(\omega) \,\mathrm{d}t$, $\int t \,\mathrm{d}\bar{W}_t^j(\omega)$, $j = 1, \dots, d$, to define a rough path are canonically defined as Riemann–Stieltjes integrals. Using Lemma A.4.6, Proposition A.4.5, and Lemma A.4.7, there exists a unique solution $(S^i, (S^i)') \in \mathcal{V}_{(\cdot,W)}^p$ to the above RDE, where $(S^i)' = (S^i b^i, S^i \sigma^{i\cdot})$, $i = 1, \dots, k$. Thus, S, and b and σ, are controlled paths with respect to $(\cdot, W)$. Moreover, $(S_t)_{t \in [0,T]}$ satisfies the RDE (4.38) if and only if it satisfies the RDE (4.35). For later reference, we also remark (again) that $(\cdot, W)$ satisfies Property (RIE) relative to p and $(\mathcal{P}^n)_{n \in \mathbb{N}}$.

4.4.1 Pathwise construction of log-optimal portfolios

As a first step to a pathwise analysis of optimal portfolios, we prove a pathwise construction of the log-optimal portfolio, supposing that the underlying price dynamics of the financial market are given by a Black–Scholes-type model. Recall that in the probabilistic setting the log-optimal portfolio is well-known and was presented in Theorem 4.1.1, which is due to [79].

We recall that, given a path W, the time-extended path is denoted by $(\cdot, W) = (t, W_t)_{t\in[0,T]}$.

Lemma 4.4.2. *For $b \in \mathcal{V}^p_W([0,T];\mathcal{L}(\mathbb{R};\mathbb{R}^k))$ and $\sigma \in \mathcal{V}^p_W([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$ such that $\sigma_t\sigma_t^\top \in \mathrm{GL}(\mathbb{R}^{k\times k})$ for all t, where each coefficient is uniformly bounded away from zero, let*

$$H^i_t := H^{(b,\sigma),i}_t := \frac{h^i_t}{S^i_t} \qquad \text{with} \qquad h_t := h^{(b,\sigma)}_t := (\sigma_t\sigma_t^\top)^{-1}b_t,$$

for $t \in [0,T]$, and set $(\varphi,\kappa) := (\varphi^{(b,\sigma)},\kappa^{(b,\sigma)}) := (\varphi^{(b,\sigma),0},\dots,\varphi^{(b,\sigma),k},\kappa^{(b,\sigma)})$, with

$$\begin{aligned}
Z_t &:= Z^{(b,\sigma)}_t := \int_0^t h_s^\top b_s \,\mathrm{d}s + \int_0^t h_s^\top \sigma_s \,\mathrm{d}\mathbf{W}_s,\\
\kappa_t &:= \kappa^{(b,\sigma)}_t := \frac{1}{K_T}\mathcal{E}(Z)_t, \qquad V_t := V^{(b,\sigma)}_t := \kappa_t(K_T - K_t),\\
\varphi^i_t &:= \varphi^{(b,\sigma),i}_t := H^i_t V_t, \qquad i = 1,\dots,k, \qquad \varphi^0_t := \varphi^{(b,\sigma),0}_t := \sum_{i=1}^d \int_0^t \varphi^i_s \,\mathrm{d}S^i_s - \sum_{i=1}^d \varphi^i_t S^i_t,
\end{aligned}$$

for $t \in [0,T]$, where $\int_0^t \varphi^i_s \,\mathrm{d}S^i_s$ is the rough integral, and $\mathcal{E}$ is the rough exponential as defined in Lemma A.4.7. Then φ, κ and V are all well-defined and are controlled paths with respect to W and, in particular, with respect to $(\cdot, W)$.

Proof. Since b and σ are controlled paths in $\mathcal{V}^p_W$, and $\det(\sigma\sigma^\top)$ is bounded away from zero by assumption, $\sigma\sigma^\top$ and h are controlled paths in $\mathcal{V}^p_W$ because the sum and product of (real-valued) controlled paths is again a controlled path (see Lemma A.4.3), as well as the inverse of a controlled path that is bounded away from zero (as a composition with the smooth function $x \mapsto \frac{1}{x}$), see, e.g., [75, Lemma 3.5]. The same holds true for $h^\top b$ and $h^\top\sigma$.

Similarly, since each component of S is bounded away from zero due to its explicit representation as a rough exponential, we obtain that H is a controlled path in $\mathcal{V}^p_W$.

Then, Z is a controlled path in $\mathcal{V}^p_W$. Lemma A.4.6 and Lemma A.4.7 give that the rough exponential $\mathcal{E}(Z)$ is a controlled path in $\mathcal{V}^p_W$, and so is κ. Since the consumption clock K is a càdlàg (deterministic) and increasing function (and thus of finite 1-variation), by Lemma A.4.3, the wealth process V is a controlled path in $\mathcal{V}^p_W$ as the product of two controlled paths.

By similar arguments, we see that $\varphi^i \in \mathcal{V}^p_W$, $i = 0,1,\dots,k$, are also all controlled paths with respect to W, and hence also with respect to $(\cdot, W)$. □

The portfolio constructed in Lemma 4.4.2 in a pathwise manner agrees, indeed, with the log-optimal portfolio for the investment-consumption problem, as considered in Section 4.1, if the underlying frictionless financial market is generated by a Black–Scholes-type model, such as the stochastic differential equation (4.37). Hence, in the following we shall call the portfolio $(\varphi, \kappa) = (\varphi^{(b,\sigma)}, \kappa^{(b,\sigma)})$ from Lemma 4.4.2 *pathwise log-optimal portfolio.*

Lemma 4.4.3. *Suppose that the discounted price process* $(\bar{S}_t)_{t\in[0,T]}$ *is modelled by the SDE* (4.37) *driven by a Brownian motion* $\bar{W}$ *on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *with respect to an underlying filtration* $(\mathcal{F}_t)_{t\in[0,T]}$*. Then the log-optimal portfolio* $(\bar{\varphi}, \bar{\kappa})$*, as provided in Remark 4.1.2, and the pathwise log-optimal portfolio* (φ, κ)*, as provided in Lemma 4.4.2, coincide* $\mathbb{P}$*-almost surely, where* (φ, κ) *is constructed given the realization* $W := \bar{W}(\omega)$ *of the Itô rough path lift of the Brownian motion* $\bar{W}$*, for almost every* $\omega \in \Omega$*.*

Proof. In this proof we consider S, φ, κ, etc., as random controlled paths in the sense that S is a stochastic process such that $S(\omega)$ is a controlled path for almost every $\omega \in \Omega$. In particular, S is defined pathwise as the solution to the RDE (4.11) and (φ, κ) is defined pathwise via Lemma 4.3.3, given a realization of the Brownian motion $\bar{W}$; more explicitly, we have $b = \bar{b}$, $\sigma = \bar{\sigma}$ and $S = \bar{S}$ etc., almost surely; see Remark 4.4.1.

We note that

$$\bar{\kappa}_t = \frac{1}{K_T}\bar{\mathcal{E}}(\bar{Z})_t, \qquad t \in [0,T],$$

for $\bar{Z} = \int_0^\cdot \bar{h}_t^\top \bar{b}_t \,\mathrm{d}t + \int_0^\cdot \bar{h}_t^\top \bar{\sigma}_t \,\mathrm{d}\bar{W}_t$, where $\bar{\mathcal{E}}$ denotes the stochastic exponential. By Proposition 4.2.4, we then have that

$$\int_0^t h_s^\top \sigma_s \,\mathrm{d}\mathbf{W}_s = \int_0^t \bar{h}_s^\top \bar{\sigma}_s \,\mathrm{d}\bar{\mathbf{W}}_s = \int_0^t \bar{h}_s^\top \bar{\sigma}_s \,\mathrm{d}\bar{W}_s,$$

almost surely, which implies that $Z = \bar{Z}$ almost surely.

Then, by Lemma A.4.6, Z admits a canonical rough path lift $\mathbf{Z} \in \mathcal{D}^p$, and since W satisfies Property (RIE) relative to p and $(\mathcal{P}^n)_{n\in\mathbb{N}}$, Z also satisfies Property (RIE) by Theorem 4.2.2. By [7, Proposition 2.18], this implies that the rough path bracket $[\mathbf{Z}]$ coincides with the quadratic variation $[Z]$ of Z along $(\mathcal{P}^n)_{n\in\mathbb{N}}$ in the sense of Föllmer. Proceeding as in the proof of Lemma 4.3.3, one can show that $[\mathbf{Z}]$ and $[\bar{Z}]$ coincide almost surely.

For the rough exponential, by Lemma A.4.7, it then follows that

$$\mathcal{E}(Z)_t = \exp(Z_t - \frac{1}{2}[\mathbf{Z}]_t) = \exp(Z_t - \frac{1}{2}[Z]_t) = \exp(\bar{Z}_t - \frac{1}{2}[\bar{Z}]_t) = \bar{\mathcal{E}}(\bar{Z})_t,$$

almost surely. Hence $\kappa = \frac{1}{K_T}\mathcal{E}(Z) = \bar{\mathcal{E}}(\bar{Z}) = \bar{\kappa}$ coincide almost surely, and so is $V_t = \kappa_t(K_T - k_t) = \bar{\kappa}_t(K_T - k_t) = \bar{V}_t$ almost surely. Consequently, as $H^i = \frac{h^i}{S^i}$ coincides with

$\bar{H}^i := \frac{\bar{h}^i}{\bar{S}^i}$ almost surely for all $i = 1, \dots, k$, we also get that $\varphi_t^i = H_t^i V_{t-}^i = \bar{H}^i \bar{V}_{t-}^i = \bar{\varphi}_t^i$ almost surely for all $i = 1, \dots, k$.

Finally, by the associativity property of rough integrals, see Proposition A.4.5, we have that

$$\begin{aligned}
&\int_0^t \varphi_s^i \,\mathrm{d}S_s^i \\
&\quad = \int_0^t \varphi_s^i S_s^i b_s^i \,\mathrm{d}s + \int_0^t \varphi_s^i S_s^i \sigma_s^{i\cdot} \,\mathrm{d}\mathbf{W}_s \\
&\quad = \int_0^t \bar{\varphi}_s^i \bar{S}_s^i \bar{b}_s^i \,\mathrm{d}s + \int_0^t \bar{\varphi}_s^i \bar{S}_s^i \bar{\sigma}_s^{i\cdot} \,\mathrm{d}\bar{W}_t,
\end{aligned}$$

almost surely, which implies that $\varphi^0 = \sum_{i=1}^k \int_0^\cdot \varphi_t^i \,\mathrm{d}S_t^i - \sum_{i=1}^d \varphi^i S^i = \int_0^\cdot \bar{\varphi}_t^\top \,\mathrm{d}\bar{S}_t - \sum_{i=1}^d \bar{\varphi}^i \bar{S}^i$ $= \bar{\varphi}^0$ almost surely. Thus, the log-optimal portfolio $(\bar{\varphi}, \bar{\kappa})$, as provided in Remark 4.1.2, and the pathwise log-optimal portfolio (φ, κ), as provided in Lemma 4.4.2, coincide almost surely. □

Remark 4.4.4. *We take $\bar{W}$ to be a Brownian motion to ensure that the pathwise log-optimal portfolio (φ, κ), as constructed in Lemma 4.4.2, is, indeed, a log-optimal portfolio for the investment-consumption problem in the setting of Black–Scholes-type models. However, again, we emphasize that the construction of the pathwise portfolio (φ, κ) as well as its pathwise analysis developed in Sections 4.4.2 and 4.4.3 works for any path W satisfying Assumption 4.3.1.*

4.4.2 Stability of pathwise log-optimal portfolios with respect to drift and volatility

Having at hand a pathwise construction of log-optimal portfolios, we are in a position to study its pathwise stability properties. In this subsection, we analyze the stability of the log-optimal portfolio and the associated capital process with respect to the model parameters, b and σ.

In particular, the following result shows that the pathwise log-optimal portfolios $(\varphi, \kappa) = (\varphi^{(b,\sigma)}, \kappa^{(b,\sigma)})$ and its associated capital process $V = V^{(b,\sigma)}$ are locally Lipschitz continuous with respect to these parameters.

Theorem 4.4.5. *For $b, \tilde{b} \in \mathcal{V}_W^p([0,T]; \mathcal{L}(\mathbb{R}; \mathbb{R}^k))$ and $\sigma, \tilde{\sigma} \in \mathcal{V}_W^p([0,T]; \mathcal{L}(\mathbb{R}^d; \mathbb{R}^k))$ such that $\sigma_t \sigma_t^\top, \tilde{\sigma}_t \tilde{\sigma}_t^\top \in \mathrm{GL}(\mathbb{R}^{k \times k})$ for all t, where each coefficient is uniformly bounded away from zero, let $(\varphi^{(b,\sigma)}, \kappa^{(b,\sigma)})$ and $(\varphi^{(\tilde{b},\tilde{\sigma})}, \kappa^{(\tilde{b},\tilde{\sigma})})$ be the corresponding pathwise log-optimal portfolios, as constructed in Lemma 4.4.2. Let M be an upper bound for*

$$\|b\|_{\mathcal{V}^p_{(\cdot,W)}}, \|\tilde{b}\|_{\mathcal{V}^p_{(\cdot,W)}}, \|\sigma\|_{\mathcal{V}^p_{(\cdot,W)}}, \|\tilde{\sigma}\|_{\mathcal{V}^p_{(\cdot,W)}}, 1/\inf_t |\det(\sigma_t \sigma_t^\top)|, 1/\inf_t |\det(\tilde{\sigma}_t \tilde{\sigma}_t^\top)| \text{ and } \|(\cdot, \mathbf{W})\|_p.$$

We then have that

$$\|(\varphi^{(b,\sigma)},\kappa^{(b,\sigma)});(\varphi^{(\tilde b,\tilde\sigma)},\kappa^{(\tilde b,\tilde\sigma)})\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|b;\tilde b\|_{\mathcal{V}^p_{(\cdot,W)}}+\|\sigma;\tilde\sigma\|_{\mathcal{V}^p_{(\cdot,W)}}$$

and

$$\|V^{(b,\sigma)};V^{(\tilde b,\tilde\sigma)}\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|b;\tilde b\|_{\mathcal{V}^p_{(\cdot,W)}}+\|\sigma;\tilde\sigma\|_{\mathcal{V}^p_{(\cdot,W)}},$$

and in particular that

$$\|(\varphi^{(b,\sigma)},\kappa^{(b,\sigma)})-(\varphi^{(\tilde b,\tilde\sigma)},\kappa^{(\tilde b,\tilde\sigma)})\|_{\infty}\lesssim \|b;\tilde b\|_{\mathcal{V}^p_{(\cdot,W)}}+\|\sigma;\tilde\sigma\|_{\mathcal{V}^p_{(\cdot,W)}}$$

and

$$\|V^{(b,\sigma)}-V^{(\tilde b,\tilde\sigma)}\|_{\infty}\lesssim \|b;\tilde b\|_{\mathcal{V}^p_{(\cdot,W)}}+\|\sigma;\tilde\sigma\|_{\mathcal{V}^p_{(\cdot,W)}},$$

where the implicit multiplicative constants depend only on p, k, d, M, s_0, $1/\inf_t|S_t^{(b,\sigma),i}|$, $1/\inf_t|S_t^{(\tilde b,\tilde\sigma),i}|$, $i=1,\dots,k$, and the consumption clock K.

Proof. Step 1. Let $c^{(b,\sigma)}:=\sigma\sigma^\top$, $c^{(\tilde b,\tilde\sigma)}=\tilde\sigma\tilde\sigma^\top$. As shown in the proof of Lemma 4.4.2, $c^{(b,\sigma)}$ is a controlled path in $\mathcal{V}^p_W$, thus in $\mathcal{V}^p_{(\cdot,W)}$. Lemma A.4.3 then yields for each component that

$$\|((c^{(b,\sigma)})^{-1})^{ij}\|_{\mathcal{V}^p_{(\cdot,W)}}\le C, \tag{4.39}$$

where $C>0$ depends only on p, k, d, and M; the same holds for $((c^{(\tilde b,\tilde\sigma)})^{-1})^{ij}$. By Lemma A.4.4 and since the inverse of a controlled path that is bounded away from zero (as a composition with the regular function $x\mapsto\frac{1}{x}$) is locally Lipschitz continuous, we can check with the estimate (4.39) that

$$\|((c^{(b,\sigma)})^{-1})^{ij};((c^{(\tilde b,\tilde\sigma)})^{-1})^{ij}\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|b;\tilde b\|_{\mathcal{V}^p_{(\cdot,W)}}+\|\sigma;\tilde\sigma\|_{\mathcal{V}^p_{(\cdot,W)}}, \tag{4.40}$$

where the implicit multiplicative constant depends only on p, k, d, and M.

Let $\vartheta^{(b,\sigma)}:=((h^{(b,\sigma)})^\top\sigma)^\top=\sigma^\top(c^{(b,\sigma)})^{-1}b$, $\vartheta^{(\tilde b,\tilde\sigma)}:=((h^{(\tilde b,\tilde\sigma)})^\top\tilde\sigma)^\top=\tilde\sigma^\top(c^{(\tilde b,\tilde\sigma)})^{-1}\tilde b$. Then, $\theta^{(b,\sigma)}=(\frac{1}{2}\vartheta^{(b,\sigma)}(\vartheta^{(b,\sigma)})^\top,(\vartheta^{(b,\sigma)})^\top)$ are controlled paths in $\mathcal{V}^p_{(\cdot,W)}$ as, again, the sum and product of controlled paths remains a controlled path. Using the same arguments as before and combining with the estimates (4.39) and (4.40), we get that

$$\|\theta^{(b,\sigma)}\|_{\mathcal{V}^p_{(\cdot,W)}}\le C, \tag{4.41}$$

where $C>0$ depends only on p, k, d, and M; the same holds for $\theta^{(\tilde b,\tilde\sigma)}$; and

$$\|\theta^{(b,\sigma)};\theta^{(\tilde b,\tilde\sigma)}\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|b;\tilde b\|_{\mathcal{V}^p_{(\cdot,W)}}+\|\sigma;\tilde\sigma\|_{\mathcal{V}^p_{(\cdot,W)}}, \tag{4.42}$$

where the implicit multiplicative constant depends only on p, k, d, and M.

Step 2. Define the rough integrals $U^{(b,\sigma)} := \int_0^{\cdot} \theta_t^{(b,\sigma)} \,\mathrm{d}(\cdot,\mathbf{W})_t$ and $U^{(\tilde{b},\tilde{\sigma})} := \int_0^{\cdot} \theta_t^{(\tilde{b},\tilde{\sigma})} \,\mathrm{d}(\cdot,\mathbf{W})_t$, which are controlled paths in $\mathcal{V}^p_{(\cdot,W)}$. Using the estimate (4.5) for the rough integral and the estimate (4.41), it holds that

$$\|U^{(b,\sigma)}\|_{\mathcal{V}^p_{(\cdot,W)}} \leq C, \tag{4.43}$$

where $C > 0$ depends only on p, k, d, and M; the same holds for $U^{(\tilde{b},\tilde{\sigma})}$. Particularly,

$$\|U^{(b,\sigma)}\|_{\infty} \leq C_0, \tag{4.44}$$

where $C_0 > 0$ depends only on p, k, d, and M. Further, using the stability for rough integrals, see, e.g. [75, Lemma 3.4], and the estimate (4.42), it immediately follows that

$$\|U^{(b,\sigma)}; U^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}} \lesssim \|b;\tilde{b}\|_{\mathcal{V}^p_{(\cdot,W)}} + \|\sigma;\tilde{\sigma}\|_{\mathcal{V}^p_{(\cdot,W)}}, \tag{4.45}$$

where the implicit multiplicative constant depends only on p, k, d, and M.

Step 3. Proceeding as in the proof of Lemma 4.3.4, it follows from Lemma A.4.7 that

$$\begin{aligned}
&\mathcal{E}(Z^{(b,\sigma)})_t \\
&\quad = \exp(Z_t^{(b,\sigma)} - \frac{1}{2}[\mathbf{Z}^{(b,\sigma)}]_t) \\
&\quad = \exp\Big(\frac{1}{2}\int_0^t (\vartheta_s^{(b,\sigma)})^\top \vartheta_s^{(b,\sigma)} \,\mathrm{d}s + \int_0^t (\vartheta_s^{(b,\sigma)})^\top \,\mathrm{d}\mathbf{W}_s\Big) \\
&\quad = \exp(U_t^{(b,\sigma)}).
\end{aligned}$$

By Lemma 4.4.2, we have that

$$\kappa_t^{(b,\sigma)} := \frac{1}{K_T}\mathcal{E}(Z^{(b,\sigma)})_t, \qquad \kappa_t^{(\tilde{b},\tilde{\sigma})} := \frac{1}{K_T}\mathcal{E}(Z^{(\tilde{b},\tilde{\sigma})})_t,$$

for $t \in [0,T]$, are the pathwise defined optimal consumption rates for the log-utility on the financial market modeled by $S^{(b,\sigma),i}$ and $S^{(\tilde{b},\tilde{\sigma}),i}$, $i = 1,\dots,k$, respectively, and are controlled paths in $\mathcal{V}^p_{(\cdot,W)}$. We therefore get with (4.43) and (4.44) (see proof of Theorem 4.3.6) that

$$\|\kappa^{(b,\sigma)}\|_{\mathcal{V}^p_{(\cdot,W)}} \leq C, \tag{4.46}$$

where $C > 0$ depends only on p, k, d, M, $\|\exp\|_{C_b^2(\{y:|y|\leq C_0\},\mathbb{R})}$, and the consumption clock K, as it is a composition of a controlled path with a regular function; see, e.g., [75, Lemma 3.5]. The same holds for $\kappa^{(\tilde{b},\tilde{\sigma})}$.

Because the composition of a controlled path with a regular function is locally Lipschitz continuous, see, e.g. [75, Lemma 3.5], it follows with (4.45) that

$$\|\kappa^{(b,\sigma)}; \kappa^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}} \lesssim \|b;\tilde{b}\|_{\mathcal{V}^p_{(\cdot,W)}} + \|\sigma;\tilde{\sigma}\|_{\mathcal{V}^p_{(\cdot,W)}}, \tag{4.47}$$

where the implicit multiplicative constant depends only on p, k, d, M, and K.

Step 4. Since K_t, $t \in [0,T]$, is a càdlàg (deterministic) and increasing function (so of finite 1-variation), we recall that by Lemma A.4.3, the wealth process $V_t^{(b,\sigma)} := \kappa_t^{(b,\sigma)}(K_T - K_t)$, $t \in [0,T]$, (as the product of two controlled paths) is a controlled path in $\mathcal{V}^p_{(\cdot,W)}$. One can derive that

$$\|V^{(b,\sigma)}\|_{\mathcal{V}^p_{(\cdot,W)}} \leq C, \tag{4.48}$$

where $C > 0$ depends only on p, k, d, M, and K. Applying Lemma A.4.4 and the estimate (4.48) yields that

$$\|V^{(b,\sigma)}; V^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}} \lesssim \|\kappa^{(b,\sigma)}; \kappa^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}},$$

where the implicit multiplicative constant depends only on p, M, K, $\|\kappa^{(b,\sigma)}\|_{\mathcal{V}^p_{(\cdot,W)}}$ and $\|\kappa^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}}$. Combining this with (4.46) and (4.47), it holds that

$$\|V^{(b,\sigma)}; V^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}} \lesssim \|b; \tilde{b}\|_{\mathcal{V}^p_{(\cdot,W)}} + \|\sigma; \tilde{\sigma}\|_{\mathcal{V}^p_{(\cdot,W)}}, \tag{4.49}$$

where the implicit multiplicative constant depends only on p, k, d, M, and K.

Step 5. Define the rough integrals $A^{(b,\sigma),i} := \int_0^\cdot (b_t^i - \frac{1}{2}\sigma_t^{i\cdot}(\sigma_t^{i\cdot})^\top, \sigma_t^{i,\cdot})\,\mathrm{d}\mathbf{W}_t$, $A^{(\tilde{b},\tilde{\sigma}),i} := \int_0^\cdot (\tilde{b}_t^i - \frac{1}{2}\tilde{\sigma}_t^{i\cdot}(\tilde{\sigma}_t^{i\cdot})^\top, \tilde{\sigma}_t^{i,\cdot})\,\mathrm{d}\mathbf{W}_t$, $i = 1, \dots, k$. Using the estimate (4.5) for the rough integral, it holds that

$$\|A^{(b,\sigma),i}\|_{\mathcal{V}^p_{(\cdot,W)}} \leq C, \tag{4.50}$$

where $C > 0$ depends only on p, d, and M; the same holds for $A^{(\tilde{b},\tilde{\sigma}),i}$. Particularly,

$$\|A^{(b,\sigma),i}\|_\infty \leq C_1, \tag{4.51}$$

where $C_1 > 0$ depends only on p, d, and M. Further, using the stability for rough integrals, see, e.g. [75, Lemma 3.4], and Lemma A.4.4, it immediately follows that

$$\|A^{(b,\sigma),i}; A^{(\tilde{b},\tilde{\sigma}),i}\|_{\mathcal{V}^p_{(\cdot,W)}} \lesssim \|b; \tilde{b}\|_{\mathcal{V}^p_{(\cdot,W)}} + \|\sigma; \tilde{\sigma}\|_{\mathcal{V}^p_{(\cdot,W)}}, \tag{4.52}$$

where the implicit multiplicative constant depends only on p, d, and M.

Step 6. Further, we recall the beginning of the section, where we derived that the solution $S = (S_t)_{t\in[0,T]}$ of the rough differential equation (4.35) is given as a rough exponential. More precisely, for $i = 1, \dots, k$ it holds that

$$S_t^{(b,\sigma)i} = s_0^i \exp\left(\int_0^t (b_s^i - \frac{1}{2}\sigma_s^{i\cdot}(\sigma_s^{i\cdot})^\top)\,\mathrm{d}s + \int_0^t \sigma_s^{i\cdot}\,\mathrm{d}\mathbf{W}_s\right) = s_0^i \exp(A_t^{(b,\sigma),i}).$$

We therefore get with (4.50) that

$$\|S^{(b,\sigma),i}\|_{\mathcal{V}^p_{(\cdot,W)}} \leq C, \tag{4.53}$$

where $C>0$ depends only on p, d, M, s_0, and $\|\exp\|_{C_b^2(\{y:|y|\leq C_1\};\mathbb{R})}$, as it is a composition of a controlled path with a regular function; see, e.g., [75, Lemma 3.5]. We note that C_1 depends only on p, d, and M, see (4.51), that is, $C>0$ depends only on p, d, M, and s_0; the same holds for $S^{(\tilde{b},\tilde{\sigma}),i}$.

Then, by the stability of regular functions of controlled paths, see, e.g., [75, Lemma 3.5], it follows with (4.52) that

$$\|S^{(b,\sigma),i};S^{(\tilde{b},\tilde{\sigma}),i}\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|b;\tilde{b}\|_{\mathcal{V}^p_{(\cdot,W)}}+\|\sigma;\tilde{\sigma}\|_{\mathcal{V}^p_{(\cdot,W)}}, \tag{4.54}$$

where the implicit multiplicative constant depends only on p, d, M, and s_0.

Step 7. Similar to Step 1, for $h^{(b,\sigma)}=(c^{(b,\sigma)})^{-1}b$, $h^{(\tilde{b},\tilde{\sigma})}=(c^{(\tilde{b},\tilde{\sigma})})^{-1}\tilde{b}$, we obtain with Lemma A.4.3 and the estimate (4.39) that

$$\|h^{(b,\sigma),i}\|_{\mathcal{V}^p_{(\cdot,W)}}\leq C, \tag{4.55}$$

where $C>0$ depends only on p, k, d, and M; the same holds for $h^{(\tilde{b},\tilde{\sigma}),i}$; and then by (4.40) and Lemma A.4.4 it follows that

$$\|h^{(b,\sigma),i};h^{(\tilde{b},\tilde{\sigma}),i}\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|b;\tilde{b}\|_{\mathcal{V}^p_{(\cdot,W)}}+\|\sigma;\tilde{\sigma}\|_{\mathcal{V}^p_{(\cdot,W)}}, \tag{4.56}$$

where the implicit multiplicative constant depends only on p, k, d, and M.

Following Lemma 4.4.2, we consider

$$H_t^{(b,\sigma),i}=\frac{h_t^{(b,\sigma),i}}{S_t^{(b,\sigma),i}}.$$

Using the same argument as in Step 1 with Lemma A.4.3, Lemma A.4.4 and the estimates (4.53) and (4.55), (4.54) and (4.56) we obtain that

$$\|H^{(b,\sigma),i}\|_{\mathcal{V}^p_{(\cdot,W)}}\leq C, \tag{4.57}$$

where $C>0$ depends only on p, k, d, M, s_0, and $1/\inf_t|S_t^{(b,\sigma),i}|$; the same holds for $H^{(\tilde{b},\tilde{\sigma}),i}$; and

$$\|H^{(b,\sigma),i};H^{(\tilde{b},\tilde{\sigma}),i}\|_{\mathcal{V}^p_{(\cdot,W)}}\lesssim \|b;\tilde{b}\|_{\mathcal{V}^p_{(\cdot,W)}}+\|\sigma;\tilde{\sigma}\|_{\mathcal{V}^p_{(\cdot,W)}}, \tag{4.58}$$

where the implicit multiplicative constant depends only on p, k, d, M, s_0, $1/\inf_t|S_t^{(b,\sigma),i}|$, and $1/\inf_t|S_t^{(\tilde{b},\tilde{\sigma}),i}|$.

Step 8. Lastly, by Lemma A.4.3, $\varphi_t^{(b,\sigma),i}:=H_t^{(b,\sigma),i}V_{t-}^{(b,\sigma)}$, $\varphi_t^{(\tilde{b},\tilde{\sigma}),i}:=H_t^{(\tilde{b},\tilde{\sigma}),i}V_{t-}^{(\tilde{b},\tilde{\sigma})}$, $i=1,\dots,k$, are controlled paths in $\mathcal{V}^p_W$, and

$$\|\varphi^{(b,\sigma),i}\|_{\mathcal{V}^p_{(\cdot,W)}}\leq C, \tag{4.59}$$

where $C > 0$ depends only on p, M, $\|H^{(b,\sigma),i}\|_{\mathcal{V}^p_{(\cdot,W)}}$, $\|V^{(b,\sigma)}\|_{\mathcal{V}^p_{(\cdot,W)}}$, that is, only on p, k, d, M, s_0, $1/\inf_t|S_t^{(b,\sigma),i}|$, and the consumption clock K, see (4.57) and (4.48); the same holds for $\varphi^{(\tilde{b},\tilde{\sigma}),i}$. By Lemma A.4.4,

$$\|\varphi^{(b,\sigma),i};\varphi^{(\tilde{b},\tilde{\sigma}),i}\|_{\mathcal{V}^p_{(\cdot,W)}} \lesssim \|H^{(b,\sigma),i};H^{(\tilde{b},\tilde{\sigma}),i}\|_{\mathcal{V}^p_{(\cdot,W)}} + \|V^{(b,\sigma)};V^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}},$$

where the implicit multiplicative constant depends only on p, M, $\|H^{(b,\sigma),i}\|_{\mathcal{V}^p_{(\cdot,W)}}$, $\|H^{(\tilde{b},\tilde{\sigma}),i}\|_{\mathcal{V}^p_{(\cdot,W)}}$, $\|V^{(b,\sigma)}\|_{\mathcal{V}^p_{(\cdot,W)}}$, $\|V^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}}$. This gives with (4.57) and (4.48), (4.58) and (4.49) that

$$\|\varphi^{(b,\sigma),i};\varphi^{(\tilde{b},\tilde{\sigma}),i}\|_{\mathcal{V}^p_{(\cdot,W)}} \lesssim \|b;\tilde{b}\|_{\mathcal{V}^p_{(\cdot,W)}} + \|\sigma;\tilde{\sigma}\|_{\mathcal{V}^p_{(\cdot,W)}}, \tag{4.60}$$

where the implicit multiplicative constant depends only on p, k, d, M, s_0, $1/\inf_t|S_t^{(b,\sigma),i}|$, $1/\inf_t|S_t^{(\tilde{b},\tilde{\sigma}),i}|$, and K.

Finally, we consider

$$\varphi_t^{(b,\sigma),0} = \sum_{i=1}^{k} \int_0^t \varphi_s^{(b,\sigma),i}\,\mathrm{d}S_s^{(b,\sigma),i} - \varphi_t^{(b,\sigma),i} S_t^{(b,\sigma),i}$$

and

$$\varphi_t^{(\tilde{b},\tilde{\sigma}),0} = \sum_{i=1}^{k} \int_0^t \varphi_s^{(\tilde{b},\tilde{\sigma}),i}\,\mathrm{d}S_s^{(\tilde{b},\tilde{\sigma}),i} - \varphi_t^{(\tilde{b},\tilde{\sigma}),i} S_t^{(\tilde{b},\tilde{\sigma}),i},$$

for $t \in [0,T]$. By the associativity property of rough integrals, it holds that

$$\begin{aligned}\int_0^{\cdot} \varphi_t^{(b,\sigma),i}\,\mathrm{d}S_t^{(b,\sigma),i} &= \int_0^{\cdot} (\varphi_t^{(b,\sigma),i} S_t^i b_t^i, \varphi_t^{(b,\sigma),i} S_t^i \sigma_t^i)\,\mathrm{d}(\cdot,\mathbf{W})_t \\ &=: \int_0^{\cdot} \psi_t^{(b,\sigma),i}\,\mathrm{d}(\cdot,\mathbf{W})_t,\end{aligned}$$

similarly for $\int_0^{\cdot} \varphi_t^{(\tilde{b},\tilde{\sigma}),i}\,\mathrm{d}S_t^{(\tilde{b},\tilde{\sigma}),i}$. Using the same arguments as above, by (4.59) and (4.53), (4.60) and (4.54), it holds that

$$\|\psi^{(b,\sigma)};\psi^{(\tilde{b},\tilde{\sigma})}\|_{\mathcal{V}^p_{(\cdot,W)}} \lesssim \|b;\tilde{b}\|_{\mathcal{V}^p_{(\cdot,W)}} + \|\sigma;\tilde{\sigma}\|_{\mathcal{V}^p_{(\cdot,W)}},$$

where the implicit multiplicative constant depends only on p, k, d, M, $\|(\cdot,\mathbf{W})\|_p$, s_0, $1/\inf_t|S_t^{(b,\sigma),i}|$, $1/\inf_t|S_t^{(\tilde{b},\tilde{\sigma}),i}|$, $i = 1,\dots,k$, and K.

Therefore, combining this with the estimate (4.5) for the rough integral, Lemma A.4.4, and the estimates (4.59), (4.60) and (4.54), we can derive that

$$\|\varphi^{(b,\sigma),0};\varphi^{(\tilde{b},\tilde{\sigma}),0}\|_{\mathcal{V}^p_{(\cdot,W)}} \lesssim \|b;\tilde{b}\|_{\mathcal{V}^p_{(\cdot,W)}} + \|\sigma;\tilde{\sigma}\|_{\mathcal{V}^p_{(\cdot,W)}},$$

where the implicit multiplicative constant depends only on p, k, d, M, $\|(\cdot,\mathbf{W})\|_p$, s_0, $1/\inf_t|S_t^{(b,\sigma),i}|$, $1/\inf_t|S_t^{(\tilde{b},\tilde{\sigma}),i}|$, $i = 1,\dots,k$, and K.

Hence, since we can bound the supremum norm by the controlled path norm, see (4.2), the (local) Lipschitz continuity for optimal portfolios and wealth processes follows. □

4.4.3 Discretization error of pathwise log-optimal portfolios

In this subsection, we introduce a time-discrete version of the pathwise log-optimal portfolio, as constructed in Lemma 4.4.2, and derive quantitative, pathwise error estimates resulting from this discretization for the portfolios as well as for their associated capital processes.

To define the time-discrete version of the pathwise log-optimal portfolio, we start by discretizing the underlying price paths. To that end, we recall that W and the sequence $(\mathcal{P}^n)$ of partitions satisfy Assumption 4.3.1, where $\mathcal{P}^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$. For $n \in \mathbb{N}$, let $W^n\colon [0,T] \to \mathbb{R}^d$ be the piecewise constant approximation of W along $\mathcal{P}^n$, as introduced in Section 4.3.3. We define b^n and σ^n in the same way, and let $\gamma^n\colon [0,T] \to \mathbb{R}$ be the time discretization path along $\mathcal{P}^n$.

To discretize the price path S, we introduce the discretization of the rough exponential S along $\mathcal{P}^n$ given by

$$S_t^{n,i} := s_0^i \exp(A_t^{n,i}) \qquad \text{with} \qquad A_t^{n,i} := \int_0^t (b_s^{n,i} - \frac{1}{2}\sigma_s^{n,i\cdot}(\sigma_s^{n,i\cdot})^\top, \sigma_s^{n,i\cdot}) \,\mathrm{d}(\gamma^n, W^n)_s,$$

for $t \in [0,T]$, $i = 1, \dots, k$. The time-discrete version (φ^n, κ^n) of the pathwise log-optimal portfolio is defined by

$$\begin{aligned}
&H_t^{n,i} := \frac{h_t^{n,i}}{S_t^{n,i}} \qquad \text{with} \qquad h_t^n := (\sigma_t^n(\sigma_t^n)^\top)^{-1} b_t^n,\\
&\kappa_t^n := \frac{1}{K_T} \exp\left(\frac{1}{2}\int_0^t ((h_s^n)^\top b_s^n, (h_s^n)^\top \sigma_s^n) \,\mathrm{d}(\gamma^n, W^n)_s\right), \qquad v_t^n := \kappa_t^n (K_T - K_t^n),\\
&\varphi_t^{n,i} := H_t^{n,i} v_{t-}^n, \qquad i = 1, \dots, k, \quad \varphi_t^{n,0} := \sum_{i=1}^k \int_0^t \varphi_s^{n,i} \,\mathrm{d}S_s^{n,i} - \varphi_t^{n,i} S_t^{n,i},\\
&V_t^n := 1 + \sum_{i=1}^k \int_0^t \varphi_s^{n,i} \,\mathrm{d}S_s^{n,i} - \int_0^t \kappa_s^n \,\mathrm{d}K_s, \qquad t \in [0,T],
\end{aligned}$$

where all above integrals are just left-point Riemann sums and K^n denotes the piecewise constant approximation of K along $\mathcal{P}^n$. For these time-discrete portfolios and their associated capital processes, we obtain the following convergence result with quantitative error estimates.

Theorem 4.4.6. *For $b \in \mathcal{V}_W^p([0,T];\mathcal{L}(\mathbb{R};\mathbb{R}^k))$ and $\sigma \in \mathcal{V}_W^p([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$ such that $\sigma_t\sigma_t^\top \in \mathrm{GL}(\mathbb{R}^{k\times k})$ for all t, where each coefficient is uniformly bounded away from zero, let $(\varphi^{(b,\sigma)}, \kappa^{(b,\sigma)})$ be the pathwise log-optimal portfolio, as constructed in Lemma 4.4.2. Then,*

$$\|(\varphi^n, \kappa^n) - (\varphi, \kappa)\|_{p'} \longrightarrow 0 \qquad \textit{as} \qquad n \longrightarrow \infty$$

and

$$\|V^n - V\|_{p'} \longrightarrow 0 \qquad \textit{as} \qquad n \longrightarrow \infty,$$

for any $p' \in (p,3)$, with a rate of convergence given by

$$
\begin{aligned}
&\|(\varphi^n,\kappa^n)-(\varphi,\kappa)\|_{p'} \\
&\quad\lesssim \|(b^n)'-b'\|_\infty^{1-\frac{p}{p'}} + \|R^{b^n}-R^b\|_\infty^{1-\frac{p}{p'}} + \|(\sigma^n)'-\sigma'\|_\infty^{1-\frac{p}{p'}} + \|R^{\sigma^n}-R^\sigma\|_\infty^{1-\frac{p}{p'}} \\
&\qquad + (|\mathcal{P}^n| + \|W^n-W\|_\infty)^{1-\frac{p}{p'}} + \left(|\mathcal{P}^n|^{1-\frac{1}{q}} + \left\|\int_0^\cdot W_t^n\otimes \mathrm{d}W_t - \int_0^\cdot W_t\otimes \mathrm{d}W_t\right\|_\infty\right)^{1-\frac{p}{p'}},
\end{aligned}
$$

and

$$
\begin{aligned}
&\|V^n-V\|_{p'} \\
&\quad\lesssim \|(b^n)'-b'\|_\infty^{1-\frac{p}{p'}} + \|R^{b^n}-R^b\|_\infty^{1-\frac{p}{p'}} + \|(\sigma^n)'-\sigma'\|_\infty^{1-\frac{p}{p'}} + \|R^{\sigma^n}-R^\sigma\|_\infty^{1-\frac{p}{p'}} \\
&\qquad + (|\mathcal{P}^n| + \|W^n-W\|_\infty)^{1-\frac{p}{p'}} + \left(|\mathcal{P}^n|^{1-\frac{1}{q}} + \left\|\int_0^\cdot W_t^n\otimes \mathrm{d}W_t - \int_0^\cdot W_t\otimes \mathrm{d}W_t\right\|_\infty\right)^{1-\frac{p}{p'}},
\end{aligned}
$$

for any $q \in (1,2)$ such that $\frac{1}{p'}+\frac{1}{q}>1$, where the implicit multiplicative constant depends only on p, p', q, k, d, $\|b\|_{\mathcal{V}^{p'}_{(\cdot,W)}}$, $\|\sigma\|_{\mathcal{V}^{p'}_{(\cdot,W)}}$, $1/\inf_t|\det(\sigma_t\sigma_t^\top)|$, $\|W\|_p$, $w(0,T)$, where w is the control function for which (4.6) *holds for $(\cdot,W)$, and the consumption clock K.*

Remark 4.4.7. *The convergence results and quantitative estimates in Theorem 4.4.6 hold true when replacing the p'-variation seminorm $\|\cdot\|_{p'}$ by the supremum seminorm $\|\cdot\|_\infty$.*

Before we present the proof, some preliminary steps are necessary. We start by recalling that, as W^n has finite 1-variation, W^n possesses a canonical rough path lift $\mathbf{W}^n = (W^n,\mathbb{W}^n) \in \mathcal{D}^p([0,T],\mathbb{R}^d)$, with $\mathbb{W}^n$ given by

$$\mathbb{W}^n_{s,t} := \int_s^t W^n_{s,u}\otimes \mathrm{d}W^n_u, \qquad (s,t)\in\Delta_T,$$

where the integral is defined as a classical limit of left-point Riemann sums. Similarly, we can define a time-space rough path $(\cdot,\mathbf{W})^n$ of $(\cdot,W)^n := (\gamma^n,W^n)$.

Since b^n is the piecewise constant approximation of b along $\mathcal{P}^n$, it is a controlled path with respect to $(\cdot,W)^n$. If $t_k^n \le s < t \le t_{k+1}^n$ for some k, then $b^n_{s,t} = b_{t_k^n,t_k^n} = 0$. Otherwise, let k_0 be the smallest k such that $t_k^n \in (s,t)$ and k_1 the largest such k. Then,

$$b^n_{s,t} = b_{t^n_{k_0},t^n_{k_1}} = b'_{t^n_{k_0}}(\cdot,W)_{t^n_{k_0},t^n_{k_1}} + R^b_{t^n_{k_0},t^n_{k_1}} = (b')^n_s(\cdot,W)^n_{s,t} + (R^b)^n_{s,t},$$

where $(b')^n$ and $(R^b)^n$ be the piecewise constant approximations of b' and R^b along $\mathcal{P}^n$, respectively. Therefore, $\sup_{n\in\mathbb{N}}\|b^n\|_{\mathcal{V}^p_{(\cdot,W)^n}} \le \|b\|_{\mathcal{V}^p_{(\cdot,W)}}$; analogously for σ.

Furthermore, proceeding as in the proof of Lemma 4.4.2, one can show that H^n, κ^n, V^n and φ^n are controlled paths in $\mathcal{V}^p_{(\cdot,W)^n}$, and $\int_0^t \varphi_s^{n,i}\,\mathrm{d}S_s^{n,i}$ is a rough integral defined as in Lemma A.4.1.

Let $\widetilde{\mathcal{P}}^m = \{0 = r_0^m < r_1^m < \cdots < r_{\widetilde{N}_m}^m = T\}$, $m \in \mathbb{N}$, be any sequence of partitions with mesh size converging to 0, such that $\mathcal{P}^n \subseteq \widetilde{\mathcal{P}}^m$ for every $m \in \mathbb{N}$. By Lemma 3.1.9, $(\cdot, W)^n$ satisfies Property (RIE) relative to p and the sequence $(\widetilde{\mathcal{P}}^m)_{m\in\mathbb{N}}$. In much the same way as in the proof of Theorem 3.1.2, one can show that the rough integral with respect to $(\cdot, \mathbf{W})^n$ is equal to a limit of left-point Riemann sums along $(\widetilde{\mathcal{P}}^m)_{m\in\mathbb{N}}$. That is, for any $t \in [0, T]$, we have

$$\begin{aligned} A_t^{n,i} &:= \sum_{k=0}^{N_n-1} (b_{t_k^n}^{n,i} - \frac{1}{2}\sigma_{t_k^n}^{n,i\cdot}(\sigma_{t_k^n}^{n,i\cdot})^\top, \sigma_{t_k^n}^{n,i\cdot})(\gamma^n, W^n)_{t_k^n\wedge t, t_{k+1}^n\wedge t} \\ &= \int_0^t (b_s^{n,i} - \frac{1}{2}\sigma_s^{n,i\cdot}(\sigma_s^{n,i\cdot})^\top, \sigma_s^{n,i\cdot}) \,\mathrm{d}(\cdot, \mathbf{W})_s^n. \end{aligned}$$

We further obtain that for any $t \in [0, T]$,

$$\begin{aligned} \kappa_t^n &:= \frac{1}{K_T} \exp\left(\frac{1}{2}\int_0^t ((h_s^n)^\top b_s^n, (h_s^n)^\top \sigma_s^n) \,\mathrm{d}(\gamma^n, W^n)_s\right) \\ &= \frac{1}{K_T} \exp\left(\frac{1}{2}\int_0^t ((h_s^n)^\top b_s^n, (h_s^n)^\top \sigma_s^n) \,\mathrm{d}(\cdot, \mathbf{W})_s^n\right). \end{aligned}$$

Similarly, the associativity property of rough integrals and Theorem 4.2.1, for any $t \in [0, T]$, it holds that the rough integral of the controlled path $\varphi^{n,i}$ against the controlled path $S^{n,i}$ is given by left-point Riemann sums, and so is the integral of κ^n against K since the paths are of finite 1-variation.

Now we are finally able to prove Theorem 4.4.6.

Proof. Let $p' \in (p, 3)$. It follows by interpolation (see, e.g., [74, Proposition 5.5]) that

$$\begin{aligned} &\|b^n; b\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}, \mathcal{V}^{p'}_{(\cdot,W)}} \\ &= \|(b^n)' - b'\|_{p'} + \|R^{b^n} - R^b\|_{\frac{p'}{2}} \\ &\leq \|(b^n)' - b'\|_\infty^{1-\frac{p}{p'}} \|(b^n)' - b'\|_p^{\frac{p}{p'}} + \|R^{b^n} - R^b\|_\infty^{1-\frac{p}{p'}} \|R^{b^n} - R^b\|_{\frac{p}{2}}^{\frac{p}{p'}} \\ &\lesssim \|(b^n)' - b'\|_\infty^{1-\frac{p}{p'}} + \|R^{b^n} - R^b\|_\infty^{1-\frac{p}{p'}}, \end{aligned}$$

where the implicit multiplicative constant depends only on p, p', and $\|b\|_{\mathcal{V}^p_{(\cdot,W)}}$ because $\sup_{n\in\mathbb{N}} \|(b^n)'\|_p \leq \|b'\|_p$ and $\sup_{n\in\mathbb{N}} \|R^{b^n}\|_{\frac{p}{2}} \leq \|R^b\|_{\frac{p}{2}}$, analogously for σ. Since b^n converges uniformly to b as $n \to \infty$, we deduce that

$$\|b^n; b\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}, \mathcal{V}^{p'}_{(\cdot,W)}} \longrightarrow 0 \qquad \text{as} \qquad n \longrightarrow \infty;$$

analogously for σ.

Following the arguments of the proof of Theorem 4.4.5 and applying the estimate in Lemma 4.3.10, one can show that

$$\|\kappa^n\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}} + \|\kappa\|_{\mathcal{V}^{p'}_{(\cdot,W)}} \leq C$$

and

$$\|V\|_{\mathcal{V}^{p'}_{(\cdot,W)}} \leq C,$$

where $C > 0$ depends only on p', k, d, T, $\|b\|_{\mathcal{V}^{p}_{(\cdot,W)}}$, $\|\sigma\|_{\mathcal{V}^{p}_{(\cdot,W)}}$, $1/\inf_t|\det(\sigma_t\sigma_t^\top)|$, $\|W\|_p$, $w(0,T)$, where w is the control function for which (4.6) holds for $(\cdot, W)$, and the consumption clock K, and

$$\|\kappa^n;\kappa\|_{\mathcal{V}^{p'}_{(\cdot,W)^n},\mathcal{V}^{p'}_{(\cdot,W)}} \lesssim \|b^n;b\|_{\mathcal{V}^{p'}_{(\cdot,W)^n},\mathcal{V}^{p'}_{(\cdot,W)}} + \|\sigma^n;\sigma\|_{\mathcal{V}^{p'}_{(\cdot,W)^n},\mathcal{V}^{p'}_{(\cdot,W)}} + \|(\cdot,\mathbf{W})^n;(\cdot,\mathbf{W})\|_{p'},$$

where the implicit multiplicative constant depends only on p', k, d, T, $\|b\|_{\mathcal{V}^{p}_{(\cdot,W)}}$, $\|\sigma\|_{\mathcal{V}^{p}_{(\cdot,W)}}$, $1/\inf_t|\det(\sigma_t\sigma_t^\top)|$, $\|W\|_p$, $w(0,T)$, and K. The same (stability) estimates hold for $S^{n,i}$, S^i, $\varphi^{n,i}$, φ^i, $i = 1,\ldots,k$, where the respective constants also depend on s_0. This allows us to also conclude the same (stability) estimate for V^n, V using standard estimates for Young and rough integration (e.g., [75, Proposition 2.4, Lemma 3.1, Lemma 3.6 and Lemma 3.7]. We can further apply Lemma A.4.2 and Lemma A.4.4 and obtain that

$$\|\varphi^{n,0} - \varphi^0\|_{p'} \lesssim \|\varphi^{n,i};\varphi^i\|_{\mathcal{V}^{p'}_{(\cdot,W)^n},\mathcal{V}^{p'}_{(\cdot,W)}} + \|S^{n,i};S^i\|_{\mathcal{V}^{p'}_{(\cdot,W)^n},\mathcal{V}^{p'}_{(\cdot,W)}} + \|(\cdot,\mathbf{W})^n;(\cdot,\mathbf{W})\|_{p'},$$

where the implicit multiplicative constant depends only on p', k, T, $\|\varphi^{n,i}\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}}$, $\|\varphi^i\|_{\mathcal{V}^{p'}_{(\cdot,W)}}$, $\|S^{n,i}\|_{\mathcal{V}^{p'}_{(\cdot,W)^n}}$, $\|S^i\|_{\mathcal{V}^{p'}_{(\cdot,W)}}$, $\|W\|_p$, and $w(0,T)$, that is,

$$\|\varphi^{n,0} - \varphi^0\|_{p'} \lesssim \|b^n;b\|_{\mathcal{V}^{p'}_{(\cdot,W)^n},\mathcal{V}^{p'}_{(\cdot,W)}} + \|\sigma^n;\sigma\|_{\mathcal{V}^{p'}_{(\cdot,W)^n},\mathcal{V}^{p'}_{(\cdot,W)}} + \|(\cdot,\mathbf{W})^n;(\cdot,\mathbf{W})\|_{p'},$$

where the implicit multiplicative constant depends only on p', k, d, T, $\|b\|_{\mathcal{V}^{p}_{(\cdot,W)}}$, $\|\sigma\|_{\mathcal{V}^{p}_{(\cdot,W)}}$, $1/\inf_t|\det(\sigma_t\sigma_t^\top)|$, s_0, $\|W\|_p$, $w(0,T)$, and K.

Combining this with the estimate derived above and the rate of convergence stated in Lemma 4.3.10, we obtain the convergence and the estimate. □

Remark 4.4.8. *If we assume stronger regularity properties of the "driving noise" path W and the sequence of partitions, we can make the quantitative estimates, provided in Theorem 4.4.5, more explicit, for instance, considering the regularity properties of Brownian sample paths, cf. Corollary 4.3.11 and Corollary 4.3.13 and the respective remarks for the local volatility model.*

Furthermore, if we assume that $b, \sigma \in D^{\frac{p}{2}}$*, i.e.,* $b', (b^n)', \sigma', (\sigma')^n = 0$ *and* $R^b = b$, $R^{b^n} = b^n$, $R^\sigma = \sigma$, $R^{\sigma^n} = \sigma^n$, $n \in \mathbb{N}$*, the rate of convergence in Theorem 4.4.6 becomes more tractable, namely,*

$$\begin{aligned}
&\|(\varphi^n, \kappa^n) - (\varphi, \kappa)\|_{p'} \\
&\quad \lesssim \|b^n - b\|_\infty^{1-\frac{p}{p'}} + \|\sigma^n - \sigma\|_\infty^{1-\frac{p}{p'}} \\
&\qquad + (|\mathcal{P}^n| + \|W^n - W\|_\infty)^{1-\frac{p}{p'}} + \left(|\mathcal{P}^n|^{1-\frac{1}{q}} + \left\| \int_0^\cdot W_t^n \otimes \mathrm{d}W_t - \int_0^\cdot W_t \otimes \mathrm{d}W_t \right\|_\infty\right)^{1-\frac{p}{p'}},
\end{aligned}$$

and

$$\begin{aligned}
&\|V^n - V\|_{p'} \\
&\quad \lesssim \|b^n - b\|_\infty^{1-\frac{p}{p'}} + \|\sigma^n - \sigma\|_\infty^{1-\frac{p}{p'}} \\
&\qquad + (|\mathcal{P}^n| + \|W^n - W\|_\infty)^{1-\frac{p}{p'}} + \left(|\mathcal{P}^n|^{1-\frac{1}{q}} + \left\| \int_0^\cdot W_t^n \otimes \mathrm{d}W_t - \int_0^\cdot W_t \otimes \mathrm{d}W_t \right\|_\infty\right)^{1-\frac{p}{p'}}.
\end{aligned}$$

Chapter 5

Existence of general pathwise stochastic integration

The theory of stochastic integration due to K. Itô has proven to be a very suitable tool for modeling dynamical systems which evolve randomly in time. It is an elegant theory that comes with properties desirable for various applications, and allows for dealing with a rich class of probabilistic models whose sample path properties capture the irregularities observed in real-world data.

However, stochastic integration is indeed stochastic and not merely analytic in the sense that the integral is constructed as a limit of approximating Riemann sums in probability, thus one is required to fix a probability measure a priori. Consequently, the stochastic integral is not necessarily well-posed for a given particular sample path of the driving process. This turns out to be a pitfall from the modeling perspective since there is usually only one time series of data available and the inherent probabilistic structure of the underlying process is not known, leading to so-called model risk.

This motivates a "state by state" notion of integration, i.e., sample path by sample path, that is able to handle paths of lower regularity appearing in classical, say, continuous-time financial models such as the sample paths of Brownian motion.

In his seminal paper [67], Föllmer introduced such a pathwise integration theory as a first deterministic analog to Itô's stochastic integration theory, which is based on a suitable notion of quadratic variation.

Assuming that a path $X\colon [0,T] \to \mathbb{R}$ has such finite quadratic variation along a given sequence of partitions $\pi = (\pi^n)_{n\in\mathbb{N}}$ of $[0,T]$ with vanishing mesh size, he showed that for any twice continuously differentiable function f the limit of left-point Riemann sums

$$\int_0^t \nabla f(X_s)\,\mathrm{d}^\pi X_s := \lim_{n\to\infty} \sum_{[u,v]\in\pi^n} \nabla f(X_u)(X_{v\wedge t} - X_{u\wedge t}), \qquad t \in [0,T], \tag{5.1}$$

exists, where ∇f denotes the gradient of f, and the integral satisfies a "pathwise Itô formula". The Föllmer integral has found various applications and extensions in the pathwise approach to stochastic analysis, see [39, 9, 40, 34], to name but a few. As it is approximated by left-point Riemann sums, the Föllmer integral comes with a clear financial interpretation as the capital gains process which is generated by continuous-time trading. It has therefore been successfully applied in the context of model-free approaches to mathematical finance, where one assumes no underlying probabilistic structure. We refer to [128, 68, 52, 152, 48].

We now aim at generalizing the notion of the Föllmer integral in the sense that for a given $\gamma \in [0,1]$ and a suitable class of (non-gradient type) integrands Y, the general pathwise integral

$$\int_0^t Y_s \,\mathrm{d}^{\gamma,\pi} X_s := \lim_{n\to\infty} \int_0^t Y_s \,\mathrm{d}^{\gamma,\pi^n} X_s, \qquad t \in [0,T], \tag{5.2}$$

exists as a uniform limit along the sequence of partitions $\pi = (\pi^n)_{n\in\mathbb{N}}$, where

$$\int_0^t Y_s \,\mathrm{d}^{\gamma,\pi^n} X_s := \sum_{[u,v]\in\pi^n} (Y_u + \gamma(Y_v - Y_u))(X_{v\wedge t} - X_{u\wedge t}).$$

That is, we consider limits of approximating Riemann sums, where the integrand is given as a convex combination $\gamma Y_u + (1-\gamma) Y_v$ of the values of Y_u and Y_v, with $[u,v]$ being a partition interval. We notice that $\gamma = 0$ corresponds to (forward) Itô-type integration, $\gamma = \frac{1}{2}$ to Stratonovich-type integration and $\gamma = 1$ to backward Itô-type integration, these being the most popular choices in applications.

For this purpose, we rely on rough path theory, initiated by Lyons [129], as it offers a general pathwise integration theory beyond the notion of Young integration able to handle paths of lower regularity.

Rough path theory is based on the insight that, since the path is not regular enough, one is required to "enhance" its informational structure to define an integral. The rough integral is then defined not as the limit of classical Riemann sums but as the limit of so-called compensated Riemann sums that involve a higher-order term mimicking the value of the iterated integral of the path against itself (its "area"). This is precisely what is required to guarantee continuity of the integral map, yielding strong stability estimates.

The task is therefore to identify a path property on the integrator which ensures that the path can be enhanced canonically so that the rough integral exists as a limit of general Riemann sums, in spirit of the Föllmer integral, see above.

It is Property (RIE), first introduced in [143] for continuous paths and extended to càdlàg paths in [7], that recovers the rough integral as a limit of left-point, i.e., non-anticipative Riemann sums. It has been applied to robust and model-free finance, see also [5], and has been extensively used in Chapter 3 and Chapter 4, in the regime of càdlàg paths.

This chapter is structured as follows. In Section 5.1 we will generalize Property (RIE) and show that it is a sufficient condition on the integrator for the rough integral to exist as a limit of general Riemann sums; this is the so-called Property γ-(RIE).

This path property turns out to be rather natural, see Section 5.2: for the Stratonovich-type integration, it is equivalent to imposing the existence of its pathwise Lévy area and a certain regularity condition of the path and along the sequence of partitions, which seems fitting in the regime of rough path theory. And, for the more general case, Property γ-(RIE) is equivalent to additionally imposing the existence of the pathwise quadratic variation, which formally links it to the Föllmer integral; all of which holds true for almost all sample paths of the Brownian motion.

Moreover, Property γ-(RIE) is satisfied by various examples of stochastic processes along suitable sequences of partitions, making the established pathwise integration theory, particularly the Stratonovich-type, applicable to the stochastic setting, see Section 5.3.

We want to mention at this point that the above construction of the pathwise integral (and Property γ-(RIE), and the quadratic variation and the Lévy area) depend strongly on the choice of the sequence of partitions. Therefore it is of interest and intended for future work to investigate the invariance of the Lévy area with respect to the sequence of partitions, in light of [38], where they derive a result about the invariance of the quadratic variation with respect to the sequence of partitions, to obtain a robust formulation of the pathwise integration constructed in this chapter.

5.1 The rough integral as a limit of general Riemann sums

Before we introduce the path properties rigorously and develop a notion of general rough integration under Property γ-(RIE), we will first recall the essentials from the theory of rough paths. For a more detailed exposition of rough path theory, we refer to [130, 74, 71].

5.1.1 Essentials on rough path theory

A *partition* $\mathcal{P}$ of an interval $[s,t]$ is a finite set of points between and including the points s and t, i.e., $\mathcal{P} = \{s = u_0 < u_1 < \cdots < u_N = t\}$ for some $N \in \mathbb{N}$, and its mesh size is denoted by $|\mathcal{P}| := \max\{|u_{i+1} - u_i| \colon i = 0, \ldots, N-1\}$.

Throughout, we let $T > 0$ be a fixed finite time horizon. We let $\Delta_T := \{(s,t) \in [0,T]^2 : s \le t\}$ denote the standard 2-simplex.

A function $w \colon \Delta_T \to [0,\infty)$ is called a *control function* if it is superadditive, in the sense that $w(s,u) + w(u,t) \le w(s,t)$ for all $0 \le s \le u \le t \le T$.

For two vectors $x = (x^1, \ldots, x^d)^\top, y = (y^1, \ldots, y^d)^\top \in \mathbb{R}^d$ we use the usual tensor product

$$x \otimes y := (x^i y^j)_{i,j=1,\ldots,d} \in \mathbb{R}^{d \times d}.$$

We shall write $a \lesssim b$ to mean that there exists a constant $C > 0$ such that $a \leq Cb$. The constant C may depend on the normed space, e.g. through its dimension or regularity parameters.

For two vector spaces, the space of linear maps from $E_1 \to E_2$ is denoted by $\mathcal{L}(E_1, E_2)$.

For a normed space $(E, |\cdot|)$, we let $C([0,T];E)$ denote the set of continuous paths from $[0,T]$ to E. For $p \geq 1$, the p-variation of a path $X \in C([0,T];E)$ is given by

$$\|X\|_p := \|X\|_{p,[0,T]} \qquad \text{with} \qquad \|X\|_{p,[s,t]} := \Bigg(\sup_{\mathcal{P} \subset [s,t]} \sum_{[u,v] \in \mathcal{P}} |X_v - X_u|^p \Bigg)^{\frac{1}{p}}, \quad (s,t) \in \Delta_T,$$

where the supremum is taken over all possible partitions $\mathcal{P}$ of the interval $[s,t]$. We recall that, given a path X, we have that $\|X\|_p < \infty$ if and only if there exists a control function c such that[1]

$$\sup_{(u,v) \in \Delta_T} \frac{|X_v - X_u|^p}{c(u,v)} < \infty.$$

We write $C^{p\text{-var}} = C^{p\text{-var}}([0,T];E)$ for the space of paths $X \in C([0,T];E)$ which satisfy $\|X\|_p < \infty$. Moreover, for a path $X \in C([0,T];\mathbb{R}^d)$, we will often use the shorthand notation:

$$X_{s,t} := X_t - X_s, \qquad \text{for} \quad (s,t) \in \Delta_T.$$

For $r \geq 1$ and a two-parameter function $\mathbb{X}\colon \Delta_T \to E$, we similarly define

$$\|\mathbb{X}\|_r := \|\mathbb{X}\|_{r,[0,T]} \qquad \text{with} \qquad \|\mathbb{X}\|_{r,[s,t]} := \Bigg(\sup_{\mathcal{P} \subset [s,t]} \sum_{[u,v] \in \mathcal{P}} |\mathbb{X}_{u,v}|^r \Bigg)^{\frac{1}{r}}, \quad (s,t) \in \Delta_T.$$

We write $C_2^{r\text{-var}} = C_2^{r\text{-var}}(\Delta_T;E)$ for the space of continuous functions $\mathbb{X}\colon \Delta_T \to E$ which satisfy $\|\mathbb{X}\|_r < \infty$.

For $p \in [2,3)$, a pair $\mathbf{X} = (X, \mathbb{X})$ is called a *(continuous) p-rough path* over $\mathbb{R}^d$ if

(i) $X \in C^{p\text{-var}}([0,T];\mathbb{R}^d)$ and $\mathbb{X} \in C_2^{\frac{p}{2}\text{-var}}(\Delta_T;\mathbb{R}^{d \times d})$, and

(ii) Chen's relation: $\mathbb{X}_{s,t} = \mathbb{X}_{s,u} + \mathbb{X}_{u,t} + X_{s,u} \otimes X_{u,t}$ holds for all $0 \leq s \leq u \leq t \leq T$.

In component form, condition (ii) states that $\mathbb{X}^{ij}_{s,t} = \mathbb{X}^{ij}_{s,u} + \mathbb{X}^{ij}_{u,t} + X^i_{s,u} X^j_{u,t}$ for every i and j. We will denote the space of p-rough paths by $\mathcal{C}^p = \mathcal{C}^p([0,T];\mathbb{R}^d)$. On the space $\mathcal{C}^p([0,T];\mathbb{R}^d)$, we use the natural seminorm

$$\|\mathbf{X}\|_p := \|\mathbf{X}\|_{p,[0,T]} \qquad \text{with} \qquad \|\mathbf{X}\|_{p,[s,t]} := \|X\|_{p,[s,t]} + \|\mathbb{X}\|_{\frac{p}{2},[s,t]},$$

for $(s,t) \in \Delta_T$.

[1] Here and throughout, we adopt the convention that $\frac{0}{0} := 0$.

Let $p \in (2,3)$ and $q > 0$ such that $\frac{2}{p} + \frac{1}{q} > 1$, and $X \in C^{p\text{-var}}([0,T];\mathbb{R}^d)$. We say that a pair (Y, Y') is a *controlled path* (with respect to X), if

$$Y \in C^{p\text{-var}}([0,T];E), \quad Y' \in C^{q\text{-var}}([0,T];\mathcal{L}(\mathbb{R}^d;E)), \quad \text{and} \quad R^Y \in C_2^{r\text{-var}}(\Delta_T;E),$$

where R^Y is defined by

$$Y_{s,t} = Y'_s X_{s,t} + R^Y_{s,t} \qquad \text{for all} \quad (s,t) \in \Delta_T,$$

and $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$. We write $\mathscr{C}_X^{p,q} = \mathscr{C}_X^{p,q}([0,T];E)$ for the space of E-valued controlled paths, which becomes a Banach space when equipped with the norm

$$(Y, Y') \mapsto |Y_0| + |Y'_0| + \|Y'\|_{q,[0,T]} + \|R^Y\|_{r,[0,T]}.$$

Given $p \in (2,3)$, $\mathbf{X} = (X, \mathbb{X}) \in \mathcal{C}^p([0,T];\mathbb{R}^d)$ and $(Y, Y') \in \mathscr{C}_X^{p,q}([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$, the (forward) rough integral

$$\int_s^t Y_r \,\mathrm{d}\mathbf{X}_r := \lim_{|\mathcal{P}| \to 0} \sum_{[u,v] \in \mathcal{P}} (Y_u X_{u,v} + Y'_u \mathbb{X}_{u,v}), \qquad (s,t) \in \Delta_T,$$

exists (in the classical mesh Riemann–Stieltjes sense), where the limit is taken along any sequence of partitions $(\mathcal{P}^n)_{n \in \mathbb{N}}$ of the interval $[s,t]$ such that $|\mathcal{P}^n| \to 0$ as $n \to \infty$. More precisely, in writing the product $Y_u X_{u,v}$, we apply the operator $Y_u \in \mathcal{L}(\mathbb{R}^d;\mathbb{R}^k)$ onto $X_{u,v} \in \mathbb{R}^d$; and in writing the product $Y'_u \mathbb{X}_{u,v}$, we use the natural identification of $\mathcal{L}(\mathbb{R}^d;\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$ with $\mathcal{L}(\mathbb{R}^d \otimes \mathbb{R}^d;\mathbb{R}^k)$. Moreover, the rough integral comes with the estimate

$$\left| \int_s^t Y_u \,\mathrm{d}\mathbf{X}_u - Y_s X_{s,t} - Y'_s \mathbb{X}_{s,t} \right| \leq C\Big(\|R^Y\|_{r,[s,t]} \|X\|_{p,[s,t]} + \|Y'\|_{q,[s,t]} \|\mathbb{X}\|_{\frac{p}{2},[s,t]} \Big)$$

for some constant C depending only on p, q and r; see e.g. [143, Theorem 4.9]. For details on the construction of the rough integral and its properties, we refer to the [129, 82, 71].

5.1.2 Rough integration using Property γ-(RIE)

We resume with the concept of quadratic variation (for a continuous path) along a sequence of partitions, which is associated with the Föllmer integral.

For this, let $\mathcal{B}([0,T])$ denote the Borel σ-algebra on $[0,T]$ and δ_t the Dirac measure at $t \in [0,T]$.

Definition 5.1.1 (Quadratic variation of a path in the sense of Föllmer)**.** *Let $X \in C([0,T];\mathbb{R})$ and $\pi = (\pi^n)_{n \in \mathbb{N}}$, with $\pi^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$, be a sequence of partitions of the interval $[0,T]$ such that* $\sup\{|X_{t_k^n,t_{k+1}^n}| \colon k = 0, \ldots, N_n - 1\}$

converges to 0 *as* $n \to \infty$*. We say that* X *has* quadratic variation along π in the sense of Föllmer *if the sequence of measures* $(\mu_n)_{n\in\mathbb{N}}$ *on* $([0,T],\mathcal{B}([0,T]))$ *defined by*

$$\mu_n := \sum_{k=0}^{N_n-1} (X_{t_k^n+1} - X_{t_k^n})^2 \delta_{t_k^n}$$

converges weakly to a measure μ*. The function* $[X]$ *given by* $[X]_t := \mu([0,t])$ *is called the* quadratic variation *of* X *along* π*. We say that a path* $X \in C([0,T];\mathbb{R}^d)$ *has quadratic variation along* π *in the sense of Föllmer if the above condition holds for* X^i *and* $X^i + X^j$ *for all* $i,j = 1,\dots,d$*. We then set*

$$[X^i, X^j] := \frac{1}{2}([X^i + X^j] - [X^i] - [X^j]).$$

As shown in e.g. [33], an equivalent characterization of the quadratic variation along a sequence of partitions in the sense of Föllmer is the following, which we will then continue with.

Assumption 5.1.2 (Quadratic variation of a path)**.** *Let* $X \in C([0,T];\mathbb{R}^d)$ *and* $\pi = (\pi^n)_{n\in\mathbb{N}}$*, with* $\pi^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$*,* $n \in \mathbb{N}$*, be a sequence of partitions of the interval* $[0,T]$ *such that* $\sup\{|X_{t_k^n,t_{k+1}^n}| \colon k = 0,\dots,N_n - 1\}$ *converges to* 0 *as* $n \to \infty$*. We assume that the quadratic variation*

$$[X]_t^\pi := \lim_{n\to\infty} [X]_t^{\pi^n} := \lim_{n\to\infty} \sum_{k=0}^{N_n-1} X_{t_k^n\wedge t, t_{k+1}^n\wedge t} \otimes X_{t_k^n\wedge t, t_{k+1}^n\wedge t}, \qquad t \in [0,T],$$

exists, where the convergence is uniform in $t \in [0,T]$*.*

We say that a path $X \in C([0,T];\mathbb{R}^d)$ possesses quadratic variation relative to π if X and π together satisfy Assumption 5.1.2.

Given this assumption then, the Föllmer integral (5.1) exists and satisfies a pathwise Itô formula, see [67, Théorème]. However, the integral is well-defined only for functions of gradient-type.

To generalize this and, moreover, to obtain a pathwise integral (5.2) as the limit of general Riemann sums is the aim of this subsection. To this end, we now introduce an additional path property which will allow us to extend the notion of the Föllmer integral. More precisely, we additionally impose the existence of the so-called Lévy area of the path and assume some regularity of the path itself and along the sequence of partitions, which implies the existence of the correct (rough) integral.

Assumption 5.1.3 (Lévy area of a path). *Let $X \in C([0,T];\mathbb{R}^d)$ and let $\pi = (\pi^n)_{n\in\mathbb{N}}$, with $\pi^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$, be a sequence of partitions of the interval $[0,T]$ such that $\sup\{|X_{t_k^n,t_{k+1}^n}| \colon k = 0,\dots,N_n-1\}$ converges to 0 as $n \to \infty$, and let $p \in (2,3)$. We assume that the Lévy area*

$$\mathcal{L}(X,\pi,[0,t]) := \lim_{n\to\infty} \mathcal{L}(X,\pi^n,[0,t]) := \lim_{n\to\infty} \sum_{k=0}^{N_n-1} (X_{t_k^n\wedge t} + X_{t_{k+1}^n\wedge t}) \otimes X_{t_k^n\wedge t, t_{k+1}^n\wedge t}$$

exists, for $t \in [0,T]$, where the convergence is uniform in $t \in [0,T]$, and that there exists a control function c such that

$$\sup_{(s,t)\in\Delta_T} \frac{|X_{s,t}|^p}{c(s,t)} + \sup_{n\in\mathbb{N}} \sup_{0\le k<\ell\le N_n} \frac{|\mathcal{L}(X,\pi^n,[t_k^n,t_\ell^n]) - (X_{t_k^n} + X_{t_\ell^n}) \otimes X_{t_k^n,t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n,t_\ell^n)} \lesssim 1.$$

We say that a path $X \in C([0,T];\mathbb{R}^d)$ possesses Lévy area relative to p and π if p, π and X together satisfy Assumption 5.1.3.

Remark 5.1.4. *The Lévy area was first introduced, in a probabilistic set-up, as the area that is enclosed by any trajectory of the Brownian motion (X^1,X^2) and its chord. It is defined by $\frac{1}{2}(\int_0^T X_t^1 \,\mathrm{d}X_t^2 - \int_0^T X_t^2 \,\mathrm{d}X_t^1)$, which makes sense as a stochastic integral, see [122].*

In our pathwise framework, assuming the respective limits exist, indeed, we obtain that

$$\begin{aligned}
&\mathcal{L}(X,\pi,[0,t])^{ij} - (X_t^i X_t^j - X_0^i X_0^j)\\
&= \lim_{n\to\infty} \sum_{k=0}^{N_n-1} (X^i_{t_k^n\wedge t} + X^i_{t_{k+1}^n\wedge t}) X^j_{t_k^n\wedge t, t_{k+1}^n\wedge t} - \sum_{k=0}^{N_n-1} (X^i_{t_{k+1}^n\wedge t} X^j_{t_{k+1}^n\wedge t} - X^i_{t_k^n\wedge t} X^j_{t_k^n\wedge t})\\
&= \lim_{n\to\infty} \sum_{k=0}^{N_n-1} (X^i_{t_k^n} + \gamma X^i_{t_k^n,t_{k+1}^n}) X^j_{t_k^n\wedge t, t_{k+1}^n\wedge t} - \sum_{k=0}^{N_n-1} (X^j_{t_k^n} + \gamma X^j_{t_k^n,t_{k+1}^n}) X^i_{t_k^n\wedge t, t_{k+1}^n\wedge t}\\
&=: \lim_{n\to\infty} \int_0^t X_s^i \,\mathrm{d}^{\gamma,\pi^n} X_s^j - \int_0^t X_s^j \,\mathrm{d}^{\gamma,\pi^n} X_s^j\\
&=: \int_0^t X_s^i \,\mathrm{d}^{\gamma,\pi} X_s^j - \int_0^t X^j \,\mathrm{d}^{\gamma,\pi} X_s^i,
\end{aligned}$$

for every $i,j = 1,\dots,d$, which thus corresponds to the usual notion of the Lévy area. We further notice that

$$\int_0^t X_s^i \,\mathrm{d}^{\gamma,\pi^n} X_s^j - \int_0^t X_s^j \,\mathrm{d}^{\gamma,\pi^n} X_s^i = \int_0^t X_s^i \,\mathrm{d}^{0,\pi^n} X_s^j - \int_0^t X_s^j \,\mathrm{d}^{0,\pi^n} X_s^i,$$

that is, in terms of general Riemann sums, the pathwise Lévy area is invariant to the choice of γ and coincides with the Itô-type one.

It turns out that an equivalent formulation of Assumption 5.1.3 together with, if $\gamma \neq \frac{1}{2}$, Assumption 5.1.2, is the following path property. It generalizes Property (RIE), as introduced in [143] and [7], which recovers the rough integral as a limit of not compensated but left-point Riemann sums, see [143, Theorem 4.19].

In Section 5.2 we will have a closer look at this and relate these assumptions to one another.

Property γ-(RIE). *Let $X \in C([0,T];\mathbb{R}^d)$ and $\pi = (\pi^n)_{n\in\mathbb{N}}$, with $\pi^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$, be a sequence of partitions of the interval $[0,T]$ such that $\sup\{|X_{t_k^n,t_{k+1}^n}| : k = 0,\dots,N_n-1\}$ converges to 0 as $n \to \infty$, and let $\gamma \in [0,1]$, $p \in (2,3)$.*

We assume that the Riemann sums $\int_0^t X_s \otimes \mathrm{d}^{\gamma,\pi^n} X_s := \sum_{k=0}^{N_n-1}(X_{t_k^n} + \gamma X_{t_k^n,t_{k+1}^n}) \otimes X_{t_k^n\wedge t,t_{k+1}^n\wedge t}$ converge uniformly as $n \to \infty$ to a limit, which we denote by $\int_0^t X_s \otimes \mathrm{d}^\pi X_s$, $t \in [0,T]$, and that there exists a control function c such that

$$\sup_{(s,t)\in\Delta_T} \frac{|X_{s,t}|^p}{c(s,t)} + \sup_{n\in\mathbb{N}} \sup_{0\le k<\ell\le N_n} \frac{|(\int_0^\cdot X_s \otimes \mathrm{d}^{\gamma,\pi^n} X_s)_{t_k^n,t_\ell^n} - X_{t_k^n} \otimes X_{t_k^n,t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n,t_\ell^n)} \lesssim 1. \qquad (5.3)$$

We say that a path $X \in C([0,T];\mathbb{R}^d)$ satisfies Property γ-(RIE) relative to γ, p and π if γ, p, π and X together satisfy Property γ-(RIE).

Under Property γ-(RIE), we now turn to rough path theory and rough integration to derive the existence of a pathwise integral that is given as a limit of general Riemann sums.

To properly define the rough integral, we first fix the correct rough path lift. Note that $\mathbf{X}^0$ corresponds to the Itô-rough path lift and $\mathbf{X}^{\frac{1}{2}}$ corresponds to the Stratonovich-rough path lift of a stochastic process, since the "iterated integral" $\mathbb{X}^0$ and $\mathbb{X}^{\frac{1}{2}}$ is given as a limit of left-point and mid-point Riemann sums, analogously to the stochastic Itô and Stratonovich integral, respectively.

Proposition 5.1.5. *Suppose that $X \in C([0,T];\mathbb{R}^d)$ satisfies Property γ-(RIE) relative to some $\gamma \in [0,1]$, $p \in (2,3)$ and a sequence of partitions $\pi = (\pi^n)_{n\in\mathbb{N}}$. Then X extends canonically to a continuous p-rough path $\mathbf{X}^\gamma := (X, \mathbb{X}^\gamma)$, where*

$$\mathbb{X}^\gamma_{s,t} := \int_0^t X_r \otimes \mathrm{d}^{\gamma,\pi} X_r - \int_0^s X_r \otimes \mathrm{d}^{\gamma,\pi} X_r - X_s \otimes X_{s,t}, \qquad (s,t) \in \Delta_T. \qquad (5.4)$$

Proof. It is straightforward to check that $(X, \mathbb{X}^\gamma)$ satisfies Chen's relation and that $\|X\|_p < \infty$. Therefore it remains to show that $\|\mathbb{X}^\gamma\|_{\frac{p}{2}} < \infty$. We define $X^n\colon [0,T] \to \mathbb{R}^d$ by

$$X_t^n = X_t \mathbf{1}_{\{T\}}(t) + \sum_{k=0}^{N_n-1} X_{t_k^n} \mathbf{1}_{[t_k^n,t_{k+1}^n)}(t), \qquad t \in [0,T].$$

By Property γ-(RIE), we know that

$$\lim_{n\to\infty} \mathbb{X}^{\gamma,\pi^n}_{s,t} = \mathbb{X}^\gamma_{s,t},$$

for $\mathbb{X}^{\gamma,\pi^n}_{s,t} := \int_0^t X_r \otimes \mathrm{d}^{\gamma,\pi^n} X_r - \int_0^s X_r \otimes \mathrm{d}^{\gamma,\pi^n} X_r - X^n_s \otimes X_{s,t}$, $(s,t) \in \Delta_T$, where the convergence is uniform in (s,t). We aim to show that $\sup_{n\in\mathbb{N}} \|\mathbb{X}^{\gamma,\pi^n}\|_{\frac{p}{2}} < \infty$, which then implies by the lower semi-continuity of the $\frac{p}{2}$-variation that

$$\|\mathbb{X}^{\gamma}\|_{\frac{p}{2}} \leq \liminf_{n\to\infty} \|\mathbb{X}^{\gamma,\pi^n}\|_{\frac{p}{2}} < \infty.$$

Let $(s,t) \in \Delta_T$. If there exists k such that $t^n_k \leq s < t \leq t^n_{k+1}$, then we estimate

$$|\mathbb{X}^{\gamma,\pi^n}_{s,t}|^{\frac{p}{2}} = |(X_{t^n_k} + \gamma X_{t^n_k,t^n_{k+1}}) \otimes X_{s,t} - X_{t^n_k} \otimes X_{s,t}|^{\frac{p}{2}} \lesssim |X_{s,t}|^p + |X_{t^n_k,t^n_{k+1}}|^p \lesssim c(t^n_k, t^n_{k+1}). \quad (5.5)$$

Otherwise, let k_0 be the smallest k such that $t^n_k \in (s,t)$, and let k_1 be the largest such k. We decompose

$$\mathbb{X}^{\gamma,\pi^n}_{s,t} = \mathbb{X}^{\gamma,\pi^n}_{s,t^n_{k_0}} + \mathbb{X}^{\gamma,\pi^n}_{t^n_{k_0},t^n_{k_1}} + \mathbb{X}^{\gamma,\pi^n}_{t^n_{k_1},t} + X^n_{s,t^n_{k_0}} \otimes X_{t^n_{k_0},t^n_{k_1}} + X^n_{s,t^n_{k_1}} \otimes X_{t^n_{k_1},t}.$$

By (5.3), we have $|\mathbb{X}^{\gamma,\pi^n}_{t^n_{k_0},t^n_{k_1}}|^{\frac{p}{2}} \lesssim c(t^n_{k_0}, t^n_{k_1})$, and we estimate

$$\begin{aligned}
&|X^n_{s,t^n_{k_0}} \otimes X_{t^n_{k_0},t^n_{k_1}}|^{\frac{p}{2}} + |X^n_{s,t^n_{k_1}} \otimes X_{t^n_{k_1},t}|^{\frac{p}{2}} \\
&\quad \lesssim |X^n_{s,t^n_{k_0}}|^p + |X_{t^n_{k_0},t^n_{k_1}}|^p + |X^n_{s,t^n_{k_1}}|^p + |X_{t^n_{k_1},t}|^p \\
&\quad = |X_{t^n_{k_0-1},t^n_{k_0}}|^p + |X_{t^n_{k_0},t^n_{k_1}}|^p + |X_{t^n_{k_0-1},t^n_{k_1}}|^p + |X_{t^n_{k_1},t}|^p \\
&\quad \lesssim 2c(t^n_{k_0-1}, t).
\end{aligned}$$

Combining this with (5.5), we deduce that $\|\mathbb{X}^{\gamma,\pi^n}\|_{\frac{p}{2}} \lesssim c(0,T)$, and the proof is complete. $\square$

We now proceed similarly to [143]. The following lemma links Property γ-(RIE) to the existence of quadratic variation, which we rely on when calculating the rough integral.

Lemma 5.1.6. *Suppose that $X \in C([0,T];\mathbb{R}^d)$ satisfies Property γ-(RIE) relative to some $\gamma \in [0,1]$, $p \in (2,3)$ and a sequence of partitions $\pi = (\pi^n)_{n\in\mathbb{N}}$. Let $1 \leq i,j \leq d$, and define for $\gamma = \frac{1}{2}$, $[X^i,X^j]^{\gamma,\pi} := 0$, and for $\gamma \neq \frac{1}{2}$,*

$$[X^i,X^j]^{\gamma,\pi}_t := X^i_t X^j_t - X^i_0 X^j_0 - \int_0^t X^i_s \,\mathrm{d}^{\gamma,\pi} X^j_s - \int_0^t X^j_s \,\mathrm{d}^{\gamma,\pi} X^i_s, \qquad t \in [0,T].$$

Then $[X^i,X^j]^{\gamma,\pi}$ is a continuous function and

$$[X^i,X^j]^{\gamma,\pi}_t = \lim_{n\to\infty} [X^i,X^j]^{\gamma,\pi^n}_t := \lim_{n\to\infty} (1-2\gamma) \sum_{k=0}^{N_n-1} X^i_{t^n_k\wedge t,t^n_{k+1}\wedge t} X^j_{t^n_k\wedge t,t^n_{k+1}\wedge t}. \quad (5.6)$$

The sequence $([X^i,X^j]^{\gamma,\pi^n})_{n\in\mathbb{N}}$ has uniformly bounded 1-variation, and in particular, $[X^i,X^j]^{\gamma,\pi}$ has finite 1-variation. We write $[X]^{\gamma,\pi} = [X,X]^{\gamma,\pi} = ([X^i,X^j]^{\gamma,\pi})_{1\leq i,j\leq d}$.

Proof. By definition, the function $[X^i, X^j]^{\gamma,\pi}$ is continuous. We observe that

$$X_t^i X_t^j - X_0^i X_0^j = \sum_{k=0}^{N_n-1} (X^i_{t^n_{k+1}\wedge t} X^j_{t^n_{k+1}\wedge t} - X^i_{t^n_k\wedge t} X^j_{t^n_k\wedge t})$$

for every $n \in \mathbb{N}$, and

$$\begin{aligned} &X^i_{t^n_{k+1}\wedge t} X^j_{t^n_{k+1}\wedge t} - X^i_{t^n_k\wedge t} X^j_{t^n_k\wedge t} \\ &\quad = (X^i_{t^n_k\wedge t} + \gamma X^i_{t^n_k\wedge t, t^n_{k+1}\wedge t}) X^j_{t^n_k\wedge t, t^n_{k+1}\wedge t} + (X^j_{t^n_k\wedge t} + \gamma X^j_{t^n_k\wedge t, t^n_{k+1}\wedge t}) X^i_{t^n_k\wedge t, t^n_{k+1}\wedge t} \\ &\quad\quad + (1-2\gamma) X^i_{t^n_k\wedge t, t^n_{k+1}\wedge t} X^j_{t^n_k\wedge t, t^n_{k+1}\wedge t}. \end{aligned}$$

Since $(\int_0^\cdot X_s \otimes \mathrm{d}^{\gamma,\pi^n} X_s)$ converges uniformly to $\int_0^\cdot X_s \otimes \mathrm{d}^{\gamma,\pi} X_s$, the convergence in (5.6) then holds. We further see that

$$X^i_{t^n_k\wedge t, t^n_{k+1}\wedge t} X^j_{t^n_k\wedge t, t^n_{k+1}\wedge t} = \frac{1}{4}(((X^i + X^j)_{t^n_k\wedge t, t^n_{k+1}\wedge t})^2 - ((X^i - X^j)_{t^n_k\wedge t, t^n_{k+1}\wedge t})^2)$$

(i.e. $[X^i, X^j]^{\gamma,\pi} = \frac{1}{4}([X^i + X^j]^{\gamma,\pi} - [X^i - X^j]^{\gamma,\pi})$). That is, $[X^i, X^j]^{\gamma,\pi^n}$ is given as the difference of two increasing functions, and its 1-variation is bounded from above by

$$\begin{aligned} &(1-2\gamma) \sum_{k=0}^{N_n-1} (((X^i + X^j)_{t^n_k, t^n_{k+1}})^2 + ((X^i - X^j)_{t^n_k, t^n_{k+1}})^2) \\ &\qquad \lesssim (1-2\gamma) \sup_{m\in\mathbb{N}} \sum_{k=0}^{N_m-1} ((X^i_{t^m_k, t^m_{k+1}})^2 + (X^j_{t^m_k, t^m_{k+1}})^2). \end{aligned}$$

Since the right-hand side is finite, we obtain that the limit $[X^i, X^j]^{\gamma,\pi}$ has finite 1-variation. □

With the quadratic variation at hand, we apply a piecewise linear interpolation to continuously approximate the path and obtain a Stratonovich-type integral, that we then translate back into a general pathwise integral.

Lemma 5.1.7. *Suppose that $X \in C([0,T];\mathbb{R}^d)$ satisfies Property γ-(RIE) relative to some $\gamma \in [0,1]$, $p \in (2,3)$ and a sequence of partitions $\pi = (\pi^n)_{n\in\mathbb{N}}$. Define $\bar{X}^n$ as the piecewise linear interpolation of X along $\pi = (\pi^n)_{n\in\mathbb{N}}$. Then*

$$\lim_{n\to\infty} \int_s^t \bar{X}^n_r \otimes \mathrm{d}\bar{X}^n_r = \lim_{n\to\infty} \sum_{k=0}^{N_n-1} (X_{t^n_k} + \frac{1}{2} X_{t^n_k, t^n_{k+1}}) \otimes X_{t^n_k\wedge t, t^n_{k+1}\wedge t} = \int_s^t X_r \otimes \mathrm{d}^{\gamma,\pi} X_r + \frac{1}{2}[X]^{\gamma,\pi}_{s,t}, \tag{5.7}$$

where the convergence is uniform in $(s,t) \in \Delta_T$. Moreover, the sequence $(\bar{\mathbb{X}}^n)_{n\in\mathbb{N}}$ has uniformly bounded $\frac{p}{2}$-variation, where $\bar{\mathbb{X}}^n_{s,t} := \int_s^t \bar{X}^n_{s,r} \otimes \mathrm{d}\bar{X}^n_r$, $(s,t) \in \Delta_T$.

Proof. Let $n \in \mathbb{N}$ and $0 \le k \le N_{n-1}$. By definition, for $t \in [t_k^n, t_{k+1}^n]$, we have

$$\bar{X}_t^n = X_{t_k^n} + \frac{t - t_k^n}{t_{k+1}^n - t_k^n} X_{t_k^n, t_{k+1}^n},$$

which gives that

$$\begin{aligned}\int_{t_k^n}^{t_{k+1}^n} \bar{X}_r^n \otimes \mathrm{d}\bar{X}_r^n &= (X_{t_k^n} + \frac{1}{2} X_{t_k^n, t_{k+1}^n}) \otimes X_{t_k^n, t_{k+1}^n} \\ &= (X_{t_k^n} + \gamma X_{t_k^n, t_{k+1}^n}) \otimes X_{t_k^n, t_{k+1}^n} + \frac{1}{2}(1 - 2\gamma) X_{t_k^n, t_{k+1}^n} \otimes X_{t_k^n, t_{k+1}^n}.\end{aligned} \tag{5.8}$$

Lemma 5.1.6 then implies the uniform convergence and (5.7).

We now show that $(\bar{\mathbb{X}}^n)_{n \in \mathbb{N}}$ has uniformly bounded $\frac{p}{2}$-variation. Let $(s,t) \in \Delta_T$. If $t_k^n \le s < t \le t_{k+1}^n$ for some k, then we estimate

$$\begin{aligned}|\bar{\mathbb{X}}_{s,t}^n|^{\frac{p}{2}} &= \Big| \int_s^t \bar{X}_{s,r}^n \otimes \mathrm{d}\bar{X}_r^n \Big|^{\frac{p}{2}} \le \Big| \int_s^t (r - s) \frac{|X_{t_k^n, t_{k+1}^n}|^2}{|t_{k+1}^n - t_k^n|^2} \, \mathrm{d}r \Big|^{\frac{p}{2}} \\ &= \frac{1}{2^{\frac{p}{2}}} |t - s|^p \frac{|X_{t_k^n, t_{k+1}^n}|^p}{|t_{k+1}^n - t_k^n|^p} \le \frac{|t - s|}{|t_{k+1}^n - t_k^n|} \|X\|_{p,[t_k^n, t_{k+1}^n]}^p.\end{aligned} \tag{5.9}$$

Otherwise, let k_0 be the smallest k such that $t_k^n \in (s,t)$, and let k_1 be the largest such k. It is straightforward to see that $(\bar{X}^n, \bar{\mathbb{X}}^n)$ satisfies Chen's relation:

$$\bar{\mathbb{X}}_{s,t}^n = \bar{\mathbb{X}}_{s,u}^n + \bar{\mathbb{X}}_{u,t}^n + \bar{X}_{s,u}^n \otimes \bar{X}_{u,t}^n$$

for all $s \le u \le t$, from which it follows that

$$\bar{\mathbb{X}}_{s,t}^n = \bar{\mathbb{X}}_{s,t_{k_0}^n}^n + \bar{\mathbb{X}}_{t_{k_0}^n, t_{k_1}^n}^n + \bar{\mathbb{X}}_{t_{k_1}^n, t}^n + \bar{X}_{s,t_{k_0}^n}^n \otimes \bar{X}_{t_{k_0}^n, t_{k_1}^n}^n + \bar{X}_{s,t_{k_1}^n}^n \otimes \bar{X}_{t_{k_1}^n, t}^n.$$

Recalling the calculation (5.8), we get that

$$|\bar{\mathbb{X}}_{t_{k_0}^n, t_{k_1}^n}^n|^{\frac{p}{2}} \lesssim \Big| \Big(\int_0^{\cdot} X_s \otimes \mathrm{d}^{\gamma, \pi^n} X_s \Big)_{t_{k_0}^n, t_{k_1}^n} - X_{t_{k_0}^n} \otimes X_{t_{k_0}^n, t_{k_1}^n} \Big|^{\frac{p}{2}} + |[X]_{t_{k_0}^n, t_{k_1}^n}^{\gamma, \pi^n}|^{\frac{p}{2}},$$

where $[X]^{\gamma, \pi^n}$ was defined in Lemma 5.1.6. Using the inequality in (5.3) and Lemma 5.1.6, we see that there exists a control function $\bar{c}$ such that the right-hand side is bounded from above by $\bar{c}(t_{k_0}^n, t_{k_1}^n)$. If we combine this with the estimate (5.9) and a simple estimate for the terms $\bar{X}_{s,t_{k_0}^n}^n \otimes \bar{X}_{t_{k_0}^n, t_{k_1}^n}^n$ and $\bar{X}_{s,t_{k_1}^n}^n \otimes \bar{X}_{t_{k_1}^n, t}^n$, we can conclude that $\|\bar{\mathbb{X}}^n\|_{\frac{p}{2}} \lesssim \bar{c}(0,T) + \|X\|_p^2$, which completes the proof. $\square$

We are now able to prove that the rough integral can be obtained as a limit of general Riemann sums given that the driving path satisfies Property γ-(RIE).

Theorem 5.1.8. *Suppose that $X \in C([0,T];\mathbb{R}^d)$ satisfies Property γ-(RIE) relative to some $\gamma \in [0,1]$, $p \in (2,3)$ and a sequence of partitions $\pi = (\pi^n)_{n\in\mathbb{N}}$. Let $q > 0$ be such that $\frac{2}{p} + \frac{1}{q} > 1$. Let $(Y,Y') \in \mathscr{C}^{p,q}_X([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$ be a controlled path such that Y is continuous. Then the rough integral $\int Y \mathrm{d}\mathbf{X}^\gamma$ satisfies*

$$\int_0^t Y_s \,\mathrm{d}\mathbf{X}^\gamma_s = \lim_{n\to\infty} \sum_{k=0}^{N_n-1} (Y_{t^n_k} + \gamma Y_{t^n_{k+1}}) X_{t^n_k\wedge t, t^n_{k+1}\wedge t},$$

where the convergence is uniform in $t \in [0,T]$.

Proof. We denote by $\bar{X}^n$ and $\bar{Y}^n$ the piecewise linear interpolation of X and Y, respectively, along $\pi = (\pi^n)_{n\in\mathbb{N}}$. Thus $(\bar{Y}^n, Y')$ is controlled by $\bar{X}^n$, with remainder $R^{\bar{Y}^n}_{s,t} = \bar{Y}^n_{s,t} - Y'_s \bar{X}^n_{s,t}$, $(s,t) \in \Delta_T$.

As shown in the proof of [143, Theorem 4.19], if $p' > p$ and $q' > q$ such that $\frac{2}{p'} + \frac{1}{q'} > 0$, then $(\bar{Y}^n, Y', R^{\bar{Y}^n})$ converges in (q',p',r')-variation to (Y,Y',R^Y), where $\frac{1}{r'} = \frac{1}{p'} + \frac{1}{q'}$.

Since the sequence $(\bar{X}^n)_{n\in\mathbb{N}}$ has uniformly bounded p-variation and $\bar{X}^n$ converges uniformly to X as $n \to \infty$, it follows by interpolation that $\bar{X}^n$ converges to X with respect to the p'-variation norm, i.e. $\|\bar{X}^n - X\|_{p'} \to 0$ as $n \to \infty$. It follows similarly using Lemma 5.1.7 that $\|(\bar{\mathbb{X}}^n - (\mathbb{X}^\gamma + \frac{1}{2}[X]^\gamma)\|_{\frac{p'}{2}} \to 0$ and hence, also that $\|(\bar{X}^n, \bar{\mathbb{X}}^n) - (X, \mathbb{X}^\gamma + \frac{1}{2}[X]^\gamma)\|_{p'} \to 0$ as $n \to \infty$.

The continuity of the Itô–Lyons map, see e.g. [71, Theorem 4.17], now yields the uniform convergence of the rough integrals $\int \bar{Y}^n \,\mathrm{d}(\bar{X}^n, \bar{\mathbb{X}}^n)$ to the rough integral $\int Y \,\mathrm{d}(X, \mathbb{X}^\gamma + \frac{1}{2}[X]^\gamma)$. But for every $t \in [0,T]$, it holds that

$$\begin{aligned}
&\lim_{n\to\infty} \int_0^t \bar{Y}^n_s \,\mathrm{d}(\bar{X}^n, \bar{\mathbb{X}}^n)_s \\
&\quad = \lim_{n\to\infty} \int_0^t \bar{Y}^n_s \,\mathrm{d}\bar{X}^n_s \\
&\quad = \lim_{n\to\infty} \sum_{k=0}^{N_n-1} (Y_{t^n_k} + \frac{1}{2} Y_{t^n_k, t^n_{k+1}}) X_{t^n_k\wedge t, t^n_{k+1}\wedge t} \\
&\quad = \lim_{n\to\infty} \left(\sum_{k=0}^{N_n-1} (Y_{t^n_k} + \gamma Y_{t^n_k, t^n_{k+1}}) X_{t^n_k\wedge t, t^n_{k+1}\wedge t} + \frac{1}{2}(1-2\gamma) \sum_{k=0}^{N_n-1} Y_{t^n_k, t^n_{k+1}} X_{t^n_k\wedge t, t^n_{k+1}\wedge t} \right).
\end{aligned}$$

Since $(Y,Y') \in \mathscr{C}^{p,q}_X$, it is immediate that the second term on the right-hand side converges uniformly to $\frac{1}{2}\int_0^t Y_s' \,\mathrm{d}[X]_s^{\gamma,\pi}$, $t \in [0,T]$. Thus,

$$
\begin{aligned}
&\lim_{n\to\infty} \sum_{k=0}^{N_n-1} (Y_{t_k^n} + \gamma Y_{t_k^n,t_{k+1}^n}) X_{t_k^n\wedge t, t_{k+1}^n \wedge t} \\
&= \lim_{n\to\infty} \int_0^t Y_s^n \,\mathrm{d}(\bar{X}^n, \bar{\mathbb{X}}^n)_s - \frac{1}{2}\int_0^t Y_s' \,\mathrm{d}[X]_s^{\gamma,\pi} \\
&= \int_0^t Y_s \,\mathrm{d}(X, \mathbb{X}^{\gamma,\pi} + \frac{1}{2}[X]^{\gamma,\pi})_s - \frac{1}{2}\int_0^t Y_s' \,\mathrm{d}[X]_s^{\gamma,\pi} \\
&= \lim_{|\mathcal{P}|\to 0} \sum_{[u,v]\in\mathcal{P}} Y_u X_{u,v} + Y_u'(\mathbb{X}^\gamma + \frac{1}{2}[X]^{\gamma,\pi})_{u,v} - \frac{1}{2} \lim_{|\mathcal{P}|\to 0} \sum_{[u,v]\in\mathcal{P}} Y_u' [X]_{u,v}^{\gamma,\pi} \\
&= \lim_{|\mathcal{P}|\to 0} \sum_{[u,v]\in\mathcal{P}} Y_u X_{u,v} + Y_u' \mathbb{X}_{u,v}^\gamma \\
&= \int_0^t Y_s \,\mathrm{d}\mathbf{X}_s^\gamma,
\end{aligned}
$$

where the limit is taken over any sequence of partitions $\mathcal{P}$ of the interval $[0,t]$ with mesh size $|\mathcal{P}| \to 0$. $\square$

5.2 On the assumption of general Riemann integrals

Theorem 5.1.8 is a generalization of [143, Theorem 4.19] which states that the rough integral can be calculated as a limit of left-point Riemann sums given that the driving path satisfies Property γ-(RIE) for $\gamma = 0$. This assumption is also known as Property (RIE) and states as follows for a continuous path:

Property (RIE). *Let $X \in C([0,T];\mathbb{R}^d)$ and $\pi = (\pi^n)_{n\in\mathbb{N}}$, with $\pi^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$, be a sequence of partitions of the interval $[0,T]$ such that $\sup\{|X_{t_k^n,t_{k+1}^n}| : k = 0,\dots,N_n - 1\}$ converges to 0 as $n \to \infty$, and let $p \in (2,3)$.*

We assume that the left-point Riemann sums $\int_0^t X_s \otimes \mathrm{d}^{\pi^n} X_s := \sum_{k=0}^{N_n-1} X_{t_k^n} \otimes X_{t_k^n\wedge t, t_{k+1}^n\wedge t}$ converge uniformly as $n \to \infty$ to a limit, which we denote by $\int_0^t X_s \otimes \mathrm{d}^\pi X_s$, $t \in [0,T]$, and that there exists a control function c such that

$$
\sup_{(s,t)\in\Delta_T} \frac{|X_{s,t}|^p}{c(s,t)} + \sup_{n\in\mathbb{N}} \sup_{0\le k<\ell\le N_n} \frac{|(\int_0^\cdot X_s \otimes \mathrm{d}^{\pi^n} X_s)_{t_k^n,t_\ell^n} - X_{t_k^n} \otimes X_{t_k^n,t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n,t_\ell^n)} \lesssim 1.
$$

We say that a path $X \in C([0,T];\mathbb{R}^d)$ satisfies Property (RIE) relative to p and π if p, π and X together satisfy Property (RIE).

In the following, we relate Property γ-(RIE) to Property (RIE), depending on the parameter γ, which determines the type of Riemann sum approximation one obtains.

Lemma 5.2.1. *Let $X \in C([0,T];\mathbb{R}^d)$, $\gamma \in [0,1]$, $p \in (2,3)$ and $\pi = (\pi^n)_{n\in\mathbb{N}}$ be a sequence of partitions of $[0,T]$.*

(i) Suppose $\gamma \neq \frac{1}{2}$. X satisfies Property (RIE) *if and only if X satisfies Property γ-*(RIE)*, both relative to p and π.*

(ii) Suppose $\gamma = \frac{1}{2}$. If X satisfies Property (RIE)*, then X satisfies Property γ-*(RIE)*, both relative to p and π.*

Proof. First note that

$$
\begin{aligned}
\int_0^t X_s \otimes \mathrm{d}^{\gamma,\pi^n} X_s &:= \sum_{k=0}^{N_n-1} (X_{t_k^n} + \gamma X_{t_k^n, t_{k+1}^n}) \otimes X_{t_k^n \wedge t, t_{k+1}^n \wedge t} \\
&= \sum_{k=0}^{N_n-1} X_{t_k^n} \otimes X_{t_k^n \wedge t, t_{k+1}^n \wedge t} + \gamma \sum_{k=0}^{N_n-1} X_{t_k^n \wedge t, t_{k+1}^n \wedge t} \otimes X_{t_k^n \wedge t, t_{k+1}^n \wedge t} \\
&\quad + \gamma \sum_{k=0}^{N_n-1} (X_{t_{k+1}^n \wedge t, t_{k+1}^n} - X_{t_k^n \wedge t, t_k^n}) \otimes X_{t_k^n \wedge t, t_{k+1}^n \wedge t} \\
&= \int_0^t X_s \otimes \mathrm{d}^{\pi^n} X_s + \gamma [X]_t^{\pi^n} + \gamma \sum_{k=0}^{N_n-1} (X_{t_{k+1}^n \wedge t, t_{k+1}^n} - X_{t_k^n \wedge t, t_k^n}) \otimes X_{t_k^n \wedge t, t_{k+1}^n, \wedge t},
\end{aligned}
\tag{5.10}
$$

for $t \in [0,T]$, where we write $[X]^{\pi^n} = \sum_{k=1}^{N_n-1} X_{t_k^n \wedge \cdot, t_{k+1}^n \wedge \cdot} \otimes X_{t_k^n \wedge \cdot, t_{k+1}^n \wedge \cdot}$. Secondly, note that, for any control function c, we have

$$
\begin{aligned}
&\sup_{0 \le k < \ell \le N_n} \frac{|(\int_0^\cdot X_s \otimes \mathrm{d}^{\gamma,\pi^n} X_s)_{t_k^n, t_\ell^n} - X_{t_k^n} \otimes X_{t_k^n, t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n, t_\ell^n)} \\
&\quad \le 2^{\frac{p}{2}} \sup_{0 \le k < \ell \le N_n} \frac{|(\int_0^\cdot X_s \otimes \mathrm{d}^{\pi^n} X_s)_{t_k^n, t_\ell^n} - X_{t_k^n} \otimes X_{t_k^n, t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n, t_\ell^n)} + (2\gamma)^{\frac{p}{2}} \sup_{0 \le k < \ell \le N_n} \frac{|[X]^{\pi^n}_{t_k^n, t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n, t_\ell^n)},
\end{aligned}
\tag{5.11}
$$

and, for $\gamma \neq \frac{1}{2}$, we get

$$
\begin{aligned}
&\sup_{0 \le k < \ell \le N_n} \frac{|(\int_0^\cdot X_s \otimes \mathrm{d}^{\pi^n} X_s)_{t_k^n, t_\ell^n} - X_{t_k^n} \otimes X_{t_k^n, t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n, t_\ell^n)} \\
&\quad \le 2^{\frac{p}{2}} \sup_{0 \le k < \ell \le N_n} \frac{|(\int_0^\cdot X_s \otimes \mathrm{d}^{\gamma,\pi^n} X_s)_{t_k^n, t_\ell^n} - X_{t_k^n} \otimes X_{t_k^n, t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n, t_\ell^n)} \\
&\qquad + \frac{(2\gamma)^{\frac{p}{2}}}{|1 - 2\gamma|^{\frac{p}{2}}} \sup_{0 \le k < \ell \le N_n} \frac{|[X]^{\gamma,\pi^n}_{t_k^n, t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n, t_\ell^n)}.
\end{aligned}
\tag{5.12}
$$

If X satisfies Property (RIE), $(\int_0^\cdot X_s \otimes \mathrm{d}^{\pi^n} X_s)_{n \in \mathbb{N}}$ converges uniformly to $(\int_0^\cdot X_s \otimes \mathrm{d}^{\pi} X_s)$ and, by [143, Lemma 4.17], $([X]^{\pi^n})_{n \in \mathbb{N}}$ converges uniformly to $[X]^\pi$ as $n \to \infty$. Moreover, again due to [143, Lemma 4.17], $([X]^{\pi^n})_{n \in \mathbb{N}}$ has uniformly bounded 1-variation. Hence, by (5.10) and (5.11), Property (RIE) implies Property γ-(RIE) for every $\gamma \in [0,1]$.

Conversely, if $\gamma \neq \frac{1}{2}$, using Lemma 5.1.6, (5.10) and (5.12) yields that Property γ-(RIE) implies Property (RIE). $\square$

To gain a better understanding of these assumptions, we make the following observation:

Remark 5.2.2. *Let $X \in C([0,T];\mathbb{R}^d)$, $\gamma \in [0,1]$ and $\pi = (\pi^n)_{n\in\mathbb{N}}$ be a sequence of partitions of $[0,T]$.*

Assuming the respective limits exist, we write $(\mathbb{X}^{\gamma,\pi}_{s,t})^{ij} = \int_0^t X^i_r \,\mathrm{d}^{\gamma,\pi} X^j_r - \int_0^s X^i_r \,\mathrm{d}^{\gamma,\pi} X^j_r - X^i_s X^j_{s,t}$, $(s,t) \in \Delta_T$. We decompose the iterated integrals into the symmetric and antisymmetric components as follows:

$$\begin{aligned}
(\mathbb{X}^{\gamma,\pi}_{s,t})^{ij} &= \frac{1}{2}((\mathbb{X}^{\gamma,\pi}_{s,t})^{ij} + (\mathbb{X}^{\gamma,\pi}_{s,t})^{ij}) + \frac{1}{2}((\mathbb{X}^{\gamma,\pi}_{s,t})^{ij} - (\mathbb{X}^{\gamma,\pi}_{s,t})^{ij}) \\
&= \frac{1}{2}(X^i_{s,t}X^j_{s,t} - [X^i, X^j]^{\gamma,\pi}_{s,t}) + \frac{1}{2}(\mathcal{L}(X,\pi,[0,\cdot])_{s,t} - (X^i_s + X^i_t)X^j_{s,t}) \\
&=: \frac{1}{2}(\mathbb{S}(X)^{\gamma,\pi}_{s,t})^{ij} + \frac{1}{2}(\mathbb{A}(X)^{\gamma,\pi}_{s,t})^{ij},
\end{aligned}$$

for every $i,j = 1,\dots,d$.

For $\gamma = \frac{1}{2}$, we notice that the symmetric part reduces to $\frac{1}{2}X_{s,t} \otimes X_{s,t}$. We realize that for the Stratonovich-type rough path lift (implying the Stratonovich-type integral) to be well-posed in the rough path sense, it is only required that the antisymmetric Riemann sums converge (which do not depend on γ, see Remark 5.1.4), and that the approximative Lévy area has uniformly bounded $\frac{p}{2}$-variation and the path has finite p-variation. For the more general case, it is additionally required that the symmetric Riemann sums converge. This suffices since the approximative quadratic variation term has uniformly bounded 1-, *thus $\frac{p}{2}$-variation by definition.*

It is therefore not surprising but rather reassuring that X satisfying Property (γ-)(RIE) *is equivalent to X possessing Lévy area together with, if $\gamma \neq \frac{1}{2}$, X possessing quadratic variation, in the sense of Assumption 5.1.2 and Assumption 5.1.3, respectively, which impose these exact assumptions. This is the content of Lemma 5.2.3, Corollary 5.2.4 and Lemma 5.2.5.*

Also from a practical perspective, these assumptions are indeed reasonable since almost all sample paths of Brownian motion, firstly, possess quadratic variation relative to π if $\pi^n \log(n) \to 0$ as $n \to \infty$, see [60], and [122], while notably having infinite p-variation for $p \leq 2$. Secondly, it follows from e.g. [74, Theorem 14.16, Exercise 15.44] and the fact that almost all sample paths of Brownian motion are $\frac{1}{p}$-Hölder continuous for $p \in (2,3)$ that almost all sample paths of Brownian motion possess Lévy area relative to p and, e.g., the dyadic partitions, i.e., $\pi^n = \{kT2^{-n}\}_{k=0}^{2^n}$, $n \in \mathbb{N}$.

Lemma 5.2.3. *Let $X \in C([0,T];\mathbb{R}^d)$, $p \in (2,3)$ and $\pi = (\pi^n)_{n\in\mathbb{N}}$ be a sequence of partitions of $[0,T]$.*

X satisfies Property (RIE) *if and only if X possesses quadratic variation and Lévy area, both relative to p and π.*

Proof. First, note that

$$\begin{aligned}\mathcal{L}(X,\pi^n,[0,t]) &= \sum_{k=0}^{N_n-1}(X_{t_k^n\wedge t}+X_{t_{k+1}^n\wedge t})\otimes X_{t_k^n\wedge t,t_{k+1}^n\wedge t}\\ &= 2\sum_{k=0}^{N_n-1}X_{t_k^n\wedge t}\otimes X_{t_k^n\wedge t,t_{k+1}^n\wedge t}+\sum_{k=0}^{N_n-1}X_{t_k^n\wedge t,t_{k+1}^n\wedge t}\otimes X_{t_k^n\wedge t,t_{k+1}^n\wedge t}\\ &= 2\int_0^t X_s\otimes \mathrm{d}^{\pi^n}X_s+2\sum_{k=0}^{N_n-1}X_{t_k^n,t_k^n\wedge t}\otimes X_{t_k^n\wedge t,t_{k+1}^n\wedge t}+[X]_t^{\pi^n},\end{aligned}\tag{5.13}$$

for $t\in[0,T]$, where we write $[X]^{\pi^n}=\sum_{k=1}^{N_n-1}X_{t_k^n\wedge\cdot,t_{k+1}^n\wedge\cdot}\otimes X_{t_k^n\wedge\cdot,t_{k+1}^n\wedge\cdot}$. Secondly, note that, for any control function c, we have

$$\begin{aligned}&\sup_{0\le k<\ell\le N_n}\frac{|\mathcal{L}(X,\pi^n,[t_k^n,t_\ell^n])-(X_{t_k^n}+X_{t_\ell^n})\otimes X_{t_k^n,t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n,t_\ell^n)}\\ &\lesssim \sup_{0\le k<\ell\le N_n}\frac{|(\int_0^\cdot X_s\otimes \mathrm{d}^{\pi^n}X_s)_{t_k^n,t_\ell^n}-X_{t_k^n}\otimes X_{t_k^n,t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n,t_\ell^n)}\\ &\quad+\sup_{0\le k<\ell\le N_n}\frac{|X_{t_k^n,t_\ell^n}\otimes X_{t_k^n,t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n,t_\ell^n)}+\sup_{0\le k<\ell\le N_n}\frac{|[X]^{\pi^n}_{t_k^n,t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n,t_\ell^n)},\end{aligned}\tag{5.14}$$

and

$$\begin{aligned}&\sup_{0\le k<\ell\le N_n}\frac{|(\int_0^\cdot X_s\otimes \mathrm{d}^{\pi^n}X_s)_{t_k^n,t_\ell^n}-X_{t_k^n}\otimes X_{t_k^n,t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n,t_\ell^n)}\\ &\lesssim \sup_{0\le k<\ell\le N_n}\frac{|\mathcal{L}(X,\pi^n,[t_k^n,t_\ell^n])-(X_{t_k^n}+X_{t_\ell^n})\otimes X_{t_k^n,t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n,t_\ell^n)}\\ &\quad+\sup_{0\le k<\ell\le N_n}\frac{|X_{t_k^n,t_\ell^n}\otimes X_{t_k^n,t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n,t_\ell^n)}+\sup_{0\le k<\ell\le N_n}\frac{|[X]^{\pi^n}_{t_k^n,t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n,t_\ell^n)}.\end{aligned}\tag{5.15}$$

If X satisfies Property (RIE), then $(\int_0^\cdot X_s\otimes \mathrm{d}^{\pi^n}X_s)_{n\in\mathbb{N}}$ converges uniformly to $(\int_0^\cdot X_s\otimes \mathrm{d}^{\pi}X_s)$ and, by [143, Lemma 4.17], X possesses quadratic variation, that is, $([X]^{\pi^n})_{n\in\mathbb{N}}$ converges uniformly to $[X]^\pi$ as $n\to\infty$. And, again due to [143, Lemma 4.17], $([X]^{\pi^n})_{n\in\mathbb{N}}$ has uniformly bounded 1-variation. Hence, by (5.13) and (5.14), if X satisfies Property (RIE), then it possesses quadratic variation and Lévy area.

Conversely, if X possesses quadratic variation, $([X]^{\pi^n})_{n\in\mathbb{N}}$ converges uniformly to $[X]^\pi$ as $n\to\infty$, and as in the proof of [143, Lemma 4.17], one can show that $([X]^{\pi^n})_{n\in\mathbb{N}}$ has uniformly bounded 1-variation. If X possesses Lévy area, $(\mathcal{L}(X,\pi^n,[0,\cdot]))_{n\in\mathbb{N}}$ converges uniformly to $\mathcal{L}(X,\pi,[0,\cdot])$ as $n\to\infty$. By (5.13) and (5.15), if X possesses quadratic variation and Lévy area, it satisfies Property (RIE). □

Corollary 5.2.4. *Let $X\in C([0,T];\mathbb{R}^d)$, $\gamma\in[0,1]$, $p\in(2,3)$ and $\pi=(\pi^n)_{n\in\mathbb{N}}$ be a sequence of partitions of $[0,T]$.*

X satisfies Property γ-(RIE) relative to $\gamma \neq \frac{1}{2}$ if and only if X possesses quadratic variation and Lévy area, both relative to p and π.

Lemma 5.2.5. *Let $X \in C([0,T];\mathbb{R}^d)$, $\gamma \in [0,1]$, $p \in (2,3)$ and $\pi = (\pi^n)_{n\in\mathbb{N}}$ be a sequence of partitions of $[0,T]$.*

X satisfies Property γ-(RIE) relative to $\gamma = \frac{1}{2}$, p and π if and only if X possesses Lévy area relative to p and π.

Proof. We note that

$$\begin{aligned}\mathcal{L}(X,\pi^n,[0,t]) &:= \sum_{k=0}^{N_n-1} (X_{t_k^n\wedge t} + X_{t_{k+1}^n\wedge t}) X_{t_k^n\wedge t, t_{k+1}^n\wedge t} \\ &= 2\int_0^t X_s \otimes \mathrm{d}^{\frac{1}{2},\pi^n} X_s + \sum_{k=0}^{N_n-1} (X_{t_k^n, t_k^n\wedge t} + X_{t_{k+1}^n, t_{k+1}^n\wedge t}) X_{t_k^n\wedge t, t_{k+1}^n\wedge t},\end{aligned} \tag{5.16}$$

for $t \in [0,T]$. And therefore, for any control function c, we have that

$$\begin{aligned}&\sup_{0\le k<\ell\le N_n} \frac{|(\int_0^\cdot X_s \otimes \mathrm{d}^{\frac{1}{2},\pi^n} X_s)_{t_k^n,t_\ell^n} - X_{t_k^n}\otimes X_{t_k^n,t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n,t_\ell^n)} \\ &\le \sup_{0\le k<\ell\le N_n} \frac{|\mathcal{L}(X,\pi^n,[t_k^n,t_\ell^n]) - (X_{t_k^n}+X_{t_\ell^n})\otimes X_{t_k^n,t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n,t_\ell^n)} + \sup_{0\le k<\ell\le N_n} \frac{|X_{t_k^n,t_\ell^n}\otimes X_{t_k^n,t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n,t_\ell^n)},\end{aligned} \tag{5.17}$$

and

$$\begin{aligned}&\sup_{0\le k<\ell\le N_n} \frac{|\mathcal{L}(X,\pi^n,[t_k^n,t_\ell^n]) - X_{t_k^n,t_\ell^n}\otimes X_{t_k^n,t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n,t_\ell^n)} \\ &\lesssim \sup_{0\le k<\ell\le N_n} \frac{|(\int_0^\cdot X_s \otimes \mathrm{d}^{\frac{1}{2},\pi^n} X_s)_{t_k^n,t_\ell^n} - X_{t_k^n}\otimes X_{t_k^n,t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n,t_\ell^n)} + \sup_{0\le k<\ell\le N_n} \frac{|X_{t_k^n,t_\ell^n}\otimes X_{t_k^n,t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n,t_\ell^n)}.\end{aligned} \tag{5.18}$$

If X possesses Lévy area, then $(\mathcal{L}(X,\pi^n,[0,\cdot]))_{n\in\mathbb{N}}$ converges uniformly to $\mathcal{L}(X,\pi,[0,\cdot])$ as $n\to\infty$, and if X satisfies Property γ-(RIE) for $\gamma = \frac{1}{2}$, then $(\int_0^\cdot X_s \otimes \mathrm{d}^{\frac{1}{2},\pi^n} X_s)_{n\in\mathbb{N}}$ converges uniformly to $(\int_0^\cdot X_s \otimes \mathrm{d}^{\frac{1}{2},\pi} X_s)$. Hence, by (5.16) and (5.17), if X possesses Lévy area, it satisfies Property γ-(RIE) relative to $\gamma = \frac{1}{2}$, and, conversely, by (5.16) and (5.18), if X satisfies Property γ-(RIE) relative to $\gamma = \frac{1}{2}$, it possesses Lévy area. $\square$

5.3 Application to stochastic integration

In this section, we apply the deterministic integration theory developed in Section 5.1 to stochastic integration. For this purpose, let X be a d-dimensional continuous semimartingale, defined on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ with a filtration $(\mathcal{F}_t)_{t\in[0,T]}$ satisfying the usual conditions, i.e., completeness and right-continuity.

It is well-known that the semimartingale X can be lifted to a random rough path via Itô integration, see [41]. We have also proven that Property (RIE) ensures that the random rough paths $\mathbf{X} = (X, \mathbb{X})$ with $\mathbb{X}$ defined pathwise via the canonical rough path lift and with $\mathbb{X}$ defined by Itô integration coincide almost surely, see part (i) of Lemma 3.2.1, and that the random rough integral of a semimartingale coincides almost surely with the associated stochastic Itô integral, see Proposition 4.2.4.

With Property γ-(RIE) for $\gamma = \frac{1}{2}$ at hand, we are now able to show this for Stratonovich integration. It is well-known that the semimartingale X can be lifted to a random rough path via Stratonovich integration, by defining $\mathbf{X} = (X, \mathbb{X}) \in \mathcal{C}^p([0,T];\mathbb{R}^d)$, $\mathbb{P}$-a.s., for any $p \in (2,3)$, where

$$\mathbb{X}_{s,t} := \int_s^t (X_r - X_s) \otimes \circ \mathrm{d}X_r = \int_s^t X_r \otimes \circ \mathrm{d}X_r - X_s \otimes X_{s,t}, \qquad (s,t) \in \Delta_T, \tag{5.19}$$

see [41]. It turns out that, if the semimartingale X satisfies Property γ-(RIE) relative to $\gamma = \frac{1}{2}$, $p \in (2,3)$ and a suitable sequence of partitions π, then the canonical random rough path coincides almost surely with the Stratonovich rough path lift and the random rough integral coincides almost surely with the associated stochastic Stratonovich integral.

Lemma 5.3.1. *Let $p \in (2,3)$ and let $\pi^n = \{\tau_k^n\}$, $n \in \mathbb{N}$, be a sequence of adapted partitions (so that each τ_k^n is a stopping time), such that for almost every $\omega \in \Omega$, $(\pi^n(\omega))_{n\in\mathbb{N}}$ is a sequence of (finite) partitions of $[0,T]$ with vanishing mesh size.*

Let X be a d-dimensional continuous semimartingale, and suppose that for almost every $\omega \in \Omega$, $\sup\{|X_{\tau_k^n(\omega),\tau_{k+1}^n(\omega)}(\omega)| : k = 0,\dots,N_n - 1\}$ converges to 0 as $n \to \infty$, and that the sample path $X(\omega)$ satisfies Property γ-(RIE) relative to $\gamma = \frac{1}{2}$, p and $(\pi^n(\omega))_{n\in\mathbb{N}}$.

(i) The random rough paths $\mathbf{X} = (X, \mathbb{X})$, with $\mathbb{X}$ defined pathwise via (5.4) for $\gamma = \frac{1}{2}$, and with $\mathbb{X}$ defined by stochastic integration as in (5.19), coincide $\mathbb{P}$-almost surely.

(ii) Let (Y, Y') be a continuous semimartingale. Suppose that, for almost every $\omega \in \Omega$, $(Y(\omega), Y'(\omega))$ is a controlled path in $\mathscr{C}^p_{X(\omega)}([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$. Then the rough and Stratonovich integrals of Y against X coincide $\mathbb{P}$-almost surely, that is,

$$\int_0^t Y_s(\omega)\,\mathrm{d}\mathbf{X}_s^{\frac{1}{2}}(\omega) = \Big(\int_0^t Y_s \circ \mathrm{d}X_s\Big)(\omega), \qquad t \in [0,T],$$

holds for almost every $\omega \in \Omega$, where $\mathbf{X}^{\frac{1}{2}}(\omega)$ is the canonical rough path lift of $X(\omega)$ as defined in Proposition 5.1.5, using Property γ-(RIE) for $\gamma = \frac{1}{2}$.

Proof. (i): By construction, the pathwise rough integral $\int_0^t X_r(\omega) \otimes \mathrm{d}^{\frac{1}{2},\pi} X_r(\omega)$ constructed via Property γ-(RIE) for $\gamma = \frac{1}{2}$ is given by the limit as $n \to \infty$ of mid-point Riemann sums:

$$\sum_{k=0}^{N_n-1} \frac{1}{2}(X_{\tau_k^n(\omega)}(\omega) + X_{\tau_{k+1}^n(\omega)}(\omega)) \otimes X_{\tau_k^n(\omega)\wedge t,\tau_{k+1}^n(\omega)\wedge t}(\omega).$$

It is known that these Riemann sums also converge uniformly in probability to the Stratonovich integral $\int_0^t X_r \otimes \circ \mathrm{d}X_r$ (see e.g. [147, Chapter II, Theorem 21, Theorem 22]), and the result thus follows from the (almost sure) uniqueness of limits.

(ii): By, e.g., [147, Chapter II, Theorem 21, Theorem 23], we have that

$$\sum_{k=0}^{N_n-1} \frac{1}{2}(Y_{\tau_k^n} + Y_{\tau_{k+1}^n}) X_{\tau_k^n \wedge t, \tau_{k+1}^n \wedge t} \longrightarrow \int_0^t Y_s \circ \mathrm{d}X_s \qquad \text{as} \quad n \to \infty,$$

where the convergence holds uniformly (in $t \in [0,T]$) in probability. By taking a subsequence if necessary, we can then assume that the (uniform) convergence holds almost surely. On the other hand, by Theorem 5.1.8, we know that for almost every $\omega \in \Omega$,

$$\sum_{k=0}^{N_n-1} \frac{1}{2}(Y_{\tau_k^n(\omega)}(\omega) + Y_{\tau_{k+1}^n(\omega)}(\omega)) X_{\tau_k^n(\omega)\wedge t, \tau_{k+1}^n(\omega)\wedge t}(\omega) \longrightarrow \int_0^t Y_s(\omega) \,\mathrm{d}\mathbf{X}_s^{\frac{1}{2}}(\omega) \quad \text{as} \quad n \to \infty$$

uniformly in $t \in [0,T]$. The result thus follows by the uniqueness of limits. $\square$

We now heuristically remark that various semimartingales (and non-semimartingales) satisfy Property γ-(RIE) (particularly for $\gamma = \frac{1}{2}$) relative to suitable sequences of partitions, making the developed theory applicable to a broad class of stochastic processes.

Due to Lemma 5.2.1, Property γ-(RIE) holds relative to any $\gamma \in [0,1]$, if Property (RIE) holds. This in turn holds true for almost all sample paths of the following stochastic processes relative to $p \in (2,3)$ and suitable sequences of partitions; for details see Section 3.2 and Section 3.3:

- Brownian motion, relative to sequences of equidistant partitions $(\pi^n)_{n\in\mathbb{N}}$ such that $|\pi^n|^{2-\frac{4}{p}} \log(n) \to 0$ as $n \to \infty$,
- Itô processes, relative to the sequence of dyadic partitions, i.e., $\pi^n = \{k2^{-n}T\}_{k=0}^{2^n}$, $n \in \mathbb{N}$,
- continuous semimartingales, relative to the sequence of partitions $\pi^n = \{\tau_k^n : k \in \mathbb{N} \cup \{0\}\}$, $n \in \mathbb{N}$, where $\tau_0^n = 0$, $\tau_k^n = \inf\{t > \tau_{k-1}^n : |t - \tau_{k-1}^n| + |X_t - X_{\tau_{k-1}^n}| \geq 2^{-n}\} \wedge T$, $k \in \mathbb{N}$.
- the pair (W, W^H), where W denotes the Brownian motion and W^H the fractional Brownian motion with Hurst parameter $H > \frac{1}{2}$, relative to sequences of equidistant partitions $(\pi^n)_{n\in\mathbb{N}}$ such that $(\pi^n)^{2-\frac{4}{p}} \to 0$ as $n \to \infty$,
- the pair (η, W), where η denotes a deterministic $\frac{1}{p}$-Hölder continuous path, relative to the sequence of dyadic partitions.

Interestingly, since Property γ-(RIE) relative to $\gamma = \frac{1}{2}$ is equivalent to the property of possessing Lévy area, we can extend the class of stochastic processes by a non-semimartingale example, namely the fractional Brownian motion for Hurst parameter $H \in (\frac{1}{3}, \frac{1}{2})$. The sample paths of fractional Brownian motion do not possess quadratic variation, see e.g. [148, 61], consequently they do not satisfy Property (RIE), see also Lemma 5.2.3, and therefore do not fit in our Itô-type setting. For the Stratonovich-type integral, however, we are able to resolve the issue since almost all sample paths of the fractional Brownian motion possess Lévy area relative to the dyadic partitions, which follows from the fact that almost all sample paths of fractional Brownian motion are of finite $\frac{1}{\alpha}$-variation for $\alpha < H$ and e.g. [42, Theorem 2], who construct a rough path lift over the fractional Brownian motion using dyadic approximations.

Chapter 6

Universal approximation with Itô-type signatures

The signature of a path plays a prominent role in the theory of rough paths, initiated by Lyons in [129], which has emerged as an improved framework for dealing with interactions in complex random evolving systems.

It can be formally defined as the enhancement of a path $X\colon [0,T] \to \mathbb{R}^d$ by all iterated integrals of the path against itself

$$\int_{0<t_1<\ldots<t_n<T} \mathrm{d}X_{t_1}^{i_1} \cdots \mathrm{d}X_{t_n}^{i_n},$$

for $i_1, \ldots, i_n \in \{1, \ldots, d\}$, $n \in \mathbb{N}$; see the early works of Chen [28, 29]. This collection of all iterated integrals (given a suitable notion of integration) summarizes the full evolution and interactions of the components of the path effectively: the signature is known to provide an intriguing nonlinear characterization of the path that is unique up to general reparametrizations, see [87] for paths of finite 1-variation. Importantly, due to its rich algebraic structure, linear functionals on the signature approximate continuous path functionals arbitrarily well on compact sets; this is known as the universal approximation theorem.

Thus the signature can be used to faithfully and tractably represent the key features from highly oscillatory streams of data, which is important in the context of machine learning. Recently, a significant strand of research has been concerned with developing data-driven methods based on the signature to exploit its desirable and rich mathematical properties for applications in mathematical finance. These are manifold and include asset pricing [127, 11, 19], optimal execution [100], and calibration of financial models [12, 45, 44], to name but a few.

In this context, the theory of signatures has been adopted to a probabilistic setting using Stratonovich integration. This is the natural choice because the Stratonovich integral satisfies the classical first order calculus, which yields the exact algebraic and geometric properties of the associated signature implying its "universal nonlinearity".

However for financial applications, from a modeling perspective, Itô integration is typically more reasonable because the stochastic process so described is a martingale, and when used as a capital process it guarantees the absence of arbitrage.

Moreover, the signature associated with Itô integration may offer statistical advantages; see [83] for a comparison with Stratonovich integration with respect to statistical consistency of the Lasso estimator using the signature.

This presents a gap between the theory of signatures and the use of Itô integration, we aim to address in this chapter.

While [88] take the signature in this regard as a universal polynomial regression basis, we show that the signature using Itô integration is able to serve as a linear regression basis for continuous functionals.

For this purpose, we make use of rough path theory and assume a path property which ensures that the rough integral exists as a limit of Riemann sums along a suitable sequence of partitions. This is Property γ-(RIE), which has been introduced in Chapter 5 and provides a unifying framework for general pathwise stochastic integration that can be applied to continuous semimartingales.

This chapter is structured as follows. In Section 6.1 we introduce a notion of the signature of the path based on Property γ-(RIE), recovering Stratonovich-type or Itô-type integration, depending on the choice of the parameter γ. When extending the path by suitable quadratic variation terms, we are able to prove a universal approximation theorem for linear functionals on the so-called *γ-signature*, see Section 6.2. This is then translated into the probabilistic setting in Section 6.3, where we derive a universal approximation theorem using Itô-signatures of continuous semimartingales.

Since this approach is motivated by its use for mathematical finance, we intend to promptly explore suitable applications thereof in future work.

6.1 The signature using general pathwise stochastic integration

We will first recall some essentials from the theory of signatures and rough paths, which we divide into the algebraic and analytic concepts. For a more detailed introduction, we refer to [130, 74].

6.1.1 Algebraic setting for signatures

The *tensor algebra* and the *extended tensor algebra* on $\mathbb{R}^d$ are defined by

$$T(\mathbb{R}^d) := \bigoplus_{n=0}^{\infty} (\mathbb{R}^d)^{\otimes n} \qquad \text{and} \qquad T((\mathbb{R}^d)) := \prod_{n=0}^{\infty} (\mathbb{R}^d)^{\otimes n},$$

where $(\mathbb{R}^d)^{\otimes n}$ denotes the n-fold tensor product of $\mathbb{R}^d$, with the convention $(\mathbb{R}^d)^{\otimes 0} := \mathbb{R}$.

We equip $T((\mathbb{R}^d))$ with the standard addition $+$, tensor multiplication $\otimes$ and scalar multiplication, which is defined for $\mathbf{a} = (a^{(n)})_{n=0}^{\infty}, \mathbf{b} = (b^{(n)})_{n=0}^{\infty} \in T((\mathbb{R}^d))$, $\lambda \in \mathbb{R}$, by setting

$$\begin{aligned}
\mathbf{a} + \mathbf{b} &:= (a^{(n)} + b^{(n)})_{n=0}^{\infty}, \\
\mathbf{a} \otimes \mathbf{b} &:= (\sum_{i+j=n} a^{(i)} \otimes b^{(j)})_{n=0}^{\infty}, \\
\lambda \mathbf{a} &:= (\lambda a^{(n)})_{n=0}^{\infty}.
\end{aligned}$$

We observe that $(T((\mathbb{R}^d)), +, \cdot, \otimes)$ is a real non-commutative algebra. The neutral element is $(1, 0, \dots, 0, \dots)$.

Let $(e_1, \dots, e_d)$ be the canonical basis of $\mathbb{R}^d$. The Lie algebra that is generated from $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$, where $\mathbf{e}_i := (0, e_i, 0, \dots) \in T(\mathbb{R}^d)$, and the commutator bracket

$$[\mathbf{a}, \mathbf{b}] = \mathbf{a} \otimes \mathbf{b} - \mathbf{b} \otimes \mathbf{a}, \qquad \mathbf{a}, \mathbf{b} \in T(\mathbb{R}^d),$$

is called the *free Lie algebra* $\mathfrak{g}(\mathbb{R}^d)$ over $\mathbb{R}^d$, see e.g. [74, Section 7.3]. It is a subalgebra of $T_0((\mathbb{R}^d))$, where we define for $c \in \mathbb{R}$, the tensor subalgebra $T_c((\mathbb{R}^d)) := \{\mathbf{a} = (a^{(n)})_{n=0}^{\infty} \in T((\mathbb{R}^d)) : a^{(0)} = c\}$.

The *free Lie group* $G((\mathbb{R}^d)) := \exp(\mathfrak{g}(\mathbb{R}^d))$ is defined as the tensor exponential of $\mathfrak{g}(\mathbb{R}^d)$, i.e., its image under the map

$$\exp_{\otimes} \colon T_0((\mathbb{R}^d)) \to T((\mathbb{R}^d)), \qquad \mathbf{a} \mapsto 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \mathbf{a}^{\otimes k}.$$

$G((\mathbb{R}^d))$ is a subgroup of $T_1((\mathbb{R}^d))$. In fact, $(G((\mathbb{R}^d)), \otimes)$ is a group with unit element $(1, 0, \dots, 0, \dots)$, and for all $\mathbf{g} = \exp_{\otimes}(\mathbf{a}) \in G((\mathbb{R}^d))$, the inverse with respect to $\otimes$ is given by $\mathbf{g}^{-1} = \exp_{\otimes}(-\mathbf{a})$, for $\mathbf{g} = \exp_{\otimes}(\mathbf{a}) \in G((\mathbb{R}^d))$. We call elements in $G((\mathbb{R}^d))$ *group-like* elements. For $N \in \mathbb{N}$, the *truncated tensor algebra* on $\mathbb{R}^d$ is defined by

$$T^N(\mathbb{R}^d) := \bigoplus_{n=0}^{N} (\mathbb{R}^d)^{\otimes n}.$$

For any $\mathbf{a} = (a^{(n)})_{n=0}^{N} \in T^N(\mathbb{R}^d)$, we set

$$|\mathbf{a}|_{T^N(\mathbb{R}^d)} := \max_{n=0,\dots,N} |a^{(n)}|_{(\mathbb{R}^d)^{\otimes n}},$$

where we write $|\cdot|$ for the Euclidean norm, on $\mathbb{R}^d$ or $(\mathbb{R}^d)^{\otimes n}$ for some $n \in \mathbb{N}$. We consider the maps $\Pi_n \colon T((\mathbb{R}^d)) \to (\mathbb{R}^d)^{\otimes n}$ and $\Pi_{\leq N} \colon T((\mathbb{R}^d)) \to T^N(\mathbb{R}^d)$, where $\Pi_n(\mathbf{a}) = a^{(n)}$ and $\Pi_{\leq N}(\mathbf{a}) = (a^{(0)}, \dots, a^{(N)})$, for $\mathbf{a} = (a^{(n)})_{n=0}^{\infty} \in T((\mathbb{R}^d))$. We set for $c \in \mathbb{R}$, $T_c^N(\mathbb{R}^d) :=$

$\{\Pi_{\leq N}(\mathbf{a}) : \mathbf{a} \in T_c((\mathbb{R}^d))\}$. Then $T_1^N(\mathbb{R}^d)$ is a Lie group under the tensor multiplication $\otimes$, truncated beyond level N. We equip $T_1^N(\mathbb{R}^d)$ with the metric

$$\rho(\mathbf{a}, \mathbf{b}) := |\mathbf{a} - \mathbf{b}|_{T^N(\mathbb{R}^d)} = \max_{n=1,\ldots,N} |(a-b)^{(n)}|_{(\mathbb{R}^d)^{\otimes n}},$$

for $\mathbf{a} = (a^{(n)})_{n=0}^N, \mathbf{b} = (b^{(n)})_{n=0}^N \in T_1^N(\mathbb{R}^d)$, which arises from the norm on $T^N(\mathbb{R}^d)$.

The *free nilpotent Lie algebra* and the *free nilpotent Lie group of order N* are defined by $\mathfrak{g}^N(\mathbb{R}^d) := \Pi_{\leq N}(\mathfrak{g}(\mathbb{R}^d))$ and $G^N(\mathbb{R}^d) := \Pi_{\leq N}(G((\mathbb{R}^d)))$, respectively. That is,

$$\mathfrak{g}^N(\mathbb{R}^d) = \{0\} \oplus \mathbb{R}^d \oplus [\mathbb{R}^d, \mathbb{R}^d] \oplus \ldots \oplus \underbrace{[\mathbb{R}^d, [\mathbb{R}^d, \ldots [\mathbb{R}^d, \mathbb{R}^d]]]}_{N-1 \text{ brackets}} \subseteq T_0^N(\mathbb{R}^d).$$

Then $G^N(\mathbb{R}^d)$ is a subgroup of $T_1^N(\mathbb{R}^d)$ with respect to $\otimes$.

Defining the truncated tensor exponential via the corresponding (finite) power series in the truncated tensor algebra, we have that $G^N(\mathbb{R}^d) = \exp_\otimes^N(\mathfrak{g}^N(\mathbb{R}^d))$.

Now, let $I = (i_1, \ldots, i_{|I|})$ be a multi-index (with entries in $\{1, \ldots, d\}$) of length $|I|$. We recall the canonical basis $(e_1, \ldots, e_d)$ of $\mathbb{R}^d$, and set $e_I := e_{i_1} \otimes \ldots \otimes e_{i_{|I|}}$. If $|I|= 1$, set $I' = \emptyset$, if $|I|\geq 1$, $I' = (i_1, \ldots, i_{|I|-1})$. Moreover, we denote by $e_\emptyset$ the basis element of $(\mathbb{R}^d)^{\otimes 0}$ and set $|\emptyset|:= 0$. This allows to write $\mathbf{a} \in T((\mathbb{R}^d))$ as

$$\mathbf{a} = \sum_{|I|\geq 0} a_I e_I,$$

for some $a_I \in \mathbb{R}$. Furthermore, for $\mathbf{a} \in T(\mathbb{R}^d)$ and $\mathbf{b} \in T((\mathbb{R}^d))$, we set

$$\langle \mathbf{a}, \mathbf{b} \rangle := \sum_{|I|\geq 0} \langle \mathbf{a}_I, \mathbf{b}_I \rangle.$$

Then $(e_I)_{\{I:|I|=n\}}$ is the canonical orthonormal basis of $(\mathbb{R}^d)^{\otimes n}$ with respect to this inner product. In particular, $\mathbf{b}_I = \langle e_I, \mathbf{b} \rangle$.

Associating $\ell \in T(\mathbb{R}^d)$ with a linear functional $\langle \ell, \cdot \rangle : T((\mathbb{R}^d)) \to \mathbb{R}$, we write

$$\langle \ell, \mathbf{a} \rangle := \sum_{0\leq |I| \leq N} \ell_I \langle e_I, \mathbf{a} \rangle, \qquad \mathbf{a} \in T((\mathbb{R}^d)),$$

for $\ell = \sum_{0\leq |I|\leq N} \ell_I e_I$, where $\ell_I := \langle e_I, \ell \rangle \in \mathbb{R}$ and $N \in \mathbb{N}_0$.

For two multi-indices $I = (i_1, \ldots, i_{|I|})$, $J = (j_1, \ldots, j_{|J|})$ with entries in $\{1, \ldots, d\}$, the *shuffle product* is recursively defined by

$$e_I \sqcup\!\sqcup e_J := (e_{I'} \sqcup\!\sqcup e_J) \sqcup\!\sqcup e_{i_{|I|}} + (e_I \sqcup\!\sqcup e_{J'}) \sqcup\!\sqcup e_{j_{|J|}},$$

with $e_I \sqcup\!\sqcup e_\emptyset := e_\emptyset \sqcup\!\sqcup e_I := e_I$. For $\mathbf{a}, \mathbf{b} \in T(\mathbb{R}^d)$, we then set

$$a \sqcup\!\sqcup b = \sum_{|I|,|J|\geq 0} a_I b_J (e_I \sqcup\!\sqcup e_J)$$

and for $\mathbf{a}, \mathbf{b} \in T((\mathbb{R}^d))$, we set

$$\langle e_I, \mathbf{a} \sqcup \mathbf{b} \rangle = \langle e_I, \Pi_{|I|}(\mathbf{a})\Pi_{|I|}(\mathbf{b}) \rangle.$$

For all $\mathbf{a} \in G((\mathbb{R}^d))$, the *shuffle product property* holds, i.e., for two multi-indices $I = (i_1, \ldots, i_{|I|})$, $J = (j_1, \ldots, j_{|J|})$, it holds that

$$\langle e_I, \mathbf{a} \rangle \langle e_J, \mathbf{a} \rangle = \langle e_I \sqcup e_J, \mathbf{a} \rangle.$$

6.1.2 Essentials on rough path theory

Throughout, we let $T > 0$ be a fixed finite time horizon. We let $\Delta_T := \{(s,t) \in [0,T]^2 : s \leq t\}$ denote the standard 2-simplex.

We shall write $a \lesssim b$ to mean that there exists a constant $C > 0$ such that $a \leq Cb$. The constant C may depend on the normed space, e.g. through its dimension or regularity parameters.

For a normed space $(E, |\cdot|)$, we let $C([0,T]; E)$ denote the set of continuous paths from $[0,T]$ to E. For $X \in C([0,T]; E)$, the supremum seminorm of the path X is given by

$$\|X\|_\infty := \sup_{t \in [0,T]} |X_t|,$$

and for $p \geq 1$, the p-variation of the path X is given by

$$\|X\|_p := \|X\|_{p,[0,T]} \qquad \text{with} \qquad \|X\|_{p,[s,t]} := \Bigg(\sup_{\mathcal{P} \subset [s,t]} \sum_{[u,v] \in \mathcal{P}} |X_v - X_u|^p \Bigg)^{\frac{1}{p}}, \quad (s,t) \in \Delta_T,$$

where the supremum is taken over all possible partitions $\mathcal{P}$ of the interval $[s,t]$. We recall that, given a path X, we have that $\|X\|_p < \infty$ if and only if there exists a control function c such that [1]

$$\sup_{(u,v) \in \Delta_T} \frac{|X_v - X_u|^p}{c(u,v)} < \infty.$$

We write $C^{p\text{-var}} = C^{p\text{-var}}([0,T]; E)$ for the space of paths $X \in C([0,T]; E)$ which satisfy $\|X\|_p < \infty$. Moreover, for a path $X \in C([0,T]; \mathbb{R}^d)$, we will often use the shorthand notation:

$$X_{s,t} := X_t - X_s, \qquad \text{for} \quad (s,t) \in \Delta_T.$$

For $r \geq 1$ and a two-parameter function $\mathbb{X} \colon \Delta_T \to E$, we similarly define

$$\|\mathbb{X}\|_r := \|\mathbb{X}\|_{r,[0,T]} \qquad \text{with} \qquad \|\mathbb{X}\|_{r,[s,t]} := \Bigg(\sup_{\mathcal{P} \subset [s,t]} \sum_{[u,v] \in \mathcal{P}} |\mathbb{X}_{u,v}|^r \Bigg)^{\frac{1}{r}}, \quad (s,t) \in \Delta_T.$$

We write $C_2^{r\text{-var}} = C_2^{r\text{-var}}(\Delta_T; E)$ for the space of continuous functions $\mathbb{X} \colon \Delta_T \to E$ which satisfy $\|\mathbb{X}\|_r < \infty$.

For $p \in [2,3)$, a pair $\mathbf{X} = (X, \mathbb{X})$ is called a *(continuous) p-rough path* over $\mathbb{R}^d$ if

[1] Here and throughout, we adopt the convention that $\frac{0}{0} := 0$.

(i) $X \in C^{p\text{-var}}([0,T];\mathbb{R}^d)$ and $\mathbb{X} \in C_2^{\frac{p}{2}\text{-var}}(\Delta_T;\mathbb{R}^{d\times d})$, and

(ii) Chen's relation: $\mathbb{X}_{s,t} = \mathbb{X}_{s,u} + \mathbb{X}_{u,t} + X_{s,u} \otimes X_{u,t}$ holds for all $0 \le s \le u \le t \le T$.

In component form, condition (ii) states that $\mathbb{X}^{ij}_{s,t} = \mathbb{X}^{ij}_{s,u} + \mathbb{X}^{ij}_{u,t} + X^i_{s,u}X^j_{u,t}$ for every i and j. We will denote the space of p-rough paths by $\mathcal{C}^p = \mathcal{C}^p([0,T];\mathbb{R}^d)$. On the space $\mathcal{C}^p([0,T];\mathbb{R}^d)$, we use the natural seminorm

$$\|\mathbf{X}\|_p := \|\mathbf{X}\|_{p,[0,T]} \qquad \text{with} \qquad \|\mathbf{X}\|_{p,[s,t]} := \|X\|_{p,[s,t]} + \|\mathbb{X}\|_{\frac{p}{2},[s,t]}$$

for $(s,t) \in \Delta_T$.

Similarly, for $p \ge 1$, and $N \in \mathbb{N}$, the p-variation of $\mathbb{X}^{\le N} \colon [0,T] \to T^{\le N}(\mathbb{R}^d)$ is given by

$$\|\mathbb{X}^{\le N}\|_{p,[s,t]} := \max_{1\le m\le N} \sup_{\mathcal{P}\subset[s,t]} \Bigg(\sum_{[u,v]\in\mathcal{P}} |\Pi_m(\mathbb{X}^{\le N}_{u,v})|^{\frac{p}{m}} \Bigg)^{\frac{m}{p}}, \qquad (s,t) \in \Delta_T,$$

where now $\mathbb{X}^{\le N}_{s,t} := (\mathbb{X}^{\le N}_s)^{-1} \otimes \mathbb{X}^{\le N}_t$, $(s,t) \in \Delta_T$, and we write $\|\mathbb{X}^{\le N}\|_p := \|\mathbb{X}^{\le N}\|_{p,[0,T]}$.

For $\mathbb{X}^{\le N}, \widetilde{\mathbb{X}}^{\le N} \colon [0,T] \to T^{\le N}(\mathbb{R}^d)$, we define the p-variation distance

$$\|\mathbb{X}^{\le N}; \widetilde{\mathbb{X}}^{\le N}\|_{p,[s,t]} := \|\mathbb{X}^{\le N} - \widetilde{\mathbb{X}}^{\le N}\|_{p,[s,t]}, \qquad (s,t) \in \Delta_T,$$

and we write $\|\mathbb{X}^{\le N}; \widetilde{\mathbb{X}}^{\le N}\|_p = \|\mathbb{X}^{\le N}; \widetilde{\mathbb{X}}^{\le N}\|_{p,[0,T]}$.

Here, we equip $G^N(\mathbb{R}^d)$ with the (inhomogeneous) subspace topology of $T^N(\mathbb{R}^d)$. In the literature, the (homogeneous) p-variation of a $G^N(\mathbb{R}^d)$-valued path is often defined in terms of the Carnot–Carathéodory metric, see e.g. [74, Chapter 8]. This is consistent because the induced topology on $G^N(\mathbb{R}^d)$ coincides with the one induced by the Carnot–Carathéodory metric, see e.g. [74, Section 8.1.2 and 8.1.3].

A continuous path $\mathbb{X}^{\le \lfloor p \rfloor} \colon [0,T] \to G^{\lfloor p \rfloor}(\mathbb{R}^d)$ is called a *weakly geometric p-rough path*, if $\mathbb{X}^{\le \lfloor p \rfloor}_0 = \mathbf{1}$ and $\|\mathbf{1}; \mathbb{X}^{\le \lfloor p \rfloor}\|_p < \infty$, where $\mathbf{1} := (1,0,\dots,0) \in T^{\lfloor p \rfloor}(\mathbb{R}^d)$.

We will denote the space of weakly geometric continuous p-rough paths by $C_o^{p\text{-var}} = C_o^{p\text{-var}}([0,T];G^{\lfloor p \rfloor}(\mathbb{R}^d))$ and equip it with the distance $\|\cdot;\cdot\|_p$.

An algebraic condition for a p-rough path to be weakly geometric is that the symmetric part of the rough path lift is determined by the increments of the path.

Lemma 6.1.1. *Let $p \in (2,3)$. Let $(X,\mathbb{X}) \in \mathcal{C}^p([0,T];\mathbb{R}^d)$ be a continuous p-rough path such that $\mathbb{S}(\mathbb{X}_{0,t}) = \frac{1}{2} X_{0,t} \otimes X_{0,t}$, $t \in [0,T]$, where we consider the decomposition into the symmetric and the antisymmetric part given by*

$$\mathbb{X}_{0,t} = \mathbb{S}(\mathbb{X}_{0,t}) + \mathbb{A}(\mathbb{X}_{0,t}) := \frac{1}{2}(\mathbb{X}_{0,t} + \mathbb{X}_{0,t}^\top) + \frac{1}{2}(\mathbb{X}_{0,t} - \mathbb{X}_{0,t}^\top),$$

where $(\cdot)^\top$ denotes matrix transposition. Then $\mathbb{X}^{\le 2}$ is a weakly geometric p-rough path, i.e., $\mathbb{X}^{\le 2} \in C_o^{p\text{-var}}$, where $\mathbb{X}^{\le 2}$ is defined by

$$\mathbb{X}^{\le 2}_t := (1, X_{0,t}, \mathbb{X}_{0,t}), \qquad t \in [0,T].$$

Proof. Recall that $G^2(\mathbb{R}^d) = \exp_\otimes^2(\mathfrak{g}^2(\mathbb{R}^d))$, where $\mathfrak{g}^2(\mathbb{R}^d) = \{0\} \oplus \mathbb{R}^d \oplus [\mathbb{R}^d, \mathbb{R}^d]$. It holds that $[\mathbb{R}^d, \mathbb{R}^d] = \operatorname{span}\{e_i \otimes e_j - e_j \otimes e_i : 1 \le i, j \le d\}$. Therefore $[\mathbb{R}^d, \mathbb{R}^d]$ equals the set of antisymmetric $d \times d$-matrices and it follows that, for any $t \in [0, T]$,

$$\mathbb{X}_t^{\le 2} = (1, X_{0,t}, \frac{1}{2} X_{0,t} \otimes X_{0,t} + \mathbb{A}(\mathbb{X}_{0,t})) = \exp_\otimes^2(0, X_{0,t}, \mathbb{A}(\mathbb{X}_{0,t})) \in G^2(\mathbb{R}^d).$$

Finally, since $(X, \mathbb{X}) \in \mathcal{C}^p([0,T]; \mathbb{R}^d)$, it particularly holds that $\|\mathbf{1}; \mathbb{X}^{\le 2}\|_p < \infty$. $\square$

Remark 6.1.2. *This condition is a consequence of "first order calculus" and therefore valid in the context of stochastic Stratonovich integration.*

6.1.3 On Property γ-(RIE)

We develop a notion of signatures using the path assumption Property γ-(RIE), which allows to construct pathwise (iterated) integrals as limits of general Riemann sums. It is an extension of Property (RIE), which we have established in detail in Chapter 5. We now give the path properties and the statements required in this chapter. For the proofs and an equivalent and more intuitive characterization of the path property, we refer to Chapter 5.

Property γ-(RIE). *Let $X \in C([0,T]; \mathbb{R}^d)$ and $\pi = (\pi^n)_{n \in \mathbb{N}}$, with $\pi^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$, be a sequence of partitions of the interval $[0,T]$ such that $\sup\{|X_{t_k^n, t_{k+1}^n}| : k = 0, \dots, N_n - 1\}$ converges to 0 as $n \to \infty$, and let $\gamma \in [0,1]$, $p \in (2,3)$.*

We assume that the Riemann sums $\int_0^t X_s \otimes \mathrm{d}^{\gamma, \pi^n} X_s := \sum_{k=0}^{N_n - 1} (X_{t_k^n} + \gamma X_{t_k^n, t_{k+1}^n}) \otimes X_{t_k^n \wedge t, t_{k+1}^n \wedge t}$ converge uniformly as $n \to \infty$ to a limit, which we denote by $\int_0^t X_s \otimes \mathrm{d}^\pi X_s$, $t \in [0,T]$, and that there exists a control function c such that

$$\sup_{(s,t) \in \Delta_T} \frac{|X_{s,t}|^p}{c(s,t)} + \sup_{n \in \mathbb{N}} \sup_{0 \le k < \ell \le N_n} \frac{|(\int_0^\cdot X_s \otimes \mathrm{d}^{\gamma, \pi^n} X_s)_{t_k^n, t_\ell^n} - X_{t_k^n} \otimes X_{t_k^n, t_\ell^n}|^{\frac{p}{2}}}{c(t_k^n, t_\ell^n)} \lesssim 1.$$

We say that a path $X \in C([0,T]; \mathbb{R}^d)$ satisfies Property γ-(RIE) relative to γ, p and π if γ, p, π and X together satisfy Property γ-(RIE).

Proposition 6.1.3 (Proposition 5.1.5)**.** *Suppose that $X \in C([0,T]; \mathbb{R}^d)$ satisfies Property γ-(RIE) relative to some $\gamma \in [0,1]$, $p \in (2,3)$ and a sequence of partitions $\pi = (\pi^n)_{n \in \mathbb{N}}$. Then X extends canonically to a continuous p-rough path $\mathbf{X}^\gamma := (X, \mathbb{X}^\gamma)$, where*

$$\mathbb{X}_{s,t}^\gamma := \int_0^t X_r \otimes \mathrm{d}^{\gamma, \pi} X_r - \int_0^s X_r \otimes \mathrm{d}^{\gamma, \pi} X_r - X_s \otimes X_{s,t}, \qquad (s,t) \in \Delta_T.$$

We note that $\mathbf{X}^0$ corresponds to the Itô-rough path lift and $\mathbf{X}^{\frac{1}{2}}$ corresponds to the Stratonovich-rough path lift of a stochastic process, since the "iterated integral" $\mathbb{X}^0$ and $\mathbb{X}^{\frac{1}{2}}$ is given as a limit of left-point and mid-point Riemann sums, analogously to the stochastic Itô and Stratonovich integral, respectively, see also Section 6.3.

Lemma 6.1.4 (Lemma 5.1.6)**.** *Suppose that $X \in C([0,T];\mathbb{R}^d)$ satisfies Property γ-(RIE) relative to some $\gamma \in [0,1]$, $p \in (2,3)$ and a sequence of partitions $\pi = (\pi^n)_{n\in\mathbb{N}}$. Let $1 \le i,j \le d$, and define for $\gamma = \frac{1}{2}$, $[X^i,X^j]^{\gamma,\pi} := 0$, and for $\gamma \neq \frac{1}{2}$,*

$$[X^i,X^j]^{\gamma,\pi}_t := X^i_t X^j_t - X^i_0 X^j_0 - \int_0^t X^i_s \,\mathrm{d}^{\gamma,\pi} X^j_s - \int_0^t X^j_s \,\mathrm{d}^{\gamma,\pi} X^i_s, \qquad t \in [0,T].$$

Then $[X^i,X^j]^{\gamma,\pi}$ is a continuous function and

$$[X^i,X^j]^{\gamma,\pi}_t = \lim_{n\to\infty} [X^i,X^j]^{\gamma,\pi^n}_t := \lim_{n\to\infty} (1-2\gamma) \sum_{k=0}^{N_n-1} X^i_{t^n_k\wedge t, t^n_{k+1}\wedge t} X^j_{t^n_k\wedge t, t^n_{k+1}\wedge t}.$$

The sequence $([X^i,X^j]^{\gamma,\pi^n})_{n\in\mathbb{N}}$ has uniformly bounded 1-variation, and in particular, $[X^i,X^j]^{\gamma,\pi}$ has finite 1-variation. We write $[X]^{\gamma,\pi} = [X,X]^{\gamma,\pi} = ([X^i,X^j]^{\gamma,\pi})_{1\le i,j\le d}$.

Lemma 6.1.5. *Suppose that $X \in C([0,T];\mathbb{R}^d)$ satisfies Property γ-(RIE) relative to some $\gamma \in [0,1]$, $p \in (2,3)$ and a sequence of partitions $\pi = (\pi^n)_{n\in\mathbb{N}}$. Define $\bar{X}^n$ as the piecewise linear interpolation of X along $\pi = (\pi^n)_{n\in\mathbb{N}}$. Then*

$$\lim_{n\to\infty} \int_s^t \bar{X}^n_r \otimes \mathrm{d}\bar{X}^n_r = \lim_{n\to\infty} \sum_{k=0}^{N_n-1} (X_{t^n_k} + \frac{1}{2} X_{t^n_k,t^n_{k+1}}) \otimes X_{t^n_k\wedge t, t^n_{k+1}\wedge t} = \int_s^t X_r \otimes \mathrm{d}^{\gamma,\pi} X_r + \frac{1}{2}[X]^{\gamma,\pi}_{s,t},$$

where the convergence is uniform in $(s,t) \in \Delta_T$. Moreover, the sequence $(\bar{\mathbb{X}}^n)_{n\in\mathbb{N}}$ has uniformly bounded $\frac{p}{2}$-variation, where $\bar{\mathbb{X}}^n_{s,t} := \int_s^t \bar{X}^n_{s,r} \otimes \mathrm{d}\bar{X}^n_r$, $(s,t) \in \Delta_T$.

We will actually continue working under Property γ-(RIE), as it is more general, but we briefly want to point out the theoretical relation to Property (RIE), which has been introduced in [143] and [5], and utilized in Chapter 3 and Chapter 4.

Property (RIE). *Let $X \in C([0,T];\mathbb{R}^d)$ and $\pi = (\pi^n)_{n\in\mathbb{N}}$, with $\pi^n = \{0 = t^n_0 < t^n_1 < \cdots < t^n_{N_n} = T\}$, $n \in \mathbb{N}$, be a sequence of partitions of the interval $[0,T]$ such that $\sup\{|X_{t^n_k,t^n_{k+1}}| : k = 0,\ldots,N_n-1\}$ converges to 0, and let $p \in (2,3)$.*

We assume that the left-point Riemann sums $\int_0^t X_s \otimes \mathrm{d}^{\pi^n} X_s := \sum_{k=0}^{N_n-1} X_{t^n_k} \otimes X_{t^n_k\wedge t, t^n_{k+1}\wedge t}$ converge uniformly as $n \to \infty$ to a limit, which we denote by $\int_0^t X_s \otimes \mathrm{d}^{\pi} X_s$, $t \in [0,T]$, and that there exists a control function c such that

$$\sup_{(s,t)\in\Delta_T} \frac{|X_{s,t}|^p}{c(s,t)} + \sup_{n\in\mathbb{N}} \sup_{0\le k<\ell\le N_n} \frac{|(\int_0^\cdot X_s \otimes \mathrm{d}^{\pi^n} X_s)_{t^n_k,t^n_\ell} - X_{t^n_k} \otimes X_{t^n_k,t^n_\ell}|^{\frac{p}{2}}}{c(t^n_k,t^n_\ell)} \lesssim 1.$$

We say that a path $X \in C([0,T];\mathbb{R}^d)$ satisfies Property (RIE) relative to p and π if p, π and X together satisfy Property (RIE).

Lemma 6.1.6 (Lemma 5.2.1)**.** *Let $X \in C([0,T];\mathbb{R}^d)$, $\gamma \in [0,1]$, $p \in (2,3)$ and $\pi = (\pi^n)_{n\in\mathbb{N}}$ be a sequence of partitions of $[0,T]$.*

(i) Suppose $\gamma \neq \frac{1}{2}$. X satisfies Property (RIE) *if and only if X satisfies Property γ-*(RIE)*, both relative to p and π.*

(ii) Suppose $\gamma = \frac{1}{2}$. If X satisfies Property (RIE)*, then X satisfies Property γ-*(RIE)*, both relative to p and π.*

Analogously to Property (RIE), see Proposition 3.1.10, Property γ-(RIE) is stable under perturbation by a path of finite q-variation for $q \in (1,2)$, which then falls into the regime of Young integration. The proof of the following lemma can be found in Appendix A.5.

Lemma 6.1.7. *Let $X \in C([0,T];\mathbb{R}^d)$, $\gamma \in [0,1]$, $p \in (2,3)$ and $\pi = (\pi^n)_{n\in\mathbb{N}}$, with $\pi^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$, be a sequence of partitions. Suppose that X satisfies Property γ-*(RIE) *relative to γ, p and π. Let $\varphi \in C^{q\text{-var}}([0,T];\mathbb{R}^d)$ for some $q \in [1,2)$ such that $\frac{1}{p} + \frac{1}{q} > 1$ and $\sup\{|\varphi_{t_k^n,t_{k+1}^n}| \colon k = 0,\dots,N_n - 1\}$ converges to 0 as $n \to \infty$. Then the path $\widehat{X} = X + \varphi$ satisfies Property γ-*(RIE) *relative to γ, p and π.*

6.1.4 The signature using Property γ-(RIE)

By Lyons' extension theorem, see e.g. [74, Theorem 9.5], for $p \in (2,3)$, any weakly geometric p-rough path admits a unique extension to a path of finite p-variation with values in $G^N(\mathbb{R}^d)$ with $N > 2$, called Lyons' extension, which allows us to define the signature of X as follows:

Definition 6.1.8. *Let $p \in (2,3)$ and $\mathbb{X}^{o,\leq 2} \in C_o^{p\text{-var}}([0,T];G^2(\mathbb{R}^d))$. The* signature *of X is defined as the unique path*

$$\mathbb{X}^{o,\infty} \colon [0,T] \to G((\mathbb{R}^d)),$$

such that for all $N \geq 3$, $\Pi_{\leq N}(\mathbb{X}^{o,\infty}) = \mathbb{X}^{o,\leq N}$, where $\mathbb{X}^{o,\leq N}$ denotes the extension of $\mathbb{X}^{o,\leq 2}$ in $G^N(\mathbb{R}^d)$. In particular, $\mathbb{X}^{o,\infty}$ is the unique path extension of $\mathbb{X}^{o,\leq 2}$ specified by Lyons' extension theorem.

We now show that the canonical rough path under Property γ-(RIE) can be corrected to a weakly geometric rough path by adding the pathwise quadratic variation term, which seems natural when comparing stochastic Itô and Stratonovich integration.

Lemma 6.1.9. *Suppose that $X \in C([0,T];\mathbb{R}^d)$ satisfies Property γ-*(RIE) *relative to some $\gamma \in [0,1]$, $p \in (2,3)$ and a sequence of partitions $\pi = (\pi^n)_{n\in\mathbb{N}}$. Let $(X,\mathbb{X}^{\gamma,o}) \in \mathcal{C}^p([0,T];\mathbb{R}^d)$ be a continuous p-rough path, with $\mathbb{X}^{\gamma,o} \colon \Delta_T \to \mathbb{R}^d$ given by*

$$\mathbb{X}^{\gamma,o}_{s,t} := \mathbb{X}^{\gamma}_{s,t} + \frac{1}{2}[X]^{\gamma,\pi}_{s,t}, \qquad (s,t) \in \Delta_T,$$

where $\mathbb{X}^\gamma$ is the canonical rough path lift defined in Proposition 6.1.3 and $[X]^{\gamma,\pi}$ is defined in Lemma 6.1.4. Then $\mathbb{X}^{\gamma,o,\leq 2} \colon [0,T] \to G^2(\mathbb{R}^d)$ is a weakly geometric p-rough path, where we define

$$\mathbb{X}^{\gamma,o,\leq 2}_t := (1, X_{0,t}, \mathbb{X}^{\gamma,o}_{0,t}), \qquad t \in [0,T].$$

Proof. Since $\mathbb{X}^\gamma$ has finite $\frac{p}{2}$-variation and $[X^i, X^j]^\gamma$ has finite 1-variation, $\mathbb{X}^{\gamma,o}$ has finite $\frac{p}{2}$-variation, see Proposition 6.1.3 and Lemma 6.1.4, and particularly, $\|\mathbf{1}; \mathbb{X}^{\gamma,o,\leq 2}\|_p < \infty$.

We show that $\mathbb{S}(\mathbb{X}^{\gamma,o}_{0,t}) = \frac{1}{2} X_{0,t} \otimes X_{0,t}$, for any $t \in [0,T]$. Then applying Lemma 6.1.1, the proof is complete.

By definition, it holds that, for any $1 \leq i, j \leq d$ and any $t \in [0,T]$,

$$
\begin{aligned}
&(\mathbb{X}^{\gamma,o}_{0,t})^{ij} + (\mathbb{X}^{\gamma,o}_{0,t})^{ji} \\
&= \int_0^t X^i_r \,\mathrm{d}^{\gamma,\pi} X^j_r - X^i_0 X^j_{0,t} + \frac{1}{2}[X^i, X^j]^{\gamma,\pi}_t + \int_0^t X^j_r \,\mathrm{d}^{\gamma,\pi} X^i_r - X^j_0 X^i_{0,t} + \frac{1}{2}[X^j, X^i]^{\gamma,\pi}_t \\
&= \lim_{n\to\infty} \sum_{k=0}^{N_n - 1} (X^i_{t^n_k} + \gamma X^i_{t^n_k, t^n_{k+1}}) X^j_{t^n_k \wedge t, t^n_{k+1} \wedge t} + (\frac{1}{2} - \gamma) X^i_{t^n_k, t^n_{k+1}} X^j_{t^n_k \wedge t, t^n_{k+1} \wedge t} - X^i_0 X^j_{0,t} \\
&\quad + \lim_{n\to\infty} \sum_{k=0}^{N_n - 1} (X^j_{t^n_k} + \gamma X^j_{t^n_k, t^n_{k+1}}) X^i_{t^n_k \wedge t, t^n_{k+1} \wedge t} + (\frac{1}{2} - \gamma) X^j_{t^n_k, t^n_{k+1}} X^i_{t^n_k \wedge t, t^n_{k+1} \wedge t} - X^j_0 X^i_{0,t} \\
&= \lim_{n\to\infty} \sum_{k=0}^{N_n - 1} \frac{1}{2} (X^i_{t^n_k} + X^i_{t^n_{k+1}}) X^j_{t^n_k, t^n_{k+1}} - X^i_0 X^j_{0,t} \\
&\quad + \lim_{n\to\infty} \sum_{k=0}^{N_n - 1} \frac{1}{2} (X^j_{t^n_k} + X^j_{t^n_{k+1}}) X^i_{t^n_k, t^n_{k+1}} - X^j_0 X^i_{0,t} \\
&= X^i_t X^j_t - X^i_0 X^j_0 - X^i_0 X^j_{0,t} - X^j_0 X^i_{0,t} \\
&= X^i_{0,t} X^j_{0,t}.
\end{aligned}
$$

□

Remark 6.1.10. *Suppose that* $\gamma = \frac{1}{2}$*. Since* $[X]^{\frac{1}{2},\pi} = 0$*, see Lemma 6.1.4, it holds that* $\mathbb{X}^{\frac{1}{2},o}_t = \mathbb{X}^{\frac{1}{2}}_{0,t}$*, which implies that* $(1, X_{0,\cdot}, \mathbb{X}^{\frac{1}{2}}_{0,\cdot}) \in C^{p\text{-var}}_o([0,T]; G^2(\mathbb{R}^d))$*. That is, the Stratonovich-type rough path is indeed a weakly geometric rough path, which is very reasonable.*

Remark 6.1.11. *If X satisfies Property γ-*(RIE)*, then the signature $\mathbb{X}^{o,\infty}$ of X defined in Definition 6.1.8 is the unique path extension of $\mathbb{X}^{\gamma,o,\leq 2}$ as defined in Lemma 6.1.9.*

A more direct approach is to define the signature as the collection of all iterated integrals over a fixed interval associated to a sufficiently regular path. Here, we utilize Property γ-(RIE) and the corresponding iterated integral, which allows for a unifying framework for Itô-type and Stratonovich-type signatures.

Definition 6.1.12. *Suppose that $X \in C([0,T]; \mathbb{R}^d)$ satisfies Property γ-*(RIE) *relative to some $\gamma \in [0,1]$, $p \in (2,3)$ and a sequence of partitions $\pi = (\pi^n)_{n\in\mathbb{N}}$.*

We recursively set

$$\langle e_\emptyset, \mathbb{X}_t^{\gamma,\infty}\rangle := 1, \quad \langle e_I, \mathbb{X}_t^{\gamma,\infty}\rangle := X_{0,t}^{i_1}, \quad I = (i_1),$$
$$\langle e_I, \mathbb{X}_t^{\gamma,\infty}\rangle := \int_0^t X_s^{i_1}\,\mathrm{d}^{\gamma,\pi} X_s^{i_2} - X_0^{i_1} X_{0,t}^{i_2} = (\mathbb{X}_{0,t}^{\gamma})^{i_1 i_2}, \quad I = (i_1, i_2),$$
$$\langle e_I, \mathbb{X}_t^{\gamma,\infty}\rangle := \int_0^t \langle e_{I'}, \mathbb{X}_s^{\gamma,\infty}\rangle\,\mathrm{d}^{\gamma,\pi} X_s^{i_{|I|}}, \qquad I = (i_1, \dots, i_{|I|}), \quad |I|> 2,$$

for $t \in [0,T]$, where

$$\int_0^t \langle e_{I'}, \mathbb{X}_s^{\gamma,\infty}\rangle\,\mathrm{d}^{\gamma,\pi} X_s^{i_{|I|}} := \lim_{n\to\infty} \sum_{k=0}^{N_n-1} \langle e_{I'}, \mathbb{X}^{\gamma,\infty}_{t_k^n+\gamma(t_{k+1}^n - t_k^n)}\rangle X^{i_{|I|}}_{t_k^n\wedge t, t_{k+1}^n\wedge t}$$

exists as a Young integral. Then $\mathbb{X}^{\gamma,\infty}\colon [0,T] \to T((\mathbb{R}^d))$ is well-defined and is called the γ-signature *of X. Its projection $\mathbb{X}^{\gamma,\le N}$ on $T^N(\mathbb{R}^d)$ is given by*

$$\mathbb{X}_t^{\gamma,\le N} = \Pi_{\le N}(\mathbb{X}_t^{\gamma,\infty}) = \sum_{|I|\le N} \langle e_I, \mathbb{X}_t^{\gamma,\infty}\rangle e_I,$$

and called γ-signature of X truncated at level N*, which takes values in $T^N(\mathbb{R}^d)$ for all $t \in [0,T]$. The increments of the γ-signature $\mathbb{X}^{\gamma,\infty}$ are defined by*

$$\mathbb{X}_{s,t}^{\gamma,\infty} := (\mathbb{X}_s^{\gamma,\infty})^{-1} \otimes \mathbb{X}_t^{\gamma,\infty}, \qquad (s,t) \in \Delta_T.$$

Remark 6.1.13. *By Property γ-(RIE), $\langle e_{(i_1,i_2)}, \mathbb{X}^{\gamma,\infty}\rangle$ has finite $\frac{p}{2}$-variation, for any multi-index $I = (i_1, i_2, i_3)$, that is, $\langle e_I, \mathbb{X}_t^{\gamma,\infty}\rangle$ is a well-defined Young integral for $t \in [0,T]$ since $\frac{2}{p} + \frac{1}{p} > 1$, and has itself finite $\frac{p}{2}$-variation. Thus it holds that $\langle e_I, \mathbb{X}_t^{\gamma,\infty}\rangle$ is a well-defined Young integral, for any multi-index of length $|I|> 4$.*

We note that for any multi-index of length $|I|> 2$, by definition of a Young integral,

$$\int_0^t \langle e_{I'}, \mathbb{X}_s^{\gamma,\infty}\rangle\,\mathrm{d}^{\gamma,\pi} X_s^{i_{|I|}} = \lim_{n\to\infty} \sum_{k=0}^{N_n-1} \langle e_{I'}, \mathbb{X}^{\gamma,\infty}_{t_k^n+\tilde\gamma(t_{k+1}^n - t_k^n)}\rangle X^{i_{|I|}}_{t_k^n\wedge t, t_{k+1}^n\wedge t}, \qquad t \in [0,T],$$

for any $\tilde\gamma \in [0,1]$.

Remark 6.1.14. *Suppose that $\gamma = \frac{1}{2}$. Then*

$$\Pi_{\le 2}(\mathbb{X}_t^{o,\infty}) = \mathbb{X}_t^{\gamma,o,\le 2} = (1, X_{0,t}, \mathbb{X}_{0,t}^{\gamma,o}) = (1, X_{0,t}, \mathbb{X}_{0,t}^{\gamma}) = \Pi_{\le 2}(\mathbb{X}_t^{\gamma,\infty}),$$

that is, the signature of X truncated at level 2 and the γ-signature of X truncated at level 2 for $\gamma = \frac{1}{2}$ coincide.

It turns out that the signature defined via Lyons' extension theorem and the γ-signature defined via iterated integrals under Property γ-(RIE) for $\gamma = \frac{1}{2}$ coincide. This is confirming in the sense that weakly geometric rough paths do align with Stratonovich-type integration.

Proposition 6.1.15. *Suppose that $X \in C([0,T];\mathbb{R}^d)$ satisfies Property γ-(RIE) relative to $\gamma = \frac{1}{2}$, some $p \in (2,3)$ and a sequence of partitions $\pi = (\pi^n)_{n\in\mathbb{N}}$. Then the $\frac{1}{2}$-signature coincides with the Lyons lift up to level $N \in \mathbb{N}$, i.e., $\mathbb{X}^{o,\leq N} = \mathbb{X}^{\frac{1}{2},\leq N}$.*

Proof. Let $n \in \mathbb{N}$. We denote by

$$\bar{X}^n_t := X_{t^n_k} + \frac{t - t^n_k}{t^n_{k+1} - t^n_k} X_{t^n_k, t^n_{k+1}}, \qquad t \in [t^n_k, t^n_{k+1}], \qquad k = 0, \dots N_n - 1,$$

the piecewise linear interpolation of X along $\pi = (\pi^n)_{n\in\mathbb{N}}$. We define its signature $\bar{\mathbb{X}}^{n,\infty}$ as the tensor series of iterated (Riemann–Stieltjes) integrals. For $N \in \mathbb{N}$ then the signature of $\bar{X}^n$ truncated at level N is the path $t \mapsto \bar{\mathbb{X}}^{n,\leq N}_t$ defined by computing all iterated integrals up to order N. Therefore $\bar{\mathbb{X}}^{n,\leq N}$ is the canonical lift of $\bar{X}^n$ to a path with values in $G^N(\mathbb{R}^d)$ (due to the integration by parts rule) and has finite 1-variation, that is, $\bar{\mathbb{X}}^{n,\leq N} \in C^{1\text{-var}}_o([0,T];G^N(\mathbb{R}^d))$. Moreover, observe that by Property γ-(RIE) relative to $\gamma = \frac{1}{2}$ we have

$$\sup_{n\in\mathbb{N}} \|\mathbf{1}; \bar{\mathbb{X}}^{n,\leq 2}\|_p < \infty, \qquad \lim_{n\to\infty} \|\bar{\mathbb{X}}^{n,\leq 2} - \mathbb{X}^{\frac{1}{2},\leq 2}\|_\infty = 0,$$

where $\mathbb{X}^{\frac{1}{2},\leq 2} = (1, X_{0,\cdot}, \mathbb{X}^{\frac{1}{2}}_{0,\cdot}) \in C^{p\text{-var}}_o([0,T];G^2(\mathbb{R}^d))$, see Lemma 6.1.5.

We aim to show that for all $N \geq 2$, $\bar{\mathbb{X}}^{n,\leq N}$ converges uniformly to $\mathbb{X}^{\frac{1}{2},\leq N}$, the γ-signature of X truncated at level N for $\gamma = \frac{1}{2}$.

Since $\mathbb{X}^{\frac{1}{2},o,\leq 2} = \mathbb{X}^{\frac{1}{2},\leq 2} \in C^{p\text{-var}}_o([0,T];G^2(\mathbb{R}^d))$, it then follows, as in the proof of [74, Theorem 9.5], by the uniqueness of Lyons' lift that $\mathbb{X}^{o,\leq N} = \mathbb{X}^{\frac{1}{2},\leq N}$, where $\mathbb{X}^{o,\leq N}$ denotes Lyons' lift of $\mathbb{X}^{\frac{1}{2},o,\leq 2}$ in $G^N(\mathbb{R}^d)$.

We proceed with an inductive argument. Let $N = 3$. For any multi-index I of length N, it holds by definition that $\|\langle e_{I'}, \mathbb{X}^{\frac{1}{2},\infty}\rangle\|_{\frac{p}{2}} < \infty$, see Remark 6.1.13, so that the following integrals can be taken as Young integrals:

$$\langle e_I, \bar{\mathbb{X}}^{n,\infty}_t\rangle = \int_0^t \langle e_{I'}, \bar{\mathbb{X}}^{n,\infty}_s\rangle \,\mathrm{d}(\bar{X}^n_s)^{i_{|I|}}, \qquad \langle e_I, \mathbb{X}^{\frac{1}{2},\infty}_t\rangle = \int_0^t \langle e_{I'}, \mathbb{X}^{\frac{1}{2},\infty}_s\rangle \,\mathrm{d}X^{i_{|I|}}_s, \qquad t \in [0,T].$$

The sequence $(\bar{X}^n)_{n\in\mathbb{N}}$ has uniformly bounded p-variation and $\bar{X}^n$ converges uniformly to X as $n \to \infty$. Similarly, by assumption, the sequence $(\langle e_{I'}, \bar{\mathbb{X}}^{n,\infty}\rangle)_{n\in\mathbb{N}}$ has uniformly bounded $\frac{p}{2}$-variation and $\langle e_{I'}, \bar{\mathbb{X}}^{n,\infty}\rangle$ converges uniformly to $\langle e_{I'}, \mathbb{X}^{\frac{1}{2},\infty}\rangle$ as $n \to \infty$.

By [74, Proposition 6.12] then,

$$\lim_{n\to\infty} \|(\bar{\mathbb{X}}^{n,\infty})^{(N)} - (\mathbb{X}^{\frac{1}{2},\infty})^{(N)}\|_\infty = 0,$$

where $(\bar{\mathbb{X}}^{n,\infty})^{(N)}$ denotes the Nth level of the signature of $\bar{X}^n$ and $(\mathbb{X}^{\frac{1}{2},\infty})^{(N)}$ denotes the Nth level of the $\frac{1}{2}$-signature of X.

Assume that the claim holds true for any order $< N$, for $N > 3$. Then it suffices to show that $(\bar{\mathbb{X}}^{n,\infty})^{(N)}$ converges uniformly to $(\mathbb{X}^{\frac{1}{2},\infty})^{(N)}$ as $n \to \infty$.

For any multi-index I of length N, it holds by definition that $\|\langle e_{I'}, \mathbb{X}^{\frac{1}{2},\infty}\rangle\|_q < \infty$, for $q > \frac{p}{3}$, see Remark 6.1.13, so that the following can be taken as Young integrals:

$$\langle e_I, \bar{\mathbb{X}}^{n,\infty}_t\rangle = \int_0^t \langle e_{I'}, \bar{\mathbb{X}}^{n,\infty}_s\rangle \,\mathrm{d}(\bar{X}^n_s)^{i_{|I|}}, \qquad \langle e_I, \mathbb{X}^{\frac{1}{2},\infty}_t\rangle = \int_0^t \langle e_{I'}, \mathbb{X}^{\frac{1}{2},\infty}_s\rangle \,\mathrm{d}X^{i_{|I|}}_s, \qquad t \in [0,T].$$

Let $p' \in (p, 3)$ and $q' \in (q, 2)$. By the standard estimate for Young integrals—see e.g. [74, Theorem 6.8]—we have that for any $t \in [0,T]$,

$$\begin{aligned}
&|\langle e_I, \bar{\mathbb{X}}^{n,\infty}_t\rangle - \langle e_I, \mathbb{X}^{\frac{1}{2},\infty}_t\rangle| \\
&\quad = \left| \int_0^t \langle e_{I'}, \bar{\mathbb{X}}^{n,\infty}_s\rangle \,\mathrm{d}(\bar{X}^n_s)^{i_{|I|}} - \int_0^t \langle e_{I'}, \mathbb{X}^{\frac{1}{2},\infty}_s\rangle \,\mathrm{d}X^{i_{|I|}}_s \right| \\
&\quad \leq \left| \int_0^t (\langle e_{I'}, \bar{\mathbb{X}}^{n,\infty}_s\rangle - \langle e_{I'}, \mathbb{X}^{\frac{1}{2},\infty}_s\rangle) \,\mathrm{d}(\bar{X}^n_s)^{i_{|I|}} \right| \\
&\qquad + \left| \int_0^t \langle e_{I'}, \mathbb{X}^{\frac{1}{2},\infty}_s\rangle \,\mathrm{d}(\bar{X}^n_s)^{i_{|I|}} - \int_0^t \langle e_{I'}, \mathbb{X}^{\frac{1}{2},\infty}_s\rangle \,\mathrm{d}X^{i_{|I|}}_s \right| \\
&\quad \leq C_{p,p',q,q'}(\|\langle e_{I'}, \bar{\mathbb{X}}^{n,\infty}\rangle - \langle e_{I'}, \mathbb{X}^{\frac{1}{2},\infty}\rangle\|_{q'} \|\bar{X}^n\|_p + \|\langle e_{I'}, \mathbb{X}^{\frac{1}{2},\infty}\rangle\|_q \|\bar{X}^n - X\|_{p'}),
\end{aligned}$$

for some constant $C_{p,p',q,q'} > 0$ depending only on p, p', q and q'. It follows by interpolation—see e.g. [74, Proposition 5.5]—that

$$\|\langle e_{I'}, \bar{\mathbb{X}}^{n,\infty}\rangle - \langle e_{I'}, \mathbb{X}^{\frac{1}{2},\infty}\rangle\|_{q'} \leq \|\langle e_{I'}, \bar{\mathbb{X}}^{n,\infty}\rangle - \langle e_{I'}, \mathbb{X}^{\frac{1}{2},\infty}\rangle\|_\infty^{1-\frac{q'}{q}} \|\langle e_{I'}, \bar{\mathbb{X}}^{n,\infty}\rangle - \langle e_{I'}, \mathbb{X}^{\frac{1}{2},\infty}\rangle\|_q^{\frac{q'}{q}}.$$

The sequence $(\langle e_{I'}, \bar{\mathbb{X}}^{n,\infty}\rangle)_{n\in\mathbb{N}}$ has uniformly bounded $\frac{p}{N-1}$-variation, see e.g. [74, Proposition 9.3], (thus q-variation) and by assumption, $\langle e_{I'}, \bar{\mathbb{X}}^{n,\infty}\rangle \to \langle e_{I'}, \mathbb{X}^{\frac{1}{2},\infty}\rangle$ uniformly on $[0,T]$ as $n \to \infty$. Similarly,

$$\|\bar{X}^n - X\|_{p'} \leq \|\bar{X}^n - X\|_\infty^{1-\frac{p}{p'}} \|\bar{X}^n - X\|_p^{\frac{p}{p'}}.$$

The sequence $(\bar{X}^n)_{n\in\mathbb{N}}$ has uniformly bounded p-variation and $\bar{X}^n$ converges uniformly to X as $n \to \infty$.

Combining these estimates implies that

$$\lim_{n\to\infty} \|(\bar{\mathbb{X}}^{n,\infty})^{(N)} - (\mathbb{X}^{\frac{1}{2},\infty})^{(N)}\|_\infty = 0.$$

Altogether, we obtain that

$$\lim_{n\to\infty} \|\bar{\mathbb{X}}^{n,\leq N} - \mathbb{X}^{\frac{1}{2},\leq N}\|_\infty = 0.$$

In addition, the Lyons' lift $\mathbb{X}^{o,\leq N}$ of $\mathbb{X}^{\frac{1}{2},o,\leq 2} = \mathbb{X}^{\frac{1}{2},\leq 2}$ is unique and as stated in [74, Exercise 9.7], it holds that $\bar{\mathbb{X}}^{n,\leq N} \to \mathbb{X}^{o,\leq N}$ uniformly on $[0,T]$ as $n \to \infty$. Therefore we obtain that

$$\mathbb{X}^{\frac{1}{2},\leq N} = \mathbb{X}^{o,\leq N},$$

which concludes the proof. □

Now, suppose that $X \in C([0,T];\mathbb{R}^d)$ satisfies Property γ-(RIE) relative to some $\gamma \in [0,1]$, $p \in (2,3)$ and a sequence of partitions $\pi = (\pi^n)_{n\in\mathbb{N}}$. We set

$$\widehat{X} := (\cdot, X, [X]^{\gamma,\pi}) \in C([0,T];\mathbb{R}^{1+d+d^2}), \tag{6.1}$$

where

$$[X]^{\gamma,\pi} := ([X^1,X^1]^{\gamma,\pi},\dots,[X^1,X^d]^{\gamma,\pi},\dots,[X^d,X^1]^{\gamma,\pi},\dots,[X^d,X^d]^{\gamma,\pi}).$$

It follows by applying Lemma 6.1.4, and Lemma 6.1.7 to $(\cdot,0,0)+(0,X,0)+(0,0,[X]^{\gamma,\pi})$ that $\widehat{X}$ satisfies Property γ-(RIE) relative to γ, p and π.

We write $(e_0,e_1,\dots,e_d,\varepsilon_{11},\dots,\varepsilon_{1d},\dots,\varepsilon_{d1},\dots,\varepsilon_{dd})$ for the canonical basis of $\mathbb{R}^{1+d+d^2}$, i.e., we use the index 0 to denote the time component, and ε_{ij} for the component of $\widehat{X}$ referring to $[X^i,X^j]^{\gamma,\pi}$, so that $\langle \varepsilon_{ij}, \widehat{\mathbb{X}}^{\gamma,\infty}_t\rangle := [X^i,X^j]^{\gamma,\pi}_t$, $i,j=1,\dots,d$, $t\in[0,T]$.

We note that $t \mapsto \langle e_0, \widehat{\mathbb{X}}^{\gamma,\infty}_t\rangle$ is strictly monotonically increasing. This is necessary so that $\widehat{\mathbb{X}}^{\gamma,\infty}_T$ uniquely characterizes $\widehat{\mathbb{X}}^{\gamma,\le 2}$, see e.g. [87, 25], which itself is uniquely determined by $\widehat{X}$. See the proof of condition (iii) in Theorem 6.2.1 for a similar argument for general signatures $\mathbb{X}^{o,\infty}$.

Extending the path X to $\widehat{X}$ by a time component and the quadratic variation terms yields that the components of the γ-signature for $\gamma = \frac{1}{2}$ can be represented as linear functionals on the γ-signature, for any γ.

Proposition 6.1.16. *Suppose that $X \in C([0,T];\mathbb{R}^d)$ satisfies Property γ-(RIE) relative to some $\gamma \in [0,1]$, $p \in (2,3)$ and a sequence of partitions $\pi = (\pi^n)_{n\in\mathbb{N}}$. Then for any multi-index I, there exists $\ell^{\gamma,I} \in T(\mathbb{R}^{1+d+d^2})$ such that*

$$\langle e_I, \widehat{\mathbb{X}}^{\frac{1}{2},\infty}_t\rangle = \langle \ell^{\gamma,I}, \widehat{\mathbb{X}}^{\gamma,\infty}_t\rangle, \qquad t\in[0,T],$$

where $\widehat{\mathbb{X}}^{\frac{1}{2},\infty}$ denotes the $\frac{1}{2}$-signature and $\widehat{\mathbb{X}}^{\gamma,\infty}$ denotes the γ-signature of $\widehat{X}$, i.e., it holds that

$$\langle e_I, \widehat{\mathbb{X}}^{\frac{1}{2},\infty}_t\rangle = \sum_{0\le |J|\le N_{\gamma,I}} \ell^{\gamma,I}_J \langle e_J, \widehat{\mathbb{X}}^{\gamma,\infty}_t\rangle, \qquad t\in[0,T],$$

for $\ell^{\gamma,I} = \sum_{0\le|J|\le N_{\gamma,I}} \ell^{\gamma,I}_J e_J$, where $\ell^{\gamma,I}_J := \langle e_J, \ell^{\gamma,I}\rangle \in \mathbb{R}$ and $N_{\gamma,I}\in\mathbb{N}_0$.

Proof. Let I be a multi-index of length $|I|$ and let $t\in[0,T]$.

For $\gamma = \frac{1}{2}$, we may consider $\ell^{\gamma,I} := e_I \in T(\mathbb{R}^{1+d+d^2})$. Clearly, we have that $\langle e_I, \widehat{\mathbb{X}}^{\frac{1}{2},\infty}_t\rangle = \langle \ell^{\gamma,I}, \widehat{\mathbb{X}}^{\gamma,\infty}_t\rangle$.

Therefore suppose that $\gamma \ne \frac{1}{2}$. First, we note that since X satisfies Property γ-(RIE) relative to $\gamma \ne \frac{1}{2}$, and so does $\widehat{X}$, it satisfies Property γ-RIE relative to $\gamma = \frac{1}{2}$, see Lemma 6.1.6, that is, the γ-signature of $\widehat{X}$ for $\gamma = \frac{1}{2}$ is well-defined.

For $|I|= 0$ and $|I|= 1$, again, considering $\ell^{\gamma,I} := e_I$, we have that $\langle e_I, \widehat{\mathbb{X}}_t^{\frac{1}{2},\infty}\rangle = \langle \ell^{\gamma,I}, \widehat{\mathbb{X}}_t^{\gamma,\infty}\rangle$ by definition of the γ-signature.

Now suppose that $|I|= 2$, that is, $I = (i_1, i_2)$, $i_1, i_2 \in \{0, \dots, d, d+1, \dots, d+d^2\}$. Then we obtain that

$$\begin{aligned}\langle e_I, \widehat{\mathbb{X}}_t^{\frac{1}{2},\infty}\rangle &= \int_0^t \widehat{X}_s^{i_1} \,\mathrm{d}^{\frac{1}{2},\pi} \widehat{X}_s^{i_2} - \widehat{X}_0^{i_1}\widehat{X}_{0,t}^{i_2} \\ &= \int_0^t \widehat{X}_s^{i_1} \,\mathrm{d}^{\gamma,\pi} \widehat{X}_s^{i_2} - \widehat{X}_0^{i_1}\widehat{X}_{0,t}^{i_2} + \frac{1}{2}[\widehat{X}^{i_1}, \widehat{X}^{i_2}]_t^{\gamma,\pi} \\ &= \langle e_I, \widehat{\mathbb{X}}_t^{\gamma,\infty}\rangle + \frac{1}{2}[\widehat{X}^{i_1}, \widehat{X}^{i_2}]_t^{\gamma,\pi}.\end{aligned}$$

Since by definition of $\widehat{X}^\gamma$ and $\widehat{\mathbb{X}}^{\gamma,\infty}$,

$$\begin{aligned}[\widehat{X}^{i_1}, \widehat{X}^{i_2}]_t^{\gamma,\pi} &= \begin{cases} [X^{i_1}, X^{i_2}]_t^{\gamma,\pi}, & i_1, i_2 = 1, \dots, d, \\ 0, & \text{else} \end{cases} \\ &= \begin{cases} \langle \varepsilon_{i_1 i_2}, \widehat{\mathbb{X}}_t^{\gamma,\infty}\rangle, & i_1, i_2 = 1, \dots, d, \\ 0, & \text{else,} \end{cases}\end{aligned}$$

we then have $\langle e_I, \widehat{\mathbb{X}}_t^{\frac{1}{2},\infty}\rangle = \langle \ell^{\gamma,I}, \widehat{\mathbb{X}}_t^{\gamma,\infty}\rangle$, for $\ell^{\gamma,I} \in T(\mathbb{R}^{1+d+d^2})$ defined by

$$\ell^{\gamma,I} := \begin{cases} e_I + \frac{1}{2}\varepsilon_{i_1 i_2}, & i_1, i_2 = 1, \dots, d, \\ e_I, & \text{else.} \end{cases}$$

We apply an inductive argument: assuming that the claim holds true for any multi-index of length $|I|< n$, for $n \geq 2$, we observe that for any multi-index I of length n, using Remark 6.1.13 in the second step and the induction hypothesis in the third step, it holds that for $i_{|I|} \in \{0, \dots, d, d+1, \dots, d+d^2\}$,

$$\begin{aligned}&\langle e_I, \widehat{\mathbb{X}}_t^{\frac{1}{2},\infty}\rangle \\ &= \int_0^t \langle e_{I'}, \widehat{\mathbb{X}}_s^{\frac{1}{2},\infty}\rangle \,\mathrm{d}^{\frac{1}{2},\pi} \widehat{X}_s^{i_{|I|}} = \int_0^t \langle e_{I'}, \widehat{\mathbb{X}}_s^{\frac{1}{2},\infty}\rangle \,\mathrm{d}^{\gamma,\pi} \widehat{X}_s^{i_{|I|}} \\ &= \int_0^t \langle \ell^{\gamma,I'}, \widehat{\mathbb{X}}_s^{\gamma,\infty}\rangle \,\mathrm{d}^{\gamma,\pi} \widehat{X}_s^{i_{|I|}} \\ &= \int_0^t \sum_{0\leq |J|\leq N_{\gamma,I'}} \ell_J^{\gamma,I'} \langle e_J, \widehat{\mathbb{X}}_s^{\gamma,\infty}\rangle \,\mathrm{d}^{\gamma,\pi} \widehat{X}_s^{i_{|I|}} \\ &= \ell_\emptyset^{\gamma,I'} \widehat{X}_{0,t}^{i_{|I|}} + \sum_{j=0}^d \ell_j^{\gamma,I'} \int_0^t \langle e_j, \widehat{\mathbb{X}}_s^{\gamma,\infty}\rangle \,\mathrm{d}^{\gamma,\pi} \widehat{X}_s^{i_{|I|}} + \sum_{1<|J|\leq N_{\gamma,I'}} \ell_J^{\gamma,I'} \int_0^t \langle e_J, \widehat{\mathbb{X}}_s^{\gamma,\infty}\rangle \,\mathrm{d}^{\gamma,\pi} \widehat{X}_s^{i_{|I|}} \\ &= \ell_\emptyset^{\gamma,I'} \widehat{X}_{0,t}^{i_{|I|}} + \sum_{j=0}^d \ell_j^{\gamma,I'} (\widehat{\mathbb{X}}_{0,t}^\gamma)^{j i_{|I|}} + \sum_{1<|J|\leq N_{\gamma,I'}} \ell_J^{\gamma,I'} \int_0^t \langle e_J, \widehat{\mathbb{X}}_s^{\gamma,\infty}\rangle \,\mathrm{d}^{\gamma,\pi} \widehat{X}_s^{i_{|I|}} \\ &= \sum_{0\leq |J|\leq N_{\gamma,I'}} \ell_J^{\gamma,I'} \langle e_J \otimes e_{i_{|I|}}, \widehat{\mathbb{X}}_t^{\gamma,\infty}\rangle.\end{aligned}$$

We then can set

$$\ell^{\gamma,I} := \sum_{0\le|J|\le N_{\gamma,I'}} \ell_J^{\gamma,I'} e_J \otimes e_{i_{|I|}} \in T(\mathbb{R}^{1+d+d^2}),$$

which concludes the proof. □

Consequently, any linear functional on the γ-signature can be written as a linear functional on the signature (defined via Lyons' extension theorem). We will use this and further comment on this in Section 6.2, when deriving a pathwise universal approximation theorem.

Corollary 6.1.17. *Suppose that $X \in C([0,T];\mathbb{R}^d)$ satisfies Property γ-*(RIE) *relative to some $\gamma \in [0,1]$, $p \in (2,3)$ and a sequence of partitions $\pi = (\pi^n)_{n\in\mathbb{N}}$. For any $\ell^o \in T(\mathbb{R}^{1+d+d^2})$, there exists $\ell^\gamma \in T(\mathbb{R}^{1+d+d^2})$ such that $\langle \ell^o, \widehat{\mathbb{X}}_t^{o,\infty}\rangle = \langle \ell^\gamma, \widehat{\mathbb{X}}_t^{\gamma,\infty}\rangle$, for $t \in [0,T]$.*

Proof. Let $\ell^o \in T(\mathbb{R}^{1+d+d^2})$, that is,

$$\langle \ell^o, \mathbf{a}\rangle := \sum_{0\le|I|\le N_o} \ell_I^o \langle e_I, \mathbf{a}\rangle, \qquad \mathbf{a} \in T((\mathbb{R}^{1+d+d^2})),$$

for $\ell^o = \sum_{0\le|I|\le N_o} \ell_I^o e_I$, where $\ell_I^o := \langle e_I, \ell^o\rangle \in \mathbb{R}$ and $N_o \in \mathbb{N}_0$.

By Proposition 6.1.15 and Proposition 6.1.16, this gives for any $t \in [0,T]$, using the notation of Proposition 6.1.16, that

$$\begin{aligned}
\langle \ell^o, \widehat{\mathbb{X}}_t^{o,\infty}\rangle &= \sum_{0\le|I|\le N_o} \ell_I^o \langle e_I, \widehat{\mathbb{X}}_t^{o,\infty}\rangle \\
&= \sum_{0\le|I|\le N_o} \ell_I^o \langle e_I, \widehat{\mathbb{X}}_t^{\frac{1}{2},\infty}\rangle \\
&= \sum_{0\le|I|\le N_o} \ell_I^o \langle \ell^{\gamma,I}, \widehat{\mathbb{X}}_t^{\gamma,\infty}\rangle \\
&= \sum_{0\le|I|\le N_o} \ell_I^o \Big(\sum_{0\le|J|\le N_{\gamma,I}} \ell_J^{\gamma,I} \langle e_J, \widehat{\mathbb{X}}_t^{\gamma,\infty}\rangle \Big) \\
&= \sum_{0\le|I|\le N_o} \sum_{0\le|J|\le N_{\gamma,I}} \ell_I^o \ell_J^{\gamma,I} \langle e_J, \widehat{\mathbb{X}}_t^{\gamma,\infty}\rangle.
\end{aligned}$$

Setting $\ell^\gamma := \sum_{0\le|I|\le N_o} \sum_{0\le|J|\le N_{\gamma,I}} \ell_I^o \ell_J^{\gamma,I} e_J \in T(\mathbb{R}^{1+d+d^2})$, we conclude the proof. □

6.2 A pathwise universal approximation theorem for signatures

The success of signature-based methods is due to a powerful property that allows for, heuristically speaking, approximating continuous functionals on the path on compact sets by linear functionals on its signature, analogously to polynomials approximating continuous real-valued functions. The corresponding result follows by an application of the Stone–Weierstrass theorem, which requires that the linear span of the signature form an algebra.

This particularly holds true for the signature defined via Lyons' lift, see Definition 6.1.8; we refer to, e.g., [121, 47]. We will therefore first recall the proof of this classical version of the universal approximation theorem before deriving the pathwise universal approximation theorem for the γ-signature for $\gamma \in [0,1]$, as an extension to a more general class of signatures.

To that end, we consider the subspace of time-extended weakly geometric p-rough paths, defined by

$$\widehat{C}_o^{p\text{-var}}([0,T]; G^2(\mathbb{R}^{d+1})) := \{\widehat{\mathbb{X}}^{o,\leq 2} \in C_o^{p\text{-var}}([0,T]; G^2(\mathbb{R}^{d+1})) : \langle e_0, \widehat{\mathbb{X}}_t^{o,\leq 2} \rangle = t,\, t \in [0,T]\}.$$

Theorem 6.2.1. *Let $p \in (2,3)$. Let $K \subset \widehat{C}_o^{p\text{-var}}([0,T]; G^2(\mathbb{R}^{d+1}))$ be a compact subset, bounded with respect to the p-variation norm and consider a continuous function $f\colon K \to \mathbb{R}$. Then for every $\varepsilon > 0$, there exists a linear functional $\ell \in T(\mathbb{R}^{d+1})$ such that*

$$\sup_{\widehat{\mathbb{X}}^{o,\leq 2} \in K} |f(\widehat{\mathbb{X}}^{o,\leq 2}) - \langle \ell, \widehat{\mathbb{X}}_T^{o,\infty} \rangle| < \varepsilon,$$

where $\widehat{\mathbb{X}}^{o,\infty}$ denotes the signature of $\widehat{X} := \Pi_1(\widehat{\mathbb{X}}^{o,\leq 2})$.

Proof. The result follows by an application of the Stone–Weierstrass theorem to the set

$$\mathcal{A} := \operatorname{span}\{K \ni \widehat{\mathbb{X}}^{o,\leq 2} \mapsto \langle e_I, \widehat{\mathbb{X}}_T^{o,\infty} \rangle \in \mathbb{R} : I \in \{1,\dots,d\}^N, N \in \mathbb{N}_0\}.$$

Therefore we have to show that $\mathcal{A}$

(i) is a vector subspace of $C(K;\mathbb{R})$,

(ii) is a subalgebra and contains a non-zero constant function, and

(iii) separates points.

(i): By [74, Corollary 9.11], the map $\widehat{\mathbb{X}}^{o,\leq 2} \mapsto \langle e_I, \widehat{\mathbb{X}}_T^{o,\infty} \rangle$ is continuous on bounded sets for every multi-index I with respect to $d_{p\text{-var}} := \|\cdot;\cdot\|_p$. More precisely, the map

$$(K, d_{p\text{-var}}) \ni \widehat{\mathbb{X}}^{o,\leq 2} \mapsto \widehat{\mathbb{X}}^{o,\leq N} \in (C_o^{p\text{-var}}([0,T]; G^N(\mathbb{R}^{d+1})), d_{p\text{-var}}),$$

is continuous on K with respect to $d_{p\text{-var}}$, for every $N \geq 3$. Moreover, the evaluation map

$$(C_o^{p\text{-var}}([0,T]; G^N(\mathbb{R}^{d+1})), d_{p\text{-var}}) \ni \widehat{\mathbb{X}}^{o,\leq N} \mapsto \widehat{\mathbb{X}}_T^{o,\leq N} \in (G^N(\mathbb{R}^{d+1}), \rho)$$

is continuous, where ρ denotes the metric induced by the norm on $T_1^N(\mathbb{R}^{d+1})$. Here, we used that we can equip $G^N(\mathbb{R}^{d+1})$ with the metric ρ, see e.g. [74, Remark 7.31]. This yields that the map

$$(K, d_{p\text{-var}}) \ni \widehat{\mathbb{X}}^{o,\leq 2} \mapsto \widehat{\mathbb{X}}_T^{o,\leq N} \in (G^N(\mathbb{R}^{d+1}), \rho)$$

is continuous. Since $\widehat{\mathbb{X}}^{o,\leq N}_T \mapsto \langle e_I, \widehat{\mathbb{X}}^{o,\leq N}_T \rangle$ is continuous for any multi-index I, we can thus conclude that the map

$$(K, d_{p\text{-var}}) \ni \widehat{\mathbb{X}}^{o,\leq 2} \mapsto \langle e_I, \widehat{\mathbb{X}}^{o,\infty}_T \rangle \in \mathbb{R}$$

is continuous with respect to $d_{p\text{-var}}$.

(ii): Since $\widehat{\mathbb{X}}^{o,\infty}_T$ is a group-like element, i.e., $\widehat{\mathbb{X}}^{o,\infty}_T \in G((\mathbb{R}^{d+1}))$, the shuffle property holds, and thus $\mathcal{A}$ is a subalgebra. Moreover, since $\langle e_\emptyset, \widehat{\mathbb{X}}^{o,\infty}_T \rangle = 1$, it contains a non-zero constant function.

(iii): For the point separation, let us consider $\widehat{\mathbb{X}}^{o,\leq 2}$, $\widehat{\mathbb{Y}}^{o,\leq 2} \in K$, with $\widehat{\mathbb{X}}^{o,\leq 2} \neq \widehat{\mathbb{Y}}^{o,\leq 2}$. We show that there exists a $k \in \mathbb{N}$, $I \in \{0, \dots, d\}^N$, $N \in \{0, 1, 2\}$ such that

$$\langle (e_I \sqcup\!\sqcup\, e_0^{\otimes k}) \otimes e_0, \widehat{\mathbb{X}}^{o,\leq 2}_T \rangle \neq \langle (e_I \sqcup\!\sqcup\, e_0^{\otimes k}) \otimes e_0, \widehat{\mathbb{Y}}^{o,\leq 2}_T \rangle.$$

We proceed with a proof by contradiction. Assume that for all $k \in \mathbb{N}$, $I \in \{0, \dots, d\}^N$, $N \in \{0, 1, 2\}$, we have

$$\langle (e_I \sqcup\!\sqcup\, e_0^{\otimes k}) \otimes e_0, \widehat{\mathbb{X}}^{o,\leq 2}_T \rangle = \langle (e_I \sqcup\!\sqcup\, e_0^{\otimes k}) \otimes e_0, \widehat{\mathbb{Y}}^{o,\leq 2}_T \rangle.$$

We first note that

$$\langle e_0^{\otimes k}, \widehat{\mathbb{X}}^{o,\leq 2}_t \rangle = \frac{t^k}{k!}.$$

Moreover, using the shuffle property, we have

$$\langle (e_I \sqcup\!\sqcup\, e_0^{\otimes k}) \otimes e_0, \widehat{\mathbb{X}}^{o,\leq 2}_T \rangle = \int_0^T \langle e_I, \widehat{\mathbb{X}}^{o,\leq 2}_t \rangle \langle e_0^{\otimes k}, \widehat{\mathbb{X}}^{o,\leq 2}_t \rangle \, \mathrm{d}t = \int_0^T \langle e_I, \widehat{\mathbb{X}}^{o,\leq 2}_t \rangle \frac{t^k}{k!} \, \mathrm{d}t.$$

Similarly, we have

$$\langle (e_I \sqcup\!\sqcup\, e_0^{\otimes k}) \otimes e_0, \widehat{\mathbb{Y}}^{o,\leq 2}_T \rangle = \int_0^T \langle e_I, \widehat{\mathbb{Y}}^{o,\leq 2}_t \rangle \frac{t^k}{k!} \, \mathrm{d}t.$$

Using the Hahn–Banach theorem, which tells us that continuous, linear functionals separate points, and since $t \mapsto t$ is strictly monotone, we obtain that

$$\langle e_I, \widehat{\mathbb{X}}^{o,\leq 2}_t \rangle = \langle e_I, \widehat{\mathbb{Y}}^{o,\leq 2}_t \rangle,$$

for all $t \in [0, T]$ and all $I \in \{0, \dots, d\}^N$, $N \in \{0, 1, 2\}$. However, this contradicts the assumption that $\widehat{\mathbb{X}}^{o,\leq 2}$, $\widehat{\mathbb{Y}}^{o,\leq 2}$ are distinct. Thus we can conclude that $\mathcal{A}$ is point separating.

□

The proof or, more precisely, the Stone–Weierstrass theorem, makes use of the shuffle product property of the signature, which holds for the signature defined via Lyons' lift since it is a group-like valued path. We aim to avoid this restriction when considering the γ-signature for a general $\gamma \in [0, 1]$. However, this is just a path with values in the extended tensor algebra so that the linear functionals on the γ-signature do not form an algebra of functionals on the path space (for $\gamma \neq \frac{1}{2}$).

We circumvent this by extending the path by the correct correction term, which is the corresponding quadratic variation term. While admittedly increasing the dimension of the path, this suffices so that any linear functional on the γ-signature can be written as a linear functional on the signature defined via Lyons' extension theorem, see Corollary 6.1.17. As a consequence, we are able to deduce the universal approximation property of linear functionals on the γ-signature from Theorem 6.2.1.

Theorem 6.2.2. *Let $\gamma \in [0,1]$, $p \in (2,3)$ and $\pi = (\pi^n)_{n\in\mathbb{N}}$ be a sequence of partitions of the interval $[0,T]$. Let $K \subset C^{p\text{-var}}([0,T];\mathbb{R}^{1+d+d^2})$ be a compact subset, bounded with respect to the p-variation norm and consider a continuous function $f\colon K \to \mathbb{R}$. Further, for some $M > 0$, let $K_M \subset K$ be the subset defined by*

$$K_M := \{\widehat{X} = (\cdot, X, [X]^{\gamma,\pi}) \in K : X \text{ satisfies Property } \gamma\text{-(RIE) relative to } \gamma,\, p \text{ and } \pi,\\ \|(\widehat{X}, \widehat{\mathbb{X}}^\gamma)\|_p + \|[\widehat{X}]^{\gamma,\pi}\|_1 \leq M\}.$$

Then for every $\varepsilon > 0$, there exists a linear functional $\ell^\gamma \in T(\mathbb{R}^{1+d+d^2})$ such that

$$\sup_{\widehat{X}\in K_M} |f(\widehat{X}) - \langle \ell^\gamma, \widehat{\mathbb{X}}^{\gamma,\infty}_T\rangle| < \varepsilon,$$

where $\widehat{\mathbb{X}}^{\gamma,\infty}$ denotes the γ-signature of $\widehat{X}$.

Proof. First, we recall that if a path $X \in C([0,T];\mathbb{R}^d)$ satisfies Property γ-(RIE) relative to γ, p and π, then so does $\widehat{X} = (\cdot, X, [X]^\gamma) \in C([0,T];\mathbb{R}^{1+d+d^2})$, see Lemma 6.1.7.

We note that if a path $\widehat{X} \in C([0,T];\mathbb{R}^{1+d+d^2})$ satisfies Property γ-(RIE) relative to γ, p and $\pi = (\pi^n)_{n\in\mathbb{N}}$, then $\widehat{X}$ extends canonically to a weakly geometric rough path via

$$\iota\colon \widehat{X} \mapsto \widehat{\mathbb{X}}^{\gamma,o,\leq 2} := (1, \widehat{X}_{0,\cdot}, \widehat{\mathbb{X}}^{\gamma,o}_{0,\cdot}) := (1, \widehat{X}_{0,\cdot}, \widehat{\mathbb{X}}^{\gamma}_{0,\cdot} + [\widehat{X}]^\gamma),$$

see Lemma 6.1.9, that is, $\iota(\widehat{X}) \in C^{p\text{-var}}_o([0,T];G^2(\mathbb{R}^{1+d+d^2}))$. Further, we observe that for any $\widehat{X} \in K_M$, it holds that

$$\|\widehat{\mathbb{X}}^{\gamma,o}\|_{\frac{p}{2}} \leq \|\widehat{\mathbb{X}}^{\gamma}\|_{\frac{p}{2}} + \|[\widehat{X}]^\gamma\|_1 \leq M,$$

thus we can embed K_M into $\iota(K_M) := \{\iota(\widehat{X}) : \widehat{X} \in K_M\}$, which is a subset of the compact subset $K_M^{\leq 2} := \{\widehat{\mathbb{X}}^{o,\leq 2} : \|\mathbf{1}; \widehat{\mathbb{X}}^{o,\leq 2}\|_p \leq M\}$ of $C^{p\text{-var}}_o([0,T];G^2(\mathbb{R}^{1+d+d^2}))$. We now consider the continuous function

$$f^{\leq 2} : K_M^{\leq 2} \to \mathbb{R}, \qquad \widehat{\mathbb{X}}^{o,\leq 2} \mapsto f(\Pi_1(\widehat{\mathbb{X}}^{o,\leq 2})).$$

Let $\varepsilon > 0$. By Theorem 6.2.1 then there exists some $\ell \in T(\mathbb{R}^{1+d+d^2})$ such that

$$\sup_{\widehat{\mathbb{X}}^{o,\leq 2} \in K_M^{\leq 2}} |f^{\leq 2}(\widehat{\mathbb{X}}^{o,\leq 2}) - \langle \ell, \widehat{\mathbb{X}}_T^{o,\infty} \rangle| < \varepsilon,$$

where $\widehat{\mathbb{X}}^{o,\infty}$ denotes the signature of $\widehat{X}$, see Definition 6.1.8.

By Lemma 6.1.17, there exists some $\ell^\gamma \in T(\mathbb{R}^{1+d+d^2})$ such that $\langle \ell, \widehat{\mathbb{X}}_T^{o,\infty} \rangle = \langle \ell^\gamma, \widehat{\mathbb{X}}_T^{\gamma,\infty} \rangle$. Thus we obtain that

$$\begin{aligned}
&\sup_{\widehat{X} \in K_M} |f(\widehat{X}) - \langle \ell^\gamma, \widehat{\mathbb{X}}_T^{\gamma,\infty} \rangle| \\
&\quad = \sup_{\widehat{X} \in K_M} |f(\widehat{X}) - \langle \ell, \widehat{\mathbb{X}}_T^{o,\infty} \rangle| \\
&\quad = \sup_{\widehat{X} \in K_M} |f^{\leq 2}(\iota(\widehat{X})) - \langle \ell, \widehat{\mathbb{X}}_T^{o,\infty} \rangle| \\
&\quad = \sup_{\iota(\widehat{X}) \in \iota(K_M)} |f^{\leq 2}(\iota(\widehat{X})) - \langle \ell, \widehat{\mathbb{X}}_T^{o,\infty} \rangle| \\
&\quad \leq \sup_{\widehat{\mathbb{X}}^{o,\leq 2} \in K_M^{\leq 2}} |f^{\leq 2}(\widehat{\mathbb{X}}^{o,\leq 2}) - \langle \ell, \widehat{\mathbb{X}}_T^{o,\infty} \rangle| \\
&\quad < \varepsilon.
\end{aligned}$$

□

6.3 Application to continuous semimartingales

In this section, we apply the deterministic theory developed in Section 6.1 and Section 6.2 to continuous semimartingales.

In fact, continuous semimartingales fit well into the theory of signatures when adopting the notion of stochastic integration. That is, the signature can be defined as the collection of iterated integrals via stochastic integration. Because it is obeying first order calculus, one usually considers Stratonovich integration, which almost surely coincides with Lyons' lift, thus implying a universal approximation theorem for continuous path functionals.

Throughout, let X be a d-dimensional continuous semimartingale, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)_{t \in [0,T]}$ satisfying the usual conditions, i.e., completeness and right-continuity.

Definition 6.3.1. *Let X be a d-dimensional continuous semimartingale. Its* Stratonovich-signature *is the stochastic process $\mathbb{X}^{\circ,\infty} = (\mathbb{X}_t^{\circ,\infty})_{t \in [0,T]}$ with values in $T_1((\mathbb{R}^d))$, whose components are recursively defined by*

$$\langle e_\emptyset, \mathbb{X}_t^{\circ,\infty} \rangle := 1, \qquad \langle e_I, \mathbb{X}_t^{\circ,\infty} \rangle := \int_0^t \langle e_{I'}, \mathbb{X}_s^{\circ,\infty} \rangle \circ \mathrm{d}X_s{}^{i_{|I|}},$$

for each $I = (i_1, \dots, i_{|I|})$ *and* $t \in [0,T]$, *where* $\circ$ *denotes the Stratonovich integral. Its projection* $\mathbb{X}^{\circ,\le N}$ *on* $T^N(\mathbb{R}^d)$ *is given by*

$$\mathbb{X}_t^{\circ,\le N} = \Pi_{\le N}(\mathbb{X}_t^{\circ,\infty}) = \sum_{|I|\le N} \langle e_I, \mathbb{X}_t^{\circ,\infty}\rangle e_I,$$

and called Stratonovich-signature of X truncated at level N, *which takes values in* $G^N(\mathbb{R}^d)$ *for all* $t \in [0,T]$. *The increments of the Stratonovich-signature* $\mathbb{X}^{\circ,\infty}$ *are defined by*

$$\mathbb{X}_{s,t}^{\circ,\infty} := (\mathbb{X}_s^{\circ,\infty})^{-1} \otimes \mathbb{X}_t^{\circ,\infty}, \qquad (s,t) \in \Delta_T.$$

It turns out that, if the semimartingale X satisfies Property γ-(RIE) relative to $\gamma \in [0,1]$, $p \in (2,3)$ and a suitable sequence of partitions, we obtain a canonical signature which corresponds $\mathbb{P}$-almost surely with the signature defined via Lyons' lift and the Stratonovich-signature.

Lemma 6.3.2. *Let* $\gamma \in [0,1]$, *let* $p \in (2,3)$ *and let* $\pi^n = \{\tau_k^n\}$, $n \in \mathbb{N}$, *be a sequence of adapted partitions (so that each* τ_k^n *is a stopping time), such that for almost every* $\omega \in \Omega$, $(\pi^n(\omega))_{n\in\mathbb{N}}$ *is a sequence of (finite) partitions of* $[0,T]$ *with vanishing mesh size.*

Let X *be a continuous* d*-dimensional semimartingale, and suppose that for almost every* $\omega \in \Omega$, $\sup\{|X_{\tau_k^n(\omega),\tau_{k+1}^n(\omega)}(\omega)| : k = 0, \dots, N_n - 1\}$ *converges to* 0 *as* $n \to \infty$, *and that the sample path* $X(\omega)$ *satisfies Property* γ-(RIE) *relative to* γ, p *and* $(\pi^n(\omega))_{n\in\mathbb{N}}$.

- *(i) The random weakly geometric rough path pathwise defined via Proposition 6.1.3 for* $\gamma = \frac{1}{2}$ *and the random weakly geometric rough path pathwise defined via Lemma 6.1.9 for* $\gamma \in [0,1]$ *coincide* $\mathbb{P}$*-almost surely.*
- *(ii) The random weakly geometric rough path pathwise defined via Lemma 6.1.9 and the Stratonovich-signature of* X *truncated at level* 2 *coincide* $\mathbb{P}$*-almost surely.*
- *(iii) The random signature* $\mathbb{X}^{o,\infty}$ *pathwise defined via Definition 6.1.8, more precisely, Remark 6.1.11, the random signature* $\mathbb{X}^{\frac{1}{2},\infty}$ *pathwise defined via Definition 6.1.12 and the Stratonovich-signature* $\mathbb{X}^{\circ,\infty}$ *of* X *coincide* $\mathbb{P}$*-almost surely.*

Proof. (i): By Lemma 6.1.6, we know that if a path satisfies Property γ-(RIE) relative to some $\gamma \in [0,1]$, then it particularly satisfies Property γ-(RIE) relative to $\gamma = \frac{1}{2}$. Then the claim holds true because of Lemma 6.1.1 and $\mathbb{X}_{0,t}^{\frac{1}{2}} = \mathbb{X}_{0,t}^{\frac{1}{2},o}$, $t \in [0,T]$.

(ii): By construction, the pathwise rough integral $\int_0^t X_r(\omega) \otimes \mathrm{d}^{\gamma,\pi} X_r(\omega)$ constructed via Property γ-(RIE) is given by the limit as $n \to \infty$ of Riemann sums:

$$\sum_{k=0}^{N_n-1} (X_{\tau_k^n(\omega)}(\omega) + \gamma X_{\tau_k^n(\omega),\tau_{k+1}^n(\omega)}) \otimes X_{\tau_k^n(\omega)\wedge t,\tau_{k+1}^n(\omega)\wedge t}(\omega).$$

Suppose that $\gamma = \frac{1}{2}$. Then it is known that these Riemann sums converge uniformly in probability to the Stratonovich integral $\int_0^t X_r \otimes \circ \mathrm{d}X_r$, see e.g. [147, Chapter II, Theorem 21, Theorem 22]. And the result follows from the (almost sure) uniqueness of limits; see also part (i) of Lemma 5.3.1.

Suppose that $\gamma \neq \frac{1}{2}$. Then adding $[X(\omega)]_{0,t}^{\gamma,\pi^n}$,

$$\begin{aligned}
&\sum_{k=0}^{N_n-1} (X_{\tau_k^n(\omega)}(\omega) + \gamma X_{\tau_k^n(\omega),\tau_{k+1}^n(\omega)}) \otimes X_{\tau_k^n(\omega)\wedge t,\tau_{k+1}^n(\omega)\wedge t}(\omega) \\
&\qquad + \frac{1}{2}(1-2\gamma) X_{\tau_k^n(\omega)\wedge t,\tau_{k+1}^n(\omega)\wedge t}(\omega) \otimes X_{\tau_k^n(\omega)\wedge t,\tau_{k+1}^n(\omega)\wedge t}(\omega) \\
&= \sum_{k=0}^{N_n-1} (X_{\tau_k^n(\omega)}(\omega) + \frac{1}{2} X_{\tau_k^n(\omega)\wedge t,\tau_{k+1}^n(\omega)\wedge t}) \otimes X_{\tau_k^n(\omega)\wedge t,\tau_{k+1}^n(\omega)\wedge t}(\omega) \\
&\qquad + \gamma (X_{\tau_k^n(\omega),\tau_{k+1}^n(\omega)}) - X_{\tau_k^n(\omega)\wedge t,\tau_{k+1}^n(\omega)\wedge t}(\omega)) \otimes X_{\tau_k^n(\omega)\wedge t,\tau_{k+1}^n(\omega)\wedge t}(\omega),
\end{aligned}$$

which again converges uniformly in probability to the Stratonovich integral $\int_0^t X_r \otimes \circ \mathrm{d}X_r$.

(iii): We first note that by Proposition 6.1.15, the random signatures pathwise defined via Definition 6.1.8 (Lyons' lift of the weakly geometric rough path) and via Definition 6.1.12 (γ-signature for $\gamma = \frac{1}{2}$) coincide $\mathbb{P}$-almost surely.

By (ii), the random weakly geometric rough path and the Stratonovich-signature of X truncated at level 2 coincide $\mathbb{P}$-almost surely, and take values in $G^2(\mathbb{R}^d)$. Since Lyons' lift is unique, see [74, Theorem 9.5], and the Stratonovich-signature of X truncated at any level $N \geq 3$ takes values in $G^N(\mathbb{R}^d)$, and so does the random signature truncated at level N pathwise defined via Lyons' lift of the weakly geometric rough path, the proof is complete. □

Corollary 6.3.3. *Let X be a d-dimensional continuous semimartingale, $\widehat{X} := (\cdot, X)$, and let $\mathcal{S}^{(2)} := \{\widehat{\mathbb{X}}^{\circ,\leq 2}(\omega) : \omega \in \Omega\}$. Further, let $p \in (2,3)$ and $K \subset \widehat{C}_o^{p\text{-var}}([0,T]; G^2(\mathbb{R}^{d+1}))$ be a compact subset of the subspace of time-extended weakly geometric p-rough paths, see Theorem 6.2.1, bounded with respect to the p-variation norm and consider a continuous function $f\colon K \to \mathbb{R}$. Then for every $\varepsilon > 0$, there exists a linear functional $\ell \in T(\mathbb{R}^{d+1})$ such that for almost every $\omega \in \Omega$,*

$$|f(\widehat{\mathbb{X}}^{\circ,\leq 2}(\omega)) - \langle \ell, \widehat{\mathbb{X}}_T^{\circ,\infty}(\omega)\rangle| < \varepsilon \qquad \text{for all} \quad \widehat{\mathbb{X}}^{\circ,\leq 2}(\omega) \in K \cap \mathcal{S}^{(2)},$$

where $\widehat{\mathbb{X}}^{\circ,\infty}$ denotes the Stratonovich-signature of $\widehat{X}$.

Analogously to the Stratonovich-signature, we now define the Itô-signature of a continuous semimartingale via iterated stochastic Itô integration, which is the preferred choice from a modeling perspective when having, for example, a financial application in mind.

Definition 6.3.4. *Let X be a d-dimensional continuous semimartingale. Its* Itô-signature *is the stochastic process $\mathbb{X}^\infty = (\mathbb{X}^\infty_t)_{t\in[0,T]}$ with values in $T_1((\mathbb{R}^d))$, whose components are recursively defined by*

$$\langle e_\emptyset, \mathbb{X}^\infty_t\rangle := 1, \qquad \langle e_I, \mathbb{X}^\infty_t\rangle := \int_0^t \langle e_{I'}, \mathbb{X}^\infty_s\rangle \,\mathrm{d}X^{i_{|I|}}_s,$$

for each $I = (i_1, \ldots, i_{|I|})$ and $t \in [0,T]$, where the integral is given as an Itô integral. Its projection $\mathbb{X}^{\le N}$ on $T^N(\mathbb{R}^d)$ is given by

$$\mathbb{X}^{\le N}_t = \Pi_{\le N}(\mathbb{X}^\infty_t) = \sum_{|I|\le N} \langle e_I, \mathbb{X}^\infty_t\rangle e_I,$$

and called Itô-signature of X truncated at level N. *The increments of the signature $\mathbb{X}^\infty$ are defined by*

$$\mathbb{X}^\infty_{s,t} := (\mathbb{X}^\infty_s)^{-1} \otimes \mathbb{X}^\infty_t, \qquad (s,t) \in \Delta_T.$$

It turns out that, if the semimartingale X satisfies Property γ-(RIE) relative to $\gamma = 0$, which is equivalent to Property (RIE), see Lemma 6.1.6, then the γ-signature for $\gamma = 0$ and the Itô-signature coincide almost surely.

Lemma 6.3.5. *Let $p \in (2,3)$ and let $\pi^n = \{\tau^n_k\}$, $n \in \mathbb{N}$, be a sequence of adapted partitions (so that each τ^n_k is a stopping time), such that for almost every $\omega \in \Omega$, $(\pi^n(\omega))_{n\in\mathbb{N}}$ is a sequence of (finite) partitions of $[0,T]$ with vanishing mesh size.*

Let X be a d-dimensional continuous semimartingale, and suppose that for almost every $\omega \in \Omega$, $\sup\{|X_{\tau^n_k(\omega),\tau^n_{k+1}(\omega)}(\omega)| : k = 0, \ldots, N_n - 1\}$ converges to 0 as $n \to \infty$, and that the sample path $X(\omega)$ satisfies Property γ-(RIE) relative to $\gamma = 0$, p and $(\pi^n(\omega))_{n\in\mathbb{N}}$.

(i) The random rough path pathwise defined via Proposition 6.1.3 for $\gamma = 0$ and the Itô-signature of X truncated at level 2 coincide $\mathbb{P}$-almost surely.

(ii) The random γ-signature $\mathbb{X}^{0,\infty}$ pathwise defined via Definition 6.1.12 for $\gamma = 0$ and the Itô-signature $\mathbb{X}^\infty$ of X coincide $\mathbb{P}$-almost surely.

Proof. (i): Since Property γ-(RIE) for $\gamma = 0$ and Property (RIE) are equivalent, see also Lemma 6.1.6, this is the statement of part (i) of Lemma 3.2.1.

(ii): By (ii), it is left to show the statement for any multi-index I of length $|I| > 2$. By definition, the pathwise integral $\int_0^t \langle e_{I'}, \mathbb{X}^{0,\infty}_r(\omega)\rangle \mathrm{d}^{0,\pi} X^{i_{|I|}}_r(\omega)$ may be taken as the limit as $n \to \infty$ of left-point Riemann sums:

$$\sum_{k=0}^{N_n-1} \langle e_{I'}, \mathbb{X}^{0,\infty}_{\tau^n_k(\omega)}(\omega)\rangle X^{i_{|I|}}_{\tau^n_k(\omega)\wedge t, \tau^n_{k+1}(\omega)\wedge t}(\omega),$$

see Remark 6.1.11. It is known that these Riemann sums converge uniformly in probability to the Itô integral $\int_0^t \langle e_{I'}, \mathbb{X}_r^\infty \rangle \,\mathrm{d}X_r^{i_{|I|}}$ (see e.g. [147, Chapter II, Theorem 21]), and the result thus follows from the (almost sure) uniqueness of limits. $\square$

Moreover, if the semimartingale X satisfies Property γ-(RIE), the pathwise quadratic variation and the stochastic quadratic variation coincide almost surely.

Lemma 6.3.6. *Let $\gamma \in [0,1]$, $p \in (2,3)$ and let $\pi^n = \{\tau_k^n\}$, $n \in \mathbb{N}$, be a sequence of adapted partitions (so that each τ_k^n is a stopping time), such that for almost every $\omega \in \Omega$, $(\pi^n(\omega))_{n\in\mathbb{N}}$ is a sequence of (finite) partitions of $[0,T]$ with vanishing mesh size.*

Let X be a d-dimensional continuous semimartingale, and suppose that for almost every $\omega \in \Omega$, $\sup\{|X_{\tau_k^n(\omega),\tau_{k+1}^n(\omega)}(\omega)| : k = 0,\dots,N_n - 1\}$ converges to 0 as $n \to \infty$, and that the sample path $X(\omega)$ satisfies Property γ-(RIE) relative to γ, p and $(\pi^n(\omega))_{n\in\mathbb{N}}$. We define the random variable

$$\begin{aligned}\widehat{X} &:= (\cdot, X, (1-2\gamma)[X]) \\ &:= (\cdot, X, (1-2\gamma)[X,X]^{11}, \dots, (1-2\gamma)[X]^{1d}, \dots, (1-2\gamma)[X]^{d1}, \dots, (1-2\gamma)[X]^{dd}),\end{aligned}$$

where $[X] = ([X]^{ij})_{1\le i,j\le d}$ denotes the quadratic variation of X. Then $\widehat{X}$ and the random variable that is pathwise defined via (6.1) coincide $\mathbb{P}$-almost surely.

Proof. This clearly holds true for $\gamma = \frac{1}{2}$. Therefore suppose that $\gamma \neq \frac{1}{2}$. By definition, the pathwise quadratic variation $[X^i(\omega), X^j(\omega)]^{\gamma,\pi}$ is given by the limit as $n \to \infty$ of:

$$(1-2\gamma) \sum_{k=0}^{N_n-1} X^i_{\tau_k^n(\omega)\wedge t, \tau_{k+1}^n(\omega)\wedge t}(\omega) X^j_{\tau_k^n(\omega)\wedge t, \tau_{k+1}^n(\omega)\wedge t}(\omega).$$

We know that these sums converge uniformly (in $t \in [0,T]$) in probability to the quadratic variation $(1-2\gamma)[X]^{ij}$, see e.g. [147, Chapter II, Theorem 22]. By taking a subsequence, if necessary, it follows the (almost sure) uniqueness of limits. $\square$

As a consequence of Theorem 6.2.2 and Lemma 6.3.5 and Lemma 6.3.6, we formulate universality of the Itô-signature of a continuous semimartingale whose sample paths almost surely satisfy Property γ-(RIE) for $\gamma = 0$ or, equivalently, Property (RIE). This holds true for various semimartingales relative to suitable sequences of partitions. We refer to Section 3.2.

Theorem 6.3.7 (Universal approximation theorem for the Itô-signature)**.** *Let $p \in (2,3)$ and let $\pi^n = \{\tau_k^n\}$, $n \in \mathbb{N}$, be a sequence of adapted partitions (so that each τ_k^n is a stopping time), such that for almost every $\omega \in \Omega$, $(\pi^n(\omega))_{n\in\mathbb{N}}$ is a sequence of (finite) partitions of $[0,T]$ with vanishing mesh size.*

Let X be a d-dimensional continuous semimartingale, and suppose that for almost every $\omega \in \Omega$, $\sup\{|X_{\tau^n_k(\omega),\tau^n_{k+1}(\omega)}(\omega)| : k = 0,\dots,N_n - 1\}$ converges to 0 as $n \to \infty$, and that the sample path $X(\omega)$ satisfies Property γ-(RIE) relative to $\gamma = 0$, p and $(\pi^n(\omega))_{n\in\mathbb{N}}$.
Let $\widehat{X} := (\cdot, X, [X])$, and $\mathcal{S}^{(1)} := \{\widehat{X}(\omega) : \omega \in \Omega\}$. Further, let $K \subset C^{p\text{-var}}([0,T];\mathbb{R}^{1+d+d^2})$ be a compact subset, bounded with respect to the p-variation norm and consider a continuous function $f\colon K \to \mathbb{R}$. For some $M > 0$, let $K_M \subset K$ be the subset defined by

$$K_M := \{\widehat{X} = (\cdot, X, [X]^{0,\pi}) \in K : X \text{ satisfies Property } \gamma\text{-(RIE) relative to } \gamma = 0,\ p \text{ and } \pi,\ \|(\widehat{X}, \widehat{\mathbb{X}}^0)\|_p + \|[\widehat{X}]^{0,\pi}\|_1 \le M\}.$$

Then for every $\varepsilon > 0$, there exists a linear functional $\ell \in T(\mathbb{R}^{1+d+d^2})$ such that for almost every $\omega \in \Omega$,

$$|f(\widehat{X}(\omega)) - \langle \ell, \widehat{\mathbb{X}}^\infty_T(\omega)\rangle| < \varepsilon \qquad \text{for all} \quad \widehat{X}(\omega) \in K_M \cap \mathcal{S}^{(1)},$$

where $\widehat{\mathbb{X}}^\infty$ denotes the Itô-signature of $\widehat{X}$.

Proof. We use that for almost every $\omega \in \Omega$, the random γ-signature of $\widehat{X}(\omega)$ for $\gamma = 0$ and the Itô-signature $\widehat{\mathbb{X}}^\infty(\omega)$ coincide, see Lemma 6.3.6 and part (ii) of Lemma 6.3.5.

The claim then immediately follows from the pathwise universal approximation theorem for linear functionals on the γ-signature, which is Theorem 6.2.2. □

Remark 6.3.8. *An analogous result also holds true when considering the Stratonovich-signature of X instead of the Itô-signature of X (also if almost all sample paths only satisfy Property γ-(RIE) relative to $\gamma = \frac{1}{2}$). This can be shown using the results of the previous sections. This is, however, weaker than the classical universal approximation theorem stated in Corollary 6.3.3 since we impose an assumption on the sample paths of the semimartingale to allow for a statement about the Itô-signature.*

Appendix

A.1 Local estimates for rough integration

The following local estimates are needed to prove the existence and uniqueness result in Theorem 2.1.3, and the continuity result in Theorem 2.1.5.

Lemma A.1.1. *Let $\mathbf{X} \in \mathcal{D}^p([0,T];\mathbb{R}^d)$ for $p \in (2,3)$ and $(Y,Y') \in \mathcal{V}^p_X([0,T];\mathbb{R}^k)$. Suppose that the non-anticipative functional $(F,F')\colon \mathcal{V}^p_X([0,T];\mathbb{R}^k) \to \mathcal{V}^p_X([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$ satisfies Assumption 2.1.1 (i) with some constant C_F. Then, we have the local estimate*

$$\|R^{\int_0^\cdot F(Y)\,\mathrm{d}\mathbf{X}}\|_{\frac{p}{2},[s,t]} \lesssim C_F(1+\|Y,Y'\|_{X,p,[s,t]})^2(1+\|X\|_{p,[s,t]})^2\|\mathbf{X}\|_{p,[s,t]},$$

for all $(s,t) \in \Delta_T$, where the implicit multiplicative constant depends only on p.

Proof. Let $(V,V') \in \mathcal{V}^p_X([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$, and set $\Xi_{u,v} := V_u X_{u,v} + V'_u \mathbb{X}_{u,v}$ and $\delta\Xi_{u,r,v} := \Xi_{u,v} - \Xi_{u,r} - \Xi_{r,v}$ for $s \le u < r < v \le t$. Here, strictly speaking, in writing $V'_u\mathbb{X}_{u,v}$, we use the canonical identification of $\mathcal{L}(\mathbb{R}^d;\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$ with $\mathcal{L}(\mathbb{R}^d \otimes \mathbb{R}^d;\mathbb{R}^k)$. We note that

$$\begin{aligned}\left|R^{\int_0^\cdot V\,\mathrm{d}\mathbf{X}}_{u,v}\right| &\le \left|\int_u^v V_r\,\mathrm{d}\mathbf{X}_r - \Xi_{u,v}\right| + |V'_u||\mathbb{X}_{u,v}|\\ &\le \left|\int_u^v V_r\,\mathrm{d}\mathbf{X}_r - \Xi_{u,v}\right| + (|V'_s|+\|V'\|_{p,[s,t]})|\mathbb{X}_{u,v}|.\end{aligned}$$

Using Chen's relation, one can show that

$$-\delta\Xi_{u,r,v} = R^V_{u,r}X_{r,v} + V'_{u,r}\mathbb{X}_{r,v},$$

which gives that

$$\begin{aligned}&|\delta\Xi_{u,r,v}|\\ &\quad\le \|R^V\|_{\frac{p}{2},[u,r]}\|X\|_{p,[r,v]} + \|V'\|_{p,[u,r]}\|\mathbb{X}\|_{\frac{p}{2},[r,v]}\\ &\quad= w_{1,1}(u,r)^{\frac{2}{p}}w_{2,1}(r,v)^{\frac{1}{p}} + w_{1,2}(u,r)^{\frac{1}{p}}w_{2,2}(r,v)^{\frac{2}{p}},\end{aligned}$$

where $w_{1,1}(s,t) := \|R^V\|^{\frac{p}{2}}_{\frac{p}{2},[s,t]}$, $w_{2,1}(s,t) := \|X\|^p_{p,[s,t]}$, $w_{1,2}(s,t) := \|V'\|^p_{p,[s,t]}$, $w_{2,2}(s,t) := \|\mathbb{X}\|^{\frac{p}{2}}_{\frac{p}{2},[s,t]}$, $(s,t) \in \Delta_T$, are control functions and $\frac{1}{p} + \frac{2}{p} > 1$. It then follows from the generalized sewing lemma, see [75, Theorem 2.5], that

$$\begin{aligned}
&\|R^{\int_0^\cdot V \mathrm{d}\mathbf{X}}\|_{\frac{p}{2},[s,t]} \\
&\quad\lesssim (\|R^V\|_{\frac{p}{2},[s,t]}\|X\|_{p,[s,t]}+\|V'\|_{p,[s,t]}\|\mathbb{X}\|_{\frac{p}{2},[s,t]}+(|V'_s|+\|V'\|_{p,[s,t]})\|\mathbb{X}\|_{\frac{p}{2},[s,t]}) \\
&\quad\lesssim \|V,V'\|_{X,p,[s,t]}\|\mathbf{X}\|_{p,[s,t]},
\end{aligned}$$

where the implicit multiplicative constant depends only on p.

For $(V,V') = (F(Y),F'(Y,Y'))$, using Assumption 2.1.1 (i), we therefore obtain the estimate. □

Lemma A.1.2. *For $p \in (2,3)$, suppose $\mathbf{X},\widetilde{\mathbf{X}} \in \mathcal{D}^p([0,T];\mathbb{R}^d)$, $(Y,Y') \in \mathcal{V}^p_X([0,T];\mathbb{R}^k)$, $(\widetilde{Y},\widetilde{Y}') \in \mathcal{V}^p_{\widetilde{X}}([0,T];\mathbb{R}^k)$, and that the non-anticipative functional $(F,F')\colon \mathcal{V}^p_X([0,T];\mathbb{R}^k) \to \mathcal{V}^p_X([0,T];\mathcal{L}(\mathbb{R}^d;\mathbb{R}^k))$ satisfies Assumption 2.1.4 (i) and (ii) given $X,\widetilde{X}$. Then, we have the local estimate*

$$\begin{aligned}
&\|R^{\int_0^\cdot F(Y)\mathrm{d}\mathbf{X}} - R^{\int_0^\cdot F(\widetilde{Y})\mathrm{d}\widetilde{\mathbf{X}}}\|_{\frac{p}{2},[s,t]} \\
&\quad\lesssim C_{F,K,X,\widetilde{X}}(|Y_s-\widetilde{Y}_s|+\|Y,Y';\widetilde{Y},\widetilde{Y}'\|_{X,\widetilde{X},p,[s,t]}+\|X-\widetilde{X}\|_{p,[s,t]})(\|\mathbf{X}\|_{p,[s,t]}\vee\|\widetilde{\mathbf{X}}\|_{p,[s,t]}) \\
&\qquad + C_F(1+K)^2(1+\|X\|_{p,[s,t]}\vee\|\widetilde{X}\|_{p,[s,t]})^2\|\mathbf{X};\widetilde{\mathbf{X}}\|_{p,[s,t]}
\end{aligned}$$

for all $(s,t) \in \Delta_T$, if $\|Y,Y'\|_{X,p,[s,t]}$, $\|\widetilde{Y},\widetilde{Y}'\|_{\widetilde{X},p,[s,t]} \leq K$, for some $K>0$, where the implicit multiplicative constant depends on p, $\|\mathbf{X}\|_p$ and $\|\widetilde{\mathbf{X}}\|_p$.

Proof. It follows from [75, Lemma 3.4] that for any $(V,V') \in \mathcal{V}^p_X$, $(\widetilde{V},\widetilde{V}') \in \mathcal{V}^p_{\widetilde{X}}$,

$$\begin{aligned}
&\|R^{\int_0^\cdot V\mathrm{d}\mathbf{X}} - R^{\int_0^\cdot \widetilde{V}\mathrm{d}\widetilde{\mathbf{X}}}\|_{\frac{p}{2},[s,t]} \\
&\quad\lesssim_p (1+\|\mathbf{X}\|_{p,[s,t]}+\|\widetilde{\mathbf{X}}\|_{p,[s,t]})(\|V,V';\widetilde{V},\widetilde{V}'\|_{X,\widetilde{X},p,[s,t]}\|\mathbf{X}\|_{p,[s,t]} \\
&\qquad + \|\widetilde{V},\widetilde{V}'\|_{\widetilde{X},p,[s,t]}\|\mathbf{X};\widetilde{\mathbf{X}}\|_{p,[s,t]}) \\
&\quad\lesssim \|V,V';\widetilde{V},\widetilde{V}'\|_{X,\widetilde{X},p,[s,t]}\|\mathbf{X}\|_{p,[s,t]}+\|\widetilde{V},\widetilde{V}'\|_{\widetilde{X},p,[s,t]}\|\mathbf{X};\widetilde{\mathbf{X}}\|_{p,[s,t]},
\end{aligned}$$

where the implicit multiplicative constant depends on p, $\|\mathbf{X}\|_p$ and $\|\widetilde{\mathbf{X}}\|_p$. For $(V,V') = (F(Y),F'(Y,Y'))$, $(\widetilde{V},\widetilde{V}') = (F(\widetilde{Y}),F'(\widetilde{Y},\widetilde{Y}'))$, using Assumption 2.1.4 (ii), we therefore obtain the estimate. □

A.2 Proof of Theorem 3.1.1

Proof of Theorem 3.1.1. Step 1. Let $L > 0$ such that $\|A\|_r, \|H\|_r, \|\mathbf{X}\|_p \leq L$, and let $w\colon \Delta_T \to [0,\infty)$ be the right-continuous control function given by

$$w(s,t) = \|A\|_{r,[s,t]}^r + \|H\|_{r,[s,t]}^r + \|X\|_{p,[s,t]}^p + \|\mathbb{X}\|_{\frac{p}{2},[s,t]}^{\frac{p}{2}}, \qquad \text{for} \quad (s,t) \in \Delta_T.$$

For $t \in (0,T]$, we define the map $\mathcal{M}_t\colon \mathcal{V}_X^{q,r}([0,t];\mathbb{R}^k) \to \mathcal{V}_X^{q,r}([0,t];\mathbb{R}^k)$ by

$$\mathcal{M}_t(Y,Y') = \bigg(y_0 + \int_0^\cdot b(H_s,Y_s)\,\mathrm{d}A_s + \int_0^\cdot \sigma(H_s,Y_s)\,\mathrm{d}\mathbf{X}_s, \sigma(H,Y)\bigg),$$

and, for $\delta \geq 1$, introduce the subset of controlled paths

$$\mathcal{B}_t^{(\delta)} = \Big\{(Y,Y') \in \mathcal{V}_X^{q,r}([0,t];\mathbb{R}^k) : (Y_0,Y_0') = (y_0,\sigma(H_0,y_0)),\ \|Y,Y'\|_{X,q,r}^{(\delta)} \leq 1\Big\},$$

where

$$\|Y,Y'\|_{X,q,r}^{(\delta)} := \|Y'\|_{q,[0,t]} + \delta\|R^Y\|_{r,[0,t]}.$$

Applying standard estimates for Young and rough integrals (e.g. [75, Proposition 2.4 and Lemma 3.6]), for any $(Y,Y') \in \mathcal{B}_t^{(\delta)}$, we deduce that

$$\|\mathcal{M}_t(Y,Y')\|_{X,q,r}^{(\delta)} \leq C_1\bigg(\frac{1}{\delta} + \delta(\|A\|_{r,[0,t]} + \|H\|_{r,[0,t]} + \|\mathbf{X}\|_{p,[0,t]})\bigg),$$

for a constant $C_1 \geq \frac{1}{2}$ which depends only on $p, q, r, \|b\|_{C_b^2}, \|\sigma\|_{C_b^3}$, and L. Let $\delta = \delta_1 := 2C_1$, so that

$$\|\mathcal{M}_t(Y,Y')\|_{X,q,r}^{(\delta_1)} \leq \frac{1}{2} + 2C_1^2(2w(0,t)^{\frac{1}{r}} + w(0,t)^{\frac{1}{p}} + w(0,t)^{\frac{2}{p}}).$$

By the right-continuity of w, we can then take $t = t_1$ sufficiently small such that

$$\|\mathcal{M}_{t_1}(Y,Y')\|_{X,q,r}^{(\delta_1)} \leq 1,$$

and we have that $\mathcal{B}_{t_1}^{(\delta_1)}$ is invariant under $\mathcal{M}_{t_1}$.

Step 2. Let $(Y,Y'), (\widetilde{Y},\widetilde{Y}') \in \mathcal{B}_t^{(\delta)}$, for some (new) $\delta \geq 1$ and $t \in (0,t_1]$. Applying standard estimates for Young and rough integrals (e.g. [75, Proposition 2.4, Lemma 3.1 and Lemma 3.7]), we deduce that

$$\begin{aligned}
&\|\mathcal{M}_t(Y,Y') - \mathcal{M}_t(\widetilde{Y},\widetilde{Y}')\|_{X,q,r}^{(\delta)} \\
&\quad \leq C_2\Big(\|R^Y - R^{\widetilde{Y}}\|_{r,[0,t]} + \delta(\|Y' - \widetilde{Y}'\|_{q,[0,t]} + \|R^Y - R^{\widetilde{Y}}\|_{r,[0,t]})(\|A\|_{r,[0,t]} + \|\mathbf{X}\|_{p,[0,t]})\Big),
\end{aligned}$$

where $C_2 > \frac{1}{2}$ depends only on $p, q, r, \|b\|_{C_b^2}, \|\sigma\|_{C_b^3}$ and L. Let $\delta = \delta_2 := 2C_2 > 1$, so that

$$\begin{aligned}
&\|\mathcal{M}_t(Y,Y') - \mathcal{M}_t(\widetilde{Y},\widetilde{Y}')\|_{X,q,r}^{(\delta_2)} \\
&\quad \leq \frac{\delta_2}{2}\|R^Y - R^{\widetilde{Y}}\|_{r,[0,t]} \\
&\qquad + 2C_2^2(\|Y' - \widetilde{Y}'\|_{q,[0,t]} + \|R^Y - R^{\widetilde{Y}}\|_{r,[0,t]})(w(0,t)^{\frac{1}{r}} + w(0,t)^{\frac{1}{p}} + w(0,t)^{\frac{2}{p}}).
\end{aligned}$$

Again by the right-continuity of w, we then take $t = t_2 \leq t_1$ sufficiently small such that

$$\begin{aligned}\|\mathcal{M}_{t_2}(Y,Y') - \mathcal{M}_{t_2}(\widetilde{Y},\widetilde{Y}')\|_{X,q,r}^{(\delta_2)} &\leq \frac{1}{2}\|Y' - \widetilde{Y}'\|_{q,[0,t_2]} + \frac{\delta_2+1}{2}\|R^Y - R^{\widetilde{Y}}\|_{r,[0,t_2]} \\ &\leq \frac{\delta_2+1}{2\delta_2}\|(Y,Y') - (\widetilde{Y},\widetilde{Y}')\|_{X,q,r}^{(\delta_2)},\end{aligned}$$

from which it follows that $\mathcal{M}_{t_2}$ is a contraction on the Banach space $(\mathcal{B}_{t_2}^{(\delta_1)}, \|\cdot\|_{X,q,r}^{(\delta_2)})$. The fixed point of this map is the unique solution of the RDE (3.4) over the time interval $[0,t_2]$.

Step 3. Now let $\widetilde{A} \in D^{q_1}$, $\widetilde{H} \in D^{q_2}$, $\widetilde{\mathbf{X}} = (\widetilde{X}, \widetilde{\mathbb{X}}) \in \mathcal{D}^p$ and $\widetilde{y}_0 \in \mathbb{R}^n$, such that $\|\widetilde{A}\|_r, \|\widetilde{H}\|_r, \|\widetilde{\mathbf{X}}\|_p \leq L$. By considering instead the control function w given by

$$\begin{aligned}w(s,t) = {} & \|A\|_{r,[s,t]}^r + \|H\|_{r,[s,t]}^r + \|X\|_{p,[s,t]}^p + \|\mathbb{X}\|_{\frac{p}{2},[s,t]}^{\frac{p}{2}} \\ & + \|\widetilde{A}\|_{r,[s,t]}^r + \|\widetilde{H}\|_{r,[s,t]}^r + \|\widetilde{X}\|_{p,[s,t]}^p + \|\widetilde{\mathbb{X}}\|_{\frac{p}{2},[s,t]}^{\frac{p}{2}}, \qquad \text{for} \quad (s,t) \in \Delta_T,\end{aligned}$$

it follows from the above that there exist unique solutions $(Y,Y') \in \mathcal{V}_X^{q,r}([0,t_2];\mathbb{R}^k)$ and $(\widetilde{Y},\widetilde{Y}') \in \mathcal{V}_{\widetilde{X}}^{q,r}([0,t_2];\mathbb{R}^k)$ of the RDE (3.4), with data $(A,H,\mathbf{X},y_0)$ and $(\widetilde{A},\widetilde{H},\widetilde{\mathbf{X}},\widetilde{y}_0)$ respectively, over a sufficiently small time interval $[0,t_2]$. Standard estimates for Young and rough integrals (e.g. [75, Proposition 2.4, Lemma 3.1 and Lemma 3.7]) imply, after some calculation, that for any $\delta \geq 1$ and $t \in (0,t_2]$,

$$\begin{aligned}&\|Y' - \widetilde{Y}'\|_{q,[0,t]} + \delta\|R^Y - R^{\widetilde{Y}}\|_{r,[0,t]} \\ &\quad \leq C_3\Big(|y_0 - \widetilde{y}_0| + |H_0 - \widetilde{H}_0| + \|H - \widetilde{H}\|_{r,[0,t]} + \|R^Y - R^{\widetilde{Y}}\|_{r,[0,t]} \\ &\qquad + \delta(\|A - \widetilde{A}\|_{r,[0,t]} + \|\mathbf{X};\widetilde{\mathbf{X}}\|_{p,[0,t]}) \\ &\qquad + \delta(|y_0 - \widetilde{y}_0| + |H_0 - \widetilde{H}_0| + \|H - \widetilde{H}\|_{r,[0,t]} + \|Y' - \widetilde{Y}'\|_{q,[0,t]} + \|R^Y - R^{\widetilde{Y}}\|_{r,[0,t]}) \\ &\qquad\quad \times (\|A\|_{r,[0,t]} + \|\mathbf{X}\|_{p,[0,t]})\Big),\end{aligned}$$

where $C_3 > 0$ depends only on $p, q, r, \|b\|_{C_b^2}, \|\sigma\|_{C_b^3}$ and L. Let $\delta = \delta_3 := C_3 + 1$, so that

$$\begin{aligned}&\|Y' - \widetilde{Y}'\|_{q,[0,t]} + \|R^Y - R^{\widetilde{Y}}\|_{r,[0,t]} \\ &\quad \leq C_3\Big(|y_0 - \widetilde{y}_0| + |H_0 - \widetilde{H}_0| + \|H - \widetilde{H}\|_{r,[0,t]} + \delta_3(\|A - \widetilde{A}\|_{r,[0,t]} + \|\mathbf{X};\widetilde{\mathbf{X}}\|_{p,[0,t]}) \\ &\qquad + \delta_3(|y_0 - \widetilde{y}_0| + |H_0 - \widetilde{H}_0| + \|H - \widetilde{H}\|_{r,[0,t]} + \|Y' - \widetilde{Y}'\|_{q,[0,t]} + \|R^Y - R^{\widetilde{Y}}\|_{r,[0,t]}) \\ &\qquad\quad \times (w(0,t)^{\frac{1}{r}} + w(0,t)^{\frac{1}{p}} + w(0,t)^{\frac{2}{p}})\Big).\end{aligned}$$

By taking $t = t_3 \leq t_2$ sufficiently small, we deduce that

$$\begin{aligned}&\|Y - \widetilde{Y}\|_{p,[0,t_3]} + \|Y' - \widetilde{Y}'\|_{q,[0,t_3]} + \|R^Y - R^{\widetilde{Y}}\|_{r,[0,t_3]} \\ &\quad \leq C_4\Big(|y_0 - \widetilde{y}_0| + |H_0 - \widetilde{H}_0| + \|H - \widetilde{H}\|_{r,[0,t_3]} + \|A - \widetilde{A}\|_{r,[0,t_3]} + \|\mathbf{X};\widetilde{\mathbf{X}}\|_{p,[0,t_3]}\Big),\end{aligned} \tag{A.2}$$

for a new constant C_4, still depending only on $p, q, r, \|b\|_{C_b^2}, \|\sigma\|_{C_b^3}$ and L.

Step 4. We infer from the above that there exists a constant $\varepsilon > 0$, which depends only on $p, q, r, \|b\|_{C_b^2}, \|\sigma\|_{C_b^3}$ and L, such that, given initial values $Y_s, \widetilde{Y}_s \in \mathbb{R}^k$, the local solutions (Y, Y') and $(\widetilde{Y}, \widetilde{Y}')$ established above exist on any interval $[s,t]$ such that $w(s,t) \leq \varepsilon$. Moreover, these local solutions satisfy an estimate on this interval of the form in (A.2).

By [75, Lemma 1.5], there exists a partition $\mathcal{P} = \{0 = t_0 < t_1 < \cdots < t_N = T\}$, such that $w(t_i, t_{i+1}-) < \varepsilon$ for every $i = 0, 1, \ldots, N-1$. We can then define the solutions (Y, Y') and $(\widetilde{Y}, \widetilde{Y}')$ on each of the half-open intervals $[t_i, t_{i+1})$. Given the solutions on $[t_i, t_{i+1})$, the values $Y_{t_{i+1}}$ and $\widetilde{Y}_{t_{i+1}}$ at the right end-point of the interval are uniquely determined by the jumps of $A, \widetilde{A}, \mathbf{X}$ and $\widetilde{\mathbf{X}}$ at time t_{i+1}. We thus deduce the existence of unique solutions (Y, Y') and $(\widetilde{Y}, \widetilde{Y}')$ of the RDE on the entire interval $[0,T]$.

Since w is superadditive, we have that

$$w(t_0, t_1-) + w(t_1-, t_1) + w(t_1, t_2-) + \cdots + w(t_{N-1}, t_N-) + w(t_N-, t_N) \leq w(0,T).$$

It is then straightforward to see that the partition $\mathcal{P}$ may be chosen such that the number of partition points in $\mathcal{P}$ may be bounded by a constant depending only on ε and $w(0,T)$. Thus, we may combine the local estimates in (A.2) on each of the subintervals, together with simple estimates on the jumps at the end-points of these subintervals, to obtain the global estimate in (3.5). $\square$

A.3 The convergence of piecewise constant approximations

In the following, we adopt the notation

$$\liminf_{n\to\infty} \mathcal{P}^n := \bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} \mathcal{P}^n$$

for the times $t \in [0,T]$ which, as $n \to \infty$, eventually belong to all subsequent partitions in the sequence $(\mathcal{P}^n)_{n\in\mathbb{N}}$. The following proposition generalizes the result of [7, Proposition 2.14] so that the sequence of partitions is no longer assumed to be nested.

Proposition A.3.1. *Let $\mathcal{P}^n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = T\}$, $n \in \mathbb{N}$, be a sequence of partitions with vanishing mesh size, so that $|\mathcal{P}^n| \to 0$ as $n \to \infty$. Let $F\colon [0,T] \to \mathbb{R}^d$ be a càdlàg path, and let*

$$F_t^n = F_T \mathbf{1}_{\{T\}}(t) + \sum_{k=0}^{N_n-1} F_{t_k^n} \mathbf{1}_{[t_k^n, t_{k+1}^n)}(t), \qquad t \in [0,T],$$

be the piecewise constant approximation of F along $\mathcal{P}^n$. Let

$$J_F := \{t \in (0,T] : F_{t-} \neq F_t\}$$

be the set of jump times of F. The following are equivalent:

(i) $J_F \subseteq \liminf_{n\to\infty} \mathcal{P}^n$,

(ii) the sequence $(F^n)_{n\in\mathbb{N}}$ *converges pointwise to* F,

(iii) the sequence $(F^n)_{n\in\mathbb{N}}$ *converges uniformly to* F.

Proof. We first show that conditions (i) and (ii) are equivalent. To this end, suppose that $J_F \subseteq \liminf_{n\to\infty} \mathcal{P}^n$ and let $t \in (0,T]$. If $t \in J_F$, then there exists $m \geq 1$ such that $t \in \mathcal{P}^n$ for all $n \geq m$. In this case we then have that $F^n_t = F_t$ for all $n \geq m$. If $t \notin J_F$, then F is continuous at time t, and, since the mesh size $|\mathcal{P}^n| \to 0$, it follows that $F^n_t \to F_t$ as $n \to \infty$.

Now suppose instead that there exists a $t \in J_F$ such that $t \notin \liminf_{n\to\infty} \mathcal{P}^n$. Then there exists a subsequence $(n_j)_{j\in\mathbb{N}}$ such that $F^{n_j}_t \to F_{t-}$ as $j \to \infty$. Since $F_{t-} \neq F_t$, it follows that $F^n_t \nrightarrow F_t$. This establishes the equivalence of (i) and (ii).

Since (iii) clearly implies (ii), it only remains to show that (ii) implies (iii). By [69, Theorem 3.3], it is enough to show that the family of paths $\{F^n : n \in \mathbb{N}\}$ is equiregulated in the sense of [69, Definition 3.1].

Step 1. Let $t \in (0,T]$ and $\varepsilon > 0$. Since the left limit F_{t-} exists, there exists $\delta > 0$ with $t - \delta > 0$, such that

$$|F_s - F_{t-}| < \frac{\varepsilon}{2} \qquad \text{for all} \quad s \in (t-\delta, t).$$

Since $|\mathcal{P}^n| \to 0$ as $n \to \infty$, there exists an $m \in \mathbb{N}$ such that, for every $n \geq m$, there exists a partition point $t^n_k \in \mathcal{P}^n$ such that $t - \delta < t^n_k < t - \frac{\delta}{2}$.

Let

$$u := \max\left(\left(t - \frac{\delta}{2}, t\right) \cap \bigcup_{n<m} \mathcal{P}^n\right),$$

where here we define $\max(\emptyset) := t - \frac{\delta}{2}$.

Take any $s \in (u,t)$ and any $n \in \mathbb{N}$. Let $i = \max\{k : t^n_k \leq s\}$ and $j = \max\{k : t^n_k < t\}$, so that $F^n_s = F_{t^n_i}$ and $F^n_{t-} = F_{t^n_j}$.

If $n \geq m$, then there exists a point $t^n_k \in \mathcal{P}^n$ such that $t - \delta < t^n_k < t - \frac{\delta}{2} \leq u < s$, and it follows that $t^n_i, t^n_j \in (t-\delta, t)$. If instead $n < m$, and if there exists a partition point $t^n_k \in (t - \frac{\delta}{2}, t)$, then $t - \frac{\delta}{2} < t^n_k \leq u < s$, and it again follows that $t^n_i, t^n_j \in (t-\delta, t)$. In either case, we then have that

$$|F^n_s - F^n_{t-}| = |F_{t^n_i} - F_{t^n_j}| \leq |F_{t^n_i} - F_{t-}| + |F_{t^n_j} - F_{t-}| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

The remaining case is when $n < m$ but $(t - \frac{\delta}{2}, t) \cap \mathcal{P}^n = \emptyset$. In this case the path F^n is constant on the interval $[t - \frac{\delta}{2}, t)$ and, since $s \in (t - \frac{\delta}{2}, t)$, we have that $F^n_s = F^n_{t-}$.

In each case, we have that $|F^n_s - F^n_{t-}| < \varepsilon$ for all $s \in (u,t)$ and all $n \in \mathbb{N}$.

Step 2. Let $t \in (J_F \cup \{0\}) \setminus \{T\}$ and $\varepsilon > 0$. Since F is right-continuous, there exists a $\delta > 0$ with $t + \delta < T$, such that

$$|F_s - F_t| < \varepsilon \qquad \text{for all} \quad s \in [t, t+\delta).$$

Since condition (ii) implies condition (i), we know that $t \in \liminf_{n\to\infty} \mathcal{P}^n$, so that there exists an $m \in \mathbb{N}$ such that $t \in \cap_{n\geq m} \mathcal{P}^n$. Let

$$u := \min\left((t, t+\delta) \cap \bigcup_{n<m} \mathcal{P}^n\right),$$

where here we define $\min(\emptyset) := t + \delta$.

Take any $s \in (t, u)$, and any $n \in \mathbb{N}$. Let $i = \max\{k : t_k^n \leq s\}$, so that $F_s^n = F_{t_i^n}$.

If $n \geq m$, then $t \in \mathcal{P}^n$, so $F_t^n = F_t$ and, moreover, $t \leq t_i^n \leq s < u \leq t + \delta$, so that in particular $t_i^n \in [t, t+\delta)$, and hence

$$|F_s^n - F_t^n| = |F_{t_i^n} - F_t| < \varepsilon.$$

If $n < m$, then there does not exist any partition point $t_k^n \in (t, u) \cap \mathcal{P}^n$. It follows that the path F^n is constant on the interval $[t, u)$, so that in particular $F_s^n = F_t^n$.

In each case, we have that $|F_s^n - F_t^n| < \varepsilon$ for all $s \in (t, v)$ and all $n \in \mathbb{N}$.

Step 3. Let $t \in (0, T) \setminus J_F$ and $\varepsilon > 0$. Since F is continuous at time t, there exists a $\delta > 0$ with $0 < t - \delta$ and $t + \delta < T$, such that

$$|F_s - F_t| < \frac{\varepsilon}{2} \qquad \text{for all} \quad s \in (t-\delta, t+\delta).$$

Since $|\mathcal{P}^n| \to 0$ as $n \to \infty$, there exists an $m \in \mathbb{N}$ such that, for every $n \geq m$, there exists a partition point $t_k^n \in \mathcal{P}^n$ such that $t - \delta < t_k^n < t$. Let

$$u := \min\left((t, t+\delta) \cap \bigcup_{n<m} \mathcal{P}^n\right),$$

where here we define $\min(\emptyset) := t + \delta$.

Take any $s \in (t, u)$ and any $n \in \mathbb{N}$. Let $i = \max\{k : t_k^n \leq s\}$ and $j = \max\{k : t_k^n \leq t\}$, so that $F_s^n = F_{t_i^n}$ and $F_t^n = F_{t_j^n}$.

If $n \geq m$, then there exists a point $t_k^n \in \mathcal{P}^n$ such that $t_k^n \in (t - \delta, t)$, and it follows that $t_i^n, t_j^n \in (t - \delta, t + \delta)$, so that

$$|F_s^n - F_t^n| = |F_{t_i^n} - F_{t_j^n}| \leq |F_{t_i^n} - F_t| + |F_{t_j^n} - F_t| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

If $n < m$, then there does not exist any partition point $t_k^n \in (t, u) \cap \mathcal{P}^n$. It follows that the path F^n is constant on the interval $[t, u)$, so that in particular $F_s^n = F_t^n$.

In each case, we have that $|F_s^n - F_t^n| < \varepsilon$ for all $s \in (t, u)$ and all $n \in \mathbb{N}$. It follows that the family of paths $\{F^n : n \in \mathbb{N}\}$ is indeed equiregulated. $\square$

Theorem A.3.2. *Let $p \in (2,3)$, $q \in [p,\infty)$ and $r \in [\frac{p}{2},2)$ such that $\frac{1}{p}+\frac{1}{r}>1$ and $\frac{1}{p}+\frac{1}{q}=\frac{1}{r}$, and let $\mathcal{P}^n=\{0=t_0^n<t_1^n<\cdots<t_{N_n}^n=T\}$, $n \in \mathbb{N}$, be a sequence of partitions with vanishing mesh size. Suppose that X satisfies Property* (RIE) *relative to p and $(\mathcal{P}^n)_{n\in\mathbb{N}}$, and let $\mathbf{X}$ be the canonical rough path lift of X, as constructed in* (3.9). *Let $(F,F') \in \mathcal{V}_X^{q,r}$ be a controlled path with respect to X, and suppose that $J_F \subseteq \liminf_{n\to\infty}\mathcal{P}^n$, where J_F is the set of jump times of F. Then the rough integral of (F,F') against $\mathbf{X}$ is given by*

$$\int_0^t F_u \,\mathrm{d}\mathbf{X}_u = \lim_{n\to\infty}\sum_{k=0}^{N_n-1} F_{t_k^n} X_{t_k^n\wedge t, t_{k+1}^n\wedge t},$$

where the convergence is uniform in $t \in [0,T]$.

The previous theorem generalizes the result of [7, Theorem 2.15] so that the sequence of partitions is no longer assumed to be nested. The proof of Theorem A.3.2 follows the proof of [7, Theorem 2.15] almost verbatim. The only difference is that, rather than using [7, Proposition 2.14] to establish the uniform convergence of F^n to F, we can instead use Proposition A.3.1 (which does not require the sequence of partitions to be nested).

A.4 Some essential results in rough path theory

In this appendix, we collect some fundamental results in the theory of càdlàg rough paths. While the analogous results are standard for stochastic Itô integration, they are less well-known and in some cases novel in the context of rough integration.

Throughout this section, we fix the following assumption.

Assumption. *Let $p \in [2,3)$, $q \in [p,\infty)$ and $r \in [\frac{p}{2},2)$ such that $\frac{1}{p}+\frac{1}{r}>1$ and $\frac{1}{p}+\frac{1}{q}=\frac{1}{r}$.*

In the following, we consider a general càdlàg rough path $\mathbf{X} = (X,\mathbb{X}) \in \mathcal{D}^p([0,T];\mathbb{R}^d)$ as introduced in Section 4.2.1, and do not impose Property (RIE) on X.

A.4.1 Rough integration with respect to controlled paths

This subsection contains slight modifications of results in [7] on rough integration with respect to controlled paths in $\mathcal{V}^{q,r}_X([0,T];\mathbb{R}^m)$.

Lemma A.4.1 (Proposition 2.4 in [7])**.** *Let $\mathbf{X} = (X,\mathbb{X}) \in \mathcal{D}^p$ be a càdlàg rough path and let $(F,F'),(G,G') \in \mathcal{V}^{q,r}_X$ be controlled paths with remainders R^F and R^G, respectively. Then the limit*[2]

$$\int_0^T F_u \,\mathrm{d}G_u := \lim_{|\mathcal{P}|\to 0} \sum_{[s,t]\in\mathcal{P}} F_s \otimes G_{s,t} + (F'_s \otimes G'_s)\mathbb{X}_{s,t} \tag{A.3}$$

exists along every sequence of partitions $\mathcal{P}$ of $[0,T]$ with mesh size $|\mathcal{P}|\to 0$, and comes with the estimate

$$\begin{aligned}&\left|\int_s^t F_u \,\mathrm{d}G_u - F_s \otimes G_{s,t} - (F'_s \otimes G'_s)\mathbb{X}_{s,t}\right| \\ &\quad\leq C\Big(\|F'\|_\infty(\|G'\|^q_{q,[s,t)}+\|X\|^p_{p,[s,t)})^{\frac{1}{r}}\|X\|_{p,[s,t]}+\|F\|_{p,[s,t)}\|R^G\|_{r,[s,t]} \\ &\qquad + \|R^F\|_{r,[s,t)}\|G'\|_\infty\|X\|_{p,[s,t]}+\|F'G'\|_{q,[s,t)}\|\mathbb{X}\|_{\frac{p}{2},[s,t]}\Big),\end{aligned}$$

for every $(s,t)\in\Delta_T$, where the constant C depends only on p,q and r.

Lemma A.4.2 (Proposition 2.7 (ii) in [7])**.** *Let $\mathbf{X} = (X,\mathbb{X})$, $\widetilde{\mathbf{X}} = (\widetilde{X},\widetilde{\mathbb{X}})$ be càdlàg rough paths, and let $(F,F'),(G,G') \in \mathcal{V}^{q,r}_X$ and $(\widetilde{F},\widetilde{F}'),(\widetilde{G},\widetilde{G}') \in \mathcal{V}^{q,r}_{\widetilde{X}}$ be controlled paths. Let $M>0$ be an upper bound for $\|F,F'\|_{\mathcal{V}^{q,r}_X}$, $\|G,G'\|_{\mathcal{V}^{q,r}_X}$, $\|\widetilde{F},\widetilde{F}'\|_{\mathcal{V}^{q,r}_{\widetilde{X}}}$, $\|\widetilde{G},\widetilde{G}'\|_{\mathcal{V}^{q,r}_{\widetilde{X}}}$, $\|\mathbf{X}\|_p$ and $\|\widetilde{\mathbf{X}}\|_p$. Then, there exists a constant C, depending only on p, q, r and M, such that*

$$\left\|\int_0^\cdot F_u\,\mathrm{d}G_u - \int_0^\cdot \widetilde{F}_u\,\mathrm{d}\widetilde{G}_u\right\|_q \leq C\Big(\|F,F';\widetilde{F},\widetilde{F}'\|_{\mathcal{V}^{q,r}_X,\mathcal{V}^{q,r}_{\widetilde{X}}}+\|G,G';\widetilde{G},\widetilde{G}'\|_{\mathcal{V}^{q,r}_X,\mathcal{V}^{q,r}_{\widetilde{X}}}+\|\mathbf{X};\widetilde{\mathbf{X}}\|_p\Big),$$

where $\int_0^\cdot F_u\,\mathrm{d}G_u$ and $\int_0^\cdot \widetilde{F}_u\,\mathrm{d}\widetilde{G}_u$ are rough integrals, as defined in (A.3).

[2] In writing $F'_s \otimes G'_s$, we technically mean the 4-tensor whose $ijk\ell$ component is given by $[F'_s \otimes G'_s]^{ijk\ell} = (F'_s)^{ij}(G'_s)^{k\ell}$, and we interpret the "multiplication" $(F'_s \otimes G'_s)\mathbb{X}_{s,t}$ as the matrix whose ik component is given by $[(F'_s \otimes G'_s)\mathbb{X}_{s,t}]^{ik} = \sum_j \sum_\ell (F'_s)^{ij}(G'_s)^{k\ell}\mathbb{X}^{j\ell}_{s,t}$.

A.4.2 The product of controlled paths

Lemma A.4.3. *Let $\mathbf{X} = (X, \mathbb{X})$ be a càdlàg rough path. The product operator, given by*

$$\begin{aligned}\mathcal{V}_X^{q,r}([0,T];\mathbb{R}^k) \times \mathcal{V}_X^{q,r}([0,T];\mathbb{R}^k) &\to \mathcal{V}_X^{q,r}([0,T];\mathbb{R}^k),\\ ((F,F'),(G,G')) &\mapsto (FG,(FG)'),\end{aligned}$$

where

$$(FG)^i := F^i G^i \qquad \text{and} \qquad ((FG)')^{ij} := (F')^{ij} G^i + F^i (G')^{ij}$$

for each $i, j = 1, \ldots, k$, is a continuous bilinear map, and comes with the estimate

$$\|FG,(FG)'\|_{\mathcal{V}_X^{q,r}} \leq C(1 + \|X\|_p)^2 \|F,F'\|_{\mathcal{V}_X^{q,r}} \|G,G'\|_{\mathcal{V}_X^{q,r}},$$

where the constant C depends only on p, q, r and the dimension k. We call $(FG,(FG)')$ the product of (F,F') and (G,G'), which we sometimes simply denote by FG.

The proof of Lemma A.4.3 is identical to the proof of the corresponding statement for continuous paths, which can be found in [5, Lemma A.1].

Lemma A.4.4. *Let $\mathbf{X} = (X, \mathbb{X})$, $\widetilde{\mathbf{X}} = (\widetilde{X}, \widetilde{\mathbb{X}})$ be càdlàg rough paths and let (F,F'), $(G,G') \in \mathcal{V}_X^{q,r}$ and $(\widetilde{F},\widetilde{F}'),(\widetilde{G},\widetilde{G}') \in \mathcal{V}_{\widetilde{X}}^{q,r}$ be controlled paths. Let $M > 0$ be an upper bound for $\|F,F'\|_{\mathcal{V}_X^{q,r}}$, $\|G,G'\|_{\mathcal{V}_X^{q,r}}$, $\|\widetilde{F},\widetilde{F}'\|_{\mathcal{V}_{\widetilde{X}}^{q,r}}$, $\|\widetilde{G},\widetilde{G}'\|_{\mathcal{V}_{\widetilde{X}}^{q,r}}$, $\|X\|_p$ and $\|\widetilde{X}\|_p$. Then, there exists a constant C, which depends only on p, q, r and M, such that*

$$\begin{aligned}&\|FG,(FG)';\widetilde{F}\widetilde{G},(\widetilde{F}\widetilde{G})'\|_{\mathcal{V}_X^{q,r},\mathcal{V}_{\widetilde{X}}^{q,r}}\\ &\quad\leq C\Big(\|F,F';\widetilde{F},\widetilde{F}'\|_{\mathcal{V}_X^{q,r},\mathcal{V}_{\widetilde{X}}^{q,r}} + \|G,G';\widetilde{G},\widetilde{G}'\|_{\mathcal{V}_X^{q,r},\mathcal{V}_{\widetilde{X}}^{q,r}} + \|X - \widetilde{X}\|_p\Big).\end{aligned}$$

Proof. For each $i, j = 1, \ldots, d$, we have that

$$\begin{aligned}|(FG)_0^i - (\widetilde{F}\widetilde{G})_0^i| &= |(F_0^i - \widetilde{F}_0^i)G_0^i + \widetilde{F}_0^i(G_0^i - \widetilde{G}_0^i)|\\ &\leq \|G,G'\|_{\mathcal{V}_X^{q,r}} \|F,F';\widetilde{F},\widetilde{F}'\|_{\mathcal{V}_X^{q,r},\mathcal{V}_{\widetilde{X}}^{q,r}} + \|\widetilde{F},\widetilde{F}'\|_{\mathcal{V}_{\widetilde{X}}^{q,r}} \|G,G';\widetilde{G},\widetilde{G}'\|_{\mathcal{V}_X^{q,r},\mathcal{V}_{\widetilde{X}}^{q,r}}\end{aligned}$$

and

$$\begin{aligned}&|((FG)')_0^{ij} - ((\widetilde{F}\widetilde{G})')_0^{ij}| \leq |(F')_0^{ij} G_0^i - (\widetilde{F}')_0^{ij} \widetilde{G}_0^i| + |F_0^i (G')_0^{ij} - \widetilde{F}_0^i (\widetilde{G}')_0^{ij}|\\ &\quad\leq |(F')_0^{ij} - (\widetilde{F}')_0^{ij}||G_0^i| + |(\widetilde{F}')_0^{ij}||G_0^i - \widetilde{G}_0^i| + |F_0^i - \widetilde{F}_0^i||(G')_0^{ij}| + |\widetilde{F}_0^i||(G')_0^{ij} - (\widetilde{G}')_0^{ij}|\\ &\quad\leq \|G,G'\|_{\mathcal{V}_X^{q,r}} \|F,F';\widetilde{F},\widetilde{F}'\|_{\mathcal{V}_X^{q,r},\mathcal{V}_{\widetilde{X}}^{q,r}} + \|\widetilde{F},\widetilde{F}'\|_{\mathcal{V}_{\widetilde{X}}^{q,r}} \|G,G';\widetilde{G},\widetilde{G}'\|_{\mathcal{V}_X^{q,r},\mathcal{V}_{\widetilde{X}}^{q,r}}.\end{aligned}$$

Further, we have that

$$
\begin{aligned}
&\|(FG)' - (\widetilde{F}\widetilde{G})'\|_q \\
&\leq \|F' - \widetilde{F}'\|_q \|G\|_\infty + \|\widetilde{F}'\|_q \|G - \widetilde{G}\|_\infty + \|F' - \widetilde{F}'\|_\infty \|G\|_q + \|\widetilde{F}'\|_\infty \|G - \widetilde{G}\|_q \\
&\quad + \|F - \widetilde{F}\|_q \|G'\|_\infty + \|\widetilde{F}\|_q \|G' - \widetilde{G}'\|_\infty + \|F - \widetilde{F}\|_\infty \|G'\|_q + \|\widetilde{F}\|_\infty \|G' - \widetilde{G}'\|_q \\
&\leq (\|F - \widetilde{F}\|_\infty + \|F - \widetilde{F}\|_q + \|F' - \widetilde{F}'\|_\infty + \|F' - \widetilde{F}'\|_q)(\|G\|_\infty + \|G\|_q + \|G'\|_\infty + \|G'\|_q) \\
&\quad + (\|\widetilde{F}\|_\infty + \|\widetilde{F}\|_q + \|\widetilde{F}'\|_\infty + \|\widetilde{F}'\|_q)(\|G - \widetilde{G}\|_\infty + \|G - \widetilde{G}\|_q + \|G' - \widetilde{G}'\|_\infty + \|G' - \widetilde{G}'\|_q) \\
&\lesssim (1 + \|X\|_p)(1 + \|\widetilde{X}\|_p)(1 + \|F, F'\|_{\mathcal{V}^{q,r}_X}) \|G, G'\|_{\mathcal{V}^{q,r}_X} (\|F, F'; \widetilde{F}, \widetilde{F}'\|_{\mathcal{V}^{q,r}_X, \mathcal{V}^{q,r}_{\widetilde{X}}} + \|X - \widetilde{X}\|_p) \\
&\quad + (1 + \|X\|_p)(1 + \|\widetilde{X}\|_p) \|\widetilde{F}, \widetilde{F}'\|_{\mathcal{V}^{q,r}_{\widetilde{X}}} (1 + \|\widetilde{G}, \widetilde{G}'\|_{\mathcal{V}^{q,r}_{\widetilde{X}}})(\|G, G'; \widetilde{G}, \widetilde{G}'\|_{\mathcal{V}^{q,r}_X, \mathcal{V}^{q,r}_{\widetilde{X}}} + \|X - \widetilde{X}\|_p).
\end{aligned}
$$

The remainder is given by $(R^{FG})^i_{s,t} = (R^F)^i_{s,t} G^i_s + F^i_s (R^G)^i_{s,t} + F^i_{s,t} G^i_{s,t}$ for each $(s,t) \in \Delta_T$ (see the proof of [5, Lemma A.1]). Using the fact that $2r \geq p$, we have

$$
\begin{aligned}
&\|R^{FG}\|_r \\
&\leq \|R^F - R^{\widetilde{F}}\|_r \|G\|_\infty + \|R^{\widetilde{F}}\|_r \|G - \widetilde{G}\|_\infty + \|F - \widetilde{F}\|_\infty \|R^G\|_r + \|\widetilde{F}\|_\infty \|R^G - R^{\widetilde{G}}\|_r \\
&\quad + \|F - \widetilde{F}\|_{2r} \|G\|_{2r} + \|\widetilde{F}\|_{2r} \|G - \widetilde{G}\|_{2r} \\
&\lesssim (1 + \|X\|_p) \|F, F'; \widetilde{F}, \widetilde{F}'\|_{\mathcal{V}^{q,r}_X, \mathcal{V}^{q,r}_{\widetilde{X}}} \|G, G'\|_{\mathcal{V}^{q,r}_X} \\
&\quad + (1 + \|X\|_p) \|\widetilde{F}, \widetilde{F}'\|_{\mathcal{V}^{q,r}_{\widetilde{X}}} (1 + \|\widetilde{G}, \widetilde{G}'\|_{\mathcal{V}^{q,r}_{\widetilde{X}}})(\|G, G'; \widetilde{G}, \widetilde{G}'\|_{\mathcal{V}^{q,r}_X, \mathcal{V}^{q,r}_{\widetilde{X}}} + \|X - \widetilde{X}\|_p) \\
&\quad + (1 + \|\widetilde{X}\|_p)(1 + \|F, F'\|_{\mathcal{V}^{q,r}_X}) \|G, G'\|_{\mathcal{V}^{q,r}_X} (\|F, F'; \widetilde{F}, \widetilde{F}'\|_{\mathcal{V}^{q,r}_X, \mathcal{V}^{q,r}_{\widetilde{X}}} + \|X - \widetilde{X}\|_p) \\
&\quad + (1 + \|\widetilde{X}\|_p) \|\widetilde{F}, \widetilde{F}'\|_{\mathcal{V}^{q,r}_X} \|G, G'; \widetilde{G}, \widetilde{G}'\|_{\mathcal{V}^{q,r}_X, \mathcal{V}^{q,r}_{\widetilde{X}}} \\
&\quad + (1 + \|X\|_p)(1 + \|\widetilde{X}\|_p)(1 + \|F, F'\|_{\mathcal{V}^{q,r}_X}) \|G, G'\|_{\mathcal{V}^{q,r}_X} (\|F, F'; \widetilde{F}, \widetilde{F}'\|_{\mathcal{V}^{q,r}_X, \mathcal{V}^{q,r}_{\widetilde{X}}} + \|X - \widetilde{X}\|_p) \\
&\quad + (1 + \|X\|_p)(1 + \|\widetilde{X}\|_p) \|\widetilde{F}, \widetilde{F}'\|_{\mathcal{V}^{q,r}_{\widetilde{X}}} (1 + \|\widetilde{G}, \widetilde{G}'\|_{\mathcal{V}^{q,r}_{\widetilde{X}}})(\|G, G'; \widetilde{G}, \widetilde{G}'\|_{\mathcal{V}^{q,r}_X, \mathcal{V}^{q,r}_{\widetilde{X}}} + \|X - \widetilde{X}\|_p) \\
&\lesssim (1 + \|X\|_p)(1 + \|\widetilde{X}\|_p)(1 + \|F, F'\|_{\mathcal{V}^{q,r}_X})(1 + \|G, G'\|_{\mathcal{V}^{q,r}_X})(1 + \|\widetilde{F}, \widetilde{F}'\|_{\mathcal{V}^{q,r}_{\widetilde{X}}}) \\
&\qquad \times (1 + \|\widetilde{G}, \widetilde{G}'\|_{\mathcal{V}^{q,r}_{\widetilde{X}}})(\|F, F'; \widetilde{F}, \widetilde{F}'\|_{\mathcal{V}^{q,r}_X, \mathcal{V}^{q,r}_{\widetilde{X}}} + \|G, G'; \widetilde{G}, \widetilde{G}'\|_{\mathcal{V}^{q,r}_X, \mathcal{V}^{q,r}_{\widetilde{X}}} + \|X - \widetilde{X}\|_p).
\end{aligned}
$$

Combining the inequalities above, we deduce the desired estimate. □

A.4.3 Associativity of rough integration

The following proposition establishes the associativity of rough integration with respect to càdlàg controlled paths.

Proposition A.4.5. *Let $\mathbf{X} = (X, \mathbb{X})$ be a càdlàg rough path and let $(Y, Y'), (F, F'), (G, G') \in \mathcal{V}^{q,r}_X$ be controlled paths. Then $(Z, Z') := (\int_0^\cdot F_u \, \mathrm{d}G_u, FG') \in \mathcal{V}^{q,r}_X$, and we have that*

$$\int_0^\cdot Y_u \, \mathrm{d}Z_u = \int_0^\cdot Y_u F_u \, \mathrm{d}G_u,$$

where on the left-hand side we have the integral of (Y, Y') *against* (Z, Z')*, and on the right-hand side we have the integral of* $(YF, (YF)')$ *against* (G, G')*, each defined in the sense of* (A.3)*.*

The proof of Proposition A.4.5 is identical to the proof of the corresponding statement for continuous paths, which can be found in [5, Proposition A.2].

A.4.4 The canonical rough path lift of a controlled path

The next lemma provides the canonical construction of a càdlàg rough path above a controlled path.

Lemma A.4.6. *Let* $\mathbf{X} = (X, \mathbb{X})$ *be a càdlàg rough path and* $(Z, Z') \in \mathcal{V}_X^{q,r}$ *be a controlled path. Then,* $\mathbf{Z} = (Z, \mathbb{Z})$ *is a càdlàg rough path, where*

$$\mathbb{Z}_{s,t} := \int_s^t Z_u \,\mathrm{d}Z_u - Z_s \otimes Z_{s,t}, \qquad (s,t) \in \Delta_T,$$

with the integral defined as in (A.3)*. We call* $\mathbf{Z} = (Z, \mathbb{Z})$ *the canonical rough path lift of* (Z, Z')*. Moreover, if* $(Y, Y') \in \mathcal{V}_Z^{q,r}$*, then* $(Y, Y'Z') \in \mathcal{V}_X^{q,r}$*, and*

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

where on the left-hand side we have the rough integral of (Y, Y') *against* $\mathbf{Z}$*, and on the right-hand side we have the integral of* $(Y, Y'Z')$ *against* (Z, Z') *in the sense of* (A.3)*.*

The proof of Lemma A.4.6 follows the proof of the corresponding statement for continuous paths verbatim; see [5, Lemma A.3].

A.4.5 The exponential of a rough path

Recall that, given a càdlàg rough path $\mathbf{X} = (X, \mathbb{X})$, one can define the so-called reduced rough path $\mathbf{X}^r = (X, [\mathbf{X}])$, where $[\mathbf{X}]_t := X_{0,t} \otimes X_{0,t} - 2\mathrm{Sym}(\mathbb{X}_{0,t})$ is the rough path bracket of $\mathbf{X}$; see, e.g., [75, Section 2.4]. If X satisfies Property (RIE) relative to p and a sequence of partitions $(\mathcal{P}^n)_{n \in \mathbb{N}}$, then, by [7, Proposition 2.18], one can see that the rough path bracket $[\mathbf{X}]$ coincides with the pathwise quadratic variation $[X]$ of X, in the sense of Föllmer; see [67]. Using this notion, one can introduce the rough exponential analogously to the stochastic exponential of Itô calculus.

In the following, given a path X, we will write $\Delta X_t := X_{t-,t}$ for the jump of X at time t.

Lemma A.4.7. *Given a one-dimensional càdlàg rough path* $\mathbf{X} = (X, \mathbb{X})$ *(so that in particular* X *is real-valued), such that* $X_0 = 0$, $\Delta[\mathbf{X}]_t = (\Delta X_t)^2$ *for every* $t \in [0,T]$, *and* $\sum_{t\in[0,T]}(\Delta X_t)^2 < \infty$, *the rough exponential* $V = \mathcal{E}(X)$ *is defined by*

$$V_t := \exp\Big(X_t - \frac{1}{2}\Gamma_t\Big) \prod_{0<s\leq t} (1 + \Delta X_s)\exp(-\Delta X_s), \qquad t \in [0,T],$$

where $\Gamma_t := [\mathbf{X}]_t - \sum_{s\leq t}(\Delta X_s)^2$ *for* $t \in [0,T]$. *We then have that* V *is the unique controlled path in* $\mathcal{V}^{q,r}_X$ *satisfying the linear rough differential equation*

$$V_t = 1 + \int_0^t V_s \,\mathrm{d}\mathbf{X}_s, \qquad t \in [0,T], \tag{A.4}$$

with Gubinelli derivative $V' = V$.

Proof. Since we assume that $\sum_{t\in[0,T]}(\Delta X_t)^2 < \infty$, and $\Delta[\mathbf{X}]_t = (\Delta X_t)^2$ for all $t \in [0,T]$, the path $\Gamma = [\mathbf{X}] - \sum_{s\leq\cdot}(\Delta X_s)^2$ is continuous and has finite $\frac{p}{2}$-variation. Let $Y := X - \frac{1}{2}\Gamma$ and $A := \prod_{s\leq\cdot}(1+\Delta X_s)\exp(-\Delta X_s)$. One can verify that A is of finite 1-variation; see, e.g., the proof of [147, Chapter II, Theorem 37]. Hence, the two-dimensional path $Z := (Y, A)$ admits a rough path lift $\mathbf{Z} = (Z, \mathbb{Z})$, such that

$$\mathbb{Z}^{1,1}_{s,t} = \mathbb{X}_{s,t} - \frac{1}{2}\int_s^t X_{s,u}\,\mathrm{d}\Gamma_u - \frac{1}{2}\int_s^t \Gamma_{s,u}\,\mathrm{d}X_u + \frac{1}{4}\int_s^t \Gamma_{s,u}\,\mathrm{d}\Gamma_u,$$
$$\mathbb{Z}^{1,2}_{s,t} = \int_s^t Y_{s,u}\,\mathrm{d}A_u, \qquad \mathbb{Z}^{2,1}_{s,t} = \int_s^t A_{s,u}\,\mathrm{d}Y_u, \qquad \mathbb{Z}^{2,2}_{s,t} = \int_s^t A_{s,u}\,\mathrm{d}A_u,$$

for $(s,t) \in \Delta_T$, where all the integrals above are interpreted as Young integrals (as in, e.g., [75, Proposition 2.4]).

We now consider the reduced rough path $(Z, [\mathbf{Z}])$ associated with $\mathbf{Z}$. By definition, we have that

$$[Y, A]_t := [\mathbf{Z}]^{1,2}_t = [\mathbf{Z}]^{2,1}_t = Y_{0,t}A_{0,t} - \bigg(\int_0^t Y_{0,u}\,\mathrm{d}A_u + \int_0^t A_{0,u}\,\mathrm{d}Y_u\bigg).$$

Since $\int_0^t Y_{0,u}\,\mathrm{d}A_u$ and $\int_0^t A_{0,u}\,\mathrm{d}Y_u$ are Young integrals, for any sequence of partitions $(\mathcal{P}^n)_{n\in\mathbb{N}}$ of $[0,T]$ with vanishing mesh size, we have that

$$\int_0^t Y_{0,u}\,\mathrm{d}A_u = \lim_{n\to\infty}\sum_{[u,v]\in\mathcal{P}^n} Y_{0,u}A_{u\wedge t,v\wedge t}, \qquad \int_0^t A_{0,u}\,\mathrm{d}Y_u = \lim_{n\to\infty}\sum_{[u,v]\in\mathcal{P}^n} A_{0,u}Y_{u\wedge t,v\wedge t}.$$

Noting that

$$Y_{0,v\wedge t}A_{0,v\wedge t} - Y_{0,u\wedge t}A_{0,u\wedge t} = Y_{0,u\wedge t}A_{u\wedge t,v\wedge t} + A_{0,u\wedge t}Y_{u\wedge t,v\wedge t} + Y_{u\wedge t,v\wedge t}A_{u\wedge t,v\wedge t}$$

and taking $\lim_{n\to\infty}\sum_{[u,v]\in\mathcal{P}^n}$ on each side, we obtain

$$[Y, A]_t = \lim_{n\to\infty}\sum_{[u,v]\in\mathcal{P}^n} Y_{u\wedge t,v\wedge t}A_{u\wedge t,v\wedge t} = \sum_{s\leq t}\Delta Y_s\Delta A_s = \sum_{s\leq t}\Delta X_s\Delta A_s,$$

and one can similarly show that $[A]_t := [\mathbf{Z}]_t^{2,2} = \sum_{s \le t} (\Delta A_s)^2$.

Since Γ is continuous and of finite $\frac{p}{2}$-variation, one can show, using the integration by parts formula for Young integrals, that $[Y]_t := [\mathbf{Z}]_t^{1,1} = [\mathbf{X}]_t$, so that $[Y]_t = \Gamma_t + \sum_{s \le t} (\Delta X_s)^2$.

Applying the Itô formula for rough paths ([75, Theorem 2.12]) to $V_t = f(Z_t)$, where $f(y, a) := a \exp(y)$, and using the expressions derived above for the rough path bracket $[\mathbf{Z}]$, a straightforward calculation (similar to the proof of [147, Chapter II, Theorem 37] in the semimartingale setting) establishes that $V_t = 1 + \int_0^t V_s \, \mathrm{d}\mathbf{X}_s$. In particular, this involves noting that $\int_0^t V_s \, \mathrm{d}Y_s = \int_0^t V_s \, \mathrm{d}\mathbf{X}_s - \frac{1}{2} \int_0^t V_s \, \mathrm{d}\Gamma_s$, where in the first integral on the right-hand side we identify (V, V) as a controlled path with respect to X.

Finally, the uniqueness of solutions to (A.4) follows from straightforward estimates using the stability of rough integration ([75, Lemma 3.4]). $\square$

A.5 Proof of Lemma 6.1.7

Proof of Lemma 6.1.7. For $\gamma \neq \frac{1}{2}$, the statement follows from Lemma 6.1.6 and Proposition 3.1.10.

Suppose that $\gamma = \frac{1}{2}$. We need to verify that the integral

$$\int_0^t \widehat{X}_r \otimes \mathrm{d}^{\gamma,\pi^n} \widehat{X}_r = \int_0^t X_r \otimes \mathrm{d}^{\gamma,\pi^n} X_r + \int_0^t X_r \otimes \mathrm{d}^{\gamma,\pi^n} \varphi_r + \int_0^t \varphi_r \otimes \mathrm{d}^{\gamma,\pi^n} X_r + \int_0^t \varphi_r \otimes \mathrm{d}^{\gamma,\pi^n} \varphi_r,$$

converges as $n \to \infty$ to the limit

$$\int_0^t \widehat{X}_r \otimes \mathrm{d}^{\gamma,\pi} X_r = \int_0^t X_r \otimes \mathrm{d}^{\gamma,\pi} X_r + \int_0^t X_r \otimes \mathrm{d}^{\gamma,\pi} \varphi_r + \int_0^t \varphi_r \otimes \mathrm{d}^{\gamma,\pi} X_r + \int_0^t \varphi_r \otimes \mathrm{d}^{\gamma,\pi} \varphi_r,$$

uniformly in $t \in [0,T]$, where the latter three integrals are defined as Young integrals.

Since X satisfies Property γ-(RIE), we have that

$$\left\| \int_0^\cdot X_r \otimes \mathrm{d}^{\gamma,\pi^n} X_r - \int_0^\cdot X_r \otimes \mathrm{d}^{\gamma,\pi} X_r \right\|_\infty \longrightarrow 0 \qquad \text{as} \qquad n \to \infty.$$

Define $\bar{X}^n$ and $\bar{\varphi}^n$ as the piecewise linear interpolation of X and φ, respectively, along $\pi = (\pi^n)_{n \in \mathbb{N}}$. Then it holds for any $t \in [0,T]$ that

$$\int_0^t X_r \otimes \mathrm{d}^{\gamma,\pi_n} \varphi_r = \sum_{k=0}^{N_n - 1} (X_{t_k^n} + \frac{1}{2} X_{t_k^n, t_{k+1}^n}) \otimes \varphi_{t_k^n \wedge t, t_{k+1}^n \wedge t} = \int_0^t \bar{X}_r^n \otimes \mathrm{d}\varphi_r.$$

Let $p' > p$ such that $\frac{1}{p'} + \frac{1}{q} > 1$. By the standard estimate for Young integrals – see e.g. [75, Proposition 2.4] – we have for all $t \in [0,T]$, that

$$\left| \int_0^t X_r \otimes \mathrm{d}^{\gamma,\pi^n} \varphi_r - \int_0^t X_r \otimes \mathrm{d}^{\gamma,\pi} \varphi_r \right| \lesssim \|\bar{X}^n - X\|_{p'} \|\varphi\|_q.$$

It follows by interpolation—see e.g. [74, Proposition 5.5]—that

$$\|\bar{X}^n - X\|_{p'} \leq \|\bar{X}^n - X\|_\infty^{1 - \frac{p}{p'}} \|\bar{X}^n - X\|_p^{\frac{p}{p'}}.$$

Since $\bar{X}^n$ converges uniformly to X as $n \to \infty$, and $\sup_{n \in \mathbb{N}} \|\bar{X}^n\|_p < \infty$, we deduce that

$$\left\| \int_0^\cdot X_u \otimes \mathrm{d}^{\gamma,\pi^n} \varphi_u - \int_0^\cdot X_u \otimes \mathrm{d}^{\gamma,\pi} \varphi_u \right\|_\infty \longrightarrow 0 \qquad \text{as} \quad n \longrightarrow \infty.$$

Similarly, for each $t \in [0,T]$, it holds that

$$\left| \int_0^t \varphi_r \otimes \mathrm{d}^{\gamma,\pi^n} X_r - \int_0^t \varphi_r \otimes \mathrm{d}^{\gamma,\pi} X_r \right| \lesssim \|\bar{\varphi}^n - \varphi\|_q \|X\|_p,$$

and

$$\left| \int_0^t \varphi_r \otimes \mathrm{d}^{\gamma,\pi^n} \varphi_r - \int_0^t \varphi_r \otimes \mathrm{d}^{\gamma,\pi} \varphi_r \right| \lesssim \|\bar{\varphi}^n - \varphi\|_q \|\varphi\|_q,$$

and, since $\|\bar{\varphi}^n - \varphi\|_q \to 0$ as $n \to \infty$, we infer the required convergence.

We further aim to find a control function c such that

$$\sup_{(s,t)\in\Delta_T} \frac{|\widehat{X}_{s,t}|^p}{c(s,t)} + \sup_{n\in\mathbb{N}} \sup_{0\le k<\ell\le N_n} \frac{|\int_{t_k^n}^{t_\ell^n} \widehat{X}_u \otimes \mathrm{d}^{\gamma,\pi^n} \widehat{X}_u - \widehat{X}_{t_k^n} \otimes \widehat{X}_{t_k^n,t_{k+1}^n}|^{\frac{p}{2}}}{c(t_k^n, t_\ell^n)} \lesssim 1, \tag{A.5}$$

where

$$\begin{aligned}\int_{t_k^n}^{t_\ell^n} \widehat{X}_u \otimes \mathrm{d}^{\gamma,\pi^n} \widehat{X}_u - \widehat{X}_{t_k^n} \otimes \widehat{X}_{t_k^n,t_{k+1}^n} &= \int_{t_k^n}^{t_\ell^n} \widehat{X}_{t_k^n,u} \otimes \mathrm{d}^{\gamma,\pi^n} \widehat{X}_u \\ &= \int_{t_k^n}^{t_\ell^n} X_{t_k^n,u} \otimes \mathrm{d}^{\gamma,\pi^n} X_u + \int_{t_k^n}^{t_\ell^n} X_{t_k^n,u} \otimes \mathrm{d}^{\gamma,\pi^n} \varphi_u \\ &\quad + \int_{t_k^n}^{t_\ell^n} \varphi_{t_k^n,u} \otimes \mathrm{d}^{\gamma,\pi^n} X_u + \int_{t_k^n}^{t_\ell^n} \varphi_{t_k^n,u} \otimes \mathrm{d}^{\gamma,\pi^n} \varphi_u.\end{aligned}$$

Let c_X be the control function with respect to which X satisfies Property γ-(RIE), and define moreover the control function c_φ, given by $c_\varphi(s,t) = \|\varphi\|_{q,[s,t]}^q$ for $(s,t) \in \Delta_T$.

We have from Property γ-(RIE) that

$$\sup_{(s,t)\in\Delta_T} \frac{|\widehat{X}_{s,t}|^p}{c_X(s,t) + c_\varphi(s,t)} \lesssim \sup_{(s,t)\in\Delta_T} \frac{|X_{s,t}|^p}{c_X(s,t)} + \sup_{(s,t)\in\Delta_T} \frac{|\varphi_{s,t}|^p}{c_\varphi(s,t)} \lesssim 1,$$

and that

$$\sup_{n\in\mathbb{N}} \sup_{0\le k<\ell\le N_n} \frac{|\int_{t_k^n}^{t_\ell^n} X_u \otimes \mathrm{d}^{\gamma,\pi^n} X_u - X_{t_k^n} \otimes X_{t_k^n,t_{k+1}^n}|^{\frac{p}{2}}}{c_X(t_k^n, t_\ell^n)} \lesssim 1.$$

By the standard estimate for Young integrals (see e.g. [75, Proposition 2.4]), for every $n \in \mathbb{N}$ and $0 \le k < \ell \le N_n$, we have

$$\begin{aligned}\left| \int_{t_k^n}^{t_\ell^n} \bar{X}_{t_k^n,u}^n \otimes \mathrm{d}\varphi_u \right|^{\frac{p}{2}} &\lesssim \|\bar{X}^n\|_{p,[t_k^n,t_\ell^n]}^{\frac{p}{2}} \|\varphi\|_{q,[t_k^n,t_\ell^n]}^{\frac{p}{2}} \\ &\le \|X\|_{p,[t_k^n,t_\ell^n]}^{\frac{p}{2}} \|\varphi\|_{q,[t_k^n,t_\ell^n]}^{\frac{p}{2}} \le c_X(t_k^n,t_\ell^n)^{\frac{1}{2}} c_\varphi(t_k^n,t_\ell^n)^{\frac{p}{2q}},\end{aligned}$$

and we can similarly obtain

$$\left| \int_{t_k^n}^{t_\ell^n} \bar{\varphi}_{t_k^n,u}^n \otimes \mathrm{d}X_u \right|^{\frac{p}{2}} \lesssim c_X(t_k^n,t_\ell^n)^{\frac{1}{2}} c_\varphi(t_k^n,t_\ell^n)^{\frac{p}{2q}}$$

and

$$\left| \int_{t_k^n}^{t_\ell^n} \bar{\varphi}_{t_k^n,u}^n \otimes \mathrm{d}\varphi_u \right|^{\frac{p}{2}} \lesssim c_\varphi(t_k^n,t_\ell^n)^{\frac{p}{q}}.$$

Since $p \in (2,3)$ and $q \in [1,2)$, we have that $\frac{1}{2} + \frac{p}{2q} > 1$ and $\frac{p}{q} > 1$, and it follows that the maps $(s,t) \mapsto c_X(s,t)^{\frac{1}{2}} c_\varphi(s,t)^{\frac{p}{2q}}$ and $(s,t) \mapsto c_\varphi(s,t)^{\frac{p}{q}}$ are superadditive and thus control functions. We deduce that (A.5) holds with a control function c of the form

$$c(s,t) = C\Big(c_X(s,t) + c_\varphi(s,t) + c_X(s,t)^{\frac{1}{2}} c_\varphi(s,t)^{\frac{p}{2q}} + c_\varphi(s,t)^{\frac{p}{q}}\Big), \qquad (s,t) \in \Delta_T,$$

where $C > 0$ is a suitable constant which depends only on p and q. $\square$

Bibliography

[1] Shigeki Aida. Reflected rough differential equations. *Stochastic Process. Appl.*, 125(9):3570–3595, 2015.

[2] Shigeki Aida. Rough differential equations containing path-dependent bounded variation terms. *J. Theoret. Probab.*, 37(3):2130–2183, 2024.

[3] Andrew L. Allan. Robust filtering and propagation of uncertainty in hidden Markov models. *Electron. J. Probab.*, 26:Paper No. 73, 37, 2021.

[4] Andrew L. Allan and Samuel N. Cohen. Pathwise stochastic control with applications to robust filtering. *Ann. Appl. Probab.*, 30(5):2274–2310, 2020.

[5] Andrew L. Allan, Christa Cuchiero, Chong Liu, and David J. Prömel. Model-free portfolio theory: a rough path approach. *Math. Finance*, 33(3):709–765, 2023.

[6] Andrew L. Allan, Anna P. Kwossek, Chong Liu, and David J. Prömel. Pathwise convergence of the Euler scheme for rough and stochastic differential equations. *arXiv preprint arXiv:2309.16489*, 2023.

[7] Andrew L. Allan, Chong Liu, and David J. Prömel. A càdlàg rough path foundation for robust finance. *Finance Stoch.*, 28(1):215–257, 2024.

[8] Anna Ananova. Rough differential equations with path-dependent coefficients. *Ann. H. Lebesgue*, 6:1–29, 2023.

[9] Anna Ananova and Rama Cont. Pathwise integration with respect to paths of finite quadratic variation. *J. Math. Pures Appl. (9)*, 107(6):737–757, 2017.

[10] David Applebaum. *Lévy processes and stochastic calculus*, volume 116 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2009.

[11] Imanol Perez Arribas. Derivatives pricing using signature payoffs. *arXiv preprint arXiv:1809.09466*, 2018.

[12] Imanol Perez Arribas, Cristopher Salvi, and Lukasz Szpruch. Sig-sdes model for quantitative finance. In *Proceedings of the First ACM International Conference on AI in Finance*, ICAIF '20, New York, NY, USA, 2021. Association for Computing Machinery.

[13] Marco Avellaneda, Arnon Levy, and Antonio Parás. Pricing and hedging derivative securities in markets with uncertain volatilities. *Appl. Math. Finance*, 2(2):73–88, 1995.

[14] Ismaël Bailleul. Flows driven by Banach space-valued rough paths. In *Séminaire de Probabilités XLVI*, volume 2123 of *Lecture Notes in Math.*, pages 195–205. Springer, Cham, 2014.

[15] Ismaël Bailleul, Antoine Brault, and Laure Coutin. Young and rough differential inclusions. *Rev. Mat. Iberoam.*, 37(4):1489–1512, 2021.

[16] D. R. Baños, G. Di Nunno, H. H. Haferkorn, and F. Proske. Stochastic functional differential equations and sensitivity to their initial path. In *Computation and combinatorics in dynamics, stochastics and control*, volume 13 of *Abel Symp.*, pages 37–70. Springer, Cham, 2018.

[17] Daniel Bartl, Michael Kupper, and Ariel Neufeld. Stochastic integration and differential equations for typical paths. *Electron. J. Probab.*, 24(97):1–21, 2019.

[18] Daniel Bartl, Michael Kupper, David J. Prömel, and Ludovic Tangpi. Duality for pathwise superhedging in continuous time. *Finance Stoch.*, 23(3):697–728, 2019.

[19] Christian Bayer, Paul P. Hager, Sebastian Riedel, and John Schoenmakers. Optimal stopping with signatures. *Ann. Appl. Probab.*, 33(1):238–273, 2023.

[20] Mireia Besalú, Giulia Binotto, and Carles Rovira. Convergence of delay equations driven by a Hölder continuous function of order $1/3 < \beta < 1/2$. *Electron. J. Differ. Equ.*, pages Paper No. 65, 27, 2020.

[21] Mireia Besalú, David Márquez-Carreras, and Carles Rovira. Delay equations with non-negativity constraints driven by a Hölder continuous function of order $\beta \in (\frac{1}{3}, \frac{1}{2})$. *Potential Anal.*, 41(1):117–141, 2014.

[22] Sara Biagini and Mustafa Ç. Pınar. The robust Merton problem of an ambiguity averse investor. *Math. Financ. Econ.*, 11(1):1–24, 2017.

[23] Klaus Bichteler. Stochastic integration and L^p-theory of semimartingales. *Ann. Probab.*, 9(1):49–89, 1981.

[24] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *J. Polit. Econ.*, 81(3):637–654, 1973.

[25] Horatio Boedihardjo, Xi Geng, Terry Lyons, and Danyu Yang. The signature of a rough path: uniqueness. *Adv. Math.*, 293:720–737, 2016.

[26] Jean Bretagnolle. p-variation de fonctions aléatoires. II. Processus à accroissements indépendants. In *Séminaire de Probabilités, VI (Univ. Strasbourg, année universitaire 1970–1971; Journées Probabilistes de Strasbourg, 1971)*, volume Vol. 258 of *Lecture Notes in Math.*, pages 64–71. Springer, Berlin-New York, 1972.

[27] Michael Caruana, Peter K. Friz, and Harald Oberhauser. A (rough) pathwise approach to a class of non-linear stochastic partial differential equations. *Ann. Inst. H. Poincaré C Anal. Non Linéaire*, 28(1):27–46, 2011.

[28] Kuo-Tsai Chen. Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula. *Ann. of Math. (2)*, 65:163–178, 1957.

[29] Kuo Tsai Chen. Iterated path integrals. *Bull. Amer. Math. Soc.*, 83(5):831–879, 1977.

[30] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 6572–6583, 2018.

[31] Ilya Chevyrev. Random walks and Lévy processes as rough paths. *Probab. Theory Related Fields*, 170(3-4):891–932, 2018.

[32] Reda Chhaibi and Ibrahim Ekren. A Hörmander condition for delayed stochastic differential equations. *Ann. H. Lebesgue*, 3:1023–1048, 2020.

[33] Henry Chiu and Rama Cont. On pathwise quadratic variation for càdlàg functions. *Electron. Commun. Probab.*, 23:Paper No. 85, 12, 2018.

[34] Henry Chiu and Rama Cont. Causal functional calculus. *Trans. London Math. Soc.*, 9(1):237–269, 2022.

[35] Vedant Choudhary, Sebastian Jaimungal, and Maxime Bergeron. FuNVol: multi-asset implied volatility market simulator using functional principal components and neural SDEs. *Quant. Finance*, 24(8):1077–1103, 2024.

[36] Samuel N. Cohen, Christoph Reisinger, and Sheng Wang. Hedging option books using neural-SDE market models. *Appl. Math. Finance*, 29(5):366–401, 2022.

[37] Samuel N. Cohen, Christoph Reisinger, and Sheng Wang. Arbitrage-free neural-SDE market models. *Appl. Math. Finance*, 30(1):1–46, 2023.

[38] Rama Cont and Purba Das. Quadratic variation and quadratic roughness. *Bernoulli*, 29(1):496–522, 2023.

[39] Rama Cont and David-Antoine Fournié. Change of variable formulas for non-anticipative functionals on path space. *J. Funct. Anal.*, 259(4):1043–1072, 2010.

[40] Rama Cont and Nicolas Perkowski. Pathwise integration and change of variable formulas for continuous paths with arbitrary regularity. *Trans. Amer. Math. Soc. Ser. B*, 6:161–186, 2019.

[41] Laure Coutin and Antoine Lejay. Semi-martingales and rough paths theory. *Electron. J. Probab.*, 10:no. 23, 761–785, 2005.

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

[43] D. Crisan, J. Diehl, P. K. Friz, and H. Oberhauser. Robust filtering: correlated noise and multidimensional observation. *Ann. Appl. Probab.*, 23(5):2139–2160, 2013.

[44] Christa Cuchiero, Guido Gazzani, Janka Möller, and Sara Svaluto-Ferro. Joint calibration to SPX and VIX options with signature-based models. *Math. Finance*, 35(1):161–213, 2025.

[45] Christa Cuchiero, Guido Gazzani, and Sara Svaluto-Ferro. Signature-based models: theory and calibration. *SIAM J. Financial Math.*, 14(3):910–957, 2023.

[46] Christa Cuchiero, Wahid Khosrawi, and Josef Teichmann. A generative adversarial network approach to calibration of local stochastic volatility models. *Risks*, 8(4), 2020.

[47] Christa Cuchiero, Francesca Primavera, and Sara Svaluto-Ferro. Universal approximation theorems for continuous functions of càdlàg paths and Lévy-type signature models. *arXiv preprint arXiv:2208.02293*, 2023.

[48] Christa Cuchiero, Walter Schachermayer, and Ting-Kam Leonard Wong. Cover's universal portfolio, stochastic portfolio theory, and the numéraire portfolio. *Math. Finance*, 29(3):773–803, 2019.

[49] Christa Cuchiero, Philipp Schmocker, and Josef Teichmann. Global universal approximation of functional input maps on weighted spaces. *arXiv preprint arXiv:2306.03303*, 2024.

[50] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems*, 2(4):303–314, 1989.

[51] A. M. Davie. Differential equations driven by rough paths: an approach via discrete approximation. *Appl. Math. Res. Express. AMRX*, pages Art. ID abm009, 40, 2008. [Issue information previously given as no. 2 (2007)].

[52] Mark Davis, Jan Obłój, and Vimal Raval. Arbitrage bounds for prices of weighted variance swaps. *Math. Finance*, 24(4):821–854, 2014.

[53] Steffen Dereich and Felix Heidenreich. A multilevel Monte Carlo algorithm for Lévy-driven stochastic differential equations. *Stochastic Process. Appl.*, 121(7):1565–1587, 2011.

[54] A. Deya, A. Neuenkirch, and S. Tindel. A Milstein-type scheme without Lévy area terms for SDEs driven by fractional Brownian motion. *Ann. Inst. Henri Poincaré Probab. Stat.*, 48(2):518–550, 2012.

[55] Aurélien Deya, Massimiliano Gubinelli, Martina Hofmanová, and Samy Tindel. One-dimensional reflected rough differential equations. *Stochastic Process. Appl.*, 129(9):3261–3281, 2019.

[56] Aurélien Deya and Samy Tindel. Rough Volterra equations. I. The algebraic integration setting. *Stoch. Dyn.*, 9(3):437–477, 2009.

[57] Joscha Diehl, Peter K. Friz, and Paul Gassiat. Stochastic control with rough paths. *Appl. Math. Optim.*, 75(2):285–315, 2017.

[58] Joscha Diehl, Harald Oberhauser, and Sebastian Riedel. A Lévy area between Brownian motion and rough paths with applications to robust nonlinear filtering and rough partial differential equations. *Stochastic Process. Appl.*, 125:161–181, 2015.

[59] Catherine Doléans-Dade and Paul-André Meyer. Intégrales stochastiques par rapport aux martingales locales. *C. R. Acad. Sci. Paris Sér. A-B*, 269:A144–A147, 1969.

[60] R. M. Dudley. Sample functions of the Gaussian process. *Ann. Probability*, 1(1):66–103, 1973.

[61] R.M. Dudley and R. Norvaiša. *An Introduction to p-Variation and Young Integrals: With Emphasis on Sample Functions of Stochastic Processes*, volume 1 of *MaPhySto Lecture notes*. Aarhus, Denmark, 1998.

[62] Bruno Dupire. Functional Itô calculus. *Quant. Finance*, 19(5):721–729, 2019.

[63] Weinan E. A proposal on machine learning via dynamical systems. *Commun. Math. Stat.*, 5(1):1–11, 2017.

[64] Lei Fan and Justin Sirignano. Machine Learning Methods for Pricing Financial Derivatives. *arXiv preprint arXiv:2406.00459*, 2024.

[65] Marco Ferrante and Carles Rovira. Stochastic delay differential equations driven by fractional Brownian motion with Hurst parameter $H > \frac{1}{2}$. *Bernoulli*, 12(1):85–100, 2006.

[66] Donald L. Fisk. Quasi-martingales. *Trans. Amer. Math. Soc.*, 120:369–389, 1965.

[67] H. Föllmer. Calcul d'Itô sans probabilités. In *Seminar on Probability, XV (Univ. Strasbourg, Strasbourg, 1979/1980) (French)*, volume 850 of *Lecture Notes in Math.*, pages 143–150. Springer, Berlin, 1981.

[68] Hans Föllmer and Alexander Schied. Probabilistic aspects of finance. *Bernoulli*, 19(4):1306–1326, 2013.

[69] Dana Fraňková. Regulated functions with values in Banach space. *Math. Bohem.*, 144(4):437–456, 2019.

[70] Peter Friz and Nicolas Victoir. Euler estimates for rough differential equations. *J. Differential Equations*, 244(2):388–412, 2008.

[71] Peter K. Friz and Martin Hairer. *A course on rough paths.* Universitext. Springer, Cham, second edition, 2020. With an introduction to regularity structures.

[72] Peter K. Friz, Antoine Hocquet, and Khoa Lê. Rough stochastic differential equations. *arXiv preprint arXiv:2106.10340*, 2024.

[73] Peter K. Friz and Atul Shekhar. General rough integration, Lévy rough paths and a Lévy-Kintchine-type formula. *Ann. Probab.*, 45(4):2707–2765, 2017.

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

[75] Peter K. Friz and Huilin Zhang. Differential equations driven by rough paths with jumps. *J. Differential Equations*, 264(10):6226–6301, 2018.

[76] Peter K. Friz and Pavel Zorin-Kranich. Rough semimartingales and p-variation estimates for martingale transforms. *Ann. Probab.*, 51(2):397–441, 2023.

[77] Paul Gassiat. Non-uniqueness for reflected rough differential equations. *Ann. Inst. Henri Poincaré Probab. Stat.*, 57(3):1369–1387, 2021.

[78] Patrick Gierjatowicz, Marc Sabate-Vidales, David Siska, Lukasz Szpruch, and Zan Zuric. Robust pricing and hedging via neural stochastic differential equations. *Journal of Computational Finance*, 26(3):1–32, 2022.

[79] Thomas Goll and Jan Kallsen. Optimal portfolios for logarithmic utility. *Stochastic Process. Appl.*, 89(1):31–48, 2000.

[80] Thomas Goll and Jan Kallsen. A complete explicit solution to the log-optimal portfolio problem. *Ann. Appl. Probab.*, 13(2):774–799, 2003.

[81] Paolo Guasoni and Scott Robertson. Portfolios and risk premia for the long run. *Ann. Appl. Probab.*, 22(1):239–284, 2012.

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

[83] Xin Guo, Binnan Wang, Ruixun Zhang, and Chaoyi Zhao. On Consistency of Signature Using Lasso. *arXiv preprint arXiv:2305.10413*, 2024.

[84] István Gyöngy. A note on Euler's approximations. *Potential Anal.*, 8(3):205–216, 1998.

[85] István Gyöngy and Miklós Rásonyi. A note on Euler approximations for SDEs with Hölder continuous diffusion coefficients. *Stochastic Process. Appl.*, 121(10):2189–2200, 2011.

[86] M. Hairer. Rough stochastic PDEs. *Comm. Pure Appl. Math.*, 64(11):1547–1585, 2011.

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

[88] Fabian A. Harang, Fred Espen Benth, and Fride Straum. Universal approximation on non-geometric rough paths and applications to financial derivatives pricing. *arXiv preprint arXiv:2412.16009*, 2024.

[89] David G. Hobson. Robust hedging of the lookback option. *Finance Stoch.*, 2:329–347, 1998.

[90] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.

[91] Akira Ishiwata and Setsuo Taniguchi. On the analyticity of stochastic flows. *Osaka J. Math.*, 36(1):139–148, 1999.

[92] Zacharia Issa, Blanka Horvath, Maud Lemercier, and Cristopher Salvi. Non-adversarial training of Neural SDEs with signature kernel scores. *Advances in Neural Information Processing Systems*, 36, 2024.

[93] Kiyosi Itô. Stochastic integral. *Proc. Imp. Acad. Tokyo*, 20:519–524, 1944.

[94] Kiyosi Ito. On stochastic differential equations. *Mem. Amer. Math. Soc.*, 4:51, 1951.

[95] Jean Jacod, Thomas G. Kurtz, Sylvie Méléard, and Philip Protter. The approximate Euler method for Lévy driven stochastic differential equations. *Ann. Inst. H. Poincaré Probab. Statist.*, 41(3):523–558, 2005.

[96] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.

[97] Robert Jarrow and Philip Protter. A short history of stochastic integration and mathematical finance: the early years, 1880–1970. In *A festschrift for Herman Rubin*, volume 45 of *IMS Lecture Notes Monogr. Ser.*, pages 75–91. Inst. Math. Statist., Beachwood, OH, 2004.

[98] Robert A. Jarrow. *Continuous-time asset pricing theory*. Springer Finance Textbooks. Springer, Cham, 2018. A martingale-based approach.

[99] Junteng Jia and Austin R Benson. Neural jump stochastic differential equations. *Advances in Neural Information Processing Systems*, 32, 2019.

[100] Jasdeep Kalsi, Terry Lyons, and Imanol Perez Arribas. Optimal execution with rough path signatures. *SIAM J. Financial Math.*, 11(2):470–493, 2020.

[101] H. Kaneko and S. Nakao. A note on approximation for stochastic differential equations. In *Séminaire de Probabilités, XXII*, volume 1321 of *Lecture Notes in Math.*, pages 155–162. Springer, Berlin, 1988.

[102] Rajeeva L. Karandikar. On almost sure convergence of modified Euler-Peano approximation of solution to an S.D.E. driven by a semimartingale. In *Séminaire de Probabilités, XXV*, volume 1485 of *Lecture Notes in Math.*, pages 113–120. Springer, Berlin, 1991.

[103] Ioannis Karatzas and Donghan Kim. Trading strategies generated pathwise by functions of market weights. *Finance Stoch.*, 24(2):423–463, 2020.

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

[105] Constantinos Kardaras and Gordan Žitković. Stability of the utility maximization problem with random endowment in incomplete markets. *Math. Finance*, 21(2):313–333, 2011.

[106] Patrick Kidger, James Foster, Xuechen (Chen) Li, and Terry Lyons. Efficient and accurate gradients for neural sdes. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 18747–18761. Curran Associates, Inc., 2021.

[107] Patrick Kidger and Terry Lyons. Universal Approximation with Deep Narrow Networks. In Jacob Abernethy and Shivani Agarwal, editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 2306–2327. PMLR, 09–12 Jul 2020.

[108] Tong Suk Kim and Edward Omberg. Dynamic nonmyopic portfolio behavior. *The Review of Financial Studies*, 9(1):141–161, 1996.

[109] P. E. Kloeden and A. Neuenkirch. The pathwise convergence of approximation schemes for stochastic differential equations. *LMS J. Comput. Math.*, 10:235–253, 2007.

[110] Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[111] Frank H. Knight. *Risk, uncertainty and profit.* Boston and New York, Houghton Mifflin Company, 1992.

[112] Ralf Korn. *Optimal portfolios: stochastic models for optimal investment and risk management in continuous time.* World Scientific, 1997.

[113] Holger Kraft. Optimal portfolios and Heston's stochastic volatility model: an explicit solution for power utility. *Quant. Finance*, 5(3):303–313, 2005.

[114] K. Kubilius. The existence and uniqueness of the solution of an integral equation driven by a p-semimartingale of special type. *Stochastic Process. Appl.*, 98(2):289–315, 2002.

[115] Hiroshi Kunita and Shinzo Watanabe. On square integrable martingales. *Nagoya Math. J.*, 30:209–245, 1967.

[116] Anna P. Kwossek, Andreas Neuenkirch, and David J. Prömel. Functional differential equations driven by càdlàg rough paths. *arXiv preprint arXiv:2403.17573*, 2024.

[117] Anna P. Kwossek, David J. Prömel, and Josef Teichmann. Universal approximation property of neural stochastic differential equations. *arXiv preprint arXiv:2503.16696*, 2025.

[118] Kasper Larsen and Gordan Žitković. Stability of utility-maximization in incomplete markets. *Stochastic Processes Appl.*, 117(11):1642–1662, 2007.

[119] Antoine Lejay. Global solutions to rough differential equations with unbounded vector fields. In *Séminaire de Probabilités XLIV*, volume 2046 of *Lecture Notes in Math.*, pages 215–246. Springer, Heidelberg, 2012.

[120] D. Lépingle. La variation d'ordre p des semi-martingales. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 36(4):295–316, 1976.

[121] Daniel Levin, Terry Lyons, and Hao Ni. Learning from the past, predicting the statistics for the future, learning an evolving system. *arXiv preprint arXiv:1309.0260*, 2016.

[122] Paul Lévy. Le mouvement brownien plan. *Amer. J. Math.*, 62:487–550, 1940.

[123] Xuechen Li, Ting-Kam Leonard Wong, Ricky T. Q. Chen, and David Duvenaud. Scalable gradients and variational inference for stochastic differential equations. In *Symposium on Advances in Approximate Bayesian Inference*, volume 118 of *Proc. Mach. Learn. Res. (PMLR)*, page 28. Proceedings of Machine Learning Research PMLR, [place of publication not identified], 2020.

[124] Chong Liu and David J. Prömel. Examples of Itô càdlàg rough paths. *Proc. Amer. Math. Soc.*, 146(11):4937–4950, 2018.

[125] Xuanqing Liu, Tesi Xiao, Si Si, Qin Cao, Sanjiv Kumar, and Cho-Jui Hsieh. Neural SDE: Stabilizing Neural ODE Networks with Stochastic Noise. *arXiv preprint arXiv:1906.02355*, 2019.

[126] Yanghui Liu and Samy Tindel. First-order Euler scheme for SDEs driven by fractional Brownian motions: the rough case. *Ann. Appl. Probab.*, 29(2):758–826, 2019.

[127] Terry Lyons, Sina Nejad, and Imanol Perez Arribas. Non-parametric pricing and hedging of exotic derivatives. *Appl. Math. Finance*, 27(6):457–494, 2020.

[128] Terry J. Lyons. Uncertain volatility and the risk-free synthesis of derivatives. *Appl. Math. Finance*, 2(2):117–133, 1995.

[129] Terry J. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana*, 14(2):215–310, 1998.

[130] Terry J. Lyons, Michael Caruana, and Thierry Lévy. *Differential equations driven by rough paths*, volume 1908 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6–24, 2004, With an introduction concerning the Summer School by Jean Picard.

[131] Xuerong Mao. *Stochastic differential equations and applications*. Horwood Publishing Limited, Chichester, second edition, 2008.

[132] Robert C. Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 51(3):247–257, 1969.

[133] Robert C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *J. Econom. Theory*, 3(4):373–413, 1971.

[134] P. A. Meyer. Un cours sur les intégrales stochastiques. In *Séminaire de Probabilités, X*, volume Vol. 511 of *Lecture Notes in Math.*, pages 245–400. Springer, Berlin-New York, 1976. (Seconde partie: Théorie des intégrales stochastiques, Univ Strasbourg, Strasbourg, année universitaire 1974/1975).

[135] Yulia S. Mishura and Georgiy M. Shevchenko. Existence and uniqueness of the solution of stochastic differential equation involving Wiener process and fractional Brownian motion with Hurst index $H > 1/2$. *Comm. Statist. Theory Methods*, 40(19-20):3492–3508, 2011.

[136] S. E. A. Mohammed. Nonlinear flows of stochastic linear delay equations. *Stochastics*, 17(3):207–213, 1986.

[137] Salah-Eldin A. Mohammed. *Stochastic functional differential equations*, volume 99 of *Research Notes in Mathematics*. Pitman (Advanced Publishing Program), Boston, MA, 1984.

[138] Salah-Eldin A. Mohammed. Stochastic differential systems with memory: theory, examples and applications. In *Stochastic analysis and related topics, VI (Geilo, 1996)*, volume 42 of *Progr. Probab.*, pages 1–77. Birkhäuser Boston, Boston, MA, 1998.

[139] Oleksii Mostovyi and Mihai Sîrbu. Sensitivity analysis of the utility maximisation problem with respect to model perturbations. *Finance Stoch.*, 23(3):595–640, 2019.

[140] A. Neuenkirch, I. Nourdin, and S. Tindel. Delay equations driven by rough paths. *Electron. J. Probab.*, 13:no. 67, 2031–2068, 2008.

[141] Ariel Neufeld and Marcel Nutz. Robust utility maximization with Lévy processes. *Math. Finance*, 28(1):82–105, 2018.

[142] R. Norvaiša. The p-variation and an extension of the class of semimartingales. *Acta Appl. Math.*, 79(1-2):47–53, 2003.

[143] Nicolas Perkowski and David J. Prömel. Pathwise stochastic integrals for model free finance. *Bernoulli*, 22(4):2486–2520, 2016.

[144] Huyên Pham. *Continuous-time stochastic control and optimization with financial applications*, volume 61 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2009.

[145] Huyên Pham, Xiaoli Wei, and Chao Zhou. Portfolio diversification and model uncertainty: a robust dynamic mean-variance approach. *Math. Finance*, 32(1):349–404, 2022.

[146] David J. Prömel and Mathias Trabs. Paracontrolled distribution approach to stochastic Volterra equations. *J. Differ. Equ.*, 302:222–272, 2021.

[147] Philip E. Protter. *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2005. Second edition. Version 2.1, Corrected third printing.

[148] L. C. G. Rogers. Arbitrage with fractional Brownian motion. *Math. Finance*, 7(1):95–105, 1997.

[149] L. C. G. Rogers. The relaxed investor and parameter uncertainty. *Finance Stoch.*, 5(2):131–154, 2001.

[150] Sylvain Rubenthaler. Numerical simulation of the solution of a stochastic differential equation driven by a Lévy process. *Stochastic Process. Appl.*, 103(2):311–349, 2003.

[151] Francesco Russo and Pierre Vallois. Forward, backward and symmetric stochastic integration. *Probab. Theory Related Fields*, 97(3):403–421, 1993.

[152] Alexander Schied, Leo Speiser, and Iryna Voloshchenko. Model-free portfolio theory and its functional master formula. *SIAM J. Financial Math.*, 9(3):1074–1101, 2018.

[153] R. L. Stratonovich. A new representation for stochastic integrals and equations. *SIAM J. Control*, 4:362–371, 1966.

[154] Daniel W. Stroock and S. R. Srinivasa Varadhan. *Multidimensional diffusion processes*, volume 233 of *Grundlehren der Mathematischen Wissenschaften.* Springer-Verlag, Berlin-New York, 1979.

[155] Revaz Tevzadze, Teimuraz Toronjadze, and Tamaz Uzunashvili. Robust utility maximization for a diffusion market model with misspecified coefficients. *Finance Stoch.*, 17(3):535–563, 2013.

[156] Samy Tindel and Iván Torrecilla. Some differential systems driven by a fBm with Hurst parameter greater than 1/4. In *Stochastic analysis and related topics*, volume 22 of *Springer Proc. Math. Stat.*, pages 169–202. Springer, Heidelberg, 2012.

[157] Belinda Tzen and Maxim Raginsky. Neural Stochastic Differential Equations: Deep Latent Gaussian Models in the Diffusion Limit. *arXiv preprint arXiv:1905.09883*, 2019.

[158] Roman Vershynin. *High-dimensional probability*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge, 2018. An introduction with applications in data science, With a foreword by Sara van de Geer.

[159] Kim Weston. Stability of utility maximization in nonequivalent markets. *Finance Stoch.*, 20(2):511–541, 2016.

[160] Toshio Yamada and Shinzo Watanabe. On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.*, 11:155–167, 1971.

[161] L. C. Young. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Math.*, 67(1):251–282, 1936.

[162] M. Zähle. Integration with respect to fractal functions and stochastic calculus. II. *Math. Nachr.*, 225:145–183, 2001.

[163] Thaleia Zariphopoulou. Optimal investment and consumption models with non-linear stock dynamics. *Math. Methods Oper. Res.*, 50(2):271–296, 1999.