Studies of Robustness in Stochastic Analysis and Mathematical Finance

DISSERTATION

zur Erlangung des akademischen Grades

Dr. rer. nat.
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät II
Humboldt-Universität zu Berlin

von
Herrn Dipl.-Math. Nicolas Simon Perkowski

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II:
Prof. Dr. Elmar Kulke

Gutachter:
1. Prof. Dr. Peter Imkeller
2. Prof. Dr. Peter Friz
3. Prof. Dr. Terry Lyons
4. Prof. Dr. Josef Teichmann

eingereicht am: 05.09.2013
Tag der mündlichen Prüfung: 13.12.2013

Abstract

This thesis deals with various problems from stochastic analysis and from mathematical finance that can best be summarized under the common theme of robustness.

We begin by studying financial market models that allow for arbitrage opportunities. This is motivated by the insight that even in the simplest models, arbitrage opportunities can be introduced by slightly changing the information structure. We identify the weaker notion of absence of arbitrage opportunities of the first kind (NA1) as the minimal property which every sensible asset price model should satisfy, and we prove that (NA1) is equivalent to the existence of a dominating probability measure that makes the asset price process a local martingale. We also show that (NA1) is relatively robust with respect to changes in the information structure. As examples of processes which satisfy (NA1) but do not admit equivalent local martingale measures, we study continuous local martingales that are conditioned not to hit zero.

We continue by working with a more robust, model free formulation of the (NA1) property, which permits to describe qualitative properties of "typical asset price trajectories". In this context we construct a pathwise Itô integral, which converges for typical price paths. The obtained results indicate that typical price paths can be used as integrators in the theory of rough paths.

This motivates us to study the rough path integral more carefully. We use a certain Fourier series expansion of continuous functions to develop an alternative approach to the theory of rough paths. Based on this expansion, the integral can be decomposed into different components with different behavior. Then it is easy to see that integrators which are only as regular as a typical sample path of the Brownian motion must be equipped with their Lévy area in order to obtain a pathwise continuous integral operator. The new approach is relatively elementary and it leads to explicit, robust, and recursive numerical algorithms with which one can calculate both Itô and Stratonovich integrals.

Based on these insights, we abstract from integration and note that the different stochastic integrals can be understood as different means of defining products of tempered distributions. If we are only considering functions of a one-dimensional index variable, then the problem of integrating two irregular functions against each other is essentially equivalent to the problem of multiplying two tempered distributions with each other. In higher index dimensions however, the problem of multiplication is more general. We now use the Littlewood-Paley decomposition of tempered distributions, to extend our previously developed approach to rough path integrals to functions of a multidimensional index variable. We construct an operator that agrees with the usual product if it is applied to smooth functions, and that is continuous in a suitable topology. Therefore, we can define the product of suitable tempered distributions in a robust way. Using this operator, we can solve stochastic partial differential equations that were previously difficult to access due to nonlinearities. Since our product operator is continuous, the solutions to these equations depend continuously on the driving stochastic signal, provided that it is approximated in a suitable topology.

Zusammenfassung

Diese Dissertation behandelt verschiedene Fragestellungen aus der stochastischen Analysis und der Finanzmathematik, die sich am besten unter dem gemeinsamen Begriff der Robustheit zusammenfassen lassen.

Zunächst betrachten wir finanzmathematische Modelle, die Arbitragemöglichkeiten zulassen. Dies ist durch die Einsicht motiviert, dass selbst in den einfachsten Modellen Änderungen in der Informationsstruktur üblicherweise Arbitragemöglichkeiten herbeiführen. Wir identifizieren den schwächeren Begriff der Abwesenheit von Arbitragemöglichkeiten der ersten Art (NA1) als die minimale Eigenschaft, die in jedem realistischen finanzmathematischen Modell gelten sollte und wir beweisen, dass (NA1) äquivalent ist zur Existenz eines dominierenden Wahrscheinlichkeitsmaßes, unter dem der Preisprozess ein lokales Martingal ist. Wir zeigen ebenfalls, dass (NA1) relativ robust ist unter Veränderungen in der Informationsstruktur. Als Beispiel für Prozesse, die (NA1) erfüllen aber kein äquivalentes Martingalmaß besitzen, studieren wir stetige lokale Martingale, die darauf bedingt werden, niemals Null zu treffen.

Anschließend wird eine robustere, modellfreie Formulierung der (NA1) Eigenschaft verwendet, die es erlaubt, qualitative Eigenschaften von "typischen Preistrajektorien" zu beschreiben. In diesem Kontext konstruieren wir für typische Preispfade ein pfadweises Itô-Integral. Die hier bewiesenen Resultate deuten darauf hin, dass typische Preispfade als Integratoren in der rough-path-Theorie verwendet werden können.

Dies motiviert ein tiefergehendes Studium des rough-path-Integrals. Zunächst verwenden wir eine bestimmte Fourierdarstellung stetiger Funktionen, um einen alternativen Zugang zur rough-path-Theorie zu entwickeln. Mit Hilfe dieser Darstellung lässt sich das Integral in verschiedene Komponenten mit unterschiedlichen Eigenschaften zerlegen. So sieht man leicht, dass Integratoren mit der Regularität einer typischen Realisierung der Brownschen Bewegung mit ihrer Lévy-Fläche versehen werden müssen, um ein pfadweise stetiges Integral zu erhalten. Der neue Ansatz ist relativ elementar und führt zu expliziten, robusten und rekursiven numerischen Algorithmen, mithilfe derer sich sowohl Itô- als auch Stratonovich-Integrale pfadweise berechnen lassen.

Darauf aufbauend abstrahieren wir vom Integral und fassen das Problem der stochastischen Integration als einen Spezialfall des Problems der Multiplikation von temperierten Distributionen auf. Mittels Integration und Differentiation lässt sich zeigen, dass die beiden Probleme im Wesentlichen äquivalent sind, solange wir nur Funktionen einer eindimensionalen Indexvariablen betrachten. In höheren Dimensionen ist das Problem der Multiplikation jedoch weitaus allgemeiner. Wir verwenden nun die Littlewood-Paley Darstellung von temperierten Distributionen, um unseren zuvor entwickelten Zugang zur rough-path-Theorie auf Funktionen mehrdimensionaler Variablen zu erweitern. Wir konstruieren einen Operator, der für glatte Funktionen mit dem üblichen Produkt übereinstimmt, und in einer geeigneten Topologie stetig ist. Somit können wir auf robuste Art und Weise das Produkt von geeigneten temperierten Distributionen definieren. Nun lassen sich stochastische partielle Differentialgleichungen lösen, die bisher aufgrund von Nichtlinearitäten nicht gut zugänglich waren. Aufgrund der Stetigkeit unseres Produktoperators hängen die Lösungen dieser Gleichungen stetig vom stochastischen Rauschen ab, solange dieses in einer geeigneten Topologie approximiert wird.

Acknowledgement

First and foremost, I would like to express my deep gratitude to Peter Imkeller. The discussions with him immensely broadened my horizon, and his enthusiasm and support made it a pleasure to work under his supervision. His remarks proved to be very fruitful (most notably in Chapter 4, which grew from a single question that he asked during a seminar talk). He gave me the freedom to pursue any research interest I had, and wherever those interests took me, he was always there to guide and encourage me (and in a few cases also to let me know in the gentlest possible way that this line of investigation was probably leading nowhere).

I would like to thank Peter Friz, Terry Lyons, and Josef Teichmann for agreeing to be co-examiners for this thesis.

Special thanks are due to Sri Namachchivaya. He trusted me as a young Ph.D. student, invited me to visit him at UIUC, and introduced me to the fascinating fields of nonlinear filtering and stochastic homogenization. While our joint work is not included in this thesis, it was an experience of great importance to me, from which I profited very much. The things I learned from our collaboration were helpful throughout the years of my thesis work.

I am also deeply grateful to Massimiliano Gubinelli for his patience and his trust in me, and for everything he has taught me. During a short research visit and in hundreds of emails, he has fundamentally reshaped my understanding of rough paths and stochastic partial differential equations. I am honored and looking forward to be able to continue working with him as a postdoctoral researcher.

I would like to thank Walter Schachermayer for inviting me to stay at the University of Vienna for three months and for the pleasant atmosphere and great team he has formed. During that time I gained insights into the forefront of financial mathematics, for which I am very thankful.

I would like to thank Johannes Ruf for collaborating with me. His energy and persistence are an inspiration and have often led me to push further, giving me a more profound understanding of the studied objects than I would have had otherwise.

Asgar Jamneshan introduced me to filtration enlargements, which were one of the main motivations for Chapter 1, and I would like to thank him for this.

I am thankful to Martin Hairer for suggesting to study the nonlinear parabolic Anderson model as an application of the product operator developed in Chapter 5.

Mathias Beiglböck pointed me to the pathwise approach of Vovk to mathematical finance and to possible connections to rough paths; this led to Chapter 3, and I would like to thank him for his suggestion.

I am grateful to my friends and colleagues in Berlin, Urbana-Champaign, and Vienna, for countless discussions and for making the work towards this thesis a joyful experience. Especially I would like to thank David Prömel for proofreading parts of the thesis.

Most importantly, I am greatly indebted to my family and to Cami. Had it not been for their unconditional love and support, I would not have been able to write this thesis.

Financial and infrastructure support from the Berlin Mathematical School is gratefully acknowledged.

To Cami

Contents

Basic notation

Numbers

- $\mathbb{R} = (-\infty, \infty) =$ the set of real numbers; $\mathbb{R}_+ = [0, \infty)$
- $\mathbb{Q} =$ the set of rational numbers; $\mathbb{Q}_+ = \mathbb{Q} \cap \mathbb{R}_+$
- $\mathbb{N} = \{0, 1, \dots\} =$ the set of nonnegative integers; $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$
- $\mathbb{Z} = \{0, -1, 1, -2, 2, \dots\} =$ the set of integers
- $\imath = \sqrt{-1} =$ the imaginary unit
- $a \wedge b = \min\{a, b\}$; $a \vee b = \max\{a, b\}$ for $a, b \in \mathbb{R}$
- $\lfloor a \rfloor = \max\{k \in \mathbb{Z} : k \le a\}$; $\lceil a \rceil = \min\{k \in \mathbb{Z} : a \le k\}$ for $a \in \mathbb{R}$
- $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$ denotes the Kronecker delta

Basic spaces

- $\mathcal{L}(\mathbb{X}, \mathbb{Y}) =$ the space of bounded linear operators between the Banach spaces $\mathbb{X}$ and $\mathbb{Y}$
- $C(\mathbb{X}, \mathbb{Y}) =$ the space of continuous functions between the topological spaces $\mathbb{X}$ and $\mathbb{Y}$
- $C^m(\mathbb{R}^d, \mathbb{R}^n) = C^m$ the space of m times continuously differentiable functions from $\mathbb{R}^d$ to $\mathbb{R}^n$, where $m, d, n \in \mathbb{N}^*$
- $C_b^m(\mathbb{R}^d, \mathbb{R}^n) =$ the space of m times continuously differentiable functions from $\mathbb{R}^d$ to $\mathbb{R}^n$ with bounded partial derivatives up to order m, where $m, d, n \in \mathbb{N}^*$

Multiindex notation

- $|\eta| = \eta_1 + \dots + \eta_d =$ the length of the multiindex $\eta \in \mathbb{N}^d$, where $d \in \mathbb{N}^*$
- $x^\eta = x^{\eta_1} \cdot \dots \cdot x^{\eta_d}$ for $x \in \mathbb{R}^d$ and $\eta \in \mathbb{N}^d$, where $d \in \mathbb{N}^*$
- $\eta! = \eta_1! \dots \eta_d!$ for $\eta \in \mathbb{N}^d$, where $d \in \mathbb{N}^*$

Derivatives

- $\mathrm{D}F(x) =$ the total differential of F in x
- $F'(x) = \mathrm{D}F(x)$
- $\mathrm{D}^m F(x) =$ the m–th order derivative of F in x, where $m \in \mathbb{N}$
- $\mathrm{D}_x F(t,x) =$ the spatial derivative of F in (t,x)
- $\partial_t F(t,x) = \frac{\partial}{\partial t} F(t,x) =$ the partial derivative of F in direction t
- $\partial_{x_k} F(x) = \frac{\partial}{\partial_{x_k}} F(x)$ the partial derivative of F in direction x_k
- $\partial^\eta F(x) = \frac{\partial^{|\eta|}}{\partial^{\eta_1}_{x_1} \dots \partial^{\eta_d}_{x_d}} F(x) =$ a higher order partial derivative of F, where $\eta \in \mathbb{N}^d$

Norms and related objects

- $xy = \sum_{k=1}^d x_k y_k =$ the inner product of x and y
- $|x| = \sqrt{xx}$ the Euclidean norm of a vector or matrix x
- $\|F\|_{C_b^m} = \sum_{|\eta| \in \mathbb{N}^d : |\eta| \leq m} \|\partial^\eta F\|_{L^\infty}$
- $|\{\dots\}| =$ the number of elements in the set $\{\dots\}$

Measures

- $\ll$ and $\gg$ absolute continuity between measures
- $\sim$ equivalence between measures

Limits

- $\lim_{s \to t-} = \lim_{s \to t, s<t}$
- $\lim_{s \to t+} = \lim_{s \to t, s>t}$

Conventions

- $a \lesssim b$ means that there exists a constant $c > 0$, such that $a \leq cb$; $a \gtrsim b$ means $b \lesssim a$; $a \simeq b$ means $a \lesssim b$ and $b \lesssim a$ – except for index variables of dyadic blocks in Chapter 5
- $a(x) \lesssim_x b(x)$ means that there exists a constant $c(x) > 0$, such that $a(x) \leq c(x)b(x)$; $a(x) \gtrsim_x b(x)$ means $b(x) \lesssim_x a(x)$; $a(x) \simeq_x b(x)$ means $a(x) \lesssim_x b(x)$ and $b(x) \lesssim_x a(x)$
- $0/0 = 0$; $\infty \cdot 0 = 0$ – except in Chapter 2

- (in-) equalities between random variables are to be understood in the almost sure sense – except in Chapter 3

Introduction

This thesis deals with questions from mathematical finance and from stochastic analysis that can best be summarized under the common theme of robustness. Mathematical models are usually an idealization of the world. Therefore, it is important to understand how robust they are with respect to changes in the underlying assumptions. On the other side, even if we have reason to believe that a model gives an accurate description of a certain phenomenon, in general this model will be infinite dimensional – at least in the applications that we have in mind. Since a computer can only store and process a finite amount of data, we might then ask how accurate a given finite dimensional approximation of the infinite dimensional model is.

One of the basic problems in mathematical finance is the derivation of "fair" prices for certain financial derivatives. For example, if $(S_t)_{t\in[0,T]}$ describes the evolution of the discounted price of a financial asset, then one might ask for fair prices for European call options. A European call option with strike $K > 0$ and maturity $T > 0$ is a contract that allows, but not obliges, its owner to buy one asset S_T at time T for the price K. Thus, under the paradigm of rational action, at time T the payoff for the owner of the contract is equal to $(S_T - K)^+$, where $x^+ = \max\{x, 0\}$ for all $x \in \mathbb{R}$.

First results in this direction have been obtained by Bachelier [Bac00], and by Black and Scholes [BS73] and Merton [Mer73]. Their derivations are based on hedging arguments: They assume that the asset price evolution is given by a Brownian motion (respectively a geometric Brownian motion), and (implicitly) use the predictable representation property to show that there exists a unique $p \in \mathbb{R}$ and a unique predictable, square-integrable strategy H, such that

$$(S_T - K)^+ = p + \int_0^T H_s \mathrm{d}S_s.$$

In other words, an investor with initial capital p can obtain the payoff $(S_T - K)^+$ by investing in S – at least in a frictionless market. Therefore, the "fair" price of the call option is equal to p. If Q denotes the unique probability measure on $\mathcal{F}_T$ that is equivalent to P and makes S a martingale, then $p = E_Q((S_T - K)^+)$.

A model that assigns to every financial derivative a unique price is called *complete*. However, most practically relevant models are *incomplete*, i.e. in those models the price process does not have the predictable representation property. Incomplete models are more realistic, because we do not expect that every financial derivative can be replicated by investing in the asset. In such a model there might not exist an integrand H such that $(S_T - K)^+ = p + \int_0^T H_s \mathrm{d}S_s$ for some $p > 0$. But usually there will exist $p > 0$, such that there is a strategy H for which $(S_T - K)^+ \le p + \int_0^T H_s \mathrm{d}S_s$. Such a p is called a *superhedging price* for $(S_T - K)^+$, and it is clear that the price of the option should be at most p. A probability measure Q that is equivalent to P and that makes S a local martingale is called an *equivalent local martingale measure* for S. If the set $\mathcal{Q}$ of equivalent local martingale measures is non-empty, then one can show that

$$\inf \left\{ p > 0 : \exists\, H \text{ s.t. } p + \int_0^T H_s \mathrm{d}S_s \geq (S_T - K)^+ \right\} = \sup_{Q \in \mathcal{Q}} E_Q((S_T - K)^+).$$

But then the question arises why such equivalent local martingale measures should exist in the first place. This can be economically motivated by the Fundamental Theorem of Asset Pricing by Delbaen and Schachermayer [DS94], see also [DS06], who show that if S is locally bounded, then $\mathcal{Q}$ is non-empty if and only if S satisfies the property *no free lunch with vanishing risk (NFLVR)*. The (NFLVR) property roughly states that it is not possible to make a risk free profit by investing in S, and that an investor who is only willing to take a small risk can only make a (relatively) small profit.

The Fundamental Theorem of Asset Pricing is one of the cornerstones of modern financial mathematics, and based on this result, most models assume that the asset price process admits an equivalent local martingale measure. While this is a reasonable assumption in most situations, in certain cases it is too restrictive. For example, if S admits an equivalent local martingale measure, and if X is a $\mathcal{F}_T$–measurable random variable, and $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(X)$, $t \in [0, T]$, is the filtration generated by $\mathcal{F}_t$ and X, then usually there will no longer exist an equivalent $Q \sim P$, such that S is a Q–local martingale in the filtration $(\mathcal{G}_t)$. If we interpret $\mathcal{G}_t$ as the information available at time t to an "informed" investor, then this shows that the (NFLVR) property is not very robust with respect to changes in the information structure. On the other side Ankirchner [Ank05] observed that the maximal expected logarithmic utility for an informed investor may well be finite, i.e. that under suitable conditions

$$\sup\{E(\log(1 + (H \cdot S)_T)) : H \text{ is a } (\mathcal{G}_t) - \text{adapted strategy}\} < \infty.$$

This indicates that, despite S not admitting an equivalent local martingale measure in the filtration $(\mathcal{G}_t)$, the model is not completely degenerate, meaning that even for an informed investor it is not possible to generate "infinite wealth" by investing in S.

Therefore, the first part of this thesis is concerned with finding a minimal property that has to be satisfied by a financial market model, such that

1. there exists an unbounded utility function U for which the maximal expected U–utility is finite, and

2. the property is (relatively) robust with respect to changes in the information structure (filtration enlargements).

In that case we would like to characterize all financial market models satisfying this property, by providing a result that is similar in spirit to the Fundamental Theorem of Asset Pricing. This is the content of Chapter 1, and it turns out that this minimal property is the (NA1) property. Intuitively, a stochastic process S admits *no arbitrage opportunities of the first kind (NA1)*, if it is not possible to make a large profit by investing in S while at the same time only taking a very small risk.

The basic example of a stochastic process satisfying only (NA1) but not (NFLVR) is given by the three dimensional Bessel process. It is well known that this process can be obtained by conditioning a Brownian motion not to hit zero. In Chapter 2 we study the dynamics of a general local martingale which is conditioned not to hit zero, and we see that this always leads to a process satisfying (NA1).

While we show in Chapter 1 that (NA1) is the minimal condition which a reasonable financial market model has to satisfy, and that (NA1) is rather robust, our results are still in the context of one fixed model. But in applications there is usually some model uncertainty. For example, consider the Black-Scholes model, where the asset price evolution is given by

$$\mathrm{d}S_t = S_t(b\mathrm{d}t + \sigma\mathrm{d}W_t)$$

for a Brownian motion W. It is possible to estimate the volatility σ given data samples from the price process. But unless the full (continuous time) sample path of S is observed, it is not possible to determine σ with absolute certainty. Therefore, in Chapter 3 we work with a model free formulation of the (NA1) property, as introduced by Vovk [Vov12], that allows to determine which properties are satisfied by "typical price paths". Here "model free" refers to the fact that we do not assume that the price process is given as a stochastic process with known distribution, but that it is an arbitrary continuous path. In this context, p is a superhedging price for the option $(S_T - K)^+$ if there exists a strategy H such that $p + \int_0^T H_s\mathrm{d}S_s \geq (S_T - K)^+$ for *every* continuous path S, or at least for a sufficiently rich set of paths. In the classical setting the inequality has to hold only almost surely, which is obviously easier to satisfy. Therefore, the results obtained in this model free context should be more robust. But while in the classical setting we could use Itô stochastic integrals, here it is not very clear what the integral $\int_0^T H_s\mathrm{d}S_s$ should mean. This is the main problem treated in Chapter 3, where we construct a pathwise Itô type integral that converges for typical price paths, and where we show that typical price paths can be used as integrators in Lyons' theory of rough paths [Lyo98].

The theory of rough paths is the focus of Chapter 4, where we develop an alternative, Fourier series based approach to rough path integration. If W is a d–dimensional Brownian motion on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$, then Itô's stochastic integral is a bounded linear operator from $L^2(\Omega \times [0,T], P \otimes \lambda)$ to the space of square integrable martingales. Here λ denotes Lebesgue measure. But it is not continuous in a pathwise sense: For example, if F is a smooth function from $\mathbb{R}^d$ to $\mathbb{R}^d$, if $W(\omega)$ is a Brownian sample path, and $W^n(\omega)$ is a sequence of smooth paths converging uniformly to $W(\omega)$, then it is in general not true that $\int_0^T F(W^n_s(\omega))\mathrm{d}W^n_s(\omega)$ converges to $(\int_0^T F(W_s)\mathrm{d}W_s)(\omega)$. And also the solution $S(\omega)$ of a stochastic differential equation $\mathrm{d}S_t = F(S_t)\mathrm{d}W_t$ with $S_0 = s$ will in general not be the limit of the solutions $S^n(\omega)$ to the equations $\mathrm{d}S^n_t(\omega) = F(S^n_t(\omega))\mathrm{d}W^n_t(\omega)$ and $S^n_0 = s$. Of course, we would not really expect to obtain an Itô integral or an Itô SDE in the limit, because the approximating paths $W^n(\omega)$ satisfy the classical integration by parts rule and not Itô's formula. But in general we will not even obtain the corresponding Stratonovich integral / SDE solution in the limit. The reason is that the uniform topology is too coarse. Lyons [Lyo98] observed that both integrals and solutions to SDEs depend continuously on the driving noise, as long as the noise is enhanced to a $(d+d^2)$–dimensional path, consisting of the noise itself, but also of its iterated integrals. We give an elementary approach to rough path integration that is based on a series representation of continuous functions. If $(\varphi_{m,k})_{m\in\mathbb{N},0\le k\le 2^m}$ are the Schauder functions, to be defined below, then every continuous function $f:[0,1]\to\mathbb{R}^d$ can be represented as $f(t)=\sum_{m,k} f_{mk}\varphi_{mk}(t)$. If $g:[0,1]\to\mathbb{R}^d$ is another continuous function with expansion $g(t)=\sum_{n,\ell} g_{n\ell}\varphi_{n\ell}(t)$, then we may formally define

$$\int_0^\cdot f(s)\mathrm{d}g(s) = \sum_{m,k}\sum_{n,\ell} f_{mk}g_{n\ell}\int_0^\cdot \varphi_{mk}(s)\mathrm{d}\varphi_{n\ell}(s),$$

because the functions $(\varphi_{n\ell})$ are of bounded variation. Examining the convergence of this double series will be the main interest of Chapter 4, and we will show that on suitable function spaces, the integral can be defined as a continuous operator.

In Chapter 5 we reformulate the results of Chapter 4 in the language of Littlewood-Paley blocks as opposed to Schauder functions. This allows us to define products of tempered distributions that have a multi dimensional index set. Moreover, our product is a *continuous* bilinear operator on suitable function spaces. If W is a Brownian motion and F is a smooth function, then the Itô integral $\int_0^\cdot F(W_s)\mathrm{d}W_s$ can be understood as a way of defining the distribution $F(W)\dot{W}$, where $\dot{W}$ is the time derivative of the Brownian motion. The same interpretation works also for the Stratonovich integral and for the rough path integral, which shows that there is a multitude of techniques to treat nonlinear operations on tempered distributions that have a one dimensional index set. Maybe somewhat surprisingly, if the index set is multi dimensional, then there are much fewer techniques available. We formulate a theory of "paracontrolled distributions" that is similar in spirit to rough paths, and nearly completely analogous to the theory developed in Chapter 4. We then apply our theory to solve two stochastic partial differential equations (SPDEs) about which previously there was not much known. Since our product is continuous in a suitable topology, we obtain automatically that the solutions to these SPDEs depend continuously on the driving noise.

So the subjects treated in this thesis can be summarized as follows. In the first Chapter we study financial market models, where we are interested in the (NA1) property, which is more robust than the classical (NFLVR) property, but still leads to economically sensible models. In Chapter 2 we derive the dynamics of a nonnegative continuous local martingale that is conditioned not to hit zero. In the third chapter we study the (NA1) property in a more robust, model free context, and we show that it allows us to define pathwise "stochastic" integrals. In Chapter 4 we present a Fourier based approach to rough path integration, which allows us to identify a topology in which the solutions of SDEs depend continuously on the driving signal. In Chapter 5 we develop a new way of defining the product between two tempered distributions. Using this product, we solve two nonlinear SPDEs that previously were not very well understood. We also show the robustness of our solutions, in the sense that under a wide range of smooth approximations of the driving noise, the solutions of the smooth equations converge to our solution.

Each chapter is relatively self-contained and can be read independently. In the following, we give a more detailed summary of the content of the single chapters.

Chapter 1: Dominating local martingale measures and arbitrage under information asymmetry

Chapter 1 is based on Imkeller and Perkowski [IP11]. Let S be a stochastic process on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$. The process S is supposed to model the evolution of a discounted asset price in a frictionless market. A λ*–admissible strategy* is a predictable process $(H_t)_{t\geq 0}$ for which the stochastic integral $H \cdot S := \int_0^\cdot H_s \mathrm{d}S_s$ exists and satisfies $P((H \cdot S)_t \geq -\lambda) = 1$ for all $t \geq 0$. In that case we write $H \in \mathcal{H}_\lambda$.

We say that S admits *no arbitrage opportunities of the first kind (S satisfies (NA1))* if the set

$$\mathcal{W}_1^\infty := \left\{1 + (H \cdot S)_\infty : H \in \mathcal{H}_1 \text{ and } \lim_{t\to\infty} (1 + (H \cdot S)_t) \text{ exists}\right\}$$

is bounded in probability, i.e. if $\lim_{m\to\infty} \sup_{X\in\mathcal{W}_1^\infty} P(X \geq m) = 0$. Heuristically, this means that an investor who is only willing to risk the initial capital of 1 is not able to make a very large profit.

The first result of this chapter is that S satisfies (NA1) if and only if there exists an unbounded utility function $U : [0, \infty) \to \mathbb{R}$, such that

$$\sup_{X\in\mathcal{W}_1^\infty} (EU(X)) < \infty.$$

The existence of such a U is a minimal requirement that every sensible model should satisfy. Otherwise any investor will be able to generate infinite utility by investing in S – no matter what her preference structure looks like!

If S admits an equivalent local martingale measure, then S satisfies (NA1). More precisely, by the Fundamental Theorem of Asset Pricing, S satisfies (NFLVR), and (NFLVR) is equivalent to (NA1) and (NA). Here (NA) means *no arbitrage*, which is satisfied if for every $X \in \mathcal{W}_1^\infty$ either $P(X < 1) > 0$ or $P(X = 1) = 1$. Heuristically, S satisfies (NA) if it is not possible to make a risk free profit by investing in S. We will show that the (NFLVR) property is usually violated after filtration enlargements. On the other side, we will show for enlargements of the type $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(X)$, where X is a random variable which satisfies Jacod's criterion, that every process satisfying (NA1) under $(\mathcal{F}_t)$ also satisfies (NA1) under $(\mathcal{G}_t)$.

In conclusion, the (NA1) property has to be satisfied by every sensible model and it is relatively robust with respect to changes in the information structure. Our next aim is to characterize all models satisfying (NA1). We know that the (NFLVR) property (which is equivalent to (NA1) and (NA)) is satisfied if and only if S admits an equivalent local martingale measure. Moreover, Delbaen and Schachermayer [DS95b] showed that if S is continuous and satisfies (NA), then S admits an absolutely continuous local martingale measure. We complement this picture by proving that S satisfies (NA1) if and only if it admits a dominating local martingale measure, i.e. a measure Q such that P is absolutely continuous with respect to Q, and such that S is a Q–local martingale.

It is not very easy to work directly on the space of dominating measures for P, and therefore we would first like to construct a type of Radon-Nikodym derivative $\mathrm{d}Q/\mathrm{d}P$. Of course, in general $\mathrm{d}Q/\mathrm{d}P$ will not exist if Q dominates P. Here we rely on a progressive Lebesgue decomposition on filtered probability spaces, the Kunita-Yoeurp decomposition, that associates to every dominating measure Q a P–supermartingale Z. Moreover, we show that if Q makes S a local martingale, then Z is a *supermartingale density*, i.e. the process $(1 + (H \cdot S))Z$ is a P–supermartingale for every $H \in \mathcal{H}_1$ (where $\mathcal{H}_1$ is defined with respect to P).

So in a first step, we show that the existence of supermartingale densities is equivalent to (NA1). In a second step, we show that we can associate dominating local martingale measures to supermartingale densities.

Not all the results here are new: (NA1) and its relation to filtration enlargements, utility maximization, supermartingale densities, and dominating local martingale measures have been studied for example by Ankirchner [Ank05], Karatzas and Kardaras [KK07], and Ruf [Ruf13] respectively. But to the best of our knowledge, here we give the first general classification of the (NA1) property. It also seems to be a new (albeit simple) result that (NA1) is the minimal property that every reasonable model should satisfy.

Chapter 2: Conditioned martingales

This chapter, which is based on Perkowski and Ruf [PR12], falls somewhat out of the theme of this thesis, in the sense that it has not much to do with robustness. One of the basic examples for a process that satisfies (NA1) but does not admit an equivalent local martingale measure is given by the three dimensional Bessel process. It is a classical result, going back at least to McKean [McK63], that the three dimensional Bessel process has the same dynamics as a Brownian motion which is conditioned not to hit zero, and that conversely a downward conditioned Bessel process has the same dynamics as a Brownian motion.

In Chapter 2 we show that a similar result holds for every continuous local martingale. Our proof is probabilistic and based on the simple observation that if M is a continuous local martingale starting in 1 and if τ_a and τ_0 denote the first hitting times of $a > 1$ and 0 respectively, then the two measures $\mathrm{d}P(\cdot|\tau_a < \tau_0)$ and $M_{\tau_a \wedge \tau_0}\mathrm{d}P$ agree on the σ–algebra $\mathcal{F}_{\tau_0 \wedge \tau_a}$. Then it only remains to let a tend to ∞, which can be done by using Parthasarathy's extension theorem. The so constructed measure Q is the Föllmer measure of M. Therefore, the main result of this chapter is that the Föllmer measure of a nonnegative local martingale M can be obtained by conditioning M not to hit zero. Under the Föllmer measure, $1/M$ is a local martingale, and therefore we can now condition M downward, which corresponds to conditioning $1/M$ upward. By the same argument as before, the downward conditioned Föllmer measure Q is equal to the original measure P.

As an application, we explicitly derive the dynamics of upward and downward conditioned diffusions.

Chapter 3: Pathwise integration in model free finance

In this chapter, we are working with a pathwise version of the (NA1) property. Let $\Omega := C([0,T],\mathbb{R}^d)$ be the space of continuous paths with values in $\mathbb{R}^d$. We interpret Ω as the space of discounted asset price trajectories. The filtration $(\mathcal{F}_t)_{t\in[0,T]}$ is defined via $\mathcal{F}_t := \sigma(\omega(s) : s \leq t)$. A *simple strategy* is a process H of the form $H_t(\omega) = \sum_n F_n(\omega) 1_{(\tau_n,\tau_{n+1}]}(t)$ for suitable stopping times $(\tau_n)_{n\in\mathbb{N}}$ and $\mathcal{F}_{\tau_n}$–measurable random variables $(F_n)_{n\in\mathbb{N}}$. For $\lambda > 0$, a simple strategy H is called λ–admissible if $(H\cdot\omega)_t \geq -\lambda$ for all $\omega\in\Omega$ and all $t\in[0,T]$. Note that this is a stronger requirement than in Chapter 1, where we only assumed that almost surely $(H\cdot\omega)_t \geq -1$ for all $t\in[0,T]$. We write $\mathcal{H}_{\lambda,s}$ for the set of λ–admissible simple strategies.

Inspired by Vovk [Vov12], we define the *outer content* of $A\subseteq\Omega$ as the cheapest superhedging price,

$$\overline{P}(A) := \inf\left\{\lambda > 0 : \exists\,(H^n)\subseteq\mathcal{H}_{\lambda,s} \text{ s.t. } \liminf_{n\to\infty}(\lambda + (H^n\cdot\omega)_T) \geq 1_A(\omega)\,\forall\omega\in\Omega\right\}.$$

A set of paths $A\subseteq\Omega$ is called a *null set* if it has outer content zero. We then show that A is a null set if and only if there exists a sequence of 1–admissible strategies $(H_n)_{n\in\mathbb{N}}$, such that $1 + \liminf_{n\to\infty}(H^n\cdot\omega)_T = \infty$ for all $\omega\in A$. Therefore, every null set A can be interpreted as a model free arbitrage opportunity of the first kind: It is possible to generate a very large profit by investing in paths from A, without ever risking to lose more than the initial capital of 1. A property (P) is said to hold for typical price paths if the set where (P) is violated is a null set.

In a model free setting, where no probability measure is given, it is a priori not clear how to define stochastic integrals for more complicated integrands than the simple strategies described above. But such stochastic integrals may be required to develop a sufficiently powerful theory (usually simple strategies are not enough). The main result of Chapter 3 states that if H is a càdlàg adapted process, such that $t\mapsto H_t(\omega)$ has the same variational regularity as ω, then for typical price processes the stochastic integral $H\cdot\omega$ can be defined as limit of Riemann sums. However, the null set where the Riemann sums do not converge depends on H. For applications, it would be convenient to exclude one null set in the beginning, and to be able to construct all stochastic integrals for all remaining paths. This can be achieved by taking H to be the coordinate mapping, so that we construct all the second order iterated integrals $(\int_0^\cdot\int_0^s \mathrm{d}\omega^i(r)\mathrm{d}\omega^j(s))_{1\leq i,j\leq d}$. These are the crucial ingredients that are needed to use Lyons' rough path integral [Lyo98]. The rough path integral is an analytic object, and therefore it can be constructed for all integrands and for all ω outside the null set where the iterated integrals do not exist.

It is remarkable that here we are not in a probabilistic context, and that typical price paths are too irregular to apply Young integration to construct their iterated integrals.

To the best of our knowledge this is the first time that the iterated integrals of paths are constructed in a nontrivial setting without using probability theory.

Chapter 4: A Fourier approach to pathwise integration

In Chapter 3 we saw that the rough path integral may be a useful tool in model free finance. Here we study the rough path integral more carefully, giving an alternative approach based on Fourier series.

It is a classical result of Ciesielski [Cie60] that $C^\alpha := C^\alpha([0,1],\mathbb{R}^d)$, the space of α-Hölder continuous functions on $[0,1]$ with values in $\mathbb{R}^d$, is isomorphic to $\ell^\infty(\mathbb{R}^d)$, the space of bounded sequences with values in $\mathbb{R}^d$. The isomorphism gives a series expansion of any Hölder continuous function f as $f(t) = \sum_{m,k} f_{mk}\varphi_{mk}(t)$. Here $(\varphi_{mk})_{mk}$ are (a rescaled version of) the Schauder functions, the primitives of the Haar wavelets, and $(f_{mk})_{m,k}$ are constant coefficients. The function f is α–Hölder continuous if and only the coefficients (f_{mk}) decay rapidly enough, more precisely if $\sup_{m,k} 2^{n\alpha}|f_{mk}| < \infty$. Since Ciesielski's work, this isomorphism has been extended to more general Fourier and wavelet bases, for which one obtains the same type of results: the regularity of a function is encoded in the decay of the coefficients of its series expansion. For details see [Tri06].

But for the applications that we have in mind, the Schauder functions have two very pleasant properties. The coefficients (f_{mk}) are (second order) increments of f, so that we understand their statistics if f is a stochastic process with known distribution. Furthermore, every φ_{mk} is piecewise linear, which makes it easy to calculate integrals of the type $\int_0^\cdot \varphi_{mk}(s)\mathrm{d}\varphi_{n\ell}(s)$.

If f and g are Hölder-continuous functions, then we formally set

$$\int_0^t f(s)\mathrm{d}g(s) := \sum_{m,k}\sum_{n,\ell} f_{mk}g_{n\ell}\int_0^t \varphi_{mk}(s)\mathrm{d}\varphi_{n\ell}(s).$$

Examining the convergence of this double series is the focus of Chapter 4. Using integration by parts, the series can be decomposed into components with different behavior:

$$\begin{aligned}\int_0^\cdot f(s)\mathrm{d}g(s) &= \sum_{m<n}\sum_{k,\ell} f_{mk}g_{n\ell}\varphi_{mk}\varphi_{n\ell} + \sum_{m,k} f_{mk}g_{mk}\int_0^\cdot \varphi_{mk}(s)\mathrm{d}\varphi_{mk}(s)\\ &\quad + \sum_{m>n}\sum_{k,\ell}(f_{mk}g_{n\ell} - f_{n\ell}g_{mk})\int_0^\cdot \varphi_{mk}(s)\mathrm{d}\varphi_{n\ell}(s)\\ &=: \pi_<(f,g) + S(f,g) + L(f,g),\end{aligned}$$

where $\pi_<$ is the *paraproduct*, S is the *symmetric part*, and L is the antisymmetric *Lévy area* (in fact we will show that L is closely related to the Lévy area of a suitable dyadic martingale). If $f \in C^\alpha$ and $g \in C^\beta$, then $\pi_<(f,g)$ is well defined and in C^β, and $S(f,g)$ is well defined and in $C^{\alpha+\beta}$. But in general $L(f,g)$ only converges if $\alpha+\beta > 1$. In that case $L(f,g) \in C^{\alpha+\beta}$. In other words, $\pi_<$ is always defined but the roughest component, S is always defined and smooth, and L is not always defined; but if it is, then it is also smooth. Since the condition $\alpha + \beta > 1$ excludes one of the most interesting examples, the case when $g = W(\omega)$ is a sample path of the Brownian motion, and $f = F(W(\omega))$ for a smooth function F, we are then looking for a way to define the Lévy area also in situations when it cannot be constructed using purely analytical arguments.

As we saw above, in case $\alpha+\beta > 1$ the integral satisfies $\int_0^\cdot f(s)\mathrm{d}g(s) - \pi_<(f,g) \in C^{\alpha+\beta}$. Similarly, we will show that for $g \in C^\alpha$ and for a smooth function F we have $F(g) \in C^\alpha$ but $F(g) - \pi_<(\mathrm{D}F(g), g) \in C^{2\alpha}$. So in both cases the rough component is given by $\pi_<$, and if it is subtracted, then the remainder is relatively smooth. Therefore we say that f is *controlled by* g if there exists f^g such that $f - \pi_<(f^g, g)$ is "smooth". Our aim is to construct the Lévy area $L(f,g)$ for f that is controlled by g. Our first main result is a sort of commutator estimate, where we show that $R(f^g, g, g) := L(\pi_<(f^g, g), g) - \int_0^\cdot f^g(s)\mathrm{d}L(g,g)(s)$ is a bounded trilinear operator provided that the regularities of the three functions add up above 1. In particular, this will be the case if $f^g, g \in C^\alpha$ for some $\alpha > 1/3$. In that case the problem of constructing $L(f,g)$ for f controlled by g reduces to constructing $L(g,g)$. If $L(g,g) \in C^{2\alpha}$, which would be its natural regularity, then we can set

$$L(f,g) := L(f - \pi_<(f,g), g) + R(f^g, g, g) + \int_0^\cdot f^g(s)\mathrm{d}L(g,g)(s).$$

This L depends continuously on f and g if the space of integrands is equipped with a "controlled path norm", and if we are keeping track of the Lévy area L in the space of integrators.

This approach provides us with a simple recursive algorithm for calculating rough path integrals, based on the series expansions of f and g. But these integrals will be of Stratonovich type, because they are obtained by smooth approximation. In a second step we compare our Schauder function integral with the integral obtained from nonanticipating Riemann sums. This leads to an expansion of the quadratic variation in terms of the Schauder functions, which can also be computed (nearly) recursively. Building on our previous results, we can show that if the nonanticipating dyadic Riemann sums of g integrated against itself converge, then also the nonanticipating dyadic Riemann sums of f integrated against g converge.

While this is not the focus of this work, it is then clear from the results of Lyons [Lyo98] and Gubinelli [Gub04] that the pathwise continuity of the integral implies the pathwise continuity of the solution flows to SDEs.

In the last part of this section we construct the Lévy area for certain hypercontractive processes and for continuous local martingales.

Chapter 5: Paracontrolled distributions and applications to SPDEs

This chapter is based on Gubinelli, Imkeller, and Perkowski [GIP12]. To motivate the results of this section, we first give a reinterpretation of the results of the previous chapter.

Let us say that we want to define the integral $\int_0^\cdot F(W_s)\mathrm{d}W_s$ for a Brownian motion W and a smooth function F. Formally, the integral can be rewritten as $\int_0^\cdot F(W_s)\dot{W}_s\mathrm{d}s$, where $\dot{W}_s$ is the white noise, i.e. the time-derivative of the Brownian motion. Hence, the integral may be constructed using three operations:

- W is differentiated;
- $F(W)$ and $\dot{W}$ are multiplied with each other;
- the result $F(W_s)\dot{W}_s$ is integrated in time.

The first and third operation are linear. Since $F(W)$ and W are tempered distributions, these linear operations pose no problem and can be treated with analytical arguments. The problem lies in the second operation, the multiplication, which is nonlinear. As we will see, it is always possible to define the product fg for tempered distributions $f \in C^\alpha$ and $g \in C^\beta$ if $\alpha + \beta > 0$ (Hölder-Besov spaces with negative regularity will be introduced in Chapter 5). The Brownian motion W is in $C^{1/2-\varepsilon}$ for every $\varepsilon > 0$, and therefore its derivative $\dot{W}$ is in $C^{-1/2-\varepsilon}$. Hence, we are just below the border $\alpha + \beta > 0$, and the product cannot be defined using classical analytical arguments. But the Itô, Stratonovich, Skorokhod, and rough path integral can all be understood as different ways of defining the product, since for any integral we can set

$$F(W_t)\dot{W}_t := \partial_t \int_0^t F(W_s)\mathrm{d}W_s.$$

Because of this natural link between integration and multiplication, it is fairly well understood how to multiply a function of one index with a derivative of a function of one index variable.

But for functions of several parameters, things get more complicated. In that case the link between integration and multiplication is not so clear any more, and therefore there are much fewer techniques available for defining the product of two tempered distributions on $\mathbb{R}^d$.

The results of Chapter 4 have a natural correspondence in terms of Littlewood-Paley blocks. More precisely, every tempered distribution on $\mathbb{R}^d$ can be decomposed with the help of *Littlewood-Paley blocks* into an infinite sum of smooth functions,

$$f = \sum_{m=-1}^{\infty} \Delta_m f,$$

where $\Delta_m f$ is infinitely often differentiable for every m. The decay of the L^∞–norm of the Littlewood-Paley blocks (Δ_m) determines the regularity of f, just as the decay of the Schauder coefficients determines the regularity of functions on $[0,1]$. If now f and g are two tempered distributions, then we formally set $fg := \sum_{m,n} \Delta_m f \Delta_n g$. Every term of this double series is well defined and it remains to study its convergence. As Bony [Bon81] observed, the series can be decomposed into terms with different behavior, just as the double series in Chapter 4:

$$fg = \pi_<(f,g) + \pi_>(f,g) + \pi_\circ(f,g),$$

where

$$\pi_<(f,g) = \sum_{m<n-1} \Delta_m f \Delta_n g, \qquad \pi_>(f,g) = \sum_{n<m-1} \Delta_m f \Delta_n g, \text{ and}$$
$$\pi_\circ(f,g) = \sum_{|m-n|\le 1} \Delta_m f \Delta_n g.$$

The terms $\pi_<$ and $\pi_>$ are always well defined and inherit the regularity of f and g, respectively. The term $\pi_\circ$ is only well defined if $f \in C^\alpha$ and $g \in C^\beta$, and $\alpha + \beta > 0$. In that case it is in $C^{\alpha+\beta}$.

Let us assume from now on that $\alpha > 0$ but $\beta < 0$. Then $\pi_<(f,g) \in C^\beta$, but the other terms are in $C^{\alpha+\beta}$ (if they are defined), and therefore they are more regular. In other words, the product fg is in C^β, but $fg - \pi_<(f,g) \in C^{\alpha+\beta}$. Similarly, if F is a smooth function, then $F(f) \in C^\alpha$, but $F(f) - \pi_<(\mathrm{D}F(f), f) \in C^{2\alpha}$. This is the content of Bony's paralinearization theorem. In the special case when $f = F(W)$ and $g = \dot W$, we obtain that $F(W) - \pi_<(\mathrm{D}F(W), W) \in C^{2\alpha}$ for all $\alpha \in (1/3, 1/2)$. We conclude that $\pi_\circ(F(W) - \pi_<(\mathrm{D}F(W), W), \dot W)$ is well defined and in $C^{3\alpha-1}$. Therefore, the term $F(W)\dot W$ can be defined if and only if $\pi_\circ(\pi_<(\mathrm{D}F(W), W), \dot W)$ can be defined. Here we prove again a commutator estimate, where we show that

$$R(F(W), W, \dot W) := \pi_\circ(\pi_<(\mathrm{D}F(W), W), \dot W) - \mathrm{D}F(W)\pi_\circ(W, \dot W)$$

is a bounded trilinear operator on $C^\alpha \times C^\alpha \times C^{\alpha-1}$. Just as in Chapter 4, we see that the problem of constructing $\pi_\circ(F(W), \dot W)$ reduces to the problem of constructing $\pi_\circ(W, \dot W)$. This extends from $F(W)$ to controlled distributions, that are defined analogously to Chapter 4, and we obtain the continuity of the product operator in suitable topologies.

The advantage of the formulation in terms of Littlewood-Paley blocks is that now the results apply for distributions on $\mathbb{R}^d$ for arbitrary $d \ge 1$, and not just for functions of one index variable. Also, we do not require that the second factor (i.e. $\dot W$) is a derivative, which was necessary to make the connection between products and integrals.

Thus, we developed a robust new way of multiplying two tempered distributions with each other. We apply our product to solve two nonlinear SPDEs. The first equation is maybe not very relevant for practical applications, but it is a perfect test bed for our techniques. We consider a multidimensional fractional Burgers type equation,

$$\partial_t u(t,x) = -(-\Delta)^\sigma u(t,x) + G(u(t,x))\mathrm{D}_x u(t,x) + \dot W(t,x), \tag{0.1}$$

where $u : [0,T] \times [-\pi,\pi]^d \to \mathbb{R}^n$, the operator $-(-\Delta)^\sigma$ is the fractional Laplacian with $\sigma > d/2 + 1/3$, the map $G : \mathbb{R}^n \to \mathcal{L}(\mathcal{L}([-\pi,\pi]^d, \mathbb{R}^n), \mathbb{R}^n)$ is smooth, D_x denotes the spatial derivative, and $\dot W(t,x)$ is a space-time white noise. The critical term is $G(u(t,x))\mathrm{D}_x u(t,x)$. We will show that the solution v to

$$\partial_t v(t,x) = -(-\Delta)^\sigma v(t,x) + \dot W(t,x)$$

satisfies $v \in C([0,T], C^\alpha([-\pi,\pi]^d, \mathbb{R}^n))$ for all $\alpha < \sigma - d/2$. We would expect u to have the same regularity as v. So if $\sigma - d/2 \leq 1/2$, then it is not possible to define the product $G(u(t,x))\mathrm{D}_x u(t,x)$ using classical arguments. But using our newly developed techniques, the product is well defined, and we can show that there exists a unique solution to (0.1). By the continuity of our product operator, it follows automatically that the solution to (0.1) depends continuously on the driving noise in a suitable topology.

The second equation that we study is a nonlinear version of the parabolic Anderson model,

$$\partial_t u(t,x) = \Delta u(t,x) + F(u(t,x))\dot{W}(x),$$

where $u : [0,T] \times [-\pi,\pi]^2 \to \mathbb{R}$, the map $F : \mathbb{R} \to \mathbb{R}$ is smooth, and $\dot{W}(x)$ is a spatial white noise. As we will see, the natural spatial regularity of the solution is $u(t,\cdot) \in C^\alpha([-\pi,\pi]^2,\mathbb{R})$ for $t \in [0,T]$ and $\alpha < 1$. Since the white noise satisfies $\dot{W} \in C^{-1-\varepsilon}$ for every $\varepsilon > 0$, the product $F(u(t,x))\dot{W}(x)$ cannot be defined by classical arguments. Here we will again use our techniques to give a meaning to the solution and to show that it depends continuously on the driving noise in a suitable topology.

1. Dominating local martingale measures and arbitrage under information asymmetry

In this chapter we study financial market models that may allow for arbitrage opportunities. We identify (NA1) as the minimal property that has to be satisfied by any reasonable asset price model, and we show that (NA1) is relatively robust under filtration enlargements. We show that a locally bounded stochastic process S satisfies (NA1) if and only if there exists a dominating measure Q such that S is a Q–local martingale.

1.1. Setting and main results

It may be argued that the foundation of financial mathematics consists in giving a mathematical characterization of market models satisfying certain financial axioms. This leads to so-called *fundamental theorems of asset pricing.* Harrison and Pliska [HP81] were the first to observe that, on finite probability spaces, the absence of arbitrage opportunities (condition *no arbitrage, (NA)*) is equivalent to the existence of an equivalent martingale measure. A definite version was shown by Delbaen and Schachermayer [DS94]. Their result, commonly referred to as *the* Fundamental Theorem of Asset Pricing, states that for locally bounded semimartingale models there exists an equivalent probability measure under which the price process is a local martingale, if and only if the market satisfies the condition *no free lunch with vanishing risk (NFLVR).* Delbaen and Schachermayer also observed that (NFLVR) is satisfied if and only if there are no arbitrage opportunities (i.e. (NA) holds), and if further it is not possible to make an unbounded profit with bounded risk (we say there are *no arbitrage opportunities of the first kind*, condition *(NA1)* holds). Since in finite discrete time, (NA) is equivalent to the existence of an equivalent martingale measure, it was then a natural question how to characterize continuous time market models satisfying only (NA) and not necessarily (NA1). For continuous price processes, this was achieved by Delbaen and Schachermayer [DS95b], who show that (NA) implies the existence of an *absolutely continuous* local martingale measure.

Here we complement this program, by proving that for locally bounded processes, (NA1) is equivalent to the existence of a *dominating* local martingale measure. Constructing dominating probability measures is rather delicate, and Föllmer's measure ([Föl72]) associated to a nonnegative supermartingale appears naturally in this context.

Let us give a more precise description of the notions of arbitrage considered in this work, and of the obtained results.

Let $S = (S_t)_{t \geq 0}$ be a d–dimensional stochastic process on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$. We assume throughout this chapter that the filtration $(\mathcal{F}_t)$ is right-continuous, and that $\mathcal{F} = \mathcal{F}_\infty = \bigvee_{t\geq 0} \mathcal{F}_t$. We think of S as the (discounted) price process of d financial assets. We consider an infinite time horizon, because this leads to more general results. The case of a finite time horizon $T > 0$ can easily be embedded in this context, by setting $\mathcal{F}_{T+t} = \mathcal{F}_T$ and $S_{T+t} = S_T$ for all $t \geq 0$.

Semimartingales are defined as usually, except that they are only almost surely (a.s.) càdlàg. A semimartingale does not need to be càdlàg for every $\omega \in \Omega$. The reason for this is that we do not assume our filtration to be complete, since our aim is to construct dominating measures which may charge P–null sets. We argue in Appendix A that the incompleteness of our filtration will not pose any problem.

A *strategy* is a predictable process $H = (H_t)_{t\geq 0}$ with values in $\mathbb{R}^d$. If S is a semimartingale and $\lambda > 0$, then a strategy H is called λ*–admissible* (for S) if the stochastic integral $H \cdot S := \int_0^\cdot H_s \mathrm{d}S_s$ exists and satisfies $P((H \cdot S)_t \geq -\lambda) = 1$ for all $t \geq 0$. Here we write $xy = \sum_{k=1}^d x_k y_k$ for the usual inner product on $\mathbb{R}^d$. We define $\mathcal{H}_\lambda$ as the set of all λ–admissible strategies. For details about vector stochastic integration we refer to Jacod and Shiryaev [JS03], Section III.6.

If S is only adapted and right-continuous, and not necessarily a semimartingale, then we can still integrate simple strategies against S. A *simple strategy* is a process of the form $H_t = \sum_{j=0}^{m-1} F_k 1_{(\tau_k,\tau_{k+1}]}(t)$ for $m \in \mathbb{N}$ and stopping times $0 \leq \tau_0 < \tau_1 < \cdots < \tau_m < \infty$, where for every $0 \leq k < m$ the random variable F_k is bounded and $\mathcal{F}_{\tau_k}$–measurable and takes its values in $\mathbb{R}^d$. If S is a right-continuous adapted process, then the integral $H \cdot S$ is defined as

$$(H \cdot S)_t = \sum_{k=0}^{m-1} F_k (S_{\tau_{k+1}\wedge t} - S_{\tau_k \wedge t}),$$

and λ–admissible strategies are defined analogously to the semimartingale case. We denote the set of simple λ–admissible strategies by $\mathcal{H}_{\lambda,s}$.

The set $\mathcal{W}_1$ consists of all wealth processes obtained by using 1–admissible strategies under initial wealth 1, and such that the terminal wealth is well defined, i.e.

$$\mathcal{W}_1 = \{1 + H \cdot S : H \in \mathcal{H}_1 \text{ and } (H \cdot S)_t \text{ a.s. converges as } t \to \infty\}. \tag{1.1}$$

Similarly $\mathcal{W}_{1,s}$ is defined as

$$\mathcal{W}_{1,s} = \{1 + H \cdot S : H \in \mathcal{H}_{1,s}\}.$$

Note that the convergence condition in (1.1) is trivially satisfied for simple strategies. We will also need $\mathcal{K}_1$, the set of terminal wealths that are attainable with initial wealth 1 and using 1–admissible strategies:

$$\mathcal{K}_1 = \{X_\infty : X \in \mathcal{W}_1\} \qquad \text{and} \qquad \mathcal{K}_{1,s} = \{X_\infty : X \in \mathcal{W}_{1,s}\}. \tag{1.2}$$

We write $L^0 = L^0(\Omega, \mathcal{F}, P)$ for the space of real-valued random variables on $(\Omega, \mathcal{F})$, where we identify random variables that are P–almost surely equal. We equip L^0 with the distance $d(X,Y) = E(|X-Y| \wedge 1)$, under which it becomes a complete metric space.

Recall that a family of random variables $\mathcal{X}$ is called *bounded in probability*, or *bounded in L^0*, if

$$\lim_{m\to\infty} \sup_{X\in\mathcal{X}} P(|X| \geq m) = 0.$$

Definition 1.1.1. We say that a semimartingale S satisfies *no arbitrage of the first kind (NA1)* if $\mathcal{K}_1$ is bounded in probability. We say that S satisfies *no arbitrage (NA)* if there is no $X \in \mathcal{K}_1$ with $X \geq 1$ and $P(X > 1) > 0$. If both (NA1) and (NA) hold, we say that S satisfies *no free lunch with vanishing risk (NFLVR).*

Similarly we say that a right-continuous adapted process S satisfies *no arbitrage of the first kind with simple strategies (NA1$_s$)*, *no arbitrage with simple strategies (NA$_s$)*, or *no free lunch with vanishing risk with simple strategies (NFLVR$_s$)*, if $\mathcal{K}_{1,s}$ satisfies the corresponding conditions.

Heuristically, (NA) says that it is not possible to make a profit without taking a risk. (NA1) says that is not possible to make unbounded profit if the risk remains bounded. This is why (NA1) is also referred to as "no unbounded profit with bounded risk" (NUPBR), see for example Karatzas and Kardaras [KK07].

The main result of this chapter is that for locally bounded semimartingales S, (NA1) is equivalent to the existence of a dominating local martingale measure. As a byproduct of the proof, we obtain that a locally bounded, right-continuous, and adapted process S that satisfies (NA1$_s$) is already a semimartingale, and in this case S also satisfies (NA1).

When constructing absolutely continuous probability measures, it suffices to work with random variables. In Section 1.2 below, we argue that dominating measures correspond to nonnegative supermartingales with strictly positive terminal values. We also show that a dominating local martingale measure corresponds to a supermartingale density in the following sense.

Definition 1.1.2. Let $\mathcal{Y}$ be a family of stochastic processes. A *supermartingale density* for $\mathcal{Y}$ is an almost surely càdlàg and nonnegative supermartingale Z with $Z_\infty = \lim_{t\to\infty} Z_t > 0$, such that YZ is a supermartingale for every $Y \in \mathcal{Y}$.

If all processes in $\mathcal{Y}$ are of the form $1 + (H \cdot S)$ for suitable integrands H, and if Z is a supermartingale density for $\mathcal{Y}$, then we will sometimes call Z a supermartingale density for S.

In the literature, supermartingale densities are usually referred to as *supermartingale deflators*. We think of a supermartingale density as the "Radon-Nikodym derivative" $\mathrm{d}Q/\mathrm{d}P$ of a dominating measure $Q \gg P$. This is why we prefer the term supermartingale density.

First we give an alternative proof of a well-known result.

Theorem 1.1.3. *Let S be a d–dimensional adapted process, almost surely right-continuous (respectively a d–dimensional semimartingale). Then (NA1$_s$) (respectively (NA1)) holds if and only if there exists a supermartingale density for $\mathcal{W}_{1,s}$ (respectively for $\mathcal{W}_1$).*

As a consequence, (NA1$_s$) implies the semimartingale property for locally bounded processes.

Corollary 1.1.4. *Let S be a d–dimensional adapted process, almost surely right-continuous. If every component S^i of $S = (S^1, \dots, S^d)$ is locally bounded from below and if S satisfies (NA1$_s$), then S is a semimartingale that satisfies (NA1), and any supermartingale density for $\mathcal{W}_{1,s}$ is also a supermartingale density for $\mathcal{W}_1$.*

Given a supermartingale density Z for S, we then apply Yoeurp's [Yoe85] results on Föllmer's measure [Föl72], to construct a dominating measure $Q \gg P$ associated to Z. Let γ be a right-continuous version of the density process $\gamma_t = \mathrm{d}P/\mathrm{d}Q|_{\mathcal{F}_t}$, and let τ be the first time that γ hits zero, $\tau = \inf\{t \geq 0 : \gamma_t = 0\}$. We define

$$S_t^{\tau-} = S_t 1_{\{t<\tau\}} + S_{\tau-} 1_{\{t \geq \tau\}} = S_t 1_{\{t<\tau\}} + \lim_{s \to \tau-} S_s 1_{\{t \geq \tau\}}.$$

Note that S and $S^{\tau-}$ are P–indistinguishable.

In the predictable case we then obtain the following result.

Theorem 1.1.5. *Let S be a predictable semimartingale. If Z is a supermartingale density for $\mathcal{W}_1$, then Z determines a probability measure $Q \gg P$ such that $S^{\tau-}$ is a Q–local martingale. Conversely, if $Q \gg P$ is a dominating local martingale measure for $S^{\tau-}$, then $\mathcal{W}_1$ admits a supermartingale density.*

Actually the current formulation is slightly too simple, we will need to impose topological conditions on the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0})$, and also we will need to enlarge the space. This will be described in more detail in Section 1.4.

Theorem 1.1.5 is false if S is not predictable, as we will demonstrate by a simple example. But in the non-predictable case we are able to exhibit a subset of supermartingale densities that do give rise to dominating local martingale measures. Conversely, every dominating local martingale measure for $S^{\tau-}$ corresponds to a supermartingale density, even for processes that are not predictable. Therefore, the following theorem, the main result of this chapter, is valid for all locally bounded processes that are adapted and almost surely right-continuous. In the non-predictable case we build on results of [Tak13] that are only formulated for processes on finite time intervals. So in the theorem we let $T_\infty = \infty$ if S is predictable, and $T_\infty \in (0, \infty)$ otherwise.

Theorem 1.1.6. *Let $(S_t)_{t \in [0, T_\infty]}$ be a locally bounded, adapted process, that is almost surely right-continuous. Then S satisfies (NA1$_s$) if and only if there exists a dominating $Q \gg P$, such that $S^{\tau-}$ is a Q–local martingale.*

This work is motivated by insights from the theory of filtrations enlargements. A filtration $(\mathcal{G}_t)$ is called *filtration enlargement* of $(\mathcal{F}_t)$ if $\mathcal{G}_t \supseteq \mathcal{F}_t$ for all $t \geq 0$. A basic question is then under which conditions all members of a given family of $(\mathcal{F}_t)$–semimartingales are $(\mathcal{G}_t)$–semimartingales. We say that *Hypothèse* (H') is satisfied if all $(\mathcal{F}_t)$–semimartingales are $(\mathcal{G}_t)$–semimartingales. Given a $(\mathcal{F}_t)$–semimartingale that satisfies (NFLVR), i.e. for which there exists an equivalent local martingale measure, one might also ask under which conditions it still satisfies (NFLVR) under $(\mathcal{G}_t)$. It is well known, and we illustrate this in an example below, that the (NFLVR) condition is usually violated after filtration enlargements.

However, we will see that (NA1) is relatively stable under filtration enlargements. If $(\mathcal{G}_t)$ is an initial enlargement of $(\mathcal{F}_t)$, i.e. $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(X)$ for some random variable X and for every $t \geq 0$, then Jacod's criterion [Jac85] is a celebrated condition on X and $(\mathcal{F}_t)$ under which Hypothèse (H') is satisfied. We show that in fact Jacod's criterion implies the existence of a universal supermartingale density. A strictly positive process Z is called *universal supermartingale density* if ZM is a $(\mathcal{G}_t)$–supermartingale for every nonnegative $(\mathcal{F}_t)$–supermartingale M. The existence of Z is much stronger than Hypothèse (H'), and in particular it shows that every process satisfying (NA1) under $(\mathcal{F}_t)$ also satisfies (NA1) under $(\mathcal{G}_t)$.

We also show that if $(\mathcal{G}_t)$ is a general (not necessarily initial) filtration enlargement of $(\mathcal{F}_t)$, and if there exists a universal supermartingale density for $(\mathcal{G}_t)$, then a generalized version of Jacod's criterion is necessarily satisfied.

Section 1.2 describes the link to filtration enlargements in more detail. In Section 1.2 we also argue that a dominating local martingale measure should correspond to a supermartingale density. In Section 1.3 we prove that the existence of supermartingale densities is equivalent to ($\mathrm{NA1}_s$). In Section 1.4 we prove that if S is predictable, then Z is a supermartingale density for S if and only if $S^{\tau-}$ is a local martingale under the Föllmer measure of Z. We also prove our main result, Theorem 1.1.6, for general locally bounded processes (not necessarily predictable). In Section 1.5 we return to filtration enlargements and examine how Jacod's criterion relates to our results.

Relevant literature

Supermartingale densities were first considered by Kramkov and Schachermayer [KS99] and Becherer [Bec01].

The semimartingale case of Theorem 1.1.3 was shown by Karatzas and Kardaras [KK07]. Their proof extensively uses the semimartingale characteristics of S, and can therefore not be applied to general processes satisfying ($\mathrm{NA1}_s$). Note that Corollary 1.1.4 states that any locally bounded process satisfying ($\mathrm{NA1}_s$) is a semimartingale. But for unbounded processes this is no longer true, as we shall demonstrate in a simple counterexample. A more general result than Theorem 1.1.3 is shown in Rokhlin [Rok10], using arguments that are related to our proof. In fact our arguments are powerful enough to imply the results of [Rok10]. We were not aware of either of these works before completing our proof. We believe that our proof gives a nice application of *convex compactness*, as introduced by Žitković [Ž10]. Oversimplifying things a bit, one can understand convex compactness as an elegant way of formalizing convergence and compactness results that are usually shown by ad-hoc considerations based on results like Lemma A1.1 of [DS94]. We also believe that our techniques may be interesting in more complicated contexts, say under transaction costs, where arbitrage considerations no longer imply the semimartingale property of the price process.

It is well known that a locally bounded process satisfying (NA1$_s$) must be a semimartingale, see Ankirchner's Ph.D. thesis [Ank05], Theorem 7.4.3, and also Kardaras and Platen [KP11]. See also [DS94] for a first result in this direction. This part of Corollary 1.1.4 is an immediate consequence of Theorem 1.1.3. We rely on [KP11] to obtain that (NA1$_s$) implies (NA1) for locally bounded processes, and that in this case supermartingale densities for $\mathcal{W}_{1,s}$ are supermartingale densities for $\mathcal{W}_1$.

Recently there has been an increased interest in Föllmer's measure, motivated by problems from mathematical finance. Föllmer's measure appears naturally in the construction and study of *strict local martingales*, i.e. local martingales that are not martingales. These are used to model bubbles in financial markets, see Jarrow, Protter, and Shimbo [JPS10]. A pioneering work on the relation between Föllmer's measure and strict local martingales is Delbaen and Schachermayer [DS95a]. Other references are Pal and Protter [PP10] and Kardaras, Kreher, and Nikeghbali [KKN11]. The work most related to ours is Ruf [Ruf13], where it is shown that, in a diffusion setting, (NA1) implies the existence of a dominating local martingale measure. All these works have in common that they study Föllmer measures of strictly positive local martingales. Carr, Fisher and Ruf [CFR12] study the Föllmer measure of a local martingale which is not strictly positive.

To the best of our knowledge, the current work is the first in which the Föllmer measure of a supermartingale which is not a local martingale is used as a local martingale measure. In Föllmer and Gundel [FG06], supermartingales Z are associated to "extended martingale measures" P^Z. But by definition, P^Z is an extended martingale measure if and only if Z is a supermartingale density. This does not obviously imply that $S^{\tau-}$ or S is a local martingale under P^Z – and in general this is not true. Here we show that if S is predictable, then any supermartingale density Z corresponds to a dominating local martingale measure P^Z – meaning that $S^{\tau-}$ is a local martingale under P^Z. For non-predictable S we give a counterexample. In that case we identify a subset of supermartingale densities that do correspond to local martingale measures.

Another related work is Kardaras [Kar10a], where it is shown that (NA1) is equivalent to the existence of a finitely additive equivalent local martingale measure. Here we construct countably additive measures, that are not equivalent but only dominating.

The main motivation for this work comes from the theory of filtrations enlargements, see for example Amendinger, Imkeller and Schweizer [AIS98], Ankirchner's Ph.D. thesis [Ank05], and Ankirchner, Dereich and Imkeller [ADI06]. In these works it is shown that if M is a continuous local martingale in a given filtration $(\mathcal{F}_t)$, then under an enlarged filtration $(\mathcal{G}_t)$, assuming suitable conditions, M is of the form $M = \widetilde{M} + \int_0^\cdot \alpha_s d\langle \widetilde{M} \rangle_s$, where $\widetilde{M}$ is a $(\mathcal{G}_t)$–local martingale. It is then a natural question whether there exists an equivalent measure Q that "eliminates" the drift, i.e. under which M is a $(\mathcal{G}_t)$–local martingale. In general, the answer to this question is negative. However, Ankirchner [Ank05], Theorem 9.2.7, observed that if there exists a well-posed utility maximization problem in the large filtration, then the *information drift* α must be locally square integrable with respect to $\widetilde{M}$. Here we show that for continuous processes, the square integrability of the information drift is equivalent to the well-posedness of a utility maximization problem in the large filtration, we relate these conditions to (NA1), and we show that this allows us to construct dominating local martingale measures. We also give the corresponding results for discontinuous processes.

1.2. Motivation

In this section we show that the (NFLVR) property is not very robust under filtration enlargements. Then we recall that if Jacod's criterion is satisfied, there still is a dominating local martingale measure. Finally we argue that under Jacod's criterion, (NA1) is often satisfied in the large filtration. We hope that this convinces the reader that (NA1) respectively ($\mathrm{NA1}_s$) should be related to the existence of dominating local martingale measures. Assuming that a dominating local martingale measure exists, we examine its Kunita-Yoeurp decomposition under P, and we see that it corresponds to a supermartingale density.

Equivalent local martingale measures and filtration enlargements

Consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ with $P(A) \in \{0,1\}$ for all $A \in \mathcal{F}_0$. Define $\mathcal{F}_\infty = \bigvee_{t\geq 0} \mathcal{F}_t$. Let S be a one dimensional semimartingale that describes a complete market (i.e. for every $X \in L^\infty(\mathcal{F}_\infty)$ there exists a predictable process H, integrable with respect to S, such that $X = X_0 + \int_0^\infty H_s \mathrm{d}S_s$ for some constant $X_0 \in \mathbb{R}$). Let X be a random variable that is $\mathcal{F}_\infty$–measurable. Assume that X is not P–almost surely constant. Define the initially enlarged filtration $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(X)$ for $t \geq 0$. This is a toy model for insider trading. At time 0, the insider has the additional knowledge of the value of X. Since X is not constant, there exists $A \in \sigma(X)$ such that $P(A) \in (0,1)$. Assume Q is an equivalent $(\mathcal{G}_t)$–local martingale measure for S. Consider the $(Q, (\mathcal{F}_t))$–martingale $N_t = E_Q(1_A|\mathcal{F}_t)$, for $t \geq 0$. Since the market is complete, 1_A can be replicated. That is, there exists a $(\mathcal{F}_t)$–predictable strategy H such that $N = Q(A) + \int_0^\cdot H_s \mathrm{d}S_s$. But then $\int_0^\cdot H_s \mathrm{d}S_s$ is a bounded $(Q, (\mathcal{G}_t))$–local martingale. Hence, it is a martingale, and since $A^c \in \mathcal{G}_0$, we obtain

$$0 = E_Q(1_{A^c} 1_A) = E_Q\left(1_{A^c}\left(Q(A) + \int_0^\infty H_s \mathrm{d}S_s\right)\right) = Q(A^c)Q(A) > 0,$$

which is absurd. The last step follows because Q was assumed to be equivalent to P.

So already in the simplest models that incorporate information asymmetry, there may not exist an equivalent local martingale measure. If S is locally bounded, then by the Fundamental Theorem of Asset Pricing at least one of the conditions (NA) or (NA1) has to be violated.

Jacod's criterion and dominating local martingale measures

Let $(\mathcal{G}_t)$ be a filtration enlargement of $(\mathcal{F}_t)$, i.e. $\mathcal{F}_t \subseteq \mathcal{G}_t$ for every $t \geq 0$. Let $\mathcal{Y}$ be a family of $(\mathcal{F}_t)$–semimartingales. One of the typical questions in filtration enlargements is under which conditions all $Y \in \mathcal{Y}$ are $(\mathcal{G}_t)$–semimartingales. *Hypothèse* (H') is satisfied if *all* $(\mathcal{F}_t)$–semimartingales are $(\mathcal{G}_t)$–semimartingales.

Jacod's criterion [Jac85] is a famous condition that implies Hypothèse (H'). Here we give an equivalent formulation, first found by Föllmer and Imkeller [FI93] and later generalized and carefully studied by Ankirchner, Dereich and Imkeller [ADI07]. Let X be a random variable and consider the initial enlargement $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(X)$. Define the product space

$$\overline{\Omega} = \Omega \times \Omega, \qquad \overline{\mathcal{G}} = \mathcal{F}_\infty \otimes \sigma(X), \qquad \overline{\mathcal{G}}_t = \mathcal{F}_t \otimes \sigma(X),\, t \geq 0.$$

We define two measures on $\overline{\Omega}$. The decoupling measure $\overline{Q} = P|_{\mathcal{F}_\infty} \otimes P|_{\sigma(X)}$, and $\overline{P} = P \circ \psi^{-1}$, where $\psi : \Omega \to \overline{\Omega}$, $\psi(\omega) = (\omega, \omega)$. In this setting the following result is a reformulation of Jacod's criterion.

Theorem (Theorem 1 in [ADI07])**.** *If $\overline{P} \ll \overline{Q}$, then Hypothèse (H') holds, i.e. any $(\mathcal{F}_t)$–semimartingale is a $(\mathcal{G}_t)$–semimartingale.*

In this formulation it is quite obvious why Jacod's criterion works. Under the measure $\overline{Q}$, the additional information from X is independent of $\mathcal{F}_\infty$. Therefore, any $(\mathcal{F}_t)$–martingale M will stay a $(\overline{\mathcal{G}}_t)$–martingale under $\overline{Q}$ (if we embed M from Ω to $\overline{\Omega}$ by setting $\overline{M}_t(\omega, \omega') = M_t(\omega)$). By assumption, $\overline{Q} \gg \overline{P}$, and therefore an application of Girsanov's theorem implies that $\overline{M}$ is a $\overline{P}$–semimartingale. But it is possible to show that if $\overline{M}$ is a $(\overline{P}, (\overline{\mathcal{G}}_t))$–semimartingale, then M is a $(P, (\mathcal{G}_t))$–semimartingale. This completes the argument.

Thus, Jacod's criterion states that there exists a dominating measure (on an enlarged space), under which any $(\mathcal{F}_t)$–martingale is a $(\mathcal{G}_t)$–martingale.

It is not hard to see that Jacod's criterion is always satisfied if X takes its values in a countable set, regardless of the structure of $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ and S. So if we recall our example of an initial filtration enlargement in a complete market from above, then we observe that Jacod's criterion may be satisfied although there is no equivalent local martingale measure in the large filtration.

Utility maximization and filtration enlargements

There are many articles devoted to calculating the additional utility of an insider. Assume S is a semimartingale in the large filtration $(\mathcal{G}_t)$. Then we define the set of attainable terminal wealths $\mathcal{K}_1(\mathcal{F}_t)$ and $\mathcal{K}_1(\mathcal{G}_t)$ as in (1.2), using $(\mathcal{F}_t)$–predictable and $(\mathcal{G}_t)$–predictable strategies respectively.

If S describes a complete market under $(\mathcal{F}_t)$, and if $(\mathcal{G}_t)$ is an initial enlargement satisfying Jacod's criterion, then it is shown in Ankirchner's Ph.D. thesis ([Ank05], Theorem 12.6.1, see also [ADI06]), that the maximal expected logarithmic utility under $(\mathcal{G}_t)$ is given by

$$\sup_{X \in \mathcal{K}_1(\mathcal{G}_t)} E(\log(X)) = \sup_{X \in \mathcal{K}_1(\mathcal{F}_t)} E(\log(X)) + I(X, \mathcal{F}_\infty),$$

where $I(X, \mathcal{F}_\infty)$ denotes the mutual information between X and $\mathcal{F}_\infty$. The mutual information is often finite, and therefore the maximal expected utility under $(\mathcal{G}_t)$ is often finite. But finite utility and (NA1) are equivalent:

Lemma 1.2.1. *The process S satisfies (NA1) under $(\mathcal{G}_t)$ if and only if there exists an unbounded increasing function U such that the maximal expected utility is finite, i.e. such that*

$$\sup_{X \in \mathcal{K}_1(\mathcal{G}_t)} E(U(X)) < \infty.$$

Proof. This follows from Proposition 1.2.2 below. □

In conclusion, we showed that (NFLVR) and thus (NA) or (NA1) is not very robust with respect to filtration enlargements. We also observed that the maximal expected logarithmic utility in an enlarged filtration may be finite, and that this is only possible under the (NA1) condition. Hence, we conclude that (NA) is the part of (NFLVR) which is less robust with respect to filtration enlargements (see Remark 1.5.4 below for a more detailed discussion). Moreover, Jacod's criterion is satisfied in the examples where (NA1) holds. As we saw above, Jacod's criterion implies the existence of a dominating local martingale measure. Hence, (NA1) seems to be related to the existence of a dominating local martingale measure. Below we prove that the two conditions are equivalent.

Supermartingale densities

Now let us assume that $Q \gg P$ is a dominating local martingale measure for S, and let us examine what type of object this gives us under P. Define γ as the right-continuous density process, $\gamma_t = \mathrm{d}P/\mathrm{d}Q|_{\mathcal{F}_t}$. Then $\tau = \inf\{t \geq 0 : \gamma_t = 0\}$ is a stopping time, and we can define the adapted process $Z_t = 1_{\{t<\tau\}}/\gamma_t$. Let H be 1–admissible for S under Q, i.e. such that $Q(\int_0^t H_s \mathrm{d}S_s \geq -1) = 1$ for all $t \geq 0$. Let $s, t \geq 0$ and let $A \in \mathcal{F}_t$. We have

$$\begin{aligned} E_P(1_A Z_{t+s}(1 + (H \cdot S)_{t+s})) &= E_Q\left(\gamma_{t+s} 1_A \frac{1_{\{t+s<\tau\}}}{\gamma_{t+s}}(1 + (H \cdot S)_{t+s})\right) \\ &\leq E_Q\left(1_A 1_{\{t<\tau\}}(1 + (H \cdot S)_{t+s})\right) \\ &\leq E_Q\left(1_A 1_{\{t<\tau\}}(1 + (H \cdot S)_{t})\right) \\ &= E_P(1_A Z_t(1 + (H \cdot S)_t)) \end{aligned} \tag{1.3}$$

using in the second line that $1_A(1 + (H \cdot S)_{t+s})$ is nonnegative, and in the third line that $1 + (H \cdot S)$ is a nonnegative Q–local martingale and therefore a Q–supermartingale. This indicates that Z should be a supermartingale density. Of course here we only considered strategies that are 1–admissible under Q, and there might be strategies that are 1–admissible under P but not under Q. The solution to this problem is to consider $S^{\tau-}$ rather than S. We will make this rigorous later.

The pair (Z, τ) is the *Kunita-Yoeurp decomposition* of Q with respect to P. The Kunita-Yoeurp decomposition is a progressive Lebesgue decomposition on filtered probability spaces. It was introduced in Kunita [Kun76] in a Markovian context, and generalized to arbitrary filtered probability spaces in Yoeurp [Yoe85]. Namely we have for every $t \geq 0$

1. $P(\tau = \infty) = 1$,
2. $Q(\cdot \cap \{\tau \leq t\})$ and P are mutually singular on $\mathcal{F}_t$,
3. for $A \in \mathcal{F}_t$ we have $Q(A \cap \{\tau > t\}) = E_P(1_A Z_t)$.

Note that the second property is a consequence of the first property.

Hence, our program will be to find a supermartingale density Z, and to construct a measure Q and a stopping time τ, such that (Z, τ) is the Kunita-Yoeurp decomposition of Q with respect to P. But the second part was already solved by [Yoe85], and Q will be the Föllmer measure of Z. After studying the relation between S and Z, we will see that $S^{\tau-}$ is a local martingale under Q.

Before starting to construct a supermartingale density, let us prove Lemma 1.2.1, which is an immediate consequence of the following de la Vallée-Poussin type result for families of random variables that are bounded in L^0.

Proposition 1.2.2. *A family of random variables $\mathcal{X}$ is bounded in probability if and only if there exists a nondecreasing and unbounded function U on $[0, \infty)$, such that*

$$\sup_{X \in \mathcal{X}} E(U(|X|)) < \infty.$$

In that case U can be chosen strictly increasing, concave, and such that $U(0) = 0$.

Proof. First, assume that such a U exists. Then we have for $m \in \mathbb{N}$

$$\sup_{X \in \mathcal{X}} P(|X| \geq m) \leq \sup_{X \in \mathcal{X}} P(U(|X|) \geq U(m)) \leq \frac{\sup_{X \in \mathcal{X}} E(U(|X|))}{U(m)}.$$

Since U is unbounded, the right hand side converges to zero as m tends to ∞.

Conversely, assume that $\mathcal{X}$ is bounded in probability. We need to construct a strictly increasing, unbounded, and concave function U with $U(0) = 0$, such that $E(U(|X|))$ is bounded for X running through $\mathcal{X}$. Our construction is inspired by the proof of de la Vallée-Poussin's theorem. That is, we will construct a function U of the form

$$U(x) = \int_0^x g(y)dy, \qquad \text{where} \qquad g(y) = g_m \text{ for } y \in [m-1, m),\ m \in \mathbb{N},$$

for a decreasing sequence of strictly positive numbers (g_m). This U will be strictly increasing, concave, and such that $U(0) = 0$. It will be unbounded if and only if $\sum_{m=1}^{\infty} g_m = \infty$.

For U of this form we have by monotone convergence and Fubini (since all terms are nonnegative)

$$\begin{aligned} E(U(|X|)) &= \sum_{m=1}^{\infty} E(U(|X|)1_{\{|X|\in[m-1,m)\}}) \leq \sum_{m=1}^{\infty} U(m)P(|X|\in[m-1,m)) \\ &= \sum_{m=1}^{\infty}\sum_{k=1}^{m} g_k P(|X|\in[m-1,m)) = \sum_{k=1}^{\infty}\sum_{m=k}^{\infty} g_k P(|X|\in[m-1,m)) \\ &= \sum_{k=1}^{\infty} g_k P(|X|\geq k-1) \leq \sum_{k=1}^{\infty} g_k F_{\mathcal{X}}(k-1), \end{aligned}$$

where $F_{\mathcal{X}}(k-1) = \sup_{X\in\mathcal{X}} P(|X|\geq k-1)$.

So the proof is complete if we can find a decreasing sequence (g_k) of positive numbers, such that $\sum_{k=1}^{\infty} g_k = \infty$ but $\sum_{k=1}^{\infty} g_k F_{\mathcal{X}}(k-1) < \infty$. Let $m \in \mathbb{N}$. By assumption, $(F_{\mathcal{X}}(k))$ converges to zero as k tends to ∞, and therefore it also converges to zero in the Cesàro sense. Hence, we obtain for large enough K_m that

$$\frac{1}{K_m}\sum_{k=1}^{K_m} F_{\mathcal{X}}(k-1) \leq \frac{1}{m}. \tag{1.4}$$

We choose an increasing sequence of numbers $(K_m)_{m\in\mathbb{N}}$, such that $K_m \geq m$ for all m, and such that every K_m satisfies (1.4). Define

$$g_k^m = \begin{cases} \frac{1}{mK_m}, & k \leq K_m, \\ 0, & k > K_m, \end{cases}$$

and let m_k denote the smallest m for which $g_k^m \neq 0$, i.e.

$$m_k := \min\{m \in \mathbb{N} : K_m \geq k\}.$$

By definition, $m_k \leq m_{k+1}$ for all k, and therefore the sequence (g_k), where

$$g_k = \sum_{m=1}^{\infty} g_k^m = \sum_{m=m_k}^{\infty} \frac{1}{mK_m} \leq \sum_{m=m_k}^{\infty} \frac{1}{m^2} < \infty,$$

is decreasing in k. Moreover, Fubini's theorem implies that

$$\sum_{k=1}^{\infty} g_k = \sum_{k=1}^{\infty}\sum_{m=1}^{\infty} g_k^m = \sum_{m=1}^{\infty}\sum_{k=1}^{\infty} g_k^m = \sum_{m=1}^{\infty}\sum_{k=1}^{K_m} \frac{1}{mK_m} = \sum_{m=1}^{\infty}\frac{1}{m} = \infty,$$

and at the same time we get from (1.4)

$$\sum_{k=1}^{\infty} g_k F_{\mathcal{X}}(k-1) = \sum_{m=1}^{\infty}\sum_{k=1}^{K_m} \frac{F_{\mathcal{X}}(k-1)}{mK_m} \leq \sum_{m=1}^{\infty}\frac{1}{m^2} < \infty,$$

which completes the proof. □

Remark 1.2.3. In Loewenstein and Willard [LW00], Theorem 1, it is shown that the utility maximization problem for Itô processes is well posed if and only if there is absence of a certain notion of arbitrage. They describe the critical arbitrage opportunities very precisely, and they consider more general utility maximization problems, allowing for intermediate consumption. Proposition 1.2.2 is much simpler and more obvious, but therefore also more robust. It is applicable in virtually any context, say to discontinuous price processes that are not semimartingales, with transaction costs, and under trading constraints. The family of portfolios need not even be convex.

Remark 1.2.4. Note that supermartingale densities are the dual variables in the duality approach to utility maximization, see [KS99]. Taking Proposition 1.2.2 into account, Theorem 1.1.3 therefore states that there exists a well posed utility maximization problem if and only if the space of dual minimizers is nonempty. This insight might also be useful in more complicated contexts, say in markets with transaction costs. As a sort of meta-theorem holding for many utility maximization problems, we expect that the space of dual variables is nonempty if and only if the space of primal variables is bounded in probability.

A first consequence is that any locally bounded process satisfying ($\mathrm{NA1}_s$) is a semimartingale.

Corollary 1.2.5. *Let S be a locally bounded, càdlàg process satisfying ($NA1_s$). Then S is a semimartingale.*

Proof. Since $\mathcal{K}_{1,s}$ is bounded in probability, Proposition 1.2.2 implies that there exists an unbounded utility function U for which $\sup_{X\in\mathcal{K}_{1,s}} E(U(X)) < \infty$. It then follows from Theorem 7.4.3 of [Ank05] that S is a semimartingale. □

This result will also follow from Theorem 1.1.3.

1.3. Existence of supermartingale densities

Now let us prove Theorem 1.1.3. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space with a right-continuous filtration. We do not require $(\mathcal{F}_t)$ to be complete, contrary to the long tradition in probability theory of only working with filtrations satisfying the usual conditions. In Appendix A there is a detailed discussion with which we hope to convince the reader that the use of incomplete filtrations does not pose any problems.

Note that Jacod [Jac79] and Jacod and Shiryaev [JS03] work without complete filtrations, as far as this is possible. See for example the discussion on page 8 of [Jac79], or Definition I.1.2 of [JS03]. Whenever we quote a result that is not from [Jac79] or [JS03], we point out why it also holds in incomplete filtrations.

After this work was finished, Alexander Gushchin pointed us to a paper by Rokhlin [Rok10], that gives a related (but not identical) proof for a stronger result than Theorem

1.1.3. In fact our arguments imply this stronger result as well. So here we prove Rokhlin's result, and as a corollary we obtain Theorem 1.1.3.

A family of nonnegative stochastic processes $\mathcal{Y}$ is called *fork-convex*, see [Ž02] or [Rok10], if every $Y \in \mathcal{Y}$ stays in zero once it hits zero, i.e. $Y_s = 0$ implies $Y_t = 0$ for all $0 \le s \le t < \infty$, and if further for all $Y^1, Y^2, Y^3 \in \mathcal{Y}$, for all $s \ge 0$, and for all $\mathcal{F}_s$–measurable random variables λ_s with values in $[0,1]$, we have that

$$Y_\cdot = 1_{[0,s)}(\cdot) Y^1_s + 1_{[s,\infty)}(\cdot) Y^1_s \left(\lambda_s \frac{Y^2_\cdot}{Y^2_s} + (1-\lambda_s) \frac{Y^3_\cdot}{Y^3_s} \right) \in \mathcal{Y}. \tag{1.5}$$

Recall that we interpret $0/0 = 0$. Note that a fork-convex family of processes with $Y_0 = 1$ for all $Y \in \mathcal{Y}$ is convex. If moreover $\mathcal{Y}$ contains the constant process 1, then $\mathcal{Y}$ is stable under stopping at deterministic times, i.e. for all $Y \in \mathcal{Y}$ and for all $t \ge 0$ also $Y_{\cdot \wedge t} \in \mathcal{Y}$.

Rokhlin's [Rok10] main result is the following.

Theorem 1.3.1. *Let $\mathcal{Y}$ be a fork-convex family of right-continuous and nonnegative processes containing the constant process 1 and such that $Y_0 = 1$ for all $Y \in \mathcal{Y}$. Let*

$$\mathcal{K} = \left\{ Y_\infty : Y \in \mathcal{Y}, Y_\infty = \lim_{t \to \infty} Y_t \text{ exists} \right\}.$$

Then $\mathcal{K}$ is bounded in probability if and only if there exists a supermartingale density for $\mathcal{Y}$.

We split up the proof in several lemmas.

Lemma 1.3.2. *Let $\mathcal{X}$ be a convex family of nonnegative random variables. Then $\mathcal{X}$ is bounded in probability if and only if there exists a strictly positive random variable Z such that*

$$\sup_{X \in \mathcal{X}} E(XZ) < \infty.$$

Proof. The sufficiency is Theorem 1 of [Yan80]. Note that Yan does not require the σ–algebra to be complete. Yan makes the additional assumption that $\mathcal{X}$ is contained in L^1. But since we are considering nonnegative random variables, this can be avoided by applying Theorem 1 of [Yan80] to the convex hull of the bounded random variables $\{X \wedge n\}$ for $n \in \mathbb{N}$, as suggested in Remark (c) of [DM82], VIII-84.

Conversely, let us assume that Z exists. Normalizing by $E(Z)$, we obtain an equivalent probability measure Q such that $\mathcal{X}$ is norm bounded in $L^1(Q)$ and therefore bounded in Q–probability. Since $P \ll Q$, it is easy to see that $\mathcal{X}$ is also bounded in P–probability. $\square$

Remark 1.3.3. Convexity is necessary: Let $\{A^m_k : 1 \le k \le 2^m, m \in \mathbb{N}\}$ be an increasing sequence of partitions of Ω, such that $P(A^m_k) = 2^{-m}$ for all m, k (take for example $\Omega = [0,1]$, equipped with the Lebesgue measure). Define the nonnegative random variables $X^m_k = 1_{A^m_k} 2^{2m}$. Then $(X^m_k : m, k)$ is bounded in probability. Let $a > 0$ and assume Z is a nonnegative random variable such that $E(ZX_k^m) \leq a$ for all m, k. Then

$$E(1_{A_k^m} Z) = E(ZX_k^m)2^{-2m} \leq a2^{-2m}.$$

Summing over k, we obtain $E(Z) \leq a2^{-m}$ for all $m \in \mathbb{N}$, and therefore $E(Z) = 0$. Since $Z \geq 0$, we have $Z = 0$.

We call a family of random variables L^p*–bounded* for $p \geq 1$ if it is norm bounded in L^p.

Remark 1.3.4. Lemma 1.3.2 states that a convex family of nonnegative random variables $\mathcal{X}$ is bounded in probability if and only if there exists a measure $Q \sim P$, such that $\mathcal{X}$ is $L^1(Q)$–bounded. One might ask if this can be improved. For example, there could exist $Q \sim P$ such that $\mathcal{X}$ is $L^p(Q)$–bounded for some $p > 1$. But this is not true in general. Even if S is a Brownian motion there might not be an absolutely continuous $Q \ll P$, such that $\mathcal{K}_1$ is uniformly integrable under Q. To see this, choose an increasing sequence of partitions $(A_k^m : 1 \leq k \leq 2^m, m \in \mathbb{N})$ of $\mathbb{R}$, such that $\mu(A_k^m) = 2^{-m}$ for all m, k, where μ denotes the standard normal distribution. Define the random variables $X_k^m = 1_{A_k^m}(S_1)2^m$. Then $X_k^m \in L^\infty$, and $E(X_k^m) = 1$ for all m, k. By the predictable representation property of Brownian motion, $X_k^m \in \mathcal{K}_1$ for all m, k. Now let $Q \ll P$, and let $g \geq 0$ be such that $\lim_{x\to\infty} g(x)/x = \infty$. If we show that $(g(X_k^m))_{m,k}$ is unbounded in $L^1(Q)$, then de la Vallée-Poussin's theorem (see [DM78], II-22) implies that $\mathcal{K}_1$ is not uniformly integrable under Q. Let $a > 0$ and let $m \in \mathbb{N}$ be such that $g(2^m) \geq a2^m$. Choose k for which $Q(S_1 \in A_k^m) \geq 2^{-m}$. Such a k must exist because Q has total mass 1. Then

$$E_Q(g(X_k^m)) \geq E_Q(1_{A_k^m}(S_1)2^m a) \geq 2^{-m}2^m a = a.$$

Since $a > 0$ was arbitrary, $E_Q(g(\cdot))$ is unbounded on $\mathcal{K}_1$.

The following Lemma establishes Theorem 1.3.1 in the case of two time steps. The general case then follows easily.

Lemma 1.3.5. *Let $\mathcal{Y}$ be a L^1–bounded family of nonnegative processes indexed by $\{0, 1\}$, adapted to a filtration $(\mathcal{F}_0, \mathcal{F}_1)$. Assume that $\mathcal{Y}$ is fork-convex and that $\mathcal{Y}$ contains a process of the form $(1, Y_1^*)$ for a strictly positive Y_1^*. Then there exists a strictly positive $\mathcal{F}_0$–measurable random variable Z, such that $(Y_0 Z, Y_1)$ is a supermartingale for every $Y \in \mathcal{Y}$. The random variable Z can be chosen such that*

$$\sup_{Y\in\mathcal{Y}} E(Y_0 Z) \leq \sup_{Y\in\mathcal{Y}} \max_{i=0,1} E(Y_i). \tag{1.6}$$

Proof. We define a nonnegative set function μ on $\mathcal{F}_0$ by setting

$$\mu(A) := \sup_{Y\in\mathcal{Y}} E(1_A Y_1/Y_0).$$

Let us apply the fork-convexity of $\mathcal{Y}$ to show that for every $Y \in \mathcal{Y}$ there exists $\widetilde{Y} \in \mathcal{Y}$, such that $Y_1/Y_0 = \widetilde{Y}_1$. We take $s = 0$ and $Y^1 = (1, Y_1^*)$ and $Y^2 = Y$ and $\lambda_s = 1$ in

(1.5). Then $\widetilde{Y} \in \mathcal{Y}$, where $\widetilde{Y}_0 = 1_{\{Y_0>0\}}$ and $\widetilde{Y}_1 = Y_1/Y_0$. In particular, it follows from the L^1–boundedness of $\mathcal{Y}$ that

$$\mu(A) = \sup_{Y\in\mathcal{Y}} E\left(1_A \frac{Y_1}{Y_0}\right) \le \sup_{\widetilde{Y}\in\mathcal{Y}} E(1_A \widetilde{Y}_1) < \infty$$

for all A, i.e. μ is finite. In fact μ is a finite measure. Let $A, B \in \mathcal{F}_0$ be two disjoint sets and let $Y^A, Y^B \in \mathcal{Y}$. We take $s = 0$, $Y^1 = (1, Y_1^*)$, $Y^2 = Y^A$, $Y^3 = Y^B$, and $\lambda_s = 1_A$ in (1.5), which implies that $\widetilde{Y} \in \mathcal{Y}$, where

$$\widetilde{Y}_t = 1_{\{0\}}(t)\left(1_A 1_{\{Y_0^A>0\}} + 1_B 1_{\{Y_0^B>0\}}\right) + 1_{\{1\}}(t)\left(1_A \frac{Y_1^A}{Y_0^A} + 1_B \frac{Y_1^B}{Y_0^B}\right).$$

Note that $\widetilde{Y}_1/\widetilde{Y}_0 = \widetilde{Y}_1$, since we set $0/0 = 0$. Because A and B are disjoint, we have

$$1_{A\cup B}\frac{\widetilde{Y}_1}{\widetilde{Y}_0} = 1_A \frac{Y_1^A}{Y_0^A} + 1_B \frac{Y_1^B}{Y_0^B}.$$

As a consequence we obtain

$$\begin{aligned}\mu(A) + \mu(B) &= \sup_{(Y^A,Y^B)\in\mathcal{Y}^2} E\left(1_A \frac{Y_1^A}{Y_0^A} + 1_B \frac{Y_1^B}{Y_0^B}\right) \\ &\le \sup_{\widetilde{Y}\in\mathcal{Y}} E\left(1_{A\cup B}\frac{\widetilde{Y}_1}{\widetilde{Y}_0}\right) = \mu(A\cup B).\end{aligned}$$

But $\mu(A \cup B) \le \mu(A) + \mu(B)$ is obvious, and therefore μ is finitely additive.

Now let (A_n) be a sequence of disjoint sets in $\mathcal{F}_0$. Then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sup_{Y\in\mathcal{Y}} \sum_{n=1}^{\infty} E\left(1_{A_n}\frac{Y_1}{Y_0}\right) \le \sum_{n=1}^{\infty} \sup_{Y^n\in\mathcal{Y}} E\left(1_{A_n}\frac{Y_1^n}{Y_0^n}\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

The opposite inequality holds for any finitely additive nonnegative set function. Therefore, μ is a finite measure on $\mathcal{F}_0$, which is absolutely continuous with respect to P. Hence, there exists a nonnegative $Z \in L^1(\mathcal{F}_0, P)$, such that

$$\mu(A) = E(1_A Z) = \sup_{Y\in\mathcal{Y}} E\left(1_A \frac{Y_1}{Y_0}\right). \tag{1.7}$$

It is easy to see that we can replace 1_A in (1.7) by any nonnegative $\mathcal{F}_0$–measurable random variable. In particular, for any $Y \in \mathcal{Y}$ and any $A \in \mathcal{F}_0$

$$E(1_A Y_0 Z) = \sup_{\widetilde{Y}\in\mathcal{Y}} E\left(1_A Y_0 \frac{\widetilde{Y}_1}{\widetilde{Y}_0}\right) \ge E\left(1_A Y_0 \frac{Y_1}{Y_0}\right) = E(1_A Y_1),$$

proving that $(Y_0 Z, Y_1)$ is a supermartingale provided that $E(Y_0 Z) < \infty$. But the bound stated in (1.6) follows immediately from the fork-convexity of $\mathcal{Y}$, because the process $\widetilde{Y} = (Y_0 1_{\{Y_0^1>0\}}, Y_0 Y_1^1/Y_0^1)$ is in $\mathcal{Y}$ for any $Y^1 \in \mathcal{Y}$, and thus

$$E(Y_0 Z) = \sup_{Y^1 \in \mathcal{Y}} E\left(Y_0 \frac{Y_1^1}{Y_0^1}\right) \leq \sup_{\widetilde{Y} \in \mathcal{Y}} E\left(\widetilde{Y}_1\right).$$

It remains to show that Z is strictly positive. But this is easy, because $(1, Y_1^*)$ is in $\mathcal{Y}$, and Y_1^* is strictly positive. Therefore, (Z, Y_1^*) is a supermartingale with strictly positive terminal value, which is only positive if also Z is strictly positive. □

Remark 1.3.6. Some type of stability assumption is necessary for Lemma 1.3.5 to hold. Even for a uniformly integrable and convex family of processes $\mathcal{Y}$, the lemma may fail without assuming fork convexity: Let again $\{A_k^m : 1 \leq k \leq 2^m, m \in \mathbb{N}\}$ be an increasing sequence of partitions of Ω, such that for every m and k we have $P(A_k^m) = 2^{-m}$. Define the random variables $X_k^m = 1_{A_k^m} 2^m/m$. Let $a > 1$ and let m_0 be such that $2^{m_0-1} < a \leq 2^{m_0}$. Then

$$\sup_{m,k} E(|X_k^m| 1_{\{|X_k^m| \geq a\}}) \leq E(|X_1^{m_0}|) = \frac{1}{m_0},$$

proving that $(X_k^m)_{m,k}$ is uniformly integrable. From de la Vallée-Poussin's theorem and Jensen's inequality we obtain that also the convex hull $\mathcal{X}$ of the X_k^m is uniformly integrable. Define $\mathcal{F}_0 = \mathcal{F}_1 = \sigma(A_k^m : 1 \leq k \leq 2^m, m \in \mathbb{N})$, and $\mathcal{Y} = \{(1, X) : X \in \mathcal{X}\}$. Assume there exists $Z > 0$ such that $E(1_A X) \leq E(1_A Z)$ for all $A \in \mathcal{F}_0$ and $X \in \mathcal{X}$. Then for every $m \in \mathbb{N}$

$$E(Z) = \sum_{k=1}^{2^m} E(1_{A_k^m} Z) \geq \sum_{k=1}^{2^m} E(1_{A_k^m} X_k^m) = \frac{2^m}{m}.$$

so that $E(Z) = \infty$. Therefore, (Z, X) cannot be a supermartingale for any $X \in \mathcal{X}$. In fact it is possible to show that $E(1_{A_k^m} Z) = \infty$ for all k, m, and since $\mathcal{F}_0$ is generated by $(A_k^m)_{m,k}$, we must have $P(Z = \infty) = 1$.

To pass from two time steps to finitely many time steps is easy and follows by induction:

Corollary 1.3.7. *Let $\mathcal{Y}$ be a L^1–bounded family of nonnegative processes indexed by $\{0, \dots, m\}$, adapted to a filtration $(\mathcal{F}_k : 0 \leq k \leq m)$. Assume that $\mathcal{Y}$ is fork-convex, and that it contains the constant process $(1, \dots, 1)$.*

Then there exists a strictly positive and adapted process $(Z_k : 0 \leq k \leq m)$, with $Z_m = 1$ and such that ZY is a supermartingale for every $Y \in \mathcal{Y}$. Z can be chosen such that

$$\sup_{Y \in \mathcal{Y}} \max_{k=0\dots,m} E(Z_k Y_k) \leq \sup_{Y \in \mathcal{Y}} \max_{k=0,\dots,m} E(Y_k). \tag{1.8}$$

Proof. For $m = 1$, this is just Lemma 1.3.5: take $Z_0 = Z, Z_1 = 1$.

Now assume the result holds for n. Let $\mathcal{Y}$ be a family of processes indexed by $\{0, \dots, n+1\}$, and assume that $\mathcal{Y}$ satisfies all requirements stated above. Then also $(Y_1, \dots, Y_{m+1})$

satisfies all those requirements. By induction hypothesis, there exists a strictly positive and adapted process $Z = (Z_1, \dots, Z_m, 1)$ such that ZY is a supermartingale for all $Y \in \mathcal{Y}$, and such that

$$\sup_{Y\in\mathcal{Y}} \max_{k=1\dots,m+1} E(Z_k Y_k) \leq \sup_{Y\in\mathcal{Y}} \max_{k=1,\dots,m+1} E(Y_k). \tag{1.9}$$

Therefore, it suffices to construct a suitable Z_0. For this purpose we apply Lemma 1.3.5 to the family of processes $\widetilde{\mathcal{Y}} = \{(Y_0, Z_1 Y_1) : Y \in \mathcal{Y}\}$. Since $\mathcal{Y}$ contains the constant process $(1, \dots, 1)$, the family $\widetilde{\mathcal{Y}}$ contains the process $(1, Z_1)$, and Z_1 is strictly positive. Furthermore, it is straightforward to check that $\widetilde{\mathcal{Y}}$ is fork-convex. L^1–boundedness of $\widetilde{\mathcal{Y}}$ follows from (1.9). Hence, we can apply Lemma 1.3.5 to $\widetilde{\mathcal{Y}}$, and the result follows. □

To prove Theorem 1.3.1, we have to go from finite discrete time to continuous time. This is achieved by means of a compactness argument. Compactness for right-continuous functions is not very easy to show, and it would require us to use some form of the Arzelà-Ascoli theorem. However, we want to construct a *supermartingale* Z, and therefore it will be sufficient to construct its "skeleton" $(Z_q : q \in \mathbb{Q}_+)$. Using standard results for supermartingales, we can then use this skeleton to construct a right-continuous supermartingale density.

We will need the notion of convex compactness as introduced by Žitković [Ž10].

Definition 1.3.8. Let $\mathbb{X}$ be a topological vector space. A closed convex subset $C \subseteq \mathbb{X}$ is called *convexly compact* if for any family $\{F_\alpha : \alpha \in A\}$ of closed convex subsets of C, we can only have $\bigcap_{\alpha\in A} F_\alpha = \emptyset$ if there exist already finitely many $\alpha_1, \dots, \alpha_m \in A$ for which $\bigcap_{k=1}^m F_{\alpha_k} = \emptyset$.

Recall that L^0 is the space of real valued random variables, equipped with the topology of convergence in probability. Žitković [Ž10] then characterizes convexly compact sets of nonnegative elements of L^0.

Lemma 1.3.9 (Theorem 3.1 of [Ž10])**.** *Let $\mathcal{X}$ be a convex set of nonnegative random variables, closed with respect to convergence in probability. Then $\mathcal{X}$ is convexly compact in L^0 if and only if it is bounded in probability.*

Note that Žitković works on a complete probability space. But completeness is not used in the proof of Theorem 3.1. There is only one point in the proof where it is not immediately clear whether completeness of the σ–algebra is needed: when Lemma A1.2 of [DS94] is applied. However, this lemma is formulated for general probability spaces.

In Proposition B.6 we prove a Tychonoff theorem for countable families of convexly compact subsets of metric spaces. This will be used in the following proof.

Lemma 1.3.10. *Let $\mathcal{Y}$ and $\mathcal{K}$ be as in Theorem 1.3.1. Then there exists a nonnegative supermartingale $(\overline{Z}_q)_{q\in\mathbb{Q}_+\cup\{\infty\}}$ with $\overline{Z}_\infty > 0$, such that $(\overline{Z}_q Y_q)_{q\in\mathbb{Q}_+}$ is a supermartingale for all $Y \in \mathcal{X}$.*

Proof. Recall that $\mathcal{Y}$ is convex, and thus $\mathcal{K}$ is convex as well. By Lemma 1.3.2 there exists $Q \sim P$ such that $\mathcal{K}$ is $L^1(Q)$–bounded. We set $a := \sup_{X\in\mathcal{K}} E_Q(X)$ and define the family of processes

$$\mathcal{Z} = \{(Z_q)_{q\in\mathbb{Q}_+\cup\{\infty\}} : Z_\infty = 1, Z_q \geq 0, Z_q \in \mathcal{F}_q, \text{ and } E_Q(Z_q) \leq a \text{ for all } q\},$$

where we write $Z_q \in \mathcal{F}_q$ to denote that Z_q is $\mathcal{F}_q$–measurable. According to Lemma 1.3.9 and Proposition B.6 in Appendix B, $\mathcal{Z}$ is a convexly compact set in $\prod_{q\in\mathbb{Q}_+\cup\{\infty\}} L^0(\mathcal{F}_q, Q)$ equipped with the product topology. Define for given $q, r \in \mathbb{Q}_+ \cup \{\infty\}$

$$\mathcal{Z}(q,r) = \{Z \in \mathcal{Z} : E_Q(Z_{q+r}Y_{q+r}/Y_q|\mathcal{F}_q) \leq Z_q \text{ for all } Y \in \mathcal{Y}\},$$

where for $r = \infty$ we only consider those $Y \in \mathcal{Y}$ for which $Y_\infty = \lim_{t\to\infty} Y_t$ exists. The sets $\mathcal{Z}(q,r)$ are convex, and by Fatou's lemma they are also closed. Furthermore, they are subsets of the convexly compact set $\mathcal{Z}$. So if

$$\bigcap_{q\in\mathbb{Q}_+, r\in\mathbb{Q}_+\cup\{\infty\}} \mathcal{Z}(q,r)$$

was empty, then already a finite intersection would have to be empty. But if a finite intersection was empty, then there would exist $0 \leq t_0 < \cdots < t_n \leq \infty$ for which it is impossible to find $(Z_{t_i} : i = 0, \ldots, n)$ with $Z_{t_n} = 1$ and such that $(Y_{t_i}Z_{t_i} : i = 0, \ldots, n)$ is a Q–supermartingale for every $Y \in \mathcal{Y}$ (respectively for every $Y \in \mathcal{Y}$ for which $\lim_{t\to\infty} Y_t$ exists in case $t_n = \infty$). This would contradict Corollary 1.3.7.

So let Z be in the intersection of all $\mathcal{Z}(q,r)$ and let $Y \in \mathcal{Y}$. Then for all $q, r \in \mathbb{Q}_+$

$$E_Q\left(Z_{q+r}\frac{Y_{q+r}}{Y_q} \middle| \mathcal{F}_q \right) \leq Z_q,$$

which shows that ZY is a Q–supermartingale indexed by $\mathbb{Q}_+$. Taking $Y \equiv 1$, we also see that $(Z_q : q \in \mathbb{Q}_+ \cup \{\infty\})$ is a supermartingale. To complete the proof it suffices now to define for $q \in \mathbb{Q}_+ \cup \{\infty\}$

$$\overline{Z}_q = Z_q \left.\frac{dP}{dQ}\right|_{\mathcal{F}_q}.$$

□

We are now ready to prove Rokhlin's result.

Proof of Theorem 1.3.1. It remains to show that given the skeleton $(\overline{Z}_q : q \in \mathbb{Q}_+ \cup \{\infty\})$ we can construct a right-continuous supermartingale density with left limits almost everywhere. This is a standard result on supermartingales. For the reader's convenience and to dispel possible concerns about the incompleteness of our filtration, we give the arguments below.

Since $(\mathcal{F}_t)$ is right-continuous, for every $t \geq 0$ there exists a nondecreasing family of sets $(\mathcal{N}_t)_{t\geq 0}$, such that $\mathcal{N}_t \in \mathcal{F}_t$ for all $t \geq 0$, and such that for $\omega \in \Omega \setminus \mathcal{N}_t$

$$\lim_{\substack{r\to s- \\ r\in\mathbb{Q}}} \overline{Z}(\omega)_r \text{ and } \lim_{\substack{r\to s+ \\ r\in\mathbb{Q}}} \overline{Z}(\omega)_r$$

exist for all $s \leq t$. See for example Ethier and Kurtz [EK86], right before Proposition 2.2.9. We define for $t \in [0,\infty)$

$$Z_t(\omega) = \begin{cases} \lim_{\substack{s\to t+ \\ s\in\mathbb{Q}}} \overline{Z}_s(\omega), & \omega \in \Omega \setminus \mathcal{N}_t \\ 0, & \text{otherwise.} \end{cases}$$

Then Z is adapted because $(\mathcal{F}_t)$ is right-continuous, and Z is right-continuous by definition. It may not have left limits everywhere. But since $(\mathcal{N}_t)$ is nondecreasing, $\tau(\omega) := \inf\{t \geq 0 : \omega \in \mathcal{N}_t\}$ defines a stopping time, such that $P(\tau = \infty) = 1$, and such that $t \mapsto Z_t(\omega)$ has left limits everywhere except at $\tau(\omega)$.

Let us show that ZY is a supermartingale for every $Y \in \mathcal{Y}$. Recall that the processes in $\mathcal{Y}$ are right-continuous. Using Fatou's Lemma in the first step and Corollary 2.2.10 of [EK86] in the second step, we obtain

$$\begin{aligned} E_Q(Z_{t+s}Y_{t+s}|\mathcal{F}_t) &\leq \liminf_{\substack{r\to(t+s)+ \\ r\in\mathbb{Q}}} E_Q(\overline{Z}_r Y_r|\mathcal{F}_t) = \liminf_{\substack{r\to(t+s)+ \\ r\in\mathbb{Q}}} \liminf_{\substack{q\to t+ \\ q\in\mathbb{Q}}} E_Q(\overline{Z}_r Y_r|\mathcal{F}_q) \\ &\leq \liminf_{\substack{r\to(t+s)+ \\ r\in\mathbb{Q}}} \liminf_{\substack{q\to t+ \\ q\in\mathbb{Q}}} \overline{Z}_q Y_q = Z_t Y_t. \end{aligned}$$

The same arguments with $s = \infty$ and $Y = 1$ show that if we set $Z_\infty = \overline{Z}_\infty$, then $(Z_t)_{t\in[0,\infty]}$ is a nonnegative supermartingale with strictly positive terminal value. By Theorem I.1.39 of [JS03], Z_t almost surely converges to a limit $\widetilde{Z}_\infty$ as $t \to \infty$. Define now $M_t = E(Z_\infty|\mathcal{F}_t)$ for $t \in [0,\infty]$. This is a uniformly integrable martingale which almost surely converges to Z_∞ as $t \to \infty$. By the supermartingale property of Z we have $M_t \leq Z_t$ for all $t \geq 0$, and therefore

$$0 < Z_\infty = \lim_{t\to\infty} M_t \leq \lim_{t\to\infty} Z_t = \widetilde{Z}_\infty.$$

It remains to show that if there exists a supermartingale density, then $\mathcal{Y}$ is bounded in probability. If Z is a supermartingale density for $\mathcal{Y}$, then for any $Y \in \mathcal{Y}$, $Z_t Y_t$ converges as $t \to \infty$, see Theorem I.1.39 of [JS03]. Since Z_t converges to a strictly positive limit, Y_t must converge as well, and we have $E(Z_\infty Y_\infty) \leq E(Z_0 Y_0) = E(Z_0)$. Now Lemma 1.3.2 shows that $\mathcal{Y}$ is bounded in probability. $\square$

Corollary 1.3.11. *If $\mathcal{Y}$ is as in Theorem 1.3.1, then every $Y \in \mathcal{Y}$ is a semimartingale for which Y_t almost surely converges as $t \to \infty$.*

Proof. Convergence was shown in the proof of Theorem 1.3.1. The semimartingale property follows from Itô's formula: Let Z be a supermartingale density for $\mathcal{Y}$. Then Z is strictly positive, and therefore $1/Z$ is a semimartingale, implying that $Y = (1/Z)(ZY)$ is a semimartingale. □

In case $\mathcal{Y} = \mathcal{W}_1$ and under the stronger assumption (NFLVR_s), Corollary 1.3.11 was already shown by Delbaen and Schachermayer [DS94]. See also Ankirchner [Ank05] and Kardaras and Platen [KP11].

Theorem 1.1.3 is now an immediate corollary of Theorem 1.3.1:

Proof of Theorem 1.1.3. It suffices to note that $\mathcal{W}_1$ and $\mathcal{W}_{1,s}$ satisfy the assumptions of Theorem 1.3.1. This is easy and shown for example in Rokhlin [Rok10], in the proof of Theorem 2. Rokhlin only treats the case of $\mathcal{W}_1$ and $\mathcal{K}_1$, but the same arguments also work for $\mathcal{W}_{1,s}$ and $\mathcal{K}_{1,s}$. □

Proof of Corollary 1.1.4. Let S have components that are locally bounded from below and assume that S satisfies ($\mathrm{NA1}_s$). Recall that local semimartingales are semimartingales, see for example Protter [Pro04], Theorem II.6. Protter works with complete filtrations, but it follows from Lemma A.5 in Appendix A that for every $(\mathcal{F}_t^P)$–semimartingale there exists an indistinguishable $(\mathcal{F}_t)$–semimartingale. Let $1 \leq k \leq d$. Since S^k is locally bounded from below, there exists an increasing sequence of stopping times (τ_m) with $\lim_{m\to\infty} \tau_m = \infty$, and a sequence of strictly positive numbers (a_m), such that $(1 + a_m S^k_{t\wedge\tau_m})_{t\geq 0} \in \mathcal{W}_{1,s}$. It follows from Corollary 1.3.11 that the stopped process $S^k_{\cdot\wedge\tau_m}$ is a semimartingale for every m, and therefore S^k is a semimartingale.

It remains to show that in the case of local boundedness from below, any supermartingale density for $\mathcal{W}_{1,s}$ is a supermartingale density for $\mathcal{W}_1$. But this is the content of Kardaras and Platen [KP11], Section 2.2. (And it will also follow from our considerations in Section 1.4.) □

Of course we could also assume that every component of S is either locally bounded from below or locally bounded from above, and we would still obtain the semimartingale property of S under ($\mathrm{NA1}_s$). But in the totally unbounded case, S is not necessarily a semimartingale. A simple counterexample is given by a one dimensional Lévy-process with jumps that are unbounded both from above and from below, to which we add an independent fractional Brownian motion with Hurst index $H \neq 1/2$. Their sum is not a semimartingale. But there are no 1–admissible simple strategies other than 0, so that $\mathcal{K}_{1,s} = \{1\}$, which is obviously bounded in probability.

1.4. Construction of dominating local martingale measures

1.4.1. The Kunita-Yoeurp problem and Föllmer's measure

Now let Z be a strictly positive supermartingale with $Z_\infty > 0$ and $E_P(Z_0) = 1$. Our aim is to construct a dominating measure Q and a stopping time τ, such that (Z, τ) is the Kunita-Yoeurp decomposition of Q with respect to P. We call this the *Kunita-Yoeurp* *problem.* Recall that (Z, τ) is the Kunita-Yoeurp decomposition of Q with respect to P if

1. $P(\tau = \infty) = 1$,
2. for $A \in \mathcal{F}_t$ we have
$$Q(A \cap \{\tau > t\}) = E_P(1_A Z_t). \tag{1.10}$$

In this case it follows for any stopping time ρ and any $A \in \mathcal{F}_\rho$ that

$$Q(A \cap \{\tau > \rho\}) = E_P\left(1_{A\cap\{\rho<\infty\}} Z_\rho\right), \tag{1.11}$$

see for example [Yoe85], Proposition 4.

In general it is impossible to construct Q and τ without making further assumptions on the underlying filtered probability space: For example, the space could be too small. Take $\Omega = \{0\}$ that consists of only one element, and define $\mathcal{F} = \mathcal{F}_t = \{\emptyset, \Omega\}$ for all $t \geq 0$. Then

$$Z_t = \frac{1}{2}(1 + e^{-t}), \qquad t \geq 0,$$

is a continuous, nonnegative supermartingale with $Z_\infty > 0$. But there exists only one probability measure on Ω, and therefore any Q would have the Kunita-Yoeurp decomposition $(1, \infty)$ with respect to P, and not (Z, τ). This is reminiscent of the Dambis Dubins-Schwarz theorem without the assumption $\langle M\rangle_\infty = \infty$ (see Revuz and Yor [RY99], Theorem V.1.7). This problem can be solved by enlarging Ω.

But even if the space is large enough, it might still not be possible to find Q and τ, because the filtration might be too large. Assume that the filtration $(\mathcal{F}_t)$ is complete with respect to P, and that $E_P(Z_0) = 1$. Then (1.10) shows that Q is absolutely continuous with respect to P on $\mathcal{F}_0$. Since $\mathcal{F}_0$ contains all P–null sets, this means that Q is absolutely continuous with respect to P, and therefore (1.10) implies that $Z_t = E_P(\mathrm{d}Q/\mathrm{d}P|\mathcal{F}_t)$. In other words Z is a uniformly integrable martingale under P. So if Z is a supermartingale, then the filtration $(\mathcal{F}_t)$ should not be completed. This problem can be avoided by assuming that $(\mathcal{F}_t)$ is the right-continuous modification of a standard system, to be defined below.

If we are allowed to enlarge Ω and if $(\mathcal{F}_t)$ is the right-continuous modification of a standard system, then the problem of constructing Q and τ has been solved by Yoeurp [Yoe85] with the help of Föllmer's measure. Let us describe Yoeurp's solution.

First we remove the second problem in constructing Q and τ by assuming that the filtration $(\mathcal{F}_t)$ is the right-continuous modification of a standard system $(\mathcal{F}^0_t)$. A filtration $(\mathcal{F}^0_t)$ is called *standard system* if

1. for all $t \geq 0$, the σ–algebra $\mathcal{F}^0_t$ is σ–isomorphic to the Borel σ–algebra of a Polish space; that is, there exists a Polish space $\mathbb{X}_t$ with Borel σ–algebra $\mathcal{B}(\mathbb{X}_t)$,

and a bijective map $\pi : \mathcal{F}_t^0 \to \mathcal{B}(\mathbb{X}_t)$, such that $\pi(\bigcup_{m\in\mathbb{N}} A_m) = \bigcup_{m\in\mathbb{N}} A_m$ and $\pi(\bigcap_{m\in\mathbb{N}} A_m) = \bigcap_{m\in\mathbb{N}} A_m$ for every sequence $(A_m)_{m\in\mathbb{N}}$;

2. if $(t_m)_{m\in\mathbb{N}}$ is a nondecreasing sequence of nonnegative times, and if $(A_m)_{m\in\mathbb{N}}$ is a nonincreasing sequence of sets, such that for every $m \in \mathbb{N}$ the set A_m is an atom of $\mathcal{F}_{t_m}^0$ (i.e. $B \in \mathcal{F}_{t_m}$ and $B \subseteq A_m$ implies $B = A_m$ or $B = \emptyset$), then $\bigcap_{m\in\mathbb{N}} A_m \neq \emptyset$.

Then $(\mathcal{F}_t)$ is defined by setting $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s^0$. Path spaces equipped with the canonical filtration are only standard systems if we allow for "explosion" to a cemetery state in finite time, see Föllmer [Föl72], Example 6.3, 2), or Meyer's result in Dellacherie [Del69], p. 100. On such a path space it is in fact always possible to solve the Kunita-Yoeurp problem without enlarging the space. This and other results will be presented in the upcoming work Perkowski and Ruf [PR13].

Here we do not assume that Ω is a path space, and therefore we continue by enlarging Ω in order to solve the possible problem of Ω being too small. Define $\overline{\Omega} := \Omega \times [0,\infty]$ and $\overline{\mathcal{F}} := \mathcal{F} \otimes \mathcal{B}[0,\infty]$, where $\mathcal{B}[0,\infty]$ denotes the Borel σ–algebra on $[0,\infty]$. We also define $\overline{P} := P \otimes \delta_\infty$, where δ_∞ is the Dirac measure at ∞. The filtration $(\overline{\mathcal{F}}_t)$ is defined as

$$\overline{\mathcal{F}}_t := \bigcap_{s>t} \mathcal{F}_s \otimes \sigma([0,r] : r \leq s).$$

Note that if $(\mathcal{F}_t)$ is the right-continuous modification of the standard system $(\mathcal{F}_t^0)$, then $(\overline{\mathcal{F}}_t)$ is the right-continuous modification of the standard system

$$\overline{\mathcal{F}}_t^0 := \mathcal{F}_t^0 \otimes \sigma([0,s] : s \leq t), \qquad t \geq 0.$$

Random variables X on Ω are embedded into $\overline{\Omega}$ by setting $\overline{X}(\omega,\zeta) := X(\omega)$.

Let us remark that $(\overline{\Omega}, \overline{\mathcal{F}}, (\overline{\mathcal{F}}_t), \overline{P})$ is an enlargement of $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ in the sense of Revuz and Yor [RY99]:

Definition 1.4.1 ([RY99], p. 182)**.** A filtered probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, (\widetilde{\mathcal{F}}_t), \widetilde{P})$ is an *enlargement* of $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ if there exists a measurable map $\pi : \widetilde{\Omega} \to \Omega$, such that $\pi^{-1}(\mathcal{F}_t) \subseteq \widetilde{\mathcal{F}}_t$ and such that $\widetilde{P} \circ \pi^{-1} = P$. In this case, random variables are embedded from $(\Omega, \mathcal{F})$ into $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ by setting $\widetilde{X}(\widetilde{\omega}) = X(\pi(\widetilde{\omega}))$.

Define $\pi(\omega,\zeta) := \omega$. Then $\pi^{-1}(A) = A \times [0,\infty] \in \overline{\mathcal{F}_t}$ for every $A \in \mathcal{F}_t$, i.e. $\pi^{-1}(\mathcal{F}_t) \subseteq \overline{\mathcal{F}_t}$. For every $A \in \mathcal{F}$ we have $\overline{P} \circ \pi^{-1}(A) = P \otimes \delta_\infty(A \times [0,\infty]) = P(A)$. And if X is a random variable on $(\Omega, \mathcal{F})$, then $\overline{X}(\omega,\zeta) = X(\omega) = X(\pi(\omega,\zeta))$.

Now we can proceed to construct $(\overline{Q}, \overline{\tau})$ on $(\overline{\Omega}, \overline{\mathcal{F}}, (\overline{\mathcal{F}}_t))$. In fact it suffices to construct $\overline{Q}$, because we will take $\overline{\tau}(\omega,\zeta) := \zeta$, so that $\overline{P}(\overline{\tau} = \infty) = 1$. However, there is one remaining problem. In general $\overline{Q}$ will not be uniquely determined by $(\overline{Z}, \overline{\tau})$. The measure $\overline{Q}$ must satisfy

$$1 = \overline{Q}(\overline{\Omega}) = \overline{Q}(\overline{\Omega} \cap \{t < \overline{\tau}\}) + \overline{Q}(\overline{\Omega} \cap \{t \geq \overline{\tau}\}) = E_{\overline{P}}(\overline{Z}_t) + \overline{Q}(t \geq \overline{\tau}) \tag{1.12}$$

for all $t \geq 0$. But $\overline{Q}$ is supposed to solve the Kunita-Yoeurp problem associated with $(\overline{Z}, \overline{\tau})$, and therefore at time $\overline{\tau}$, the measure $\overline{Q}$ should stop being absolutely continuous with respect to $\overline{P}$, and (1.12) implies that $\overline{Q}(\overline{\tau} < \infty) > 0$ if $\overline{Z}$ is not a martingale. So knowing $\overline{P}$, $\overline{Z}$, and $\overline{\tau}$, in general we can only hope to determine $\overline{Q}$ uniquely on the σ–field

$$\begin{aligned}\overline{\mathcal{F}}_{\overline{\tau}-} &:= \sigma(\overline{\mathcal{F}}^0_0, \{\overline{A}_t \cap \{\overline{\tau} > t\} : \overline{A}_t \in \overline{\mathcal{F}}_t, t > 0\}) \\ &= \sigma(\{A_0 \times \{0\}, A_t \times (t, \infty] : A_t \in \mathcal{F}^0_t, t \geq 0\}). \end{aligned} \tag{1.13}$$

For the second equality we refer to [Föl72]. Note that $\overline{\mathcal{F}}_{\overline{\tau}-}$ is the predictable σ–algebra over $(\mathcal{F}^0_t)$ on $\Omega \times [0, \infty]$. The reason for taking $\overline{\mathcal{F}}^0_0$ rather than $\overline{\mathcal{F}}_0$ lies in the fact that $\overline{\mathcal{F}}^0_0$ is countably generated but in general $\overline{\mathcal{F}}_0$ is not. It then follows from our definition that $\overline{\mathcal{F}}_{\overline{\tau}-}$ is countably generated, a condition which is needed to apply Parthasarathy's extension theorem [Par67], on which Föllmer's construction is based.

In conclusion, in order to construct τ and Q, we need to assume that $(\mathcal{F}_t)$ is the right-continuous modification of a standard system, we need to enlarge $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ as described above, and we have to accept that Q will only be defined on $\overline{\mathcal{F}}_{\overline{\tau}-}$. Under these conditions, we can take $\overline{Q}$ as the Föllmer measure of Z. Given a nonnegative supermartingale Z with $E_P(Z_0) = 1$, Föllmer [Föl72], see also Meyer [Mey72], constructs a measure P^Z on $(\overline{\Omega}, \overline{\mathcal{F}}_{\tau-})$, which satisfies $P^Z(\overline{A}_t \cap \{\overline{\tau} > t\}) = E_{\overline{P}}(\overline{Z}_t 1_{\overline{A}_t})$ for all $t \geq 0$ and all $\overline{A}_t \in \overline{\mathcal{F}}_t$. This is exactly the relation (1.10), and therefore $(\overline{Z}, \overline{\tau})$ is the Kunita-Yoeurp decomposition of $\overline{Q} := P^Z$ with respect to $\overline{P}$.

Note that it is possible to extend $\overline{Q}$ from $\overline{\mathcal{F}}_{\overline{\tau}-}$ to $\overline{\mathcal{F}}$, generally in a non-unique way, see p. 9 of [KKN11]. See also Perkowski and Ruf [PR13], where we will describe very precisely under which conditions and in which sense the extension is unique. From now on we assume that $\overline{Q}$ is one of these extensions, i.e. that $\overline{Q}$ denotes a probability measure on $(\overline{\Omega}, \overline{\mathcal{F}})$ that satisfies $\overline{Q}|_{\overline{\mathcal{F}}_{\overline{\tau}-}} = P^Z$.

It remains to show that $\overline{Q}$ dominates $\overline{P}$. But this is a consequence of the following general result:

Lemma 1.4.2. *Let P and Q be two probability measures on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0})$. Let (Z, τ) be the Kunita-Yoeurp decomposition of Q with respect to P. Define $\mathcal{F}_\infty = \bigvee_{t\geq 0} \mathcal{F}_t$ and assume that $P(Z_\infty > 0) = 1$. Then $P|_{\mathcal{F}_\infty} \ll Q|_{\mathcal{F}_\infty}$.*

Proof. Let $A \in \bigcup_{t\geq 0} \mathcal{F}_t$. We use the σ–continuity of P, (1.10), and Fatou's lemma, to obtain

$$Q(A \cap \{\tau = \infty\}) = \lim_{t\to\infty} Q(A \cap \{\tau > t\}) = \lim_{t\to\infty} E_P(Z_t 1_A) \geq E_P(Z_\infty 1_A).$$

By the monotone class theorem, this inequality extends to all $A \in \mathcal{F}_\infty$. Since $P(Z_\infty > 0) = 1$, we conclude that $P|_{\mathcal{F}_\infty} \ll Q|_{\mathcal{F}_\infty}$. □

In the following we make the standing assumption that we work on a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ where it is possible to solve the Kunita-Yoeurp problem of associating a probability measure and a stopping time to a given supermartingale, and we omit the notation $\overline{(\cdot)}$.

Calculating expectations under Q

Here we collect important results of Yoeurp [Yoe85] that allow to rewrite certain expectations under Q as expectations under P. More precisely, let Z be a nonnegative supermartingale with $E(Z_0) = 1$ and with Doob-Meyer decomposition $Z = Z_0 + M - D$, where M is a local martingale starting in zero, and D is an adapted process, almost surely nondecreasing and càdlàg. Let τ and Q be a stopping time and a probability measure, such that (Z, τ) is the Kunita-Yoeurp decomposition of Q with respect to P.

Recall that if N is a local martingale, then a nondecreasing sequence of stopping times $(\rho_m)_{m\in\mathbb{N}}$ is called *localizing sequence* for N if N^{ρ_m} is a uniformly integrable martingale for every $m \in \mathbb{N}$, and if $P(\lim_{m\to\infty} \rho_m = \infty) = 1$.

Lemma 1.4.3. *Let $Z = Z_0 + M - D$, and let τ and Q be as described above. Let $(\rho_m)_{m\in\mathbb{N}}$ be a localizing sequence for M, such that every ρ_m is finite. Then we have for every bounded predictable process Y and for every $m \in \mathbb{N}$ that*

$$E_Q(Y_\tau^{\rho_m}) = E_P\left(Y_{\rho_m} Z_{\rho_m} + \int_0^{\rho_m} Y_s \mathrm{d}D_s\right). \tag{1.14}$$

Proof. This is part of Proposition 9 of [Yoe85]. For the convenience of the reader, we provide a proof. First consider a simple process of the form $Y_s(\omega) = X(\omega)1_{(t,\infty]}(s)$ for some bounded $\mathcal{F}_t$–measurable X. For such Y we get from (1.10) that

$$\begin{aligned} E_Q(Y_\tau^{\rho_m}) &= E_Q(X1_{(t,\infty]}(\rho_m \wedge \tau)) = E_Q(X1_{\{t<\rho_m\}}1_{\{t<\tau\}}) = E_P(X1_{\{t<\rho_m\}}Z_t) \\ &= E_P(X1_{\{t<\rho_m\}}(Z_0 + M_t - D_t)) = E_P(X1_{\{t<\rho_m\}}(Z_0 + M_t^{\rho_m} - D_t)). \end{aligned}$$

Now we use that M^{ρ_m} is a uniformly integrable martingale, and that $X1_{\{t<\rho_m\}}$ is $\mathcal{F}_t$–measurable, to replace $M_t^{\rho_m}$ by $M_\infty^{\rho_m} = M_{\rho_m}$. Moreover, we have

$$\begin{aligned} X1_{\{t<\rho_m\}}D_t &= X1_{\{t<\rho_m\}}D_{\rho_m} - X1_{\{t<\rho_m\}}(D_{\rho_m} - D_t) \\ &= X1_{\{t<\rho_m\}}D_{\rho_m} - \int_0^{\rho_m} Y_s \mathrm{d}D_s, \end{aligned}$$

which proves (1.14) for such simple Y. The general case now follows from the monotone class theorem. $\square$

Corollary 1.4.4. *Let Y be a bounded adapted process that is P–almost surely càdlàg. Define*

$$Y_t^{\tau-}(\omega) := Y_t(\omega)1_{\{t<\tau(\omega)\}} + \limsup_{s\to\tau(\omega)-} Y_s(\omega)1_{\{t\geq\tau(\omega)\}}.$$

Let Z and (ρ_m) be as in Lemma 1.4.3. Then

$$E_Q(Y_{\rho_m}^{\tau-}) = E_P\left(Y_{\rho_m} Z_{\rho_m} + \int_0^{\rho_m} Y_{s-} \mathrm{d}D_s\right).$$

Proof. This is a slight generalization of (2.4) in [Yoe85]. Define $Y_t^-(\omega) = Y_{t-}(\omega) = \limsup_{s\to t-} Y_s$ for $t > 0$, and $Y_0^- = Y_0$. Then Y^- is a predictable process, because it is the pointwise limit of the step functions

$$Y_t^n = Y_0 1_{\{0\}}(t) + \sum_{k\geq 0} \limsup_{s\to k2^{-n}-} Y_s 1_{(k2^{-n},(k+1)2^{-n}]}(t).$$

Therefore, we can apply Lemma 1.4.3 to Y^-. Observe that

$$Y_{\rho_m}^{\tau-} = Y_{\rho_m} 1_{\{\tau>\rho_m\}} + Y_{\tau-} 1_{\{\tau\leq\rho_m\}} = Y_{\rho_m} 1_{\{\tau>\rho_m\}} + (Y^-)_\tau 1_{\{\tau\leq\rho_m\}}.$$

Now (1.11) implies that $E_Q(Y_{\rho_m} 1_{\{\tau>\rho_m\}}) = E_P(Y_{\rho_m} Z_{\rho_m})$, whereas (1.14) and then again (1.11) applied to the second term on the right hand side give

$$\begin{aligned} E_Q((Y^-)_\tau 1_{\{\tau\leq\rho_m\}}) &= E_Q((Y^-)_\tau^{\rho_m}) - E_Q((Y^-)_{\rho_m} 1_{\{\tau>\rho_m\}}) \\ &= E_P\left(Y_{\rho_m-} Z_{\rho_m} + \int_0^{\rho_m} Y_{s-} \mathrm{d}D_s\right) - E_P\left(Y_{\rho_m-} Z_{\rho_m}\right) \\ &= E_P\left(\int_0^{\rho_m} Y_{s-} \mathrm{d}D_s\right). \end{aligned}$$

□

1.4.2. The predictable case

We still assume that $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ is a probability space on which it is possible to solve the Kunita-Yoeurp problem. Let S be a d–dimensional predictable semimartingale, let $\mathcal{W}_1$ be defined as in (1.1), and let Z be a supermartingale density for $\mathcal{W}_1$. Here we examine the structure of S and Z closer. This will allow us to apply Lemma 1.4.3 to deduce that $S^{\tau-}$ is a local martingale under the dominating measure associated to Z. Note that Yoeurp [Yoe85] also establishes a generalized Girsanov formula, which we could apply directly rather than using Lemma 1.4.3. However, the use of Lemma 1.4.3 turns out to be rather instructive, and it allows us to obtain some insight into why the non-predictable case is more complicated.

Remark 1.4.5. Observe that, thanks to predictability, $S - S_0$ is almost surely locally bounded. This follows from I.2.16 of [JS03], which says that for $a > 0$ there exists an announcing sequence for the entrance time of S into $\{x \in \mathbb{R}^d : |x| \geq a\}$. In view of Corollary 1.1.4 it would therefore suffice to assume that Z is a supermartingale density for $\mathcal{W}_{1,s}$. Then S is a semimartingale and Z is a supermartingale density for $\mathcal{W}_1$.

Since $S - S_0$ is locally bounded, it is even a special semimartingale (see [JS03], I.4.23 (iv)). That is, there exists a unique decomposition

$$S = S_0 + M + D, \tag{1.15}$$

where M is a local martingale with $M_0 = 0$, and D is a predictable process of finite variation with $D_0 = 0$. Thus, $M = S - S_0 - D$ is predictable. But any predictable right-continuous local martingale is continuous ([JS03], Corollary I.2.31). Therefore, S is of the form (1.15) with continuous M. But then also D must be continuous, because (NA1) implies $\mathrm{d}D^i \ll \mathrm{d}\langle M^i\rangle$ for $i = 1, \dots, d$, where $M = (M^1, \dots, M^d)$ and $D = (D^1, \dots, D^d)$. This is a well known fact, see for example Ankirchner's Ph.D. thesis [Ank05], Lemma 9.1.2. Otherwise one could find a predictable process H^i which satisfies $H^i \cdot M^i \equiv 0$, but for which $H^i \cdot D^i$ is increasing; this would contradict $\mathcal{K}_1$ being bounded in probability. Therefore, D and then also S must be continuous.

In fact S must satisfy the *structure condition* as defined by Schweizer [Sch95]. Recall that $L^2_{\mathrm{loc}}(M)$ is the space of progressively measurable processes $(\lambda_t)_{t\geq 0}$ that are locally square integrable with respect to M, i.e. such that

$$\int_0^t \sum_{i,j=1}^d \lambda^i_s \lambda^j_s \mathrm{d}\langle M^i, M^j\rangle_s < \infty$$

for every $t > 0$. For details see [JS03], III.4.3.

Definition 1.4.6. Let $S = S_0 + M + D$ be a d–dimensional special semimartingale with locally square-integrable M. Define

$$C_t = \sum_{i=1}^d \langle M^i\rangle_t \qquad \text{and for } 1 \leq i, j \leq d: \qquad \sigma^{ij}_t = \frac{\mathrm{d}\langle M^i, M^j\rangle_t}{\mathrm{d}C_t}.$$

Note that σ exists by the Kunita-Watanabe inequality. Then S satisfies the *structure condition* if $\mathrm{d}D^i \ll \mathrm{d}\langle M^i\rangle$ for all $1 \leq i \leq d$, with predictable derivative $\alpha^i_t = \mathrm{d}D^i_t/\mathrm{d}\langle M^i\rangle_t$, and if there exists a predictable process $\lambda_t = (\lambda^1_t, \dots, \lambda^d_t) \in L^2_{\mathrm{loc}}(M)$, such that for $i = 1, \dots, d$ we have $\mathrm{d}C(\omega) \otimes P(\mathrm{d}\omega)$–almost everywhere

$$(\sigma\lambda)^i = \alpha^i \sigma^{ii}. \tag{1.16}$$

Note that λ might not be uniquely determined, but the stochastic integral $\int \lambda \mathrm{d}M$ does not depend on the choice of λ, see [Sch95]. If

$$\int_0^\infty \sum_{i,j=1}^d \lambda^i_t \sigma^{ij}_t \lambda^j_t \mathrm{d}C_t < \infty, \tag{1.17}$$

then we say that S satisfies the *structure condition until* ∞.

Recall that two one dimensional local martingales L and N are called *strongly orthogonal* if LN is a local martingale. If L and N are multidimensional, then we call them strongly orthogonal if all their components are strongly orthogonal. Also recall that the *stochastic exponential* of a semimartingale X is defined by the SDE

$$\mathcal{E}(X)_t = 1 + \int_0^t \mathcal{E}(X)_{s-} \mathrm{d}X_s, \qquad t \geq 0.$$

Finally we recall that every nonnegative supermartingale Y satisfies $Y_{\tau+t} \equiv 0$ for all

$t \geq 0$, where $\tau = \inf\{t \geq 0 : Y_{t-} = 0 \text{ or } Y_t = 0\}$.

Let us write $\mathrm{d}X_t \sim \mathrm{d}Y_t$ if $\mathrm{d}(X-Y)_t$ is the differential of a local martingale.

Lemma 1.4.7. *Let $S = S_0 + M + D$ be a predictable semimartingale and suppose that Z is a supermartingale density for S. Then S satisfies the structure condition until ∞, and*

$$\mathrm{d}Z_t = Z_{t-}(-\lambda_t \mathrm{d}M_t + \mathrm{d}N_t - \mathrm{d}B_t), \tag{1.18}$$

where λ satisfies (1.16) *and* (1.17)*, N is a local martingale that is strongly orthogonal to M, B is increasing, and $\mathcal{E}(N-B)_\infty > 0$.*

Conversely, if a predictable process S satisfies the structure condition until ∞, and if Z is defined by (1.18) *with $Z_0 = 1$, then Z is a supermartingale density for S.*

In particular, for predictable S, the structure condition until ∞ is equivalent to (NA1).

Proof. This is essentially Proposition 3.2 of Larsen and Žitković [LŽ07] in infinite time. We provide a slightly simplified version of their proof, because later we will need some results obtained during the proof.

Let Z be a supermartingale density. Since Z is strictly positive, it is of the form $dZ_t = Z_{t-}(dL_t - dB_t)$ for a local martingale L and a predictable increasing process B. Since M is continuous, there exists a predictable process $\lambda \in L^2_{\mathrm{loc}}(M)$, such that $dL_t = \lambda_t dM_t + dN_t$, where N is a local martingale that is strongly orthogonal to all components of M, see [JS03], Theorem III.4.11. Moreover,

$$0 < Z_\infty = Z_0 \mathcal{E}(\lambda \cdot M + N - B)_\infty = Z_0 \mathcal{E}(\lambda \cdot M)_\infty \mathcal{E}(N-B)_\infty,$$

which is only possible if λ satisfies (1.17) and if $\mathcal{E}(N-B)_\infty > 0$. It only remains to show that λ also satisfies (1.16).

Let H be a 1–admissible strategy. Write $W^H := 1 + H \cdot S$ for the wealth process generated by H. Then $W^H Z$ is a nonnegative supermartingale. Since Z is strictly positive, we must have $W^H_t \equiv 0$ for $t \geq \tau^H := \inf\{s \geq 0 : W^H_{s-} = 0 \text{ or } W^H_s = 0\}$. Therefore, we may assume without loss of generality that $H_t = H_t 1_{\{t<\tau^H\}}$ for all $t \geq 0$. Define $\pi_t := H_t / W^H_{t-}$, where we interpret $0/0 = 0$ as before. Then

$$W^H_t = 1 + (H \cdot S)_t = 1 + \int_0^t \pi_s W^H_{s-} \mathrm{d}S_s.$$

In other words, every wealth process is of the form $W^H = \mathcal{E}(\pi \cdot S)$ for a suitable integrand π. To simplify matters, we slightly abuse notation and write W^π instead of W^H.

Integration by parts applied to ZW^π gives

$$\begin{aligned} \mathrm{d}(ZW^\pi)_t &= W^\pi_{t-}\mathrm{d}Z_t + Z_{t-}\pi_t W^\pi_{t-}\mathrm{d}S_t + \mathrm{d}[W^\pi, Z]_t \\ &= W^\pi_{t-}Z_{t-}(\lambda_t \mathrm{d}M_t + \mathrm{d}N_t - \mathrm{d}B_t) + Z_{t-}\pi_t W^\pi_{t-}(\mathrm{d}M_t + \mathrm{d}D_t) \\ &\quad + W^\pi_{t-}Z_{t-}\mathrm{d}[\pi \cdot (M+D), \lambda \cdot M + N - B]_t \\ &\sim -W^\pi_{t-}Z_{t-}\mathrm{d}B_t + Z_{t-}\pi_t W^\pi_{t-}\mathrm{d}D_t + W^\pi_{t-}Z_{t-}\mathrm{d}\langle \pi \cdot M, \lambda \cdot M\rangle_t, \end{aligned} \tag{1.19}$$

where we used that M and D are continuous. By orthogonality, $[H\cdot M, N] = \langle H\cdot M, N\rangle$ is a continuous martingale of finite variation, starting at zero, which must therefore vanish.

Let now C and σ be as described in Definition 1.4.6. Then Theorem III.4.5 of [JS03] implies that the bracket $\langle \pi\cdot M, \lambda\cdot M\rangle$ can be rewritten as

$$\begin{aligned} \mathrm{d}(ZW^\pi)_t &\sim W^\pi_{t-}Z_{t-}\left(-\mathrm{d}B_t + \pi_t\mathrm{d}D_t + \mathrm{d}\langle \pi\cdot M, \lambda\cdot M\rangle_t\right) \\ &= W^\pi_{t-}Z_{t-}\left(-\mathrm{d}B_t + \sum_{i=1}^d \pi^i_t\left(\mathrm{d}D^i_t + \sum_{j=1}^d \sigma^{ij}_t\lambda^j_t\mathrm{d}C_t\right)\right). \end{aligned} \tag{1.20}$$

Assume now that there exists an $i\in\{1,\dots,d\}$ for which the continuous process of finite variation

$$X^i_t = D^i_t + \sum_{j=1}^d \int_0^\cdot \sigma^{ij}_s\lambda^j_s\mathrm{d}C_s$$

is not evanescent. We claim that then there exists a 1–admissible strategy π for which the finite variation part of (ZW^π) is increasing on a small time interval: By the predictable Radon-Nikodym theorem of Delbaen and Schachermayer [DS95b], Theorem 2.1 b), there exists a predictable γ^i with values in $\{-1,1\}$, such that $\int_0^\cdot \gamma^i_s\mathrm{d}D^i_s = V^i$, where V^i denotes the total variation process of X^i. Note that [DS95b] work with complete filtrations, but given the $(\mathcal{F}^P_t)$–predictable $\widetilde{\gamma}^i$ that they construct, we can apply Lemma A.4 to obtain a $(\mathcal{F}_t)$–predictable γ^i that is indistinguishable from $\widetilde{\gamma}^i$.

Let now $m\in\mathbb{N}$ and set $\pi^j_t := m\delta_{ij}\gamma^i_t$ for $j=1,\dots,d.$ into (1.20). Then

$$\mathrm{d}(ZW^\pi)_t \sim W^\pi_{t-}Z_{t-}\left(-\mathrm{d}B_t + m\mathrm{d}V^i_t\right).$$

Since V^i is an increasing process that is not constant, there exists $m\in\mathbb{N}$ such that $-\mathrm{d}B_t + m\mathrm{d}V^i_t$ is locally strictly increasing with positive probability. Since π is bounded, we obtain that $W^\pi_{t-} > 0$ for all $t\geq 0$, and of course also $Z_{t-} > 0$ for all $t\geq 0$. Therefore, the finite variation part of $W^\pi Z$ is locally strictly increasing with positive probability, a contradiction to ZW^π being a supermartingale.

Thus, X^i is evanescent. Recall that $\mathrm{d}D^i \ll \mathrm{d}\langle M^i\rangle = \sigma^{ii}\mathrm{d}C$, and therefore there exists a predictable process α^i for which

$$0 \equiv \left(\mathrm{d}D^i_t + \sum_{j=1}^d \sigma^{ij}_t\lambda^j_t\mathrm{d}C_t\right) = \left(\alpha^i_t\sigma^{ii}_t + (\sigma_t\lambda_t)^i\right)\mathrm{d}C_t,$$

so that

$$\alpha^i\sigma^{ii} = -(\sigma\lambda)^i \quad \mathrm{d}C(\omega)\otimes P(\mathrm{d}\omega) - \text{almost everywhere}, \tag{1.21}$$

i.e. (1.16) is satisfied, and the proof of the first part is complete.

The converse direction is easy and follows directly from (1.19). □

Remark 1.4.8. For later reference we remark that if M and D are not necessarily continuous, then a priori we only know that $\mathrm{d}Z_t = Z_{t-}(\mathrm{d}N_t - \mathrm{d}B_t)$, for a local martingale N and a predictable process of finite variation B. The finite variation part of $W^\pi Z$ is then given by

$$\mathrm{d}(W^\pi Z)_t \sim W^\pi_{t-} Z_{t-}(-\mathrm{d}B_t + \pi_t \mathrm{d}D_t + \mathrm{d}[\pi \cdot M, N]_t - \mathrm{d}[\pi \cdot D, B]_t), \tag{1.22}$$

which can be derived similarly as (1.19). Here we used that if L is a local martingale and if D is predictable process of finite variation, then $[L, D]$ is a local martingale, see Proposition I.4.49 of [JS03].

Next we will show that $S^{\tau-}$ is a local martingale under the measure Q that is associated to Z. But first we observe that if Z is a supermartingale density, then SZ is not necessarily a local martingale.

Corollary 1.4.9. *Let Z and S be as in Lemma 1.4.7. Then ZS^i is a local supermartingale if and only if $S^i \geq 0$ on the support of the measure $\mathrm{d}B$. If $S^i \geq 0$ identically, then ZS^i is a supermartingale.*

The process ZS^i is a local martingale if and only if $S^i = 0$ on the support of the measure $\mathrm{d}B$.

Proof. Integration by parts and (1.18) imply that

$$\begin{aligned}
\mathrm{d}(ZS^i)_t &= Z_{t-}\mathrm{d}S^i_t + S^i_{t-}\mathrm{d}Z_t + \mathrm{d}[S^i, Z]_t \\
&= Z_{t-}(\mathrm{d}M^i_t + \alpha^i_t \sigma^{ii}_t \mathrm{d}C_t) + S^i_{t-} Z_{t-}(-\lambda \mathrm{d}M_t + \mathrm{d}N_t - \mathrm{d}B_t) - Z_{t-}(\sigma\lambda)^i_t \mathrm{d}C_t \\
&\sim -S^i_{t-} Z_{t-} \mathrm{d}B_t,
\end{aligned}$$

where we used (1.16) in the last step.

The claim now follows easily since nonnegative local supermartingales are supermartingales by Fatou's lemma. □

Another consequence of Lemma 1.4.7 is that in the predictable case, the maximal elements among the supermartingale densities are always local martingales. This is important in the duality approach to utility maximization. For details we refer to [LŽ07].

We are now ready to prove Theorem 1.1.5 under the assumption that it is possible to solve the Kunita-Yoeurp problem.

Corollary 1.4.10. *Let S be a predictable semimartingale, and let Z be a supermartingale density for S. Let τ be a stopping time and Q be a probability measure, such that $(Z/E_P(Z_0), \tau)$ is the Kunita-Yoeurp decomposition of Q with respect to P. Then $S^{\tau-}$ is a Q–local martingale.*

Conversely, if $Q \gg P$ with Kunita-Yoeurp decomposition (Z, τ) with respect to P, and if $S^{\tau-}$ is a local martingale under Q, then Z is a supermartingale density for S.

Proof. We first show that $S^{\tau-}$ is Q–almost surely locally bounded: Let $\widetilde{\rho}_n := \inf\{t \geq 0 : |S^{\tau-}_t| \geq n\}$ for $n \in \mathbb{N}$. Since $S^{\tau-}$ was only required to be right-continuous P–almost surely and not identically, $\widetilde{\rho}_n$ is not necessarily a stopping time. But it is a $(\mathcal{F}_t^Q)$–stopping time. According to Lemma A.2, there exists a sequence of $(\mathcal{F}_t)$–stopping times $(\rho_n)_{n\in\mathbb{N}}$ such that $Q(\rho_n = \widetilde{\rho}_n) = 1$ for all $n \in \mathbb{N}$. Then $\sup_n \rho_n$ is a stopping time, and we obtain from (1.11) that

$$Q\left(\sup_n \rho_n < \tau\right) = E_P\left(Z_{\sup_n \rho_n} 1_{\{\sup_n \rho_n < \infty\}}\right) = 0, \tag{1.23}$$

where we used that $P(\sup_n \rho_n < \infty) = 0$. But $S_t^{\tau-}$ is constant for $t \geq \tau$, and therefore $\{\sup_n \rho_n \geq \tau\}$ is Q–almost surely contained in $\{\sup_n \rho_n = \infty\}$, showing that $S^{\tau-}$ is Q–almost surely locally bounded.

Let now $(\sigma_n)_{n\in\mathbb{N}}$ be a localizing sequence of finite stopping times for M, where $Z = Z_0 + M - D$. We define $\tau_n := \rho_n \wedge \sigma_n$. Let H be a strategy that is 1–admissible for $(S^{\tau-})^{\tau_n}$ under Q. Since $(H \cdot S)^{\tau-} = (H \cdot S^{\tau-})$, we can apply Corollary 1.4.4 (which extends from bounded Y to nonnegative Y via monotone convergence), to obtain

$$E_Q(1 + (H \cdot S^{\tau-})_{\tau_n}) = E_P\left((1 + (H \cdot S)_{\tau_n})Z_{\tau_n} + \int_0^{\tau_n} (1 + (H \cdot S)_{s-}) \mathrm{d}D_s\right).$$

But now (1.20) and (1.16) imply that

$$(1 + (H \cdot S))Z + \int_0^{\cdot} (1 + (H \cdot S)_{s-}) \mathrm{d}D_s = W^\pi Z + \int_0^{\cdot} W_{s-}^\pi \mathrm{d}D_s$$

is a nonnegative P–local martingale starting in 1, and therefore $E_Q((H \cdot S^{\tau-})_{\tau_n}) \leq 0$ for every strategy H that is 1–admissible under Q. Since $(S^{\tau-})^{\tau_n}$ is bounded, we easily conclude that it is a martingale.

The only remaining problem is that we only know $Q(\sup_n \tau_n \geq \tau) = 1$ and not $Q(\sup_n \tau_n = \infty) = 1$. But the same arguments as used above also show that $(S^{\tau-})^{\rho_n \wedge \tau_m}$ is a martingale for all $n, m \in \mathbb{N}$. Therefore, we can apply bounded convergence to obtain for all $s, t \geq 0$ that

$$E_Q((S^{\tau-})_{t+s}^{\rho_n} | \mathcal{F}_t) = \lim_{m\to\infty} E_Q((S^{\tau-})_{t+s}^{\rho_n\wedge\tau_m} | \mathcal{F}_t) = \lim_{m\to\infty} (S^{\tau-})_t^{\rho_n\wedge\tau_m} = (S^{\tau-})_t^{\rho_n}.$$

As we argued above, $Q(\sup_n \rho_n = \infty) = 1$, and therefore $S^{\tau-}$ is a Q–local martingale.

Conversely, let $S^{\tau-}$ be a Q–local martingale, and let H be a 1–admissible strategy for S under P. Define $\rho := \inf\{t \geq 0 : (H \cdot S^{\tau-})_t < -1\}$. Then $P(\rho < \infty) = 0$ and therefore $Q(\rho < \tau) = 0$ by the same argument as in (1.23). Hence, H is 1–admissible for $S^{\tau-}$ under Q. Now we can repeat the arguments in (1.3), to obtain that $Z_t = 1_{\{t<\tau\}}/\gamma_t$ is a supermartingale density for S, where we denoted $\gamma_t := (\mathrm{d}P/\mathrm{d}Q)|_{\mathcal{F}_t}$. □

Remark 1.4.11. For later reference, note that we only used once that S is predictable: it was only needed to obtain

$$E_P\left((1 + (H \cdot S)_{\tau_n})Z_{\tau_n} + \int_0^{\tau_n} (1 + (H \cdot S)_{s-}) \mathrm{d}Ds\right) \leq 1,$$

for which we applied Lemma 1.4.7 (and formula (1.20) from the proof of that lemma).

The general version of Theorem 1.1.5 is as follows:

Corollary 1.4.12 ("Correct formulation of Theorem 1.1.5")**.** *Let $(\mathcal{F}_t)$ be the right-continuous modification of a standard system. Let S be a predictable stochastic process that is almost surely right-continuous. Then S satisfies (NA1$_s$) if and and only there exists an enlarged probability space $(\overline{\Omega}, \overline{\mathcal{F}}, (\overline{\mathcal{F}}_t), \overline{P})$ and a dominating measure $\overline{Q} \gg \overline{P}$ with Kunita-Yoeurp decomposition $(\overline{Z}, \overline{\tau})$ with respect to $\overline{P}$, such that $\overline{S}^{\overline{\tau}-}$ is a $\overline{Q}$–local martingale.*

Proof. It remains to show that if $\overline{Q}$ exists, then S satisfies (NA1$_s$). But if $\overline{Q}$ exists, then Corollary 1.4.10 and Theorem 1.1.3 show that $\overline{S}$ even satisfies (NA1) on $(\overline{\Omega}, \overline{\mathcal{F}}, (\overline{\mathcal{F}}_t), \overline{P})$. Since this space is an enlargement of $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$, the process S must also satisfy (NA1). □

Remark 1.4.13. We argued above that a predictable process satisfying (NA1) must be continuous. Therefore, Corollary 1.4.12 is not much more general than Ruf [Ruf13], where it is shown that a diffusion S that satisfies (NA1) admits a dominating measure Q under which $S^{\tau-}$ is a local martingale. However, one difference is that [Ruf13] only shows for supermartingale densities which are local martingales that they correspond to dominating local martingale measures. Here we show that in the predictable case this is true for all supermartingale densities. Also, we show equivalence between (NA1) and the existence of a dominating local martingale measure, and not only that (NA1) implies the existence of Q. Of course, as it is usually the case for this type of result, the reverse direction is much easier.

1.4.3. The general case

We start the treatment of the non-predictable case with two examples that illustrate why it is natural to consider dominating local martingale measures for $S^{\tau-}$ rather than for S.

Example 1.4.14. If S is optional and if Q is a dominating local martingale measure for S rather than for $S^{\tau-}$, then S does not need to satisfy (NA1): Let τ be exponentially distributed with parameter 1 under Q. Define $S_t = e^t 1_{\{t<\tau\}}$ for $t \in [0,1]$. Since time is finite, S is a uniformly integrable martingale. Therefore, $\mathrm{d}P = S_1 \mathrm{d}Q$ is absolutely continuous with respect to Q. But under P we have $S_t = e^t$ for all $t \in [0,1]$. Clearly S does not satisfy (NA1) under P, even though Q is a dominating martingale measure for S. Note that $S^{\tau-}$ is not a local martingale under Q because $S^{\tau-}_t = e^t$ for all $t \in [0,1]$.

Recall that a stopping time τ is called *foretellable* under a probability measure P if there exists an increasing sequence (τ_n) of stopping times, such that $P(\tau_n < \tau) = 1$ for every n, and such that $P(\sup_n \tau_n = \tau) = 1$. In this case (τ_n) is called an *announcing sequence* for τ. Every predictable time is foretellable under any probability measure, see Theorem I.2.15 and Remark I.2.16 of [JS03].

Example 1.4.15. Let S be a semimartingale under P and let $Q \gg P$ be a dominating measure with Kunita-Yoeurp decomposition (Z, τ) with respect to P. Assume that

τ is not foretellable under Q. Then there exists an adapted process $\widetilde{S}$ which is P–indistinguishable from S, such that $\widetilde{S}$ is not a Q–local martingale: Let $x \in \mathbb{R}^d$ and define $\widetilde{S}^x_t = S_t 1_{\{t<\tau\}} + x 1_{\{t \geq \tau\}}$, which is P–indistinguishable from S since $P(\tau = \infty) = 1$. If $\widetilde{S}^x$ is a Q–local martingale, then $\tau^x_n = \inf\{t \geq 0 : |\widetilde{S}^x_t| \geq n\}$, $n \in \mathbb{N}$, defines a localizing sequence. In particular, (τ^x_n) converges Q–almost surely to infinity as n tends to ∞, and thus $Q(\lim_{n\to\infty} \tau^x_n \geq \tau) = 1$. Since τ is not foretellable under Q, there must exist $n \in \mathbb{N}$ for which $Q(\tau^x_n \geq \tau) > 0$. Moreover, we have

$$E_Q(S_0) = E_Q(\widetilde{S}^x_{\tau^x_n}) = E_Q(S_{\tau^x_n} 1_{\{\tau^x_n < \tau\}}) + xQ(\tau^x_n \geq \tau).$$

Since $\tau^x_n = \tau^y_n$ for all $|x| < n$, $|y| < n$, we obtain a contradiction by letting x vary through the ball of radius $n-1$.

These two examples show that given $Q \gg P$, it is important to choose a good version of S if we want to obtain a Q–local martingale. All the results obtained so far indicate that this good version should be $S^{\tau-}$. Maybe somewhat surprisingly, this is not true in general, as we demonstrate in the following example.

Example 1.4.16. Let $(L_t)_{t\in[0,1]}$ be a Lévy process under Q, with jump measure $\nu = \delta_1 + \delta_{-1}$ and drift $b \in \mathbb{R}$. That is, $L_t = N^1_t - N^2_t + bt$, where N^1 and N^2 are independent Poisson processes. Let $a > |b|$ and let ρ be an exponential random variable with parameter a, such that ρ is independent from L. Define $\tau = \rho$ if $\rho \leq 1$, and $\tau = \infty$ otherwise. Then $(e^{at} 1_{\{t<\tau\}})_{t\in[0,1]}$ is a uniformly integrable martingale, and therefore it defines a probability measure $\mathrm{d}P = e^a 1_{\{1<\tau\}} \mathrm{d}Q$. Since τ and L are independent, L has the same distribution under P as under Q. The Kunita-Yoeurp decomposition of Q with respect to P is given by $((e^{-at})_{t\in[0,1]}, \tau)$.

We claim that $Z = e^{-a\cdot}$ is a supermartingale density for L. Let $(\pi_t W^\pi_{t-})$ be a strategy for L, where W^π is the wealth process obtained by investing in this strategy. Such a strategy is 1–admissible if and only if $|\pi_t| \leq 1$ for all $t \in [0,1]$. Moreover, we get from (1.22) that

$$d(ZW^\pi)_t \sim -W^\pi_{t-} Z_{t-} a dt + Z_{t-} \pi_t W^\pi_{t-} b dt = W^\pi_{t-} Z_{t-} (\pi_t b - a) dt.$$

Since $W^\pi Z \geq 0$ and since $\pi_t b - a < 0$ (recall that $a > |b|$), the drift rate is negative. Therefore, ZW^π is a local supermartingale, and since it is a nonnegative process, it is a supermartingale.

Now τ is independent from L under Q, and L has no fixed jump times. Hence

$$Q(\Delta L_\tau \neq 0, \tau \leq 1) = \int_{[0,1]} Q(\Delta L_t \neq 0)(Q \circ \tau^{-1})(\mathrm{d}t) = 0,$$

which implies that $L^{\tau-} = L^\tau$, and this is clearly no Q–local martingale.

Remark 1.4.17. In the preceding example it is possible to show that the modified process

$$\widetilde{L}_t = L^{\tau-}_t - \frac{b}{a} 1_{\{t \geq \tau\}} \tag{1.24}$$

is a Q–martingale. More generally, we expect that given a semimartingale S, a supermartingale density Z for S, and a measure $Q \gg P$ with Kunita-Yoeurp decomposition (Z, τ) with respect to P, there should always exist a version $\widetilde{S}$ that is P–indistinguishable from S, such that $\widetilde{S}$ is a Q–local martingale. But as (1.24) shows, we will need to take different $\widetilde{S}$ for different supermartingale densities. Therefore, this seems somewhat unnatural, and we will not pursue it further.

Note that all three examples had one thing in common: τ was not foretellable under Q. It turns out that things get much simpler if τ is foretellable under Q. But if $(\tau_n)_{n\in\mathbb{N}}$ is an announcing sequence for τ, then we obtain from (1.11) that

$$\begin{aligned} 1 &= Q(\tau_n < \tau) = E_P(Z_{\tau_n} 1_{\{\tau_n < \infty\}}) \text{ for all } n \in \mathbb{N}, \text{ and} \\ 0 &= Q\left(\sup_n \tau_n < \tau\right) = E_P(Z_{\sup_n \tau_n} 1_{\{\sup_n \tau_n < \infty\}}). \end{aligned}$$

Since Z is strictly positive, we conclude that $(\tau_n)_{n\in\mathbb{N}}$ is a localizing sequence for Z under P, i.e. Z is a P–local martingale.

Therefore, we should look for supermartingale densities that are local martingales. We call such supermartingale densities *local martingale densities.* If $(S_t)_{t\in[0,T]}$ is one dimensional with finite terminal time $T < \infty$, it is shown by Kardaras [Kar12], Theorem 1.1, that local martingale densities exist if and only if (NA1) is satisfied. The proof is in the spirit of the article [KK07]. Takaoka [Tak13] solves the multidimensional case with finite terminal time. More precisely, it is easily deduced from Remark 7 of [Tak13] that for a locally bounded d–dimensional semimartingale $(S_t)_{t\in[0,T]}$, (NA1) is satisfied if and only if there exists a local martingale density. Takoaka's proof is based on the insight of Delbaen and Schachermayer [DS95c], that a change of numéraire can induce the (NA) property, even if previously there were arbitrage opportunities in the market. [Tak13] continues to show that a clever choice of numéraire preserves the (NA1) property, so that then the condition (NA) + (NA1) = (NFLVR) is satisfied, which permits to apply the Fundamental Theorem of Asset Pricing [DS94]. See also the recent preprint Song [Son13], where an alternative proof of Takaoka's result is given that does not use the Fundamental Theorem of Asset Pricing. Roughly speaking, this is achieved by combining the philosophies behind [KK07] and [Tak13].

Of course [Kar12], [Tak13], and [Son13] all work with complete filtrations, but given a local martingale density $\widetilde{Z}$ that is $(\mathcal{F}_t^P)$–adapted, there exists an indistinguishable process Z that is $(\mathcal{F}_t)$–adapted, see Lemma A.4.

Lemma 1.4.18. *Let $(S_t)_{t\in[0,T]}$ be a locally bounded semimartingale on a finite time horizon $T < \infty$, and let Z be a local martingale density for S. Let τ be a stopping time and Q be a probability measure, such that $(Z/E_P(Z_0), \tau)$ is the Kunita-Yoeurp decomposition of Q with respect to P. Then $S^{\tau-}$ is a Q–local martingale.*

Conversely, if $Q \gg P$ has Kunita-Yoeurp decomposition (Z, τ) with respect to P, and if $S^{\tau-}$ is a Q–local martingale, then Z is a supermartingale density for S.

Proof. The proof is very similar to the one of Corollary 1.4.10. Recall from Remark 1.4.11 that we only used the predictability of S once in the proof of Corollary 1.4.10, to obtain

$$E_Q((H \cdot S^{\tau-})_{\sigma_n}) \leq 0 \tag{1.25}$$

for all strategies H that are 1–admissible for $(S^{\tau-})^{\sigma_n}$ under Q. Here (σ_n) was a localizing sequence of finite stopping times for M under P, where $Z = Z_0 + M - D$. Therefore, it suffices to show that (1.25) always holds if Z has the decomposition $Z = Z_0 + M$, i.e. if $D = 0$, even if S is not predictable.

So let (σ_n) be a localizing sequence of finite stopping times for the local martingale Z under P, and let H be a strategy that is 1–admissible for $(S^{\tau-})^{\sigma_n}$ under Q (and then also for S^{σ_n} under P). We apply Corollary 1.4.4 with $D = 0$, and obtain

$$E_Q(1 + (H \cdot S^{\tau-})_{\sigma_n}) = E_P((1 + (H \cdot S)_{\sigma_n})Z_{\sigma_n}) \leq 1,$$

where the last step follows because Z is a supermartingale density. From here on we can just copy the proof of Corollary 1.4.10. □

We obtain our main result, a weak fundamental theorem of asset pricing:

Corollary 1.4.19 ("Correct formulation of Theorem 1.1.6")**.** *Let $(\mathcal{F}_t)_{t\in[0,T]}$ be the right-continuous modification of a standard system. Let $S = (S_t)_{t\in[0,T]}$ be a locally bounded, right-continuous stochastic process. Then S satisfies (NA1$_s$) if and and only there exists an enlarged probability space $(\overline{\Omega}, \overline{\mathcal{F}}, (\overline{\mathcal{F}}_t)_{t\in[0,T]}, \overline{P})$, and a dominating measure $\overline{Q} \gg \overline{P}$ with Kunita-Yoeurp decomposition $(\overline{Z}, \overline{\tau})$ with respect to $\overline{P}$, such that $\overline{S}^{\overline{\tau}-}$ is a $\overline{Q}$–local martingale.*

Remark 1.4.20. There is another subset of supermartingale densities of which one might expect that they correspond to local martingale measures for $S^{\tau-}$: the maximal elements among the supermartingale densities. A supermartingale density Z is called *maximal* if it is indistinguishable from any supermartingale density Y that satisfies $Y_t \geq Z_t$ for all $t \geq 0$. If S is not continuous, then some maximal supermartingale densities are supermartingales and not local martingales, see Example 5.1' of Kramkov and Schachermayer [KS99].

But such Z will usually *not* correspond to local martingale measures for S. Assume for example that we are in the situation described in Theorem 2.2 of [KS99], i.e. we have a dual optimizer Z and a primal optimizer H for a certain utility maximization problem. Then point iii) of this Theorem 2.2 states that $(1 + (H \cdot S))Z$ is a uniformly integrable martingale. If we assume now that Z is not a local martingale, as is the case in Example 5.1' of [KS99], and if (τ_n) is a localizing sequence of finite stopping times for the local martingale part M of $Z = Z_0 + M - D$, then we obtain from Corollary 1.4.4 that

$$\begin{aligned} E_Q(1 + (H \cdot S^{\tau-})_{\tau_n}) &= E_P((1 + (H \cdot S)_{\tau_n})Z_{\tau_n}) + E_P\left(\int_0^{\tau_n} (1 + (H \cdot S)_{s-})\mathrm{d}D_s\right) \\ &= 1 + E_P\left(\int_0^{\tau_n} (1 + (H \cdot S)_{s-})\mathrm{d}D_s\right), \end{aligned} \tag{1.26}$$

where we used that $(1+(H\cdot S))Z$ is a uniformly integrable martingale. Since H is optimal, the wealth process $(1+(H\cdot S)_{s-})_{s\geq 0}$ will be strictly positive with positive probability. Since also $\mathrm{d}D \neq 0$ with positive probability, the expectation in (1.26) is strictly positive for sufficiently large n, and therefore $(H\cdot S^{\tau-})^{\tau_n}$ cannot be a Q–supermartingale, i.e. $S^{\tau-}$ cannot be a Q–local martingale.

1.5. Relation to filtration enlargements

Here we show that Jacod's criterion for initial filtration enlargements is in fact a criterion for the existence of a universal supermartingale density (to be defined below). We also treat general filtration enlargements. We show that if there exists a universal supermartingale density in an enlarged filtration, then a generalized version of Jacod's criterion is satisfied.

1.5.1. Jacod's criterion and universal supermartingale densities

Let $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},P)$ be a filtered probability space, and let $(\mathcal{G}^0_t)_{t\geq 0}$ be an initial filtration enlargement of $(\mathcal{F}_t)$, by which we mean that there there exists a random variable X such that $\mathcal{G}^0_t = \mathcal{F}_t \vee \sigma(X)$ for all $t\geq 0$. We define the right-continuous regularization of $(\mathcal{G}^0_t)$ by setting $\mathcal{G}_t := \bigcap_{s>t}\mathcal{G}^0_s$ for all $t\geq 0$. Note that we do not require $\mathcal{F}$, $(\mathcal{F}_t)_{t\geq 0}$, or $(\mathcal{G}_t)_{t\geq 0}$ to be complete, contrary to Jacod [Jac85] (although we *allow* them to be complete).

Recall that Hypothèse (H') is satisfied if all $(\mathcal{F}_t)$–semimartingales are $(\mathcal{G}_t)$-semimartingales.

We now give the classical formulation of Jacod's criterion, see [Jac85]. For this purpose we need to assume that X takes its values in a standard Borel space, which we denote by $(\mathbb{X},\mathcal{B})$. For the definition of standard Borel spaces see Parthasarathy [Par67], Definition V.2.2. For a detailed discussion see also Dellacherie [Del69], where standard Borel spaces are referred to as Lusin spaces. Note that $(\mathbb{X},\mathcal{B})$ is a standard Borel space provided that $\mathbb{X}$ is a Polish space and $\mathcal{B}$ its Borel σ–algebra.

If X takes its values in the standard Borel space $(\mathbb{X},\mathcal{B})$, then the regular conditional distribution

$$P_t(\omega,\mathrm{d}x) := P(X\in \mathrm{d}x|\mathcal{F}_t)(\omega)$$

exists for all $t\geq 0$, see Durrett [Dur10], Theorem 5.1.9 (Durrett calls standard Borel spaces "nice spaces"). We write P_X for the distribution of X. Jacod's criterion states that Hypothèse (H') is satisfied provided that for every $t\geq 0$ almost surely

$$P_t(\omega,\mathrm{d}x) \ll P_X(\mathrm{d}x). \tag{1.27}$$

Note that this statement only makes sense if the set $\{\omega : P_t(\omega,\mathrm{d}x)\ll P_X(\mathrm{d}x)\}$ is $\mathcal{F}$–measurable. But since the σ–algebra of a standard Borel space is countably generated (see also [PR13]), it is easily verified that this is indeed the case. Below we give an alternative proof of Jacod's result, and we relate it to the existence of a *universal supermartingale density*.

First observe that Hypothèse (H') is satisfied if and only if all nonnegative $(\mathcal{F}_t)$–martingales are $(\mathcal{G}_t)$–semimartingales: This follows by decomposing every $(\mathcal{F}_t)$–local martingale into a sum of a locally bounded local martingale and a local martingale of finite variation, by observing that every bounded process can be made nonnegative by adding a deterministic constant, and from the fact that local semimartingales are semimartingales (see Protter [Pro04], Theorem II.6).

Definition 1.5.1. Let $(\mathcal{G}_t)$ be a filtration enlargement of $(\mathcal{F}_t)$. Let Z be a $(\mathcal{G}_t)$–adapted process that is almost surely càdlàg, such that $P(Z_t > 0) = 1$ for all $t \geq 0$. Then Z is called *universal supermartingale density* for $(\mathcal{G}_t)$ if ZM is a $(\mathcal{G}_t)$–supermartingale for every nonnegative $(\mathcal{F}_t)$–supermartingale M.

Note that here we do not require Z_∞ to be positive, unlike in the previous sections. This is because here we are interested in the semimartingale property and not primarily in the (NA1) property. Local semimartingales are semimartingales, and therefore it suffices to verify the $(\mathcal{G}_t)$–semimartingale property of M on $[0, t]$ for every $t \geq 0$. Hence, it suffices if $Z_t > 0$ for every $t \geq 0$.

Also note that we required ZM to be a $(\mathcal{G}_t)$–supermartingale for every nonnegative $(\mathcal{F}_t)$–supermartingale M, and not just for nonnegative $(\mathcal{F}_t)$–martingales. This has the advantage that now we see immediately that in finite time every process satisfying (NA1) under $(\mathcal{F}_t)$ satisfies also (NA1) under $(\mathcal{G}_t)$, provided that there exists a universal supermartingale density Z: if Y is a $(\mathcal{F}_t)$–supermartingale density for S, then ZY is a $(\mathcal{G}_t)$–supermartingale density for S. To extend this result to infinite time, we would have to change the definition of a universal supermartingale density by additionally requiring that $P(Z_\infty > 0) = 1$. To obtain a universal supermartingale density under Jacod's criterion, we would then have to assume that also almost surely $P_\infty(\omega, \cdot) \ll P_X(\cdot)$. Here we do not pursue this further.

The first result of this section shows that Jacod's criterion is not so much a criterion for Hypothèse (H') to hold, but rather a criterion for the existence of a universal supermartingale density.

Proposition 1.5.2. *Let $(\mathcal{G}_t)$ be the right-continuous regularization of an initial enlargement of $(\mathcal{F}_t)$ with a random variable X taking its values in a standard Borel space. Assume Jacod's criterion* (1.27) *is satisfied. Then there exists a universal supermartingale density for $(\mathcal{G}_t)$.*

Proof. 1. Let $t \geq 0$. Without loss of generality we may assume that $\mathrm{d}P_t(\omega, \cdot) \ll \mathrm{d}P_X(\cdot)$ for *all* $\omega \in \Omega$. This can be achieved by setting $P_t(\omega, \cdot) := 0$ on the measurable set $\{\omega : P_t(\omega)$ does not satisfy $P_t(\omega, \cdot) \ll P_X(\cdot)\}$. Now we can apply a theorem of Doob, see [YM78], according to which there exists a $\mathcal{F}_t \otimes \mathcal{B}$–measurable random variable $Y_t : \Omega \times \mathbb{X} \to \mathbb{R}_+$, such that for every $\omega \in \Omega$ we have P_X–almost surely

$$Y_t(\omega, x) = \frac{\mathrm{d}P_t(\omega, \cdot)}{\mathrm{d}P_X}(x).$$

Note that Yor and Meyer [YM78] do not require complete σ–algebras. Let now $t, s \geq 0$. We first show that $P \otimes P_X$–almost surely

$$\{(\omega, x) : Y_t(\omega, x) = 0\} \subseteq \{(\omega, x) : Y_{t+s}(\omega, x) = 0\}. \tag{1.28}$$

Note that $Y_{t+s} \geq 0$, and therefore Fubini's theorem and the tower property of conditional expectations imply that

$$\begin{aligned}
&\int_{\Omega \times \mathbb{X}} 1_{\{Y_t(\omega,x)=0\}} Y_{t+s}(\omega, x) P \otimes P_X(\mathrm{d}\omega, \mathrm{d}x) \\
&\quad = \int_\Omega \int_{\mathbb{X}} 1_{\{Y_t(\omega,x)=0\}} P_{t+s}(\omega, \mathrm{d}x) P(\mathrm{d}\omega) = \int_\Omega \int_{\mathbb{X}} 1_{\{Y_t(\omega,X(\omega))=0\}} P(\mathrm{d}\omega) \\
&\quad = \int_\Omega \int_{\mathbb{X}} 1_{\{Y_t(\omega,x)=0\}} P_t(\omega, \mathrm{d}x) P(\mathrm{d}\omega) = 0,
\end{aligned}$$

where we used that $P_t(\omega, \cdot)$–almost surely $Y_t(\omega, \cdot) > 0$.

2. Define $\widetilde{Z}_t(\omega, x) := 1_{\{Y_t(\omega,x)>0\}} / Y_t(\omega, x)$ and

$$Z_t(\omega) := \widetilde{Z}_t(\omega, X(\omega)).$$

This Z is $(\mathcal{G}_t)$–adapted by construction. Let now M be a nonnegative $(\mathcal{F}_t)$–supermartingale. Let $s, t \geq 0$, let $A \in \mathcal{F}_t$, and $B \in \mathcal{B}(\mathbb{X})$. Then we can apply the tower property to obtain

$$\begin{aligned}
&E\left(1_A 1_B(X) M_{t+s} Z_{t+s}\right) = E\left(1_A M_{t+s} E(1_B(X) \tilde{Z}_{t+s}(\cdot, X) | \mathcal{F}_{t+s})\right) \\
&\quad = \int_\Omega 1_A(\omega) M_{t+s}(\omega) \int_{\mathbb{X}} 1_B(x) \widetilde{Z}_{t+s}(\omega, x) P_{t+s}(\omega, \mathrm{d}x) P(\mathrm{d}\omega) \\
&\quad = \int_\Omega 1_A(\omega) M_{t+s}(\omega) \int_{\mathbb{X}} 1_B(x) \frac{Y_{t+s}(\omega, x)}{Y_{t+s}(\omega, x)} 1_{\{Y_{t+s}(\omega,x)>0\}} P_X(\mathrm{d}x) P(\mathrm{d}\omega) \\
&\quad \leq \int_\Omega 1_A(\omega) M_{t+s}(\omega) \int_{\mathbb{X}} 1_B(x) 1_{\{Y_t(\omega,x)>0\}} P_X(\mathrm{d}x) P(\mathrm{d}\omega).
\end{aligned}$$

In the last step we used (1.28) and that $1_A(\omega) 1_B(x) M_{t+s}(\omega)$ is $P_X \otimes P$–almost surely nonnegative. Using the $(\mathcal{F}_t)$–supermartingale property of M in conjunction with Fubini's theorem, we obtain

$$\begin{aligned}
&\int_\Omega 1_A(\omega) M_{t+s}(\omega) \int_{\mathbb{X}} 1_B(x) 1_{\{Y_t(\omega,x)>0\}} P_X(\mathrm{d}x) P(\mathrm{d}\omega) \\
&\quad \leq \int_{\mathbb{X}} 1_B(x) \int_\Omega 1_A(\omega) M_t(\omega) 1_{\{Y_t(\omega,x)>0\}} P(\mathrm{d}\omega) P_X(\mathrm{d}x)
\end{aligned}$$

$$= \int_\Omega 1_A(\omega) \int_{\mathbb{X}} 1_B(x) M_t(\omega) \frac{Y_t(\omega,x)}{Y_t(\omega,x)} 1_{\{Y_t(\omega,x)>0\}} P_X(\mathrm{d}x) P(\mathrm{d}\omega)$$
$$= \int_\Omega 1_A(\omega) \int_{\mathbb{X}} 1_B(x) M_t(\omega) \widetilde{Z}_t(\omega,x) P_t(\omega,\mathrm{d}x) P(\mathrm{d}\omega) = E\left(1_A 1_B(X) M_t Z_t\right)$$

The monotone class theorem allows to pass from sets of the form $A \cap X^{-1}(B)$ to general sets in $(\mathcal{G}_t^0)$, and therefore MZ is a $(\mathcal{G}_t^0)$–supermartingale. Taking $M \equiv 1$, we see that also Z is a $(\mathcal{G}_t^0)$–supermartingale.

3. Let us show that Z_t is P–almost surely strictly positive for every $t \geq 0$. For this purpose it suffices to show that $P(\omega : Y_t(\omega, X(\omega)) = 0) = 0$. By the tower property we have

$$E(1_{\{Y_t(\cdot,X(\cdot))=0\}}) = \int_\Omega \int_{\mathbb{X}} 1_{\{Y_t(\omega,x)=0\}} P_t(\omega,\mathrm{d}x) P(\mathrm{d}\omega) = 0.$$

4. Z is not necessarily right-continuous, and also we did not show yet that ZM is a $(\mathcal{G}_t)$–supermartingale and not just a $(\mathcal{G}_t^0)$–supermartingale. But the construction of a right-continuous universal supermartingale density is now done exactly as in the proof of Theorem 1.3.1. The $(\mathcal{G}_t)$–supermartingale property of ZM follows also in the same way as in the proof of Theorem 1.3.1.

$\square$

Remark 1.5.3. If we are only interested whether Hypothèse (H') holds and not whether there exists a universal supermartingale density, then we can also work with the filtration $(\mathcal{G}_t^0)$ and not with its right-continuous regularization $(\mathcal{G}_t)$. Since Hypothèse (H') holds for $(\mathcal{G}_t)$ and since $(\mathcal{G}_t^0)$ is a filtration shrinkage of $(\mathcal{G}_t)$, Stricker's theorem implies that Hypothèse (H') is also satisfied for $(\mathcal{G}_t^0)$.

Remark 1.5.4. We could replace assumption (1.27) by $P_t(\omega,\mathrm{d}x) \gg P_X(\mathrm{d}x)$ or $P_t(\omega,\mathrm{d}x) \sim P_X(\mathrm{d}x)$. In the first case we could use the same proof as for Proposition 1.5.2 to obtain the existence of a nonnegative martingale Z, not necessarily strictly positive, such that ZM is a $(\mathcal{G}_t)$–supermartingale for every nonnegative $(\mathcal{F}_t)$–supermartingale M. In particular, then there exists an absolutely continuous measure $Q \ll P$, such that every locally bounded $(P,(\mathcal{F}_t))$–local martingale is a $(Q,(\mathcal{G}_t))$–local martingale. Since (NA) is related to the existence of absolutely continuous local martingale measures, see [DS95b], this indicates that the (NA) property may be stable under initial filtration enlargements that satisfy this "reverse Jacod condition". Note that it is much harder to satisfy this assumption. For example it will never be satisfied if X is $\mathcal{F}_t$–measurable for some $t \geq 0$.

If $P_t(\omega,\mathrm{d}x) \sim P_X(\mathrm{d}x)$, then the same proof as for Proposition 1.5.2 yields the existence of an equivalent measure $Q \sim P$, such that every nonnegative $(P,(\mathcal{F}_t))$–supermartingale is a nonnegative $(Q,(\mathcal{G}_t))$–supermartingale. In particular, then every locally bounded $(P,(\mathcal{F}_t))$–local martingale is a $(Q,(\mathcal{G}_t))$–local martingale. This condition has been studied by Amendinger, Imkeller and Schweizer [AIS98], as well as Amendinger [Ame00]. Obviously it is harder to satisfy than Jacod's condition or the reverse Jacod condition. In financial applications one may however assume that the knowledge of the "insider" is perturbed by a small Gaussian noise that is independent of $\mathcal{F}_\infty$ (or more generally by an independent noise with strictly positive density with respect to Lebesgue measure). Then $P_t(\omega, \mathrm{d}x) \sim P_X(\mathrm{d}x)$ is always satisfied.

1.5.2. Universal supermartingale densities and the generalized Jacod criterion

In the previous section we saw that for initial enlargements, Jacod's criterion is a *sufficient* condition for the existence of a universal supermartingale density. Here we show that for general filtration enlargements, a generalized version of Jacod's criterion is a *necessary* condition for the existence of a universal supermartingale density.

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ be a filtered probability space, such that $(\Omega, \mathcal{F})$ is a standard Borel space. We assume that there exists a filtration $(\mathcal{F}^0_t)_{t\geq 0}$, such that $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}^0_s$ for all $t \geq 0$. We also assume that $(\mathcal{G}^0_t)_{t\geq 0}$ is a filtration enlargement of $(\mathcal{F}^0_t)$, such that $\mathcal{G}^0_t \subseteq \mathcal{F}$ is countably generated for every $t \geq 0$, i.e. there exists a sequence of sets $(B^t_n)_{n\in\mathbb{N}}$ such that $\mathcal{G}^0_t = \sigma(B^t_1, B^t_2, \dots)$. Then $(\mathcal{G}_t)_{t\geq 0}$, defined by $\mathcal{G}_t := \bigcap_{s>t} \mathcal{G}^0_t$, is a filtration enlargement of $(\mathcal{F}_t)$.

The reason for choosing such a complicated set-up is that $\mathcal{G}_t$ will in general not be countably generated, even if $\mathcal{G}^0_s$ is countably generated for every $s \geq 0$. But in our argumentation below we will need $\mathcal{G}^0_t$ to be countably generated. On the other side, if we would only work with the non right-continuous filtration $(\mathcal{G}^0_t)$, then there would be little hope of constructing a right-continuous universal supermartingale density in the first place.

Since $(\Omega, \mathcal{F})$ is a standard Borel space, the regular conditional probabilities

$$P_t(\omega, \cdot) := P(\cdot|\mathcal{F}_t)(\omega)$$

exist. We say that the *generalized Jacod criterion* is satisfied if for all $s, t \geq 0$ almost surely

$$P_{t+s}|_{\mathcal{G}^0_t}(\omega, \cdot) \ll P_t|_{\mathcal{G}^0_t}(\omega, \cdot).$$

It is known that neither Jacod's criterion nor the generalized Jacod criterion are necessary conditions for Hypothèse (H') to hold. But the generalized Jacod criterion is a necessary condition for the existence of a universal supermartingale density for $(\mathcal{G}_t)$:

Proposition 1.5.5. *Assume that there exists a universal supermartingale density Z for $(\mathcal{G}_t)$. Then the generalized Jacod criterion is satisfied.*

Proof. 1. For every $A \in \mathcal{F}$ the process $M^A_t := E_P(1_A|\mathcal{F}_t)$, $t \geq 0$, is a nonnegative $(\mathcal{F}_t)$–martingale. Therefore, $M^A Z$ is a $(\mathcal{G}_t)$–supermartingale. Fix $s, t \geq 0$. Let

$A \in \mathcal{F}_{t+s}$ and $B \in \mathcal{G}_t$. Then for every $n \in \mathbb{N}$ we have that

$$\begin{aligned} E\left(1_A 1_B \frac{Z_{t+s}}{Z_t} 1_{\{Z_t \geq 1/n\}}\right) &= E\left(\frac{1_B 1_{\{Z_t \geq 1/n\}}}{Z_t} M^A_{t+s} Z_{t+s}\right) \\ &\leq E\left(\frac{1_B 1_{\{Z_t \geq 1/n\}}}{Z_t} M^A_t Z_t\right) \\ &= E(1_A E(1_B 1_{\{Z_t \geq 1/n\}} | \mathcal{F}_t)). \end{aligned}$$

Applying monotone convergence on both sides, we obtain

$$E\left(1_A 1_B \frac{Z_{t+s}}{Z_t}\right) \leq E(1_A E_P(1_B | \mathcal{F}_t)).$$

The same inequality holds if we replace Z_{t+s}/Z_t by a version $\widetilde{Z}_{t+s}/\widetilde{Z}_t$ that is strictly positive for *every* $\omega \in \Omega$. Since the inequality holds for all $A \in \mathcal{F}_{t+s}$, we conclude that

$$\int 1_B(\omega') \frac{\widetilde{Z}_{t+s}}{\widetilde{Z}_t}(\omega') P_{t+s}(\omega, \mathrm{d}\omega') \leq P_t(\omega, B) \text{ for almost every } \omega \in \Omega. \tag{1.29}$$

This looks promising. The only problem is that the null set outside of which the inequality holds may depend on B.

2. Now we use the assumption that $\mathcal{G}^0_t$ is countably generated: there exists an increasing sequence of finite σ–algebras $(\mathcal{H}_n)_{n \in \mathbb{N}}$ on Ω, such that $\mathcal{G}^0_t = \bigvee_n \mathcal{H}^n$. Since $\bigcup_n \mathcal{H}^n$ is countable and since $\mathcal{G}^0_t \subseteq \mathcal{G}_t$, we can use (1.29) to obtain a null set $\mathcal{N}$ such that for all $\omega \in \Omega \setminus \mathcal{N}$ and all $B \in \bigcup_n \mathcal{H}_n$ we have

$$\int 1_B(\omega') \frac{\widetilde{Z}_{t+s}}{\widetilde{Z}_t}(\omega') P_{t+s}(\omega, \mathrm{d}\omega') \leq P_t(\omega, B). \tag{1.30}$$

Now $\bigcup_n \mathcal{H}_n$ is stable under finite intersections (it even is an algebra), and therefore the monotone class theorem implies that (1.30) holds for all $B \in \bigvee_n \mathcal{H}_n = \mathcal{G}^0_t$. Since $\widetilde{Z}_{t+s}(\omega')/\widetilde{Z}_t(\omega') > 0$ for every $\omega' \in \Omega$, the proof is complete.

$\square$

Corollary 1.5.6. *Suppose that there exists a one dimensional continuous local martingale M that has the predictable representation property under $(\mathcal{F}_t)$. If under $(\mathcal{G}_t)$, the semimartingale decomposition of M is of the form*

$$M_t = \widetilde{M}_t + \int_0^t \alpha_s \mathrm{d}\langle \widetilde{M} \rangle_s$$

for a $(\mathcal{G}_t)$–local martingale $\widetilde{M}$ and a predictable integrand $\alpha \in L^2_{loc}(\widetilde{M})$, then the generalized Jacod criterion holds.

Proof. In this case the stochastic exponential

$$Z_t := \exp\left(- \int_0^t \alpha_s \mathrm{d}\widetilde{M}_s - \frac{1}{2}\int_0^t \alpha_s^2 \mathrm{d}\langle \widetilde{M} \rangle_s \right)$$

is a universal supermartingale density. □

Corollary 1.5.6 was previously shown by Imkeller, Pontier and Weisz [IPW01] for initial enlargements and under the stronger assumption

$$E\left(\int_0^\infty \alpha_s^2 \mathrm{d}\langle \widetilde{M} \rangle_s \right) < \infty.$$

For simplicity we gave the one dimensional formulation of Corollary 1.5.6. Of course the same argument works in the multidimensional setting: if $M = (M^1, \dots, M^d)$ has the predictable representation property under $(\mathcal{F}_t)$, and if M satisfies the structure condition under $(\mathcal{G}_t)$, then the generalized Jacod criterion is satisfied.

2. Conditioned martingales

In Chapter 1 we studied the (NA1) condition and showed that it is a robust, natural condition to impose on an asset price model. But we did not give many examples of processes satisfying (NA1), apart from the classical examples for which an equivalent local martingale measure exists. Another example that we gave was that of a complete market under an initial filtration enlargement by a countable random variable: in that case (NA1) is satisfied but there exists no equivalent local martingale measure.

A more concrete example of a process satisfying (NA1) but not (NA) is given by the three dimensional Bessel process in finite time. This was first observed by Delbaen and Schachermayer [DS95a]. On a finite time horizon $[0, T]$, the three dimensional Bessel process can be constructed from a Brownian motion W under P started at 1 and stopped at 0: under the measure $\mathrm{d}Q := W_T \mathrm{d}P$, the process W is a Bessel process. Delbaen and Schachermayer [DS95a] noted that the inverse Bessel process $1/W$ is a local martingale deflator for W under Q. More precisely, they showed that $1/W$ defines a dominating measure (namely P), under which W is a local martingale (a Brownian motion). At least in the canonical filtration of W it is then clear that W cannot admit an equivalent local martingale measure, because it follows from the predictable representation property of W that $1/W$ is the only candidate for the density of an equivalent local martingale measure. Since $1/W$ is a strict local martingale (i.e. a local martingale that is not a martingale), there cannot exist an equivalent local martingale measure. This was generalized to arbitrary filtrations by Karatzas and Kardaras [KK07], Example 3.6, who constructed an explicit strategy that realizes an arbitrage.

Therefore, the Bessel process is an interesting object when studying (NA1). It is well known, and was possibly first observed by McKean [McK63], that in infinite time it can be obtained by conditioning a Brownian motion not to hit zero. Below we give a simple probabilistic proof for this result, that extends to continuous nonnegative local martingales.

2.1. Introduction

We study the law Q of a continuous nonnegative P–local martingale M starting in 1, if conditioned never to hit zero. The key step in our analysis is the simple observation that the conditional measure Q, on the corresponding σ–algebra, is given by $M_\tau \mathrm{d}P$, where τ denotes the first hitting time of either 0 or another value $x > 1$. This observation relates the change of measure over an infinite time horizon (through a conditioning argument) to the change of measure in finite time (via the Radon-Nikodym derivative M_τ).

Under the conditional measure Q, the process M diverges to ∞, and $1/M$ is a local martingale. This insight allows us to condition M downwards, which corresponds to conditioning $1/M$ upwards and can therefore be treated with our previously developed arguments. In the case of a diffusion it is possible to write down the dynamics of the upward conditioned process explicitly, defined via its scale function, - and similarly for a downward conditioned diffusion.

For example, if M is a P–Brownian motion stopped in 0, then M is a Q–three dimensional Bessel process. This connection of Brownian motion and Bessel process has been well known, at least since the work of McKean [McK63], building on Doob [Doo57]. Following McKean, several different proofs were given for this result, mostly embedding the statement in a more general result such as the one about path decompositions in Williams [Wil74]. Most of these proofs are analytical and rely strongly on the Markov property of Brownian motion and Bessel process - or even on the fact that the transition densities are known for these processes.

As the study of the law of upward and downward conditioned processes has usually not been the main focus of these papers, results have, to the best of our knowledge, not been proven in the full generality of this paper, and the underlying arguments were often only indirect. Our proof uses only elementary arguments, it is probabilistic, and works for every continuous local martingale. We show that in finite time it is not possible to obtain a Bessel process by conditioning a Brownian motion not to hit zero and we point out that conditioning a Brownian motion upward and conditioning a Bessel process downward can be understood using the same result.

In Subsection 2.2.1 we treat the case of upward conditioning of local martingales and in Subsection 2.2.2 the case of downward conditioning. In Section 2.3 we study the implications of these results for diffusions. In Appendix C we illustrate that conditioning on a null set (such as the Brownian motion never hitting zero) is highly sensitive with respect to the approximating sequence of sets.

Relevant literature

The connection of Brownian motion and the three dimensional Bessel process has been studied in several important and celebrated papers. Most of these studies have focused on more general statements than this connection only. To provide a complete list of references is beyond this note. In the following paragraphs, we try to give an overview of some of the most relevant and influential work in this area.

For a Markov process X, Doob [Doo57] studies its h–transform, where h denotes an excessive function such that, in particular, $h(X)$ is a supermartingale. Using $h(X)/h(X_0)$ as a Radon-Nikodym density, a new (sub-probability) measure is constructed. Doob shows, among many other results, that, if h is harmonic (and additionally "minimal," as defined therein), the process X converges under the new measure to the points on the extended real line where h takes the value infinity. In this sense, changing the measure corresponds to conditioning the process to the event that X converges to these points. For example, if X is Brownian motion started in 1, then $h(x) = x$ is harmonic and leads to a probability measure, under which X, now distributed as a Bessel process, tends to infinity. Our results also yield this observation; furthermore they contain the case of non-Markovian processes X that are nonnegative local martingales only.

An analytic proof of the fact that upward conditioned Brownian motion is a three dimensional Bessel process is given in McKean's work [McK63] on Brownian excursions. He shows that if W is a Brownian motion started in 1, if $B \in \mathcal{F}_s$, where $\mathcal{F}_s$ is the σ–algebra generated by W up to time s for some $s > 0$, and if τ_0 is the hitting time of 0, then $P(W \in B|\tau_0 > t) \to P(X \in B)$ as $t \to \infty$, where X is a three dimensional Bessel process. The proof is based on techniques from partial differential equations. In that article, also a path decomposition is given for excursions of Brownian motion in terms of two Bessel processes, one run forward in time, and the other one run backward. McKean already generalizes all these results to regular diffusions.

Knight [Kni69] computes the dynamics of Brownian motion conditioned to stay either in the interval $[-a, a]$ or $(-\infty, a]$ for some $a > 0$, and thus also derives the Bessel dynamics. To obtain these results, Knight uses a very astute argument based on inverting Brownian local time. He moreover illustrates the complications arising from conditioning on null sets by providing an insightful example; we shall present two other examples based on direct arguments, without the necessity of any computations, in Appendix C to illustrate this point further.

In his seminal paper on path decompositions, Williams [Wil74] shows that Brownian motion conditioned not to hit zero corresponds to the Bessel process. His results extend to diffusions and reach far beyond this observation. For example, he shows that “stitching” a Brownian motion up to a certain stopping time and a three dimensional Bessel process together yields another Bessel process. In Pitman and Yor [PY81] this approach is generalized to killed diffusions. A diffusion process is killed with constant rate and conditioned to hit infinity before the killing time. This allows the interpretation of a two-parameter Bessel process as an upward conditioned one-parameter Bessel process.

Pitman [Pit75] proves essentially our Lemma 2.2.1 below in the Brownian case. This is achieved by approximating the continuous processes by random walks, whose paths can be counted. For the continuous case, the statement then follows by a weak convergence argument. The main result of that article is Pitman's famous theorem that $2W^* - W$ is a Bessel process if W is a Brownian motion and W^* its running maximum.

Baudoin [Bau02] takes a different approach. Given a Brownian motion, a functional Y of its path and a distribution μ, Baudoin constructs a probability measure under which Y is distributed as μ. The recent monograph by Roynette and Yor [RY09] studies penalizations of Brownian paths, which can be understood as a generalization of conditioned Brownian motion. Under the penalized measure, the coordinate process can have radically different behavior than under the Wiener measure. In our example it does not hit zero. In Roynette and Yor [RY09] there is an example of a penalized measure under which the supremum process stays almost surely bounded.

2.2. General case: continuous local martingales

Let $\Omega = C_{\text{abs}} := C_{\text{abs}}(\mathbb{R}_+, [0, \infty])$ be the space of $[0, \infty]$–valued functions ω that are absorbed in 0 and ∞, and that are continuous on $[0, \tau_\infty(\omega))$, where $\tau_\infty(\omega)$ denotes the first hitting time of ∞ by ω, to be specified below. The reason for considering this space is that it allows for the application of Parthasarathy's extension theorem, which we will need below. Let M be the coordinate process, that is, $M_t(\omega) = \omega(t)$. Define, for the sake of notational simplicity, $M_\infty := \sqrt{\limsup_{t\to\infty} M_t \liminf_{t\to\infty} M_t}$ (with $\infty \cdot 0 := 1$).[1] Denote the canonical filtration by $(\mathcal{F}_t)_{t\geq 0}$ with $\mathcal{F}_t := \sigma(M_s : s \leq t)$, and write $\mathcal{F} := \bigvee_{t\geq 0} \mathcal{F}_t$. For all $a \in [0,\infty]$, define τ_a as the first hitting time of a, to wit,

$$\tau_a := \inf\{t \in [0,\infty] : M_t = a\} \tag{2.1}$$

with $\inf \emptyset := \mathfrak{T}$, representing a time "beyond infinity." The introduction of $\mathfrak{T}$ allows for a unified approach to treat examples like geometric Brownian motion. We shall extend the natural ordering to $[0,\infty] \cup \{\mathfrak{T}\}$ by $t < \mathfrak{T}$ for all $t \in [0,\infty]$. For all stopping times τ, define the σ–algebras $\mathcal{F}_\tau$ as

$$\begin{aligned}\mathcal{F}_\tau :=& \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t \quad \forall t \in [0,\infty)\}\\ =& \sigma(M_s^\tau : s < \infty) = \sigma(M_s^{\tau\wedge\tau_0} : s < \infty),\end{aligned}$$

where $M^\tau \equiv M^{\tau\wedge\tau_0}$ is the process M stopped at the stopping time τ. For the equality between the σ–algebras see Stroock and Varadhan [SV06], Lemma 1.3.3. Let P be a probability measure on $(\Omega, \mathcal{F})$, such that M is a nonnegative local martingale with $P(M_0 = 1) = 1$.

2.2.1. Upward conditioning

In this section, we study the law of the local martingale M conditioned never to hit zero. This event can be expressed as

$$\{\tau_0 = \mathfrak{T}\} = \bigcap_{a\in[0,\infty)} \{\tau_a \leq \tau_0\} \supset \bigcup_{a\in(0,\infty]} \{\tau_a \wedge \tau_0 = \mathfrak{T}\}. \tag{2.2}$$

The core of this chapter is the following simple observation:

Lemma 2.2.1 (Upward conditioning). *If $P(\tau_a \wedge \tau_0 < \mathfrak{T}) = 1$ for some $a \in (1,\infty)$, we have that*

$$\mathrm{d}P(\cdot|\tau_a \leq \tau_0) = M_{\tau_a}\mathrm{d}P.$$

Proof. Note that M^{τ_a} is bounded and thus a uniformly integrable martingale. In particular,

$$1 = E_P(M_\infty^{\tau_a}) = aP(\tau_a \leq \tau_0) + 0,$$

[1] Note that this definition differs from the convention in the remainder of this thesis. The definition of M_∞ is not further relevant as M converges (or diverges to infinity) almost surely under all measures that we shall consider. We chose this definition of M_∞ since it commutes with taking the reciprocal $1/M_\infty$.

which implies that, for all $A \in \mathcal{F}$,

$$P(A|\tau_a \le \tau_0) = \frac{P(A \cap \{\tau_a \le \tau_0\})}{P(\tau_a \le \tau_0)} = \frac{P(A \cap \{\tau_a \le \tau_0\})}{\frac{1}{a}} = E_P\left(M_\infty^{\tau_a} 1_A\right),$$

yielding the statement. $\square$

Three different probability measures

Consider three possible probability measures:

1. The local martingale M introduces an *h–transform* Q of P. This is the unique probability measure Q on $(\Omega, \mathcal{F})$ that satisfies $\mathrm{d}Q|_{\mathcal{F}_\tau} = M_\tau \mathrm{d}P|_{\mathcal{F}_\tau}$ for all stopping times τ for which M^τ is a uniformly integrable martingale. The probability measure Q is called the *Föllmer measure* of M, see Föllmer [Föl72] and Meyer [Mey72].[2] Note that the construction of this measure does not require the density process M to be the canonical process on Ω - the extension only relies on the topological structure of $\Omega = C_{\text{abs}}$. This will be important later, when we consider diffusions. We remark that, in the case of M being a P–martingale, we could also use a standard extension theorem, such as Theorem 1.3.5 in Stroock and Varadhan [SV06].
2. If $P(\tau_0 = \mathfrak{T}) = 0$, Lemma 2.2.1 in conjunction with (2.2) directly yields the consistency of the family of probability measures $\{P(\cdot|\tau_a \le \tau_0)\}_{a>1}$ on the filtration $(\mathcal{F}_{\tau_a})_{a>1}$. By Föllmer's construction again, there exists a unique probability measure $\widetilde{Q}$ on $(\Omega, \mathcal{F})$, such that $\widetilde{Q}|_{\mathcal{F}_{\tau_a}} = P(\cdot|\tau_a \le \tau_0)|_{\mathcal{F}_{\tau_a}}$.
3. If $P(\tau_0 = \mathfrak{T}) > 0$, we can define the probability measure $\widehat{Q}(\cdot) := P(\cdot|\tau_0 = \mathfrak{T})$ via the Radon-Nikodym derivative $1_{\{\tau_0 = \mathfrak{T}\}}/P(\tau_0 = \mathfrak{T})$.

Since in the case $P(\tau_0 = \mathfrak{T}) = 0$, we have $\{\tau_a \le \tau_0\} =_{P-a.s.} \{\tau_a < \tau_0\}$ for all $a \in (0, \infty]$, the measure $\widetilde{Q}$ is also called *upward conditioned* measure since it is constructed by iteratively conditioning the process M to hit any level a before hitting 0.

Relationship of probability measures

We are now ready to relate the three probability measures constructed above:

Theorem 2.2.2 (Identity of measures)**.** *Set $p := P(\tau_0 = \mathfrak{T}) = P(M_\infty > 0)$. If $p = 0$, then $Q = \widetilde{Q}$. If $p > 0$, then $Q = \widehat{Q}$ if and only if M is a uniformly integrable martingale with $P(M_\infty \in \{0, 1/p\}) = 1$.*

Proof. First, consider the case $p = 0$. Both Q and $\widetilde{Q}$ satisfy, for all $a > 1$,

$$\mathrm{d}\widetilde{Q}|_{\mathcal{F}_{\tau_a}} = M_{\tau_a} \mathrm{d}P|_{\mathcal{F}_{\tau_a}} = \mathrm{d}Q|_{\mathcal{F}_{\tau_a}}.$$

[2] See also Delbaen and Schachermayer [DS95a] for a discussion of this measure, Pal and Protter [PP10] for the extension to infinite time horizons and Carr, Fisher, and Ruf [CFR12] for allowing nonnegative local martingales.

Thus, Q and $\widetilde{Q}$ agree on $\bigvee_{a>1} \mathcal{F}_{\tau_a} = \bigvee_{a>1} \sigma(M_t^{\tau_a} : t \geq 0) = \mathcal{F}$.

Next, consider the case $p > 0$. Then, $Q = \widehat{Q}$ and $\mathrm{d}\widehat{Q}/\mathrm{d}P|_{\mathcal{F}_t} \leq 1/p$ imply that $M_t \leq 1/p$, yielding that M is a uniformly integrable martingale with $M_\infty = \mathrm{d}Q/\mathrm{d}P \in \{0, 1/p\}$. For the reverse direction, observe that $M_\infty = 1_{\{\tau_0 = \mathfrak{T}\}}/p$. This observation together with its uniform integrability completes the proof. □

This theorem implies, in particular, that in finite time the three dimensional Bessel process cannot be obtained by conditioning a Brownian motion not to hit zero. However, over finite time horizons, a Bessel-process can be constructed via the h–transform $M_T \mathrm{d}P$, when M is P–Brownian motion started in 1 and stopped in 0. Over infinite time horizons, one has two choices; the first one is using an extension theorem for the h–transforms, the second one is conditioning M not to hit 0 by approximating this null set by the sequence of events that M hits any $a > 0$ before it hits 0.

Remark 2.2.3 (Conditioning on null sets). We remark that the interpretation of the measure $\widetilde{Q}$ as P conditioned on a null set requires specifying an approximating sequence of that null set. In Appendix C we illustrate this subtle but important point.

Remark 2.2.4 (The trans-infinite time $\mathfrak{T}$). The introduction of $\mathfrak{T}$ in this subsection allows us to introduce the upward-conditioned measure $\widetilde{Q}$ and to show its equivalence to the h–transform Q if M converges to zero but not necessarily hits zero in finite time, such as P–geometric Brownian motion. If one is only interested in processes as, say, stopped Brownian motion, then one could formulate all results in this subsection in the standard way when $\inf \emptyset := \infty$ in (2.1). One would then need to exchange $\mathfrak{T}$ by ∞ throughout this subsection; in particular, one would have to assume in Lemma 2.2.1 that $P(\tau_a \wedge \tau_0 < \infty) = 1$ and replace the condition $P(\tau_0 = \mathfrak{T}) = 0$ by $P(\tau_0 = \infty) = 0$ for the construction of the upward-conditioned measure $\widetilde{Q}$.

2.2.2. Downward conditioning

In this subsection, we consider the converse case of conditioning M downward instead of upward. Towards this end, we first provide a well-known result; see for example [CFR12]. For the sake of completeness, we provide a proof.

Lemma 2.2.5 (Local martingale property of $1/M$). *Under the h–transformed measure Q, the process $1/M$ is a nonnegative local martingale and $Q(\tau_\infty = \mathfrak{T}) = E_P(M_\infty)$.*

Proof. Observe that for $s, t \geq 0$ and $A \in \mathcal{F}_t$ we have

$$\begin{aligned} E_Q\left(1_A \frac{1}{M_{t+s}^{\tau_{1/n}}}\right) &= \lim_{m\to\infty} E_Q\left(1_{A\cap\{\tau_m > t\}} \frac{1}{M_{t+s}^{\tau_{1/n}\wedge\tau_m}}\right) + E_Q\left(1_{A\cap\{\tau_\infty \leq t\}} \frac{1}{M_{t+s}^{\tau_{1/n}}}\right) \\ &= \lim_{m\to\infty} E_P\left(1_{A\cap\{\tau_m > t\}} \frac{1}{M_{t+s}^{\tau_{1/n}\wedge\tau_m}} M_{t+s}^{\tau_m}\right) + E_Q\left(1_{A\cap\{\tau_\infty \leq t\}} \frac{1}{M_t^{\tau_{1/n}}}\right) \end{aligned} \tag{2.3}$$

Now we consider the two events $\{\tau_{1/n} \leq t\}$ and $\{\tau_{1/n} > t\}$ separately and used the P–martingale property of M^{τ_m} after conditioning on $\mathcal{F}_t$ and $\mathcal{F}_{\tau_{1/n}}$, respectively (note that $A \cap \{\tau_m > t\} \cap \{\tau_{1/n} > t\} \in \mathcal{F}_{\tau_{1/n}}$), to obtain

$$E_P\left(1_{A\cap\{\tau_m>t\}} \frac{1}{M_{t+s}^{\tau_{1/n}\wedge\tau_m}} M_{t+s}^{\tau_m}\right) = E_P\left(1_{A\cap\{\tau_m>t\}} \frac{1}{M_t^{\tau_{1/n}\wedge\tau_m}} M_t^{\tau_m}\right).$$

Plugging this back into (2.3), we have

$$\begin{aligned}
E_Q\left(1_A \frac{1}{M_{t+s}^{\tau_{1/n}}}\right) &= \lim_{m\to\infty} E_P\left(1_{A\cap\{\tau_m>t\}} \frac{1}{M_t^{\tau_{1/n}\wedge\tau_m}} M_t^{\tau_m}\right) + E_Q\left(1_{A\cap\{\tau_\infty\leq t\}} \frac{1}{M_t^{\tau_{1/n}}}\right) \\
&= \lim_{m\to\infty} E_Q\left(1_{A\cap\{\tau_m>t\}} \frac{1}{M_t^{\tau_{1/n}\wedge\tau_m}}\right) + E_Q\left(1_{A\cap\{\tau_\infty\leq t\}} \frac{1}{M_t^{\tau_{1/n}}}\right) \\
&= E_Q\left(1_A \frac{1}{M_t^{\tau_{1/n}}}\right).
\end{aligned}$$

The local martingale property of $1/M$ then follows from

$$\begin{aligned}
Q\left(\lim_{n\to\infty} \tau_{1/n} < \infty\right) &= \lim_{m\to\infty} Q\left(\lim_{n\to\infty} \tau_{1/n} < \tau_m \wedge \infty\right) \\
&= \lim_{m\to\infty} E_P\left(1_{\{\lim_{n\to\infty}\tau_{1/n}<\tau_m\}} M_\infty^{\tau_m}\right) = 0.
\end{aligned}$$

Therefore, $1/M$ converges Q–almost surely to some random variable $1/M_\infty$. We observe that

$$\begin{aligned}
Q(\tau_\infty = \mathfrak{T}) &= 1 - \lim_{m\to\infty} Q(\tau_m < \infty) = 1 - \lim_{m\to\infty} E_P(1_{\{\tau_m<\infty\}} M_\infty^{\tau_m}) \\
&= \lim_{m\to\infty} E_P(1_{\{\tau_m\geq\infty\}} M_\infty) = E_P(M_\infty),
\end{aligned}$$

where we use that M converges P–almost surely, since it is a nonnegative supermartingale. $\square$

The last lemma directly implies the following observation.

Corollary 2.2.6 (Mutual singularity)**.** *We have $P(M_\infty = 0) = 1$ if and only if $Q(M_\infty = \infty) = 1$.*

This observation is consistent with our understanding that either condition implies that the two measures are supported on two disjoint sets. Corollary 2.2.6 is also consistent with Theorem 2.2.2, which yields that $P(M_\infty = 0) = 1$ implies the identity $Q = \widetilde{Q}$, where $\widetilde{Q}$ denotes the upward conditioned measure.

Lemma 2.2.5 indicates that we can condition M *downward* under Q, corresponding to conditioning $1/M$ *upward.* The proof of the next result is exactly along the lines of the arguments in Subsection 2.2.1; however, now with the Q–local martingale $1/M$ taking the place of the P–local martingale M.

Theorem 2.2.7 (Downward conditioning). *If p of Theorem 2.2.2 satisfies $p = 0$, then*

$$\mathrm{d}Q(\cdot|\tau_{1/a} \leq \tau_\infty) = \frac{1}{M_{\tau_{1/a}}}\mathrm{d}Q$$

for all $a > 1$. In particular, there exists a unique probability measure $\widetilde{P}$, such that $\widetilde{P}|_{\mathcal{F}_{\tau_{1/a}}} = Q(\cdot|\tau_{1/a} < \mathfrak{T})$; in fact, $\widetilde{P} = P$.

2.3. Diffusions

In this section, we apply Theorems 2.2.2 and 2.2.7 to diffusions.

2.3.1. Definition and h-transform for diffusions

We call *diffusion* any time-homogeneous strong Markov process $X : C_{\text{abs}} \times [0, \infty) \to [\ell, r]$ with continuous paths in a possibly infinite interval $[\ell, r]$ with $-\infty \leq \ell < r \leq \infty$. Note that we explicitly allow X to take the values ℓ and r; we stop X once it hits the boundary of $[\ell, r]$. We define τ_a for all $a \in [\ell, r]$ as in (2.1) with M replaced by X. We denote the probability measure under which $X_0 = x \in [\ell, r]$ by P_x.

Since X is Markovian it has an *infinitesimal generator* (see page 161 in Ethier and Kurtz [EK86]). As we do not assume any regularity of the semigroup of X, we find it convenient to work with the following *extended infinitesimal generator*: A continuous function $f : [\ell, r] \to \mathbb{R} \cup \{-\infty, \infty\}$, such that f restricted to $\mathbb{R}$ only takes finite values, is in the *domain* of the extended infinitesimal generator $\mathcal{L}$ of X if there exists a continuous function $g : [\ell, r] \to \mathbb{R} \cup \{-\infty, \infty\}$, such that g restricted to $\mathbb{R}$ only takes finite values, and an increasing sequence of stopping times (ρ_n), such that $P_x(\lim_{n\to\infty} \rho_n \geq \tau_\ell \wedge \tau_r) = 1$ and

$$f(X_\cdot^{\rho_n}) - f(x) - \int_0^{\cdot\wedge\rho_n} g(X_s)\mathrm{d}s$$

is a P_x–martingale for all $x \in (\ell, r)$. In that case we write $f \in \mathrm{dom}(\mathcal{L})$ and $\mathcal{L}f = g$.

Throughout this section we shall work with a *regular* diffusion X; that is, for all $x, z \in (\ell, r)$ we have that $P_x(\tau_z < \infty) > 0$. In that case there always exists a continuous, strictly increasing function $s : (\ell, r) \to \mathbb{R} \cup \{-\infty, \infty\}$, uniquely determined up to an affine transformation, such that $s(X)$ is a local martingale (see Propositions VII.3.2 and VII.3.5 in Revuz and Yor [RY99]). We call every such s a *scale function* for X, and we extend its domain to $[\ell, r]$ by taking limits. The next result summarizes Proposition VII.3.2 in [RY99] and describes the relationship of the scale function s and the limiting behavior of X:

Lemma 2.3.1 (Scale function). *We have*

1. *$P_x(\tau_\ell = \mathfrak{T}) = 0$ for one (and then for all) $x \in (\ell, r)$ if and only if $s(\ell) \in \mathbb{R}$ and $s(r) = \infty$;*

2. *$P_x(\tau_r = \mathfrak{T}) = 0$ for one (and then for all) $x \in (\ell, r)$ if and only if $s(\ell) = -\infty$ and $s(r) \in \mathbb{R}$;*

3. *$P_x(\tau_\ell \wedge \tau_r = \mathfrak{T}) = 0$ and $P_x(\tau_\ell < \mathfrak{T}) \in (0, 1)$ for one (and then for all) $x \in (\ell, r)$ if and only if $s(\ell) \in \mathbb{R}$ and $s(r) \in \mathbb{R}$.*

Throughout this section, we shall work with the standing assumption that the scale function s satisfies $s(\ell) > -\infty$ (Assumption L) or $s(r) < \infty$ (Assumption R). Without loss of generality, we shall assume that then $s(\ell) = 0$ or $s(r) = 0$, respectively, and that $\mathcal{F} = \mathcal{F}_{\tau_\ell \wedge \tau_r}$.

Since by assumption $s(X)$ is a local martingale, it defines, under each P_x, a Föllmer measure Q_x as in Section 2.2, where we would set $M := s(X)/s(x)$, for all $x \in [\ell, r]$ (with $0/0 := \infty/\infty := 1$, again contrary to the convention in the remainder of this thesis). The following proposition illustrates how the extended infinitesimal generators of X under P_x and Q_x are related:

Proposition 2.3.2 (h–transform for diffusions)**.** *The process X is a regular diffusion under the probability measures $\{Q_x\}_{x\in[\ell,r]}$. Its extended infinitesimal generator $\mathcal{L}^s$ under $\{Q_x\}_{x\in[\ell,r]}$ is given by* $\mathrm{dom}(\mathcal{L}^s) = \{\varphi : s\varphi \in \mathrm{dom}(\mathcal{L})\}$ *and*

$$\mathcal{L}^s\varphi(x) = \frac{1}{s(x)}\mathcal{L}[s\varphi](x).$$

Proof. We only discuss the case $s(\ell) = 0$ since the case $s(r) = 0$ is treated in the same way. In order to show the Markov property of X under Q_x, we need to prove that

$$E_{Q_x}(f(X_{\rho+t})|\mathcal{F}_\rho) = E_{Q_x}(f(X_{\rho+t})|X_\rho)$$

for all $t \geq 0$, for all bounded and continuous functions $f : [\ell, r] \to \mathbb{R}$, and for all finite stopping times ρ. On the event $\{\rho \geq \tau_r\}$, the equality holds trivially as X gets absorbed in ℓ and r. On the event $\{\rho < \tau_r\}$, observe that

$$\begin{aligned}
E_{Q_x}(f(X_{\rho+t})|\mathcal{F}_\rho) &= \lim_{a\to r-} E_{Q_x}(f(X^{\tau_a}_{\rho+t})|\mathcal{F}_\rho) \\
&= \lim_{a\to r-} E_{Q_x}(f(X^{\tau_a}_{\rho+t})|X^{\tau_a}_\rho) = E_{Q_x}(f(X_{\rho+t})|X_\rho),
\end{aligned}$$

where the second equality follows from the generalized Bayes' formula in Proposition C.2 in [CFR12] and the Markov property of X^{τ_a} under P_x. Therefore, X is strongly Markovian under Q_x. Since X is also time-homogeneous under any of the measures Q_x, we have shown that X is a diffusion under $\{Q_x\}_{x\in[\ell,r]}$.

As for the regularity, fix $a \in (\ell, x)$ and $b \in (x, r)$. Observe that Q_x is equivalent to P_x on $\mathcal{F}_{\tau_a\wedge\tau_b}$. This fact in conjunction with the regularity of X under P and Proposition VII.3.2 in [RY99] yields that $Q_x(\tau_a < \infty) > 0$ as well as $Q_x(\tau_b < \infty) > 0$.

Denote now the extended infinitesimal generator of X under $\{Q_x\}_{x\in[\ell,r]}$ by $\mathcal{G}$, let $\varphi \in \mathrm{dom}(\mathcal{G})$ with localizing sequence $(\rho_n)_{n\in\mathbb{N}}$, and fix $x \in (\ell, r)$. Fix two sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ with $a_n \to \ell+$ and $b_n \to r-$ as $n \to \infty$. We may assume, without loss of generality, that $\rho_n \le \tau_{a_n} \wedge \tau_{b_n}$. By definition of the extended infinitesimal generator,

$$\varphi(X_\cdot^{\rho_n}) - \varphi(x) - \int_0^{\cdot\wedge\rho_n} \mathcal{G}\varphi(X_s)\mathrm{d}s$$

is a Q_x–martingale. Since $\varphi(\cdot)$ and $\mathcal{G}\varphi(\cdot)$ are bounded on $[a_n, b_n]$ this fact, in conjunction with Fubini's theorem, yields that

$$\frac{1}{s(x)}\left(\varphi(X_\cdot^{\rho_n})s(X_\cdot^{\rho_n}) - \varphi(x)s(x) - \int_0^{\cdot\wedge\rho_n} \mathcal{G}\varphi(X_u^{\rho_n})s(X_u^{\rho_n})\mathrm{d}u\right)$$

is a P_x–martingale. Since $(\rho_n)_{n\in\mathbb{N}}$ converges P_x–almost surely to $\tau_\ell \wedge \tau_r$ for all $x \in (\ell, r)$ this implies that $\varphi s \in \mathrm{dom}(\mathcal{L})$ and $\mathcal{L}[s\varphi](x) = \mathcal{G}\varphi(x)s(x)$. The other inclusion can be shown in the same manner, which completes the proof. $\square$

The following observation is a direct consequence of Lemma 2.2.5 and the fact that X is a regular diffusion under the probability measures $\{Q_x\}_{x\in[\ell,r]}$:

Lemma 2.3.3 (Scale function for h–transform)**.** *Under $\{Q_x\}_{x\in[\ell,r]}$, the function $\widetilde{s}(\cdot) = -1/s(\cdot)$ is, with the appropriate definition of* $1/0$, *a scale function for X with $\widetilde{s}(\ell) = -\infty$, $\widetilde{s}(r) \in \mathbb{R}$ under Assumption L and with $\widetilde{s}(r) = \infty$, $\widetilde{s}(\ell) \in \mathbb{R}$ under Assumption R.*

2.3.2. Conditioned diffusions

We now are ready to formulate and prove a version of the statements of Section 2.2 for diffusions.

Corollary 2.3.4 (Conditioning of diffusions)**.** *Fix $x \in (\ell, r)$ and make Assumption L.*

1. *Suppose that $P_x(\tau_\ell = \mathfrak{T}) = 0$, which is equivalent to $s(r) = \infty$. Then the family of probability measures $\{P_x(\cdot|\tau_a \le \tau_\ell)|_{\mathcal{F}_{\tau_a}}\}_{x<a<r}$ is consistent and thus has an extension $\widetilde{Q}_x$ on $\mathcal{F}$. Moreover, the extension satisfies $\widetilde{Q}_x = Q_x$.*

2. *Suppose that $P_x(\tau_\ell = \mathfrak{T}) > 0$, which is equivalent to $s(r) < \infty$, and define $\widehat{Q}_x = P_x(\cdot|\tau_\ell = \mathfrak{T})$. Then $\widehat{Q}_x$ satisfies $\widehat{Q}_x = Q_x$.*

Furthermore, provided that $s(r) = \infty$, the family $\{Q_x(\cdot|\tau_a \le \tau_r)|_{\mathcal{F}_{\tau_a}}\}_{\ell<a<x}$ of probability measures is consistent. Its unique extension is P_x.

Under Assumption R, all statements still hold with r replaced with ℓ and, implicitly, $x < a < r$ replaced with $\ell < a < x$.

Proof. We only consider the case of Assumption L, as Assumption R requires the same arguments. We write $M = s(X)/s(x)$. The hitting times σ_a of M are defined as in (2.1). Since s is strictly increasing, we have that, for all $x < a < r$,

$$\{\tau_a \le \tau_\ell\} = \{\sigma_{s(a)/s(x)} \le \sigma_0\}.$$

Since M is a nonnegative local martingale with $P_x(M_0 = 1) = 1$, the statements in 1. and 2. follow immediately from Theorem 2.2.2 and Lemma 2.3.1, which shows that $s(X)_\infty$ takes exactly two values. The remaining assertions follow from Lemma 2.3.3 and Theorem 2.2.7. □

It is clear that the measure Q under Assumption L corresponds to the upward conditioned diffusion X, while under Assumption R it corresponds to the downward conditioned diffusion.

After finishing this work, we learned about Kardaras [Kar10b]. Therein, by similar techniques it is shown that X under Q tends to infinity if $s(r) = \infty$; see Section 6.2 in [Kar10b]. In Section 5 therein, a similar probability measure is constructed for a Lévy process X that drifts to $-\infty$. After a change of measure of the form $s(X)$ for a harmonic function s, the process X under the new measure drifts to infinity.

2.3.3. Explicit generators

Here we formally derive the dynamics of upward conditioned and downward conditioned diffusions. For this purpose suppose that X is a diffusion with extended infinitesimal generator $\mathcal{L}$, such that $\mathrm{dom}(\mathcal{L}) \supseteq C^2$, where C^2 denotes the space of twice continuously differentiable functions on (ℓ, r), and

$$\mathcal{L}\varphi(x) = b(x)\varphi'(x) + \frac{1}{2}a(x)\varphi''(x), \quad \varphi \in C^2$$

for some locally bounded, measurable functions b and a such that $a(x) > 0$ for all $x \in (\ell, r)$.

Finding the scale function then at least formally corresponds to solving the linear ordinary differential equation

$$b(x)s'(x) + \frac{1}{2}a(x)s''(x) = 0. \tag{2.4}$$

This is for example done in Section 5.5.B of Karatzas and Shreve [KS88]. From now on, we continue under either Assumption L or Assumption R, with s being either nonnegative or nonpositive. We plug s into the definition of $\mathcal{L}^s$. Towards this end, let $\varphi \in C^2$. Then we have

$$\begin{aligned}
\mathcal{L}^s\varphi(x) &= \frac{1}{s(x)}\mathcal{L}(s\varphi)(x) = \frac{1}{s(x)}\left(b(x)(s\varphi)'(x) + \frac{1}{2}a(x)(s\varphi)''(x)\right) \\
&= \frac{1}{s(x)}\Big(b(x)(s'(x)\varphi(x) + s(x)\varphi'(x)) \\
&\qquad\qquad + \frac{1}{2}a(x)(s''(x)\varphi(x) + 2s'(x)\varphi'(x) + s(x)\varphi''(x))\Big) \\
&= \left(b(x) + \frac{a(x)s'(x)}{s(x)}\right)\varphi'(x) + \frac{1}{2}a(x)\varphi''(x),
\end{aligned}$$

since $s'' = -2(b/a)s'$ due to (2.4). Therefore, the upward or downward conditioned process has an additional drift of $(as')/s$. This drift is always positive (or always negative), as is to be expected.

Now, under Assumption L (upward conditioning) with $\ell = 0$, if $b = 0$, then $s(x) = x$; therefore the additional drift of the upward conditioned diffusion is $a(x)/x$. Under Assumption R (downward conditioning) with $\ell = 0$ and $r = \infty$, if $b(x) = a(x)/x$, then (2.4) yields $s(x) = -\frac{1}{x}$ and thus an additional drift of $-a(x)/x = -b(x)$. These observations lead to the following well-known fact.

Corollary 2.3.5 ((Geometric) Brownian motion)**.** *A Brownian motion conditioned on hitting* ∞ *before hitting* 0 *is a three dimensional Bessel process. Vice versa, a three dimensional Bessel process conditioned to hit 0 is a Brownian motion. Moreover, a geometric Brownian motion conditioned on hitting* ∞ *before hitting 0 is a geometric Brownian motion with unit drift.*

3. Pathwise integration in model free finance

Here we use Vovk's [Vov12] pathwise, hedging based approach to finance to describe "typical price paths". Roughly speaking, a property (P) holds for typical price paths if it is possible to make an arbitrarily large profit by investing in those paths where (P) is violated, without ever risking to lose much. This can be interpreted as a model free version of the (NA1) property. Just as we can dismiss stochastic models that violate (NA1) because they always lead to infinite utility, we can dismiss sets of paths that allow an investor to make too much profit. We show that for typical price paths it is possible to define a pathwise Itô type integral. We also indicate that typical price paths can be used as integrators in Lyons's theory of rough paths.

3.1. Motivation

We saw in Chapter 1 that (NA1) (see Definition 1.1.1) is a natural condition to impose on an asset price model, because it is equivalent to the existence of a non-degenerate utility maximization problem (see Proposition 1.2.2). We also saw that (NA1) is rather robust under changes in the information structure (see Section 1.5). It is also preserved when switching to an equivalent probability measure. However, (NA1) may be violated after passing to an absolutely continuous measure, see Example 1.4.14. Another example is $S_0 = 1$ and $P(S_1 = 0) = P(S_1 = 2) = 1/2$: the process S is a martingale, and $\mathrm{d}Q = S_1 \mathrm{d}P$ is absolutely continuous with respect to P. But S violates both (NA) and (NA1) under Q. Since (NA1) may already fail when passing to an absolutely continuous probability measure, there is no hope to show that (NA1) is preserved when passing to a singular measure Q.

Of course in practice the probability measure P that describes the statistical behavior of the asset price process is usually not known with absolute certainty. Therefore, in recent years there has been a lot of interest in mathematical finance under model uncertainty, where one has to argue simultaneously for uncountably many mutually singular probability measures, and in model free finance, where one does not assume any statistical knowledge about the asset price process. A model free formulation of the (NA1) property, that we will work with below, was given by Vovk [Vov12].

Maybe the simplest example of model uncertainty is given by the Black-Scholes model under volatility uncertainty. Here it is assumed that the (discounted) price process of a given asset is described by a geometric Brownian motion with drift,

$$\mathrm{d}S_t^\sigma = S_t^\sigma(\sigma \mathrm{d}W_t + b\mathrm{d}t),$$

where W is a one dimensional standard Brownian motion, and where $\sigma > 0$ and $b \in \mathbb{R}$. In contrast to the classical theory, here it is not assumed that the volatility σ is known. Rather it is assumed that σ lies in some interval $[a, c]$ for $0 < a < c$. Note that if P^σ is a probability measure on $C([0,T],\mathbb{R})$ for which the coordinate process has the distribution of S^σ, then P^{σ_1} and P^{σ_2} are mutually singular for $\sigma_1 \neq \sigma_2$. The reason for only keeping track of σ lies in the fact that the prices for European options on S^σ are independent of b, because the martingale measure for S^σ does not depend on b.

One of the basic problems in mathematical finance is to calculate “fair” prices for financial derivatives of the underlying asset price process S. The minimal superhedging price of a derivative (i.e. a random variable) F is defined as

$$p(F) := \inf \left\{ \lambda \in \mathbb{R} : \exists\, H \in \mathcal{H}_\lambda : \lambda + \int_0^T H_s \mathrm{d}S_s \geq F \text{ a.s.} \right\}, \tag{3.1}$$

where we denote by $\mathcal{H}_\lambda$ the λ–admissible strategies, i.e. all H for which the stochastic integral $H \cdot S$ exists and satisfies $(H \cdot S)_t \geq -\lambda$ for all $t \in [0,T]$. It can be shown that under suitable conditions

$$p(F) = \sup\{E_Q(F) : Q \text{ is an equivalent local martingale measure for } S\}.$$

In order to obtain a similar result under volatility uncertainty, we first have to define superhedging prices in this context. It would be natural to replace the “a.s.” assumption in (3.1) by “a.s under every P^σ, $\sigma \in [a,c]$”. But then the stochastic integral $(H \cdot S)_T$ has to be constructed simultaneously under all the measures P^σ. First results in this direction have been obtained by [ALP95] and [Lyo95]. In recent years, such problems have been tackled with the help of “quasi-sure analysis”, see for example [DM06].

The mutual singularity of the measure P^σ for different values of σ requires new techniques to handle the stochastic integrals against general (not necessarily simple) integrands. Such integrals are needed to develop a sufficiently strong theory of mathematical finance under model uncertainty. But in the model uncertainty context we can essentially still rely on Itô’s integration techniques, because while we have to deal with many probability measures at once, the price process is a semimartingale under every given measure. Model free mathematical finance no longer assumes any model structure. Instead it is assumed that some basic facts about the financial market are known (for example some European call and put prices), and the aim is to calculate all prices for a given derivative that are compatible with these known facts.

In [BHLP11] it is assumed that $S = (S_t)_{t=0,\dots,T}$ is a discrete time process, and that the prices for all European call options with payoff $(S_t - K)^+$ for $0 \leq t \leq T$ and $K \in \mathbb{R}$ are known. This determines the marginal distribution μ_t of S_t for $0 \leq t \leq T$ under every compatible pricing measure, but not the joint distribution $(S_0, S_1, \dots, S_T)$. The aim in [BHLP11] is to calculate the arbitrage free prices of a path-dependent derivative $\varphi(S_0, \dots, S_T)$, where $\varphi : \mathbb{R}^{T+1} \to \mathbb{R}$ is a given function. A real number p is called *subhedging price* for φ if there exists a strategy with initial capital p that invests in S and in European call options on S_t, $t = 1, \dots, T$, such that the payoff generated by this strategy if $(S_0, \dots, S_T) = (s_0, \dots, s_T) \in \mathbb{R}^{T+1}$ is bounded from above by $\varphi(s_0, \dots, s_T)$. This has to hold for all $(s_0, \dots, s_T) \in \mathbb{R}^{T+1}$, in contrast to the classical theory, where such an inequality has to be satisfied only almost surely. Their main result, shown by using techniques from optimal transport, is that the maximal subhedging price for φ is equal to the minimal martingale expectation of φ, i.e. to $\inf_Q E_Q(\varphi(S_0, \dots, S_T))$, where Q runs through all probability measures on $\mathbb{R}^{T+1}$, with marginals $\mathrm{law}_Q(S_t) = \mu_t$, that make S a martingale.

Since S is a discrete time process, here the stochastic integrals do not pose any problem and can be defined pathwise. In continuous time however, it is not a priori clear how to define stochastic integrals without a probability measure. In [DS12] this problem is resolved by only considering strategies that are of bounded variation, so that the integrals can be defined in a pathwise sense, for example by formally applying integration by parts. In [DOR13], Föllmer's pathwise Itô calculus [Föl79] is used to define pathwise stochastic integrals.

Föllmer assumes that S is a continuous real-valued path, and that the quadratic variation of S exists along a given sequence of partitions $\pi^n = \{t_0^n, \dots, t_{N_n}^n\}$ of $[0, T]$, i.e. that

$$[S, S]_n(t) := \sum_{k=0}^{N_n - 1} (S_{t \wedge t_{k+1}^n} - S_{t \wedge t_k^n})^2$$

converges for every $t \in [0, T]$ to a limit $[S, S](t)$ as $n \to \infty$. Of course the mesh size of the partition, $\max_{k=1,\dots,N_n} |t_{k+1}^n - t_k^n|$, should converge to zero as n tends to ∞. Föllmer shows that under these assumptions, if $F \in C^1(\mathbb{R}, \mathbb{R})$, then the non-anticipating Riemann sums

$$\sum_{k=1}^{N_n} F(S_{t_k^n})(S_{t_{k+1}^n \wedge t} - S_{t_k^n \wedge t})$$

converge to a limit that we denote by $\int_0^t F(S_s) \mathrm{d}S_s$. This is an analytical result, and the obtained integral satisfies Itô's formula. It is possible to generalize Föllmer's result into various directions. For example, continuity is not actually necessary. It suffices that S is càdlàg. In [DOR13] it is shown, building on the unpublished diploma thesis [Wue80], that it is possible to take F only weakly differentiable, with a derivative in $L^2([0, T])$. This requires a notion of pathwise local time. It is also possible to take a path-dependent functional F, see [CF10]. However, a basic limitation of Föllmer's pathwise integral is that it can essentially only handle one dimensional integrators. It is also possible to consider integrators with values in $\mathbb{R}^d$, but in that case F must be a gradient, i.e. $F = \nabla\varphi$ for some $\varphi : \mathbb{R}^d \to \mathbb{R}$.

Lyons' theory of rough paths [Lyo98] is somewhat similar in spirit to Föllmer's pathwise Itô calculus, but it reaches far beyond that, and also works in the multi dimensional case. Lyons does not assume that the quadratic variation of S exists, but rather that the iterated integrals

$$\left(\int_0^t S_s^i \mathrm{d}S_s^j\right)_{1\leq i,j\leq d}$$

can be constructed, and that S and its integrals are sufficiently regular. Since

$$[S^i, S^j](t) = S_t^i S_t^j - S_0^i S_0^j - \int_0^t S_s^i \mathrm{d}S_s^j - \int_0^t S_s^j \mathrm{d}S_s^i, \tag{3.2}$$

this is a more restrictive assumption than the one made by Föllmer. Lyons [Lyo98] and Gubinelli [Gub04] are then able to construct integrals of the type $\int_0^t G_s \mathrm{d}S_s$ if G is *controlled* by S. For further details see Section 4.2.2 below, but let us remark here that for example $G = F(S_\cdot)$ is controlled by S if $F \in C^2(\mathbb{R}^d, \mathbb{R}^d)$. Lyons [Lyo98] also shows that the Itô-Lyons map, which maps a path S to the solution to an SDE of the form $\mathrm{d}X_t = F(X_t)\mathrm{d}S_t$, depends continuously on S and its iterated integrals in a suitable topology, and that it is impossible to find a topology on a path space (without equipping paths with their iterated integrals), such that the space contains typical sample paths of Brownian motion, and such that the Itô-Lyons map is continuous. Moreover, while Föllmer's approach can only handle the "semimartingale setting", where S has finite $(2+\varepsilon)$–variation for every $\varepsilon > 0$, Lyons' approach allows S to have arbitrarily low regularity (finite p–variation for some $p < \infty$), provided that sufficiently many iterated integrals of S are given.

Let us also remark that (3.2) shows that the symmetric part of $(\int_0^t S_s^i \mathrm{d}S_s^j)_{1\leq i,j\leq d}$ can be recovered from S_t and $[S,S](t)$. If now F is a smooth function and $n \in \mathbb{N}^*$, then

$$\begin{aligned}
\int_0^t F(S_s)\mathrm{d}S_s &= \sum_{k=0}^{n-1} \int_{\frac{kt}{n}}^{\frac{(k+1)t}{n}} F(S_s)\mathrm{d}S_s \\
&\simeq \sum_{k=0}^{n-1} \int_{\frac{kt}{n}}^{\frac{(k+1)t}{n}} \left(F(S_{\frac{kt}{n}}) + \sum_{i=1}^{d} \partial_{x_i} F(S_{\frac{kt}{n}})(S_s^i - S_{\frac{kt}{n}}^i)\right) \mathrm{d}S_s \\
&= \sum_{k=0}^{n-1} \left(F(S_{\frac{kt}{n}})(S_{\frac{(k+1)t}{n}} - S_{\frac{kt}{n}}) + \sum_{i,j=1}^{d} \partial_{x_i} F^j(S_{\frac{kt}{n}}) \int_{\frac{kt}{n}}^{\frac{(k+1)t}{n}} (S_s^i - S_{\frac{kt}{n}}^i)\mathrm{d}S_s^j \right).
\end{aligned} \tag{3.3}$$

Rough path theory is essentially built on this heuristic argument. We see that the second term in the last line depends on the iterated integrals of S. However, if $F = \nabla\varphi$ for some smooth φ, then the derivative of F is the Hessian of φ, and therefore it is symmetric, i.e. $\partial_{x_i} F^j = \partial_{x_j} F^i$. So in that case the last addend in (3.3) only depends on the symmetric part of the iterated integrals of S, which, as we argued above, can be reconstructed from S and its quadratic variation. Therefore, it is not surprising that Föllmer can integrate gradients given only the quadratic variation of the integrator, but not its iterated integrals.

To summarize, Lyons' theory of rough paths allows to extend Föllmer's pathwise Itô integral to the multidimensional case, and it gives pathwise continuity results for the solutions to SDEs.

In a recent series of papers, Vovk [Vov11, Vov12] has introduced a model free, hedging based approach to mathematical finance that uses arbitrage considerations to examine which properties are satisfied by "typical price paths". One of the most important results is that typical price paths are either constant, or they possess a nontrivial quadratic variation. This gives an axiomatic justification for the use of Föllmer's pathwise Itô calculus in model free finance. Here we construct the iterated integrals of typical prices paths, giving the first steps towards an axiomatic justification for the use of rough path integrals in model free finance. To complete the argument, it is still necessary to show that the iterated integrals are sufficiently regular, which will be done in the upcoming work [PP13].

In Section 3.2 we define an outer content and introduce the notion of "typical price paths". In Section 3.3 we construct an Itô type integral that converges for typical price paths. We also indicate that typical price paths can be taken as integrators for the rough path integral.

3.2. Superhedging and typical price paths

Vovk's hedging based, model free approach to finance [Vov12] is based on a notion of outer content, which is given by the cheapest superhedging price.

Let $T > 0$ and let $\Omega = C([0,T], \mathbb{R}^d)$ be the space of d–dimensional continuous paths. The filtration $(\mathcal{F}_t)_{t\in[0,T]}$ is defined as $\mathcal{F}_t := \sigma(X_s : s \leq t)$, where $X_s(\omega) = \omega(s)$ denotes the coordinate process, and we set $\mathcal{F} := \mathcal{F}_T$. Stopping times τ and the associated σ–algebras $\mathcal{F}_\tau$ are defined as usually. A process $H : \Omega \times [0,T] \to \mathbb{R}^d$ is called a *simple strategy* if there exist stopping times $0 = \tau_0 < \tau_1 < \dots$, such that for every $\omega \in \Omega$ we have $\tau_n(\omega) = \infty$ for all but finitely many n, and bounded functions $F_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, such that F_n is $\mathcal{F}_{\tau_n}$–measurable for every n, for which

$$H_t(\omega) = \sum_{n=0}^{\infty} F_n(\omega) 1_{(\tau_n(\omega), \tau_{n+1}(\omega)]}(t).$$

In that case the integral

$$(H \cdot \omega)_t = \sum_{n=0}^{\infty} F_n(\omega)(\omega(t \wedge \tau_{n+1}(\omega)) - \omega(t \wedge \tau_n(\omega)))$$

is well defined for every $\omega \in \Omega$ and every $t \in [0,T]$. Here $F_n(\omega)(\omega(t \wedge \tau_{n+1}(\omega)) - \omega(t \wedge \tau_n(\omega)))$ denotes the usual inner product on $\mathbb{R}^d$.

Let $\lambda > 0$. A simple strategy H is called λ*–admissible* if $(H \cdot \omega)_t \geq -\lambda$ for all $\omega \in \Omega$ and all $t \in [0,T]$. The set of λ–admissible simple strategies is denoted by $\mathcal{H}_{\lambda,s}$.

Definition 3.2.1. The *outer content* of $A \subseteq \Omega$ is defined as the cheapest superhedging price,

$$\overline{P}(A) := \inf\Big\{\lambda > 0 : \exists\, (H^n)_{n\in\mathbb{N}} \subseteq \mathcal{H}_{\lambda,s} \text{ s.t. } \liminf_{n\to\infty}(\lambda + (H^n \cdot \omega)_T) \geq 1_A(\omega)\, \forall \omega \in \Omega\Big\}.$$

A set of paths $A \subseteq \Omega$ is called a *null set* if it has outer content zero.

Remark 3.2.2. By definition, every Itô stochastic integral is the limit of stochastic integrals against simple functions. Therefore, our definition of a superhedging price is essentially the same as in the classical setting, see (3.1). However, there is one important difference: Here we require superhedging with respect to all $\omega \in \Omega$, and not just almost surely.

Remark 3.2.3. Our definition is not quite the same as Vovk's. See Section 3.2.1 below for a discussion.

Remark 3.2.4 ([Vov12], p. 564)*.* An equivalent definition of $\overline{P}$ would be

$$\widetilde{P}(A) := \inf\Big\{\lambda > 0 : \exists\, (H^n)_{n\in\mathbb{N}} \subseteq \mathcal{H}_{\lambda,s} \text{ s.t. } \liminf_{n\to\infty} \sup_{t\in[0,T]} (\lambda + (H^n \cdot \omega)_t) \geq 1_A(\omega)\, \forall \omega \in \Omega\Big\}.$$

Clearly $\widetilde{P} \leq \overline{P}$. To see the opposite inequality, let $\widetilde{P}(A) < \lambda$. Let $(H^n)_{n\in\mathbb{N}} \subset \mathcal{H}_{\lambda,s}$ be a sequence of simple strategies such that $\liminf_{n\to\infty} \sup_{t\in[0,T]}(\lambda + (H^n \cdot \omega)_t) \geq 1_A(\omega)$, and let $\varepsilon > 0$. Define $\tau_n(\omega) := \inf\{t \in [0,T] : \lambda + \varepsilon + (H^n \cdot \omega)_t \geq 1\}$. Then the stopped strategy $G^n_t(\omega) := H^n_t(\omega) 1_{[0,\tau_n(\omega))}(t)$, $t \in [0,T]$, is in $\mathcal{H}_{\lambda,s} \subseteq \mathcal{H}_{\lambda+\varepsilon,s}$, and

$$\liminf_{n\to\infty}(\lambda + \varepsilon + (G^n \cdot \omega)_T) \geq \liminf_{n\to\infty} 1_{\{\omega' : \lambda+\varepsilon+\sup_{t\in[0,T]}(H^n\cdot\omega')_t \geq 1\}}(\omega) \geq 1_A(\omega).$$

Therefore, $\overline{P}(A) \leq \lambda + \varepsilon$, and since $\varepsilon > 0$ was arbitrary, we conclude that $\overline{P} \leq \widetilde{P}$ and therefore $\overline{P} = \widetilde{P}$.

Lemma 3.2.5. *The outer content $\overline{P}$ is countably subadditive. That is, if $(A_n)_{n\in\mathbb{N}}$ is a sequence of subsets of Ω, then $\overline{P}(\bigcup_n A_n) \leq \sum_n \overline{P}(A_n)$.*

Proof. Write $p_n := \overline{P}(A_n)$ for $n \in \mathbb{N}$. Let $\varepsilon > 0$ and let $(H^{n,m})_{m\in\mathbb{N}}$ be a sequence of $(p_n + \varepsilon 2^{-n})$–admissible simple strategies such that $\liminf_{m\to\infty}(p_n + \varepsilon 2^{-n} + (H^{n,m} \cdot \omega)_T) \geq 1_{A_n}(\omega)$ for all $\omega \in A_n$. Define for $m \in \mathbb{N}$ the 1–admissible simple strategy

$G^m := \sum_{n=0}^m H^{n,m}$. Let $k \in \mathbb{N}$. Then by Fatou's lemma

$$\begin{aligned}\liminf_{m\to\infty}\left(\sum_{n=0}^{\infty} p_n + 2\varepsilon + (G^m \cdot \omega)_T\right) &= \sum_{n=0}^{\infty} p_n + 2\varepsilon + \liminf_{m\to\infty}\sum_{n=0}^{m}(H^{n,m}\cdot\omega)_T \\ &\geq \sum_{n=0}^{k}(p_n + \varepsilon 2^{-n} + \liminf_{m\to\infty}(H^{n,m}\cdot\omega)_T) \\ &\geq 1_{\bigcup_{n=0}^k A_n}(\omega).\end{aligned}$$

Since the left hand side does not depend on k, we can replace $1_{\bigcup_{n=0}^k A_n}$ by $1_{\bigcup_n A_n}$, and the proof is complete. □

Maybe the most important property of $\overline{P}$ is that there exists an arbitrage interpretation for sets with outer content zero:

Lemma 3.2.6. *A set $A \subseteq \Omega$ is a null set if and only if there exists a sequence of 1–admissible simple strategies $(H^n)_n \subset \mathcal{H}_{1,s}$, such that*

$$\liminf_{n\to\infty}(1 + (H^n \cdot S)_T) \geq \infty \cdot 1_A(\omega), \tag{3.4}$$

where we set $\infty \cdot 0 = 0$.

Proof. If such a sequence exists, then we can scale it down by an arbitrary factor $\varepsilon > 0$ to obtain a sequence of strategies in $\mathcal{H}_{\varepsilon,s}$ that superhedge A. Therefore, $\overline{P}(A) = 0$.

If conversely $\overline{P}(A) = 0$, then for every $n \in \mathbb{N}$ there exists a sequence of simple strategies $(H^{n,m})_{m\in\mathbb{N}} \subset \mathcal{H}_{2^{-n-1},s}$ such that $2^{-n-1} + \liminf_{m\to\infty}(H^{n,m}\cdot\omega)_T \geq 1_A(\omega)$ for all $\omega \in \Omega$. For $m \in \mathbb{N}$ we define $G^m := \sum_{n=0}^m H^{n,m}$, so that $G^m \in \mathcal{H}_{1,s}$. For $k \in \mathbb{N}$ we obtain

$$\begin{aligned}\liminf_{m\to\infty}(1 + (G^m\cdot\omega)_T) &\geq \liminf_{m\to\infty}\sum_{n=0}^{m}(2^{-n-1} + (H^{n,m}\cdot\omega)_T) \\ &\geq \sum_{n=0}^{k}(2^{-n-1} + \liminf_{m\to\infty}(H^{n,m}\cdot\omega)_T) \geq k1_A(\omega).\end{aligned}$$

Since the left hand side does not depend on k, the sequence (G^m) satisfies (3.4). □

Remark 3.2.7. We interpret (3.4) as a model free version of the (NA1) property. More precisely, we interpret a set of paths $A \subseteq \Omega$ where (3.4) is satisfied as a model free arbitrage opportunity of the first kind.

We say that a property (P) holds for *typical price paths* if the set A where (P) is violated is a null set. In other words, if (P) holds for typical price paths, then it is possible to make an arbitrarily large profit by investing in paths that violate (P), without ever risking to lose more than the initial capital 1.

We can relate this model free notion of (NA1) to the classical (NA1) property. Every set of paths with outer content zero is in fact a "universal null set" that has measure zero under every probability measure for which the coordinate process satisfies (NA1).

Proposition 3.2.8. *Let $A \in \mathcal{F}$ be a null set, and let P be a probability measure on $(\Omega, \mathcal{F})$ such that the coordinate process satisfies (NA1). Then $P(A) = 0$.*

Proof. Let $(H^n)_{n \in \mathbb{N}}$ be a sequence of 1–admissible simple strategies such that for all $\omega \in A$ we have $\lim_{n \to \infty} (H^n \cdot \omega)_T = \infty$. For every $c > 0$ we obtain

$$
\begin{aligned}
P(A) &= P\Big(A \cap \{\omega : \liminf_{n \to \infty} (H^n \cdot \omega)_T > c\}\Big) \\
&\leq P\left(A \cap \left(\bigcup_{n \geq 0} \bigcap_{k \geq n} \{\omega : (H^k \cdot \omega)_T > c\}\right)\right) \\
&= \lim_{n \to \infty} P\left(A \cap \left(\bigcap_{k \geq n} \{\omega : (H^k \cdot \omega)_T > c\}\right)\right) \\
&\leq \sup_{H \in \mathcal{H}_{1,s}} P(\{\omega : (H \cdot \omega)_T > c\}).
\end{aligned}
$$

By assumption, the right hand side converges to 0 as $c \to \infty$, and therefore $P(A) = 0$. □

Remark 3.2.9. The proof shows that the measurability assumption on A can be relaxed: if $\overline{P}(A) = 0$, then A is contained in a measurable set of the form $\{\omega : \lim_{n \to \infty} (H^n \cdot \omega)_T = \infty\}$, and this set has P–measure zero for every P under which the coordinate process satisfies (NA1). Therefore, A is contained in the P–completion of $\mathcal{F}$, and gets assigned mass 0 by the unique extension of P to the completion.

Corollary 3.2.10. *Let $A \in \mathcal{F}$ be a null set, and let P be a probability measure on $(\Omega, \mathcal{F})$ such that the coordinate process is a P–local martingale. Then $P(A) = 0$.*

If under P the coordinate process satisfies only (NA) but not (NA1), then we do not expect that $P(A) = 0$ for every $A \in \mathcal{F}$ with $\overline{P}(A) = 0$.

3.2.1. Relation to Vovk's outer content

Our definition of the outer content $\overline{P}$ is not exactly the same as Vovk's [Vov12]. We find the definition given above more intuitive, but since we rely on some of the results established by Vovk, let us compare the two notions.

For $\lambda > 0$ we define the set of processes

$$
\mathcal{S}_\lambda := \left\{ \sum_{k=0}^{\infty} H^k : H^k \in \mathcal{H}_{\lambda_k, s}, \lambda_k > 0, \sum_{k=0}^{\infty} \lambda_k = \lambda \right\}.
$$

For every $G = \sum_{k \geq 0} H^k \in \mathcal{S}_\lambda$, every $\omega \in \Omega$, and every $t \in [0, T]$, the integral

$$
(G \cdot \omega)_t := \sum_{k \geq 0} (H^k \cdot \omega)_t = \sum_{k \geq 0} (\lambda_k + (H^k \cdot \omega)_t) - \lambda
$$

is well defined and takes values in $[-\lambda, \infty]$. Vovk then defines for $A \subseteq \Omega$ the cheapest superhedging price as

$$\overline{Q}(A) := \inf\{\lambda > 0 : \exists\, G \in \mathcal{S}_\lambda \text{ s.t. } \lambda + (G \cdot \omega)_T \geq 1_A(\omega)\, \forall \omega \in \Omega\}.$$

It is easy to see that $\overline{P}$ is dominated by $\overline{Q}$:

Lemma 3.2.11. *Let $A \subseteq \Omega$. Then $\overline{P}(A) \leq \overline{Q}(A)$.*

Proof. Let $G = \sum_k H^k$, with $H^k \in \mathcal{H}_{\lambda_k, s}$, and $\sum_k \lambda_k = \lambda$, and assume that $\lambda + (G \cdot \omega)_T \geq 1_A(\omega)$. Then $(\sum_{k=0}^n H^k)_{n \in \mathbb{N}}$ defines a sequence of simple strategies in $\mathcal{H}_{\lambda,s}$, such that

$$\liminf_{n \to \infty} \Big(\lambda + \Big(\Big(\sum_{k=0}^n H^k \Big) \cdot \omega \Big)_T \Big) = \lambda + (G \cdot \omega)_T \geq 1_A(\omega).$$

So if $\overline{Q}(A) < \lambda$, then also $\overline{P}(A) \leq \lambda$, and therefore $\overline{P}(A) \leq \overline{Q}(A)$. □

Remark 3.2.12. At least it is not easy to show that $\overline{P} = \overline{Q}$. Therefore it seems like we obtain a weaker result in Section 3.3, when we prove that a set A satisfies $\overline{P}(A) = 0$, compared to showing that it satisfies $\overline{Q}(A) = 0$. But actually we will (implicitly) work with a third notion of outer content, $\overline{R}$, defined as

$$\overline{R}(A) := \inf \Big\{ \sum_{n=0}^\infty \overline{S}(A_n) : A \subseteq \bigcup_{n \in \mathbb{N}} A_n \Big\},$$

where

$$\overline{S}(A) := \inf\{\lambda > 0 : \exists H \in \mathcal{H}_{\lambda,s} \text{ s.t. } \lambda + (H \cdot \omega)_T \geq 1_A(\omega)\},$$

and we will show $\overline{R}(A) = 0$. Since $\overline{P}$ and $\overline{Q}$ are countably subadditive, it is easy to see that they are both controlled by $\overline{R}$.

Recall that for $p \geq 1$ the p–variation $\|\cdot\|_{p-\text{var}}$ of a path $f : [0,T] \to \mathbb{R}^d$ is defined as

$$\|f\|_{p-\text{var}} := \sup \left\{ \left(\sum_{k=1}^n |f(t_k) - f(t_{k-1})|^p \right)^{1/p} : 0 = t_0 < \cdots < t_n = T, n \in \mathbb{N} \right\}.$$

Corollary 3.2.13. *For every $p > 2$, the set $A_p := \{\omega \in \Omega : \|\omega\|_{p-\text{var}} = \infty\}$ has outer content zero, i.e. $\overline{P}(A_p) = 0$.*

Proof. It is shown in Theorem 1 of Vovk [Vov08] that $\overline{Q}(A_p) = 0$, so the result follows from Lemma 3.2.11. □

It is a remarkable result of [Vov12] that if $\Omega = C([0,\infty), \mathbb{R})$ (i.e. if the asset price process is one dimensional), and if $A \subseteq \Omega$ is "invariant under time changes" and such that $\omega(0) = 0$ for all $\omega \in A$, then $A \in \mathcal{F}$, and $\overline{Q}(A) = \mu(A)$, where μ denotes the Wiener measure. This can be interpreted as a pathwise Dambis / Dubins-Schwarz theorem.

3.3. A pathwise Itô integral for typical price paths

Here we give a pathwise construction of an Itô type integral for typical price paths in $C([0,T],\mathbb{R}^d)$. The integral is in the spirit of Karandikar [Kar95]. If H is a suitable process, then we define a sequence of stopping times $(\tau_k^n)_{n,k\in\mathbb{N}}$, such that $\{\tau_k^n : k \in \mathbb{N}\} \subseteq \{\tau_k^{n+1} : k \in \mathbb{N}\}$ for all $n \in \mathbb{N}$, and such that the mesh size $\sup_{k\in\mathbb{N}} |\tau_{k+1}^n(\omega) - \tau_k^n(\omega)|$ converges to zero for every $\omega \in \Omega$, except possibly on intervals where ω is constant. We will then construct a sequence of simple 1–admissible strategies (G^n), such that for every $\omega \in \Omega$ either the Riemann sums

$$\sum_{k=0}^{\infty} H_{\tau_k^n}(\omega)(\omega(\tau_{k+1}^n \wedge \cdot) - \omega(\tau_k^n \wedge \cdot))$$

converge uniformly, or $(G^n \cdot \omega)_T$ diverges to ∞. This proves that for typical price paths the integral $(H \cdot \omega)$ can be defined as a continuous function.

Definition 3.3.1. A process $H : \Omega \times [0,T] \to \mathbb{R}^d$ is called *càdlàg* if $t \mapsto H_t(\omega)$ is càdlàg for every $\omega \in \Omega$. The process is called *adapted* if $\omega \mapsto H_t(\omega)$ is $\mathcal{F}_t$–measurable for every $t \in [0,T]$. For $p \geq 1$ it is called *p–variation preserving* if $t \mapsto H_t(\omega)$ has finite p–variation for every ω with finite p–variation.

Recall that if H is càdlàg and adapted, and if τ is a stopping time, then $H_\tau 1_{\{\tau \leq T\}}$ is $\mathcal{F}_\tau$–measurable; see for example [JS03], Proposition 1.1.21.

Let now H be a càdlàg and adapted process and let $n \in \mathbb{N}$. We define a sequence of stopping times $(\widetilde{\tau}_k^n)_{k\in\mathbb{N}}$ by $\widetilde{\tau}_0^n := 0$, and for $k \in \mathbb{N}$

$$\widetilde{\tau}_{k+1}^n := \inf\{t \in [\widetilde{\tau}_k^n, T] : |H_t(\omega) - H_{\widetilde{\tau}_k^n}(\omega)| + |\omega(t) - \omega(\widetilde{\tau}_k^n)| \geq 2^{-n}\}.$$

Since $t \mapsto H_t(\omega)$ and $t \mapsto \omega(t)$ are càdlàg, we obtain for every $\omega \in \Omega$ that $\widetilde{\tau}_k^n(\omega) = \infty$ for all but finitely many $k \in \mathbb{N}$. Write $\widetilde{\pi}_H^n := \{\widetilde{\tau}_k^n : k \in \mathbb{N}\}$. To obtain an increasing sequence of partitions, we take the union of the (π_H^n). More precisely, for $n \in \mathbb{N}$ we define $\tau_0^n := 0$ and then for $k \in \mathbb{N}$

$$\tau_{k+1}^n(\omega) := \min\left\{\tau(\omega) : \tau \in \bigcup_{m=0}^{n} \widetilde{\pi}_H^m, \tau(\omega) > \tau_k^n(\omega)\right\}.$$

If we set $\pi_H^n := \{\tau_k^n : k \in \mathbb{N}\}$, then $(\pi_H^n)_{n\in\mathbb{N}}$ is an increasing sequence of partitions. It is not necessarily true that the mesh size of this sequence of partitions converges to 0, because H and ω may be constant on some intervals. But for every $0 \leq s < t \leq T$ and every $\omega \in \Omega$ that is not constant on $[s,t]$ there exist $n, k \in \mathbb{N}$ such that $\tau_k^n(\omega) \in [s,t]$.

We define $N_t^n(\omega) := \max\{k \in \mathbb{N} : \tau_k^n(\omega) \leq t\}$, so that for every $\omega \in \Omega$ there are $N_t^n(\omega) + 1$ stopping times in π_H^n with values in $[0,t]$. We have the following estimate for $N_T^n(\omega)$:

Lemma 3.3.2. *Let $p \geq 1$. There exists a constant $C > 0$ such that for every $\omega \in \Omega$ and* *every $n \in \mathbb{N}$*

$$N_T^n(\omega) = \max\{k \in \mathbb{N} : \tau_k^n(\omega) < \infty\} \leq C 2^{np} \left(\|\omega\|_{p-\text{var}}^p + \|H(\omega)\|_{p-\text{var}}^p \right).$$

Proof. By definition, for every $k \in \mathbb{N}$ there exist $m \leq n$ and $\ell \in \mathbb{N}$, such that $\tau_k^n(\omega) = \widetilde{\tau}_\ell^m(\omega)$. Fix $m \leq n$ and write $\widetilde{N}_T^m(\omega) := \max\{\ell \in \mathbb{N} : \widetilde{\tau}_\ell^m(\omega) \in [0,T]\}$. The definition of $\widetilde{\tau}_{\ell+1}^m$ and the right-continuity of H and ω imply that

$$\begin{aligned}
\widetilde{N}^m(\omega) &\leq \sum_{\ell=0}^{\widetilde{N}_T^m(\omega)-1} 2^{mp} \left(|H_{\widetilde{\tau}_{\ell+1}^m}(\omega) - H_{\widetilde{\tau}_\ell^m}(\omega)| + |\omega(\widetilde{\tau}_{\ell+1}^m) - \omega(\widetilde{\tau}_\ell^m)| \right)^p \\
&\leq 2^{mp} C_p \sum_{\ell=1}^{\widetilde{N}_T^m(\omega)-1} \left(|H_{\widetilde{\tau}_{\ell+1}^m}(\omega) - H_{\widetilde{\tau}_\ell^m}(\omega)|^p + |\omega(\widetilde{\tau}_{\ell+1}^m) - \omega(\widetilde{\tau}_\ell^m)|^p \right) \\
&\leq 2^{mp} C_p \left(\|\omega\|_{p-\text{var}}^p + \|H(\omega)\|_{p-\text{var}}^p \right).
\end{aligned}$$

The result now follows by noting that

$$N_T^n(\omega) \leq \sum_{m=0}^{n} \widetilde{N}_T^m(\omega) \leq 2^{(n+1)p} C_p \left(\|\omega\|_{p-\text{var}}^p + \|H(\omega)\|_{p-\text{var}}^p \right),$$

so that we can set $C := 2^p C_p$. ☐

The idea of relating the number of upcrossings to the p–variation goes at least back to Bruneau [Bru79], and Lemma 3.3.2 can be seen as a crude adaption of Bruneau's result.

In Lemma D.1 in the Appendix we present a pathwise version of the Hoeffding inequality that is due to Vovk. This will be needed in the proof below.

At this point we are ready to state and prove the main result of this section. The following construction is inspired by Karandikar [Kar95], whereas the proof follows [Vov12], Lemma 8.1.

Theorem 3.3.3. *Let H be a càdlàg, adapted process that is p–variation preserving for some $p \in (2,3)$. Define for $n \in \mathbb{N}$ the partition $\pi_H^n = \{\tau_k^n : k \in \mathbb{N}\}$ as above. Then for typical price paths, the non-anticipating Riemann sums*

$$I_n(H, \mathrm{d}\omega)(t) := \sum_{k=0}^{\infty} H_{\tau_k^n}(\omega)(\omega(\tau_{k+1}^n \wedge t) - \omega(\tau_k^n \wedge t))$$

converge uniformly to a limit that we denote by $\int_0^\cdot H_s \mathrm{d}\omega_s$.

Proof. For every $n \in \mathbb{N}$ we define the process

$$H_t^n := \sum_{k=0}^{\infty} H_{\tau_k^n}(\omega) 1_{[\tau_k^n, \tau_{k+1}^n)}(t).$$

Since H is right-continuous, we have $\sup_{t\in[0,T]}|H_t - H_t^n| \leq 2^{-n}$, and thus $\sup_{t\in[0,T]}|H_t^n - H_t^{n-1}| \leq 2^{-n+2}$ for all $n \in \mathbb{N}$. Moreover, $\pi_H^{n-1} \subseteq \pi_H^n$ for all $n \geq 1$, which leads to

$$\begin{aligned} &I_n(H, \mathrm{d}\omega)(t) - I_{n-1}(H, \mathrm{d}\omega)(t) \\ &\qquad = \sum_{k=0}^{\infty} \left(H^n_{\tau_k^{n-1}}(\omega) - H^{n-1}_{\tau_k^{n-1}}(\omega) \right) \left(\omega(\tau_{k+1}^n \wedge t) - \omega(\tau_k^n \wedge t) \right). \end{aligned}$$

By definition of the stopping times $(\tau_k^n)_k$, we have

$$\begin{aligned} &\sup_{t\in[0,T]} \left| \left(H^n_{\tau_k^{n-1}}(\omega) - H^{n-1}_{\tau_k^{n-1}}(\omega) \right) \left(\omega(\tau_{k+1}^n \wedge t) - \omega(\tau_k^n \wedge t) \right) \right| \\ &\qquad \leq 2^{-n+2} 2^{-n} = 2^{-2n+2}. \end{aligned}$$

Hence, the pathwise Hoeffding inequality, Lemma D.1 in Appendix D, implies for every $\lambda \in \mathbb{R}$ the existence of a 1–admissible simple strategy $G^\lambda \in \mathcal{H}_{1,s}$, such that

$$\begin{aligned} 1 + (G^\lambda \cdot \omega)_t &\geq \exp\left(\lambda (I_n(H, \mathrm{d}\omega)(t) - I_{n-1}(H, \mathrm{d}\omega)(t)) - \frac{\lambda^2}{2} N_t^n(\omega) 2^{-4n+4} \right) \\ &=: \mathcal{E}_t^{\lambda,n}(\omega) \end{aligned} \tag{3.5}$$

for all $\omega \in \Omega$ and all $t \in [0,T]$. If $a > 0$, then the strategies $G^{2^n}/(2^n a)$ and $G^{-2^n}/(2^n a)$ are both in $\mathcal{H}_{2^{-n}/a,s}$, and therefore we can apply Remark 3.2.4 in conjunction with (3.5) to obtain that

$$\begin{aligned} &\overline{P}\left(\sup_{t\in[0,T]} 2^{-n} \mathcal{E}_t^{2^n,n}(\omega) + \sup_{t\in[0,T]} 2^{-n} \mathcal{E}_t^{-2^n,n}(\omega) \geq 2a \right) \\ &\quad \leq \overline{P}\left(\sup_{t\in[0,T]} \frac{\mathcal{E}_t^{2^n,n}(\omega)}{2^n a} \geq 1 \right) + \overline{P}\left(\sup_{t\in[0,T]} \frac{\mathcal{E}_t^{-2^n,n}(\omega)}{2^n a} \geq 1 \right) \leq \frac{2^{-n+1}}{a}. \end{aligned} \tag{3.6}$$

Summing (3.6) over n and letting a tend to ∞, we see that

$$\overline{P}\left(\sup_{n\in\mathbb{N}} \left(\sup_{t\in[0,T]} 2^{-n} \mathcal{E}_t^{2^n,n}(\omega) + \sup_{t\in[0,T]} 2^{-n} \mathcal{E}_t^{-2^n,n}(\omega) \right) = \infty \right) = 0. \tag{3.7}$$

Hence, for typical price paths $\omega \in \Omega$, there exists $n_0(\omega)$ such that for all $n \geq n_0(\omega)$ we have $\sup_{t\in[0,T]} \mathcal{E}_t^{2^n,n}(\omega) < 2^n n$. Note that $N_t^n(\omega)$ is increasing in t, and therefore we can take the logarithm to see that this implies

$$\sup_{t\in[0,T]} 2^n (I_n(H, \mathrm{d}\omega)(t) - I_{n-1}(H, \mathrm{d}\omega)(t)) < \frac{2^{2n}}{2} N_T^n(\omega) 2^{-4n+4} + n\log(2) + \log(n). \tag{3.8}$$

If $\sup_{t\in[0,T]} \mathcal{E}_t^{-2^n,n}(\omega) < 2^n n$, then we obtain the same inequality as in (3.8), only that the sign of the left hand side is reversed. So if both $\sup_{t\in[0,T]} \mathcal{E}_t^{2^n,n}(\omega) < 2^n n$ and

$\sup_{t\in[0,T]} \mathcal{E}_t^{-2^n,n}(\omega) < 2^n n$, then

$$
\begin{aligned}
\sup_{t\in[0,T]} |I_n(H, \mathrm{d}\omega)(t) - I_{n-1}(H, \mathrm{d}\omega)(t)| &< N_T^n(\omega) 2^{-3n+3} + 2^{-n}(n \log(2) + \log(n)) \\
&\leq C \left(\|\omega\|_{p-\mathrm{var}}^p + \|H(\omega)\|_{p-\mathrm{var}}^p \right) 2^{-n(3-p)+3} + 2^{-n}(n \log(2) + \log(n)),
\end{aligned} \tag{3.9}
$$

where the last step follows from Lemma 3.3.2.

Since $p < 3$, we can combine (3.7) and (3.9) to obtain

$$
\begin{aligned}
&\overline{P}\left(\sum_{n=1}^{\infty} \sup_{t\in[0,T]} |I_n(H, \mathrm{d}\omega)(t) - I_{n-1}(H, \mathrm{d}\omega)(t)| = \infty \right) \\
&\qquad\qquad \leq \overline{P}(\|\omega\|_{p-\mathrm{var}}^p + \|H(\omega)\|_{p-\mathrm{var}}^p = \infty) = \overline{P}(\|\omega\|_{p-\mathrm{var}} = \infty) = 0,
\end{aligned}
$$

where the second to last step uses that H is p–variation preserving, and the last step is Corollary 3.2.13, which can be applied because $p > 2$. □

Remark 3.3.4. While the integral $\int_0^\cdot H_s \mathrm{d}\omega_s$ converges for all typical price paths, the strategies that we constructed in the proof depend on H. Therefore, also the null set where $\int_0^\cdot H_s \mathrm{d}\omega_s$ does not exist depends on H. Since there are uncountably many processes H, it is a priori not clear whether a "universal null set" exists, outside of which all integrals can be constructed. It is possible to obtain such a universal null set by using an analytic construction of the integral, such as Föllmer's or Lyons' constructions.

Remark 3.3.5. At some points our analysis was rather crude, and therefore we did not obtain optimal results. For example, it is not actually necessary to assume that H is p–variation preserving. Also, here we just considered one fixed sequence of partitions $(\pi_H^n)_{n\in\mathbb{N}}$. It is possible to show that the Riemann sums over any sequence of partitions converge to the same limit, as long as the mesh size of the partition converges rapidly enough to 0 (in a way that depends on H, uniformly in ω). Furthermore, one can show that for $p > 2$ the "area"

$$
\Phi_{s,t}(\omega) := \left(\Phi_{s,t}^{i,j}(\omega) \right)_{1\leq i,j\leq d} := \left(\int_s^t \omega^i(r) \mathrm{d}\omega^j(r) - \omega^i(s)(\omega^j(t) - \omega^j(s)) \right)_{1\leq i,j\leq d}
$$

satisfies

$$
\|\Phi\|_{p/2-\mathrm{var}} := \sup \left\{ \sum_{k=1}^{n} |\Phi_{t_{k-1},t_k}|^{p/2} : 0 = t_0 < \cdots < t_n = T, n \in \mathbb{N} \right\} < \infty
$$

for typical price paths ω. This condition is required to use ω as an integrator for Lyons' rough path integral. The rough path integral is for example defined for $F \in C^2$ as uniform limit of the Riemann sums

$$
\lim_{n\to\infty} \sum_{k=0}^{n-1} \left(F(\omega(t_k^n))(\omega(t_{k+1}^n \wedge t) - \omega(t_k^n \wedge t)) + \mathrm{D}F(\omega(t_k^n)) \Phi_{t_k^n \wedge t, t_{k+1}^n \wedge t} \right), \tag{3.10}
$$

where $0 = t_0^n < \cdots < t_n^n = T$, $n \in \mathbb{N}$, is an arbitrary sequence of partitions with mesh size converging to 0. Here we see that there is a small problem with the use of the rough path integral in finance: the term $\mathrm{D}F(\omega(t_k^n))\Phi_{t_k^n \wedge t, t_{k+1}^n \wedge t}$ in (3.10) is not an increment of ω, and therefore it is technically not possible to interpret the integral process as capital obtained by investing in ω. But this can be resolved, because we can show that if $(\tau_k^n)_{n,k\in\mathbb{N}}$ is a double sequence of stopping times such that

$$\int_0^{\cdot} \omega(s)\mathrm{d}\omega(s) = \lim_{n\to\infty} \sum_{k=0}^{\infty} \omega(\tau_k^n)(\omega(\tau_{k+1}^n \wedge \cdot) - \omega(\tau_k^n \wedge \cdot)),$$

then under some additional conditions also the rough path integral $\int_0^{\cdot} F(\omega(s))\mathrm{d}\omega(s)$ is given by

$$\int_0^{\cdot} F(\omega(s))\mathrm{d}\omega(s) = \lim_{n\to\infty} \sum_{k=0}^{\infty} F(\omega(\tau_k^n))(\omega(\tau_{k+1}^n \wedge \cdot) - \omega(\tau_k^n \wedge \cdot)).$$

These and other results will be presented in the upcoming work [PP13].

4. A Fourier approach to pathwise stochastic integration

Here we use the decomposition of continuous functions in terms of the Schauder functions, $f(t) = \sum_{pm} f_{pm}\varphi_{pm}(t)$, to give a pathwise definition of the integral $\int_0^t f(s)\mathrm{d}g(s)$ as

$$\int_0^t f(s)\mathrm{d}g(s) := \sum_{pm}\sum_{qn} f_{pm}g_{qn}\int_0^t \varphi_{pm}(s)\mathrm{d}\varphi_{qn}(s).$$

If f is α–Hölder continuous and g is β–Hölder continuous and $\alpha+\beta > 1$, then we recover Young's integral. For $\alpha+\beta \leq 1$ we define a rough path integral in terms of the Schauder decomposition.

This new approach to rough paths is quite elementary, and it becomes obvious why paths have to be enhanced with their Lévy area if we want to obtain a pathwise continuous stochastic integral. It also leads to simple recursive algorithms for the calculation of stochastic integrals.

In the setting of Itô integration, we show that under suitable conditions, the Itô rough path integral can be obtained as limit of nonanticipating Riemann sums involving only the integrator and not its iterated integrals.

4.1. Introduction

It is a classical result of Ciesielski [Cie60] that $C^\alpha := C^\alpha([0,1],\mathbb{R}^d)$, the space of α–Hölder continuous functions on $[0,1]$ with values in $\mathbb{R}^d$, is isomorphic to $\ell^\infty(\mathbb{R}^d)$, the space of bounded sequences with values in $\mathbb{R}^d$. The isomorphism gives a Fourier decomposition of a Hölder-continuous function f as

$$f = \sum_{p,m}\langle H_{pm}, \mathrm{d}f\rangle G_{pm}$$

where (H_{pm}) are the Haar functions and (G_{pm}) are the Schauder functions. Ciesielski proved that a continuous function f is in $C^\alpha([0,1],\mathbb{R}^d)$ if and only if the coefficients $(\langle H_{pm}, \mathrm{d}f\rangle)_{p,m}$ satisfy $\sup_{p,m} 2^{p(\alpha-1/2)}|\langle H_{pm},\mathrm{d}f\rangle| < \infty$.

Since then this isomorphism has been extended to many other Fourier and wavelet bases, where one can show the same type of results: classical function spaces, such as the space of Hölder continuous functions, or the space of functions with a certain Besov regularity, are in one-to-one correspondence with those functions for which the coefficients in a fixed basis have the correct decay. See for example Triebel [Tri06].

The isomorphism based on Schauder functions still plays a special role in stochastic analysis, because the coefficients in the Schauder basis have the pleasant property that they are just rescaled second order increments of f. So if f is a stochastic process with known distribution, then also the distribution of its coefficients in the Schauder basis is known explicitly. This makes the Schauder functions a very useful tool in stochastic analysis. For example, one of the most elegant constructions of Brownian motion, the Lévy-Ciesielski construction, is based on them. Ciesielski's isomorphism can also be used to give a simple proof of Kolmogorov's continuity criterion. An incomplete list with applications of Schauder functions in stochastic analysis will be given below.

Another convenient property of the Schauder functions is that they are piecewise linear, and therefore their iterated integrals $\int_0^\cdot G_{pm}(s)\mathrm{d}G_{qn}(s)$, can be easily calculated. This makes them an ideal tool for our purpose of studying pathwise stochastic integrals.

If we are given two Hölder-continuous functions f and g on $[0,1]$ with values in $\mathcal{L}(\mathbb{R}^d,\mathbb{R}^n)$ and $\mathbb{R}^d$ respectively, then we formally define

$$\int_0^t f(s)\mathrm{d}g(s) := \sum_{p,m}\sum_{q,n}\langle H_{pm},\mathrm{d}f\rangle\langle H_{qn},\mathrm{d}g\rangle \int_0^t G_{pm}(s)\mathrm{d}G_{qn}(s),$$

provided the limit exists. Our first observation, which is of course well known, is that the integral introduces a bounded operator from $C^\alpha([0,1],\mathcal{L}(\mathbb{R}^d,\mathbb{R}^n))\times C^\beta([0,1],\mathbb{R}^d)$ to $C^\beta([0,1],\mathbb{R}^n)$ if and only if $\alpha+\beta>1$. In this case we recover Young's integral. In the derivation of the Young integral, we identify different components of the integral that exhibit different behavior: we have

$$\int_0^t f(s)\mathrm{d}g(s) = S(f,g)(t)+\pi_<(f,g)(t)+L(f,g)(t),$$

where S is the *symmetric part*, $\pi_<$ is the *paraproduct*, and $L(f,g)$ is the *Lévy area*. The operators S and $\pi_<$ are defined for arbitrary α and β, and it is only the Lévy area that requires $\alpha+\beta>1$. We are therefore looking for a pathwise way of defining $L(f,g)$ for suitable g. Considering the regularity of the three operators, we have $S(f,g)\in C^{\alpha+\beta}$ and $\pi_<(f,g)\in C^\beta$ and $L(f,g)\in C^{\alpha+\beta}$, whenever the latter is defined. Therefore, in the Young regime $\int_0^\cdot f(s)\mathrm{d}g(s)-\pi_<(f,g)\in C^{\alpha+\beta}$. Similarly we can show that for smooth functions F we have $F(f)\in C^\alpha$ but $F(f)-\pi_<(\mathrm{D}F(f),f)\in C^{2\alpha}$. In both cases the "rough component" is given by $\pi_<$. This inspires us to call a function $f\in C^\beta$ *controlled* by g if there exists a function $f^g\in C^\beta$ such that $f-\pi_<(f^g,g)\in C^{2\beta}$. Our aim is then to construct the Lévy area $L(f,g)$ for $\beta<1/2$ and f controlled by g. If $\beta>1/3$, then the term $L(f-\pi_<(f^g,g),g)$ is well defined, and it suffices to make sense of the term $L(\pi_<(f^g,g),g)$. This is achieved with the following commutator estimate:

$$\left\| L(\pi_<(f^g,g),g)-\int_0^\cdot f^g(s)\mathrm{d}L(g,g)(s)\right\|_{3\beta} \le \|f^g\|_\beta\|g\|_\beta\|g\|_\beta.$$

Therefore, the integral $\int_0^\cdot f(s)\mathrm{d}g(s)$ can be constructed for all f that are controlled by g, provided that $L(g,g)$ can be constructed. In other words, we have found an alternative formulation of Lyons' [Lyo98] rough path integral, at least for Hölder continuous functions of Hölder exponent larger than 1/3.

Since we approximate f and g by functions of bounded variation, our integral is of Stratonovich type, i.e. it satisfies the usual integration by parts rule. We also consider a non-anticipating Itô type integral, that can essentially be reduced to the Stratonovich case with the help of the quadratic variation.

The last remaining problem is then to construct the Lévy area $L(g, g)$ for suitable stochastic processes g. We construct the Lévy area for certain hypercontractive processes. For continuous martingales that possess sufficiently many moments we give a construction of the Itô iterated integrals that allows us to use them as integrators for our pathwise Itô integral.

Below we give some references to the use of Schauder functions in stochastic analysis, and to rough paths. In Section 4.2 we recall some details on Ciesielski's isomorphism, and we give a short overview on rough paths and Young integration. In Section 4.3 we develop a paradifferential calculus in terms of Schauder functions, and we examine the different components of Young's integral. In Section 4.4 we construct the rough path integral based on Schauder functions. Section 4.5 develops the pathwise Itô integral. And in Section 4.6 we construct the Lévy area for suitable stochastic processes.

Relevant literature

Starting with the Lévy-Ciesielski construction of Brownian motion, Schauder functions have been a very popular tool in stochastic analysis. They can be used to prove in a comparatively easy way that stochastic processes belong to Besov spaces; see for example Ciesielski, Kerkyacharian, and Roynette [CKR93], Roynette [Roy93], and Rosenbaum [Ros09]. Baldi and Roynette [BR92] have used Schauder functions to extend the large deviation principle for Brownian motion, Schilder's theorem, from the uniform to the Hölder topology; see also Ben Arous and Ledoux [BL94] for the extension to diffusions, Eddahbi, N'zi, and Ouknine [ENO99] for the large deviation principle for diffusions in Besov spaces, and Andresen, Imkeller, and Perkowski [AIP13] for the large deviation principle for a Hilbert space valued Wiener process in Hölder topology. Ben Arous, Grădinaru, and Ledoux [BGL94] use Schauder functions to extend the Stroock-Varadhan support theorem for diffusions from the uniform to the Hölder topology. Lyons and Zeitouni [LZ99] use Schauder functions to prove exponential moment bounds for Stratonovich iterated integrals of a Brownian motion if the Brownian motion is conditioned to stay in a small ball. Gantert [Gan94] uses Schauder functions to associate to every sample path of the Brownian bridge a sequence of probability measures on path space, and continues to show that for almost all sample paths these measures converge to the distribution of the Brownian bridge. This shows that the law of the Brownian bridge can be reconstructed from a single "typical sample path".

Concerning integrals based on Schauder functions, there are three important references: Roynette [Roy93] constructs a version of Young's integral on Besov spaces and shows that in the one dimensional case the Stratonovich integral $\int_0^\cdot F(W_s)\mathrm{d}W_s$, where W is a Brownian motion, and $F \in C^2$, can be defined in a deterministic manner with the help of Schauder functions. Roynette also constructs more general Stratonovich integrals with the help of Schauder functions, but in that case only almost sure convergence is established, where the null set depends on the integrand, and the integral is not a deterministic operator. Ciesielski, Kerkyacharian, and Roynette [CKR93] slightly extend the Young integral of [Roy93], and simplify the proof by developing the integrand in the Haar basis and not in the Schauder basis. They also construct pathwise solutions to SDEs driven by fractional Brownian motions with Hurst index $H > 1/2$. Kamont [Kam94] extends the approach of [CKR93] to define a multiparameter Young integral for functions in anisotropic Besov spaces.

Rough paths have been introduced by Lyons [Lyo98], see also [Lyo95, LQ96, LQ97] for previous results. Lyons observed that solution flows to SDEs (or more generally ordinary differential equations (ODEs) driven by rough signals) can be defined in a pathwise, continuous way, if paths are equipped with sufficiently many iterated integrals. More precisely, if a path has finite p–variation for some $p \geq 1$, then one needs to associate $\lfloor p \rfloor$ iterated integrals to it to obtain an object which can be taken as the driving signal in an ODE, such that the solution to the ODE depends continuously on the signal. Gubinelli [Gub04, Gub10] simplified the theory of rough paths by introducing the concept of controlled paths, on which we will strongly rely in what follows. Roughly speaking, a path f is controlled by the reference path g if the small scale fluctuations of f "look like those of g". Good monographs on rough paths are [LQ02, LCL07, FV10b], and Friz and Hairer [FH13], which is currently in preparation.

4.2. Preliminaries

4.2.1. Ciesielski's isomorphism

Here we present Ciesielski's isomorphism between $C^\alpha([0,1],\mathbb{R}^d)$ and $\ell^\infty(\mathbb{R}^d)$.

The *Haar functions* $(H_{pm}, p \in \mathbb{N}, 1 \leq m \leq 2^p)$ are defined as

$$H_{pm}(t) := \begin{cases} \sqrt{2^p}, & t \in \left[\frac{m-1}{2^p}, \frac{2m-1}{2^{p+1}}\right), \\ -\sqrt{2^p}, & t \in \left[\frac{2m-1}{2^{p+1}}, \frac{m}{2^p}\right), \\ 0, & \text{otherwise.} \end{cases}$$

If completed by $H_{00} \equiv 1$, the Haar functions are an orthonormal basis of $L^2([0,1], \mathrm{d}t)$. We also define $H_{p0} \equiv 0$ for $p \geq 1$, which will allow us to write expressions such as $\sum_{p\geq 0}\sum_{m=0}^{2^p} H_{pm}$. The primitives of the Haar functions are called *Schauder functions*, and they are given by $G_{pm}(t) := \int_0^t H_{pm}(s)\mathrm{d}s$ for $t \in [0,1]$, $p \in \mathbb{N}$, $0 \leq m \leq 2^p$. More explicitly $G_{00}(t) = t$ and for $p \in \mathbb{N}$, $1 \leq m \leq 2^p$

$$G_{pm}(t) = \begin{cases} 2^{p/2}\left(t - \frac{m-1}{2^p}\right), & t \in \left[\frac{m-1}{2^p}, \frac{2m-1}{2^{p+1}}\right), \\ -2^{p/2}\left(t - \frac{m}{2^p}\right), & t \in \left[\frac{2m-1}{2^{p+1}}, \frac{m}{2^p}\right), \\ 0, & \text{otherwise.} \end{cases}$$

Since every G_{pm} satisfies $G_{pm}(0) = 0$, we are only able to expand functions f with $f(0) = 0$ in terms of this family (G_{pm}). Therefore, we complete (G_{pm}) once more, by defining $G_{-10}(t) := 1$ for all $t \in [0,1]$.

To abbreviate notation, we define the times t^i_{pm}, $i = 0,1,2$, by setting

$$t^0_{pm} := \frac{m-1}{2^p}, \quad t^1_{pm} := \frac{2m-1}{2^{p+1}}, \quad t^2_{pm} := \frac{m}{2^p},$$

for $p \in \mathbb{N}$ and $1 \le m \le 2^p$. For $(p,m) = (-1,0)$ and $(p,m) = (0,0)$ we set $t^0_{-10} := 0$, $t^1_{-10} := 0$, $t^2_{-10} := 1$ and $t^0_{00} := 0$, $t^1_{00} := 1$, $t^2_{00} := 1$. The definition of t^i_{-10} and t^i_{00} for $i \neq 1$ is rather arbitrary, but the definition for $i = 1$ simplifies for example the statement of Lemma 4.2.1 below. It is also convenient to define $t^i_{p0} := 0$ for $p \ge 1$ and $i = 0,1,2$.

If f is a continuous function on $[0,1]$ with values in $\mathbb{R}^d$, then we define for $p \in \mathbb{N}$ and $1 \le m \le 2^p$ by formally applying integration by parts

$$\begin{aligned}\langle H_{pm}, \mathrm{d}f\rangle :=& 2^{\frac{p}{2}}\left[\left(f\left(t^1_{pm}\right) - f\left(t^0_{pm}\right)\right) - \left(f\left(t^2_{pm}\right) - f\left(t^1_{pm}\right)\right)\right]\\ =& 2^{\frac{p}{2}}\left[2f\left(t^1_{pm}\right) - f\left(t^0_{pm}\right) - f\left(t^2_{pm}\right)\right]\end{aligned}$$

and $\langle H_{00}, \mathrm{d}f\rangle := f(1) - f(0)$ as well as $\langle H_{-10}, \mathrm{d}f\rangle := f(0)$. Note that we only defined G_{-10} and not H_{-10}, and that the definition of $\langle H_{-10}, \mathrm{d}f\rangle$ is to be understood as convention.

Lemma 4.2.1. *The function*

$$\begin{aligned}f_k :=& \langle H_{-10}, \mathrm{d}f\rangle G_{-10} + \langle H_{00}, \mathrm{d}f\rangle G_{00} + \sum_{p=0}^{k}\sum_{m=1}^{2^p}\langle H_{pm}, \mathrm{d}f\rangle G_{pm}\\ =& \sum_{p=-1}^{k}\sum_{m=0}^{2^p}\langle H_{pm}, \mathrm{d}f\rangle G_{pm}\end{aligned}$$

is the linear interpolation of f between the points $t^1_{-10}, t^1_{00}, t^1_{pm}$, $0 \le p \le k, 1 \le m \le 2^p$. So if f is continuous, then f_k converges uniformly to f as $k \to \infty$.

Proof. The statement follows easily by induction. □

Ciesielski [Cie60] observed that if f is *Hölder-continuous*, then the series f_k converges absolutely, and the speed of convergence of f_k to f can be estimated in terms of the Hölder norm of f. The norms $\|\cdot\|_\infty$ and $\|\cdot\|_{C^\alpha}$ are defined respectively as

$$\|f\|_\infty := \sup_{t\in[0,1]} |f(t)| \qquad \text{and} \qquad \|f\|_{C^\alpha} := \|f\|_\infty + \sup_{0\le s<t\le 1}\frac{|f_{s,t}|}{|t-s|^\alpha},$$

where we write $f_{s,t} := f(t) - f(s)$.

Lemma 4.2.2 ([Cie60]). *Let $\alpha \in (0,1)$. A continuous function $f : [0,1] \to \mathbb{R}^d$ is in C^α if and only if $\sup_{p,m} 2^{p(\alpha-1/2)}|\langle H_{pm}, \mathrm{d}f\rangle| < \infty$. In this case*

$$\sup_{p,m} 2^{p(\alpha-1/2)}|\langle H_{pm}, \mathrm{d}f\rangle| \simeq \|f\|_\alpha \text{ and} \tag{4.1}$$

$$\|f - f_{N-1}\|_\infty = \left\| \sum_{p=N}^{\infty} \sum_{m=0}^{2^p} |\langle H_{pm}, \mathrm{d}f\rangle| G_{pm} \right\|_\infty \lesssim \|f\|_\alpha 2^{-\alpha N}.$$

Before we continue, we slightly adapt the notation. We want to get rid of the factor $2^{-p/2}$ in (4.1), and therefore we define for $p \in \mathbb{N}$ and $0 \le m \le 2^p$ the rescaled functions

$$\chi_{pm} := 2^{\frac{p}{2}} H_{pm} \qquad \text{and} \qquad \varphi_{pm} := 2^{\frac{p}{2}} G_{pm},$$

as well as $\varphi_{-10} := G_{-10} \equiv 1$. Note that for $p \in \mathbb{N}$ and $1 \le m \le 2^p$

$$\max_{t\in[0,1]} |\varphi_{pm}(t)| = \varphi_{pm}(t^1_{pm}) = 2^{\frac{p}{2}} \int_{t^0_{pm}}^{t^1_{pm}} 2^{\frac{p}{2}} \mathrm{d}s = 2^p \left(\frac{2m-1}{2^{p+1}} - \frac{2m-2}{2^{p+1}} \right) = \frac{1}{2},$$

so that $\|\varphi_{pm}\|_\infty \le 1$ for all p, m. The expansion of f in terms of (φ_{pm}) is given by $f_k = \sum_{p=0}^{k} \sum_{m=0}^{2^p} f_{pm}\varphi_{pm}$, where $f_{-10} := f(1)$, and $f_{00} := f(1) - f(0) = f_{0,1}$ and for $p \in \mathbb{N}$ and $m \ge 1$

$$f_{pm} := 2^{-p}\langle \chi_{pm}, \mathrm{d}f\rangle = 2f\left(t^1_{pm}\right) - f\left(t^0_{pm}\right) - f\left(t^2_{pm}\right) = f_{t^0_{pm}, t^1_{pm}} - f_{t^1_{pm}, t^2_{pm}}.$$

We also write $\langle \chi_{pm}, \mathrm{d}f\rangle := 2^p f_{pm}$ for all values of (p,m), also for $(p,m) = (-1,0)$, despite not having defined χ_{-10}.

We will mainly measure the regularity of functions by the size of their coefficients in the Schauder series expansion:

Definition 4.2.3. For $\alpha > 0$ and continuous $f : [0,1] \to \mathbb{R}^d$ the norm $\|\cdot\|_\alpha$ is defined as $\|f\|_\alpha := \sup_{pm} 2^{p\alpha}|f_{pm}|$. We then define the space

$$\mathcal{C}^\alpha := \mathcal{C}^\alpha(\mathbb{R}^d) = \left\{ f : [0,1] \to \mathbb{R}^d : f \text{ is continuous and } \|f\|_\alpha < \infty \right\}.$$

It is easy to see that $\mathcal{C}^\alpha$ is isomorphic to $\ell^\infty(\mathbb{R}^d)$. In particular, $\mathcal{C}^\alpha$ is a Banach space.

For $\alpha \in (0,1)$, Ciesielski's isomorphism, Lemma 4.2.2, implies that $\mathcal{C}^\alpha = C^\alpha([0,1], \mathbb{R}^d)$. For $\alpha = 1$ it can be shown that $\mathcal{C}^1$ is the Zygmund space of continuous functions f satisfying $|2f(x) - f(x+h) - f(x-h)| \lesssim h$. But for $\varepsilon > 0$, there is no reasonable identification of $\mathcal{C}^{1+\varepsilon}$ with a classical function space. The space $C^{1+\varepsilon}([0,1], \mathbb{R}^d)$ consists of all continuously differentiable functions f with ε–Hölder continuous derivative $\mathrm{D}f$. But since the tent shaped functions φ_{pm} are not continuously differentiable, even an f with a finite expansion in terms of (φ_{pm}) is generally not in $C^{1+\varepsilon}$, despite being in $\mathcal{C}^\alpha$ for all $\alpha > 0$.

One might ask if the a priori requirement of f being continuous could be relaxed. It can, but not much. To obtain continuity of f from its coefficients (f_{pm}) is only possible if f is uniquely determined by the values $(f(t^i_{pm}))_{i,p,m}$. This is the case if f is right- or left-continuous, but in general it is false, because we may always choose a point t_0 that is not dyadic and define $\widetilde{f}(t) := f(t)$ for all $t \neq t_0$, and $\widetilde{f}(t_0) := f(t_0) + 1$. Since the set $(t^i_{pm})_{i,p,m}$ is countable, it is not even true that the coefficients of f determine the function Lebesgue-almost everywhere.

Littlewood-Paley notation. We will employ the notation from Littlewood-Paley theory. For $p \geq -1$ and $f \in C([0,1])$ we define

$$\Delta_p f := \sum_{m=0}^{2^p} f_{pm}\varphi_{pm} \qquad \text{and} \qquad S_p f := \sum_{q \leq p} \Delta_q f.$$

We will occasionally refer to $(\Delta_p f)$ as the Schauder blocks of f. Note that $\mathcal{C}^\alpha$ consists exactly of those $f = \sum_p \Delta_p f$ for which

$$\|2^{p\alpha}\|\Delta_p f\|_\infty\|_{\ell^\infty} < \infty.$$

4.2.2. Young integration and rough paths

Here we present the main concepts of Young integration and of rough path theory. The results presented in this section will not be applied in the remainder of this chapter, but we feel that it could be useful for the reader to be familiar with the basic concepts of rough paths, since it is the main inspiration for the constructions developed below.

Young's integral [You36] allows to define $\int f \mathrm{d}g$ for $f \in C^\alpha$, $g \in C^\beta$, and $\alpha + \beta > 1$. More precisely, let $f \in C^\alpha$ and $g \in C^\beta$ be given, let $t \in [0,1]$, and let $\pi = \{t_0, \dots, t_N\}$ be a partition of $[0,t]$, i.e. $0 = t_0 < t_1 < \dots < t_N = t$. Then it can be shown that the Riemann sums

$$\sum_{t_k \in \pi} f(t_k)(g(t_{k+1}) - g(t_k)) := \sum_{k=0}^{N-1} f(t_k)(g(t_{k+1}) - g(t_k))$$

converge as the mesh size $\max_{k=0,\dots,N-1} |t_{k+1} - t_k|$ tends to zero, and that the limit does not depend on the approximating sequence of partitions. We denote the limit by $\int_0^t f(s)\mathrm{d}g(s)$, and we define $\int_s^t f(r)\mathrm{d}g(r) := \int_0^t f(r)\mathrm{d}g(r) - \int_0^s f(r)\mathrm{d}g(r)$. The function $t \mapsto \int_0^t f(s)\mathrm{d}g(s)$ is uniquely characterized by the fact that

$$\left| \int_s^t f(r)\mathrm{d}g(r) - f(s)(g(t) - g(s)) \right| \lesssim |t-s|^{\alpha+\beta}\|f\|_\alpha\|g\|_\beta$$

for all $s, t \in [0,1]$. The condition $\alpha + \beta > 1$ is sharp, in the sense that there exist $f, g \in C^{1/2}$, and a sequence of partitions $(\pi_n)_{n\in\mathbb{N}}$ with mesh size going to zero, for which the Riemann sums $\sum_{t_k \in \pi_n} f(t_k)(g(t_{k+1}) - g(t_k))$ do not converge as n tends to ∞.

The condition $\alpha + \beta > 1$ excludes one of the most important examples: we would like to take g as a sample path of Brownian motion, and $f = F(g)$. Lyons' theory of rough paths [Lyo98] overcomes this restriction by stipulating the "existence" of basic integrals and by defining a large class of related integrals as their functionals. Here we present the approach of Gubinelli [Gub04].

Let $\alpha \in (1/3, 1)$ and assume that we are given two functions $v, w \in C^\alpha$, as well as an associated "Riemann integral" $I^{v,w}_{s,t} = \int_s^t v(r)\mathrm{d}w(r)$ that satisfies the estimate

$$|\Phi^{v,w}_{s,t}| := |I^{v,w}_{s,t} - v(s)w_{s,t}| \lesssim |t-s|^{2\alpha}. \tag{4.2}$$

The remainder $\Phi^{v,w}$ is often (incorrectly) called the *area* of v and w. This name has its origin in the fact that its antisymmetric part $1/2(\Phi^{v,w}_{s,t} - \Phi^{w,v}_{s,t})$ corresponds to the algebraic area spanned by the curve $((v(r), w(r)) : r \in [s,t])$ in the plane $\mathbb{R}^2$.

If $\alpha \leq 1/2$, then the integral $I^{v,w}$ cannot be constructed using Young's theory of integration, and also $I^{v,w}$ is not uniquely characterized by (4.2). But let us assume nonetheless that we are given such an integral $I^{v,w}$ satisfying (4.2). A function $f \in C^\alpha$ is *controlled* by $v \in C^\alpha$ if there exists $f^v \in C^\alpha$, such that for all $s, t \in [0,1]$

$$|f_{s,t} - f^v_s v_{s,t}| \lesssim |t-s|^{2\alpha}. \tag{4.3}$$

Proposition 4.2.4 ([Gub04], Theorem 1)**.** *Let $\alpha > 1/3$, let $v, w \in C^\alpha$, and let $I^{v,w}$ satisfy* (4.2). *Let f and g be controlled by v and w respectively, with derivatives f^v and g^w. Then there exists a unique function $I(f,g) = \int_0^\cdot f(s)\mathrm{d}g(s)$ that satisfies for all $s, t \in [0,1]$*

$$|I(f,g)_{s,t} - f(s)g_{s,t} - f^v(s)g^w(s)\Phi^{v,w}_{s,t}| \lesssim |t-s|^{3\alpha}.$$

If (π_n) is a sequence of partitions of $[0,t]$, with mesh size going to zero, then

$$I(f,g)(t) = \lim_{n\to\infty} \sum_{t_k \in \pi_n} \left(f(t_k)g_{t_k,t_{k+1}} + f^v_{t_k} g^w_{t_k} \Phi^{v,w}_{t_k,t_{k+1}} \right).$$

The integral $I(f,g)$ coincides with the Riemann-Stieltjes integral and with the Young integral, whenever these are defined. Moreover, the integral map is self-consistent, in the sense that if we consider v and w as controlled by themselves, with derivatives $v^v = w^w \equiv 1$, then $I(v,w) = I^{v,w}$.

The only remaining problem is the construction of the integral $I^{v,w}$. This is usually achieved with probabilistic arguments. If v and w are Brownian motions, then we can for example use Itô or Stratonovich integration to define $I^{v,w}$. Already in this simple example we see that the integral $I^{v,w}$ is not unique if v and w are outside of the Young regime.

It is possible to go beyond $\alpha > 1/3$ by stipulating the existence of higher order iterated integrals. For details see [Gub10] or any book on rough paths, such as [LQ02, LCL07, FV10b].

Note that the rough path integral is similar in spirit to Föllmer's pathwise Itô calculus, see Chapter 3, that stipulates the existence of the quadratic variation and uses this to give a pathwise construction of stochastic integrals.

4.3. Paradifferential calculus and Young integration

In this section we develop the basic tools that will be required for the rough path integral in terms of Schauder functions, and we study Young's integral and its different components.

4.3.1. Paradifferential calculus with Schauder functions

Here we introduce a "paradifferential calculus" in terms of Schauder functions. Paradifferential calculus is usually formulated in terms of Littlewood-Paley blocks, and was initiated by Bony [Bon81]. For details see Bahouri, Chemin, and Danchin [BCD11], or Chapter 5 below.

We will need to study the regularity of $\sum_{p,m} u_{pm}\varphi_{pm}$, where u_{pm} are functions and not constant coefficients. For this purpose we define the following space of sequences of functions.

Definition 4.3.1. If $(u_{pm})_{p\geq -1, 0\leq m\leq 2^p}$ is a family of affine functions of the form $u_{pm} : [t^0_{pm}, t^2_{pm}] \to \mathbb{R}^d$, where $u_{pm}(s) = a_{pm} + (s - t^0_{pm})b_{pm}$, then we define for $\alpha > 0$

$$\|(u_{pm})\|_{\mathcal{A}^\alpha} := \sup_{p,m} 2^{p\alpha} \|u_{pm}\|_\infty,$$

where it is understood that $\|u_{pm}\|_\infty := \max_{t\in[t^0_{pm}, t^2_{pm}]} |u_{pm}(t)|$. The space $\mathcal{A}^\alpha$ is then defined as

$$\begin{aligned}\mathcal{A}^\alpha &:= \mathcal{A}^\alpha(\mathbb{R}^d) \\ &:= \Big\{(u_{pm})_{p\geq -1, 0\leq m\leq 2^p} : u_{pm} \in C([t^0_{pm}, t^2_{pm}], \mathbb{R}^d) \text{ is affine and } \|(u_{pm})\|_{\mathcal{A}^\alpha} < \infty\Big\}.\end{aligned}$$

In Appendix E we prove the following regularity estimate:

Lemma 4.3.2. *Let $\alpha \in (0, 2)$ and let $(u_{pm}) \in \mathcal{A}^\alpha$. Then $\sum_{p,m} u_{pm}\varphi_{pm} \in \mathcal{C}^\alpha$, and*

$$\Big\|\sum_{p,m} u_{pm}\varphi_{pm}\Big\|_\alpha \lesssim \|(u_{pm})\|_{\mathcal{A}^\alpha}.$$

Before we get to paraproducts and paralinearization in terms of Schauder functions, let us show that $\mathcal{C}^\alpha$ is stable under the application of smooth functions.

Lemma 4.3.3. *Let $\alpha \in (0, 2)$ and let $v \in \mathcal{C}^\alpha(\mathbb{R}^d)$. Let $F \in C^{\lfloor\alpha+1\rfloor}_b(\mathbb{R}^d, \mathbb{R})$. Then $F(v) \in \mathcal{C}^\alpha$, and*

$$\|F(v)\|_\alpha \lesssim \|F\|_{C^{\lfloor\alpha+1\rfloor}_b} (1 + \|v\|_\alpha)^{\lfloor\alpha+1\rfloor}. \tag{4.4}$$

Proof. We have to estimate the coefficients $(F(v))_{pm} = 2^{-p}\langle \chi_{pm}, \mathrm{d}F(v)\rangle$. For $(p,m) = (-1,0)$ and $(p,m) = (0,0)$ we estimate

$$|(F(v))_{pm}| \lesssim \|F(v)\|_\infty \leq \|F\|_\infty.$$

For all other values of (p,m) we have $(F(v))_{pm} = (F(v))_{t^0_{pm},t^1_{pm}} - (F(v))_{t^1_{pm},t^2_{pm}}$. If $\alpha \in (0,1)$, then we apply a first order Taylor expansion with integral remainder, to obtain

$$\begin{aligned}(F(v))_{t^0_{pm},t^1_{pm}} - (F(v))_{t^1_{pm},t^2_{pm}} &= \sum_{|\eta|=1} \int_0^1 \partial^\eta F\left(v(t^0_{pm}) + r v_{t^0_{pm},t^1_{pm}}\right)(v_{t^0_{pm},t^1_{pm}})^\eta \mathrm{d}r \\ &\quad - \sum_{|\eta|=1} \int_0^1 \partial^\eta F\left(v(t^1_{pm}) + r v_{t^1_{pm},t^2_{pm}}\right)(v_{t^1_{pm},t^2_{pm}})^\eta \mathrm{d}r\end{aligned}$$

According to Lemma 4.2.2 we have $\|v\|_{C^\alpha} \simeq \|v\|_\alpha$, and therefore $|v_{t^0_{pm},t^1_{pm}}| \lesssim \|v\|_\alpha 2^{-p\alpha}$, which yields (4.4).

For $\alpha \in [1,2)$ we apply a second order Taylor expansion, which implies that

$$\begin{aligned}&(F(v))_{t^0_{pm},t^1_{pm}} - (F(v))_{t^1_{pm},t^2_{pm}} \\ &\quad = \sum_{|\eta|=1} \partial^\eta F(v(t^1_{pm}))(v_{t^0_{pm},t^1_{pm}})^\eta + R^1_{pm} - \sum_{|\eta|=1} \partial^\eta F(v(t^1_{pm}))(v_{t^1_{pm},t^2_{pm}})^\eta - R^2_{pm},\end{aligned} \tag{4.5}$$

where we use $\mathcal{C}^\alpha \subseteq \mathcal{C}^{1-\varepsilon} = C^{1-\varepsilon}$ to obtain that

$$|R^1_{pm}| + |R^2_{pm}| \lesssim \|F\|_{C^2_b}(|v_{t^0_{pm},t^1_{pm}}|^2 + |v_{t^0_{pm},t^1_{pm}}|^2) \lesssim \|F\|_{C^2_b}\|v\|^2_{1-\varepsilon} 2^{-2p(1-\varepsilon)}$$

for all $\varepsilon > 0$. Choose $\varepsilon > 0$ small enough so that $2 - 2\varepsilon > \alpha$. Then

$$\begin{aligned}|(F(v))_{t^0_{pm},t^1_{pm}} - (F(v))_{t^1_{pm},t^2_{pm}}| &\leq \left|\sum_{|\eta|=1} \partial^\eta F(v(t^1_{pm}))(v_{pm})^\eta\right| + |R^1_{pm}| + |R^2_{pm}| \\ &\lesssim 2^{-p\alpha}\|F\|_{C^2_b}\|v\|_\alpha(1 + \|v\|_\alpha).\end{aligned}$$

$\square$

Remark 4.3.4. Since v has compact support, it actually suffices if $F \in C^{\lfloor \alpha+1 \rfloor}$, without assuming that F and its partial derivatives are bounded. Of course then the estimate (4.4) would have to be adapted. For simplicity we only consider the case $F \in C^{\lfloor \alpha+1 \rfloor}_b$.

Let us define a paraproduct in terms of Schauder functions.

Lemma 4.3.5. *Let $\beta \in (0,2)$, let $v \in C([0,1], \mathcal{L}(\mathbb{R}^d, \mathbb{R}^n))$, and $w \in \mathcal{C}^\beta(\mathbb{R}^d)$. Then*

$$\pi_<(v,w) := \sum_{p=0}^\infty S_{p-1} v \Delta_p w \in \mathcal{C}^\beta(\mathbb{R}^n) \quad \text{and} \quad \|\pi_<(v,w)\|_\beta \lesssim \|v\|_\infty \|w\|_\beta. \tag{4.6}$$

Proof. We have

$$\pi_<(v,w) = \sum_{p,m} u_{pm}\varphi_{pm} \qquad \text{with} \qquad u_{pm} = (S_{p-1}v)|_{[t^0_{pm},t^2_{pm}]} w_{pm}.$$

For every (p,m), the function $(S_{p-1}v)|_{[t^0_{pm},t^2_{pm}]}$ is the linear interpolation of v between t^0_{pm} and t^2_{pm}. As $\|(S_{p-1}v)|_{[t^0_{pm},t^2_{pm}]} w_{pm}\|_\infty \le 2^{-p\beta}\|v\|_\infty\|w\|_\beta$, the statement follows from Lemma 4.3.2. □

Remark 4.3.6. If $v \in \mathcal{C}^\alpha$ and $w \in \mathcal{C}^\beta$, then we can decompose the product vw into three components, $vw = \pi_<(v,w) + \pi_>(v,w) + \pi_\circ(v,w)$, where

$$\pi_>(v,w) := \sum_p \Delta_p v S_{p-1} w, \qquad \|\pi_>(v,w)\|_\alpha \lesssim \|v\|_\alpha \|w\|_\infty, \qquad \text{and}$$
$$\pi_\circ(v,w) := \sum_p \Delta_p v \Delta_p w, \qquad \|\pi_\circ(v,w)\|_{\alpha+\beta} \lesssim \|v\|_\alpha \|w\|_\beta.$$

The estimate for $\pi_\circ$ only holds for $\alpha+\beta<2$, and it is easy to show. Since we will not use it, we omit the proof.

The paralinearization theorem in terms of Schauder functions is as follows.

Proposition 4.3.7. *Let $\alpha \in (0,1)$, let $v \in \mathcal{C}^\alpha(\mathbb{R}^d)$, and $F \in C^2_b(\mathbb{R}^d,\mathbb{R})$. Define*

$$\pi_<(\mathrm{D}F(v),v) := \sum_{|\eta|=1} \pi_<(\partial^\eta F(v), v^\eta).$$

Then $F(v) - \pi_<(\mathrm{D}F(v),v) \in \mathcal{C}^{2\alpha}$, and

$$\|F(v) - \pi_<(\mathrm{D}F(v),v)\|_{2\alpha} \lesssim \|F\|_{C^2_b}(1+\|v\|_\alpha)^2. \tag{4.7}$$

Proof. First note that $\|F(v)\|_\infty \le \|F\|_\infty$, which implies the estimate required for (4.7) if $(p,m)=(-1,0)$ or $(p,m)=(0,0)$. For all other values of (p,m) we apply a second order Taylor expansion to obtain

$$(F(v))_{pm} = \sum_{|\eta|=1} \partial^\eta F(v(t^1_{pm}))(v_{pm})^\eta + R_{pm} = \mathrm{D}F(v(t^1_{pm}))v_{pm} + R_{pm},$$

where $|R_{pm}| \lesssim \|F\|_{C^2_b} 2^{-2p\alpha}\|v\|^2_\alpha$. Therefore, $F(v) = \sum_{pm} \mathrm{D}F(v(t^1_{pm}))v_{pm}\varphi_{pm} + R$, with $R \in \mathcal{C}^{2\alpha}$ and $\|R\|_{2\alpha} \lesssim \|F\|_{C^2_b}\|v\|^2_\alpha$. Subtracting $\pi_<(\mathrm{D}F(v),v)$ gives

$$F(v) - \pi_<(\mathrm{D}F(v),v) = \sum_{pm} [\mathrm{D}F(v(t^1_{pm})) - (S_{p-1}\mathrm{D}F(v))|_{[t^0_{pm},t^2_{pm}]}] v_{pm}\varphi_{pm} + R.$$

$(S_{p-1}\mathrm{D}F(v))|_{[t^0_{pm},t^2_{pm}]}$ is the linear interpolation of $\mathrm{D}F(v)$ between t^0_{pm} and t^2_{pm}, so according to Lemma 4.3.2 it suffices to note that

$$\begin{aligned}\|[\mathrm{D}F(v(t^1_{pm})) - (S_{p-1}\mathrm{D}F(v))|_{[t^0_{pm},t^2_{pm}]}]v_{pm}\|_\infty &\lesssim 2^{-p\alpha}\|\mathrm{D}F(v)\|_\alpha 2^{-p\alpha}\|v\|_\alpha \\ &\lesssim 2^{-2p\alpha}\|F\|_{C^2_b}(1+\|v\|_\alpha)\|v\|_\alpha,\end{aligned}$$

where we used the estimate $\|\mathrm{D}F(v)\|_{C^\alpha} \simeq \|\mathrm{D}F(v)\|_\alpha$ and Lemma 4.3.3. □

Remark 4.3.8. The same proof shows that if f is controlled by v in the sense of Section 4.2.1, i.e. $f_{s,t} = f^v(s)v_{s,t} + R_{s,t}$ with $f^v \in \mathcal{C}^\alpha$ and $|R_{s,t}| \le \|R\|_{2\alpha}|t-s|^{2\alpha}$, then $f - \pi_<(f^v, v) \in \mathcal{C}^{2\alpha}$.

4.3.2. Young's integral and its different components

In this section we construct Young's integral using the Schauder expansion. If $v \in \mathcal{C}^\alpha$ and $w \in \mathcal{C}^\beta$, then we formally define

$$\int_0^\cdot v(s)\mathrm{d}w(s) := \sum_{p,m}\sum_{q,n} v_{pm}w_{qn}\int_0^\cdot \varphi_{pm}(s)\mathrm{d}\varphi_{qn}(s) = \sum_{p,q}\int_0^\cdot \Delta_p v(s)\mathrm{d}\Delta_q w(s).$$

We show that this definition makes sense provided that $\alpha+\beta>1$, and we identify three components of the integral that all have a different behavior. This will be our starting point towards a definition of controlled paths in our setting, and towards an extension of the integral beyond the Young regime.

In a first step, let us calculate the iterated integrals of Schauder functions.

Lemma 4.3.9. *Let $p > q \ge 0$. Then*

$$\int_0^1 \varphi_{pm}(s)\mathrm{d}\varphi_{qn}(s) = 2^{-p-2}\chi_{qn}(t^0_{pm}) \tag{4.8}$$

for all m, n. If $p = q$, then $\int_0^1 \varphi_{pm}(s)\mathrm{d}\varphi_{pn}(s) = 0$, except if $p = q = 0$, in which case the integral is bounded by 1. If $0 \le p < q$, then for all (m,n) we have

$$\int_0^1 \varphi_{pm}(s)\mathrm{d}\varphi_{qn}(s) = -2^{-q-2}\chi_{pm}\left(t^0_{qn}\right). \tag{4.9}$$

If $p = -1$, then the integral is bounded by 1.

Proof. The cases $p = q$ and $p = -1$ are easy, so let $p > q \ge 0$. Since $\chi_{qn} \equiv \chi_{qn}(t^0_{pm})$ on the support of φ_{pm}, we have

$$\int_0^1 \varphi_{pm}(s)\mathrm{d}\varphi_{qn}(s) = \chi_{qn}(t^0_{pm})\int_0^1 \varphi_{pm}(s)\mathrm{d}s = \chi_{qn}(t^0_{pm})2^{-p-2}.$$

If $0 \le p < q$, then integration by parts and (4.8) yield (4.9). □

Next we estimate the coefficients of iterated integrals in the Schauder basis.

Lemma 4.3.10. *Let $i, p \geq -1$, $q \geq 0$, $0 \leq j \leq 2^i$, $0 \leq m \leq 2^p$, $0 \leq n \leq 2^q$. Then*

$$\left|\left\langle \chi_{ij}, \mathrm{d}\left(\int_0^{\cdot} \varphi_{pm}\chi_{qn}\mathrm{d}s\right)\right\rangle\right| \leq 2^{-2(i\vee p\vee q)+i+p+q}, \tag{4.10}$$

except if $p < q = i$. In this case we only have the worse estimate

$$\left|\left\langle \chi_{ij}, \mathrm{d}\left(\int_0^{\cdot} \varphi_{pm}\chi_{qn}\mathrm{d}s\right)\right\rangle\right| \leq 2^{i}. \tag{4.11}$$

Proof. We have $\langle \chi_{-10}, \mathrm{d}(\int_0^\cdot \varphi_{pm}\chi_{qn}\mathrm{d}s)\rangle = 0$ for all (p, m) and (q, n). So let $i \geq 0$. If $i < p \vee q$, then χ_{ij} is constant on the support of $\varphi_{pm}\chi_{qn}$, and therefore

$$|\langle \chi_{ij}, \varphi_{pm}\chi_{qn}\rangle| \leq 2^i\, |\langle \varphi_{pm}, \chi_{qn}\rangle| \leq 2^{-2(p\vee q)+p+q+i} = 2^{-2(i\vee p\vee q)+p+q+i},$$

where we used Lemma 4.3.9.

Now let $i > q$. Then χ_{qn} is constant on the support of χ_{ij}, and therefore another application of Lemma 4.3.9 implies that

$$|\langle \chi_{ij}, \varphi_{pm}\chi_{qn}\rangle| \leq 2^q 2^{-2(p\vee i)+p+i} = 2^{-2(i\vee p\vee q)+p+q+i}.$$

The only remaining case is $i = q \geq p$. If $p = i = q$, then

$$|\langle \chi_{ij}, \varphi_{pm}\chi_{qn}\rangle| \leq 2^{2p} \int_0^1 \varphi_{pm}(s)\mathrm{d}s \leq 2^{2p}2^{-p} = 2^{-2(i\vee p\vee q)+p+q+i}.$$

Otherwise, if $i = q > p$, then

$$|\langle \chi_{ij}, \varphi_{pm}\chi_{qn}\rangle| \leq 2^{2i} \int_{t_{ij}^0}^{t_{ij}^2} \varphi_{pm}(s)\mathrm{d}s \leq 2^i \|\varphi_{pm}\|_\infty \leq 2^i.$$

□

We use this result to estimate the iterated integrals of Schauder blocks.

Corollary 4.3.11. *Let $i, p \geq -1$ and $q \geq 0$. Let $v \in C([0,1], \mathcal{L}(\mathbb{R}^d, \mathbb{R}^n))$ and $w \in C([0,1], \mathbb{R}^d)$. Then*

$$\left\|\Delta_i\left(\int_0^{\cdot} \Delta_p v(s)\mathrm{d}\Delta_q w(s)\right)\right\|_\infty \lesssim 2^{-(i\vee p\vee q)-i+p+q}\|\Delta_p v\|_\infty \|\Delta_q w\|_\infty, \tag{4.12}$$

except if $i = q > p$. In this case we only have the worse estimate

$$\left\|\Delta_i\left(\int_0^{\cdot} \Delta_p v(s)\mathrm{d}\Delta_q w(s)\right)\right\|_\infty \lesssim \|\Delta_p v\|_\infty \|\Delta_q w\|_\infty. \tag{4.13}$$

Proof. The case $i = -1$ is again easy, so let $i \geq 0$. We have

$$\Delta_i \left(\int_0^\cdot \Delta_p v(s) \mathrm{d}\Delta_q w(s) \right) = \sum_{j,m,n} v_{pm} w_{qn} \langle 2^{-i} \chi_{ij}, \varphi_{pm} \chi_{qn} \rangle \varphi_{ij}.$$

For fixed j, there are at most $2^{(i \vee p \vee q)-i}$ non-vanishing terms in the double sum. Furthermore, we have $|v_{pm}| \lesssim \|\Delta_p v\|_\infty$ and similarly for $|w_{qn}|$. Hence, we obtain from Lemma 4.3.10 that

$$\begin{aligned} \left\| \sum_{m,n} v_{pm} w_{qn} \langle 2^{-i} \chi_{ij}, \varphi_{pm} \chi_{qn} \rangle \varphi_{ij} \right\|_\infty &\lesssim 2^{(i \vee p \vee q)-i} \|\Delta_p v\|_\infty \|\Delta_q w\|_\infty 2^{-i} 2^{-2(i \vee p \vee q)+i+p+q} \\ &= 2^{-(i \vee p \vee q)-i+p+q} \|\Delta_p v\|_\infty \|\Delta_q w\|_\infty, \end{aligned}$$

except if $i = q > p$. In that case Lemma 4.3.10 yields

$$\begin{aligned} \left\| \sum_{m,n} v_{pm} w_{qn} \langle 2^{-i} \chi_{ij}, \varphi_{pm} \chi_{qn} \rangle \varphi_{ij} \right\|_\infty &\lesssim 2^{i-i} \|\Delta_p v\|_\infty \|\Delta_q w\|_\infty 2^{-i} 2^{i} \\ &= \|\Delta_p v\|_\infty \|\Delta_q w\|_\infty. \end{aligned}$$

□

Corollary 4.3.12. *Let $i, p, q \geq -1$. Let $v \in C([0,1], \mathcal{L}(\mathbb{R}^d, \mathbb{R}^n))$ and $w \in C([0,1], \mathbb{R}^d)$. Then for $p \vee q \leq i$ we have*

$$\|\Delta_i (\Delta_p v \Delta_q w)\|_\infty \lesssim 2^{-(i \vee p \vee q)-i+p+q} \|\Delta_p v\|_\infty \|\Delta_q w\|_\infty, \tag{4.14}$$

except if $i = q > p$ or $i = p > q$, in which case we only have the worse estimate

$$\|\Delta_i (\Delta_p v \Delta_q w)\|_\infty \lesssim \|\Delta_p v\|_\infty \|\Delta_q w\|_\infty. \tag{4.15}$$

If $p > i$ or $q > i$, then $\Delta_i(\Delta_p v \Delta_q w) \equiv 0$.

Proof. The case $p = -1$ or $q = -1$ is easy. Otherwise we apply integration by parts and note that the estimates (4.12) and (4.13) are symmetric in p and q.

If for example $p > i$, then $\Delta_p(v)(t^k_{ij}) = 0$ for all k, j, which implies that $\Delta_i(\Delta_p v \Delta_q w) = 0$. □

The estimates (4.12) and (4.13) allow us to identify different components of the integral $\int_0^\cdot v(s) \mathrm{d}w(s)$. More precisely, (4.13) indicates that the series $\sum_{p<q} \int_0^\cdot \Delta_p v(s) \mathrm{d}\Delta_q w(s)$ is rougher than the remainder $\sum_{p \geq q} \int_0^\cdot \Delta_p v(s) \mathrm{d}\Delta_q w(s)$. If we apply integration by parts to $\int_0^\cdot \Delta_p v(s) \mathrm{d}\Delta_q w(s)$, then we obtain

$$\sum_{p<q} \int_0^\cdot \Delta_p v(s) \mathrm{d}\Delta_q w(s) = \pi_<(v, w) - \sum_{p<q} \sum_{m,n} v_{pm} w_{qn} \int_0^\cdot \varphi_{qn}(s) \mathrm{d}\varphi_{pm}(s).$$

This motivates us to decompose the integral into three components, namely

$$\sum_{p,q} \int_0^{\cdot} \Delta_p v(s) \mathrm{d}\Delta_q w(s) = L(v,w) + S(v,w) + \pi_<(v,w).$$

Here L is defined as the antisymmetric "Lévy area" (we will justify the name below by showing that L is closely related to the Lévy area of certain dyadic martingales)

$$\begin{aligned} L(v,w) &:= \sum_{p>q} \sum_{m,n} (v_{pm} w_{qn} - v_{qn} w_{pm}) \int_0^{\cdot} \varphi_{pm} \mathrm{d}\varphi_{qn} \\ &= \sum_p \left(\int_0^{\cdot} \Delta_p v \mathrm{d} S_{p-1} w - \int_0^{\cdot} \mathrm{d}(S_{p-1} v) \Delta_p w \right). \end{aligned}$$

The symmetric part S is defined as

$$\begin{aligned} S(v,w) &:= \sum_{m,n \le 1} v_{0m} w_{0n} \int_0^{\cdot} \varphi_{0m} \mathrm{d}\varphi_{0n} + \sum_{p \ge 1} \sum_m v_{pm} w_{pm} \int_0^{\cdot} \varphi_{pm} \mathrm{d}\varphi_{pm} \\ &= \sum_{m,n \le 1} v_{0m} w_{0n} \int_0^{\cdot} \varphi_{0m} \mathrm{d}\varphi_{0n} + \frac{1}{2} \sum_{p \ge 1} \Delta_p v \Delta_p w, \end{aligned}$$

and $\pi_<$ is the paraproduct, as defined in (4.6). As we observed in Lemma 4.3.5, $\pi_<(v,w)$ is always well defined, and it inherits the regularity of w. Let us examine under which conditions S and L are defined, and how regular they are.

Lemma 4.3.13. *Let $\alpha, \beta \in (0,1)$ be such that $\alpha + \beta > 1$, and let $v \in \mathcal{C}^\alpha$ and $w \in \mathcal{C}^\beta$. Then $L(v,w)$ is well defined and in $\mathcal{C}^{\alpha+\beta}$, and moreover*

$$\|L(v,w)\|_{\alpha+\beta} \lesssim_{\alpha+\beta} \|v\|_\alpha \|w\|_\beta.$$

Proof. We only argue for $\sum_p \int_0^{\cdot} \Delta_p v \mathrm{d} S_{p-1} w$, because the term $- - \int_0^{\cdot} \mathrm{d}(S_{p-1} v) \Delta_p w$ can be treated with the same arguments. Corollary 4.3.11 (more precisely (4.12)) implies that

$$\begin{aligned} &\left\| \sum_p \Delta_i \left(\int_0^{\cdot} \Delta_p v \mathrm{d} S_{p-1} w \right) \right\|_\infty \\ &\quad \le \sum_{p \le i} \sum_{q < p} \left\| \Delta_i \left(\int_0^{\cdot} \Delta_p v \mathrm{d}\Delta_q w \right) \right\|_\infty + \sum_{p > i} \sum_{q < p} \left\| \Delta_i \left(\int_0^{\cdot} \Delta_p v \mathrm{d}\Delta_q w \right) \right\|_\infty \\ &\quad \le \left(\sum_{p \le i} \sum_{q < p} 2^{-2i+p+q} 2^{-p\alpha} \|v\|_\alpha 2^{-q\beta} \|w\|_\beta + \sum_{p > i} \sum_{q < p} 2^{-i+q} 2^{-p\alpha} \|v\|_\alpha 2^{-q\beta} \|w\|_\beta \right) \\ &\quad \lesssim_{\alpha+\beta} 2^{-i(\alpha+\beta)} \|v\|_\alpha \|w\|_\beta, \end{aligned}$$

where we used $1 - \alpha < 0$ and $1 - \beta < 0$ for both series, and for the second series we also used that $\alpha + \beta > 1$. $\square$

Unlike the Lévy area L, the symmetric part S is always well defined. It is also smooth.

Lemma 4.3.14. *Let $\alpha, \beta \in (0,1)$, and let $v \in \mathcal{C}^\alpha$ and $w \in \mathcal{C}^\beta$. Then $S(v,w) \in \mathcal{C}^{\alpha+\beta}$, and*

$$\|S(v,w)\|_{\alpha+\beta} \lesssim \|v\|_\alpha \|w\|_\beta.$$

Proof. This is shown using the same arguments as in the proof of Lemma 4.3.13. $\square$

In conclusion, the integral consists of three components. The Lévy area $L(v,w)$ is only defined if $\alpha+\beta > 1$, but then it is smooth. The symmetric part $S(v,w)$ is always defined and smooth. And the paraproduct $\pi_<(v,w)$ is always defined, but it is rougher than the other components. To summarize:

Theorem 4.3.15 (Young's integral)**.** *Let $\alpha, \beta \in (0,1)$ be such that $\alpha + \beta > 1$, and let $v \in \mathcal{C}^\alpha$ and $w \in \mathcal{C}^\beta$. Then*

$$I(v, \mathrm{d}w) := \sum_{p,q} \int_0^\cdot \Delta_p v \mathrm{d}\Delta_q w = L(v,w) + S(v,w) + \pi_<(v,w) \in \mathcal{C}^\beta$$

satisfies $\|I(v,\mathrm{d}w)\|_\beta \lesssim \|v\|_\alpha \|w\|_\beta$ and

$$\|I(v,\mathrm{d}w) - \pi_<(v,w)\|_{\alpha+\beta} \lesssim \|v\|_\alpha \|w\|_\beta. \tag{4.16}$$

Lévy area and dyadic martingales

Here we show that the Lévy area $L(v,w)(1)$ can be expressed in terms of the Lévy area of suitable dyadic martingales. To simplify notation, we assume that $v(0) = w(0) = 0$, so that we do not have to bother with the components v_{-10} and w_{-10}.

We define the filtration $(\mathcal{F}_n)_{n \ge 0}$ on $[0,1]$ by setting

$$\mathcal{F}_n = \sigma(\chi_{pm} : 0 \le p \le n, 0 \le m \le 2^p),$$

we set $\mathcal{F} = \bigvee_n \mathcal{F}_n$, and we consider the Lebesgue measure on $([0,1], \mathcal{F})$. On this space, the process $M_n = \sum_{p=0}^n \sum_{m=0}^{2^p} \chi_{pm}$, $n \in \mathbb{N}$, is a martingale. For any continuous function $v : [0,1] \to \mathbb{R}$ with $v(0) = 0$, the process

$$M_n^v = \sum_{p=0}^n \sum_{m=0}^{2^p} \langle 2^{-p} \chi_{pm}, \mathrm{d}v \rangle \chi_{pm} = \sum_{p=0}^n \sum_{m=0}^{2^p} v_{pm} \chi_{pm},$$

$n \in \mathbb{N}$, is a martingale transform of M, and therefore a martingale as well. Since it will be convenient later, we also define $\mathcal{F}_{-1} = \{\emptyset, [0,1]\}$ and $M_{-1}^v = 0$ for every v.

Assume now that v and w are continuous real-valued functions with $v(0) = w(0) = 0$, and that the Lévy area $L(v, w)(1)$ exists. Then it is given by

$$
\begin{aligned}
L(v,w)(1) &= \sum_{p=0}^{\infty}\sum_{q=0}^{p-1}\sum_{m,n}(v_{pm}w_{qn} - v_{qn}w_{pm})\int_0^1 \varphi_{pm}(s)\chi_{qn}(s)\mathrm{d}s \\
&= \sum_{p=0}^{\infty}\sum_{q=0}^{p-1}\sum_{m,n}(v_{pm}w_{qn} - v_{qn}w_{pm})\chi_{qn}(t^0_{pm})\int_0^1 \varphi_{pm}(s)\mathrm{d}s \\
&= \sum_{p=0}^{\infty}\sum_{q=0}^{p-1}\sum_{m,n}(v_{pm}w_{qn} - v_{qn}w_{pm})2^p\int_0^1 \chi_{qn}(s)1_{[t^0_{pm},t^2_{pm})}(s)\mathrm{d}s\langle\varphi_{pm},1\rangle \\
&= \sum_{p=0}^{\infty}\sum_{q=0}^{p-1}\sum_{m,n}(v_{pm}w_{qn} - v_{qn}w_{pm})2^{-p}\int_0^1 \chi_{qn}(s)\chi^2_{pm}(s)\mathrm{d}s2^{-p-2} \\
&= \sum_{p=0}^{\infty}\sum_{q=0}^{p-1}2^{-2p-2}\int_0^1\sum_{m,n}\sum_{m'}(v_{pm}w_{qn} - v_{qn}w_{pm})\chi_{qn}(s)\chi_{pm}(s)\chi_{pm'}(s)\mathrm{d}s,
\end{aligned}
$$

where in the fourth line we used that $\langle\varphi_{pm}, 1\rangle = 2^{-p-2}$ for all $p \geq 1$, and in the last step we used that χ_{pm} and $\chi_{pm'}$ have disjoint support for $m \neq m'$. Recall that the p–th Rademacher function (or "square wave") is defined for $p \geq 1$ as

$$r_p(t) := \sum_{m'=1}^{2^p} 2^{-p}\chi_{pm'}(t).$$

The martingale associated to the Rademacher functions is given by $R_0 := 0$ and $R_p := \sum_{k=1}^{p} r_k$ for $p \geq 1$. Let us write $\Delta M^v_p = M^v_p - M^v_{p-1}$ and similarly for M^w and R and all other discrete time processes that arise. This notation somewhat clashes with the expression $\Delta_p v$ for the dyadic blocks of v, but we will only use it in the following lines, where we do not directly work with dyadic blocks. The quadratic covariation of two dyadic martingales is defined as $[M, N]_n := \sum_{k=0}^{n}\Delta M_k\Delta N_k$, and the discrete time stochastic integral is defined as $(M \cdot N)_n := \sum_{k=0}^{n} M_{k-1}\Delta N_k$. Writing $E(\cdot)$ for the integral $\int_0^1 \cdot\mathrm{d}s$, we obtain

$$
\begin{aligned}
L(v,w)(1) &= \sum_{p=0}^{\infty}\sum_{q=0}^{p-1}2^{-p-2}E\left(\Delta M^v_p\Delta M^w_q\Delta R_p - \Delta M^v_q\Delta M^w_p\Delta R_p\right) \\
&= \sum_{p=0}^{\infty}2^{-p-2}E\left(\left(M^w_{p-1}\Delta M^v_p - M^v_{p-1}\Delta M^w_p\right)\Delta R_p\right) \\
&= \sum_{p=0}^{\infty}2^{-p-2}E\left(\Delta\left[M^w\cdot M^v - M^v\cdot M^w, R\right]_p\right).
\end{aligned}
$$

Hence, $L(v, w)(1)$ is closely related to the Lévy area $1/2(M^w \cdot M^v - M^v \cdot M^w)$ of the dyadic martingale (M^v, M^w).

4.4. Controlled paths and pathwise integration beyond Young

In this section we construct a rough path integral in terms of Schauder functions.

4.4.1. Controlled paths

We observed in Section 4.3 that for $\beta \in (0,1)$, for $w \in \mathcal{C}^\beta$, and for $F \in C_b^2$ we have $F(w) - \pi_<(\mathrm{D}F(w), w) \in \mathcal{C}^{2\beta}$. In Section 4.3.2 we observed that if moreover $v \in \mathcal{C}^\alpha$, where $\alpha + \beta > 1$, then the Young integral $I(v, \mathrm{d}w)$ satisfies $I(v, \mathrm{d}w) - \pi_<(v, w) \in \mathcal{C}^{\alpha+\beta}$. Hence, in both cases the function under consideration can be written as $\pi_<(f^w, w)$ for suitable f^w, plus a smooth remainder. We make this our definition of controlled paths:

Definition 4.4.1. Let $\alpha > 0$ and $v \in \mathcal{C}^\alpha(\mathbb{R}^d)$. We define

$$\begin{aligned}\mathcal{D}_v^\alpha &:= \mathcal{D}_v^\alpha(\mathbb{R}^n)\\ &:= \left\{ f \in \mathcal{C}^\alpha(\mathbb{R}^n) : \exists f^v \in \mathcal{C}^\alpha(\mathcal{L}(\mathbb{R}^d, \mathbb{R}^n)) \text{ s.t. } f^\sharp = f - \pi_<(f^v, v) \in \mathcal{C}^{2\alpha}(\mathbb{R}^n) \right\}.\end{aligned}$$

If $f \in \mathcal{D}_v^\alpha$, then f is called *controlled* by v. The function f^v is called the *derivative* of f with respect to v. We equip $\mathcal{D}_v^\alpha$ with the norm

$$\|f\|_{v,\alpha} := \|f\|_\alpha + \|f^v\|_\alpha + \|f^\sharp\|_{2\alpha}.$$

Remark 4.4.2. In general the derivative f^v is not uniquely determined by f and v. For example, if $v \in \mathcal{C}^{2\alpha}$, then $0 \in \mathcal{D}_v^\alpha$, and every $f^v \in \mathcal{C}^\alpha$ can be taken as its derivative. So the correct definition would be $(f, f^v) \in \mathcal{D}_v^\alpha$, and $\|(f, f^v)\|_{v,\alpha} = \|f\|_\alpha + \|f^v\|_\alpha + \|f^\sharp\|_{2\alpha}$. But usually there will be no confusion about the derivative that we have in mind, and therefore we will continue writing $f \in \mathcal{D}_v^\alpha$ and $\|f\|_{v,\alpha}$.

Example 4.4.3. Let $\alpha \in (0,1)$ and $v \in \mathcal{C}^\alpha$. Then $F(v) \in \mathcal{D}_v^\alpha$ for every $F \in C_b^2$, with derivative $\mathrm{D}F(v)$. This follows from Proposition 4.3.7.

Example 4.4.4. Let $\alpha \in (1/2, 1)$ and $v, w \in \mathcal{C}^\alpha$. Then the Young integral $I(v, \mathrm{d}w)$ is in $\mathcal{D}_w^\alpha$, with derivative v. This follows from (4.16).

The following lemma relates our notion of controlled paths to the classical one. This will allow us to simplify some of the proofs below.

Lemma 4.4.5. *If $\alpha \in (0, 1/2)$ and $f, v \in \mathcal{C}^\alpha$, then*

$$|\pi_<(f,v)(t) - \pi_<(f,v)(s) - f(s)(v(t) - v(s))| \lesssim \|f\|_\alpha \|v\|_\alpha |t-s|^{2\alpha}.$$

As a consequence, $f \in \mathcal{D}_v^\alpha$ if and only if for all $0 \le s < t \le 1$

$$f_{s,t} = f^v(s) v_{s,t} + R_{s,t}$$

for some $f^v \in C^\alpha$ and R that satisfies $|R_{s,t}| \lesssim |t-s|^{2\alpha}$. In other words, $f \in \mathcal{D}_v^\alpha$ if and only if f is controlled by v in the sense of Gubinelli [Gub04] (see Section 4.2.2).

Proof. Let $s, t \in [0,1]$ and let i be such that $2^{-i} \le |t-s| < 2^{-i+1}$. Note that

$$|(\Delta_p f)_{s,t}| \le \max_{m=0,\dots,2^p} |f_{pm}| 2^p |t-s| \le 2^{p(1-\alpha)} |t-s| \|f\|_\alpha.$$

for all $p \ge -1$ and $f \in C^\alpha$. This implies for $p \le i$

$$\begin{aligned}
|(S_{p-1} f \Delta_p v)_{s,t} - f(s)(\Delta_p v)_{s,t}| &\le |(S_{p-1} f)_{s,t} \Delta_p v(t)| + |(S_{p-1} f(s) - f(s))(\Delta_p v)_{s,t}| \\
&\lesssim \sum_{q<p} |(\Delta_q f)_{s,t}| \|\Delta_p v\|_\infty + \|S_{p-1} f - f\|_\infty |(\Delta_p v)_{s,t}| \\
&\lesssim \sum_{q<p} 2^{q(1-\alpha)} |t-s| \|f\|_\alpha 2^{-p\alpha} \|v\|_\alpha + 2^{-p\alpha} \|f\|_\alpha 2^{p(1-\alpha)} |t-s| \|v\|_\alpha \\
&\lesssim |t-s| 2^{p(1-2\alpha)} \|f\|_\alpha \|v\|_\alpha.
\end{aligned}$$

For $p > i$ we obtain

$$\begin{aligned}
&|(S_{p-1} f \Delta_p v)_{s,t} - f(s)(\Delta_p v)_{s,t}| \\
&\quad \lesssim \sum_{q \le i} |(\Delta_q f)_{s,t}| \|\Delta_p v\|_\infty + \sum_{q=i+1}^{p-1} \|\Delta_q f\|_\infty \|\Delta_p v\|_\infty + \|S_{p-1} f - f\|_\infty \|\Delta_p v\|_\infty \\
&\quad \lesssim \left(\sum_{q \le i} 2^{q(1-\alpha)} |t-s| 2^{-p\alpha} + \sum_{q=i+1}^{p-1} 2^{-q\alpha} 2^{-p\alpha} + 2^{-2p\alpha} \right) \|f\|_\alpha \|v\|_\alpha \\
&\quad \lesssim \big(2^{i(1-\alpha)} |t-s| 2^{-p\alpha} + 2^{-i\alpha} 2^{-p\alpha} + 2^{-2p\alpha}\big) \|f\|_\alpha \|v\|_\alpha \\
&\quad \simeq \big(|t-s|^\alpha 2^{-p\alpha} + 2^{-2p\alpha}\big) \|f\|_\alpha \|v\|_\alpha,
\end{aligned}$$

where we used that $|t-s| \simeq 2^{-i}$. We combine the two estimates to obtain

$$\begin{aligned}
&|\pi_<(f,v)(t) - \pi_<(f,v)(s) - f(s)(v(t) - v(s))| \\
&\quad \le \sum_{p \le i} |(S_{p-1} f \Delta_p v)_{s,t} - f(s)(\Delta_p v)_{s,t}| + \sum_{p>i} |(S_{p-1} f \Delta_p v)_{s,t} - f(s)(\Delta_p v)_{s,t}| \\
&\quad \lesssim \sum_{p \le i} |t-s| 2^{p(1-2\alpha)} \|f\|_\alpha \|v\|_\alpha + \sum_{p>i} \big(|t-s|^\alpha 2^{-p\alpha} + 2^{-2p\alpha}\big) \|f\|_\alpha \|v\|_\alpha \\
&\quad \lesssim |t-s|^{2\alpha} \|f\|_\alpha \|v\|_\alpha,
\end{aligned}$$

where we used that $1 - 2\alpha > 0$, and also that $2^{-i} \simeq |t-s|$. This implies that every $f \in \mathcal{D}_v^\alpha$ is controlled in the sense of Gubinelli [Gub04]. The opposite inclusion is the content of Remark 4.3.8. □

This connection between $\mathcal{D}_v^\alpha$ and the controlled paths of [Gub04] allows us to easily obtain several useful properties of $\mathcal{D}_v^\alpha$.

Corollary 4.4.6. *Let $\alpha \in (0, 1/2)$ and let $f \in \mathcal{D}_v^\alpha$ with derivative f^v. Let $F \in C_b^2$. Then*

$F(f) \in \mathcal{D}^\alpha_v$ with derivative $\mathrm{D}F(f)f^v$, and

$$\|F(f)\|_{v,\alpha} \lesssim \|F\|_{C^2_b}(1+\|v\|_\alpha)(1+\|f\|_{v,\alpha})^2.$$

Proof. We have

$$\|F(f)\|_{v,\alpha} = \|F(f)\|_\alpha + \|\mathrm{D}F(f)f^v\|_\alpha + \|F(f) - \pi_<(\mathrm{D}F(f)f^v, v)\|_{2\alpha}.$$

The first term on the right hand side can be estimated with Lemma 4.3.3. For the second and third term we use that $\|\cdot\|_\alpha \simeq \|\cdot\|_{C^\alpha}$, i.e. we work with the classical Hölder norm. We apply a Taylor expansion to $\mathrm{D}F(f)$ to obtain

$$\|\mathrm{D}F(f)f^v\|_{C^\alpha} \lesssim \|\mathrm{D}F\|_{C^1_b}(1+\|f\|_{C^\alpha})\|f^v\|_{C^\alpha}.$$

For the remainder we have

$$\|F(f) - \pi_<(\mathrm{D}F(f)f^v, v)\|_\infty \lesssim \|F\|_\infty + \|\pi_<(\mathrm{D}F(f)f^v, v)\|_\infty \lesssim \|F\|_{C^1_b}(1+\|f^v\|_\infty\|v\|_\alpha),$$

where we applied Lemma 4.3.5 to the second term. Moreover, for $0 \le s < t \le 1$, a first order Taylor expansion yields

$$\begin{aligned}
|(F(f))_{s,t} - \mathrm{D}F(f(s))f^v(s)v_{s,t}| &\lesssim |\mathrm{D}F(f(s))f_{s,t} - \mathrm{D}F(f(s))f^v(s)v_{s,t}| \\
&\qquad + \|F\|_{C^2_b}\|f\|_\alpha|t-s|^{2\alpha} \\
&\lesssim \|\mathrm{D}F\|_\infty|f_{s,t} - (\pi_<(f^v,v))_{s,t}| \\
&\qquad + \|\mathrm{D}F\|_\infty|(\pi_<(f^v,v))_{s,t} - f^v(s)v_{s,t}|) \\
&\qquad + \|F\|_{C^2_b}\|f\|_\alpha|t-s|^{2\alpha} \\
&\lesssim \|\mathrm{D}F\|_\infty\|f\|_{v,\alpha}(1+\|v\|_\alpha)|t-s|^{2\alpha} \\
&\qquad + \|F\|_{C^2_b}\|f\|_\alpha|t-s|^{2\alpha},
\end{aligned}$$

where we applied Lemma 4.4.5 in the last step. This shows that $F(f)$ is controlled in the sense of [Gub04]. The claim now follows by another application of Lemma 4.4.5. □

The space of controlled paths is an algebra:

Corollary 4.4.7. *Let $\alpha \in (0,1/2)$, let $v \in \mathcal{C}^\alpha(\mathbb{R}^d)$, and $f, g \in \mathcal{D}^\alpha_v(\mathbb{R})$, with derivatives f^v and g^v respectively. Then $fg \in \mathcal{D}^\alpha_v(\mathbb{R})$, with derivative $f^v g + f g^v$, and $\|fg\|_{v,\alpha} \lesssim \|f\|_{v,\alpha}\|g\|_{v,\alpha}(1+\|v\|_\alpha)$.*

Proof. We only concentrate on $(fg)^\sharp$. For this term it suffices to note that

$$\begin{aligned}
|(fg)_{s,t} - f^v(s)g(s)v_{s,t} - f(s)g^v(s)v_{s,t}| &\le |g(s)(f_{s,t} - f^v(s)v_{s,t})| \\
&\qquad + |f(s)(g_{s,t} - g^v(s)v_{s,t}) + f_{s,t}g_{s,t}| \\
&\lesssim \|g\|_\infty\|f\|_{v,\alpha}(1+\|v\|_\alpha)|t-s|^{2\alpha} \\
&\qquad + \|f\|_\infty\|g\|_{v,\alpha}(1+\|v\|_\alpha)|t-s|^{2\alpha},
\end{aligned}$$

where we applied Lemma 4.4.5 in the second step. Another application of Lemma 4.4.5 now completes the argument. □

Controlled paths satisfy the following transitivity condition.

Corollary 4.4.8. *Let $\alpha \in (0, 1/2)$. Let $w \in \mathcal{C}^\alpha(\mathbb{R}^m)$, let $v \in \mathcal{D}^\alpha_w(\mathbb{R}^n)$ with derivative $v^w \in \mathcal{C}^\alpha(\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n))$, and let $f \in \mathcal{D}^\alpha_v(\mathbb{R}^d)$ with derivative $f^v \in \mathcal{C}^\alpha(\mathcal{L}(\mathbb{R}^n, \mathbb{R}^d))$. Then $f \in \mathcal{D}^\alpha_w$, with derivative $f^w = f^v v^w \in \mathcal{C}^\alpha(\mathcal{L}(\mathbb{R}^m, \mathbb{R}^d))$, and*

$$\|f\|_{w,\alpha} \lesssim \|f\|_{v,\alpha} + \|f^v\|_\alpha \|v\|_{w,\alpha}(1 + \|w\|_\alpha).$$

Proof. Again we only argue for the remainder term. It suffices to note that

$$|f_{s,t} - f^v(s) v^w(s) w_{s,t}| \le |f_{s,t} - f^v(s) v_{s,t}| + |f^v(s)(v_{s,t} - v^w(s) w_{s,t})|$$

and to apply Lemma 4.4.5. □

Remark 4.4.9. Corollaries 4.4.6 – 4.4.8 hold in fact for $\alpha \in (0, 1)$, and can be shown using our definition of controlled paths rather than the equivalent characterization of Lemma 4.4.5. But then the proofs are longer. Since we will only need these results for $\alpha < 1/2$, we decided to give the short proofs based on Lemma 4.4.5.

We will need that if f is controlled by v, then $S_N f$ is controlled by $S_N v$.

Lemma 4.4.10. *Let $\alpha \in (0, 1)$, let $v \in \mathcal{C}^\alpha$, and let $f \in \mathcal{D}^\alpha_v$ with derivative f^v. Let $N \in \mathbb{N}$. Then $S_N f \in \mathcal{D}^\alpha_{S_N v}$ with derivative f^v, and for all $\varepsilon \in [0, \alpha)$ we have*

$$\|S_N f - f\|_{\alpha - \varepsilon} + \|(S_N f)^\sharp - f^\sharp\|_{2\alpha - \varepsilon} \lesssim 2^{-N\varepsilon} \|f\|_{v,\alpha}(1 + \|v\|_\alpha).$$

In particular, $\|S_N f\|_{S_N v, \alpha} \lesssim \|f\|_{v,\alpha}(1 + \|v\|_\alpha)$.

Proof. The estimate for $\|S_N f - f\|_{\alpha-\varepsilon}$ is straightforward, so let us concentrate on $\|(S_N f)^\sharp - f^\sharp\|_{2\alpha-\varepsilon}$. We have

$$\|(S_N f)^\sharp - f^\sharp\|_{2\alpha-\varepsilon} \le \|S_N(f^\sharp) - f^\sharp\|_{2\alpha-\varepsilon} + \|\pi_<(f^v, S_N v) - S_N(\pi_<(f^v, v))\|_{2\alpha-\varepsilon}.$$

It is easy to see that $\|S_N(f^\sharp) - f^\sharp\|_{2\alpha-\varepsilon} \lesssim 2^{-N\varepsilon}\|f^\sharp\|_{2\alpha}$, and therefore it only remains to estimate the second addend. Recall that $\Delta_i(\Delta_p f^v \Delta_q v) = 0$ for $q > i$ or $p > i$, which leads to

$$\begin{aligned}
&\pi_<(f^v, S_N v) - S_N(\pi_<(f^v, v)) \\
&\qquad = \sum_{i=-1}^{\infty} \sum_{q=-1}^{N} \sum_{p=-1}^{q-1} \Delta_i(\Delta_p f^v \Delta_q v) - \sum_{i=-1}^{N} \sum_{q=-1}^{N} \sum_{p=-1}^{q-1} \Delta_i(\Delta_p f^v \Delta_q v) \\
&\qquad = \sum_{i=N+1}^{\infty} \sum_{q=-1}^{N} \sum_{p=-1}^{q-1} \Delta_i(\Delta_p f^v \Delta_q v).
\end{aligned}$$

Now let $n \in \mathbb{N}$. Then we obtain from Corollary 4.3.12 that

$$
\begin{aligned}
&\|\Delta_n[\pi_<(f^v, S_N v) - S_N(\pi_<(f^v, v))]\|_\infty \\
&\qquad \lesssim 1_{[N+1,\infty)}(n) \sum_{q=-1}^{N} \sum_{p=-1}^{q-1} 2^{-2n+p(1-\alpha)+q(1-\alpha)} \|f^v\|_\alpha \|v\|_\alpha \\
&\qquad \simeq 1_{[N+1,\infty)}(n) 2^{-2n} 2^{N(2-2\alpha)} \|f^v\|_\alpha \|v\|_\alpha \le 1_{[N+1,\infty)}(n) 2^{-2\alpha n} \|f^v\|_\alpha \|v\|_\alpha,
\end{aligned}
$$

where we used that $2 - 2\alpha > 0$, and from where the claimed estimate follows. □

4.4.2. A basic commutator estimate

Here we prove a basic commutator estimate, which will be the main ingredient needed for constructing the integral $I(f, \mathrm{d}g)$, where f is controlled by v, and g is controlled by w, and where we assume that the integral $I(v, \mathrm{d}w)$ exists.

Proposition 4.4.11. *Let $\alpha, \beta, \gamma \in (0,1)$, and assume that $\alpha + \beta + \gamma > 1$. We also assume that $\beta + \gamma \neq 1$ and $\alpha + \beta + \gamma \neq 2$. Let $f \in \mathcal{C}^\alpha$, $v \in \mathcal{C}^\beta$, and $w \in \mathcal{C}^\gamma$. Then the "commutator"*

$$
\begin{aligned}
R(f, v, w) &:= L(\pi_<(f, v), w) - I(f, \mathrm{d}L(v, w)) \qquad (4.17)\\
&:= \lim_{N\to\infty} [L(S_N(\pi_<(f, v)), S_N w) - I(f, \mathrm{d}L(S_N v, S_N w))] \\
&= \lim_{N\to\infty} \sum_{p=-1}^{N} \sum_{q=-1}^{p-1} \Bigg[\int_0^\cdot \Delta_p(\pi_<(f, v))(s) \mathrm{d}\Delta_q w(s) - \int_0^\cdot \mathrm{d}(\Delta_q(\pi_<(f, v)))(s) \Delta_p w(s) \\
&\qquad - \Big(\int_0^\cdot f(s) \Delta_p v(s) \mathrm{d}\Delta_q w(s) - \int_0^\cdot f(s) \mathrm{d}(\Delta_q v)(s) \Delta_p w(s) \Big) \Bigg]
\end{aligned}
$$

converges in $\mathcal{C}^{\alpha+\beta+\gamma-\varepsilon}$ for all $\varepsilon > 0$. Moreover

$$
\|R(f, v, w)\|_{\alpha+\beta+\gamma} \lesssim \|f\|_\alpha \|v\|_\beta \|w\|_\gamma.
$$

Proof. We only argue for the difference of the positive terms in (4.17), i.e. for

$$
X_N := \sum_{p=-1}^{N} \sum_{q=-1}^{p-1} \left[\int_0^\cdot \Delta_p(\pi_<(f, v))(s) \mathrm{d}\Delta_q w(s) - \int_0^\cdot f(s) \Delta_p v(s) \mathrm{d}\Delta_q w(s) \right]. \qquad (4.18)
$$

The difference of the negative terms in (4.17) can be handled with the same arguments. First we prove that X_N converges uniformly, then we show that $\|X_N\|_{\alpha+\beta+\gamma}$ stays uniformly bounded. This will imply the desired result, because it is easy to see that bounded sets in $\mathcal{C}^{\alpha+\beta+\gamma}$ are relatively compact in $\mathcal{C}^{\alpha+\beta+\gamma-\varepsilon}$ for all $\varepsilon > 0$.

To prove uniform convergence, note that

$$\begin{aligned} X_N - X_{N-1} &= \sum_{q=-1}^{N-1} \left[\int_0^\cdot \Delta_N(\pi_<(f,v))(s)\mathrm{d}\Delta_q w(s) - \int_0^\cdot f(s)\Delta_N v(s)\mathrm{d}\Delta_q w(s) \right] \\ &= \sum_{q=-1}^{N-1} \Bigg[\sum_{j=-1}^{N} \sum_{i=-1}^{j-1} \int_0^\cdot \Delta_N(\Delta_i f \Delta_j v)(s)\mathrm{d}\Delta_q w(s) \\ &\qquad - \sum_{j=N}^{\infty} \sum_{i=-1}^{j} \int_0^\cdot \Delta_j(\Delta_i f \Delta_N v)(s)\mathrm{d}\Delta_q w(s) \Bigg], \end{aligned} \tag{4.19}$$

where for the second term it is possible to take the infinite sum over j outside of the integral, because $\sum_j \Delta_j g$ converges uniformly to g for every continuous function g, and because $\Delta_q w$ is a finite variation path. The restrictions of the range of summation are justified because $\Delta_N(\Delta_i f \Delta_j v) = 0$ for $i > N$ or $j > N$.

Only very few terms in (4.19) actually cancel. Nonetheless the cancellations are crucial, because they eliminate most terms for which we only have the worse estimate (4.15) in Corollary 4.3.12. We obtain

$$\begin{aligned} X_N - X_{N-1} &= \sum_{q=-1}^{N-1} \sum_{j=-1}^{N-1} \sum_{i=-1}^{j-1} \int_0^\cdot \Delta_N(\Delta_i f \Delta_j v)(s)\mathrm{d}\Delta_q w(s) \\ &\quad - \sum_{q=-1}^{N-1} \int_0^\cdot \Delta_N(\Delta_N f \Delta_N v)(s)\mathrm{d}\Delta_q w(s) \\ &\quad - \sum_{q=-1}^{N-1} \sum_{j=N+1}^{\infty} \sum_{i=-1}^{j-1} \int_0^\cdot \Delta_j(\Delta_i f \Delta_N v)(s)\mathrm{d}\Delta_q w(s) \\ &\quad - \sum_{q=-1}^{N-1} \sum_{j=N+1}^{\infty} \int_0^\cdot \Delta_j(\Delta_j f \Delta_N v)(s)\mathrm{d}\Delta_q w(s). \end{aligned} \tag{4.20}$$

Note that $\|\partial_t \Delta_q w\|_\infty \lesssim 2^q \|\Delta_q w\|_\infty$. Hence, an application of Corollary 4.3.12, where we use (4.14) for the first three terms and (4.15) for the fourth term, yields

$$\begin{aligned} \|X_N - X_{N-1}\|_\infty &\lesssim \|f\|_\alpha \|v\|_\beta \|w\|_\gamma \Bigg[\sum_{q=-1}^{N-1} \sum_{j=-1}^{N-1} \sum_{i=-1}^{j-1} 2^{-2N+i+j} 2^{-i\alpha} 2^{-j\beta} 2^{q(1-\gamma)} \\ &\quad + \sum_{q=-1}^{N-1} 2^{-N(\alpha+\beta)} 2^{q(1-\gamma)} + \sum_{q=-1}^{N-1} \sum_{j=N+1}^{\infty} \sum_{i=-1}^{j-1} 2^{-2j+i+N} 2^{-i\alpha} 2^{-N\beta} 2^{q(1-\gamma)} \\ &\quad + \sum_{q=-1}^{N-1} \sum_{j=N+1}^{\infty} 2^{-j\alpha} 2^{-N\beta} 2^{q(1-\gamma)} \Bigg]. \end{aligned}$$

Now we use that $\alpha, \beta, \gamma < 1$, to conclude that

$$\|X_N - X_{N-1}\|_\infty \lesssim \|f\|_\alpha \|v\|_\beta \|w\|_\gamma 2^{-N(\alpha+\beta+\gamma-1)}. \tag{4.21}$$

Since $\alpha + \beta + \gamma > 1$, this is summable in N, and therefore (X_N) converges uniformly.

Next we need to show that $\|X_N\|_{\alpha+\beta+\gamma} \lesssim \|f\|_\alpha \|v\|_\beta \|w\|_\gamma$ for all N. Similarly to (4.20) we obtain for $n \in \mathbb{N}$ that

$$\Delta_n X_N = \sum_{p \le N} \sum_{q<p} \Delta_n \Bigg[\sum_{j<p} \sum_{i<j} \int_0^\cdot \Delta_p(\Delta_i f \Delta_j v)(s) \mathrm{d}\Delta_q w(s) - \int_0^\cdot \Delta_p(\Delta_p f \Delta_p v)(s) \mathrm{d}\Delta_q w(s)$$
$$- \sum_{j>p} \sum_{i \le j} \int_0^\cdot \Delta_j(\Delta_i f \Delta_p v)(s) \mathrm{d}\Delta_q w(s) \Bigg],$$

and therefore by Corollary 4.3.11

$$\|\Delta_n X_N\|_\infty \lesssim \sum_p \sum_{q<p} \Bigg[\sum_{j<p} \sum_{i<j} 2^{-(n\vee p)-n+p+q} \|\Delta_p(\Delta_i f \Delta_j v)\|_\infty \|\Delta_q w\|_\infty$$
$$+ 2^{-(n\vee p)-n+p+q} \|\Delta_p(\Delta_p f \Delta_p v)\|_\infty \|\Delta_q w\|_\infty$$
$$+ \sum_{j>p} \sum_{i\le j} 2^{-(n\vee j)-n+j+q} \|\Delta_j(\Delta_i f \Delta_p v)\|_\infty \|\Delta_q w\|_\infty \Bigg].$$

We apply Corollary 4.3.12, where for the last addend we distinguish the cases $i < j$ and $i = j$. Moreover, we use that $1 - \gamma > 0$. This leads to

$$\|\Delta_n X_N\|_\infty \lesssim \|f\|_\alpha \|v\|_\beta \|w\|_\gamma \sum_p 2^{p(1-\gamma)} \Bigg[\sum_{j<p} \sum_{i<j} 2^{-(n\vee p)-n+p} 2^{-2p} 2^{i(1-\alpha)} 2^{j(1-\beta)}$$
$$+ 2^{-(n\vee p)-n+p} 2^{-p\alpha} 2^{-p\beta}$$
$$+ \sum_{j>p} \sum_{i<j} 2^{-(n\vee j)-n+j} 2^{-2j+i(1-\alpha)+p(1-\beta)}$$
$$+ \sum_{j>p} 2^{-(n\vee j)-n+j} 2^{-j\alpha-p\beta} \Bigg].$$

Now a straightforward calculation, using $\alpha, \beta < 1$, yields

$$\|\Delta_n X_N\|_\infty \lesssim \|f\|_\alpha \|v\|_\beta \|w\|_\gamma \sum_p \Bigg[2^{-(n\vee p)-n} 2^{p(2-\alpha-\beta-\gamma)} + 2^{-(n\vee p)-n} 2^{p(2-\alpha-\beta-\gamma)}$$
$$+ 2^{-(n\vee p)-n} 2^{p(2-\alpha-\beta-\gamma)} + 2^{p(1-\beta-\gamma)} 2^{-n} 2^{-\alpha(n\vee p)} \Bigg].$$

Hence, we have

$$\|\Delta_n X_N\|_\infty \lesssim \|f\|_\alpha \|v\|_\beta \|w\|_\gamma \left[\sum_{p=-1}^{n} \left[2^{-2n} 2^{p(2-\alpha-\beta-\gamma)} + 2^{p(1-\beta-\gamma)} 2^{-n(1+\alpha)} \right] + \sum_{p=n+1}^{\infty} 2^{-n} 2^{p(1-\alpha-\beta-\gamma)} \right].$$

Now we distinguish the cases $\alpha+\beta+\gamma < 2$ and $\alpha+\beta+\gamma > 2$, as well as $\beta+\gamma < 1$ and $\beta+\gamma > 1$, and we use that $1-\alpha-\beta-\gamma < 0$, to conclude that

$$\|\Delta_n X_N\|_\infty \lesssim \|f\|_\alpha \|v\|_\beta \|w\|_\gamma 2^{-n(\alpha+\beta+\gamma)},$$

which completes the proof. □

Remark 4.4.12. If $\beta+\gamma$ or $\alpha+\beta+\gamma$ happen to be integers, then we can apply Proposition 4.4.11 with $\beta-\varepsilon$ to obtain that $R(f,v,w) \in \mathcal{C}^{\alpha+\beta+\gamma-\varepsilon}$ for every sufficiently small $\varepsilon > 0$.

For later reference, we collect the following result from the proof of Proposition 4.4.11:

Lemma 4.4.13. *Let $\alpha, \beta, \gamma, f, v, w$ be as in Proposition 4.4.11. Then*

$$\|R(f,v,w) - L(S_N(\pi_<(f,v)), S_N w) - I(f, \mathrm{d}L(S_N v, S_N w))\|_\infty \lesssim 2^{-N(\alpha+\beta+\gamma-1)} \|f\|_\alpha \|v\|_\beta \|w\|_\gamma.$$

Proof. This follows by summing up (4.21) over N. □

Corollary 4.4.14. *Let $\alpha, \beta, \gamma \in (0,1)$, and assume that $\alpha+\beta+\gamma > 1$. We also assume that $\beta+\gamma \neq 1$ and $\alpha+\beta+\gamma \neq 2$. Let $f \in \mathcal{C}^\alpha(\mathcal{L}(\mathbb{R}^d,\mathbb{R}))$, $g \in \mathcal{C}^\alpha(\mathcal{L}(\mathbb{R}^n,\mathbb{R}))$, $v \in \mathcal{C}^\beta(\mathbb{R}^d)$, and $w \in \mathcal{C}^\gamma(\mathbb{R}^n)$. Then*

$$\begin{aligned}
\widetilde{R}(f,g,v,w) &:= L(\pi_<(f,v), \pi_<(g,w)) - I(fg, \mathrm{d}L(v,w)) \\
&:= \sum_{k=1}^{d} \sum_{\ell=1}^{n} [L(\pi_<(f^k,v^k), \pi_<(g^\ell,w^\ell)) - I(f^k g^\ell, \mathrm{d}L(v^k,w^\ell))] \\
&:= \sum_{k=1}^{d} \sum_{\ell=1}^{n} \lim_{N\to\infty} [L(S_N(\pi_<(f^k,v^k)), S_N(\pi_<(g^\ell,w^\ell))) - I(f^k g^\ell, \mathrm{d}L(S_N v^k, S_N w^\ell))]
\end{aligned}$$

converges in $\mathcal{C}^{\alpha+\beta+\gamma-\varepsilon}$ for all $\varepsilon > 0$. Moreover

$$\|\widetilde{R}(f,g,v,w)\|_{\alpha+\beta+\gamma} \lesssim \|f\|_\alpha \|g\|_\alpha \|v\|_\beta \|w\|_\gamma.$$

Proof. For fixed k, ℓ we have, for R as in Proposition 4.4.11,

$$
\begin{aligned}
\lim_{N\to\infty}&\big[L(S_N(\pi_<(f^k, v^k)), S_N(\pi_<(g^\ell, w^\ell))) - I(f^k g^\ell, \mathrm{d}L(S_N v^k, S_N w^\ell))\big] \\
&= R(f^k, v^k, \pi_<(g^\ell, w^\ell)) + \lim_{N\to\infty}\big[I(f^k, \mathrm{d}L(S_N v^k, S_N(\pi_<(g^\ell, w^\ell)))) \\
&\qquad\qquad\qquad\qquad\qquad\qquad - I(f^k, \mathrm{d}I(g^\ell, \mathrm{d}L(S_N v^k, S_N w^\ell)))\big] \\
&= R(f^k, v^k, \pi_<(g^\ell, w^\ell)) - \lim_{N\to\infty}\big[I(f^k, \mathrm{d}L(S_N(\pi_<(g^\ell, w^\ell)), S_N v^k)) \\
&\qquad\qquad\qquad\qquad\qquad\qquad - I(f^k, \mathrm{d}I(g^\ell, \mathrm{d}L(S_N w^\ell, S_N v^k)))\big] \\
&= R(f^k, v^k, \pi_<(g^\ell, w^\ell)) - I(f^k, \mathrm{d}R(g^\ell, w^\ell, v^k)),
\end{aligned}
$$

where we used that L is antisymmetric. The claimed estimate now follows from Proposition 4.4.11 and from Theorem 4.3.15. □

4.4.3. Pathwise integration for rough paths

In this section we apply our commutator estimates Proposition 4.4.11 respectively Corollary 4.4.14 to show that if $v, w \in \mathcal{C}^\alpha$ for $\alpha > 1/3$, and if $I(v, \mathrm{d}w) \in \mathcal{C}^\alpha$ is given and satisfies $I(v, \mathrm{d}w) - \pi_<(v, w) \in \mathcal{C}^{2\alpha}$, then we can construct the pathwise integral $I(f, \mathrm{d}g)$ for all $f \in \mathcal{D}^\alpha_v$ and $g \in \mathcal{D}^\alpha_w$.

Theorem 4.4.15. *Let $\alpha \in (1/3, 1)$, $\alpha \neq 1/2$, $\alpha \neq 2/3$. Let $v \in \mathcal{C}^\alpha(\mathbb{R}^d)$ and $w \in \mathcal{C}^\beta(\mathbb{R}^n)$, and assume that the Lévy area*

$$L(v, w) := \lim_{N\to\infty}\big(L(S_N v^k, S_N w^\ell)\big)_{1\le k\le d, 1\le \ell\le n}$$

converges uniformly, such that $\sup_N \|L(S_N v, S_N w)\|_{2\alpha} < \infty$. Let $f \in \mathcal{D}^\alpha_v(\mathcal{L}(\mathbb{R}^m, \mathbb{R}^e))$ and $g \in \mathcal{D}^\alpha_w(\mathbb{R}^m)$. Then

$$I(S_N f, \mathrm{d}S_N g) = \sum_{p\le N}\sum_{q\le N}\int_0^\cdot \Delta_p f(s)\mathrm{d}\Delta_q g(s)$$

converges in $\mathcal{C}^{\alpha-\varepsilon}$ for all $\varepsilon > 0$. We denote the limit by $I(f, \mathrm{d}g)$. Then

$$\|I(f, \mathrm{d}g)\|_\alpha \lesssim \|f\|_{v,\alpha}\|g\|_{w,\alpha}(1 + \|v\|_\alpha + \|w\|_\alpha + \|v\|_\alpha\|w\|_\alpha + \|L(v, w)\|_{2\alpha}).$$

Moreover, $I(f, \mathrm{d}g) \in \mathcal{D}^\alpha_w$ with derivative fg^w and

$$\|I(f, \mathrm{d}g)\|_{w,\alpha} \lesssim \|f\|_{v,\alpha}(1 + \|g\|_{w,\alpha})(1 + \|v\|_\alpha + \|w\|_\alpha + \|v\|_\alpha\|w\|_\alpha + \|L(v, w)\|_{2\alpha}).$$

Proof. We have

$$I(S_N f, \mathrm{d}S_N g) = S(S_N f, S_N g) + \pi_<(S_N f, S_N g) + L(S_N f, S_N g), \tag{4.22}$$

where S and $\pi_<$ are bounded bilinear operators, with

$$\|S(f,g)\|_{2\alpha} + \|\pi_<(f,g)\|_\alpha \lesssim \|f\|_\alpha \|g\|_\alpha, \tag{4.23}$$

see Lemma 4.3.5 and Lemma 4.3.14. Therefore, the result follows once we show that $L(S_N f, S_N g)$ converges in $\mathcal{C}^{2\alpha-\varepsilon}$ as $N \to \infty$, and that the limit $L(f,g)$ is in $\mathcal{C}^{2\alpha}$. But since $f \in \mathcal{D}^\alpha_v$ and $g \in \mathcal{D}^\alpha_w$, we obtain

$$\begin{aligned} L(S_N f, S_N g) = L(S_N f^\sharp, S_N g) + L(S_N \pi_<(f^v, v), S_N g^\sharp) \\ + L(S_N \pi_<(f^v, v), S_N \pi_<(g^w, w)). \end{aligned} \tag{4.24}$$

Now $f^\sharp, g^\sharp \in \mathcal{C}^{2\alpha}$ and $3\alpha > 1$. Hence, we can apply Lemma 4.3.13 to obtain the convergence of $L(S_N f^\sharp, S_N g) + L(S_N \pi_<(f^v, v), S_N g^\sharp)$ in $\mathcal{C}^{3\alpha-\varepsilon}$, as well as the estimate

$$\begin{aligned} \|L(f^\sharp, g)\|_{3\alpha} + \|L(\pi_<(f^v, v), g^\sharp)\|_{3\alpha} &\lesssim \|f^\sharp\|_{2\alpha}\|g\|_\alpha + \|\pi_<(f^v, v)\|_\alpha \|g^\sharp\|_{2\alpha} \\ &\lesssim (1 + \|v\|_\alpha)\|f\|_{v,\alpha}\|g\|_{w,\alpha}. \end{aligned} \tag{4.25}$$

Only the term $L(S_N \pi_<(f^v, v), S_N \pi_<(g^w, w))$ remains to be treated. But for this term we obtain from Corollary 4.4.14 that

$$\lim_{N\to\infty} L(S_N \pi_<(f^v, v), S_N \pi_<(g^w, w)) = \widetilde{R}(f^v, g^w, v, w) + \lim_{N\to\infty} I(f^v g^w, \mathrm{d}L(S_N v, S_N w)).$$

By assumption, $L(S_N v, S_N w)$ converges in $\mathcal{C}^{2\alpha-\varepsilon}$ to $L(v,w)$. The continuity of the Young integral, see Theorem 4.3.15, therefore implies the convergence in $\mathcal{C}^{2\alpha-\varepsilon}$ of $L(S_N f, S_n g)$ to $L(f,g)$, as well as the estimate

$$\begin{aligned} \|L(\pi_<(f^v, v), \pi_<(g^w, w))\|_{2\alpha} &\lesssim \|\widetilde{R}(f^v, g^w, v, w)\|_{3\alpha} + \|I(f^v g^w, \mathrm{d}L(v,w))\|_{2\alpha} \\ &\lesssim \|f^v\|_\alpha \|g^w\|_\alpha \big(\|v\|_\alpha \|w\|_\alpha + \|L(v,w)\|_{2\alpha}\big). \end{aligned} \tag{4.26}$$

Combining (4.22)–(4.26), we obtain $I(f, \mathrm{d}g) \in \mathcal{D}^\alpha_g$ with derivative f, and

$$\|I(f, \mathrm{d}g)\|_\alpha \lesssim \|f\|_{v,\alpha}\|g\|_{w,\alpha}(1 + \|v\|_\alpha + \|w\|_\alpha + \|v\|_\alpha\|w\|_\alpha + \|L(v,w)\|_{2\alpha})$$

as well as

$$\|I(f, \mathrm{d}g)\|_{g,\alpha} \lesssim \|f\|_{v,\alpha}(1 + \|g\|_{w,\alpha})\big(1 + \|v\|_\alpha + \|v\|_\alpha\|w\|_\alpha + \|L(v,w)\|_{2\alpha}\big).$$

Now we only need to apply Corollary 4.4.8 to obtain that $I(f, \mathrm{d}g) \in \mathcal{D}^\alpha_w$, and the estimate

$$\|I(f, \mathrm{d}g)\|_{w,\alpha} \lesssim \|f\|_{v,\alpha}(1 + \|g\|_{w,\alpha})\big(1 + \|v\|_\alpha + \|w\|_\alpha + \|v\|_\alpha\|w\|_\alpha + \|L(v,w)\|_{2\alpha}\big).$$

□

The integral I is a bounded bilinear operator, and therefore it is continuous. The difference of two integrals can be estimated:

Corollary 4.4.16. *Let α, v, w, f, g be as described in Theorem 4.4.15. Let $\tilde{v} \in \mathcal{C}^\alpha(\mathbb{R}^d)$ and $\tilde{w} \in \mathcal{C}^\alpha(\mathbb{R}^n)$, and assume that the Lévy area $L(S_N\tilde{v}, S_N\tilde{w})$ converges uniformly and with uniformly bounded $\mathcal{C}^{2\alpha}$ norm to $L(\tilde{v}, \tilde{w})$. Let $\tilde{f} \in \mathcal{D}^\alpha_{\tilde{v}}(\mathcal{L}(\mathbb{R}^m, \mathbb{R}^e))$ and $\tilde{g} \in \mathcal{D}^\alpha_{\tilde{w}}(\mathbb{R}^m)$. Then*

$$
\begin{aligned}
\|I(f, \mathrm{d}g) - I(\tilde{f}, \mathrm{d}\tilde{g})\|_\alpha \lesssim{} & (\|f - \tilde{f}\|_\alpha + \|f^v - \tilde{f}^{\tilde{v}}\|_\alpha + \|f^\sharp - \tilde{f}^\sharp\|_{2\alpha})\|g\|_{w,\alpha} \\
& \quad \times (1 + \|v\|_\alpha + \|v\|_\alpha\|w\|_\alpha + \|L(v, w)\|_{2\alpha}) \\
& + (\|g - \tilde{g}\|_\alpha + \|g^w - \tilde{g}^{\tilde{w}}\|_\alpha + \|g^\sharp - \tilde{g}^\sharp\|_{2\alpha})\|\tilde{f}\|_{\tilde{v},\alpha} \\
& \quad \times (1 + \|v\|_\alpha + \|v\|_\alpha\|w\|_\alpha + \|L(v, w)\|_{2\alpha}) \\
& + (\|v - \tilde{v}\|_\alpha + \|w - \tilde{w}\|_\alpha + \|L(v, w) - L(\tilde{v}, \tilde{w})\|_{2\alpha})\|\tilde{f}\|_{\tilde{v},\alpha}\|\tilde{g}\|_{\tilde{w},\alpha} \\
& \quad \times (1 + \|\tilde{v}\|_\alpha + \|w\|_\alpha).
\end{aligned}
$$

Proof. We decompose $I(f, \mathrm{d}g) - I(\tilde{f}, \mathrm{d}\tilde{g})$ in the same way as $I(f, \mathrm{d}g)$ was decomposed in the proof of Theorem 4.4.15. The claimed estimate then follows from multilinearity and boundedness of the involved operators. □

We can apply Corollary 4.4.16 to estimate $\|I(S_N f, \mathrm{d}S_N g) - I(f, \mathrm{d}g)\|_\alpha$. But usually we are more interested how close $I(S_N f, \mathrm{d}S_N g)$ and $I(f, \mathrm{d}g)$ are in uniform distance.

Corollary 4.4.17. *Let $\alpha \in (1/3, 1/2)$ and let v, w, f, g be as described in Theorem 4.4.15. Then we have for all $\varepsilon \in (0, 3\alpha - 1)$ that*

$$
\begin{aligned}
\|I(S_N f, \mathrm{d}S_N g) - I(f, \mathrm{d}g)\|_\infty \lesssim_\varepsilon{} & 2^{-N(3\alpha-1-\varepsilon)}\|f\|_{v,\alpha}\|g\|_{w,\alpha}(1 + \|v\|_\alpha + \|v\|_\alpha\|w\|_\alpha) \\
& + \|f\|_\alpha\|g\|_\alpha\|L(S_N v, S_N w) - L(v, w)\|_{2\alpha}.
\end{aligned}
$$

Proof. We decompose $I(S_N f, \mathrm{d}S_N g)$ as described in the proof of Theorem 4.4.15. This gives us for example the term

$$
S(S_N f, S_N g) - S(f, g) = S(S_N f - f, S_N g) + S(f, S_N g - g).
$$

Let $\delta > 0$ be such that $\alpha < (1-\delta)/2$. We have $\|\cdot\|_\infty \lesssim \|\cdot\|_{\alpha+\delta}$, and therefore Lemma 4.3.14 implies that

$$
\begin{aligned}
\|S(S_N f, S_N g) - S(f, g)\|_{\alpha+\delta} &\lesssim \|S_N f - f\|_\delta\|S_N g\|_\alpha + \|f\|_\alpha\|S_N g - g\|_\delta \\
&\lesssim 2^{-N(\alpha-\delta)}\|f\|_\alpha\|g\|_\alpha.
\end{aligned}
$$

By choice of δ we have $\alpha - \delta > 3\alpha - 1$, and therefore $2^{-N(\alpha-\delta)} < 2^{-N(3\alpha-1)} < 2^{-N(3\alpha-1-\varepsilon)}$.

Let us treat one of the critical terms, say $L(S_N f^\sharp, S_N g)$. Since $3\alpha - \varepsilon > 1$, we can apply Lemma 4.3.13 to obtain

$$
\begin{aligned}
\|L(S_N f^\sharp, S_N g) - L(f^\sharp, g)\|_\infty &\lesssim \|L(S_N f^\sharp - f^\sharp, S_N g)\|_{1+\varepsilon} + \|L(f^\sharp, S_N g - g)\|_{1+\varepsilon} \\
&\lesssim_\varepsilon \|S_N f^\sharp - f^\sharp\|_{1+\varepsilon-\alpha}\|g\|_\alpha + \|f^\sharp\|_{2\alpha}\|S_N g - g\|_{1+\varepsilon-2\alpha} \\
&\lesssim 2^{-N(2\alpha-(1+\varepsilon-\alpha))}\|f^\sharp\|_{2\alpha}\|g\|_\alpha + 2^{-N(\alpha-(1+\varepsilon-2\alpha))}\|f^\sharp\|_{2\alpha}\|g\|_\alpha
\end{aligned}
$$

$$\lesssim 2^{-N(3\alpha-1-\varepsilon)} \|f^\sharp\|_{2\alpha} \|g\|_\alpha.$$

By rewriting $\widetilde{R}$ in terms of R as in the proof of Corollary 4.4.14 and then applying Lemma 4.4.13, we see that

$$\begin{aligned}\|L(S_N \pi_<(f, v), S_N \pi_<(g, w))\|_\infty \lesssim\ & 2^{-N(3\alpha-1)} \|f\|_\alpha \|g\|_\alpha \|v\|_\alpha \|w\|_\alpha \\ & + \|I(fg, \mathrm{d}L(S_N v, S_N w)) - I(fg, \mathrm{d}L(v, w))\|_\infty.\end{aligned}$$

For the second term on the right hand side we use the continuity of the Young integral to obtain

$$\begin{aligned}\|I(fg, \mathrm{d}L(S_N v, S_N w)) - I(fg, \mathrm{d}L(v, w))\|_\infty & \lesssim \|I(fg, \mathrm{d}L(S_N v, S_N w) - \mathrm{d}L(v, w))\|_{2\alpha} \\ & \lesssim \|f\|_\alpha \|g\|_\alpha \|L(S_N v, S_N w) - L(v, w)\|_{2\alpha}.\end{aligned}$$

The other terms are treated with similar arguments, and the claimed estimate follows. □

Remark 4.4.18. In Lemma 4.4.13 we saw that the rate of convergence of

$$\begin{aligned}& L(S_N \pi_<(f^v, v), S_N \pi_<(g^w, w)) - I(fg, \mathrm{d}L(S_N v, S_N w)) \\ & \quad - (L(\pi_<(f^v, v), \pi_<(g^w, w)) - I(fg, \mathrm{d}L(v, w)))\end{aligned}$$

is in fact $2^{-N(3\alpha-1)}$ when measured in uniform distance, and not just $2^{-N(3\alpha-1-\varepsilon)}$. It is possible to show that this optimal rate is attained by the other terms as well, so that

$$\begin{aligned}\|I(S_N f, \mathrm{d}S_N g) - I(f, \mathrm{d}g)\|_\infty \lesssim\ & 2^{-N(3\alpha-1)} \|f\|_{v,\alpha} \|g\|_{w,\alpha} (1 + \|v\|_\alpha + \|v\|_\alpha \|w\|_\alpha) \\ & + \|f\|_\alpha \|g\|_\alpha \|L(S_N v, S_N w) - L(v, w)\|_{2\alpha-\varepsilon}.\end{aligned}$$

Since this requires a rather lengthy calculation, we decided not to include the arguments here.

Since we approximate f and g by the piecewise smooth functions $S_N f$ and $S_N g$ when defining the integral $I(f, \mathrm{d}g)$, it is not surprising that we obtain a Stratonovich type integral.

Proposition 4.4.19. *Let $\alpha \in (1/3, 1)$ and $v \in \mathcal{C}^\alpha(\mathbb{R}^d)$. Let $F \in C^3(\mathbb{R}^d, \mathbb{R})$. Then*

$$F(v(t)) - F(v(0)) = I(\mathrm{D}F(v), \mathrm{d}v)(t) := \lim_{N\to\infty} I(S_N \mathrm{D}F(v), \mathrm{d}S_N v)(t)$$

for all $t \in [0, 1]$.

Proof. The function $S_N v$ is Lipschitz continuous for all $N \in \mathbb{N}$. Therefore, we can apply integration by parts to obtain

$$F(S_N v(t)) - F(S_N v(0)) = I(\mathrm{D}F(S_N v), \mathrm{d}S_N v)(t).$$

The left hand side converges to $F(v(t)) - F(v(0))$ as N tends to ∞. To complete the proof it therefore suffices to show that $I(S_N \mathrm{D}F(v) - \mathrm{D}F(S_N v), \mathrm{d}S_N v)$ converges to zero as $N \to \infty$. By continuity of the Young integral, Theorem 4.3.15, it suffices to show that there exists a sufficiently small $\varepsilon > 0$ such that $\lim_{N\to\infty} \|S_N \mathrm{D}F(v) - \mathrm{D}F(S_N v)\|_{2\alpha-\varepsilon} = 0$. Recall that $S_N v$ is the linear interpolation of v between the points (t^1_{pm}) for $p \le N$ and $0 \le m \le 2^p$, and therefore $\Delta_p \mathrm{D}F(S_N v) = \Delta_p \mathrm{D}F(v) = \Delta_p S_N \mathrm{D}F(v)$ for all $p \le N$. For $p > N$ and $1 \le m \le 2^p$ we apply a first order Taylor expansion to both terms to see that

$$|[S_N \mathrm{D}F(v) - \mathrm{D}F(S_N v)]_{pm}| \le C_F 2^{-2p\alpha} \|S_N v\|_\alpha,$$

where $C_F > 0$ is a constant such that F and its partial derivatives up to order 3 are bounded by C_F on the support of v (and thus of $S_N v$). Therefore, we have for all $\varepsilon \in [0, 2\alpha)$ that

$$\|S_N \mathrm{D}F(v) - \mathrm{D}F(S_N v)\|_{2\alpha-\varepsilon} \le C 2^{-N\varepsilon} \|v\|_\alpha,$$

which completes the proof. □

Remark 4.4.20. Note that here we did not need any assumption on the convergence of the area $L(v, v)$. The reason are cancellations that arise due to the symmetric structure of the derivative of $\mathrm{D}F$, i.e. of the Hessian of F. It is possible to show directly for the Lévy area $L(S_N \mathrm{D}F(v), S_N v)$ that it converges, and in this approach the importance of the symmetry of $\mathrm{D}F$ becomes very obvious: After a first order Taylor expansion, the fact that L is antisymmetric implies, in conjunction with the symmetry of the Hessian of F, that all terms cancel except the remainder, which is smooth enough to be accessible with Young integration. However, the disadvantage of this approach is that then it is not trivial to identify the limit as $F(v(t)) - F(v(0))$. That is why we chose the approach presented above.

Proposition 4.4.19 was previously obtained by Roynette [Roy93], Proposition 1, except that Roynette assumed that v is one dimensional and in the Besov space $B^{1/2}_{1,\infty}$.

4.5. Pathwise Itô integration

In the previous section we saw that our pathwise integral $I(f, \mathrm{d}g)$ is of Stratonovich type, i.e. it satisfies the usual integration by parts rule. But in applications it may be interesting to have an Itô integral. Here we show that a slight modification of $I(f, \mathrm{d}g)$ allows us to treat non-anticipating Itô-type integrals.

In our dyadic context, a natural approximation of a non-anticipating integral is given for $k \in \mathbb{N}$ by

$$\begin{aligned} I^{\text{Itô}}_k(f, \mathrm{d}g)(t) &:= \sum_{\ell=0}^{2^k} f(t^0_{k\ell})(g(t^2_{k\ell} \wedge t) - g(t^0_{k\ell} \wedge t)) \\ &= \sum_{\ell=0}^{2^k} \sum_{p,q} \sum_{m,n} f_{pm} g_{qn} \varphi_{pm}(t^0_{k\ell})(\varphi_{qn}(t^2_{k\ell} \wedge t) - \varphi_{qn}(t^0_{k\ell} \wedge t)). \end{aligned}$$

Let us assume for the moment that $t = m2^{-k}$ for some $0 \leq m \leq 2^k$. In that case we obtain for $p \geq k$ or $q \geq k$ that $\varphi_{pm}(t^0_{k\ell})(\varphi_{qn}(t^2_{k\ell} \wedge t) - \varphi_{qn}(t^0_{k\ell} \wedge t)) = 0$. For $p, q < k$, both φ_{pm} and φ_{qn} are affine functions on $[t^0_{k\ell} \wedge t, t^2_{k\ell} \wedge t]$. And for affine v and w and $s < t$ it is not hard to see that

$$v(s)(w(t) - w(s)) = \int_s^t v(r)\mathrm{d}w(r) - \frac{1}{2}[v(t) - v(s)][w(t) - w(s)].$$

Hence, we conclude that for $t = m2^{-k}$ we have

$$I_k^{\text{Itô}}(f, \mathrm{d}g)(t) = I(S_{k-1}f, \mathrm{d}S_{k-1}g)(t) - \frac{1}{2}[f, g]_k(t), \tag{4.27}$$

where $[f, g]_k$ is the k–th dyadic approximation of the quadratic covariation $[f, g]$, i.e.

$$[f, g]_k(t) := \sum_{\ell=0}^{2^k} [f(t^2_{k\ell} \wedge t) - f(t^0_{k\ell} \wedge t)][g(t^2_{k\ell} \wedge t) - g(t^0_{k\ell} \wedge t)].$$

For the moment let us continue by studying the right-hand side of (4.27). Later we will show how to return from there to $I_k^{\text{Itô}}(f, \mathrm{d}g)(t)$ for general t, not necessarily of the form $t = m2^{-k}$.

We write $[w, w] := ([w^i, w^j])_{1 \leq i,j \leq d}$ and $L(w, w) := (L(w^i, w^j))_{1 \leq i,j \leq d}$, and similarly for all expressions of the same type.

Theorem 4.5.1. *Let $\alpha \in (1/3, 1/2)$ and let $w \in \mathcal{C}^\alpha(\mathbb{R}^d)$ and $f, g \in \mathcal{D}_w^\alpha(\mathbb{R})$. Assume that $(L(S_k w, S_k w))$ converges uniformly, with uniformly bounded $\mathcal{C}^{2\alpha}$ norm. Also assume that $([w, w]_k)$ converges uniformly. Then $I(S_{k-1}f, \mathrm{d}S_{k-1}g)(t) - 1/2[f, g]_k(t)$ converges uniformly to a limit $I^{\text{Itô}}(f, \mathrm{d}g)$ that satisfies*

$$\|I^{\text{Itô}}(f, \mathrm{d}g)\|_\infty \lesssim \|f\|_{w,\alpha}\|g\|_{w,\alpha}(1 + \|w\|_\alpha^2 + \|L(w, w)\|_{2\alpha} + \|[w, w]\|_\infty).$$

The quadratic variation of $I^{\text{Itô}}(f, \mathrm{d}g)$ is given by

$$[f, g] = \sum_{i,j=1}^d \int_0^\cdot f^{w,i}(s) g^{w,j}(s) \mathrm{d}[w^i, w^j](s). \tag{4.28}$$

Moreover, for $\varepsilon \in (0, 3\alpha - 1)$ the speed of convergence can be estimated by

$$\begin{aligned} \Big\| I^{\text{Itô}}(f, \mathrm{d}g) - \Big(I(S_{k-1}f, \mathrm{d}S_{k-1}g) - \frac{1}{2}[f, g]_k \Big) \Big\|_\infty & \\ \lesssim_\varepsilon 2^{-k(3\alpha - 1 - \varepsilon)} \|f\|_{w,\alpha}\|g\|_{w,\alpha}(1 + \|w\|_\alpha + \|w\|_\alpha^2) & \\ + \|f\|_\alpha \|g\|_\alpha \|L(S_{k-1}w, S_{k-1}y) - L(w, y)\|_{2\alpha} & \\ + \|f^w\|_\infty \|g^w\|_\infty \|[w, w]_k - [w, w]\|_\infty. & \end{aligned}$$

Proof. Let us first treat the quadratic variation. Recall from Lemma 4.4.5 that $f \in \mathcal{D}_w^\alpha(\mathbb{R})$ if and only if there exists $R^f : [0, 1]^2 \to \mathbb{R}$, such that $|R^f_{s,t}| \lesssim |t - s|^{2\alpha}$, and such that for all $0 \le s < t \le 1$ we have $f_{s,t} = f^w(s) w_{s,t} + R^f_{s,t}$. An analogous statement holds for g. Hence

$$
\begin{aligned}
[f,g]_k(t) &= \sum_\ell f_{t^0_{k\ell}\wedge t, t^2_{k\ell}\wedge t} g_{t^0_{k\ell}\wedge t, t^2_{k\ell}\wedge t} \\
&= \sum_\ell R^f_{t^0_{k\ell}\wedge t, t^2_{k\ell}\wedge t} g_{t^0_{k\ell}\wedge t, t^2_{k\ell}\wedge t} + \sum_\ell f^w(t^0_{k\ell}\wedge t) w_{t^0_{k\ell}\wedge t, t^2_{k\ell}\wedge t} R^g_{t^0_{k\ell}\wedge t, t^2_{k\ell}\wedge t} \\
&\quad + \sum_{i,j}\sum_\ell f^{w,i}(t^0_{k\ell}) g^{w,j}(t^0_{k\ell}) w^i_{t^0_{k\ell}\wedge t, t^2_{k\ell}\wedge t} w^j_{t^0_{k\ell}\wedge t, t^2_{k\ell}\wedge t}.
\end{aligned}
$$

It is easy to see that there exists $C > 0$, such that the first two terms on the right hand side are uniformly bounded by $C2^{-k(3\alpha-1)}\|f\|_{w,\alpha}\|g\|_{w,\alpha}$. For the third term, let us fix i and j. Then this is just the integral of $f^{w,i}g^{w,j}$ with respect to the measure $\mu^k_t = \sum_\ell \delta_{t^0_{k\ell}} w^i_{t^0_{k\ell}\wedge t, t^2_{k\ell}\wedge t} w^j_{t^0_{k\ell}\wedge t, t^2_{k\ell}\wedge t}$. We can decompose the measure μ^k_t into a positive and negative part as

$$
\mu^k_t = \frac{1}{4}\left[\sum_\ell \delta_{t^0_{k\ell}}[(w^i + w^j)_{t^0_{k\ell}\wedge t, t^2_{k\ell}\wedge t}]^2 - \sum_\ell \delta_{t^0_{k\ell}}[(w^i - w^j)_{t^0_{k\ell}\wedge t, t^2_{k\ell}\wedge t}]^2\right] =: \mu^{k,+}_t - \mu^{k,-}_t.
$$

Hence, we can estimate

$$
\begin{aligned}
&\left|\int_0^1 f^{w,i}(s) g^{w,j}(s)\mu^k_t(\mathrm{d}s) - \int_0^1 f^{w,i}(s) g^{w,j}(s)\mu_t(\mathrm{d}s)\right| \\
&\quad \lesssim \left\|f^{w,i}g^{w,j}\right\|_\infty \left(\left\|[w^i + w^j]_k - [w^i + w^j]\right\|_\infty + \left\|[w^i - w^j]_k - [w^i - w^j]\right\|_\infty\right) \\
&\quad \lesssim \left\|f^{w,i}g^{w,j}\right\|_\infty \|[w,w]_k - [w,w]\|_\infty,
\end{aligned}
$$

where we write $[u] := [u,u]$ and similarly for $[u]_k$. By assumption, the right hand side converges to zero, from where we get the uniform convergence of $[f,g]_k$ to $[f,g]$. Moreover, we have the explicit representation

$$
[f,g](t) = \sum_{i,j}\int_0^t f^{w,i}(s) g^{w,j}(s)\mathrm{d}[w^i,w^j](s),
$$

and therefore $\|[f,g]\|_\infty \lesssim \|f^w\|_\infty \|g^w\|_\infty \|[w,w]\|_\infty$, where we use the decomposition of $[w^i,w^j]$ into the difference of two nondecreasing processes, $[w^i,w^j] = 1/4([w^i + w^j] - [w^i - w^j])$.

Now let us come to the integral $I(S_{k-1}f, \mathrm{d}S_{k-1}g)$. Here we only need to apply Theorem 4.4.15, to obtain convergence to a limit $I(f,\mathrm{d}g)$ that satisfies

$$
\|I(f,\mathrm{d}g)\|_\infty \lesssim \|f\|_{w,\alpha}\|g\|_{w,\alpha}(1 + \|w\|^2_\alpha + \|L(w,w)\|_{2\alpha}),
$$

where we used that $1 + \|w\|_\alpha + \|w\|^2_\alpha \lesssim 1 + \|w\|^2_\alpha$. According to Corollary 4.4.17, the speed of convergence can be estimated by

$$\begin{aligned}\|I(f,\mathrm{d}g) - I(S_{k-1}f,\mathrm{d}S_{k-1}g)\|_\infty \lesssim_\varepsilon\ & 2^{-k(3\alpha-1-\varepsilon)}\|f\|_{w,\alpha}\|g\|_{w,\alpha}(1+\|w\|_\alpha+\|w\|_\alpha^2)\\ &+\|f\|_\alpha\|g\|_\alpha\|L(S_{k-1}w,S_{k-1}w)-L(w,w)\|_{2\alpha}.\end{aligned}$$

□

Note that $[w,w]$ is always a continuous function of bounded variation, but a priori it is not clear whether it is in $\mathcal{C}^{2\alpha}$. Under this additional assumption we have the following stronger result.

Corollary 4.5.2. *In addition to the conditions of Theorem 4.5.1, assume that also $[w,w]\in\mathcal{C}^{2\alpha}$. Then $I^{\text{Itô}}(f,\mathrm{d}g)\in\mathcal{D}^\alpha_w$ with derivative fg^w, and*

$$\|I^{\text{Itô}}(f,\mathrm{d}g)\|_{w,\alpha}\lesssim\|f\|_{w,\alpha}(1+\|g\|_{w,\alpha})(1+\|w\|_\alpha^2+\|L(w,w)\|_{2\alpha}+\|[w,w]\|_{2\alpha}).$$

Let moreover $\tilde{w}\in\mathcal{C}^\alpha(\mathbb{R}^d)$ with Lévy area $L(S_k\tilde{w},S_k\tilde{w})$ that converges uniformly and with uniformly bounded $\mathcal{C}^{2\alpha}$ norm to $L(\tilde{w},\tilde{w})$, and with quadratic variation $[\tilde{w},\tilde{w}]_k$ that converges uniformly to $[\tilde{w},\tilde{w}]\in\mathcal{C}^{2\alpha}$. Let $\tilde{f},\tilde{g}\in\mathcal{D}^\alpha_{\tilde{w}}(\mathbb{R})$. Then

$$\begin{aligned}\|I^{\text{Itô}}(f,\mathrm{d}g)-I^{\text{Itô}}(\tilde{f},d\tilde{g})\|_\alpha\lesssim\ & (\|f-\tilde{f}\|_\alpha+\|f^w-\tilde{f}^{\tilde{w}}\|_\alpha+\|f^\sharp-\tilde{f}^\sharp\|_{2\alpha})\|g\|_{w,\alpha}\\ &\qquad\times(1+\|w\|_\alpha^2+\|L(w,w)\|_{2\alpha}+\|[w,w]\|_{2\alpha})\\ &+(\|g-\tilde{g}\|_\alpha+\|g^w-\tilde{g}^{\tilde{w}}\|_\alpha+\|g^\sharp-\tilde{g}^\sharp\|_{2\alpha})\|\tilde{f}\|_{\tilde{w},\alpha}\\ &\qquad\times(1+\|w\|_\alpha^2+\|L(w,w)\|_{2\alpha}+\|[w,w]\|_{2\alpha})\\ &+(\|w-\tilde{w}\|_\alpha+\|L(w,w)-L(\tilde{w},\tilde{w})\|_{2\alpha}+\|[w,w]-[\tilde{w},\tilde{w}]\|_{2\alpha})\\ &\qquad\times\|\tilde{f}\|_{\tilde{w},\alpha}\|\tilde{g}\|_{\tilde{w},\alpha}(1+\|\tilde{w}\|_\alpha+\|w\|_\alpha).\end{aligned}$$

Proof. This is a combination of Theorem 4.4.15 and Corollary 4.4.16, and the explicit representation (4.28) for the quadratic variation. We also need continuity of the Young integral, Theorem 4.3.15, for example to estimate $\|[f,g]\|_{2\alpha}$. □

The term $I(S_{k-1}f,\mathrm{d}S_{k-1}g)$ has the pleasant property that if we want to refine our calculation by passing from k to $k+1$, then we can build on our existing calculation and only add the additional terms $I(S_{k-1}f,\mathrm{d}\Delta_k g)+I(\Delta_k f,\mathrm{d}S_k g)$. For the quadratic variation $[f,g]_k$ this is not exactly true. But note that $[f,g]_k(m2^{-k})=[S_{k-1}f,S_{k-1}g]_k(m2^{-k})$ for $m=0,\dots,2^k$. And there is a recursive way of calculating $[S_{k-1}f,S_{k-1}g]_k$:

Lemma 4.5.3. *Let $f,g\in C([0,1],\mathbb{R})$. Then we have for all $k\geq 1$ and all $t\in[0,1]$ that*

$$[S_kf,S_kg]_{k+1}(t)=\frac{1}{2}[S_{k-1}f,S_{k-1}g]_k(t)+[S_{k-1}f,\Delta_kg]_{k+1}(t)+[\Delta_kf,S_kg]_{k+1}(t)+R_k(t),\tag{4.29}$$

where

$$R_k(t) := -\frac{1}{2} f_{\llcorner t^k \lrcorner, t} g_{\llcorner t^k \lrcorner, t} + f_{\llcorner t^k \lrcorner, \ulcorner t^{k+1} \urcorner \wedge t} g_{\llcorner t^k \lrcorner, \ulcorner t^{k+1} \urcorner \wedge t} + f_{\ulcorner t^{k+1} \urcorner \wedge t, t} g_{\ulcorner t^{k+1} \urcorner \wedge t, t}$$

and $\llcorner t^k \lrcorner := \lfloor t 2^k \rfloor 2^{-k}$ *and* $\ulcorner t^k \urcorner := \llcorner t^k \lrcorner + 2^{-(k+1)}$*. In particular, we obtain for* $t = 1$ *that*

$$[f, g]_{k+1}(1) = \frac{1}{2}[f, g]_k(1) + \frac{1}{2} \sum_m f_{km} g_{km} = \frac{1}{2^{k+1}} \sum_{p \le k} \sum_m 2^p f_{pm} g_{pm}. \tag{4.30}$$

If moreover $\alpha \in (0, 1)$ *and* $f, g \in \mathcal{C}^\alpha$*, then*

$$\| [S_{k-1} f, S_{k-1} g]_k - [f, g]_k \|_\infty \lesssim 2^{-2k\alpha} \|f\|_\alpha \|g\|_\alpha.$$

Proof. By subtracting $[S_{k-1} f, \Delta_k g]_{k+1}(t) + [\Delta_k f, S_k g]_{k+1}(t)$ on both sides of (4.29), we see that it suffices to show $[S_{k-1} f, S_{k-1} g]_{k+1} = 1/2 [S_{k-1} f, S_{k-1} g]_k + R_k$. Let us assume that $t = m 2^{-k}$. In that case $R_k(t) = 0$, and for every $\ell \le 2^k$ we obtain

$$\begin{aligned} ([S_{k-1} f, S_{k-1} g]_{k+1})_{t^0_{k\ell}, t^2_{k\ell}} &= \Big((S_{k-1} f)_{t^0_{k\ell}, t^1_{k\ell}} (S_{k-1} g)_{t^0_{k\ell}, t^1_{k\ell}} + (S_{k-1} f)_{t^1_{k\ell}, t^2_{k\ell}} (S_{k-1} g)_{t^1_{k\ell}, t^2_{k\ell}} \Big) \\ &= \frac{1}{2} (S_{k-1} f)_{t^0_{k\ell}, t^2_{k\ell}} (S_{k-1} g)_{t^0_{k\ell}, t^2_{k\ell}} = \frac{1}{2} ([S_{k-1} f, S_{k-1} g]_k)_{t^0_{k\ell}, t^2_{k\ell}}, \end{aligned}$$

where we used that $S_{k-1} f$ and $S_{k-1} g$ are linear on $[t^0_{k\ell}, t^2_{k\ell}]$, and that the two intervals $[t^0_{k\ell}, t^1_{k\ell}]$ and $[t^1_{k\ell}, t^2_{k\ell}]$ have the same length 2^{-k-1}. The term R_k is now chosen exactly so that we also obtain the right expression for $t \in [0, 1]$ that is not of the form $m 2^{-k}$.

The formula for $[f, g]_{k+1}(1)$ follows because $[f, g]_{k+1}(1) = [S_k f, S_k g]_{k+1}(1)$, and because it is easy to see that $[\Delta_p f, \Delta_q g]_{k+1}(1) = 0$ unless $p = q$, and that $[\Delta_k f, \Delta_k g]_{k+1} = 1/2 \sum_m f_{km} g_{km}$.

The estimate for $\| [S_{k-1} f, S_{k-1} g]_k - [f, g]_k \|_\infty$ holds because the two functions agree in all dyadic points of the form $m 2^{-k}$, and because between two such points the quadratic variation can pick up mass of at most $2^{-2k\alpha} \|f\|_\alpha \|g\|_\alpha$. □

Remark 4.5.4. The Cesàro mean formula (4.30) makes the study of existence of the quadratic variation accessible to ergodic theory. This was previously observed by Gantert [Gan94]. See also Gantert's thesis [Gan91], Beispiel 3.29, where it is shown that ergodicity alone (of the distribution of w with respect to suitable transformations on path space) is not sufficient to obtain convergence of $([w, w]_k(1))$ as k tends to ∞.

Recall that we defined $I^{\text{Itô}}_k(f, \mathrm{d}g)(t) = \sum_\ell f(t^0_{k\ell}) g_{t^0_{k\ell} \wedge t, t^2_{k\ell} \wedge t}$.

Remark 4.5.5. Let $\alpha \in (0, 1)$. If $f \in C([0, 1])$ and $g \in \mathcal{C}^\alpha$, then

$$\Big\| I^{\text{Itô}}_k(f, \mathrm{d}g) - \Big(I(S_{k-1} f, \mathrm{d} S_{k-1} g) - \frac{1}{2} [S_{k-1} f, S_{k-1} g]_k \Big) \Big\|_\infty \lesssim 2^{-k\alpha} \|f\|_\infty \|g\|_\alpha.$$

This holds because both functions agree in all dyadic points of the form $m 2^{-k}$, and because between those points the integrals can pick up mass of at most $\|f\|_\infty 2^{-k\alpha} \|g\|_\alpha$.

It follows from Remark 4.5.5 that our pathwise Itô type integral constructed in Theorem 4.5.1 is the limit of non-anticipating Riemann sums. Therefore, it would be more natural to assume that also for the controlling path w the non-anticipating Riemann sums converge, rather than assuming that $(L(S_k w, S_k w))_k$ and $([w,w]_k)$ converge. Below we show that this is sufficient, as long as a uniform Hölder estimate is satisfied by the Riemann sums. In that case all the conditions of Theorem 4.5.1 and of Corollary 4.5.2 are satisfied.

We first show that the existence of the Itô iterated integrals implies the existence of the quadratic variation.

Lemma 4.5.6. *Let $\alpha \in (0,1/2)$ and let $w \in \mathcal{C}^\alpha(\mathbb{R}^d)$. Assume that the non-anticipating Riemann sums $(I_k^{\text{Itô}}(w, \mathrm{d}w))_k$ converge uniformly to $I^{\text{Itô}}(w,\mathrm{d}w)$. Then also $([w,w]_k)_k$ converges uniformly to a limit $[w,w]$. Moreover, for all $0 \le s < t \le 1$*

$$|[w,w]_k(t) - [w,w]_k(s)| \lesssim |I_k^{\text{Itô}}(w,\mathrm{d}w)_{s,t} - w(s)w_{s,t}| + |w_{s,t}|^2. \tag{4.31}$$

If moreover

$$\sup_k \sup_{0\le \ell < \ell' \le 2^k} \frac{|I_k^{\text{Itô}}(w,\mathrm{d}w)(\ell' 2^{-k}) - I_k^{\text{Itô}}(w,\mathrm{d}w)(\ell 2^{-k}) - w(\ell 2^{-k})(w(\ell' 2^{-k}) - w(\ell 2^{-k}))|}{|(\ell' - \ell)2^{-k}|^{2\alpha}} = C < \infty,$$

then $[w,w] \in \mathcal{C}^{2\alpha}$, and

$$\|[w,w]\|_{2\alpha} \lesssim C + \|w\|_\alpha^2. \tag{4.32}$$

Proof. Let $t \in [0,1]$ and $1 \le i,j \le d$. Then

$$\begin{aligned} w^i(t)w^j(t) - w^i(0)w^j(0) &= \sum_{\ell=1}^{2^k} \Big[w^i(t^2_{k\ell}\wedge t) w^j(t^2_{k\ell}\wedge t) - w^i(t^0_{k\ell}\wedge t)w^j(t^0_{k\ell}\wedge t)\Big] \\ &= \sum_{\ell=1}^{2^k} \Big[w^i(t^0_{k\ell}) w^j_{t^0_{k\ell}\wedge t, t^2_{k\ell}\wedge t} + w^j(t^0_{k\ell}) w^i_{t^0_{k\ell}\wedge t, t^2_{k\ell}\wedge t} + w^i_{t^0_{k\ell}\wedge t, t^2_{k\ell}\wedge t} w^j_{t^0_{k\ell}\wedge t, t^2_{k\ell}\wedge t}\Big] \\ &= I_k^{\text{Itô}}(w^i, \mathrm{d}w^j)(t) + I_k^{\text{Itô}}(w^j,\mathrm{d}w^i)(t) + [w^i,w^j]_k(t), \end{aligned} \tag{4.33}$$

which implies the convergence of $([w,w]_k)_k$ as k tends to ∞. For $0 \le s < t \le 1$ we obtain from (4.33) that

$$\begin{aligned} ([w^i,w^j]_k)_{s,t} &= \big(w^i w^j\big)_{s,t} - I_k^{\text{Itô}}(w^i,\mathrm{d}w^j)_{s,t} - I_k^{\text{Itô}}(w^j,\mathrm{d}w^i)_{s,t} \\ &= \Big[w^i(s)w^j_{s,t} - I_k^{\text{Itô}}(w^i,\mathrm{d}w^j)_{s,t}\Big] + \Big[w^j(s)w^i_{s,t} - I_k^{\text{Itô}}(w^j,\mathrm{d}w^i)_{s,t}\Big] + w^i_{s,t}w^j_{s,t}, \end{aligned}$$

leading to (4.31). Given (4.31) it is now easy to estimate $\|[w,w]\|_{2\alpha}$. We estimate the classical Hölder norm, not the $\mathcal{C}^{2\alpha}$ norm. Let $0 \le s < t \le 1$. Using the continuity of $[w,w]$, we choose k large enough such that there exist $s < s_k = \ell_s 2^{-k} < t$ and

$s < t_k = \ell_t 2^{-k} < t$ with

$$|[w,w]_{s,s_k}| + |[w,w]_{t_k,t}| + \|[w,w]_k - [w,w]\|_\infty \leq \|w\|_\alpha^2 |t-s|^{2\alpha}.$$

Since

$$|[w,w]_{s,t}| \leq |[w,w]_{s,s_k}| + |[w,w]_{t_k,t}| + \|[w,w]_k - [w,w]\|_\infty,$$

we obtain (4.32) as a consequence of (4.31) and the hypothesis. □

Let us show that convergence of $(I_k^{\text{Itô}}(w, \mathrm{d}w))$ implies convergence of $(L(S_k w, S_k w))_k$:

Lemma 4.5.7. *Let $\alpha \in (0, 1/2)$, and let $w \in \mathcal{C}^\alpha(\mathbb{R}^d)$. Assume that the non-anticipating integrals $(I_k^{\text{Itô}}(w, \mathrm{d}w))_k$ converge uniformly, and that*

$$\sup_k \sup_{0 \leq \ell < \ell' \leq 2^k} \frac{|I_k^{\text{Itô}}(w,\mathrm{d}w)(\ell' 2^{-k}) - I_k^{\text{Itô}}(w,\mathrm{d}w)(\ell 2^{-k}) - w(\ell 2^{-k})(w(\ell' 2^{-k}) - w(\ell 2^{-k}))|}{|(\ell' - \ell)2^{-k}|^{2\alpha}}$$
$$= C < \infty.$$

Then $L(S_k w, S_k w)$ converges uniformly as $k \to \infty$, and

$$\sup_k \|L(S_k w, S_k w)\|_{2\alpha} \lesssim C + \|w\|_\alpha^2.$$

Proof. Let $k \in \mathbb{N}$ and $0 \leq \ell \leq 2^k$, and write $t = \ell 2^{-k}$. Then we obtain from (4.27) that

$$\begin{aligned} &L(S_{k-1}w, S_{k-1}w)(t) \\ &\quad = I(S_{k-1}w, \mathrm{d}S_{k-1}w)(t) - \pi_<(S_{k-1}w, S_{k-1}w)(t) - S(S_{k-1}w, S_{k-1}w)(t) \\ &\quad = I_k^{\text{Itô}}(w, \mathrm{d}w)(t) + \frac{1}{2}[w,w]_k(t) - \pi_<(S_{k-1}w, S_{k-1}w)(t) - S(S_{k-1}w, S_{k-1}w)(t). \end{aligned} \tag{4.34}$$

Let now $s, t \in [0,1]$. We first assume that there exists ℓ such that $t_{k\ell}^0 \leq s < t \leq t_{k\ell}^2$. Then we use that $\|\partial_t \Delta_q w\|_\infty \lesssim 2^{q(1-\alpha)} \|w\|_\alpha$ to obtain

$$\begin{aligned} |L(S_{k-1}w, S_{k-1}w)_{s,t}| &\leq \sum_{p<k} \sum_{q<p} \left| \int_s^t \Delta_p w(r) \mathrm{d}\Delta_q w(r) - \int_s^t \mathrm{d}\Delta_q w(r) \Delta_p w(r) \right| \\ &\lesssim \sum_{p<k} \sum_{q<p} |t-s| 2^{-p\alpha} 2^{q(1-\alpha)} \|w\|_\alpha^2 \lesssim |t-s| 2^{-k(2\alpha-1)} \|w\|_\alpha^2 \leq |t-s|^{2\alpha} \|w\|_\alpha^2, \end{aligned} \tag{4.35}$$

where we used that $2\alpha - 1 < 0$, and also that $|t-s| \leq 2^{-k}$ by assumption.

Combining (4.34) and (4.35), we obtain the uniform convergence of $(L(S_{k-1}w, S_{k-1}w))$ from Lemma 4.5.6 and from the continuity of $\pi_<$ and S.

For s and t that do not lie in the same dyadic interval of generation k, let $\ulcorner s^k \urcorner = \ell_s 2^{-k}$ and $\llcorner t^k \lrcorner = \ell_t 2^{-k}$ be such that $\ulcorner s^k \urcorner - 2^{-k} < s \leq \ulcorner s^k \urcorner$ and $\llcorner t^k \lrcorner \leq t < \llcorner t^k \lrcorner + 2^{-k}$. In particular, $\ulcorner s^k \urcorner \leq \llcorner t^k \lrcorner$. Moreover

$$\begin{aligned}|L(S_{k-1}w, S_{k-1}w)_{s,t}| \leq{}& |L(S_{k-1}w, S_{k-1}w)_{s,\ulcorner s^k \urcorner}| + |L(S_{k-1}w, S_{k-1}w)_{\ulcorner s^k \urcorner, \llcorner t^k \lrcorner}| \\ &+ |L(S_{k-1}w, S_{k-1}w)_{\llcorner t^k \lrcorner, t}|.\end{aligned}$$

According to (4.35), the first and third term on the right hand side can be estimated by $(|\ulcorner s^k \urcorner - s|^{2\alpha} + |t - \llcorner t^k \lrcorner|^{2\alpha})\|w\|_\alpha^2 \lesssim |t-s|^{2\alpha}\|w\|_\alpha^2$. For the middle term we apply (4.34) to obtain

$$\begin{aligned}|L(S_{k-1}w, S_{k-1}w)_{\ulcorner s^k \urcorner, \llcorner t^k \lrcorner}| \leq{}& \left| I_k^{\text{Itô}}(w, \mathrm{d}w)_{\ulcorner s^k \urcorner, \llcorner t^k \lrcorner} - w(\ulcorner s^k \urcorner)(w(\llcorner t^k \lrcorner) - w(\ulcorner s^k \urcorner)) \right| \\ &+ \left| w(\ulcorner s^k \urcorner) w_{\ulcorner s^k \urcorner, \llcorner t^k \lrcorner} - \pi_<(S_{k-1}w, S_{k-1}w)_{\ulcorner s^k \urcorner, \llcorner t^k \lrcorner} \right| \\ &+ \frac{1}{2}\left| ([w,w]_k)_{\ulcorner s^k \urcorner, \llcorner t^k \lrcorner} \right| + \left| S(S_{k-1}w, S_{k-1}w)_{\ulcorner s^k \urcorner, \llcorner t^k \lrcorner} \right| \\ \lesssim{}& |\llcorner t^k \lrcorner - \ulcorner s^k \urcorner|^{2\alpha} \left(C + \|w\|_\alpha^2 \right) \leq |t-s|^{2\alpha} \left(C + \|w\|_\alpha^2 \right),\end{aligned}$$

where we used Lemma 4.4.5, Lemma 4.5.6, and Lemma 4.3.14. □

Combining Lemma 4.5.6 and Lemma 4.5.7 with Theorem 4.5.1, we see that uniform convergence of $(I_k^{\text{Itô}}(w, \mathrm{d}w))_k$ to $I^{\text{Itô}}(w, \mathrm{d}w)$ implies the uniform convergence of $(I_k^{\text{Itô}}(f, \mathrm{d}g))_k$ to $I^{\text{Itô}}(f, \mathrm{d}g)$ for f and g controlled by w:

Corollary 4.5.8. *Let $\alpha \in (1/3, 1/2)$ and let $w \in \mathcal{C}^\alpha(\mathbb{R}^d)$ and $f, g \in \mathcal{D}_w^\alpha(\mathbb{R})$. Assume that the non-anticipating Riemann sums $(I_k^{\text{Itô}}(w, \mathrm{d}w))_k$ converge uniformly to $I^{\text{Itô}}(w, \mathrm{d}w)$, and that furthermore*

$$\begin{aligned}\sup_k \sup_{0\leq \ell < \ell' \leq 2^k} & \frac{|I_k^{\text{Itô}}(w, \mathrm{d}w)(\ell' 2^{-k}) - I_k^{\text{Itô}}(w, \mathrm{d}w)(\ell 2^{-k}) - w(\ell 2^{-k})(w(\ell' 2^{-k}) - w(\ell 2^{-k}))|}{|(\ell' - \ell)2^{-k}|^{2\alpha}} \\ &= C < \infty.\end{aligned}$$

Then the non-anticipating Riemann sums $(I_k^{\text{Itô}}(f, \mathrm{d}g))_k$ converge to a limit $I^{\text{Itô}}(f, \mathrm{d}g)$ that satisfies

$$\|I^{\text{Itô}}(f, \mathrm{d}g)\|_\infty \lesssim \|f\|_{w,\alpha} \|g\|_{w,\alpha}(1 + \|w\|_\alpha^2 + C).$$

Remark 4.5.9. Observe that we calculate the pathwise Itô integral $I^{\text{Itô}}(f, \mathrm{d}g)$ as limit of Riemann sums involving only f and g, and not the Lévy area of $L(w,w)$ or the quadratic variation $[w,w]$. The classical rough path integral, see Proposition 4.2.4, is obtained as a "compensated Riemann sum" that involves f and g, but also their derivatives with respect to w, as well as the iterated integrals of w. For applications in mathematical finance, it is more convenient to have an integral that is the limit of Riemann sums involving only f and g, because then this integral can be interpreted as capital process obtained by investing in g.

It follows from the work of Föllmer [Föl79] that our pathwise Itô integral satisfies Itô's formula:

Corollary 4.5.10. *Let $\alpha \in (1/3, 1/2)$ and let $w \in \mathcal{C}^\alpha(\mathbb{R}^d)$ and $f, g \in \mathcal{D}^\alpha_w(\mathbb{R})$. Assume that the non-anticipating Riemann sums $(I^{\mathrm{Itô}}_k(w, \mathrm{d}w))_k$ converge uniformly to $I^{\mathrm{Itô}}(w, \mathrm{d}w)$, and that furthermore*

$$\sup_k \sup_{0 \le \ell < \ell' \le 2^k} \frac{|I^{\mathrm{Itô}}_k(w, \mathrm{d}w)(\ell' 2^{-k}) - I^{\mathrm{Itô}}_k(w, \mathrm{d}w)(\ell 2^{-k}) - w(\ell 2^{-k})(w(\ell' 2^{-k}) - w(\ell 2^{-k}))|}{|(\ell' - \ell) 2^{-k}|^{2\alpha}}$$
$$= C < \infty.$$

Let $F \in C^2(\mathbb{R}^d, \mathbb{R})$. Then $(I^{\mathrm{Itô}}(\mathrm{D}F(w), \mathrm{d}w))_k$ converges to a limit $I^{\mathrm{Itô}}(\mathrm{D}F(w), \mathrm{d}w)$ that satisfies for all $t \in [0, 1]$

$$F(w(t)) - F(w(0)) = I^{\mathrm{Itô}}(\mathrm{D}F(w), \mathrm{d}w)(t) + \int_0^t \sum_{k,\ell=1}^d \partial_{x_k} \partial_{x_\ell} F(w(s)) \mathrm{d}[w^k, w^\ell](s).$$

Proof. This is Remarque 1 of Föllmer [Föl79] in combination with Lemma 4.5.6. □

Remark 4.5.11. Note that $\mathrm{D}F \in C^1$, and therefore $\mathrm{D}F(w)$ is not controlled by w. Just as in the Stratonovich case, see Remark 4.4.20, the symmetry of the derivative of $\mathrm{D}F$ leads to crucial cancellations that allow to take $\mathrm{D}F$ less regular than in the non-gradient case.

4.6. Construction of the Lévy area

To apply our theory, it remains to construct the Lévy area respectively the pathwise Itô iterated integrals for suitable stochastic processes. In Section 4.6.1 we construct the Lévy area for hypercontractive stochastic processes whose covariance function satisfies a certain "finite variation" property. In Section 4.6.2 we construct the pathwise Itô iterated integrals for some continuous martingales.

4.6.1. Hypercontractive processes

Let $X : [0, 1] \to \mathbb{R}^d$ be a centered continuous stochastic process, such that X^i is independent of X^j for $i \neq j$. We write R for its covariance function, i.e. $R : [0, 1]^2 \to \mathbb{R}^{d \times d}$ and $R(s, t) := (E(X^i_s X^j_t))_{1 \le i,j \le d}$. The increment of R over a rectangle $[s, t] \times [u, v] \subseteq [0, 1]^2$ is defined as

$$R_{[s,t] \times [u,v]} = R(t, v) + R(s, u) - R(s, v) - R(t, u) = (E(X^i_{s,t} X^j_{u,v}))_{1 \le i,j \le d}.$$

Let us make the following two assumptions.

(ρ–var) There exist $\rho \in [1, 2)$ and $C > 0$ such that for all $0 \le s < t \le 1$ and for every partition $s = t_0 < t_1 < \cdots < t_n = t$ of $[s,t]$

$$\sum_{i,j=1}^{n} |R_{[t_{i-1},t_i]\times[t_{j-1},t_j]}|^\rho \le C|t-s|.$$

(HC) The process X is hypercontractive, i.e. for every $m, n \in \mathbb{N}$ and every $p > 2$ there exists $C_{p,m,n} > 0$ such that for every polynomial $P : \mathbb{R}^n \to \mathbb{R}$ of degree m, for all $i_1, \ldots, i_n \in \{1, \ldots, d\}$, and for all $t_1, \ldots, t_n \in [0,1]$

$$E(|P(X^{i_1}_{t_1}, \ldots, X^{i_n}_{t_n})|^{2p}) \le C_{p,m,n} E(|P(X^{i_1}_{t_1}, \ldots, X^{i_n}_{t_n})|^2)^p.$$

These conditions are taken from [FV10a], where under even more general assumptions it is shown that it is possible to construct the iterated integrals $I(X, \mathrm{d}X)$, and that $I(X, \mathrm{d}X)$ is the limit of $(I(X^n, \mathrm{d}X^n))_{n\in\mathbb{N}}$ under a wide range of smooth approximations $(X^n)_n$ that converge to X.

We first construct the Lévy area $L(X,X)$ for X satisfying (ρ–var) and (HC). Then we give some examples in which these conditions are satisfied.

Lemma 4.6.1. *Assume that the stochastic process $X : [0,1] \to \mathbb{R}$ satisfies (ρ–var). Then we have for all $p \ge -1$ and for all $M, N \in \mathbb{N}$ with $M \le N \le 2^p$ that*

$$\sum_{m_1,m_2=M}^{N} |E(X_{pm_1} X_{pm_2})|^\rho \lesssim (N - M + 1)2^{-p}. \tag{4.36}$$

Proof. Let $p \ge 1$. It suffices to note that

$$\begin{aligned} E(X_{pm_1} X_{pm_2}) &= E\left((X_{t^0_{pm_1}, t^1_{pm_1}} - X_{t^1_{pm_1}, t^2_{pm_1}})(X_{t^0_{pm_2}, t^1_{pm_2}} - X_{t^1_{pm_2}, t^2_{pm_2}})\right) \\ &= \sum_{i_1,i_2=0,1} (-1)^{i_1+i_2} R_{[t^{i_1}_{pm_1}, t^{i_1+1}_{pm_1}]\times[t^{i_2}_{pm_2}, t^{i_2+1}_{pm_2}]}, \end{aligned}$$

and that $\{t^i_{pm} : i = 0,1,2, m = M, \ldots, N\}$ partitions the interval $[(M-1)2^{-p}, N2^{-p}]$.

Now the cases $p = -1$ and $p = 0$ can be included by enlarging the (implicit) constant on the right hand side of (4.36). □

Lemma 4.6.2. *Let $X, Y : [0,1] \to \mathbb{R}$ be independent, centered, continuous processes, both satisfying (ρ–var) for some $\rho \in [1,2]$. Then for all $i, p \ge -1$ and all $q < p$, and for all $0 \le j \le 2^i$*

$$E\left[\left|\sum_{m=0}^{2^p}\sum_{n=0}^{2^q} X_{pm} Y_{qn} \langle 2^{-i}\chi_{ij}, \varphi_{pm}\chi_{qn}\rangle\right|^2\right] \lesssim 2^{(p\vee i)(1/\rho-4)} 2^{(q\vee i)(1-1/\rho)} 2^{-i} 2^{p(4-3/\rho)} 2^{q/\rho}.$$

Proof. Since $p > q$, for every m there exists exactly one $n(m)$, such that $\varphi_{pm}\chi_{qn(m)}$ is not identically zero. Hence, we can apply the independence of X and Y to obtain

$$
\begin{aligned}
&E\left[\left|\sum_{m=0}^{2^p}\sum_{n=0}^{2^q} X_{pm}Y_{qn}\langle 2^{-i}\chi_{ij},\varphi_{pm}\chi_{qn}\rangle\right|^2\right]\\
&\quad\le \sum_{m_1,m_2=0}^{2^p}\left|E(X_{pm_1}X_{pm_2})E(Y_{qn(m_1)}Y_{qn(m_2)})\langle 2^{-i}\chi_{ij},\varphi_{pm_1}\chi_{qn(m_1)}\rangle\langle 2^{-i}\chi_{ij},\varphi_{pm_2}\chi_{qn(m_2)}\rangle\right|.
\end{aligned}
$$

Let us write $M_j := \{m : 0 \le m \le 2^p, \langle\chi_{ij},\varphi_{pm}\chi_{qn(m)}\rangle \neq 0\}$. We also write ρ' for the conjugate exponent of ρ, i.e. $1/\rho+1/\rho' = 1$. Hölder's inequality and Lemma 4.3.10 imply

$$
\begin{aligned}
&\sum_{m_1,m_2\in M_j}\left|E(X_{pm_1}X_{pm_2})E(Y_{qn(m_1)}Y_{qn(m_2)})\langle 2^{-i}\chi_{ij},\varphi_{pm_1}\chi_{qn(m_1)}\rangle\langle 2^{-i}\chi_{ij},\varphi_{pm_2}\chi_{qn(m_2)}\rangle\right|\\
&\quad\lesssim \left(\sum_{m_1,m_2\in M_j}|E(X_{pm_1}X_{pm_2})|^\rho\right)^{1/\rho}\left(\sum_{m_1,m_2\in M_j}|E(Y_{qn(m_1)}Y_{qn(m_2)})|^{\rho'}\right)^{1/\rho'}(2^{-2(p\vee i)+p+q})^2.
\end{aligned}
$$

Write N_j for the set of n for which $\chi_{ij}\chi_{qn}$ is not identically zero. For given $\bar{n}\in N_j$ there are 2^{p-q} numbers $m \in M_j$ for which $n(m) = \bar{n}$. Hence

$$
\begin{aligned}
&\left(\sum_{m_1,m_2\in M_j}|E(Y_{qn(m_1)}Y_{qn(m_2)})|^{\rho'}\right)^{1/\rho'}\\
&\quad\lesssim (2^{2(p-q)})^{1/\rho'}\left(\left(\max_{n_1,n_2\in N_j}|E(Y_{qn_1}Y_{qn_2})|\right)^{\rho'-\rho}\sum_{n_1,n_2\in N_j}|E(Y_{qn_1}Y_{qn_2})|^\rho\right)^{1/\rho'},
\end{aligned}
$$

where we used that $\rho\in[1,2]$ and therefore $\rho'-\rho\ge 0$. Lemma 4.6.1 implies that $(|E(Y_{qn_1}Y_{qn_2})|^{\rho'-\rho})^{1/\rho'} \lesssim 2^{-q(1/\rho-1/\rho')}$. Similarly we apply Lemma 4.6.1 to the sum over n_1, n_2, and we obtain

$$
\begin{aligned}
&(2^{2(p-q)})^{1/\rho'}\left(\left(\max_{n_1,n_2\in N_j}|E(Y_{qn_1}Y_{qn_2})|\right)^{\rho'-\rho}\sum_{n_1,n_2\in N_j}|E(Y_{qn_1}Y_{qn_2})|^\rho\right)^{1/\rho'}\\
&\quad\lesssim (2^{2(p-q)})^{1/\rho'}2^{-q(1/\rho-1/\rho')}(|N_j|2^{-q})^{1/\rho'} = 2^{(q\vee i)/\rho'}2^{-i/\rho'}2^{2p/\rho'}2^{q(-2/\rho'-1/\rho)}\\
&\quad= 2^{(q\vee i)(1-1/\rho)}2^{i(1/\rho-1)}2^{2p(1-1/\rho)}2^{q(1/\rho-2)},
\end{aligned}
$$

where we used that $|N_j| = 2^{(q\vee i)-i}$. Since $|M_j| = 2^{(p\vee i)-i}$, another application of Lemma 4.6.1 yields

$$
\left(\sum_{m_1,m_2\in M_j}|E(X_{pm_1}X_{pm_2})|^\rho\right)^{1/\rho} \lesssim 2^{(p\vee i)/\rho}2^{-i/\rho}2^{-p/\rho}.
$$

The result now follows by combining these estimates:

$$
\begin{aligned}
&E\left[\left|\sum_{m=0}^{2^p}\sum_{n=0}^{2^q} X_{pm}Y_{qn}\langle 2^{-i}\chi_{ij},\varphi_{pm}\chi_{qn}\rangle\right|^2\right] \\
&\lesssim \left(\sum_{m_1,m_2\in M_j}\left|E(X_{pm_1}X_{pm_2})\right|^{\rho}\right)^{1/\rho}\left(\sum_{m_1,m_2\in M_j}\left|E(Y_{qn(m_1)}Y_{qn(m_2)})\right|^{\rho'}\right)^{1/\rho'}(2^{-2(p\vee i)+p+q})^2 \\
&\lesssim \left(2^{(p\vee i)/\rho}2^{-i/\rho}2^{-p/\rho}\right)\left(2^{(q\vee i)(1-1/\rho)}2^{i(1/\rho-1)}2^{2p(1-1/\rho)}2^{q(1/\rho-2)}\right)\left(2^{-4(p\vee i)+2p+2q}\right) \\
&= 2^{(p\vee i)(1/\rho-4)}2^{(q\vee i)(1-1/\rho)}2^{-i}2^{p(4-3/\rho)}2^{q/\rho}.
\end{aligned}
$$

□

Theorem 4.6.3. *Let $X:[0,1]\to\mathbb{R}^d$ be a continuous, centered stochastic process with independent components, and assume that X satisfies (ρ–var) for some $\rho\in[1,2)$ and (HC). Then for every $\alpha\in(0,1/\rho)$ almost surely*

$$\sum_{N\geq 0}\|L(S_NX,S_NX)-L(S_{N-1}X,S_{N-1}X)\|_\alpha<\infty,$$

and therefore the limit $L(X,X)=\lim_{N\to\infty}L(S_NX,S_NX)$ is almost surely an α–Hölder continuous process.

Proof. First note that L is antisymmetric, and in particular the diagonal of the matrix $L(S_NX,S_NX)$ is constantly zero. For $k,\ell\in\{1,\dots,d\}$ with $k\neq\ell$ we have

$$
\begin{aligned}
&\|L(S_NX^k,S_NX^\ell)-L(S_{N-1}X^k,S_{N-1}X^\ell)\|_\alpha \\
&\quad=\left\|\sum_{q=-1}^{N-1}\sum_{m,n}(X^k_{Nm}X^\ell_{qn}-X^k_{qn}X^\ell_{Nm})\int_0^{\cdot}\varphi_{Nm}(s)\mathrm{d}\varphi_{qn}(s)\right\|_\alpha \\
&\quad\leq\sum_{q=-1}^{N-1}\left\|\sum_{m,n}X^k_{Nm}X^\ell_{qn}\int_0^{\cdot}\varphi_{Nm}(s)\mathrm{d}\varphi_{qn}(s)\right\|_\alpha \\
&\qquad+\sum_{q=-1}^{N-1}\left\|\sum_{m,n}X^\ell_{Nm}X^k_{qn}\int_0^{\cdot}\varphi_{Nm}(s)\mathrm{d}\varphi_{qn}(s)\right\|_\alpha
\end{aligned}
$$

Let us argue for the first addend on the right hand side, the arguments for the second addend being identical. Let $r\geq 1$. Using the hypercontractivity condition (HC), we obtain

$$
\begin{aligned}
&\sum_{i=-1}^{\infty}\sum_{j=0}^{2^i}\sum_{N=-1}^{\infty}\sum_{q=-1}^{N-1}P\left(\left|\sum_{m,n}X^\ell_{Nm}X^k_{qn}\langle 2^{-i}\chi_{ij},\varphi_{Nm}\chi_{qn}\rangle\right|>2^{-i\alpha}2^{-N/(2r)}2^{-q/(2r)}\right) \\
&\quad\leq\sum_{i=-1}^{\infty}\sum_{j=0}^{2^i}\sum_{N=-1}^{\infty}\sum_{q=-1}^{N-1}E\left(\left|\sum_{m,n}X^\ell_{Nm}X^k_{qn}\langle 2^{-i}\chi_{ij},\varphi_{Nm}\chi_{qn}\rangle\right|^{2r}\right)2^{i\alpha 2r}2^{N+q}
\end{aligned}
$$

$$\lesssim \sum_{i=-1}^{\infty}\sum_{j=0}^{2^i}\sum_{N=-1}^{\infty}\sum_{q=-1}^{N-1} E\Big(\Big|\sum_{m,n} X^\ell_{Nm}X^k_{qn}\langle 2^{-i}\chi_{ij}, \varphi_{Nm}\chi_{qn}\rangle\Big|^2\Big)^r 2^{i\alpha 2r}2^{N+q}.$$

Now we can apply Lemma 4.6.2 to bound this expression by

$$\begin{aligned}
&\sum_{i=-1}^{\infty}\sum_{j=0}^{2^i}\sum_{N=-1}^{\infty}\sum_{q=-1}^{N-1} \big(2^{(N\vee i)(1/\rho-4)}2^{(q\vee i)(1-1/\rho)}2^{-i}2^{N(4-3/\rho)}2^{q/\rho}\big)^r 2^{i\alpha 2r}2^{N+q}\\
&\quad\lesssim \sum_{i=-1}^{\infty} 2^i \sum_{N=-1}^{i}\sum_{q=-1}^{N-1} 2^{ir(2\alpha-4)}2^{Nr(4-3/\rho+1/r)}2^{qr(1/\rho+1/r)}\\
&\qquad + \sum_{i=-1}^{\infty} 2^i \sum_{N=i+1}^{\infty}\sum_{q=-1}^{i} 2^{ir(2\alpha-1/\rho)}2^{Nr(1/r-2/\rho)}2^{qr(1/\rho+1/r)}\\
&\qquad + \sum_{i=-1}^{\infty} 2^i \sum_{N=i+1}^{\infty}\sum_{q=i+1}^{N-1} 2^{ir(2\alpha-1)}2^{Nr(1/r-2/\rho)}2^{qr(1+1/r)}\\
&\quad\lesssim \sum_{i=-1}^{\infty} 2^{ir(2\alpha+3/r-2/\rho)} + \sum_{i=-1}^{\infty}\sum_{N=i+1}^{\infty} 2^{ir(2\alpha+2/r)}2^{Nr(1/r-2/\rho)}\\
&\qquad + \sum_{i=-1}^{\infty}\sum_{N=i+1}^{\infty} 2^{ir(2\alpha+1/r-1)}2^{Nr(1+2/r-2/\rho)}.
\end{aligned}$$

Note that for all $r \geq 1$ we have $1/r - 2/\rho < 0$, because $\rho < 2$. Therefore, the sum over N in the second addend on the right hand side converges. If now we choose $r > 1$ large enough so that $1+3/r-2/\rho < 0$ (and then also $2\alpha+3/r-2/\rho < 0$), then all three series on the right hand side are finite. Hence, the Borel-Cantelli lemma implies the existence of $C(\omega) > 0$, such that for almost all $\omega \in \Omega$ and for all N, q, i, j

$$\Big|\sum_{m,n} X^\ell_{Nm}(\omega)X^k_{qn}(\omega)\langle 2^{-i}\chi_{ij}, \varphi_{Nm}\chi_{qn}\rangle\Big| \leq C(\omega)2^{-i\alpha}2^{-N/(2r)}2^{-q/(2r)}.$$

From here it is straightforward to see that for these ω we have

$$\sum_{N=0}^{\infty} \|L(S_N X(\omega), S_N X(\omega)) - L(S_{N-1}X(\omega), S_{N-1}X(\omega))\|_\alpha < \infty.$$

□

Example 4.6.4. Condition (HC) is satisfied by all Gaussian processes. More generally, it is satisfied by every process "living in a fixed Gaussian chaos". Slightly oversimplifying things, this is the case if X is given by polynomials of fixed degree and iterated integrals of fixed order with respect to a Gaussian reference process. For details about hypercontractivity for random variables living in a fixed Gaussian chaos, we refer to [FV10b], Appendix D.4.

Prototypical examples of processes living in a fixed chaos are Hermite processes. They are defined for $H \in (1/2, 1)$ and $k \in \mathbb{N}$, $k \geq 1$ as

$$Z_t^{k,H} = C(H,k) \int_{\mathbb{R}^k} \left(\int_0^t \prod_{i=1}^k (s - y_i)_+^{-\left(\frac{1}{2} + \frac{1-H}{k}\right)} \mathrm{d}s \right) \mathrm{d}B_{y_1} \dots \mathrm{d}B_{y_k},$$

where $(B_y)_{y \in \mathbb{R}}$ is a standard Brownian motion, and $C(H,k)$ is a normalization constant. In particular, $Z^{k,H}$ lives in the Wiener chaos of order k. The covariance of $Z^{k,H}$ is

$$E(Z_s^{k,H} Z_t^{k,H}) = \frac{1}{2} \left(t^{2H} + s^{2H} + |t-s|^{2H} \right)$$

Since $Z^{1,H}$ is Gaussian, it is exactly the fractional Brownian motion with Hurst parameter H. For $k = 2$ we obtain the Rosenblatt process. For further details about Hermite processes see [PT11]. However, we should point out that it follows from Kolmogorov's continuity criterion that $Z^{k,H}$ is α–Hölder continuous for every $\alpha < H$. Since $H \in (1/2, 1)$, Hermite processes are amenable to Young integration, and it is trivial to construct $L(Z^{k,H}, Z^{k,H})$.

Example 4.6.5. Condition (ρ–var) is satisfied by Brownian motion with $\rho = 1$. More generally it is satisfied by the fractional Brownian motion with Hurst index $H > 1/4$. In that case we have $\rho = 1/(2H)$. It is also satisfied by the fractional Brownian bridge with Hurst index $H > 1/4$. A general criterion that implies condition (ρ–var) is the one of Coutin and Qian [CQ02]: If $E(|X_{s,t}^i|^2) \lesssim |t-s|^{2H}$ and $|E(X_{s,s+h}^i X_{t,t+h}^i)| \lesssim |t-s|^{2H-2} h^2$ for $i = 1, \dots, d$, then (ρ–var) is satisfied for $\rho = 1/(2H)$. For proofs of these claims and for further examples see [FV10b], Section 15.2.

4.6.2. Continuous martingales

Here we assume that $(X_t)_{t \in [0,1]}$ is a d–dimensional continuous martingale. Of course in that case it is no problem to construct the Itô iterated integrals $I^{\text{Itô}}(X, \mathrm{d}X)$ of X. But in order to apply Corollary 4.5.8, we still need the pathwise convergence of $I_k^{\text{Itô}}(X, \mathrm{d}X)$ to $I^{\text{Itô}}(X, \mathrm{d}X)$, and we need to prove the uniform Hölder continuity along the dyadics of the approximating integrals. We are not claiming the greatest generality and work under rather restrictive conditions. The main example that we have in mind is Brownian motion.

Recall that for a d–dimensional semimartingale $X = (X^1, \dots, X^d)$, the quadratic variation is defined as $[X] = ([X^i, X^j])_{1 \leq i,j \leq d}$. We also write $X_s X_{s,t} := (X_s^i X_{s,t}^j)_{1 \leq i,j \leq d}$ for $s, t \in [0,1]$.

Theorem 4.6.6. *Let $X = (X^1, \dots, X^d)$ be a d–dimensional continuous martingale indexed by $[0,1]$. Assume that there exists $p \geq 2$ and $\beta > 1/3 + 1/p$, such that $p\beta > 7/2$, and such that*

$$E(|[X]_{s,t}|^p) \lesssim |t-s|^{2p\beta} \tag{4.37}$$

for all $s, t \in [0,1]$. Then $I_k^{\text{Itô}}(X, \mathrm{d}X)$ almost surely converges uniformly to $I^{\text{Itô}}(X, \mathrm{d}X)$.

Furthermore, we have for all $\alpha \in (0, \beta - 1/p)$ that $X \in \mathcal{C}^\alpha$ and that almost surely

$$\sup_k \sup_{0 \le \ell < \ell' \le 2^k} \frac{|I_k^{\text{Itô}}(X, \mathrm{d}X)_{\ell 2^{-k}, \ell' 2^{-k}} - X_{\ell 2^{-k}} X_{\ell 2^{-k}, \ell' 2^{-k}}|}{|(\ell' - \ell) 2^{-k}|^{2\alpha}} < \infty. \tag{4.38}$$

In particular, X almost surely satisfies all the conditions of Corollary 4.5.8.

Proof. The Hölder continuity of X follows from Kolmogorov's continuity criterion, because by the Burkholder-Davis-Gundy inequality and using (4.37) we have

$$E(|X_{s,t}|^{2p}) \lesssim \sum_{i=1}^d E(|X^i_{s,t}|^{2p}) \lesssim \sum_{i=1}^d E(|[X^i]_{s,t}|^p) \lesssim E(|[X]_{s,t}|^p) \lesssim |t-s|^{2p\beta}.$$

Kolmogorov's continuity criterion now shows that $X \in \mathcal{C}^\alpha$ for all $\alpha \in (0, \beta - 1/(2p))$ and in particular for all $\alpha \in (0, \beta - 1/p)$. Since we will need it below, let us also study the regularity of the Itô integral $I^{\text{Itô}}(X, \mathrm{d}X)$: A similar application of the Burkholder-Davis-Gundy inequality implies that

$$E(|I^{\text{Itô}}(X, \mathrm{d}X)_{s,t} - X_s X_{s,t}|^p) \lesssim E\Big(\Big|\int_s^t |X_r - X_s|^2 \mathrm{d}|[X]|_s\Big|^{\frac{p}{2}}\Big).$$

We apply Jensen's inequality (here we need $p \ge 2$) to obtain

$$\begin{aligned} E\Big(\Big|\int_s^t |X_r - X_s|^2 \mathrm{d}|[X]|_s\Big|^{\frac{p}{2}}\Big) &= E\Big(\Big|\int_s^t |[X]|_{s,t} |X_r - X_s|^2 \frac{\mathrm{d}|[X]|_s}{|[X]|_{s,t}}\Big|^{\frac{p}{2}}\Big) \\ &\lesssim E\Big(\int_s^t |[X]|_{s,t}^{\frac{p}{2}-1} |X_r - X_s|^p \mathrm{d}|[X]|_s\Big), \end{aligned}$$

where we set $0/0 = 0$. Now Cauchy-Schwarz's and then Burkholder-Davis-Gundy's inequalities yield

$$\begin{aligned} E\Big(\int_s^t |[X]|_{s,t}^{\frac{p}{2}-1} |X_r - X_s|^p \mathrm{d}|[X]|_s\Big) &\lesssim E\Big(\sup_{r \in [s,t]} |X_r - X_s|^p |[X]|_{s,t}^{\frac{p}{2}}\Big) \\ &\le \sqrt{E\Big(\sup_{r \in [s,t]} |X_r - X_s|^{2p}\Big)} \sqrt{E(|[X]|_{s,t}^p)} \\ &\lesssim E(|[X]_{s,t}|^p) \lesssim |t-s|^{2p\beta}. \end{aligned}$$

The Kolmogorov criterion for rough paths, Theorem 3.1 of [FH13], then implies that almost surely

$$|I^{\text{Itô}}(X, \mathrm{d}X)_{s,t} - X_s X_{s,t}| \lesssim |t-s|^{2\alpha} \tag{4.39}$$

for all $\alpha \in (0, \beta - 1/p)$.

Let us continue with the proof of our claim. We need to show that $I_k^{\text{Itô}}(X, \mathrm{d}X)$ almost surely converges uniformly to $I^{\text{Itô}}(X, \mathrm{d}X)$, and that the uniform Hölder condition (4.38) holds. Using similar arguments as before, we can show that

$$
\begin{aligned}
&E(|I^{\text{Itô}}(X,\mathrm{d}X)_{\ell 2^{-k},\ell' 2^{-k}} - I_k^{\text{Itô}}(X,\mathrm{d}X)_{\ell 2^{-k},\ell' 2^{-k}}|^p) \\
&\qquad = E\Big(\Big|\int_{\ell 2^{-k}}^{\ell' 2^{-k}} \sum_{m=\ell}^{\ell'-1} \mathbf{1}_{[m2^{-k},(m+1)2^{-k})}(r) X_{m2^{-k},r}\mathrm{d}X_s\Big|^p\Big) \\
&\qquad \lesssim E\Big(|[X]|_{\ell 2^{-k},\ell' 2^{-k}}^{\frac{p}{2}-1} \int_{\ell 2^{-k}}^{\ell' 2^{-k}} \Big|\sum_{m=\ell}^{\ell'-1} \mathbf{1}_{[m2^{-k},(m+1)2^{-k})}(r)|X_{m2^{-k},r}|^2\Big|^{\frac{p}{2}}\mathrm{d}|[X]|_s\Big).
\end{aligned}
$$

Since the terms in the sum all have disjoint support, we can pull the exponent $p/2$ into the sum, from where we conclude using once again Cauchy-Schwarz's and Burkholder-Davis-Gundy's inequalities

$$
\begin{aligned}
&E\Big(|[X]|_{\ell 2^{-k},\ell' 2^{-k}}^{\frac{p}{2}-1} \int_{\ell 2^{-k}}^{\ell' 2^{-k}} \sum_{m=\ell}^{\ell'-1} \mathbf{1}_{[m2^{-k},(m+1)2^{-k})}(r)|X_{m2^{-k},r}|^p\mathrm{d}|[X]|_s\Big) \\
&\qquad \lesssim \sqrt{E\Bigg(\sup_{r\in[s,t]}\Big|\sum_{m=\ell}^{\ell'-1} \mathbf{1}_{[m2^{-k},(m+1)2^{-k})}(r)|X_{m2^{-k},r}|^p\Big|^2\Bigg)}\sqrt{E(|[X]|_{\ell 2^{-k},\ell' 2^{-k}}^p)} \\
&\qquad \lesssim \sqrt{\sum_{m=\ell}^{\ell'-1} E(|[X]_{m2^{-k},(m+1)2^{-k}}|^p)}\sqrt{E(|[X]_{\ell 2^{-k},\ell' 2^{-k}}|^p)} \\
&\qquad \lesssim \sqrt{(\ell'-\ell)(2^{-k})^{2p\beta}}\sqrt{|(\ell'-\ell)2^{-k}|^{2p\beta}} = (\ell'-\ell)^{\frac{1}{2}+p\beta}2^{-k2p\beta}.
\end{aligned}
$$

Hence, we obtain for $\alpha \in \mathbb{R}$ that

$$
\begin{aligned}
&P\left(|I^{\text{Itô}}(X,\mathrm{d}X)_{\ell 2^{-k},\ell' 2^{-k}} - I_k^{\text{Itô}}(X,\mathrm{d}X)_{\ell 2^{-k},\ell' 2^{-k}}| > |(\ell'-\ell)2^{-k}|^{2\alpha}\right) \\
&\qquad \lesssim \frac{(\ell'-\ell)^{\frac{1}{2}+p\beta}2^{-k2p\beta}}{(\ell'-\ell)^{2p\alpha}2^{-k2p\alpha}} = (\ell'-\ell)^{\frac{1}{2}+p\beta-2p\alpha}2^{-k2p(\beta-\alpha)}.
\end{aligned}
$$

If we set $\alpha = \beta - 1/(2p) - \varepsilon$ for sufficiently small $\varepsilon > 0$, then

$$1/2 + p\beta - 2p\alpha = 3/2 - p\beta - 2p\varepsilon.$$

Now by assumption, $p\beta > 7/2$, and therefore we can find $\alpha \in (0, \beta - 1/(2p))$ such that

$$1/2 + p\beta - 2p\alpha < -2. \tag{4.40}$$

Estimating the double sum by a double integral, we easily see that

$$\sum_{\ell=1}^{2^k}\sum_{\ell'=\ell+1}^{2^k}(\ell'-\ell)^\gamma \lesssim 2^k$$

for $\gamma < -2$. Therefore, we have for $\alpha \in (0, \beta - 1/(2p))$ satisfying (4.40)

$$\sum_{\ell=1}^{2^k} \sum_{\ell'=\ell+1}^{2^k} P\left(|I^{\text{Itô}}(X, \mathrm{d}X)_{\ell 2^{-k}, \ell' 2^{-k}} - I_k^{\text{Itô}}(X, \mathrm{d}X)_{\ell 2^{-k}, \ell' 2^{-k}}| > |(\ell' - \ell)2^{-k}|^{2\alpha}\right)$$
$$\lesssim 2^k 2^{-k2p(\beta-\alpha)}.$$

Since $\alpha < \beta - 1/(2p)$, this is summable in k, and therefore Borel-Cantelli's lemma implies that almost surely

$$\sup_k \sup_{0 \le \ell < \ell' \le 2^k} \frac{|I^{\text{Itô}}(X, \mathrm{d}X)_{\ell 2^{-k}, \ell' 2^{-k}} - I_k^{\text{Itô}}(X, \mathrm{d}X)_{\ell 2^{-k}, \ell' 2^{-k}}|}{|(\ell' - \ell)2^{-k}|^{2\alpha}} < \infty. \tag{4.41}$$

We only proved this for α close enough to $\beta - 1/(2p)$, but of course then it also holds for all $\alpha' \le \alpha$, since $(\ell' - \ell)2^{-k} \le 1$. The estimate (4.38) now follows by combining (4.39) and (4.41). The uniform convergence of $I_k^{\text{Itô}}(X, \mathrm{d}X)$ to $I^{\text{Itô}}(X, \mathrm{d}X)$ follows from (4.41) in combination with the Hölder continuity of X. □

Example 4.6.7. The conditions of Theorem 4.6.6 are satisfied by the d–dimensional standard Brownian motion. Here we can take $\beta = 1/2$, and p can be taken arbitrarily large. More generally, the conditions are satisfied by all Itô martingales of the form $X_t = X_0 + \int_0^t \sigma_s dW_s$, as long as σ satisfies

$$E\left(\sup_{s \in [0,1]} |\sigma_s|^{2p}\right) < \infty$$

for some $p > 7$. In that case we can take $\beta = 1/2$.

5. Paracontrolled distributions and applications to SPDEs

Here we build on the ideas developed in Chapter 4, to develop an extension of rough path theory that works for functions of a multi dimensional index variable. We apply this to solve two nonlinear SPDEs, for which previously it was not well understood how to make sense of the nonlinearity.

5.1. Introduction

One way of interpreting the rough path integral of Chapters 3 and 4, but also the Itô and Stratonovich integral, is as a way of defining products of tempered distributions. Conversely, if we are able to multiply suitable tempered distributions with each other, then we can integrate the result in time to obtain a "stochastic" integral. Schwartz's theory of distributions gives a robust framework for defining linear operations on irregular generalized functions. But when trying to handle nonlinear operations, we quickly run into problems. For example, in Schwartz' theory it is not possible to define the product $\varphi(W_t)\dot{W}_t$, where φ is a smooth function, W is a Brownian motion, and $\dot{W}$ its derivative. But using for example Itô's stochastic integral, the product can be defined as

$$\varphi(W_t)\dot{W}_t := \partial_t \int_0^t \varphi(W_s)\mathrm{d}W_s.$$

The Itô integral requires an "arrow of time" (a filtration and adapted integrands), a probability measure (it is defined as L^2–limit), and L^2–orthogonal increments of the integrator (the integrator needs to be a (semi-) martingale). If one or several of these assumptions are violated, then the rough path integral can be a useful alternative. For example, we saw in Chapter 3 that the rough path integral can be applied in a model free approach to finance, where no probability measure is given. In Chapter 4 we constructed a pathwise integral for, among other processes, fractional Brownian motion, which is not a semimartingale.

The "arrow of time" condition is typically violated if the index is a spatial variable and not a temporal variable. It is a remarkable observation of Hairer [Hai11], that in such cases sometimes the rough path integral can be used to handle nonlinear operations. In [Hai11], Hairer studies the following Burgers type SPDE:

$$\partial_t u(t,x) = \Delta u(t,x) + G(u(t,x))\partial_x u(t,x) + \dot{W}(t,x),$$

where $(t,x) \in [0,T] \times [-\pi,\pi]$, and where $\dot{W}(t,x)$ is a space-time white noise. This problem is motivated by insights from path sampling, where one can formally derive the equation as an SPDE whose invariant measure describes the law of a certain conditioned diffusion. For every fixed $t > 0$, the solution v to

$$\partial_t v(t,x) = \Delta v(t,x) + \dot{W}(t,x)$$

is α–Hölder continuous in space for every $\alpha < 1/2$. We would expect u to have the same regularity as v. But then the product $G(u(t,x))\partial_x u(t,x)$ is ill-defined: we expect $G(u) \in C^\alpha$ and $\partial_x u \in C^{\alpha-1}$. Since $\alpha < 1/2$, the sum of the regularities of $G(u)$ and $\partial_x u$ is negative, and therefore their product cannot be defined using classical analytic methods (see Section 5.2 below). Since x is a spatial variable, there is no natural filtration associated to the problem, and the integral cannot be treated with Itô's theory. But Hairer showed that the rough path integral can be used to define the product, and that with this definition, the SPDE has a unique solution. Furthermore, this solution is the limit as $\varepsilon \to 0$ of the solutions u_ε to

$$\partial_t u_\varepsilon(t,x) = \Delta u_\varepsilon(t,x) + G(u_\varepsilon(t,x))\partial_x u_\varepsilon(t,x) + \dot{W}_\varepsilon(t,x),$$

where $\dot{W}_\varepsilon(t,x)$ are suitable smooth approximations that converge to $\dot{W}(t,x)$.

Since x is a spatial variable, it is natural to ask about extensions of Hairer's approach to higher dimensions. In one dimension, all techniques presented above made use of integrals to define products. In that setting, defining the product $G(u)\partial_x u$ is essentially equivalent to defining the integral $\int G(u)\mathrm{d}_x u$, because in one dimension the integral is an "antiderivative", i.e. an inverse operation to differentiation. In the multidimensional case, there usually exists no antiderivative, and therefore the link between integrals and products is not so clear. In other words, for multidimensional index variables it is more natural to work directly on the level of products, rather than working on the level of integrals.

Here, we adapt the techniques of Chapter 4 to develop an extension of rough path theory that operates on the level of products, and that works for arbitrary index dimensions. More precisely, we use the Littlewood-Paley decomposition of tempered distributions, and not the Schauder decomposition. We then combine Bony's paraproduct, a concept from functional analysis, with ideas from the theory of controlled rough paths, in order to develop an algebraic theory for certain types of distributions that we call *controlled*. This is similar to the construction of Chapter 4, but since we do not use any integrals to define our products, the approach presented here works in any index dimension and constitutes a flexible generalization of rough path theory that allows to handle problems which were well out of reach with previously known methods.

To exemplify the applicability of our ideas, we will consider two SPDEs for which previously it was not known how to describe solutions:

1. The first example is the generalization of Hairer's Burgers type SPDE to a higher dimensional spatial index variable. While this equation is maybe not very relevant for applications, it is a perfect test bed for our techniques. We consider the following

equation on the d–dimensional torus $\mathbb{T}^d := [-\pi,\pi]^d := (\mathbb{R}/2\pi\mathbb{Z})^d$ with periodic boundary conditions:

$$\partial_t u(t,x) = -Au(t,x) + G(u(t,x))\mathrm{D}_x u(t,x) + \dot{W}(t,x).$$

Here $u : \mathbb{R}_+ \times \mathbb{T}^d \to \mathbb{R}^n$ is a vector valued function, $-A = -(-\Delta)^\sigma$ is the fractional Laplacian with $\sigma \geq 1$ such that $\sigma > 1/3 + d/2$, the Gaussian noise $\dot{W}$ is white in space and time with values in $\mathbb{R}^n$, and D_x denotes the spatial derivative. Moreover, $G : \mathbb{R}^n \to \mathcal{L}(\mathcal{L}(\mathbb{T}^d, \mathbb{R}^n), \mathbb{R}^n)$ is a smooth field of linear transformations.

2. The second example is a nonlinear version of the parabolic Anderson model,

$$\partial_t u(t,x) = \Delta u(t,x) + F(u(t,x))\dot{W}(x),$$

where $u : \mathbb{R}_+ \times \mathbb{T}^2 \to \mathbb{R}$, we consider a white noise potential $\dot{W}$ which does not depend on time, and $F : \mathbb{R} \to \mathbb{R}$ is a smooth function.

In both cases we will exhibit a space of controlled distribution where the equations are well posed (in a suitable sense), and admit a local solution.

Below we give some references to other articles that apply rough path techniques to SPDEs. In Section 5.2 we recall the main concepts of Littlewood-Paley theory and of Bony's paraproduct, and we present our basic ideas. Section 5.3 develops the paradifferential calculus of controlled distributions. In Section 5.4 we solve Burgers equation driven by white noise, and in Section 5.5 we solve a nonlinear version of the parabolic Anderson model.

It may be helpful to go through Section 4.2.2 in Chapter 4 before continuing to read, to get a basic overview on rough paths.

Relevant literature

Even if only implicitly, the relevance of paraproducts to rough paths has been remarked before in the work of Unterberger on the renormalization of rough paths [Unt10a, Unt10b], where it is referred to as "Fourier normal-ordering", and in the related work of Nualart and Tindel [NT11].

Before we developed the paraproduct approach, there were several other papers that applied rough path ideas to treat SPDEs and more generic stochastic processes. But they all relied on special features of the problem at hand in order to be able to apply the integration theory provided by the rough path machinery:

Deya, Gubinelli, Lejay, and Tindel [GLT06, Gub12, DGT12] deal with SPDEs of the form

$$\partial_t u(t,x) = \Delta u(t,x) + \sigma(u(t,x))\eta(t,x),$$

where $x \in [-\pi,\pi]$, the noise η is a space-time Gaussian distribution (for example white in time and colored in space), and σ is some nonlinear coefficient. They interpret this as an evolution equation (in time), taking values in a space of functions (with respect to the space variable). They extend the rough path machinery to handle the convolution integrals that appear when applying the heat flow to the noise.

Friz, Caruana, Diehl, and Oberhauser [CF09, CFO11, FO11, DF12] deal with fully nonlinear stochastic PDEs with a special structure. Among others, of the form

$$\partial_t u(t,x) = F(u, \partial_x u, \partial_x^2 u) + \sigma(t,x)\partial_x u(t,x)\eta(t),$$

where the spatial index x can be multidimensional, but the noise η only depends on time. Such an SPDE can be reinterpreted as a standard PDE with random coefficients via a change of variables involving the flow of the stochastic characteristics associated to σ.

Teichmann [Tei11] studies semilinear SPDEs of the form

$$\partial_t u(t,x) = Au(t,x) + \sigma(u(t,x))\eta(t,x),$$

where A is a suitable linear operator, in general unbounded. The SPDE is transformed into an SDE with bounded coefficients by applying a suitable transformation based on the (semi-) group generated by A. This is called the method of the moving frame.

Bessaih, Gubinelli, and Russo [BGR05] and Brzezniak, Gubinelli, and Neklyudov [BGN10] consider a PDE motivated by the description of the motion of a vortex line in an incompressible fluid. Rough path theory allows to make sense of this equation with random irregular initial vortex configurations. Here, the irregularities appear along the direction of the (one dimensional) variable parameterizing the vortex line.

Hairer, Maas, and Weber [Hai11, HW13, Hai13b, HMW12] build on the insight of Hairer that rough path theory allows to make sense of SPDEs that are ill-defined in standard function spaces due to spatial irregularities. Hairer and Weber [HW13] extend the Burgers type SPDE that we presented in the introduction to the case of multiplicative noise. Hairer, Maas, and Weber [HMW12] study approximations to this equation, where they discretize the spatial derivative as $\partial_x u(t,x) \simeq 1/\varepsilon(u(t,x+\varepsilon) - u(t,x))$. They show that in the limit $\varepsilon \to 0$, this introduces a Stratonovich type corrector term to the equation. Hairer [Hai13b] solves the KPZ equation, an SPDE of one spatial index variable that describes the random growth of an interface. This equation was introduced by Kardar, Parisi, and Zhang [KPZ86], and before Hairer's work, it could only be solved by applying a spatial transform (the Cole-Hopf transform) that linearizes the equation.

In all these works, the intrinsic one dimensional nature of rough path theory severely limits possible improvements or applications to other contexts. To the best of our knowledge, the first attempt to remove these limitations is the still unpublished work by Chouk and Gubinelli, extending rough path theory to handle the Brownian sheet (a two-parameter stochastic process akin to Brownian motion).

In the recent paper [Hai13a] however, Hairer has introduced a "theory of regularity structures", that fundamentally redefines the notion of regularity. Hairer's theory is also inspired by the theory of controlled rough paths, and also extends it to functions of a multidimensional index variable. The crucial insight is that the regularity of the solution to an equation driven by - say - Gaussian space time white noise should not be described in the classical way. Usually we say that a function is smooth if it can be approximated around every point by a polynomial of a given degree (the Taylor polynomial). In other words, smooth functions locally look like polynomials. Since the solution to an SPDE does not look like a polynomial at all, this is not the correct way of describing its regularity. We rather expect that the solution locally looks like the driving noise (more precisely like the noise convoluted with the Green kernel of the linear part of the equation; in the case of ODEs this is the time integral of the white noise, i.e. the Brownian motion). Therefore, in Hairer's theory a function is called smooth if it can locally be well approximated by this convolution (and higher order terms depending on the noise). This notion of smoothness induces a natural topology in which the solutions to semilinear SPDEs depend continuously on the driving signal. Hairer's approach is very general, and allows to handle more complicated problems than the ones we treat below. The merit of our approach is its relative simplicity, the fact that it seems to be very adaptable so that it can be easily modified to treat problems with a different structure, and that we make the connection between Fourier analysis and rough paths – although Hairer also uses wavelets, to show that for every consistent "generalized Taylor expansion" in terms of polynomials and the noise, there exists a tempered distribution which has this expansion.

5.2. Preliminaries

Littlewood-Paley theory

Littlewood-Paley theory allows for an elegant way of characterizing the regularity of functions and distributions. Compared to the characterization of regularity based on increments, the Littlewood-Paley approach has the advantage that it also applies to distributions that are not functions.

The space of real valued infinitely differentiable functions of compact support is denoted by $\mathcal{D}(\mathbb{R}^d)$ or $\mathcal{D}$. The space of Schwartz functions, which consists of the smooth functions all of whose derivatives are rapidly decreasing, is denoted by $\mathcal{S}(\mathbb{R}^d)$ or $\mathcal{S}$. Its dual, the space of tempered distributions, is $\mathcal{S}'(\mathbb{R}^d)$ or $\mathcal{S}'$. If u is a vector of n tempered distributions on $\mathbb{R}^d$, then we write $u \in \mathcal{S}'(\mathbb{R}^d, \mathbb{R}^n)$. The Fourier transform is defined with the normalization

$$\mathcal{F}u(\xi) := \widehat{u}(\xi) := \int_{\mathbb{R}^d} e^{-\imath\langle \xi, x\rangle} u(x)\mathrm{d}x,$$

so that the inverse Fourier transform is given by $\mathcal{F}^{-1}u(\xi) = (2\pi)^{-d}\mathcal{F}u(-\xi)$. Recall that for any $u \in \mathcal{S}'$ and $\varphi \in \mathcal{S}$ we have

$$\mathcal{F}u\mathcal{F}\varphi = \mathcal{F}(u * \varphi) \qquad \text{and thus} \qquad \mathcal{F}^{-1}(\mathcal{F}u\mathcal{F}\varphi) = u * \varphi, \tag{5.1}$$

see for example Proposition 1.24 of [BCD11].

An *annulus* is a set of the form $\mathcal{A} = \{x \in \mathbb{R}^d : a \leq |x| \leq b\}$ for some $0 < a < b$. A *ball* is a set of the form $\mathcal{B} = \{x \in \mathbb{R}^d : |x| \leq b\}$.

Definition 5.2.1. A pair $(\chi, \rho) \in \mathcal{D}^2$ of nonnegative radial functions is called *dyadic partition of unity* if

1. the support of χ is contained in a ball and the support of ρ is contained in an annulus;
2. $\chi(\xi) + \sum_{j\geq 0} \rho(2^{-j}\xi) = 1$ for all $\xi \in \mathbb{R}^d$;
3. $\mathrm{supp}(\chi) \cap \mathrm{supp}(\rho(2^{-j}\cdot)) = \emptyset$ for $j \geq 1$ and $\mathrm{supp}(\rho(2^{-i}\cdot)) \cap \mathrm{supp}(\rho(2^{-j}\cdot)) = \emptyset$ for $|i-j| > 1$.

In that case we also write $\rho_{-1} := \chi$ and $\rho_j := \rho(2^{-j}\cdot)$ for $j \geq 0$.

For the existence of dyadic partitions of unity see [BCD11], Proposition 2.10.

If φ is a smooth function, such that φ and all its derivatives are at most of polynomial growth at infinity, then we define $\varphi(\mathrm{D})u := \mathcal{F}^{-1}(\varphi \mathcal{F} u)$ for any $u \in \mathcal{S}'$. More generally we define $\varphi(\mathrm{D})u$ in this way whenever the right hand side makes sense. Operators of the form $\varphi(\mathrm{D})$ are called *Fourier multipliers*. The Littlewood-Paley blocks are now defined as

$$\Delta_{-1} u := \chi(\mathrm{D})u = \rho_{-1}(\mathrm{D})u \qquad \text{and for } j \geq 0: \qquad \Delta_j u := \rho_j(\mathrm{D})u.$$

Then $\Delta_{-1}u = \tilde{h} * u$ and for $j \geq 0$ we have $\Delta_j u = h_j * u$, where $\tilde{h} = \mathcal{F}^{-1}\chi$ and $h_j = \mathcal{F}^{-1}\rho_j$. In particular, $\Delta_j u$ is an infinitely differentiable function for every $j \geq -1$. We also use the notation

$$S_i f := \sum_{j \leq i-1} \Delta_j f.$$

It is not hard to see that $u = \sum_{j\geq -1} \Delta_j u = \lim_{i\to\infty} S_i u$ for every $u \in \mathcal{S}'$, where the convergence holds in the topology of $\mathcal{S}'$.

For $N \in \mathbb{N}$ we define the set $A_N := \{(i,j) \in \{-1,0,1,\dots\}^2 : i \leq j + N\}$. The notation $i \lesssim j$ then means that there exists $N \in \mathbb{N}$ such that $(i,j) \in A_N$ for all values of i and j under consideration. Similarly $i \gtrsim j$ means $j \lesssim i$, and $i \sim j$ means $i \lesssim j$ and $j \lesssim i$. This notation will only be applied to index variables of Littlewood-Paley blocks.

For $\alpha \in \mathbb{R}$, the Hölder-Besov space C^α is given by $C^\alpha := B^\alpha_{\infty,\infty}(\mathbb{R}^d, \mathbb{R}^n)$, where for $p, q \in [1,\infty]$ we define the norm $\|\cdot\|_{B^\alpha_{p,q}}$ and the space $B^\alpha_{p,q}$ as

$$B^\alpha_{p,q} := B^\alpha_{p,q}(\mathbb{R}^d, \mathbb{R}^n) := \left\{ u \in \mathcal{S}'(\mathbb{R}^d, \mathbb{R}^n) : \|u\|_{B^\alpha_{p,q}} := \left(\sum_{j\geq -1} (2^{j\alpha} \|\Delta_j u\|_{L^p})^q \right)^{\frac{1}{q}} < \infty \right\},$$

with the usual interpretation as ℓ^∞ norm in case $q = \infty$. The $\|\cdot\|_{L^p}$ norm is taken with respect to Lebesgue measure on $\mathbb{R}^d$. While the norm $\|\cdot\|_{B^\alpha_{p,q}}$ depends on the dyadic partition of unity (χ, ρ), the space $B^\alpha_{p,q}$ does not, and any other dyadic partition of unity corresponds to an equivalent norm. We will usually write $\|\cdot\|_\alpha$ instead of $\|\cdot\|_{B^\alpha_{\infty,\infty}}$.

If $\alpha \in (0,\infty) \setminus \mathbb{N}$, then C^α is the space of $\lfloor \alpha \rfloor$ times differentiable functions, whose partial derivatives up to order $\lfloor \alpha \rfloor$ are bounded, and whose partial derivatives of order $\lfloor \alpha \rfloor$ are $(\alpha - \lfloor \alpha \rfloor)$-Hölder continuous, see p. 99 of [BCD11]. For $m \in \mathbb{N}$, the Hölder-Besov space C^m is strictly larger than C^m_b, the space of m times continuously differentiable functions, bounded with bounded derivatives.

We will use without comment that $\|\cdot\|_\alpha \le \|\cdot\|_\beta$ for $\alpha \le \beta$, that $\|\cdot\|_{L^\infty} \lesssim \|\cdot\|_\alpha$ for $\alpha > 0$, and that $\|\cdot\|_\alpha \lesssim \|\cdot\|_{L^\infty}$ for $\alpha \le 0$. For $\alpha < 0$ and $u \in C^\alpha$ we also use that $\|S_j u\|_{L^\infty} \lesssim 2^{j\alpha}\|u\|_\alpha$.

The following Bernstein inequalities are for example useful for calculating the regularity of derivatives.

Lemma 5.2.2 (Lemma 2.1 of [BCD11])**.** *For $k \in \mathbb{N}$, for $u \in \mathcal{S}'$, and for $1 \le p \le q \le \infty$ we have*

$$\max_{\eta \in \mathbb{N}^d : |\eta| = k} \|\partial^\eta \Delta_j u\|_{L^q} = \max_{\eta \in \mathbb{N}^d : |\eta| = k} \|\Delta_j \partial^\eta u\|_{L^q} \lesssim 2^{jk + d(\frac{1}{p} - \frac{1}{q})} \|\Delta_j u\|_{L^p}$$

for all $j \ge -1$, and for all $j \ge 0$ we moreover have

$$2^{jk} \|\Delta_j u\|_{L^p} \lesssim \max_{\eta \in \mathbb{N}^d : |\eta| = k} \|\partial^\eta \Delta_j u\|_{L^p}.$$

We will often use the following criterion to show that a function is in a certain Hölder-Besov space:

Lemma 5.2.3 (Lemma 2.69 and 2.84 of [BCD11])**.** *Let $\mathcal{A}$ be an annulus and let $\mathcal{B}$ be a ball.*

1. *Let $\alpha \in \mathbb{R}$, and let (u_j) be a sequence of smooth functions such that $\mathcal{F}u_j$ has its support in $2^j\mathcal{A}$, and such that $\|u_j\|_{L^\infty} \lesssim 2^{-j\alpha}$. Then*
$$u = \sum_{j \ge -1} u_j \in C^\alpha \qquad \textit{and} \qquad \|u\|_\alpha \lesssim \sup_{j \ge -1} \{2^{j\alpha}\|u_j\|_{L^\infty}\}.$$

2. *Let $\alpha > 0$, and let (u_j) be a sequence of smooth functions such that $\mathcal{F}u_j$ has its support in $2^j\mathcal{B}$, and such that $\|u_j\|_{L^\infty} \lesssim 2^{-j\alpha}$. Then*
$$u = \sum_{j \ge -1} u_j \in C^\alpha \qquad \textit{and} \qquad \|u\|_\alpha \lesssim \sup_{j \ge -1} \{2^{j\alpha}\|u_j\|_{L^\infty}\}.$$

We should point out that everything above and all that follows can (and will) be applied to distributions on the torus. More precisely, define $\mathbb{T}^d := [-\pi,\pi]^d := (\mathbb{R}/2\pi\mathbb{Z})^d$ and let $\mathcal{D}'(\mathbb{T}^d)$ be the space of distributions on $\mathbb{T}^d$. Any $u \in \mathcal{D}'(\mathbb{T}^d)$ can be interpreted as a tempered distribution on $\mathbb{R}^d$ that is 2π–periodic in every direction, with frequency spectrum contained in $\mathbb{Z}^d$ - and vice versa. For details see [ST87], Chapter 3.2. In particular, $\Delta_j u$ is a 2π–periodic smooth function, and therefore $\|\Delta_j u\|_{L^\infty} = \|\Delta_j u\|_{L^\infty(\mathbb{T}^d)}$.

In other words, we can define

$$C^\alpha(\mathbb{T}^d) := \{u \in C^\alpha : u \text{ is } (2\pi) - \text{periodic}\}.$$

For $p \neq \infty$ however, this definition is not very useful, because no nontrivial periodic function is in L^p for $p < \infty$. Therefore, the general Besov space $B^\alpha_{p,q}(\mathbb{T}^d)$ is defined as

$$B^\alpha_{p,q}(\mathbb{T}^d) := \left\{ u \in \mathcal{D}'(\mathbb{T}^d) : \|u\|_{B^\alpha_{p,q}(\mathbb{T}^d)} := \left(\sum_{j \geq -1} (2^{j\alpha} \|\Delta_j u\|_{L^p(\mathbb{T}^d)})^q \right)^{\frac{1}{q}} < \infty \right\}.$$

Note that for $u \in \mathcal{D}'(\mathbb{T}^d)$, the Fourier transform is supported in $\mathbb{Z}^d$, and

$$u(x) = (2\pi)^{-d} \sum_{k \in \mathbb{Z}^d} \widehat{u}(k) e^{\imath \langle k, x \rangle} = \mathcal{F}^{-1}(\widehat{u})(x).$$

Apart from that, $\Delta_j u = \mathcal{F}^{-1}(\rho_j \mathcal{F} u)$ is defined exactly as in the non-periodic case.

Strictly speaking we will not work with $B^\alpha_{p,q}(\mathbb{T}^d)$ for $(p,q) \neq (\infty, \infty)$. But we will need the Besov embedding theorem on the torus:

Lemma 5.2.4. *Let $1 \leq p_1 \leq p_2 \leq \infty$ and $1 \leq q_1 \leq q_2 \leq \infty$, and let $\alpha \in \mathbb{R}$. Then $B^\alpha_{p_1,q_1}(\mathbb{T}^d)$ is continuously embedded in $B^{\alpha - d(1/p_1 - 1/p_2)}_{p_2,q_2}(\mathbb{T}^d)$.*

For the embedding theorem on $\mathbb{R}^d$ see [BCD11], Proposition 2.71. The result on the torus can be shown using the same arguments, see for example [CG06]. In both cases, the proof is based on the Bernstein inequalities, Lemma 5.2.2.

For further details concerning Littlewood-Paley theory, Besov spaces, and paraproducts, we refer to the nice book of Bahouri, Chemin, and Danchin [BCD11].

Bony's paraproduct and Young integrals

In general, the product fg of two distributions $f \in C^\alpha$ and $g \in C^\beta$ is not well defined unless $\alpha + \beta > 0$. In terms of Littlewood-Paley blocks, a product can be (at least formally) decomposed as

$$fg = \sum_{j \geq -1} \sum_{i \geq -1} \Delta_i f \Delta_j g = \pi_<(f,g) + \pi_>(f,g) + \pi_\circ(f,g).$$

Here $\pi_<(f,g)$ is the part of the double sum with $i < j - 1$, and $\pi_>(f,g)$ the part with $i > j + 1$, while $\pi_\circ(f,g)$ is the "diagonal part" where $|i - j| \leq 1$:

$$\pi_<(f,g) := \sum_{j \geq -1} \sum_{i=-1}^{j-2} \Delta_i f \Delta_j g, \qquad \pi_>(f,g) := \sum_{i \geq -1} \sum_{j=-1}^{i-2} \Delta_i f \Delta_j g := \pi_<(g,f), \qquad \text{and}$$

$$\pi_\circ(f,g) := \sum_{|i-j| \leq 1} \Delta_i f \Delta_j g.$$

We also introduce the notation

$$\pi_{\lessgtr}(f,g) := \pi_<(f,g) + \pi_>(f,g).$$

This decomposition is referred to as paraproduct [Bon81], and it behaves nicely with respect to Littlewood-Paley theory. Of course the decomposition depends on the dyadic partition of unity used to define the blocks Δ_j, and also on the particular choice of the pairs (i,j) in the diagonal part. Our choice of taking all (i,j) with $|i-j| \leq 1$ into the diagonal part corresponds to property 3 in the definition of dyadic partition of unity: $\mathrm{supp}(\rho(2^{-i}\cdot)) \cap \mathrm{supp}(\rho(2^{-j}\cdot)) = \emptyset$ for $|i-j| > 1$. In conjunction with (5.1), this property implies that every term in the series

$$\pi_<(f,g) = \sum_{j\geq -1} \sum_{i=-1}^{j-2} \Delta_i f \Delta_j g = \sum_{j \geq -1} S_{j-1} f \Delta_j g$$

has a Fourier transform that is supported in a suitable annulus, and of course the same holds true for $\pi_>(f,g)$. On the other side, the terms in the diagonal part have a Fourier transform that is supported in a ball.

Bony's crucial observation is that $\pi_<(f,g)$ (and thus $\pi_>(f,g)$) is always a well-defined distribution. In particular, if $\alpha > 0$ and $\beta \in \mathbb{R}$, then $\pi_<$ is a bounded bilinear operator from $C^\alpha \times C^\beta$ to C^β. Heuristically $\pi_<(f,g)$ behaves at large frequencies like g (and thus retains the same regularity), and f provides only a modulation of g at larger scales. The only difficulty in defining fg for arbitrary distributions lies in handling the diagonal term $\pi_\circ(f,g)$. A basic result about this bilinear operation is given by Bony's paraproduct estimates.

Lemma 5.2.5 (Theorem 2.82 and 2.85 of [BCD11])**.** *1. For any $\beta \in \mathbb{R}$, we have*

$$\|\pi_<(f,g)\|_\beta \lesssim_\beta \|f\|_{L^\infty} \|g\|_\beta.$$

For $\alpha < 0$, we have

$$\|\pi_<(f,g)\|_{\alpha+\beta} \lesssim_{\alpha,\beta} \|f\|_\alpha \|g\|_\beta.$$

2. For $\alpha + \beta > 0$, we have

$$\|\pi_\circ(f,g)\|_{\alpha+\beta} \lesssim_{\alpha,\beta} \|f\|_\alpha \|g\|_\beta.$$

We conclude that the product fg of two elements $f \in C^\alpha$ and $g \in C^\beta$ is well defined as soon as $\alpha + \beta > 0$. The attentive reader will note immediately the analogy of this statement with one of the possible incarnations of Young's theory of integration, see Chapter 4. For $\alpha, \beta > 0$ and functions $f \in C^\alpha(\mathbb{R})$ and $g \in C^\beta(\mathbb{R})$, the Young integral $I(f, \mathrm{d}g)(t) = \int_0^t f(s)\mathrm{d}g(s)$ is well defined as soon as $\alpha + \beta > 1$. On the other side it is clear from the paraproduct estimates that if $\alpha + \beta > 1$, then the distribution $f\partial_t g$ is well defined and belongs to $C^{\alpha+\beta-1}$. Another basic fact of Young integration is that

$$|I(f, \mathrm{d}g)(t) - I(f, \mathrm{d}g)(s) - f(s)(g(t) - g(s))| \lesssim |t-s|^{\alpha+\beta}\|f\|_\alpha \|g\|_\beta, \tag{5.2}$$

for all $s, t \in \mathbb{R}$, which means that the increments of the integral I behave locally (in the parameter) like the increments of g, modulo a small remainder. Taking derivatives in the sense of distributions, we have $\partial_t I(f, \mathrm{d}g) = f\partial_t g$. Moreover, Bony's paraproduct estimates allow us to see that

$$\partial_t I(f, \mathrm{d}g) - \pi_<(f, \partial_t g) = \pi_\circ(f, \partial_t g) + \pi_>(f, \partial_t g) \in C^{\alpha+\beta-1}.$$

This is an alternative form of the Young estimate (5.2). By now it should be clear why we insist on distinguishing the two "symmetric" paraproducts $\pi_<$ and $\pi_>$. Indeed in this application of paraproducts to Young integration, which will motivate all the developments that follow, the two terms $\pi_<(f, \partial_t g)$ and $\pi_>(f, \partial_t g)$ play very different roles. The first one, $\pi_<(f, \partial_t g)$, is akin to the first-order approximation $f(s)(g(t) - g(s))$ of the integral $I(f, \mathrm{d}g)(t) - I(f, \mathrm{d}g)(s)$, in the sense that the distribution $\pi_<(f, \partial_t g)$ is always well defined and behaves "locally" (here meaning for high Fourier modes) like $\partial_t g$. On the other side the term $\pi_>(f, \partial_t g)$ belongs to $C^{\alpha+\beta-1}$ and is therefore "smoother" than the main contribution $\pi_<(f, \partial_t g)$, which only belongs to $C^{\beta-1}$.

Controlled distributions and the Besov area

Recall from Chapter 4 that the theory of controlled paths is based on three basic building blocks:

1. The integral $I(v, \mathrm{d}w)$ for reference paths v and w is assumed to exist.
2. If f is controlled by v and if F is a smooth function, then $F(f)$ is controlled by v.
3. If g is controlled by w, then $I(f, \mathrm{d}g)$ is controlled by w.

The last two properties allow to consider the space of controlled paths as a Banach algebra which is stable under nonlinear maps and integration. In particular the space of paths controlled by a reference path w (which may take its values in a finite dimensional vector space) is the natural setting where to solve rough differential equations of the form

$$f(t) = f_0 + \int_0^t \varphi(f(s))\mathrm{d}w(s)$$

with fixed point methods. All that is needed is the existence of sufficiently regular iterated integrals $I(w, \mathrm{d}w)$ for w.

Consider now the product of distributions $f\partial_t w$ where $f, w \in C^\alpha$. A priori this is not well defined if $\alpha \leq 1/2$. As we have already seen, Bony's paraproduct induces the decomposition

$$f\partial_t w = \pi_<(f, \partial_t w) + \pi_\circ(f, \partial_t w) + \pi_>(f, \partial_t w),$$

where the only problematic term is $\pi_\circ(f, \partial_t w)$. Assume now that the distribution f is controlled by $v \in C^\alpha$ in the sense that there exists $f^v \in C^\alpha$, such that

$$f^\sharp = f - \pi_<(f^v, v) \in C^{2\alpha}.$$

Then (at least formally) we can decompose the problematic term as

$$\pi_\circ(f, \partial_t w) = \pi_\circ(\pi_<(f^v, v), \partial_t w) + \pi_\circ(f^\sharp, \partial_t w),$$

where $\pi_\circ(f^\sharp, \partial_t w) \in C^{3\alpha-1}$ is well defined if we assume that $3\alpha - 1 > 0$. It turns out (see Lemma 5.3.3 below) that, although $\pi_\circ(\pi_<(f^v, v), \partial_t w)$ is a-priori not defined, the "commutator"

$$R(f^v, v, \partial_t w) := \pi_\circ(\pi_<(f^v, v), \partial_t w) - f^v \pi_\circ(v, \partial_t w)$$

can be defined and belongs to $C^{3\alpha-1}$. If we assume that $\pi_\circ(v, \partial_t w) \in C^{2\alpha-1}$ is given, then the product $f^v \pi_\circ(v, \partial_t w)$ is well defined under the assumption $3\alpha - 1 > 0$. In this way we have reduced the problem of defining the product $f\partial_t w$ for f controlled by v essentially to the problem of defining the product $v\partial_t w$, exactly as in the theory of controlled rough paths. Moreover, we get the expansion formula

$$f\partial_t w = \pi_<(f, \partial_t w) + \pi_>(f, \partial_t w) + \pi_\circ(f^\sharp, \partial_t w) + R(f^v, v, \partial_t w) + f^v \pi_\circ(v, \partial_t w) \in C^{\alpha-1}$$

and

$$f\partial_t w - \pi_<(f, \partial_t w) \in C^{2\alpha-1}.$$

This shows that f is controlled by $\partial_t w$ (modulo a slight adaption of the definition of controlled distributions).

We point out that the argumentation above works for general distributions that are defined on $\mathbb{R}^d$, without assuming that one of them is a derivative: If $\pi_\circ(v, w)$ is given, smooth enough, and f and g are controlled by v and w respectively, then also $\pi_\circ(f, g)$ (and thus fg) can be defined in a sensible way.

5.3. Paracontrolled calculus

Motivated by the considerations of the previous section here we lay out some elements of a calculus of controlled distributions. We start from the analysis of the commutator, which allows to define the diagonal part of the product as a function of the diagonal part for special reference distributions. Then we show that controlled distributions are stable under nonlinear maps. At the end of this section, which can be skipped at first reading, we gather further commutator estimates, we extend the definition of the product, and we establish some continuity properties of the product.

5.3.1. A basic commutator estimate

In this section we prove the basic commutator estimate for $\pi_\circ(\pi_<(f,v),w) - f\pi_\circ(v,w)$, where f,v,w are suitable tempered distributions, that will allow us to extend the definition of the product by a continuity argument.

We will occasionally prove bounds for operators acting on functions with Fourier transform of compact support, and then argue that the domain of the operator can be extended by approximation. This can be achieved in the following way. Let $A : \mathcal{S}' \to \mathcal{S}'$ be an operator such that for all f with Fourier transform of compact support we have $\|Af\|_\alpha \lesssim \|f\|_\beta$, where $\alpha, \beta \in \mathbb{R}$. Then it is possible to extend A in a continuous way to

$$C^{\beta,0} := \left\{ f \in C^\beta : \lim_{j\to\infty} 2^{j\beta} \|\Delta_j f\|_{L^\infty} = 0 \right\} :$$

$S_j f$ converges to f in C^β for every $f \in C^{\beta,0}$, and since $\|AS_i f - AS_j f\|_\alpha \lesssim \|S_i f - S_j f\|_\beta$, we obtain that $(AS_j f)_j$ is a Cauchy sequence. It is clear that the definition of Af does not depend on the specific sequence of smooth approximations $(S_j f)_j$. If furthermore we have the estimate $\|Af\|_{\alpha'} \lesssim \|f\|_{\beta'}$ for some $\alpha' \le \alpha$ and $\beta' < \beta$, then we can also extend A continuously to C^β, because $C^\beta \subset C^{\beta',0}$.

We will often need the following commutator estimate. Recall that we set $\varphi(\mathrm{D})u = \mathcal{F}^{-1}(\varphi \mathcal{F} u)$ whenever the right hand side is defined.

Lemma 5.3.1 (Lemma 2.97 of [BCD11])**.** *Let φ be a continuously differentiable function, such that $(1+|\cdot|)\mathcal{F}\varphi \in L^1$. Then for any Lipschitz continuous function u with $\mathrm{D}u \in L^\infty$, any $v \in L^\infty$, and any $\lambda > 0$ we have*

$$\|[u, \varphi(\lambda \mathrm{D})]v\|_{L^\infty} \lesssim \lambda \max_{\eta \in \mathbb{N}^d : |\eta| = 1} \|\partial^\eta u\|_{L^\infty} \|v\|_{L^\infty},$$

where $[u, \varphi(\lambda\mathrm{D})]v := u \cdot (\varphi(\lambda\mathrm{D})v) - \varphi(\lambda\mathrm{D})(u \cdot v)$.

If we apply this for $\varphi = \rho$ or $\varphi = \chi$, where (χ, ρ) is our dyadic partition of unity, and for $\lambda = 2^{-j}$, then we obtain the following corollary:

Corollary 5.3.2. *Let $u \in C^1_b$ and $v \in L^\infty$. Then for all $j \ge -1$ we have*

$$\Delta_j(uv) = u\Delta_j v + B_j(u,v),$$

where

$$\|B_j(u,v)\|_\infty \lesssim 2^{-j} \max_{\eta \in \mathbb{N}^d : |\eta| = 1} \|\partial^\eta u\|_\infty \|v\|_\infty.$$

We are now in a position to prove our main commutator estimate.

Lemma 5.3.3. *Let $\alpha \in (0,1)$, $\alpha + \beta + \gamma > 0$, but $\beta + \gamma < 0$. Then the remainder*

$$R(f,v,w) := \pi_\circ(\pi_<(f,v),w) - f\pi_\circ(v,w)$$

is well defined for all $f \in C^\alpha$*,* $v \in C^\beta$*, and* $w \in C^\gamma$*, and belongs to* $C^{\alpha+\beta+\gamma}$*. Moreover*

$$\|R(f,v,w)\|_{\alpha+\beta+\gamma} \lesssim \|f\|_\alpha \|v\|_\beta \|w\|_\gamma.$$

Proof. As explained above, it suffices to argue for f, v, w with Fourier transform of compact support, so that all sums are finite and exchanging the order of summation is justified. Since the Fourier transform of $S_{k-1} f \Delta_k v$ has its support in an annulus of the form $2^k \mathcal{A}$, there exists $N > 0$ such that

$$\pi_\circ(\pi_<(f,v),w) = \sum_{i,j,k} [-1, k-1)(\ell) 1_{[-1,1]}(|i-j|) 1_{[-N,N]}(|i-k|) \Delta_i (S_{k-1} f \Delta_k v) \Delta_j w.$$

Now Corollary 5.3.2 states that

$$\Delta_i (S_{k-1} f \Delta_k v) = S_{k-1} f \Delta_i \Delta_k v + B_i(S_{k-1} f, \Delta_k v).$$

Hence

$$\begin{aligned} &\pi_\circ(\pi_<(f,v),w) - f\pi_\circ(v,w) \\ &\quad = \sum_{i,j,k,\ell} (1_{[-1,k-1)}(\ell) 1_{[-1,1]}(|i-j|) 1_{[-N,N]}(|i-k|) - 1_{[-1,1]}(|i-j|)) \Delta_\ell f \Delta_i \Delta_k v \Delta_j w \\ &\qquad + \sum_{i,j,k} 1_{[-1,1]}(|i-j|) 1_{[-N,N]}(|i-k|) B_i(S_{k-1} f, \Delta_k v) \Delta_j w. \end{aligned} \tag{5.3}$$

Let us consider the first series in (5.3). We have $\Delta_i \Delta_k v = 0$ for $|i-k| > 1$, and therefore the series can be rewritten as

$$\begin{aligned} &\sum_{i,j,k,\ell} (1_{[-1,k-1)}(\ell) 1_{[-1,1]}(|i-j|) 1_{[-N,N]}(|i-k|) - 1_{[-1,1]}(|i-j|)) \Delta_\ell f \Delta_i \Delta_k v \Delta_j w \\ &\qquad = -\sum_{i,j,k,\ell} (1_{[k-1,\infty)}(\ell) 1_{[-1,1]}(|i-j|) 1_{[-1,1]}(|i-k|)) \Delta_\ell f \Delta_i \Delta_k v \Delta_j w \\ &\qquad = -\sum_\ell \left(\sum_{i,j,k} (1_{[-1,\ell+1]}(k) 1_{[-1,1]}(|i-j|) 1_{[-1,1]}(|i-k|)) \Delta_\ell f \Delta_i \Delta_k v \Delta_j w \right). \end{aligned}$$

Every term of this series in ℓ has a Fourier transform that is supported in a suitable ball $2^\ell \mathcal{B}$. By Lemma 5.2.3 it suffices to control its L^∞ norm, which can be estimated by

$$\begin{aligned} &\left\| \sum_{i,j,k} (1_{[-1,\ell+1]}(k) 1_{[-1,1]}(|i-j|) 1_{[-1,1]}(|i-k|)) \Delta_\ell f \Delta_i \Delta_k v \Delta_j w \right\|_\infty \\ &\qquad \lesssim \sum_{k \le \ell+1} 2^{-\ell\alpha} \|f\|_\alpha 2^{-k\beta} \|v\|_\beta 2^{-k\gamma} \|w\|_\gamma \lesssim 2^{-\ell(\alpha+\beta+\gamma)} \|f\|_\alpha \|v\|_\beta \|w\|_\gamma, \end{aligned}$$

where in the last step we used that $\beta + \gamma < 0$, which implies $\sum_{k \le \ell+1} 2^{-k(\beta+\gamma)} \simeq 2^{-\ell(\alpha+\beta)}$.

Since $\alpha+\beta+\gamma>0$, Lemma 5.2.3 now yields

$$\left\|\sum_{\ell}\left(\sum_{i,j,k}(1_{[-1,\ell+1]}(k)1_{[-1,1]}(|i-j|)1_{[-1,1]}(|i-k|))\Delta_\ell f\Delta_i\Delta_k v\Delta_j w\right)\right\|_{\alpha+\beta+\gamma} \lesssim \|f\|_\alpha\|v\|_\beta\|w\|_\gamma.$$

It remains to estimate the second series in (5.3), i.e.

$$\sum_{i,j,k}1_{[-1,1]}(|i-j|)1_{[-N,N]}(|i-k|)B_i(S_{k-1}f,\Delta_k v)\Delta_j w = \sum_k\left(\sum_{i,j}1_{[-1,1]}(|i-j|)1_{[-N,N]}(|i-k|)B_i(S_{k-1}f,\Delta_k v)\Delta_j w\right). \tag{5.4}$$

Recall that by definition of $B_i(S_{k-1}f,\Delta_k v)$ we have

$$B_i(S_{k-1}f,\Delta_k v)=\Delta_i(S_{k-1}f\Delta_k v)-S_{k-1}f\Delta_i\Delta_k v=\Delta_i(S_{k-1}f\Delta_k v)-S_{k-1}f\Delta_k\Delta_i v,$$

and thus for every $i,k\geq -1$ the Fourier transform of $B_i(S_{k-1}f,\Delta_k v)$ has its support in an annulus of the form $2^k\mathcal{A}$. Hence, for $i\sim j\sim k$, the Fourier transform of $B_i(S_{k-1}f,\Delta_k v)\Delta_j w$ is supported in a ball $2^k\mathcal{B}$. Since $\alpha+\beta+\gamma>0$, Lemma 5.2.3 implies that it suffices to control the L^∞–norm of the terms of the series (5.4). But we obtain from Corollary 5.3.2 that

$$\left\|\sum_{i:i\sim k}\sum_{j:j\sim i}B_i(S_{k-1}f,\Delta_k v)\Delta_j w\right\|_{L^\infty} \lesssim 2^{-k}\max_{|\eta|=1}\|\partial^\eta S_{k-1}f\|_{L^\infty}\|\Delta_k v\|_{L^\infty}\|\Delta_k w\|_{L^\infty}$$
$$\lesssim 2^{-k}2^{k(1-\alpha)}\|f\|_\alpha 2^{-k\beta}\|v\|_\beta 2^{-k\gamma}\|w\|_\gamma = 2^{-k(\alpha+\beta+\gamma)}\|f\|_\alpha\|v\|_\beta\|w\|_\gamma,$$

where we used that $1-\alpha>0$, and therefore $\sum_{\ell<k-1}2^{\ell(1-\alpha)}\simeq 2^{k(1-\alpha)}$. □

Remark 5.3.4. The restriction $\beta+\gamma<0$ is not problematic. If $\beta+\gamma\geq 0$, then for every $\varepsilon>0$ we have $\beta+\gamma'<0$, where $\gamma'=-\beta-\varepsilon$, and therefore

$$\|R(f,v,w)\|_{\alpha-\varepsilon}\lesssim\|f\|_\alpha\|v\|_\beta\|w\|_{-\beta-\varepsilon}\leq\|f\|_\alpha\|v\|_\beta\|w\|_\gamma.$$

Since the full product fw will not be in $C^{\alpha+\beta+\gamma}$ but much rougher than that, this loss of regularity does not bother us.

The restriction $\alpha<1$ is a real one, unless $1+\beta+\gamma>0$. Here, this will not affect us, because we will deal with distributions that are rougher than C^1. But if we are given a relatively smooth v, say $v\in C^{3/2}$, and a smooth function F, and we want to multiply $F(v)$ with a derivative of a high order, say $\mathrm{D}^4 v$, then our current commutator estimate cannot be applied. Even if we assume that $\pi_\circ(v,\mathrm{D}^4 v)\in C^{-1}$ is given, and despite the fact that $3/2-1>0$, the restriction $\alpha<1$ prevents us from defining $F(v)\mathrm{D}^4 v$ using Lemma 5.3.3.

It is possible to overcome this restriction by introducing higher order Littlewood-Paley blocks, in terms of which one can define higher order correctors. But to simplify the presentation, and since we are not dealing with problems where the higher order Littlewood-Paley blocks would be helpful, we decided to not present these results here.

Lemma 5.3.3 allows us to define the product fw for f controlled by v, under the assumption that we already constructed vw. The next result will allow us to define the product fg for f controlled by v, for g controlled by w, and under the assumption that vw is given.

Corollary 5.3.5. *Let $\alpha \in (0,1)$, $\alpha+\beta+\gamma > 0$, but $\beta+\gamma < 0$. Let $v = (v^1, \dots, v^m) \in C^\beta(\mathbb{R}^d, \mathbb{R}^m)$, $w = (w^1, \dots, w^n) \in C^\gamma(\mathbb{R}^d, \mathbb{R}^n)$, $f = (f^1, \dots, f^m) \in C^\alpha(\mathbb{R}^d, \mathcal{L}(\mathbb{R}^m, \mathbb{R}))$, and $g = (g^1, \dots, g^n) \in C^\alpha(\mathbb{R}^d, \mathcal{L}(\mathbb{R}^n, \mathbb{R}))$. Then*

$$\begin{aligned}\overline{R}(f,g,v,w) &:= \pi_\circ(\pi_<(f,v), \pi_<(g,w)) - fg\pi_\circ(v,w) \\ &:= \sum_{k=1}^m \sum_{\ell=1}^n \Big(\pi_\circ(\pi_<(f^k, v^k), \pi_<(g^\ell, w^\ell)) - f^k g^\ell \pi_\circ(v^k, w^\ell) \Big)\end{aligned}$$

is well defined and in $C^{\alpha+\beta+\gamma}$, and

$$\|\overline{R}(f,g,v,w)\|_{\alpha+\beta+\gamma} \lesssim \|f\|_\alpha \|g\|_\alpha \|v\|_\beta \|w\|_\gamma.$$

Proof. Lemma 5.3.3 implies that for fixed $1 \le k \le m$ and $1 \le \ell \le n$ we have

$$\begin{aligned}\pi_\circ(\pi_<(f^k, v^k), \pi_<(g^\ell, w^\ell)) &= R(f^k, v^k, \pi_<(g^\ell, w^\ell)) + f^k \pi_\circ(v^k, \pi_<(g^\ell, w^\ell)) \\ &= R(f^k, v^k, \pi_<(g^\ell, w^\ell)) + f^k \pi_\circ(\pi_<(g^\ell, w^\ell), v^k) \\ &= R(f^k, v^k, \pi_<(g^\ell, w^\ell)) + f^k R(g^\ell, w^\ell, v^k) + f^k g^\ell \pi_\circ(w^\ell, v^k) \\ &= R(f^k, v^k, \pi_<(g^\ell, w^\ell)) + f^k R(g^\ell, w^\ell, v^k) + f^k g^\ell \pi_\circ(v^k, w^\ell).\end{aligned}$$

The result now follows from Lemma 5.3.3 and from the paraproduct estimates. □

5.3.2. Product of controlled distributions

In this section we define the product of two controlled distributions, and we prove continuity properties of the product operator. For simplicity we restrict our attention to one dimensional controlled distributions. The *controlling* distributions can be multi dimensional. The general (finite dimensional) case can then be treated by considering each component separately.

Definition 5.3.6. Let $\alpha \in \mathbb{R}$ and $\beta > 0$. A distribution $f \in C^\alpha(\mathbb{R}^d, \mathbb{R})$ is controlled by $v \in C^\alpha(\mathbb{R}^d, \mathbb{R}^n)$ if there exists $f^v \in C^\beta(\mathbb{R}^d, \mathcal{L}(\mathbb{R}^n, \mathbb{R}))$, such that

$$f^\sharp = f - \pi_<(f^v, v) \in C^{\alpha+\beta}.$$

We write $f \in \mathcal{D}_v^{\alpha,\beta} := \mathcal{D}_v^{\alpha,\beta}(\mathbb{R})$ and define

$$\|f\|_{v,\alpha,\beta} := \|f\|_\alpha + \|f^v\|_\beta + \|f^\sharp\|_{\alpha+\beta}.$$

Remark 5.3.7. In general, f^v is not uniquely determined by f and v. For example, if $\alpha = \beta > 0$ and $v \in C^{2\alpha}$, then $0 \in \mathcal{D}_v^{\alpha,\alpha}$, and every $f \in C^\alpha$ can be taken as its derivative. So the correct definition would be $(f, f^v) \in \mathcal{D}_v^{\alpha,\beta}$. We use $f \in \mathcal{D}_v^{\alpha,\beta}$ as an abbreviation, because usually it will be clear from the context which derivative we have in mind.

Example 5.3.8. If α is the regularity of a first order derivative, say $\alpha = \gamma - 1$ for some $\gamma \in (0,1)$, then we will usually take $\beta = \gamma$, so that $\alpha + \beta = 2\gamma - 1$.

Remark 5.3.9. If $(\widetilde{\chi}, \widetilde{\rho})$ is another dyadic partition of unity, and if $\widetilde{\pi}_<(f^v, v)$ is the paraproduct based on $(\widetilde{\chi}, \widetilde{\rho})$, then $\|\pi_<(f^v, v) - \widetilde{\pi}_<(f^v, v)\|_{\alpha+\beta} \lesssim \|f^v\|_\beta \|v\|_\alpha$, see Lemma F.1 in the appendix. Therefore, only the norm $\|\cdot\|_{v,\alpha,\beta}$ depends on the specific dyadic partition of unity, but not the space $\mathcal{D}_v^{\alpha,\beta}$. Since v is fixed, every other dyadic partition of unity corresponds to an equivalent norm.

To fix ideas, we first define the product fw for given $\pi_\circ(v, w)$ and f controlled by v, where v has positive regularity and w has negative regularity. The problem of defining fg for g controlled by w can then be treated using similar arguments, as we will show below.

Let $v = (v^1, \dots, v^m) \in C^\beta$ and $w = (w^1, \dots, w^n) \in C^\gamma$. If $0 < -\gamma < \beta$, then Bony's paraproduct estimates imply that $vw = (v^k w^\ell)_{1 \le k \le m, 1 \le \ell \le n} \in C^\gamma$, but $vw - \pi_<(v, w) = (v^k w^\ell - \pi_<(v^k, w^\ell))_{k,\ell} \in C^{\beta+\gamma}$. This motivates our standing assumption, which is that we are given a distribution $vw \in C^\gamma$, such that

$$vw - \pi_<(v, w) \in C^{\beta+\gamma} \tag{5.5}$$

for one (and then according to Lemma F.1 for every) dyadic partition of unity. Note that if $\beta + \gamma \le 0$, then this is an assumption on the existence of vw, which cannot be constructed using Bony's arguments.

Definition 5.3.10. Let $\gamma < 0 < \beta$, and let $v \in C^\beta(\mathbb{R}^d, \mathbb{R}^m)$ and $w \in C^\gamma(\mathbb{R}^d, \mathbb{R}^m)$. A distribution $vw \in C^\gamma(\mathbb{R}^d, \mathbb{R}^{m \times n})$ is called *Besov area* for v and w if it satisfies (5.5). In that case we define

$$\pi_\circ(v, w) := vw - \pi_<(v, w) - \pi_>(v, w).$$

Theorem 5.3.11. *Let $\gamma < 0 < \beta$ and let $\alpha \in (0,1)$ be such that $\alpha + \beta + \gamma > 0$ and $\alpha > \beta + \gamma$. Let $v \in C^\beta(\mathbb{R}^d, \mathbb{R}^m)$, $w \in C^\gamma(\mathbb{R}^d, \mathbb{R}^n)$, with associated Besov area $vw = (v^k w^\ell)_{1 \le k \le m, 1 \le \ell \le n} \in C^\gamma(\mathbb{R}^d, \mathbb{R}^{m \times n})$. For $f \in \mathcal{D}_v^{\beta,\alpha}(\mathbb{R})$ we define*

$$f \circ_v w := (f \circ w^\ell)_{1 \le \ell \le n} := \pi_{\lessgtr}(f, w) + \pi_\circ(f^\sharp, w) + R(f^v, v, w) + f^v \pi_\circ(v, w).$$

Then $f \circ_v w \in \mathcal{D}_w^{\gamma,\beta}(\mathbb{R}^n)$ with derivative f, and

$$\|f \circ_v w\|_{w,\gamma,\beta} \lesssim \|f\|_{v,\beta,\alpha} \left(1 + \|w\|_\gamma + \|v\|_\beta \|w\|_\gamma + \|\pi_\circ(v, w)\|_{\beta+\gamma}\right).$$

If there exist sequences of smooth functions $(v_j)_{j\in\mathbb{N}}$ and $(w_j)_{j\in\mathbb{N}}$, such that (v_j) converges to v in C^β, (w_j) converges to w in $C^{\beta+\gamma}$, and $(\pi_\circ(v_j, w_j))$ converges to $\pi_\circ(v, w)$ in $C^{\beta+\gamma}$, then the definition of $f \circ_v w$ does not depend on the dyadic partition of unity used to define the paraproduct and the commutator R.

Proof. Since $\alpha, \beta > 0$ and $\gamma < 0$, the paraproduct estimates, Lemma 5.2.5, imply

$$\begin{aligned}\|\pi_{\lessgtr}(f, w) + \pi_\circ(f^\sharp, w)\|_\gamma &\lesssim \|f\|_{L^\infty}\|w\|_\gamma + \|\pi_>(f, w)\|_{\beta+\gamma} + \|\pi_\circ(f^\sharp, w)\|_{\beta+\alpha+\gamma}\\ &\lesssim \|f\|_\beta\|w\|_\gamma + \|f^\sharp\|_{\beta+\alpha}\|w\|_\gamma,\end{aligned}$$

where we used that $\beta + \alpha + \gamma > 0$ to estimate $\pi_\circ(f^\sharp, w)$. Let $\varepsilon > 0$ be such that $\alpha - \varepsilon > \beta + \gamma$. Lemma 5.3.3 and Remark 5.3.4 imply that

$$\|R(f^v, v, w)\|_\gamma \le \|R(f^v, v, w)\|_{(\alpha-\varepsilon)\wedge(\alpha+\beta+\gamma)} \lesssim \|f^v\|_\alpha\|v\|_\beta\|w\|_\gamma.$$

Since $\alpha + \beta + \gamma > 0$ and $\beta + \gamma < \alpha$, another application of the paraproduct estimates yields

$$\|f^v\pi_\circ(v, w)\|_\gamma \le \|f^v\pi_\circ(v, w)\|_{\beta+\gamma} \lesssim \|f^v\|_\alpha\|\pi_\circ(v, w)\|_{\beta+\gamma}.$$

Note that for all terms we could estimate at least the $\|\cdot\|_{\beta+\gamma}$–norm, except for $\pi_<(f, w)$, which is only in C^γ. Hence, we have

$$\|f \circ_v w\|_\gamma + \|f \circ_v w - \pi_<(f, w)\|_{\gamma+\beta} \lesssim \|f\|_{v,\beta,\alpha}(\|w\|_\gamma + \|v\|_\beta\|w\|_\gamma + \|\pi_\circ(v, w)\|_{\beta+\gamma}).$$

The estimate for the derivative $(f \circ_v w)^w = f$ is trivial.

Let now $(v_j)_{j\in\mathbb{N}}$ and $(w_j)_{j\in\mathbb{N}}$ be sequences of smooth functions, such that (v_j) converges to v in C^β, (w_j) converges to w in $C^{\beta+\gamma}$, and $(\pi_\circ(v, w_j))$ converges to $\pi_\circ(v, w)$ in $C^{\beta+\gamma}$. Let $(\widetilde{\chi}, \widetilde{\rho})$ be another dyadic partition of unity, and let $\widetilde{\pi}_\circ$, $\widetilde{\pi}_<$, $\widetilde{\pi}_>$, and $f\widetilde{\circ}_v w$ be the operators defined in terms of this partition. We define

$$\begin{aligned}P_1(f, f^v, v, v_j, w_j) &:= \pi_{\lessgtr}(f, w_j) + \pi_\circ(f - \pi_<(f^v, v), w_j) + R(f^v, v_j, w_j) + f^v\pi_\circ(v_j, w_j)\\ &= f w_j + \pi_\circ(\pi_<(f^v, v_j - v), w_j),\end{aligned}$$

where $f w_j$ denotes the classical product, defined for example using Bony's arguments. By continuity of the operators involved, $P_1(f, f^v, v, v_j, w_j)$ converges to $f \circ_v w$ as j tends to ∞. Similarly we define $P_2(f, f^v, v, v_j, w_j) := f w_j + \widetilde{\pi}_\circ(\widetilde{\pi}_<(f^v, v_j - v), w_j)$, which converges to $f\widetilde{\circ}_v w$ as j tends to ∞, because by Lemma F.1 also $(\widetilde{\pi}_\circ(v_j, w_j))$ converges to $\widetilde{\pi}_\circ(v, w)$. For the difference $P_1 - P_2$, Lemma F.1 and a straightforward calculation imply that

$$\begin{aligned}&\|P_1(f, f^v, v, v_j, w_j) - P_2(f, f^v, v, v_j, w_j)\|_{\beta+\gamma}\\ &\quad\lesssim \|\pi_\circ(\pi_<(f^v, v_j - v) - \widetilde{\pi}_<(f^v, v_j - v), w_j)\|_{\beta+\gamma}\\ &\qquad + \|\pi_\circ(\widetilde{\pi}_<(f^v, v_j - v), w_j) - \widetilde{\pi}_\circ(\widetilde{\pi}_<(f^v, v_j - v), w_j)\|_{\beta+\gamma}\\ &\quad\lesssim \|f^v\|_\alpha\|v_j - v\|_\beta\|w_j\|_\gamma,\end{aligned}$$

which converges to zero by assumption. $\square$

Corollary 5.3.12. *Let α, β, γ and f, v, w be as in Theorem 5.3.11. Let furthermore $\widetilde{v} \in C^\beta$ and $\widetilde{w} \in C^\gamma$ with associated Besov area $\widetilde{v}\widetilde{w}$, and let $\widetilde{f} \in \mathcal{D}^{\beta,\alpha}_{\widetilde{v}}$. Then*

$$\begin{aligned} &\|f \circ_v w - \widetilde{f} \circ_{\widetilde{v}} \widetilde{w}\|_\gamma \qquad (5.6) \\ &\quad \lesssim (\|f - \widetilde{f}\|_\beta + \|f^v - \widetilde{f}^{\widetilde{v}}\|_\alpha + \|f^\sharp - \widetilde{f}^\sharp\|_{\beta+\alpha})(\|w\|_\gamma + \|v\|_\beta \|w\|_\gamma + \|\pi_\circ(v, w)\|_{\beta+\gamma}) \\ &\qquad + (\|w - \widetilde{w}\|_\gamma + \|v - \widetilde{v}\|_\beta + \|\pi_\circ(v, w) - \pi_\circ(\widetilde{v}, \widetilde{w})\|_{\beta+\gamma})\|\widetilde{f}\|_{\widetilde{v},\beta,\alpha}(1 + \|v\|_\beta + \|w\|_\gamma). \end{aligned}$$

Proof. The multilinearity of the involved operators leads to the decomposition

$$\begin{aligned} f \circ_v w - \widetilde{f} \circ_{\widetilde{v}} \widetilde{w} = \pi_\lessgtr(f - \widetilde{f}, w) &+ \pi_\lessgtr(\widetilde{f}, w - \widetilde{w}) + \pi_\circ(f^\sharp - \widetilde{f}^\sharp, w) + \pi_\circ(\widetilde{f}^\sharp, w - \widetilde{w}) \\ &+ R(f^v - \widetilde{f}^{\widetilde{v}}, v, w) + R(\widetilde{f}^{\widetilde{v}}, v - \widetilde{v}, w) + R(\widetilde{f}^{\widetilde{v}}, \widetilde{v}, w - \widetilde{w}) \\ &+ (f^v - \widetilde{f}^{\widetilde{v}})\pi_\circ(v, w) + \widetilde{f}^{\widetilde{v}}(\pi_\circ(v, w) - \pi_\circ(\widetilde{v}, \widetilde{w})). \end{aligned}$$

The estimate now follows from the paraproduct estimates and from Lemma 5.3.3. $\square$

Remark 5.3.13. To lighten the notation, we will write fv instead of $f \circ_v w$ from now on, unless we want to stress that the product of controlled distributions is considered.

Note that the definition of fv depends on the special choice of vw. From the theory of rough paths it is known that there is no canonical definition of vw beyond the Young/Bony setting, and in fact there are infinitely many possible choices.

Next we define the product fg for f controlled by v and g controlled by w. For this purpose we need the following commutator estimate, that is due to Bony:

Lemma 5.3.14 (Theorem 2.3 of [Bon81])**.** *Let $\alpha > 0$ and $\beta \in \mathbb{R}$. Let $f, v \in C^\alpha$ and $w \in C^\beta$. Then*

$$\|\pi_<(f, \pi_<(v, w)) - \pi_<(fv, w)\|_{\alpha+\beta} \lesssim \|f\|_\alpha \|v\|_\alpha \|w\|_\beta.$$

Let us now define the product $f \circ_{v,w} g$ for f controlled by v and g controlled by w. We restrict our attention to $f \in \mathcal{D}^{\beta,\beta}_v$ and $g \in \mathcal{D}^{\gamma,\beta}_w$. The situation $f \in \mathcal{D}^{\beta,\alpha}_v$ and $g \in \mathcal{D}^{\gamma,\alpha}_w$ can be treated analogously, at the price of distinguishing the cases $\alpha \le \beta$ and $\alpha > \beta$.

Theorem 5.3.15. *Let $\gamma < 0$ and $\beta \in (0, 1)$ be such that $\gamma + 2\beta > 0$. Let $v \in C^\beta(\mathbb{R}^d, \mathbb{R}^m)$, $w \in C^\gamma(\mathbb{R}^d, \mathbb{R}^n)$, with associated Besov area $vw = (v^k w^\ell)_{1 \le k \le m, 1 \le \ell \le n} \in C^\gamma(\mathbb{R}^d, \mathbb{R}^{m \times n})$. For $f \in \mathcal{D}^{\beta,\beta}_v(\mathbb{R})$ and $g \in \mathcal{D}^{\gamma,\beta}_w(\mathbb{R})$ we define*

$$f \circ_{v,w} g := \pi_\lessgtr(f, g) + \pi_\circ(f^\sharp, g) + \pi_\circ(\pi_<(f^v, v), g^\sharp) + \overline{R}(f^v, g^w, v, w) + f^v g^w \pi_\circ(v, w).$$

Then $f \circ_{v,w} g \in \mathcal{D}^{\gamma,\beta}_w$ with derivative fg^w, and

$$\|f \circ_{v,w} g\|_{w,\gamma,\beta} \lesssim \|f\|_{v,\beta,\beta} \|g\|_{w,\gamma,\beta}(1 + \|v\|_\beta + \|w\|_\gamma + \|v\|_\beta \|w\|_\gamma + \|\pi_\circ(v, w)\|_{\beta+\gamma}).$$

If there exist sequences of smooth functions $(v_j)_{j\in\mathbb{N}}$ and $(w_j)_{j\in\mathbb{N}}$, such that (v_j) converges to v in C^β, (w_j) converges to w in $C^{\beta+\gamma}$, and $(\pi_\circ(v_j,w_j))$ converges to $\pi_\circ(v,w)$ in $C^{\beta+\gamma}$, then the definition of $f\circ_{v,w} g$ does not depend on the dyadic partition of unity used to define the paraproduct and the commutator $\overline{R}$.

If furthermore $\widetilde{v}\in C^\beta$, $\widetilde{w}\in C^\gamma$, $\pi_\circ(\widetilde{v},\widetilde{w})\in C^{\beta+\gamma}$, and $\widetilde{f}\in\mathcal{D}^{\beta,\beta}_{\widetilde{v}}$, $\widetilde{g}\in\mathcal{D}^{\gamma,\beta}_{\widetilde{w}}$, then

$$
\begin{aligned}
\|f\circ_{v,w} g-\widetilde{f}\circ_{\widetilde{v},\widetilde{w}}\widetilde{g}\|_\gamma \lesssim{} & (\|f-\widetilde{f}\|_\beta+\|f^v-\widetilde{f}^{\widetilde{v}}\|_\beta+\|f^\sharp-\widetilde{f}^\sharp\|_{2\beta}) \\
&\quad\times\|g\|_{w,\gamma,\beta}(1+\|v\|_\beta+\|v\|_\beta\|w\|_\gamma+\|\pi_\circ(v,w)\|_{\beta+\gamma}) \\
&+(\|g-\widetilde{g}\|_\gamma+\|g^w-\widetilde{g}^{\widetilde{w}}\|_\beta+\|g^\sharp-\widetilde{g}^\sharp\|_{\gamma+\beta}) \\
&\quad\times\|\widetilde{f}\|_{\widetilde{v},\beta,\beta}(1+\|v\|_\beta+\|v\|_\beta\|w\|_\gamma+\|\pi_\circ(v,w)\|_{\beta+\gamma}) \\
&+(\|w-\widetilde{w}\|_\gamma+\|v-\widetilde{v}\|_\beta+\|\pi_\circ(v,w)-\pi_\circ(\widetilde{v},\widetilde{w})\|_{\beta+\gamma}) \\
&\quad\times\|\widetilde{f}\|_{\widetilde{v},\beta,\beta}\|\widetilde{g}\|_{\widetilde{w},\gamma,\beta}(1+\|v\|_\beta+\|w\|_\gamma).
\end{aligned}
\tag{5.7}
$$

Proof. By the same arguments as in the proof of Theorem 5.3.11, using Corollary 5.3.5 instead of Lemma 5.3.3, we obtain that $f\circ_{v,w} g$ is controlled by g, with derivative f, and

$$
\begin{aligned}
&\|f\circ_{v,w} g\|_\gamma+\|f\circ_{v,w} g-\pi_<(f,g)\|_{\gamma+\beta} \\
&\qquad\lesssim\|f\|_{v,\beta,\beta}\|g\|_{w,\gamma,\beta}(1+\|v\|_\beta+\|v\|_\beta\|w\|_\gamma+\|\pi_\circ(v,w)\|_{\beta+\gamma}),
\end{aligned}
$$

and that the definition of the product does not depend on the dyadic partition of unity if $\pi_\circ(v,w)$ is given as limit of $(\pi_\circ(v_j,w_j))$ for sequences of smooth functions (v_j) and (w_j) that converge to v and w respectively.

Let us show that $f\circ_{v,w} g$ is controlled by w, with derivative fg^w. Note that

$$
\begin{aligned}
\|f\circ_{v,w} g-\pi_<(fg^w,w)\|_{\gamma+\beta} \le{} & \|f\circ_{v,w} g-\pi_<(f,g)\|_{\gamma+\beta}+\|\pi_<(f,g-\pi_<(g^w,w))\|_{\gamma+\beta} \\
&+\|\pi_<(f,\pi_<(g^w,w))-\pi_<(fg^w,w)\|_{\gamma+\beta} \\
\lesssim{} & \|f\|_{v,\beta,\beta}\|g\|_{w,\gamma,\beta}(1+\|v\|_\beta+\|v\|_\beta\|w\|_\gamma+\|\pi_\circ(v,w)\|_{\beta+\gamma}) \\
&+\|f\|_{L^\infty}\|g^\sharp\|_{\gamma+\beta}+\|f\|_\beta\|g^w\|_\beta\|w\|_\gamma,
\end{aligned}
$$

where we applied Lemma 5.3.14 to the third term. The estimate for the derivative fg^w is easily derived, and therefore

$$
\|f\circ_{v,w} g\|_{w,\gamma,\beta}\lesssim\|f\|_{v,\beta,\beta}\|g\|_{w,\gamma,\beta}(1+\|v\|_\beta+\|w\|_\gamma+\|v\|_\beta\|w\|_\gamma+\|\pi_\circ(v,w)\|_{\beta+\gamma}).
$$

The estimate for $\|f\circ_{v,w} g-\widetilde{f}\circ_{\widetilde{v},\widetilde{w}}\widetilde{g}\|_\gamma$ is derived in the same way as Corollary 5.3.12. $\square$

5.3.3. Stability under nonlinear maps

Here we establish the stability of controlled distributions under nonlinear maps, and we show that the space of controlled distributions is an algebra. We still assume the controlled distributions to be one dimensional, whereas the controlling distributions can take their values in $\mathbb{R}^n$ for arbitrary n.

In the appendix we prove a simple version of the paralinearization theorem for $u \in C^\alpha$ with $\alpha \in (0,1/2)$, that is $F(u) - \pi_<(\mathrm{D}F(u), u) \in C^{2\alpha}$ for any $F \in C_b^2$. In fact we slightly generalize this statement, and allow for smooth (meaning $C^{2\alpha}$) perturbations of u. The crucial point is that the norm of the smooth perturbation appears only linearly in the estimate. This will be needed later to obtain global solutions for Burgers type SPDEs. Our paralinearization theorem is also a generalization of the classical result in the sense that we only require $F \in C_b^2$ and not $F \in C^\infty$, and our estimate for $\|F(u) - \pi_<(\mathrm{D}F(u), u)\|_{2\alpha}$ is more precise than the estimates that we could find in the literature. On the other side we restrict our attention to $\alpha \in (0,1/2)$.

Lemma 5.3.16 (Lemma G.3 in Appendix G)**.** *Let $\alpha \in (0,1/2)$, let $f \in C^\alpha(\mathbb{R}^d,\mathbb{R}^n)$, $u \in C^{2\alpha}(\mathbb{R}^d,\mathbb{R}^n)$, and $F \in C_b^2(\mathbb{R}^n,\mathbb{R})$. Then*

$$\|F(f+u) - \pi_<(\mathrm{D}F(f+u), f)\|_{2\alpha} \lesssim \|F\|_{C_b^2}(1+\|u\|_{2\alpha})(1+\|f\|_\alpha)^2.$$

In the appendix we also examine the Hölder-Besov regularity of $F(u)$ for $u \in C^\alpha$.

Lemma 5.3.17 (Lemma G.2 in Appendix G)**.** *Let $\alpha \in (0,1)$. If $f \in C^\alpha(\mathbb{R}^d,\mathbb{R}^n)$ and $F \in C_b^1(\mathbb{R}^n,\mathbb{R})$, then*

$$\|F(f)\|_\alpha \lesssim \|\mathrm{D}F\|_{L^\infty}\|f\|_\alpha + |F(0)| \lesssim \|F\|_{C_b^1}\|f\|_\alpha.$$

In fact Lemma 5.3.16 and Lemma 5.3.17 hold for general $\alpha > 0$, but the estimates get worse (polynomial rather than linear / quadratic) as α increases.

Now we are in a position to prove the first main result of this section, about the stability of controlled distributions under nonlinear maps.

Lemma 5.3.18. *Let $\alpha \in (0,1/2)$, let $v \in C^\alpha$, and let $f \in \mathcal{D}_v^{\alpha,\alpha}$, with derivative f^v. Let $u \in C^{2\alpha}$ and $F \in C_b^2$. Then $F(f+u) \in \mathcal{D}_v^{\alpha,\alpha}$ with derivative $\mathrm{D}F(f+u)f^v$, and*

$$\|F(f+u)\|_{v,\alpha,\alpha} \lesssim \|F\|_{C_b^2}(1+\|v\|_\alpha)(1+\|u\|_{2\alpha})(1+\|f\|_{v,\alpha,\alpha})^2.$$

Proof. The estimates for $\|F(f+u)\|_\alpha$ and $\|\mathrm{D}F(u+v)f^v\|_\alpha$ are straightforward, using Lemma 5.3.17 and the paraproduct estimates. For the remainder we have

$$\begin{aligned} F(f+u) - \pi_<(\mathrm{D}F(f+u)f^v, v) = {} & F(f+u) - \pi_<(\mathrm{D}F(f+u), f) \\ & + \pi_<(\mathrm{D}F(f+u), f - \pi_<(f^v, v)) \\ & + \pi_<(\mathrm{D}F(f+u), \pi_<(f^v, v)) - \pi_<(\mathrm{D}F(f+u)f^v, v). \end{aligned} \tag{5.8}$$

According to Lemma 5.3.16, the first term on the right hand side can be estimated by

$$\|F(f+u) - \pi_<(\mathrm{D}F(f+u), f)\|_{2\alpha} \lesssim \|F\|_{C_b^2}(1+\|u\|_{2\alpha})(1+\|f\|_\alpha)^2.$$

The paraproduct estimates imply for the second term on the right hand side of (5.8) that

$$\|\pi_<(\mathrm{D}F(f+u), f - \pi_<(f^v, v))\|_{2\alpha} \lesssim \|\mathrm{D}F\|_{L^\infty}\|f - \pi_<(f^v, v)\|_{2\alpha} \le \|F\|_{L^\infty}\|f\|_{v,\alpha,\alpha}.$$

Finally Lemma 5.3.14 yields that the third term on the right hand side of (5.8) can be estimated by

$$\begin{aligned}\|\pi_<(\mathrm{D}F(f+u),\pi_<(f^v,v))-\pi_<(\mathrm{D}F(f+u)f^v,v)\|_{2\alpha} &\lesssim \|\mathrm{D}F(f+u)\|_\alpha\|f^v\|_\alpha\|v\|_\alpha \\ &\lesssim \|F\|_{C_b^2}(\|f\|_\alpha+\|u\|_\alpha)\|f^v\|_\alpha\|v\|_\alpha,\end{aligned}$$

where we applied Lemma 5.3.17 in the last step. □

Next we prove the second main result of this section: If $\alpha\in(0,1)$, then the space of controlled distributions is an algebra.

Lemma 5.3.19. *Let $\alpha\in(0,1)$, $v\in C^\alpha(\mathbb{R}^d,\mathbb{R}^n)$, and $f,g\in\mathcal{D}_v^{\alpha,\alpha}(\mathbb{R})$, with derivatives f^v,g^v respectively. Then $fg\in\mathcal{D}_v^{\alpha,\alpha}$ with derivative f^vg+fg^v, and*

$$\|fg\|_{v,\alpha,\alpha}\lesssim\|f\|_{v,\alpha,\alpha}\|g\|_{v,\alpha,\alpha}(1+\|v\|_\alpha).$$

Proof. The estimates for $\|fg\|_\alpha$ and $\|f^vg+fg^v\|_\alpha$ are straightforward. The remainder can be decomposed as

$$fg=\pi_<(f,\pi_<(g^v,v))+\pi_>(\pi_<(f^v,v),g)+r,$$

where

$$\|r\|_{2\alpha}=\|\pi_<(f,g^\sharp)+\pi_\circ(f,g)+\pi_>(f^\sharp,g)\|_{2\alpha}\lesssim\|f\|_{v,\alpha,\alpha}\|g\|_{v,\alpha,\alpha}.$$

Therefore, it suffices to control $\|\pi_<(f,\pi_<(g^v,v))+\pi_>(\pi_<(f^v,v),g)-\pi_<(f^vg+fg^v,v)\|_{2\alpha}$. Lemma 5.3.14 implies that

$$\|\pi_<(f,\pi_<(g^v,v))-\pi_<(fg^v,v)\|_{2\alpha}\lesssim\|f\|_\alpha\|g^v\|_\alpha\|v\|_\alpha\lesssim\|f\|_{v,\alpha,\alpha}\|g\|_{v,\alpha,\alpha}\|v\|_\alpha.$$

Since f and g are both one dimensional, another application of Lemma 5.3.14 yields

$$\begin{aligned}\|\pi_>(\pi_<(f^v,v),g)-\pi_<(f^vg,v)\|_{2\alpha} &= \|\pi_<(g,\pi_<(f^v,v))-\pi_<(gf^v,v)\|_{2\alpha} \\ &\lesssim \|g\|_\alpha\|f^v\|_\alpha\|v\|_\alpha\lesssim\|f\|_{v,\alpha,\alpha}\|g\|_{v,\alpha,\alpha}\|v\|_\alpha,\end{aligned}$$

which completes the proof. □

5.3.4. Heat flow, paraproducts, and Fourier multipliers

In this section, which can be skipped at first reading, we examine how Fourier multipliers interact with the paraproduct. This will allow us to obtain commutator estimates between heat flow and paraproduct. We also quantify the smoothing effect of certain Fourier multipliers.

Lemma 5.3.20. *Let $\alpha<1$ and $\beta\in\mathbb{R}$. Let $\varphi\in\mathcal{S}$, let $u\in C^\alpha$, and $v\in C^\beta$. Then for* *every $\varepsilon > 0$ and every $\delta \geq -1$ we have*

$$\|\varphi(\varepsilon \mathrm{D})\pi_<(u,v) - \pi_<(u,\varphi(\varepsilon \mathrm{D})v)\|_{\alpha+\beta+\delta} \lesssim \varepsilon^{-\delta}\|u\|_\alpha \|v\|_\beta.$$

Proof. We have

$$\varphi(\varepsilon \mathrm{D})\pi_<(u,v) - \pi_<(u,\varphi(\varepsilon \mathrm{D})v) = \sum_{j\geq -1} (\varphi(\varepsilon \mathrm{D})(S_{j-1}u\Delta_j v) - S_{j-1}u\Delta_j\varphi(\varepsilon \mathrm{D})v),$$

and every term of this series has a Fourier transform with support in an annulus of the form $2^j\mathcal{A}$. Lemma 5.2.3 implies that it suffices to control the L^∞ norm of each term. Let $\psi \in \mathcal{D}$ with support in an annulus be such that $\psi \equiv 1$ on $\mathcal{A}$. We have

$$\begin{aligned}
&\varphi(\varepsilon \mathrm{D})(S_{j-1}u\Delta_j v) - S_{j-1}u\Delta_j\varphi(\varepsilon \mathrm{D})v \\
&\qquad = (\psi(2^{-j}\cdot)\varphi(\varepsilon\cdot))(\mathrm{D})(S_{j-1}u\Delta_j v) - S_{j-1}u(\psi(2^{-j}\cdot)\varphi(\varepsilon\cdot))(\mathrm{D})\Delta_j v \\
&\qquad = [(\psi(2^{-j}\cdot)\varphi(\varepsilon\cdot))(\mathrm{D}), S_{j-1}u]\Delta_j v,
\end{aligned}$$

where $[(\psi(2^{-j}\cdot)\varphi(\varepsilon\cdot))(\mathrm{D}), S_{j-1}u]$ denotes the commutator. In the proof of Lemma 2.97 in [BCD11] (our Lemma 5.3.1), it is shown that writing the Fourier multiplier as a convolution operator and applying a first order Taylor expansion and then Young's inequality yields

$$\begin{aligned}
&\|[(\psi(2^{-j}\cdot)\varphi(\varepsilon\cdot))(\mathrm{D}), S_{j-1}u]\Delta_j v\|_{L^\infty} \\
&\qquad \lesssim \sum_{\eta\in\mathbb{N}^d:|\eta|=1} \|x^\eta \mathcal{F}^{-1}(\psi(2^{-j}\cdot)\varphi(\varepsilon\cdot))\|_{L^1}\|\partial^\eta S_{j-1}u\|_{L^\infty}\|\Delta_j v\|_{L^\infty}.
\end{aligned} \tag{5.9}$$

Now $\mathcal{F}^{-1}(f(2^{-j}\cdot)g(\varepsilon\cdot)) = 2^{jd}\mathcal{F}^{-1}(fg(\varepsilon 2^j\cdot))(2^j\cdot)$ for every f, g, and thus we have for every multi-index η of order one

$$\begin{aligned}
&\|x^\eta \mathcal{F}^{-1}(\psi(2^{-j}\cdot)\varphi(\varepsilon\cdot))\|_{L^1} \\
&\qquad \leq 2^{-j}\|\mathcal{F}^{-1}((\partial^\eta\psi)(2^{-j}\cdot)\varphi(\varepsilon\cdot))\|_{L^1} + \varepsilon\|\mathcal{F}^{-1}(\psi(2^{-j}\cdot)\partial^\eta\varphi(\varepsilon\cdot))\|_{L^1} \\
&\qquad = 2^{-j}\|\mathcal{F}^{-1}((\partial^\eta\psi)\varphi(\varepsilon 2^j\cdot))\|_{L^1} + \varepsilon\|\mathcal{F}^{-1}(\psi\partial^\eta\varphi(\varepsilon 2^j\cdot))\|_{L^1} \\
&\qquad \lesssim 2^{-j}\|(1+|\cdot|)^{2d}\mathcal{F}^{-1}((\partial^\eta\psi)\varphi(\varepsilon 2^j\cdot))\|_{L^\infty} + \varepsilon\|(1+|\cdot|)^{2d}\mathcal{F}^{-1}(\psi\partial^\eta\varphi(\varepsilon 2^j\cdot))\|_{L^\infty} \\
&\qquad = 2^{-j}\|\mathcal{F}^{-1}((1-\Delta)^d((\partial^\eta\psi)\varphi(\varepsilon 2^j\cdot)))\|_{L^\infty} + \varepsilon\|\mathcal{F}^{-1}((1-\Delta)^d(\psi\partial^\eta\varphi(\varepsilon 2^j\cdot))\|_{L^\infty} \\
&\qquad \lesssim 2^{-j}\|(1-\Delta)^d((\partial^\eta\psi)\varphi(\varepsilon 2^j\cdot))\|_{L^\infty} + \varepsilon\|(1-\Delta)^d(\psi\partial^\eta\varphi(\varepsilon 2^j\cdot))\|_{L^\infty},
\end{aligned} \tag{5.10}$$

where the last step follows because ψ has compact support. For j satisfying $\varepsilon 2^j \geq 1$ we obtain

$$\|x^\eta \mathcal{F}^{-1}(\varphi(\varepsilon\cdot)\psi(2^{-j}\cdot))\|_{L^1} \lesssim (\varepsilon + 2^{-j})(\varepsilon 2^j)^{2d} \sum_{\eta:|\eta|\leq 2d+1} \|\partial^\eta\varphi(\varepsilon 2^j\cdot)\|_{L^\infty(\mathrm{supp}(\psi))}, \tag{5.11}$$

where we used that ψ and all its partial derivatives are bounded, and where $L^\infty(\mathrm{supp}(\psi))$

means that the supremum is taken over the values of $\partial^\eta \varphi(\varepsilon 2^j \cdot)$ restricted to $\operatorname{supp}(\psi)$. Now φ is a Schwartz function, and therefore it decays faster than any polynomial. Hence, there exists a ball $\mathcal{B}_\delta$ such that for all $x \notin \mathcal{B}_\delta$ and all $|\eta| \le 2d+1$ we have

$$|\partial^\eta \varphi(x)| \le |x|^{-2d-1-\delta}. \tag{5.12}$$

Let $j_0 \in \mathbb{N}$ be minimal such that $2^{j_0}\varepsilon\mathcal{A} \cap \mathcal{B}_\delta = \emptyset$ and $\varepsilon 2^{j_0} \ge 1$. Then the combination of (5.9), (5.11), and (5.12) shows for all $j \ge j_0$ that

$$\begin{aligned}
&\|[(\psi(2^{-j}\cdot)\varphi(\varepsilon\cdot))(\mathrm{D}), S_{j-1}u]\Delta_j v\|_{L^\infty} \\
&\qquad \lesssim (\varepsilon + 2^{-j})(\varepsilon 2^j)^{2d} \sum_{\eta:|\eta|\le 2d+1} \|(\partial^\eta\varphi)(\varepsilon 2^j\cdot)\|_{L^\infty(\operatorname{supp}(\psi))} 2^{j(1-\alpha)}\|u\|_\alpha 2^{-j\beta}\|v\|_\beta \\
&\qquad \lesssim (\varepsilon + 2^{-j})(\varepsilon 2^j)^{2d}(\varepsilon 2^j)^{-2d-1-\delta} 2^{j(1-\alpha-\beta)}\|u\|_\alpha\|v\|_\beta \\
&\qquad \lesssim (1 + (\varepsilon 2^j)^{-1})\varepsilon^{-\delta} 2^{-j(\alpha+\beta+\delta)}\|u\|_\alpha\|v\|_\beta.
\end{aligned}$$

Here we used that $\alpha < 1$ in order to obtain $\|\partial^\eta S_{j-1}u\|_{L^\infty} \lesssim 2^{j(1-\alpha)}\|u\|_{L^\infty}$. Since $\varepsilon 2^j \ge 1$, we have shown the desired estimate for $j \ge j_0$. On the other side Lemma 5.3.1 implies for every $j \ge -1$ that

$$\|[\varphi(\varepsilon\mathrm{D}), S_{j-1}u]\Delta_j v\|_{L^\infty} \lesssim \varepsilon \max_{\eta\in\mathbb{N}^d:|\eta|=1} \|\partial^\eta S_{j-1}u\|_{L^\infty}\|\Delta_j v\|_{L^\infty} \lesssim \varepsilon 2^{j(1-\alpha-\beta)}\|u\|_\alpha\|v\|_\beta.$$

Hence, we obtain for $j < j_0$, i.e. for j satisfying $2^j\varepsilon \lesssim 1$, that

$$\|[\varphi(\varepsilon\mathrm{D}), S_{j-1}u]\Delta_j v\|_{L^\infty} \lesssim (\varepsilon 2^j)^{1+\delta}\varepsilon^{-\delta}2^{-j(\alpha+\beta+\delta)}\|u\|_\alpha\|v\|_\beta \lesssim \varepsilon^{-\delta}2^{-j(\alpha+\beta+\delta)}\|u\|_\alpha\|v\|_\beta,$$

where we used that $\delta \ge -1$. This completes the proof. □

The same arguments allow us to quantify the smoothing properties of Fourier multipliers that behave like Schwartz functions for large $|x|$. In particular, we will be able quantify the smoothing properties of the semigroup generated by the fractional Laplacian. Of course these are well known, but at this point we can give a short proof.

Corollary 5.3.21. *Let $\alpha \in \mathbb{R}$. Let φ be a continuous function, such that φ is smooth outside of a ball centered at 0, and such that φ and all its partial derivatives decay faster than any polynomial at infinity. Assume also that $\mathcal{F}\varphi \in L^1$. Then for all $u \in C^\alpha$, $\varepsilon > 0$, and $\delta \ge 0$ we have*

$$\|\varphi(\varepsilon\mathrm{D})u\|_{\alpha+\delta} \lesssim \varepsilon^{-\delta}\|u\|_\alpha.$$

Proof. Let $\psi \in \mathcal{D}$ with support in an annulus be such that $\psi\rho = \rho$, where (χ, ρ) is our dyadic partition of unity. For $j \ge 0$ we have

$$\varphi(\varepsilon\mathrm{D})\Delta_j u = \mathcal{F}^{-1}(\varphi(\varepsilon\cdot)\psi(2^{-j}\cdot)) * \Delta_j u,$$

and therefore Young's inequality implies that

$$\|\varphi(\varepsilon\mathrm{D})\Delta_j u\|_{L^\infty} \lesssim \|\mathcal{F}^{-1}(\varphi(\varepsilon\cdot)\psi(2^{-j}\cdot))\|_{L^1} 2^{-j\alpha}\|u\|_\alpha.$$

Hence, it suffices to prove $\|\mathcal{F}^{-1}(\varphi(\varepsilon\cdot)\psi(2^{-j}\cdot))\|_{L^1} \lesssim \varepsilon^{-\delta}2^{-j\delta}$. For j large enough so that $2^j\varepsilon \geq 1$ and $|\partial^\eta\varphi(x)| \leq |x|^{-2d-1-\delta}$ for all $\eta \in \mathbb{N}^d$ with $|\eta| \leq 2d$ and for all $x \in \mathrm{supp}(\psi(2^j\varepsilon\cdot))$, this is shown exactly as in the proof of Lemma 5.3.20: just omit the factor x^η in the derivation of (5.10). But for $\varepsilon 2^j \lesssim 1$ we have

$$\begin{aligned}\|\varphi(\varepsilon\mathrm{D})\Delta_j u\|_{L^\infty} &\leq \|\mathcal{F}^{-1}(\varphi(\varepsilon\cdot))\|_{L^1}\|\Delta_j u\|_{L^\infty} \lesssim \|\mathcal{F}^{-1}\varphi\|_{L^1}2^{-j\alpha}\|u\|_\alpha \\ &\lesssim 2^{-j\alpha}\|u\|_\alpha = (\varepsilon 2^j)^\delta\varepsilon^{-\delta}2^{-j(\alpha+\delta)}\|u\|_\alpha \lesssim \varepsilon^{-\delta}2^{-j(\alpha+\delta)}\|u\|_\alpha,\end{aligned}$$

where in the last step we used that $\delta \geq 0$. □

5.4. Rough Burgers type equation

A first example on which to test our theory is the Burgers type SPDE

$$\partial_t u(t,x) = -Au(t,x) + G(u(t,x))\mathrm{D}_x u(t,x) + \dot{W}(t,x), \tag{5.13}$$

where $-A = -(-\Delta)^\sigma$ is the fractional Laplacian for a sufficiently large σ, the Gaussian noise $\dot{W}$ is white in space and time and takes its values in $\mathbb{R}^n$, and the spatial derivative is denoted by D_x. Moreover, $G : \mathbb{R}^n \to \mathcal{L}(\mathcal{L}(\mathbb{T}^d, \mathbb{R}^n), \mathbb{R}^n)$ is a smooth map. We consider this equation on the d–dimensional torus $\mathbb{T}^d = [-\pi,\pi]^d$ with periodic boundary conditions. Recall from Section 5.2 that our results on $\mathbb{R}^d$ carry over to this setting without problem. Note that if $n > 1$, then in general there exists no function $\Gamma : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$G(u(t,x))\mathrm{D}_x u(t,x) = \mathrm{D}_x(\Gamma(u(t,x))).$$

This prevents a direct definition of the term $G(u(t,x))\mathrm{D}_x u(t,x)$ as a distribution. Moreover, we will see that the natural regularity of u is too low to define $G(u(t,x))\mathrm{D}_x u(t,x)$ using Bony's paraproduct.

In the case of a one dimensional spatial index set, Hairer [Hai11] uses a flow decomposition to make sense of (5.13): Consider the *stationary* solution ψ to

$$\partial_t\psi(t,x) = (-A-\lambda)\psi(t,x) + \dot{W}(t,x)$$

for a correction term $\lambda > 0$ (needed for the existence of stationary solutions). If we set $v := u - \psi$, then v formally solves

$$\partial_t v(t,x) = -Av(t,x) + G(v(t,x)+\psi(t,x))\mathrm{D}_x(v(t,x)+\psi(t,x)) + \lambda\psi(t,x).$$

The main problem then consists of making sense of the term

$$G(v(s,\cdot)+\psi(s,\cdot))\mathrm{D}_x(v(s,\cdot)+\psi(s,\cdot)).$$

Hairer's key insight is that for $d = 1$, the theory of controlled rough paths can be applied to make sense of the product. An extension of this approach to higher dimensions is naturally provided by the theory of paracontrolled distributions. As we will show, we can make sense of the product $G(v(s,\cdot)+\psi(s,\cdot))\mathrm{D}_x(v(s,\cdot)+\psi(s,\cdot))$ by devising a suitable area distribution for ψ. We then combine the continuity properties of our product operator with smoothing properties of the heat flow to prove the existence and uniqueness of solutions to (5.13).

5.4.1. Construction of the Besov area

The first step is to analyze the stationary solution ψ to

$$\partial_t \psi(t,x) = (-A-\lambda)\psi(t,x) + \dot{W}(t,x). \tag{5.14}$$

For the existence of ψ, we suppose that the space-time white noise $\dot{W}$ is defined on $\mathbb{R}\times\mathbb{T}^d$. Recall that $\dot{W}$ is a space-time white noise if it is a mean zero Gaussian process with values in $\mathcal{S}'(\mathbb{R}\times\mathbb{T}^d)$, such that

$$E(\langle \dot{W}^j, \varphi\rangle\langle \dot{W}^\ell, \vartheta\rangle) = \delta_{j\ell}\int_{\mathbb{R}\times\mathbb{T}^d}\varphi(t,x)\vartheta(t,x)\mathrm{d}t\mathrm{d}x$$

for all test function $\varphi, \vartheta \in L^2(\mathbb{R}\times\mathbb{T}^d)$. Formally we write $E(\dot{W}^j(t,x)\dot{W}^\ell(t',x')) = \delta_{j\ell}\delta(t-t')\delta(x-x')$, where δ denotes the Dirac delta but $\delta_{j\ell}$ denotes the Kronecker delta.

In this setting, the stationary solution ψ is given as

$$\psi(t,x) = \int_{-\infty}^{t}(P^\lambda_{t-s}\dot{W}(s,\cdot))(x)\mathrm{d}s,$$

where $(P^\lambda_t)_{t\geq 0} = (e^{-\lambda t}(e^{-t|\cdot|^{2\sigma}})(\mathrm{D}))_{t\geq 0}$ is the semigroup generated by $-A-\lambda$. The reason for considering the stationary solution is that its Fourier transform has a particularly simple covariance structure, which is convenient in the following calculations.

Lemma 5.4.1. *Let $\widehat{\psi}$ be the spatial Fourier transform of ψ, i.e.*

$$\widehat{\psi}(t,k) = \int_{\mathbb{T}^d} e^{-\imath\langle k,x\rangle}\psi(t,x)\mathrm{d}x \text{ for } k \in \mathbb{Z}^d.$$

Then $\widehat{\psi}$ is a complex-valued stationary Gaussian process with zero mean and covariance

$$E(\widehat{\psi}^j(t,k)\widehat{\psi}^\ell(t',k')) = (2\pi)^d\delta_{j\ell}\delta_{k(-k')}\frac{e^{-|t'-t|(\lambda+|k|^{2\sigma})}}{2(\lambda+|k|^{2\sigma})} \tag{5.15}$$

for $j,\ell = 1,\dots,n$, $k,k' \in \mathbb{Z}^d$ and $t,t' \in \mathbb{R}$.

Proof. We give a formal derivation, which can be rendered rigorous by considering the action on test functions. For the spatial Fourier transform $\widehat{\dot{W}}$ of $\dot{W}$ we obtain

$$\begin{aligned}
E\Big(\widehat{\dot{W}^j}(t,k)\widehat{\dot{W}^\ell}(t',k')\Big) &= \int_{\mathbb{T}^d}\int_{\mathbb{T}^d} e^{-\imath\langle k,x\rangle}e^{-\imath\langle k',x'\rangle}E(\dot{W}^j(t,x)\dot{W}^\ell(t',x'))\mathrm{d}x'\mathrm{d}x \\
&= \delta_{j\ell}\delta(t-t')\int_{\mathbb{T}^d} e^{-\imath\langle k,x\rangle}e^{-\imath\langle k',x\rangle}\mathrm{d}x \\
&= (2\pi)^d\delta_{j\ell}\delta(t-t')\delta_{k(-k')}.
\end{aligned}$$

Hence, the covariance of $\widehat{\psi}$ is given by

$$\begin{aligned}
&E(\widehat{\psi}^j(t,k)\widehat{\psi}^\ell(t',k')) \\
&\qquad = \int_{-\infty}^{t}\int_{-\infty}^{t'} e^{-(t-s)(\lambda+|k|^{2\sigma})}e^{-(t'-s')(\lambda+|k'|^{2\sigma})}E\Big(\widehat{\dot{W}^j}(s,k)\widehat{\dot{W}^\ell}(s',k')\Big)\mathrm{d}s'\mathrm{d}s \\
&\qquad = (2\pi)^d\delta_{j\ell}\delta_{k(-k')}\int_{-\infty}^{t\wedge t'} e^{-(t-s)(\lambda+|k|^{2\sigma})}e^{-(t'-s)(\lambda+|k|^{2\sigma})}\mathrm{d}s \\
&\qquad = (2\pi)^d\delta_{j\ell}\delta_{k(-k')}\frac{e^{2(t\wedge t')(\lambda+|k|^{2\sigma})}}{2(\lambda+|k|^{2\sigma})}e^{-(t+t')(\lambda+|k|^{2\sigma})} \\
&\qquad = (2\pi)^d\delta_{j\ell}\delta_{k(-k')}\frac{e^{-|t'-t|(\lambda+|k|^{2\sigma})}}{2(\lambda+|k|^{2\sigma})}.
\end{aligned}$$

□

Our first concern is to study the Hölder-Besov regularity of the process ψ.

Lemma 5.4.2. *The process ψ is almost surely in $C([0,T],C^\alpha(\mathbb{T}^d,\mathbb{R}^n))$ for any $\alpha < \sigma - d/2$.*

Proof. Let $s,t\in[0,T]$ and $\ell\geq -1$. Recall that $f_{s,t} := f(t)-f(s)$, for any f defined on $[0,T]$. Writing Δ_ℓ as a convolution operator, we see that $\Delta_\ell\psi_{s,t}$ is a Gaussian process indexed by $\mathbb{T}^d$, with values in $\mathbb{R}^n$. Using Gaussian hypercontractivity we obtain for $p\geq 1$ that

$$E\left(\|\Delta_\ell\psi_{s,t}\|^{2p}_{L^{2p}(\mathbb{T}^d)}\right) \lesssim_p \|E(|\Delta_\ell\psi_{s,t}(x)|^2)\|^p_{L^p_x(\mathbb{T}^d)}, \tag{5.16}$$

where we denote $\|f(x)\|^p_{L^p_x(\mathbb{T}^d)} := \int_{\mathbb{T}^d}|f(x)|^p\mathrm{d}x$ for any $f\in L^1(\mathbb{T}^d)$. By definition we have

$$\Delta_\ell\psi_{s,t}(x) = (2\pi)^{-d}\sum_{k\in\mathbb{Z}^d}\rho_\ell(k)\widehat{\psi}_{s,t}(k)e^{\imath\langle k,x\rangle},$$

and therefore

$$E(|\Delta_\ell\psi_{s,t}(x)|^2) = (2\pi)^{-2d}\sum_{k,k'}\rho_\ell(k)\rho_\ell(k')e^{\imath\langle k+k',x\rangle}E(\widehat{\psi}_{s,t}(k)\widehat{\psi}_{s,t}(k')).$$

Since ρ_ℓ has compact support we are considering finite sums, and therefore exchanging expectation and summation is justified. Now (5.15) implies for $1 \leq j \leq n$ and independently of $x \in \mathbb{T}^d$ that

$$E(|\Delta_\ell \psi^j_{s,t}(x)|^2) = (2\pi)^{-d} \sum_{k\in\mathbb{Z}^d} \rho^2_\ell(k) \frac{1 - e^{-|t-s|(\lambda+|k|^{2\sigma})}}{\lambda + |k|^{2\sigma}}.$$

For any $\varepsilon \in (0, 1]$ we have $1 - e^{-x} \lesssim x^\varepsilon$, and therefore

$$\begin{aligned}(2\pi)^{-d} \sum_{k\in\mathbb{Z}^d} \rho^2_\ell(k) \frac{1 - e^{-|t-s|(\lambda+|k|^{2\sigma})}}{\lambda + |k|^{2\sigma}} &\lesssim |t-s|^\varepsilon \sum_{k\in\operatorname{supp}(\rho_\ell)} \frac{1}{(\lambda + |k|^{2\sigma})^{1-\varepsilon}} \\ &\lesssim |t-s|^\varepsilon 2^{\ell(d-2\sigma(1-\varepsilon))}.\end{aligned}$$

Hence, we obtain from (5.16) that

$$\begin{aligned}E\left(\|\psi(t,\cdot) - \psi(s,\cdot)\|^{2p}_{B^\alpha_{2p,2p}(\mathbb{T}^d)}\right) &\lesssim \sum_{\ell\geq -1} 2^{\ell\alpha 2p} E\left(\|\Delta_\ell \psi_{s,t}\|^{2p}_{L^{2p}(\mathbb{T}^d)}\right) \\ &\lesssim \sum_{\ell\geq -1} 2^{\ell\alpha 2p} \left(|t-s|^\varepsilon 2^{\ell(d-2\sigma(1-\varepsilon))}\right)^p\end{aligned}$$

for any $\alpha \in \mathbb{R}$ and any $p \geq 1$. For $\alpha < \sigma - d/2$ there exists $\varepsilon \in (0, 1]$ small enough so that the series converges. Since we can choose p arbitrarily large, Kolmogorov's continuity criterion for Banach space valued processes, Theorem I.2.1 of [RY99], implies that ψ has a continuous version such that $\psi \in C([0,T], B^\alpha_{2p,2p}(\mathbb{T}^d))$ for all $\alpha < \sigma - d/2$. Now we use again that p can be chosen arbitrarily large, so that the Besov embedding theorem, Lemma 5.2.4, shows that this continuous version takes its values in $C([0,T], C^\alpha(\mathbb{T}^d))$ for all $\alpha < \sigma - d/2$. □

Contrary to the one dimensional case, in higher dimensions the Laplacian does not sufficiently smoothen the white noise. In fact Lemma 5.4.2 shows that if $\sigma = 1$, then already for $d = 2$ we only have a distribution valued solution $\psi \in C([0,T], C^{-\varepsilon})$ for any $\varepsilon > 0$. Hence, in that case it is not even clear how to define $G(\psi(t,\cdot))$, let alone $G(\psi(t,\cdot))\mathrm{D}_x\psi(t,\cdot)$. So in higher dimensions we need to consider the fractional Laplacian of a sufficiently high order.

If $\sigma - d/2 \in (0, 1/2)$, then $G(\psi)$ makes sense, but the product $G(\psi)\mathrm{D}_x\psi$ cannot be defined using classical analytical arguments. This is why prior to [Hai11] it was not known how to describe solutions to (5.13). Hairer solved the case $d = 1$, and in the following we show how to use our paracontrolled calculus in order to solve (5.13) for $d \geq 1$. For this purpose we need to construct the area process of $\psi(t,\cdot)$.

Lemma 5.4.3. *Let ψ be the stationary solution to* (5.14)*. If $1 + d/2 - 2\sigma < 0$, then for any $\alpha < \sigma - d/2$ almost surely*

$$\pi_\circ(\psi, \mathrm{D}_x\psi) = (\pi_\circ(\psi^i, \mathrm{D}_x\psi^j))_{1\leq i,j\leq n} \in C([0,T]; C^{2\alpha-1}(\mathbb{T}^d; \mathbb{R}^{n\times n})).$$

Proof. Without loss of generality we can argue for $\pi_\circ(\psi^1, \mathrm{D}_x\psi^2)$. The case $\pi_\circ(\psi^1, \mathrm{D}_x\psi^1)$ is easy, because Leibniz's rule implies that $\pi_\circ(\psi^1, \mathrm{D}_x\psi^1) = \frac{1}{2}\mathrm{D}_x(\pi_\circ(\psi^1, \psi^1))$.

Note that if i is smaller than $\ell - N$ for sufficiently large N, and if $|i-j| \leq 1$, then $\Delta_\ell(\Delta_i f \Delta_j g) = 0$ for all $f, g \in \mathcal{S}'$. Hence, the projection of $\pi_\circ(\psi^1, \mathrm{D}_x\psi^2)$ onto the ℓ–th dyadic Fourier block is given by

$$\Delta_\ell \pi_\circ(\psi^1, \mathrm{D}_x\psi^2) = \sum_{|i-j|\leq 1} \Delta_\ell(\Delta_i\psi^1 \Delta_j \mathrm{D}_x\psi^2) = \sum_{|i-j|\leq 1} 1_{\ell \lesssim i}\Delta_\ell(\Delta_i\psi^1 \Delta_j \mathrm{D}_x\psi^2),$$

Therefore, we can apply Gaussian hypercontractivity (see [FV10b], Appendix D.4) to obtain

$$E\|[\Delta_\ell\pi_\circ(\psi^1, \mathrm{D}_x\psi^2)]_{s,t}\|^{2p}_{L^{2p}(\mathbb{T}^d)} \lesssim \left\| E\Big(\Big(\sum_{|i-j|\leq 1} 1_{\ell\lesssim i}[\Delta_\ell(\Delta_i\psi^1\Delta_j\mathrm{D}_x\psi^2)(x)]_{s,t}\Big)^2\Big)\right\|^p_{L^p_x(\mathbb{T}^d)}. \tag{5.17}$$

Let us start by estimating

$$\begin{aligned} &E\Big(\Big(\sum_{|i-j|\leq 1} 1_{\ell\lesssim i}\Delta_\ell(\Delta_i\psi^1(t,\cdot)\Delta_j\mathrm{D}_x\psi^2_{s,t})(x)\Big)^2\Big) \qquad (5.18)\\ &\quad = \sum_{|i-j|\leq 1}\sum_{|i'-j'|\leq 1} 1_{\ell\lesssim i}1_{\ell\lesssim i'} E(\Delta_\ell(\Delta_i\psi^1(t,\cdot)\Delta_j\mathrm{D}_x\psi^2_{s,t})(x)\Delta_\ell(\Delta_{i'}\psi^1(t,\cdot)\Delta_{j'}\mathrm{D}_x\psi^2_{s,t})(x)). \end{aligned}$$

Taking the infinite sums outside of the expectation can be justified a posteriori, because for every finite partial sum we will obtain a bound on the L^2–norm below, which does not depend on the number of terms that we sum up. The Gaussian hypercontractivity (5.17) then provides a uniform L^p–bound for any $p \geq 2$, which implies that the squares of the partial sums are uniformly integrable, and thus allows us to exchange summation and expectation.

Let us write $\mathcal{F}$ for the spatial Fourier transform. Recall that $\mathcal{F}(\mathcal{F}^{-1}u\mathcal{F}^{-1}v)(k) = (2\pi)^{-d}\sum_{k'} u(k')v(k-k')$, and $\mathcal{F}(\mathrm{D}_x u)(k) = \imath k\mathcal{F}(u)(k)$, and therefore

$$\begin{aligned} \Delta_\ell(\Delta_i\psi^1(t,\cdot)\Delta_j\mathrm{D}_x\psi^2_{s,t})(x) &= (2\pi)^{-d}\sum_{k\in\mathbb{Z}^d}\rho_\ell(k)e^{\imath\langle k,x\rangle}\mathcal{F}(\Delta_i\psi^1(t,\cdot)\Delta_j\mathrm{D}_x\psi^2_{s,t})(k)\\ &= (2\pi)^{-d}\sum_{k,k'\in\mathbb{Z}^d}\rho_\ell(k)e^{\imath\langle k,x\rangle}\rho_i(k')\mathcal{F}(\psi^1)(t,k')\rho_j(k-k')\imath(k-k')\mathcal{F}(\psi^2)_{s,t}(k-k'). \end{aligned}$$

Using the explicit covariance (5.15) and the independence of ψ^1 and ψ^2, we thus obtain

$$\begin{aligned} &E\Big(\Big(\sum_{|i-j|\leq 1} 1_{\ell\lesssim i}\Delta_\ell(\Delta_i\psi^1(t,\cdot)\Delta_j\mathrm{D}_x\psi^2_{s,t})(x)\Big)^2\Big)\\ &\quad = \sum_{|i-j|\leq 1}\sum_{|i'-j'|\leq 1} 1_{\ell\lesssim i}1_{\ell\lesssim i'}\sum_{k,k'\in\mathbb{Z}^d}\rho^2_\ell(k+k')\rho_i(k)\rho_{i'}(k)\rho_j(k')\rho_{j'}(k') \end{aligned}$$

$$\times |k'|^2 \frac{1-e^{-|t-s|(\lambda+|k'|^{2\sigma})}}{\lambda+|k'|^{2\sigma}} \frac{1}{2(\lambda+|k|^{2\sigma})}$$
$$\lesssim \sum_{|i-j|\le 1} 1_{\ell\lesssim i} \sum_{k\in \operatorname{supp}(\rho_i), k'\in\operatorname{supp}(\rho_j)} \rho_\ell^2(k+k')|k'|^2 \frac{1-e^{-|t-s|(\lambda+|k'|^{2\sigma})}}{2(\lambda+|k'|^{2\sigma})^2}.$$

For any $\varepsilon \in (0,1]$ we can estimate the right hand side by

$$\lesssim \sum_{|i-j|\le 1} 1_{\ell\lesssim i} \sum_{k\in \operatorname{supp}(\rho_i), k'\in\operatorname{supp}(\rho_j)} \rho_\ell^2(k+k')|k'|^2 \frac{|t-s|^\varepsilon(\lambda+|k'|^{2\sigma})^\varepsilon}{2(\lambda+|k'|^{2\sigma})^2}$$
$$\lesssim \sum_{i:i\gtrsim \ell} 2^{di}2^{d\ell}(2^i)^2 \frac{|t-s|^\varepsilon (2^{i2\sigma})^\varepsilon}{(2^{i2\sigma})^2} = 2^{d\ell} \sum_{i:i\gtrsim\ell} 2^{2i(1+d/2-2\sigma(1-\varepsilon))}|t-s|^\varepsilon.$$

So if $1+d/2-2\sigma<0$, then we obtain for sufficiently small $\varepsilon>0$ that

$$E\Big(\Big(\sum_{|i-j|\le 1} 1_{\ell\lesssim i}\Delta_\ell(\Delta_i\psi^1(t,\cdot)\Delta_j \mathrm{D}_x\psi^2_{s,t})(x)\Big)^2\Big) \lesssim 2^{2\ell(1+d-2\sigma(1-\varepsilon))}|t-s|^\varepsilon,$$

and by the same arguments

$$E\Big(\Big(\sum_{|i-j|\le 1} 1_{\ell\lesssim i}\Delta_\ell(\Delta_i\psi^1_{s,t}\Delta_j \mathrm{D}_x\psi^2(s,\cdot))(x)\Big)^2\Big) \lesssim 2^{2\ell(1+d-2\sigma(1-\varepsilon))}|t-s|^\varepsilon.$$

Since

$$\Delta_i\psi^1(t,\cdot)\Delta_j\mathrm{D}_x\psi^2(t,\cdot) - \Delta_i\psi^1(s,\cdot)\Delta_j\mathrm{D}_x\psi^2(s,\cdot)$$
$$= \Delta_i\psi^1(t,\cdot)\Delta_j\mathrm{D}_x\psi^2_{s,t} + \Delta_i\psi^1_{s,t}\Delta_j\mathrm{D}_x\psi^2(s,\cdot),$$

we get for sufficiently small $\varepsilon>0$ and for arbitrarily large $p\ge 1$ that

$$E\|\Delta_\ell\pi_\circ(\psi^1_t,\mathrm{D}_x\psi^2_t) - \Delta_\ell\pi_\circ(\psi^1_s,\mathrm{D}_x\psi^2_s)\|^{2p}_{L^{2p}(\mathbb{T}^d)} \lesssim 2^{\ell(1+d-2\sigma(1-\varepsilon))2p}|t-s|^{\varepsilon p}.$$

Now we use the same arguments as in the proof of Lemma 5.4.2 to conclude that $\pi_\circ(\psi^1,\mathrm{D}_x\psi^2_t)$ has a version that almost surely belongs to $C([0,T],C^{2\alpha-1})$ for all $\alpha<\sigma-d/2$. $\square$

Remark 5.4.4. We constructed the Besov area $\psi \mathrm{D}_x\psi$ using one fixed partition of unity. But if $(\widetilde{\chi},\widetilde{\rho})$ is another dyadic partition of unity, with dyadic blocks $(\widetilde{\Delta}_i)_{i\ge -1}$, and if $\psi\widetilde{\circ}\mathrm{D}_x\psi$ is constructed using that partition, then $\psi\mathrm{D}_x\psi$ and $\psi\widetilde{\circ}\mathrm{D}_x\psi$ almost surely coincide. This can be seen by considering the difference

$$\sum_{i,j\le N}\Delta_i\psi\Delta_j\mathrm{D}_x\psi - \sum_{i',j'\le N}\widetilde{\Delta}_{i'}\psi\widetilde{\Delta}_{j'}\mathrm{D}_x\psi,$$

and by showing that it converges to 0 as N tends to ∞.

More generally, we expect that for a wide range of smooth approximations $(\psi_j)_{j\in\mathbb{N}}$ to ψ, the product $(\pi_\circ(\psi_j, \mathrm{D}_x\psi_j))$ almost surely converges to $\pi_\circ(\psi, \mathrm{D}_x\psi)$. In the setting of Gaussian rough paths, such results have been obtained by Friz and Victoir [FV10a]. See also [FGGR12] for the case of SPDEs with a one dimensional spatial index variable, and Chapter 10 of [Hai13a] for some general results in the multidimensional case.

5.4.2. Picard iteration

Having constructed the Besov area of the driving noise, we can now solve the fractional Burgers type equation. For this purpose, we first have to define what we mean by a solution.

For $t \geq 0$ and $\alpha \in \mathbb{R}$ we define the space $C_T^\alpha := C([0,T], C^\alpha(\mathbb{T}^d, \mathbb{R}^n))$ with norm

$$\|u\|_{C_T^\alpha} := \sup_{s\in[0,T]} \|u_s\|_\alpha,$$

where we write $u_s := u(s,\cdot)$.

Definition 5.4.5. Let $d \in \mathbb{N}$, $\sigma - d/2 > 1/3$, $\varepsilon > 0$, and $\alpha \in (1/3, \sigma - d/2)$. Let $u_0 \in C^\alpha$. A function $u \in C_T^\alpha$ is called *mild solution* to

$$\partial_t u(t,x) = -Au(t,x) + G(u(t,x))\mathrm{D}_x u(t,x) + \dot{W}(t,x), \tag{5.19}$$

with initial condition u_0, if $v := u - \psi$ is in $C_T^{1+\varepsilon}$, and

$$\begin{aligned} v(t,\cdot) = P_t(u_0 - \psi_0)(x) &+ \int_0^t P_{t-s}[G(v(s,\cdot) + \psi(s,\cdot))\mathrm{D}_x(v(s,\cdot) + \psi(s,\cdot))]\mathrm{d}s \\ &+ \lambda \int_0^t P_{t-s}\psi(s,\cdot)\mathrm{d}s, \end{aligned}$$

where the product $G(v(s,\cdot)+\psi(s,\cdot))\mathrm{D}_x(v(s,\cdot)+\psi(s,\cdot))$ is as in Section 5.3: Since $v(s,\cdot) \in C^{1+\varepsilon}$, we have $v(s,\cdot) + \psi(s,\cdot) \in \mathcal{D}_\psi^{\alpha,\alpha}$, and $\mathrm{D}_x(v(s,\cdot) + \psi(s,\cdot)) \in \mathcal{D}_{\mathrm{D}_x\psi}^{\alpha-1,\alpha}$.

Remark 5.4.6. A priori this definition depends on the constant $\lambda > 0$ that we introduced to obtain a stationary solution to the linear part of the equation. But if ψ solves $\partial_t\psi(t,x) = -(A+\lambda)\psi(t,x) + \dot{W}(t,x)$, and $\widetilde{\psi}$ solves $\partial_t\widetilde{\psi}(t,x) = -(A+\widetilde{\lambda})\widetilde{\psi}(t,x) + \dot{W}(t,x)$, then

$$\partial_t(\psi - \widetilde{\psi})(t,x) = -A(\psi - \widetilde{\psi})(t,x) + \widetilde{\lambda}\widetilde{\psi}(t,x) - \lambda\psi(t,x),$$

and from here it is easy to see that $\psi - \widetilde{\psi}$ is smooth. Therefore, any $w \in \mathcal{D}_\psi^{\alpha,\alpha}$ is also in $\mathcal{D}_{\widetilde{\psi}}^{\alpha,\alpha}$, which implies that the definition of the product $P_{t-s}G(v(s,\cdot) + \psi(s,\cdot))\mathrm{D}_x(v(s,\cdot) + \psi(s,\cdot))$ does not depend on the special choice of λ.

We also could have taken

$$\widetilde{\psi}(t,x) = \int_0^t (P_{t-s}\dot{W}(s,\cdot))(x)\mathrm{d}s$$

as reference distribution, and not the stationary solution ψ. While it is harder to derive the regularity of $\widetilde{\psi}$ directly, it is easy to show that the difference $\psi - \widetilde{\psi}$ is smooth.

We could also define weak solutions. It should be no problem to show that mild and weak solutions coincide.

For further details on these questions see [HW13], Definition 3.1 and the following discussion.

Before proving the existence of a unique solution to (5.19), we establish some a priori estimates. To lighten the notation, we introduce a "rough path norm".

Definition 5.4.7. Let $\alpha > 1/3$ and $w \in C^\alpha(\mathbb{T}^d, \mathbb{R}^n)$ with associated Besov area $w\mathrm{D}_x w \in C^{2\alpha-1}(\mathbb{T}^d, \mathbb{R}^n \otimes \mathcal{L}(\mathbb{T}^d, \mathbb{R}^n))$. In that case we write $\mathbb{w} := (w, w\mathrm{D}_x w) \in \mathscr{C}^\alpha$, and we define the "norm"

$$\|\mathbb{w}\|_{\mathscr{C}^\alpha} := \|\mathrm{D}_x w\|_{\alpha-1} + \|w\|_\alpha \|\mathrm{D}_x w\|_{\alpha-1} + \|\pi_\circ(w, \mathrm{D}_x w)\|_{2\alpha-1}.$$

For $t > 0$ we also introduce the space $\mathscr{C}^\alpha_t := C([0,t], \mathscr{C}^\alpha)$ with "norm"

$$\|\mathbb{W}\|_{\mathscr{C}^\alpha_t} := \sup_{s\in[0,t]} \|\mathbb{W}(s,\cdot)\|_{\mathscr{C}^\alpha}.$$

We first recall the smoothing properties of the semigroup generated by the fractional Laplacian.

Lemma 5.4.8. *Let $\sigma > 1/2$ if d is odd, and $\sigma > 1$ if d is even. Let $A = -(-\Delta)^\sigma$ with periodic boundary conditions on $\mathbb{T}^d$ and let $(P_t)_{t\ge 0}$ be the semigroup generated by A. Let $\alpha \in \mathbb{R}$, and let $u \in C^\alpha$, $t > 0$, $\delta \ge 0$. Then*

$$\|P_t u\|_{\alpha+\delta} \lesssim t^{-\frac{\delta}{2\sigma}} \|u\|_\alpha.$$

For $\alpha \le 1$, $u \in C^{\alpha-1}$, $v \in C^\alpha$, and $\varepsilon \ge 0$ we obtain

$$\|P_t u\|_{1+\varepsilon} \lesssim t^{-\frac{2+\varepsilon-\alpha}{2\sigma}} \|u\|_{\alpha-1} \qquad \text{and} \qquad \|P_t v\|_{1+\varepsilon} \lesssim t^{-\frac{1+\varepsilon-\alpha}{2\sigma}} \|v\|_\alpha.$$

Proof. The semigroup is given by $P_t = \varphi(t^{1/(2\sigma)}\mathrm{D})$, $t \ge 0$, where $\varphi(x) = e^{-|x|^{2\sigma}}$. Hence, it suffices to show that φ satisfies the assumptions of Corollary 5.3.21. Outside of every ball that contains the singularity 0 (where φ is not infinitely differentiable), φ behaves like a Schwartz function. Let us show that $\mathcal{F}\varphi \in L^1$. Let m be the smallest even integer that is strictly larger than d. Since $\|\mathcal{F}\varphi\|_{L^\infty} \lesssim \|\varphi\|_{L^1}$, it suffices to show that $|\cdot|^m \mathcal{F}\varphi \in L^\infty$. But

$$|x|^m |\mathcal{F}\varphi(x)| = |\mathcal{F}(\Delta^{m/2}\varphi)(x)| \lesssim \|\Delta^{m/2}\varphi\|_{L^1} \lesssim \sum_{\eta\in\mathbb{N}^d:|\eta|\le m} \|\partial^\eta \varphi\|_{L^1}.$$

If d is odd, then $2\sigma - m > 1-(d+1) = -d$, and therefore the right hand side is finite. The argument for even d is similar, and therefore the proof is complete. $\square$

Remark 5.4.9. We only used the lower bounds on σ to prove that $\mathcal{F}\varphi \in L^1$. This holds for every $\sigma > 0$, in fact it is well known that $|\mathcal{F}\varphi(x)| \lesssim |x|^{-d-2\sigma}$. But eventually we are only interested in $\sigma - d/2 > 1/3$, so the lower bounds in Lemma 5.4.8 do not impose any additional restrictions. On the other side they simplify the proof.

Based on these estimates, we can establish an a priori estimate for the action of the fractional heat kernel on the nonlinear part.

Lemma 5.4.10. *Let $\alpha \in (1/3,1)$ and $\mathbb{w} = (w, w\mathrm{D}_x w) \in \mathscr{C}^\alpha$. Let $\varepsilon > 0$ and let $v \in C^{1+\varepsilon}$ and $G \in C_b^2$. Then for all $t \in [0,T]$ we have*

$$\|P_t[G(v+w)\mathrm{D}_x(v+w)]\|_{1+\varepsilon} \lesssim_{G,\mathbb{w}} t^{-\frac{2+\varepsilon-\alpha}{2\sigma}}(1+\|v\|_{2\alpha}) + t^{-\frac{1+\varepsilon}{2\sigma}}\|v\|_{1+\varepsilon}.$$

Proof. Since $v+w$ is controlled by w, and $\mathrm{D}_x(v+w)$ is controlled by $\mathrm{D}_x w$, it would be possible to directly apply Theorem 5.3.15 to define $G(v+w)\mathrm{D}_x(v+w)$. But this would only give us a quadratic estimate in v, which would lead to problems when trying to construct global solutions. Therefore, we consider the terms $G(v+w)\mathrm{D}_x v$ and $G(v+w)\mathrm{D}_x w$ separately.

The semigroup estimate Lemma 5.4.8 implies for the smooth term

$$\begin{aligned}\|P_t[G(v+w)\mathrm{D}_x v]\|_{1+\varepsilon} &\lesssim t^{-\frac{1+\varepsilon}{2\sigma}}\|G(v+w)\mathrm{D}_x v\|_0 \lesssim t^{-\frac{1+\varepsilon}{2\sigma}}\|G(v+w)\mathrm{D}_x v\|_{L^\infty} \\ &\lesssim t^{-\frac{1+\varepsilon}{2\sigma}}\|G\|_{L^\infty}\|\mathrm{D}_x v\|_\varepsilon \lesssim t^{-\frac{1+\varepsilon}{2\sigma}}\|G\|_{L^\infty}\|v\|_{1+\varepsilon}. \end{aligned} \tag{5.20}$$

Theorem 5.3.11 applied with $\beta = \alpha$ and $\gamma = \alpha - 1$ lets us estimate the rough term by

$$\|G(v+w)\mathrm{D}_x w\|_{\alpha-1} \le \|G(v+w)\mathrm{D}_x w\|_{\mathrm{D}_x w,\alpha-1,\alpha} \lesssim \|G(v+w)\|_{w,\alpha,\alpha}(1+\|\mathbb{w}\|_{\mathscr{C}^\alpha}). \tag{5.21}$$

Since $G \in C_b^2$, Lemma 5.3.18 yields

$$\|G(v+w)\|_{w,\alpha,\alpha} \lesssim_G (1+\|w\|_\alpha)(1+\|v\|_{2\alpha})(1+\|w\|_{w,\alpha,\alpha})^2. \tag{5.22}$$

We combine (5.21) and (5.22) with the semigroup estimate Lemma 5.4.8 (where we take $\delta = 2+\varepsilon-\alpha$), and obtain

$$\|P_t(G(v+w)\mathrm{D}_x w)\|_{1+\varepsilon} \lesssim_{G,\mathbb{w}} t^{-\frac{2+\varepsilon-\alpha}{2\sigma}}(1+\|v\|_{2\alpha}). \tag{5.23}$$

The proof is completed by combining (5.20) and (5.23) and noting that $\alpha < 1$ implies $2+\varepsilon-\alpha > 1+\varepsilon$, so that we can replace $t^{-\frac{1+\varepsilon}{2\sigma}}$ by $t^{-\frac{2+\varepsilon-\alpha}{2\sigma}}$. $\square$

Next we establish a contraction property for the semigroup acting on the nonlinear part.

Lemma 5.4.11. *Let $\alpha \in (1/3, 1)$ and $\mathbb{w} \in \mathscr{C}^\alpha$. Let $\varepsilon > 0$ and let $v_1, v_2 \in C^{1+\varepsilon}$ and $G \in C_b^3$. Then for all $t \in [0,T]$ we have*

$$\begin{aligned}&\|P_t[G(v_1+w)\mathrm{D}_x(v_1+w)] - P_t[G(v_2+w)\mathrm{D}_x(v_2+w)]\|_{1+\varepsilon}\\&\quad\lesssim_{G,\mathbb{w}} t^{-\frac{2+\varepsilon-\alpha}{2\sigma}}(1+\|v_1\|_{2\alpha}+\|v_2\|_{2\alpha})\|v_1-v_2\|_{2\alpha} + t^{-\frac{1+\varepsilon}{2\sigma}}(1+\|v_1\|_{1+\varepsilon})\|v_1-v_2\|_{1+\varepsilon}.\end{aligned}$$

Proof. We decompose

$$G(v_1+w)\mathrm{D}_x(v_1+w) - G(v_2+w)\mathrm{D}_x(v_2+w) = g_1+g_2+g_3,$$

where

$$g_1 := (G(v_1+w)-G(v_2+w))\mathrm{D}_x w, \qquad g_2 := (G(v_1+w)-G(v_2+w))\mathrm{D}_x v_1,$$
$$g_3 := G(v_2+w)\mathrm{D}_x(v_1-v_2).$$

Using Theorem 5.3.11 with $\beta = \alpha$ and $\gamma = \alpha - 1$, the term g_1 can be estimated by

$$\|g_1\|_{\alpha-1} = \|(G(v_1+w)-G(v_2+w))\mathrm{D}_x w\|_{\alpha-1} \lesssim_{\mathbb{w}} \|G(v_1+w)-G(v_2+w)\|_{w,\alpha,\alpha}.$$

We apply a Taylor expansion in the first step, Lemma 5.3.19 in the second step, and Lemma 5.3.18 in the third step, to obtain

$$\begin{aligned}&\|G(v_1+w)-G(v_2+w)\|_{w,\alpha,\alpha}\\&\qquad = \Big\| \sum_{\eta\in\mathbb{N}^n:|\eta|=1}\int_0^1 (\partial^\eta G)(v_2+w+r(v_1-v_2))(v_1-v_2)^\eta \mathrm{d}r\Big\|_{w,\alpha,\alpha}\\&\qquad \lesssim_w \sum_{|\eta|=1}\int_0^1 \|(\partial^\eta G)(v_2+w+r(v_1-v_2))\|_{w,\alpha,\alpha}\|v_1-v_2\|_{2\alpha}\mathrm{d}r\\&\qquad \lesssim_{G,w} (1+\|v_1\|_{2\alpha}+\|v_2\|_{2\alpha})\|v_1-v_2\|_{2\alpha}.\end{aligned}$$

Hence

$$\|g_1\|_{\alpha-1} \lesssim_{G,\mathbb{w}} (1+\|v_1\|_{2\alpha}+\|v_2\|_{2\alpha})\|v_1-v_2\|_{2\alpha}.$$

We apply a Taylor expansion to g_2 and obtain

$$\begin{aligned}\|g_2\|_{L^\infty}+\|g_3\|_{L^\infty} &\lesssim \|(G(v_1+w)-G(v_2+w))\mathrm{D}_x v_1\|_{L^\infty} + \|G(v_2+w)\mathrm{D}_x(v_1-v_2)\|_{L^\infty}\\&\lesssim_{G,w} \|v_1-v_2\|_{L^\infty}\|v_1\|_{1+\varepsilon} + \|v_1-v_2\|_{1+\varepsilon}.\end{aligned}$$

The statement now follows from the semigroup estimate Lemma 5.4.8, which we apply with $\delta = 2+\varepsilon-\alpha$ to estimate $P_t g_1$, and with $\delta = 1+\varepsilon$ to estimate $P_t g_2$ and $P_t g_3$. For the last two terms we also need that $\|\cdot\|_0 \lesssim \|\cdot\|_{L^\infty}$. □

We are now ready to prove the main result of this section.

Theorem 5.4.12. *Let $T > 0$, $d \in \mathbb{N}$, let $\sigma \geq 1$ be such that $\sigma - d/2 > 1/3$, and let*

$\alpha \in (1/3, (\sigma - d/2) \wedge 1)$. Let $u_0 \in C^\alpha$ and $G \in C_b^3$. Then there exists a unique mild solution $u \in C_T^\alpha$ to equation (5.19)*.*

Proof. Inspired by [Hai11], we subtract the contribution of the initial condition and solve for $v_t = u_t - \psi_t - P_t(u_0 - \psi_0)$.

We set up a Picard iteration in $C^{1+\varepsilon}$ for some small $\varepsilon > 0$, to be specified below. We define $v^0 := 0$, and

$$\begin{aligned} v_t^{n+1} := & \int_0^t P_{t-s}[G(v_s^n + \psi_s + P_s(u_0 - \psi_0))\mathrm{D}_x(v_s^n + \psi_s + P_s(u_0 - \psi_0))]\mathrm{d}s \\ & + \lambda \int_0^t P_{t-s}\psi_s \mathrm{d}s. \end{aligned}$$

In Lemma 5.4.2 and Lemma 5.4.3 we showed that $(\psi, \psi \mathrm{D}_x \psi) \in \mathscr{C}_T^\alpha$. For Lemma 5.4.3 we needed that $1 + d/2 - 2\sigma < 0$. But since $\sigma - d/2 > 1/3$, this is satisfied.

Let $\beta \in (1/3, \alpha)$ and $t \in [0, T]$. We apply the a priori estimate Lemma 5.4.10 with β in the place of α, to obtain

$$\begin{aligned} \|v_t^{n+1}\|_{1+\varepsilon} \le & \int_0^t \|P_{t-s}[G(v_s^n + \psi_s + P_s(u_0 - \psi_0))\mathrm{D}_x(v_s^n + \psi_s + P_s(u_0 - \psi_0))]\|_{1+\varepsilon}\, \mathrm{d}s \\ & + \lambda \int_0^t \|P_{t-s}\psi_s\|_{1+\varepsilon} \mathrm{d}s \\ \lesssim_{G,\psi} & \int_0^t (t-s)^{-\frac{2+\varepsilon-\beta}{2\sigma}} (1 + \|v_s^n + P_s(u_0 - \psi_0)\|_{2\beta}) \mathrm{d}s \\ & + \int_0^t \Big[(t-s)^{-\frac{1+\varepsilon}{2\sigma}} \|v_s^n + P_s(u_0 - \psi_0)\|_{1+\varepsilon} + \lambda \|P_{t-s}\psi_s\|_{1+\varepsilon} \Big] \mathrm{d}s. \end{aligned}$$

Now the semigroup estimate Lemma 5.4.8, applied with $\delta = 2\beta - \alpha$ and $\delta = 1 + \varepsilon - \alpha$ respectively, yields

$$\begin{aligned} \|v_t^{n+1}\|_{1+\varepsilon} \lesssim_{G,\psi} & \int_0^t (t-s)^{-\frac{2+\varepsilon-\beta}{2\sigma}} (1 + \|v_s^n\|_{1+\varepsilon} + s^{-\frac{2\beta-\alpha}{2\sigma}} \|u_0 - \psi_0\|_\alpha) \mathrm{d}s \\ & + \int_0^t \Big[(t-s)^{-\frac{1+\varepsilon}{2\sigma}} (\|v_s^n\|_{1+\varepsilon} + s^{-\frac{1+\varepsilon-\alpha}{2\sigma}} \|u_0 - \psi_0\|_\alpha) + \lambda (t-s)^{-\frac{1+\varepsilon-\alpha}{2\sigma}} \|\psi_s\|_\alpha \Big] \mathrm{d}s. \end{aligned}$$

For $a, b \ge 0$ the integral $\int_0^t (t-s)^{-a} s^{-b} \mathrm{d}s$ converges to zero as t tends to 0 if and only if $a + b < 1$. So if we choose $\varepsilon > 0$ small enough, then there exists $T_1 \in (0, T]$, independent of u_0, such that

$$\|v^n\|_{C_{T_1}^{1+\varepsilon}} \le \frac{\|u_0\|_\alpha + 1}{2} \tag{5.24}$$

for all $n \in \mathbb{N}$.

Let us write $U_s := P_s(u_0 - \psi_0)$. The contraction estimate Lemma 5.4.11, applied with

β in the place of α, implies for all $t \in [0,T]$ that

$$\begin{aligned}\|v_t^{n+1} - v_t^n\|_{1+\varepsilon} \le \int_0^t \Big\| & P_{t-s}[G(v_s^n + \psi_s + U_s)\mathrm{D}_x(v_s^n + \psi_s + U_s)] \\ & - P_{t-s}[G(v_s^{n-1} + \psi_s + U_s)\mathrm{D}_x(v_s^{n-1} + \psi_s + U_s)]\Big\|_{1+\varepsilon} \mathrm{d}s \\ \lesssim_{G,\psi} \int_0^t & (t-s)^{-\frac{2+\varepsilon-\beta}{2\sigma}}(1 + \|v_s^n + U_s\|_{2\beta} + \|v_s^{n-1} + U_s\|_{2\beta})\|v_s^n - v_s^{n-1}\|_{2\beta}\mathrm{d}s \\ + \int_0^t & (t-s)^{-\frac{1+\varepsilon}{2\sigma}}(1 + \|v_s^n + U_s\|_{1+\varepsilon})\|v_s^n - v_s^{n-1}\|_{1+\varepsilon}\end{aligned}$$

Now the same arguments as before, in combination with (5.24), prove the contraction property on $[0,T_2]$ for a suitable $T_2 > 0$. Hence, there exists a unique solution on $[0,T_2]$.

It remains to show the existence of a global solution. Let $T_{\max} \in (0,T]$ be the maximum time for which there exists a solution v on $[0,T_{\max})$. First assume that $\lim_{t\to T_{\max}} \|v\|_{C_t^{1+\varepsilon}} < \infty$. Then we can iterate the construction of a local solution on small intervals of length T_3, where $T_3 > 0$ is fixed, because the initial condition in each iteration will be bounded by $\sup_{t<T_{\max}} \|v\|_{C_t^{1+\varepsilon}} < \infty$. Hence, $T_{\max} = T$. On the other side $\lim_{t\to T_{\max}} \|v\|_{C_t^{1+\varepsilon}} = \infty$ is impossible, because (5.24) yields

$$\|v\|_{C_{T_{\max}}^{1+\varepsilon}} \le F^{\circ\lceil T_{\max}/T_1\rceil}(\|u_0\|_\alpha) < \infty,$$

where $F(x) := (x+1)/2$, and $F^{\circ m}$ is the m–fold iterative application of F. □

Remark 5.4.13. The continuity of all operators involved in the Picard iteration enables us to show that solutions to (5.19) depend continuously on the reference path ψ: If $(\psi^j)_{j\in\mathbb{N}}$ converges to ψ, such that $(\pi_\circ(\psi^j, \mathrm{D}_x\psi^j))$ converges to $\pi_\circ(\psi, \mathrm{D}_x\psi)$, then the solutions v^j to

$$\begin{aligned}v^j(t,\cdot) = P_t(u_0 - \psi_0^j)(x) &+ \int_0^t P_{t-s}[G(v^j(s,\cdot) + \psi^j(s,\cdot))\mathrm{D}_x(v^j(s,\cdot) + \psi^j(s,\cdot))]\mathrm{d}s \\ &+ \lambda \int_0^t P_{t-s}\psi^j(s,\cdot)\mathrm{d}s\end{aligned}$$

converge to v. This can be seen by similar arguments as the ones used in the proof of Proposition 8 in [Gub04].

As explained in Remark 5.4.4, we expect that if we approximate $\dot{W}$ by suitable sequences $(\dot{W}^j)$ of smooth functions, then $(\pi_\circ(\psi^j, \mathrm{D}_x\psi^j))$ converges to $\pi_\circ(\psi, \mathrm{D}_x\psi)$. This indicates that our solution u is the limit of the classical solutions u^j to

$$\partial_t u^j(t,x) = -Au^j(t,x) + G(u^j(t,x))\mathrm{D}_x u^j(t,x) + \dot{W}^j(t,x).$$

5.5. Non-linear parabolic Anderson model

In this section we study the following nonlinear version of the heat equation

$$\partial_t u(t,x) = \Delta u(t,x) + F(u(t,x))\dot{W}(x), \tag{5.25}$$

where $\dot{W}$ is a spatial white noise on $\mathbb{T}^2$ without time dependence, $F : \mathbb{R} \to \mathbb{R}$ is a smooth function, and $u : [0,T] \times \mathbb{T}^2 \to \mathbb{R}$. This is a nonlinear generalization of the parabolic Anderson model, for which $F(u) = u$. See for example [CM94] for a comprehensive treatment of the spatially discrete parabolic Anderson model. There is not much work on the spatially continuous case in dimension 2 or higher, at least not for a white noise potential. Most papers that deal with this problem use Wick products to make sense of the product $u\dot{W}$, see for example [LR09]. To the best of our knowledge, the only other work on the nonlinear version (5.25) is the recent paper by Hairer [Hai13a] on regularity structures, in which, among many other problems, a slightly more general version of the equation is solved.

The regularity of the spatial white noise is $\dot{W} \in C^{-d/2-}$, meaning that $\dot{W} \in C^{-d/2-\varepsilon}$ for every $\varepsilon > 0$. So for $d = 2$ we have $\dot{W} \in C^{-1-}$ and the regularization provided by the Laplacian would be sufficient to have $u \in C^{1-}$. Heuristically we gain (almost) 2 degrees of spatial regularity. Then the product $F(u)\dot{W}$ is not defined, because the regularities of $F(u)$ and $\dot{W}$ sum up to a negative value. The mild solution reads

$$u(t,x) = P_t u_0(x) + \int_0^t P_{t-s}[F(u_s)\dot{W}](x)\mathrm{d}s,$$

where $(P_t)_{t\geq 0}$ is the heat kernel and where $u_s := u(s,\cdot)$. Now

$$F(u_s)\dot{W} = \pi_<(F(u_s),\dot{W}) + \pi_\circ(F(u_s),\dot{W}) + \pi_>(F(u_s),\dot{W})$$

and the critical diagonal part can be further decomposed as

$$\begin{aligned}\pi_\circ(F(u_s),\dot{W}) &= \pi_\circ(\pi_<(F'(u_s),u_s),\dot{W}) + \pi_\circ(r_s,\dot{W})\\ &= F'(u_s)\pi_\circ(u_s,\dot{W}) + R(F'(u_s),u_s,\dot{W}) + \pi_\circ(r_s,\dot{W}),\end{aligned}$$

where the paralinearization theorem implies that $r_s = F(u_s) - \pi_<(F'(u_s),u_s)$ is smooth. Again the difficult term is $\pi_\circ(\pi_<(F'(u_s),u_s),\dot{W})$, and R is as in Lemma 5.3.3. Since the commutator between semigroup and paraproduct is smooth (see Lemma 5.5.7 below), we have

$$\begin{aligned}u_t = P_t u_0 + \int_0^t P_{t-s}(F(u_s)\dot{W})\mathrm{d}s &\simeq P_t u_0 + \int_0^t P_{t-s}(\pi_<(F(u_s),\dot{W}))\mathrm{d}s\\ &\simeq P_t u_0 + \int_0^t \pi_<(F(u_s),P_{t-s}\dot{W})\mathrm{d}s,\end{aligned}$$

where we write $f \simeq g$ if $f - g$ is smooth. Therefore, u is called *controlled by* $\dot{W}$ if there exists u'_s such that

$$u_t \simeq P_t u_0 + \int_0^t \pi_<(u'_s, P_{t-s}\dot{W})\mathrm{d}s.$$

In this case (and if for the moment we ignore the initial condition), we get

$$\pi_\circ(u_t, \dot{W}) \simeq \pi_\circ\Big(\int_0^t \pi_<(u'_s, P_{t-s}\dot{W})\mathrm{d}s, \dot{W}\Big) \simeq \int_0^t u'_s \pi_\circ(P_{t-s}\dot{W}, \dot{W})\mathrm{d}s.$$

So here $\pi_\circ(P_t\dot{W}, \dot{W})$ takes the role of the area. Of course for every $t > 0$ this is a well defined smooth function, but we need an estimate as $t \to 0$.

5.5.1. Regularity of the Besov area and renormalized products

Motivated by this heuristic discussion, let us now study the regularity of the Besov area $\pi_\circ(P_t\dot{W}, \dot{W})$ and construct the product $F(u_t)\dot{W}$. It will turn out that we have to renormalize the product by “subtracting an infinite constant” in order to obtain a well-defined object.

Recall that $\dot{W}$ is a spatial white noise if it is a mean zero Gaussian process with values in $\mathcal{S}'(\mathbb{T}^2)$, such that

$$E(\langle \dot{W}, \varphi\rangle\langle \dot{W}, \vartheta\rangle) = \int_{\mathbb{T}^2} \varphi(x)\vartheta(x)\mathrm{d}x$$

for all test function $\varphi, \vartheta \in L^2(\mathbb{T}^2)$. Formally we write $E(\dot{W}(x)\dot{W}(x')) = \delta(x - x')$, $x, x' \in \mathbb{R}^2$.

Note that for $t > 0$ the function $P_t\dot{W}$ is smooth, and therefore the Besov area $P_t\dot{W}\dot{W}$ is a well-defined smooth function. We have to study how its regularity depends on t.

Lemma 5.5.1. *For any $x \in \mathbb{T}^2$ and $t > 0$ we have*

$$g_t := E(\pi_\circ(P_t\dot{W}, \dot{W})(x)) = E(\Delta_{-1}(\pi_\circ(P_t\dot{W}, \dot{W}))(x)) = \sum_{k\in\mathbb{Z}^2} e^{-t|k|^2} \simeq \frac{1}{t}.$$

In particular, g_t does not depend on the dyadic partition of unity under consideration.

Proof. Let $x \in \mathbb{T}^2$, $t > 0$, and $m \geq -1$. Then

$$E(\Delta_m(\pi_\circ(P_t\dot{W}, \dot{W}))(x)) = \sum_{|i-j|\leq 1} E(\Delta_m(\Delta_i(P_t\dot{W})\Delta_j\dot{W})(x)),$$

where exchanging summation and expectation is justified because it can be easily verified that the partial sums of $\Delta_m(\pi_\circ(P_t\dot{W}, \dot{W}))(x)$ are uniformly L^p–bounded for any $p \geq 1$. Similarly as in the proof of Lemma 5.4.1 we can show that $(\widehat{\dot{W}}(k))_{k\in\mathbb{Z}^2}$ is a complex valued centered Gaussian process with covariance $E(\widehat{\dot{W}}(k)\widehat{\dot{W}}(k')) = (2\pi)^2\delta_{k(-k')}$. Recall that $P_t = e^{-t|\cdot|^2}(\mathrm{D})$. Therefore

$$\begin{aligned}
&E(\Delta_m(\Delta_i(P_t\dot{W})\Delta_j\dot{W})(x)) \\
&\qquad = (2\pi)^{-2}\sum_{k,k'\in\mathbb{Z}^2} e^{\imath\langle k+k',x\rangle}\rho_m(k+k')\rho_i(k)e^{-t|k|^2}\rho_j(k')E(\widehat{\dot{W}}(k)\widehat{\dot{W}}(k')) \\
&\qquad = \sum_{k\in\mathbb{Z}^2}\rho_m(0)\rho_i(k)e^{-t|k|^2}\rho_j(k) = \delta_{(-1)m}\sum_{k\in\mathbb{Z}^2}\rho_i(k)\rho_j(k)e^{-t|k|^2}.
\end{aligned}$$

For $|i-j|>1$ we have $\rho_i(k)\rho_j(k)=0$. This implies, independently of $x\in\mathbb{T}^2$, that

$$g_t = E(\pi_\circ(P_t\dot{W},\dot{W})(x)) = \sum_{k\in\mathbb{Z}^2}\sum_{i,j}\rho_i(k)\rho_j(k)e^{-t|k|^2} = \sum_{k\in\mathbb{Z}^2}e^{-t|k|^2} \simeq \int_{\mathbb{R}^2}e^{-t|x|^2}\mathrm{d}x \simeq \frac{1}{t},$$

while $E(\pi_\circ(P_t\dot{W},\dot{W})(x) - \Delta_{-1}(\pi_\circ(P_t\dot{W},\dot{W}))(x)) = 0$ for every $x\in\mathbb{T}^2$. □

The factor $1/t$ leads to a diverging time integral. This motivates us to study the renormalized product $\pi_\circ(P_t\dot{W},\dot{W}) - g_t$.

Lemma 5.5.2. *For $t>0$ we define*

$$\Xi_t(x) := \pi_\circ(P_t\dot{W},\dot{W})(x) - g_t.$$

Then for all $\varepsilon\in(0,1)$ and for all $t>0$ we have $E(\|\Xi_t\|_{-\varepsilon}) \lesssim t^{-1+\frac{\varepsilon}{4}}$. In particular

$$E\Big(\int_0^T \|\Xi_t\|_{-\varepsilon}\mathrm{d}t\Big) < \infty$$

for all $T,\varepsilon>0$.

Proof. We use similar arguments as in the proofs of Lemma 5.4.2 and Lemma 5.4.3. Let $t>0$. By Gaussian hypercontractivity we obtain for $p\geq 1$ and $m\geq -1$ that

$$E\|\Delta_m\Xi_t\|^{2p}_{L^{2p}(\mathbb{T}^2)} \lesssim_p \Big\|E(|\Delta_m\Xi_t(x)|^2)\Big\|^p_{L^p_x(\mathbb{T}^2)}. \tag{5.26}$$

Lemma 5.5.1 implies that $\Delta_m g_t = 0 = E(\Delta_m(\pi_\circ(P_t\dot{W},\dot{W}))(x))$ for $m\geq 0$, and therefore

$$E(|\Delta_m\Xi_t(x)|^2) = E(|\Delta_m(\pi_\circ(P_t\dot{W},\dot{W}))(x)|^2) = \mathrm{Var}(\Delta_m(\pi_\circ(P_t\dot{W},\dot{W}))(x)), \tag{5.27}$$

where $\mathrm{Var}(X)$ denotes the variance of X. Lemma 5.5.1 also gives $\Delta_{-1}g_t = g_t = E(\Delta_{-1}(\pi_\circ(P_t\dot{W},\dot{W}))(x))$, leading to

$$E(|\Delta_{-1}\Xi_t(x)|^2) = E(|\Delta_{-1}(\pi_\circ(P_t\dot{W},\dot{W}))(x) - g_t|^2) = \mathrm{Var}(\Delta_{-1}(\pi_\circ(P_t\dot{W},\dot{W}))(x)). \tag{5.28}$$

Therefore, it suffices to control $\mathrm{Var}(\Delta_m(\pi_\circ(P_t\dot W,\dot W))(x))$. We have

$$\mathcal{F}(\pi_\circ(P_t\dot W,\dot W))(k)=(2\pi)^{-2}\sum_{k'\in\mathbb{Z}^2}\sum_{|i-j|\le 1}\rho_i(k-k')e^{-t|k-k'|^2}\widehat{\dot W}(k-k')\rho_j(k')\widehat{\dot W}(k'),$$

implying for all $m\ge -1$ and $x\in\mathbb{T}^2$ that

$$\begin{aligned}\Delta_m(\pi_\circ(P_t\dot W,\dot W))(x)&=(2\pi)^{-2}\sum_{k\in\mathbb{Z}^2}e^{\imath\langle k,x\rangle}\rho_m(k)\mathcal{F}(\pi_\circ(P_t\dot W,\dot W))(k)\\&=(2\pi)^{-4}\sum_{k_1,k_2\in\mathbb{Z}^2}\sum_{|i-j|\le 1}e^{\imath\langle k_1+k_2,x\rangle}\rho_m(k_1+k_2)\rho_i(k_1)e^{-t|k_1|^2}\widehat{\dot W}(k_1)\rho_j(k_2)\widehat{\dot W}(k_2).\end{aligned}$$

Hence

$$\begin{aligned}&\mathrm{Var}(\Delta_m(\pi_\circ(P_t\dot W,\dot W))(x))\\&\quad=(2\pi)^{-8}\sum_{k_1,k_2\in\mathbb{Z}^2}\sum_{k_1',k_2'\in\mathbb{Z}^2}\sum_{|i-j|\le 1}\sum_{|i'-j'|\le 1}e^{\imath\langle k_1+k_2,x\rangle}\rho_m(k_1+k_2)\rho_i(k_1)e^{-t|k_1|^2}\rho_j(k_2)\\&\qquad\times e^{\imath\langle k_1'+k_2',x\rangle}\rho_m(k_1'+k_2')\rho_i(k_1')e^{-t|k_1'|^2}\rho_j(k_2')\,\mathrm{cov}(\widehat{\dot W}(k_1)\widehat{\dot W}(k_2),\widehat{\dot W}(k_1')\widehat{\dot W}(k_2')),\end{aligned}$$

where exchanging summation and expectation can again be justified a posteriori by the uniform L^p-boundedness of the partial sums, and where $\mathrm{cov}(X,Y)$ denotes the covariance of X and Y.

Now $(\widehat{\dot W}(k))_{k\in\mathbb{Z}^2}$ is a centered Gaussian process, and therefore Wick's theorem ([Jan97], Theorem 1.28) yields

$$\begin{aligned}&\mathrm{cov}(\widehat{\dot W}(k_1)\widehat{\dot W}(k_2),\widehat{\dot W}(k_1')\widehat{\dot W}(k_2'))=E(\widehat{\dot W}(k_1)\widehat{\dot W}(k_2)\widehat{\dot W}(k_1')\widehat{\dot W}(k_2'))\\&\qquad\qquad-E(\widehat{\dot W}(k_1)\widehat{\dot W}(k_2))E(\widehat{\dot W}(k_1')\widehat{\dot W}(k_2'))\\&\quad=\mathrm{cov}(\widehat{\dot W}(k_1),\widehat{\dot W}(k_2))\,\mathrm{cov}(\widehat{\dot W}(k_1'),\widehat{\dot W}(k_2'))+\mathrm{cov}(\widehat{\dot W}(k_1),\widehat{\dot W}(k_1'))\,\mathrm{cov}(\widehat{\dot W}(k_2),\widehat{\dot W}(k_2'))\\&\qquad+\mathrm{cov}(\widehat{\dot W}(k_1),\widehat{\dot W}(k_2'))\,\mathrm{cov}(\widehat{\dot W}(k_2),\widehat{\dot W}(k_1'))-\mathrm{cov}(\widehat{\dot W}(k_1),\widehat{\dot W}(k_2))\,\mathrm{cov}(\widehat{\dot W}(k_1'),\widehat{\dot W}(k_2'))\\&\quad=(2\pi)^4(\delta_{k_1(-k_1')}\delta_{k_2(-k_2')}+\delta_{k_1(-k_2')}\delta_{k_2(-k_1')}).\end{aligned}$$

This implies that

$$\begin{aligned}&\mathrm{Var}(\Delta_m(\pi_\circ(P_t\dot W,\dot W))(x))\\&\quad=\sum_{k_1,k_2\in\mathbb{Z}^2}\sum_{|i-j|\le 1}\sum_{|i'-j'|\le 1}1_{m\lesssim i}1_{m\lesssim i'}\rho_m^2(k_1+k_2)\rho_i(k_1)\rho_j(k_2)\rho_{i'}(k_1)\rho_{j'}(k_2)e^{-2t|k_1|^2}\\&\qquad+\sum_{k_1,k_2\in\mathbb{Z}^2}\sum_{|i-j|\le 1}\sum_{|i'-j'|\le 1}1_{m\lesssim i}1_{m\lesssim i'}\rho_m^2(k_1+k_2)\rho_i(k_1)\rho_j(k_2)\rho_{i'}(k_2)\rho_{j'}(k_1)e^{-t|k_1|^2-t|k_2|^2}.\end{aligned}$$

There exists $c>0$ such that $e^{-2t|k|^2}\lesssim e^{-tc2^{2i}}$ for all $k\in\mathrm{supp}(\rho_i)$ and for all $i\ge -1$. In the remainder of the proof the value of this strictly positive c may change from line to line. If $|i-j| \leq 1$, then we also have $e^{-t|k|^2} \lesssim e^{-tc2^{2i}}$ for all $k \in \operatorname{supp}(\rho_j)$. Thus

$$\begin{aligned}
&\operatorname{Var}(\Delta_m(\pi_\circ(P_t\dot{W},\dot{W}))(x)) \\
&\quad\lesssim \sum_{i,j,i',j'} 1_{m\lesssim i} 1_{i\sim j\sim i'\sim j'} \sum_{k_1,k_2\in\mathbb{Z}^2} 1_{\operatorname{supp}(\rho_m)}(k_1+k_2) 1_{\operatorname{supp}(\rho_i)}(k_1) 1_{\operatorname{supp}(\rho_j)}(k_2) e^{-2tc2^{2i}} \\
&\quad\lesssim \sum_{i:i\gtrsim m} 2^{2i} 2^{2m} e^{-tc2^{2i}} \lesssim \frac{2^{2m}}{t} \sum_{i:i\gtrsim m} e^{-tc2^{2i}} \lesssim \frac{2^{2m}}{t} e^{-tc2^{2m}}, \qquad (5.29)
\end{aligned}$$

where we used that $t2^{2i} \lesssim e^{t(c-c')2^{2i}}$ for any $c' < c$.

Now let $\varepsilon \in (0,1)$. We apply Jensen's inequality and combine (5.26), (5.27), (5.28), and (5.29) to obtain

$$\begin{aligned}
E\Big(\|\Xi_t\|_{B^{-\varepsilon}_{2p,2p}}\Big) &\lesssim \Big(\sum_{m\geq -1} 2^{-\varepsilon m 2p} E\Big(\|\Delta_m \Xi_t\|^{2p}_{L^{2p}(\mathbb{T}^2)}\Big)\Big)^{\frac{1}{2p}} \\
&\lesssim t^{-1/2}\Big(\sum_{m\geq -1} 2^{-\varepsilon m 2p} 2^{2mp} e^{-tcp2^{2m}}\Big)^{\frac{1}{2p}} \simeq t^{-1/2}\Big(\int_{-1}^{\infty} (2^x)^{2p(1-\varepsilon)} e^{-ctp(2^x)^2} \mathrm{d}x\Big)^{\frac{1}{2p}}.
\end{aligned}$$

The change of variables $y = \sqrt{t}2^x$ then yields

$$E\Big(\|\Xi_t\|_{B^{-\varepsilon}_{2p,2p}}\Big) \lesssim t^{-1/2}\Big(t^{-p(1-\varepsilon)} \int_0^\infty y^{2p(1-\varepsilon)-1} e^{-cpy^2} \mathrm{d}y\Big)^{\frac{1}{2p}}.$$

Since $\varepsilon < 1$, the integral is finite, and therefore finally

$$E\Big(\|\Xi_t\|_{B^{-\varepsilon}_{2p,2p}}\Big) \lesssim_p t^{-1+\frac{\varepsilon}{2}}.$$

Since p can be chosen arbitrarily large, the result now follows from the Besov embedding theorem, Lemma 5.2.4. □

Remark 5.5.3. The renormalized product Ξ_t has the natural regularity of $\pi_\circ(P_t\dot{W},\dot{W})$: Since $P_t\dot{W} \in C^{1-}$ and $\dot{W} \in C^{-1-}$, we would expect $\pi_\circ(P_t\dot{W},\dot{W})$ to be in C^{0-}.

Lemma 5.5.2 suggests that we should really consider $\pi_\circ(u_t,\dot{W}) - u'_t G_t$ in the definition of the product, where $G_t := \int_0^t g_{t-s}\mathrm{d}s$ is formally defined as an infinite constant. Motivated by this discussion we now give a meaning to the product in an appropriate space of controlled distributions.

For $\alpha > 0$ and $\beta \in \mathbb{R}$ we define $C^{\alpha,\beta}_T := C^\alpha([0,T], C^\beta(\mathbb{T}^2,\mathbb{R}))$ as the space of α–Hölder continuous functions taking their values in C^β, equipped with the norm

$$\|u\|_{C^{\alpha,\beta}_T} = \sup_{t\in[0,T]} \|u_t\|_\beta + \sup_{s\neq t} \frac{\|u_t - u_s\|_\beta}{|t-s|^\alpha}.$$

Definition 5.5.4. Let $\alpha, \beta > 0$. We say that $u \in C_T^{\alpha,1/2+\beta}$ is *controlled by* $\dot{W}$ if

$$u(t,\cdot) = \int_0^t \pi_<(u'(s,\cdot), P_{t-s}\dot{W})\mathrm{d}s + u^\sharp(t,\cdot)$$

for all $t \in [0,T]$, where $u' \in C_T^{\alpha,1/2+\beta}$ and $u^\sharp \in C([0,T], C^{1+\beta})$. In this case we write $u \in \mathcal{D}_{\dot{W}}^{\alpha,\beta,T}$ and define the norm

$$\|u\|_{\dot{W},\alpha,\beta,T} := \|u\|_{C_T^{\alpha,1/2+\beta}} + \|u'\|_{C_T^{\alpha,1/2+\beta}} + \|u^\sharp\|_{C_T^{1+\beta}}.$$

In Section 5.3.3 we formulated the paralinearization theorem for $\alpha \in (0,1/2)$. Here, we need the case $\alpha \in (1/2,1)$, for which we can give a very simple proof.

Lemma 5.5.5. *Let $\alpha \in (1/2,1)$. If $u \in C^\alpha$ and $F \in C_b^2$, then*

$$\|F(u) - \pi_<(F'(u),u)\|_{2\alpha} \lesssim \|F\|_{C_b^2}(1+\|u\|_\alpha)^2.$$

Proof. Since $2\alpha \in (1,2)$, we have $\|u\|_{2\alpha} \simeq \|u\|_{L^\infty} + \|\mathrm{D}u\|_{2\alpha-1}$, see Example 2 on p. 99 of [BCD11]. The estimate for the L^∞–norm is straightforward, so let us estimate the derivative. The chain rule together with the paraproduct decomposition implies that

$$\mathrm{D}(F(u)) = \pi_<(F'(u),\mathrm{D}u) + \pi_>(F'(u),\mathrm{D}u) + \pi_\circ(F'(u),\mathrm{D}u).$$

The first term on the right hand side is the critical one, and it follows from the Leibniz rule that it is cancelled by one of the terms of $\mathrm{D}\pi_<(F'(u),u)$. So we end up with

$$\begin{aligned}\|\mathrm{D}(F(u) - \pi_<(F'(u),u))\|_{2\alpha-1} &\le \|\pi_>(F'(u),\mathrm{D}u)\|_{2\alpha-1} + \|\pi_\circ(F'(u),\mathrm{D}u)\|_{2\alpha-1} \\ &\quad + \|\pi_<(\mathrm{D}(F'(u)),u)\|_{2\alpha-1} \\ &\lesssim \|F'(u)\|_\alpha \|u\|_\alpha \lesssim \|F\|_{C_b^2}(1+\|u\|_\alpha)\|u\|_\alpha,\end{aligned}$$

where we used Lemma 5.3.17 in the last step. □

The paralinearization theorem will be needed in the following lemma, where we implicitly use that $F(u)$ is controlled by $\dot{W}$ if u is controlled by $\dot{W}$.

For $\varepsilon > 0$ we define $G_t^\varepsilon := \int_0^t g_{t-s+\varepsilon}\mathrm{d}s$.

Lemma 5.5.6. *Let $\alpha > 0$, $\beta \in (0,1/2)$, and assume that $u \in \mathcal{D}_{\dot{W}}^{\alpha,\beta,T}$ and $F \in C_b^2$. Then the renormalized product*

$$F(u_t) \cdot P_\varepsilon \dot{W} = F(u_t)P_\varepsilon \dot{W} - F'(u_t)u_t' G_t^\varepsilon$$

converges as $\varepsilon \to 0$ to a well defined distribution which we denote by $F(u_t)\cdot\dot{W}$. For any *dyadic partition of unity we have the explicit representation*

$$\begin{aligned}
F(u_t)\cdot\dot W &= \pi_<(F(u_t),\dot W)+\pi_>(F(u_t),\dot W)+\pi_\circ(F(u_t)-\pi_<(F'(u_t),u_t),\dot W)\\
&\quad + R(F'(u_t),u_t,\dot W)+F'(u_t)\int_0^t R(u'_s,P_{t-s}\dot W,\dot W)\mathrm{d}s+F'(u_t)\int_0^t u'_s\Xi_{t-s}\mathrm{d}s\\
&\quad + F'(u_t)\int_0^t (u'_s-u'_t)g_{t-s}\mathrm{d}s+F'(u_t)\pi_\circ(u^\sharp_t,\dot W),
\end{aligned}$$

where each of these terms is well defined. Moreover, we have for $\gamma\in(0,\beta)$ *that*

$$\begin{aligned}
&\|F(u_t)\cdot\dot W\|_{C_T^{-1-\gamma}}+\|F(u_t)\cdot\dot W-\pi_<(F(u_t),\dot W)\|_{C_T^{\beta-1/2-\gamma}} \qquad (5.30)\\
&\quad\lesssim (1+T)^{\alpha\vee\frac{\gamma}{2}}\|F\|_{C^2_b}(1+\|u\|_{\dot W,\alpha,\beta,T})^2(1+\|\dot W\|_{-1-\gamma})^2\Big(1+\int_0^T\|\Xi_s\|_{-\gamma}\mathrm{d}s\Big).
\end{aligned}$$

Proof. We will perform a formal expansion of the quantity $F(u_t)\dot W$, which can be rendered rigorous by considering the regularized product. The subtraction then eliminates the only term which diverges in the limit. According to the discussion above, the only non-trivial term in $F(u_t)\cdot\dot W$ is $F'(u_t)\pi_\circ(u_t,\dot W)$. Since $F'(u_t)\in C^{1/2+\beta}$, this term is well defined as long as $\pi_\circ(u_t,\dot W)$ is well defined and in $C^{-\varepsilon}$ for some $\varepsilon<1/2+\beta$. But we have

$$\begin{aligned}
\pi_\circ(u_t,\dot W) &= \int_0^t \pi_\circ(\pi_<(u'_s,P_{t-s}\dot W),\dot W)\mathrm{d}s+\pi_\circ(u^\sharp_t,\dot W)\\
&= \int_0^t R(u'_s,P_{t-s}\dot W,\dot W)\mathrm{d}s+\int_0^t u'_s\pi_\circ(P_{t-s}\dot W,\dot W)\mathrm{d}s+\pi_\circ(u^\sharp_t,\dot W)\\
&= \int_0^t R(u'_s,P_{t-s}\dot W,\dot W)\mathrm{d}s+\int_0^t u'_s\Xi_{t-s}\mathrm{d}s+\int_0^t u'_s g_{t-s}\mathrm{d}s+\pi_\circ(u^\sharp_t,\dot W)\\
&= \int_0^t R(u'_s,P_{t-s}\dot W,\dot W)\mathrm{d}s+\int_0^t u'_s\Xi_{t-s}\mathrm{d}s+\int_0^t (u'_s-u'_t)g_{t-s}\mathrm{d}s+u'_tG_t+\pi_\circ(u^\sharp_t,\dot W).
\end{aligned}$$

The contribution $F'(u_t)u'_tG_t$ is exactly canceled by the correction term in the definition of the product. Putting all together we obtain the claimed result. The term

$$\int_0^t (u'_t-u'_s)g_{t-s}\mathrm{d}s$$

is well defined because Lemma 5.5.1 implies $g_{t-s}\simeq(t-s)^{-1}$ and we have $u'\in C_T^{\alpha,1/2+\beta}$, and therefore $\|u'_s-u'_t\|_{1-\beta}\lesssim|t-s|^\alpha\|u\|_{C_T^{\alpha,1/2+\beta}}$. The estimate (5.30) is shown by a somewhat lengthy but elementary calculation based on the paraproduct estimates and Lemma 5.3.3. □

5.5.2. Picard iteration

Now that we constructed the area and made sense of the product on a space of suitable controlled distributions, we can start studying the SPDE under consideration. Before we get to the existence and uniqueness of solutions, let us first establish some a priori estimates. Our first result is a commutator estimate between heat flow and paraproduct

Lemma 5.5.7. *Let $\alpha < 1$, $\beta \in \mathbb{R}$, and $t > 0$. Then for all $\delta \geq 0$ we have*

$$\|P_t(\pi_<(u,v)) - \pi_<(u, P_t v)\|_{\alpha+\beta+\delta} \lesssim t^{-\frac{\delta}{2}} \|u\|_\alpha \|v\|_\beta.$$

Proof. The semigroup is given by $P_t = \varphi(\sqrt{t})$, where $\varphi(z) = e^{-|z|^2}$ is a Schwartz function. The estimate is therefore a special case of Lemma 5.3.20. □

The next result will allow us to examine the Hölder continuity of $t \mapsto \int_0^t P_{t-s} u \mathrm{d}s$.

Lemma 5.5.8. *Let $\alpha \in [0,1]$ and $\beta \in \mathbb{R}$. Then we have for any $u \in C^\beta$ that*

$$\|(P_t - \mathrm{id})u\|_{\beta-2\alpha} \lesssim t^\alpha \|u\|_\beta.$$

Proof. We only sketch the proof. Following the proof of Lemma 2.4 in [BCD11] it is possible to show that

$$\|\Delta_j (P_t - \mathrm{id})u\|_{L^\infty} \lesssim t 2^{2j} \|\Delta_j u\|_{L^\infty}$$

for any $j \geq -1$, so that in particular

$$\|(P_t - \mathrm{id})u\|_{\beta-2} \lesssim t\|u\|_\beta$$

for any $\beta \in \mathbb{R}$. On the other side $P_t - \mathrm{id}$ is clearly bounded on C^β, and therefore the interpolation inequalities, Theorem 2.80 of [BCD11], imply that

$$\begin{aligned}\|(P_t - \mathrm{id})u\|_{\beta-2\alpha} &= \|(P_t - \mathrm{id})u\|_{\alpha(\beta-2)+(1-\alpha)\beta} \\ &\leq \|(P_t - \mathrm{id})u\|_{\beta-2}^\alpha \|(P_t - \mathrm{id})u\|_\beta^{1-\alpha} \lesssim t^\alpha \|u\|_\beta.\end{aligned}$$

□

Corollary 5.5.9. *Let $\delta \in (0,2)$ and $\gamma \in \mathbb{R}$. Then for all $T > 0$, for every integrable $u : [0,T] \to C^\gamma$, and for all $\alpha \in (0, 1-\delta/2)$ we have*

$$\Big\|\int_0^\cdot P_{\cdot-s} u_s \mathrm{d}s\Big\|_{C_T^{\alpha}, \gamma+\delta} \lesssim \sup_{t\in[0,T]} \int_0^t (t-s)^{-\frac{\delta}{2}-\alpha} \|u_s\|_\gamma \mathrm{d}s. \tag{5.31}$$

Proof. Let $s, t \in [0,T]$ with $s < t$. By the semigroup property of $(P_r)_{r\geq 0}$ we have

$$\int_0^t P_{t-r} u_r \mathrm{d}r - \int_0^s P_{s-r} u_r \mathrm{d}r = \int_s^t P_{t-r} u_r \mathrm{d}r + (P_{t-s} - \mathrm{id}) \int_0^s P_{s-r} u_r \mathrm{d}r.$$

Therefore, Lemma 5.4.8 and Lemma 5.5.8 imply for $\delta, \alpha > 0$ that

$$
\begin{aligned}
\Big\| \int_0^t P_{t-r} u_r \mathrm{d}r - \int_0^s P_{s-r} u_r \mathrm{d}r \Big\|_{\gamma+\delta} &\lesssim \int_s^t (t-r)^{-\frac{\delta}{2}} \|u_r\|_\gamma \mathrm{d}r \\
&\quad + (t-s)^\alpha \int_0^s (s-r)^{-\frac{\delta}{2}-\alpha} \|u_r\|_\gamma \mathrm{d}r \\
&\lesssim (t-s)^\alpha \sup_{t\in[0,T]} \int_0^t (t-r)^{-\frac{\delta}{2}-\alpha} \|u_r\|_\gamma \mathrm{d}r.
\end{aligned}
$$

□

Since we will solve again for $u - P_t u_0$, we need to allow for smooth perturbations in the following a priori estimate.

Lemma 5.5.10. *Let $\beta \in (0, 1/4)$ and $\alpha \in (0, \beta/4)$. Let $u \in \mathcal{D}^{\alpha,\beta,T}_{\dot W}$ and $w \in C^{1+\beta}_T \cap C^{\alpha,1/2+\beta}_T$. Let $F \in C^2_b$ and define*

$$
v_t := \int_0^t P_{t-s}(F(u_s + w_s) \cdot \dot W) \mathrm{d}s.
$$

Then $v \in \mathcal{D}^{\alpha,\beta,T}_{\dot W}$ with derivative $v'_t = F(u_t + w_t)$. For $T \in [0,1]$ we have

$$
\begin{aligned}
&\|v\|_{C^{\alpha,1/2+\beta}_T} + \Big\| v - \int_0^\cdot \pi_<(F(u_s+w_s), P_{\cdot-s}\dot W)\mathrm{d}s \Big\|_{C^{1+\beta}_T} \qquad (5.32)\\
&\qquad \lesssim_{F,\dot W,\Xi} (1 + \|u\|_{\dot W,\alpha,\beta,T})^2 \sup_{t\in[0,T]} \int_0^t (t-s)^{-\frac{3}{4}(1+\beta)-\alpha} (1 + \|w_s\|_{1+\beta})^2 \mathrm{d}s
\end{aligned}
$$

and

$$
\|v'\|_{C^{\alpha,1/2+\beta}_T} = \|F(u+w)\|_{C^{\alpha,1/2+\beta}_T} \lesssim_F \Big(1 + \|u\|_{C^{\alpha,1/2+\beta}_T} + \|w\|_{C^{\alpha,1/2+\beta}_T}\Big)^2. \qquad (5.33)
$$

Proof. For fixed s we can consider $t \mapsto u_t + w_s$ as controlled path, with derivative u', and controlled path norm $\|u\|_{\dot W,\alpha,\beta,T} + \|w_s\|_{1+\beta}$. Lemma 5.5.6 applied with $\gamma = \beta/2$ shows that $F(u_s + w_s) \cdot \dot W \in C^{-1-\beta/2}$, and that

$$
\|F(u_t + w_s) \cdot \dot W\|_{C^{-1-\beta/2}_T} \lesssim_{F,\dot W,\Xi} (1+T)^{\alpha \vee \frac{\beta}{4}} (1 + \|u\|_{\dot W,\alpha,\beta,T} + \|w_s\|_{1+\beta})^2.
$$

for $t \in [0,T]$. Next we apply Corollary 5.5.9 with $\gamma = -1-\beta/2$ and $\delta = 3/2(1+\beta)$ to get for $T \in [0,1]$ that

$$
\begin{aligned}
\|v\|_{C^{\alpha,1/2+\beta}_T} &= \Big\| \int_0^t P_{t-s}(F(u_s+w_s)\cdot \dot W)\mathrm{d}s \Big\|_{C^{\alpha,1/2+\beta}_T} \\
&\lesssim \sup_{t\in[0,T]} \int_0^t (t-s)^{-\frac{3}{4}(1+\beta)-\alpha} \|F(u_s+w_s)\cdot \dot W\|_{-1-\beta/2} \mathrm{d}s
\end{aligned}
$$

$$\lesssim_{F,\dot W,\Xi} (1+\|u\|_{\dot W,\alpha,\beta,T})^2 \sup_{t\in[0,T]} \int_0^t (t-s)^{-\frac{3}{4}(1+\beta)-\alpha}(1+\|w_s\|_{1+\beta})^2 \mathrm{d}r.$$

Since $\beta < 1/4$ and $\alpha < \beta/4$, the right hand side is finite. Furthermore, we clearly have

$$\Big\|v - \int_0^\cdot P_{\cdot-s}\pi_<(F(u_s+w_s),\dot W)\mathrm{d}s\Big\|_{C_T^{\alpha,1+\beta}} \lesssim \Big\|v - \int_0^\cdot P_{\cdot-s}\pi_<(F(u_s+w_s),\dot W)\mathrm{d}s\Big\|_{C_T^{\alpha,1+2\beta}}$$

The estimate (5.32) for this remainder now follows from Lemma 5.5.6 and Corollary 5.5.9 in the same manner as the estimate for $\|v\|_{C_T^{\alpha,1-\beta}}$, using that $F(u_s+w_s)\cdot\dot W - \pi_<(F(u_s+w_s),\dot W) \in C^{(\beta-1)/2}$, and that for $\delta = 3/2(1+\beta)$ as defined above we have $(\beta-1)/2+\delta = 1+2\beta$. But of course we did not subtract the right corrector, and therefore we need to consider the commutator between heat flow and paraproduct. Lemma 5.5.7 with $\delta = 3/2+\beta/2$ gives

$$\begin{aligned}\Big\|\int_0^t P_{t-s}\pi_<(F(u_s+w_s),\dot W)\mathrm{d}s - \int_0^t \pi_<(F(u_s+w_s),P_{t-s}\dot W)\mathrm{d}s\Big\|_{1+\beta}& \\ \lesssim \int_0^t (t-s)^{-\frac{3}{4}-\frac{\beta}{4}}\|F(u_s+w_s)\|_{1/2+\beta}\|\dot W\|_{-1-\beta/2}\mathrm{d}s& \\ \lesssim_{F,\dot W} (1+\|u\|_{C_T^{1/2+\beta}}) \sup_{t\in[0,T]}\int_0^t (t-s)^{-\frac{3}{4}-\frac{\beta}{4}}(1+\|w_s\|_{1/2+\beta})\mathrm{d}s.&\end{aligned}$$

Of course we increase the right hand side by replacing $(t-s)^{-\frac{3}{4}-\frac{\beta}{4}}(1+\|w_s\|_{1/2+\beta})$ with $(t-s)^{-\frac{3}{4}(1+\beta)-\alpha}(1+\|w_s\|_{1+\beta})^2$, and therefore estimate (5.32) follows.

It remains to control $v' = F(u+w)$. But using a first order Taylor expansion, it is easy to see that

$$\|F(u+w)\|_{C_T^{\alpha,1/2+\beta}} \lesssim \|F\|_{C_b^2}(1+\|u\|_{C_T^{\alpha,1/2+\beta}} + \|w\|_{C_T^{\alpha,1/2+\beta}})^2.$$

□

The estimate (5.33) for the derivative of v cannot be controlled by choosing T small. We will solve this problem by going back one step further in the Picard iteration.

Theorem 5.5.11. *Let $\beta \in (0,1/12)$ and $\alpha \in (0,\beta/4)$. Let $F \in C_b^3$ and $u_0 \in C^{1-}$. Then there exists a $\sigma(\dot W)\vee\sigma(u_0)$–measurable random time $\tau > 0$, such that equation* (5.25) *has a unique mild solution $u \in C^\alpha([0,\tau), C^{1-\beta})$.*

Proof. We solve for $v_t := u_t - P_t u_0$. Since $u_0 \in C^{1-}$ we have in particular that $u_0 \in C^{1-\beta}$. Let $T \in (0,1]$ to be specified below. We define the map

$$\Gamma_T : \mathcal{D}_{\dot W}^{\alpha,\beta,T} \to \mathcal{D}_{\dot W}^{\alpha,\beta,T} \qquad \text{by} \qquad \Gamma_T(v)_t := \int_0^t P_{t-s}(F(v_s+P_su_0)\cdot\dot W)\mathrm{d}s.$$

Lemma 5.5.10 implies that

$$\begin{aligned}\|\Gamma_T(v)\|_{C_T^{\alpha,1/2+\beta}} + \Big\|\Gamma_T(v) - \int_0^\cdot \pi_<(F(v_s+P_su_0), P_{\cdot-s}\dot W)\mathrm{d}s\Big\|_{C_T^{1+\beta}} \\ \lesssim_{F,\dot W,\Xi} (1+\|v\|_{\dot W,\alpha,\beta,T})^2 \sup_{t\in[0,T]}\int_0^t (t-s)^{-\frac{3}{4}(1+\beta)-\alpha}(1+\|P_su_0\|_{1+\beta})^2\mathrm{d}r.\end{aligned}$$

Now Lemma 5.4.8 yields $\|P_su_0\|_{1+\beta}\lesssim s^{-\beta}\|u_0\|_{1-\beta}$, and the assumptions on β and α are chosen exactly so that $3/4(1+\beta)+\alpha+2\beta<1$. Thus, we have

$$\sup_{t\in[0,T]}\int_0^t (t-s)^{-\frac{3}{4}(1+\beta)-\alpha}(1+\|P_su_0\|_{1+\beta})^2\mathrm{d}r \lesssim G(T)(1+\|u_0\|_{1-\beta})^2$$

for some continuous function G with $G(0)=0$. In particular, there exists a constant $C(f,\dot W,\Xi)>0$ such that

$$\begin{aligned}\|\Gamma_T(v)\|_{C_T^{\alpha,1/2+\beta}} + \Big\|\Gamma_T(v) - \int_0^\cdot \pi_<(F(v_s+P_su_0), P_{\cdot-s}\dot W)\mathrm{d}s\Big\|_{C_T^{1+\beta}} \\ \le C(f,\dot W,\Xi)G(T)(1+\|u_0\|_{1-\beta})^2(1+\|v\|_{\dot W,\alpha,\beta,T})^2.\end{aligned} \tag{5.34}$$

Moreover, Lemma 5.5.10 yields for the derivative $(\Gamma_T(v))'$ that

$$\begin{aligned}\|(\Gamma_T(v))'\|_{C_T^{\alpha,1/2+\beta}} = \|F(v+P_\cdot u_0)\|_{C_T^{\alpha,1/2+\beta}} &\lesssim_F (1+\|v\|_{C_T^{\alpha,1/2+\beta}}+\|P_\cdot u_0\|_{C_T^{\alpha,1/2+\beta}})^2 \\ &\lesssim_F (1+\|v\|_{C_T^{\alpha,1/2+\beta}})^2(1+\|P_\cdot u_0\|_{C_T^{\alpha,1/2+\beta}})^2.\end{aligned}$$

Now Lemma 5.5.8 implies for every $s<t\in[0,T]$ that

$$\|P_tu_0-P_su_0\|_{1/2+\beta} = \|P_s(P_{t-s}-\mathrm{id})u_0\|_{1-\beta-2(1/4-\beta)} \lesssim (t-s)^{1/4-\beta}\|u_0\|_{1-\beta}.$$

Since $\alpha<\beta/4$ and $\beta<1/12$, we have $\alpha<1/4-\beta$, and therefore $\|P_\cdot u_0\|_{C_T^{\alpha,1/2+\beta}}\lesssim \|u_0\|_{1-\beta}$. We conclude that there exists a constant $C(F)>0$ for which

$$\|(\Gamma_T(v))'\|_{C_T^{\alpha,1/2+\beta}} \le C(F)(1+\|v\|_{C_T^{\alpha,1/2+\beta}})^2(1+\|u_0\|_{C^{1-\beta}})^2. \tag{5.35}$$

Now choose $M>1$ such that $(C(F,\dot W,\Xi)+C(F))(1+\|u_0\|_{1-\beta})^2\le M$. We start the Picard iteration with $v^0\equiv 0$. Then (5.34) and (5.35) imply for $v^1:=\Gamma_T(v^0)$ that

$$1+\|v^1\|_{\dot W,\alpha,\beta,T}\le 1+G(T)M+M.$$

Since $M>1$, there exists $T>0$ for which the right hand side is smaller than $2M$. Let now $n\ge 1$, and suppose $1+\|v^i\|_{\dot W,\alpha,\beta,T}\le 2M$ for $i=n,n-1$. Define $v^{n+1}:=\Gamma_T(v^n)$. Then (5.34) gives

$$\|v^{n+1}\|_{C_T^{\alpha,1/2+\beta}} + \|(v^{n+1})^\sharp\|_{C_T^{1+\beta}} \le G(T)2M^3.$$

Moreover, (5.35) and then (5.34) imply for the derivative

$$\|(v^{n+1})'\|_{C_T^{\alpha,1/2+\beta}} \le M(1+\|v^n\|_{C_T^{\alpha,1/2+\beta}})^2 \le M(1+G(T)2M^3)^2.$$

It is now clear that there exists $T>0$, only depending on M, so that the Picard iteration sequence $(v^n)_{n\in\mathbb{N}}$ satisfies $\sup_n(1+\|v^n\|_{\dot{W},\alpha,\beta,T}) \le 2M$.

It remains to show the contraction property, but given the estimate on (v^n), this is now easy and follows by standard arguments.

Therefore, we obtain the existence of a unique solution u on $[0,T]$ for suitably small $T>0$. We controlled only the $C_T^{\alpha,1/2+\beta}$ norm of u, but from Lemma 5.5.10 it is clear that we actually have $u \in C_T^{\alpha,1-\varepsilon}$ for any $\varepsilon > 0$. In particular, we have $u_T \in C^{1-}$, and thus we can continue the construction on the next small time interval. Since we only have a quadratic a priori estimate, we are not able to prove the existence of global solutions. But we obtain the existence of a unique solution up to a random explosion time $\tau > 0$. □

Remark 5.5.12. The assumption $F \in C_b^3$ of course excludes the most interesting case when $F(x) = x$. That case can be included by a simple refinement of the analysis. For example, in Lemma 5.5.6 it is not strictly necessary to impose boundedness assumptions on F. The image of any $u \in \mathcal{D}_T^{\alpha,\beta}$ is a compact subset of $\mathbb{R}$, so for $F \in C^2$ we could consider a C_b^2 function that coincides with F on the image of u. In this way we would be able to obtain the existence of unique local solutions under the assumption $F \in C^3$.

Appendix

A. Incomplete filtrations

Here we collect some classical observations that allow us to transfer results which were obtained under complete filtrations to a probability space with an incomplete filtration.

It is an established tradition in probability theory to only work with filtrations satisfying the usual conditions. The most important reasons to consider complete filtrations are that the cross-section theorem ([DM78], III-44) only holds in complete σ–algebras, and as a consequence entrance times into Borel sets are generally only stopping times with respect to complete filtrations, and that supermartingales only have càdlàg modifications in complete filtrations.

But there are at least two classical monographs on stochastic analysis that avoid using complete filtrations as far as possible: Jacod [Jac79] (see the discussion on p. 8), and Jacod and Shiryaev [JS03] (see Definition I.1.2). We follow [JS03] in presenting results that allow us to pass from complete filtrations to incomplete filtrations.

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space with a right-continuous filtration $(\mathcal{F}_t)$. Write $\mathcal{F}^P$ for the P–completion of $\mathcal{F}$, and $\mathcal{N}^P$ for the P–null sets of $\mathcal{F}^P$. Then $\mathcal{F}^P_t := \mathcal{F}_t \vee \mathcal{N}^P$ is the σ–algebra generated by $\mathcal{F}_t$ and $\mathcal{N}^P$, and $(\mathcal{F}^P_t)$ satisfies the usual conditions. We call $(\Omega, \mathcal{F}^P, (\mathcal{F}^P_t), P)$ the *completion* of $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$.

Recall that the optional σ–algebra over $(\mathcal{F}_t)$ is the σ–algebra on $\Omega \times [0, \infty)$ that is generated by all processes of the form $X_t(\omega) = 1_A(\omega)1_{[r,s)}(t)$ for some $0 \leq r < s < \infty$ and $A \in \mathcal{F}_r$. The predictable σ–algebra over $(\mathcal{F}_t)$ is the σ–algebra on $\Omega \times [0, \infty)$ that is generated by all processes of the form $X_t(\omega) = 1_A(\omega)1_{\{0\}}(t) + 1_B(\omega)1_{(r,s]}(t)$ for some $0 \leq r < s < \infty$, for $A \in \mathcal{F}_0$, and $B \in \mathcal{F}_r$. Similarly we define the predictable and optional σ–algebra over $(\mathcal{F}^P_t)$.

The first result is not a precise mathematical statement, and is intended to reassure the reader in this critical point:

Lemma A.1 ([JS03], Section I.4.d and Section III.6)**.** *Stochastic integration does not require a complete filtration.*

Next we relate stopping times under $(\mathcal{F}_t)$ and under $(\mathcal{F}^P_t)$.

Lemma A.2 (Lemma I.1.19 of [JS03])**.** *Any stopping time on the completion $(\Omega, (\mathcal{F}^P_t))$ is almost surely equal to a stopping time on $(\Omega, (\mathcal{F}_t))$.*

Most entrance times that are practically relevant are $(\mathcal{F}_t)$–stopping times:

Lemma A.3. *If S is a right-continuous adapted process with values in a metric space $(\mathbb{X}, d)$, and if A is an open or closed set, then the entrance time*

$$\tau_A := \inf\{t \geq 0 : S_t \in A\}$$

is a $(\mathcal{F}_t)$–stopping time.

The proof is easy and therefore omitted.

Lemma A.4. *Any predictable (respectively optional) $\mathbb{R}^d$–valued process on the completion $(\Omega, (\mathcal{F}_t^P))$ is indistinguishable from a predictable (respectively optional) process on $(\Omega, (\mathcal{F}_t))$.*

Proof. The predictable case is Lemma I.2.17 of [JS03]. The proof of the optional case works exactly in the same way: the claim is trivial for the generating processes described above, and we can use the monotone class theorem to pass to indicator functions of general optional sets. Then we use monotone convergence to pass to general optional processes. □

As a consequence, we obtain a similar result for càdlàg processes.

Lemma A.5. *Let S be an $\mathbb{R}^d$–valued, $(\mathcal{F}_t^P)$–adapted process that it almost surely càdlàg. Then S is indistinguishable from a $(\mathcal{F}_t)$–adapted process $\widetilde{S}$.*

The process $\widetilde{S}$ can be chosen such that $t \mapsto \widetilde{S}_t(\omega)$ is right-continuous for every $\omega \in \Omega$, and has left limits everywhere except at $\tau(\omega)$, where τ is a stopping time with $P(\tau = \infty) = 1$.

Proof. Since $(\mathcal{F}_t^P)$ is complete, S admits an indistinguishable version $\overline{S}$ that is $(\mathcal{F}_t^P)$–adapted and càdlàg for *every* $\omega \in \Omega$. This $\overline{S}$ is optional, so the existence of an indistinguishable version S' that is $(\mathcal{F}_t)$–adapted follows from Lemma A.4. Using S', the version $\widetilde{S}$ with the desired properties can now be constructed in the same way as in the proof of Theorem 1.3.1. □

B. Convex compactness and Tychonoff's theorem

Here we prove Tychonoff's theorem for countable products of convexly compact spaces. Recall the following definitions.

Definition B.1.

1. A set A is called *directed* if it is partially ordered and if for all $\alpha, \beta \in A$ there exists $\gamma \in A$ such that $\alpha \leq \gamma$ and $\beta \leq \gamma$.
2. Let $\mathbb{X}$ be a topological space. A *net* in $\mathbb{X}$ is a map from some directed set A to $\mathbb{X}$.
3. A net $\{x_\alpha\}_{\alpha \in A}$ in $\mathbb{X}$ *converges* to a point $x \in \mathbb{X}$ if for every open neighborhood U of x there exists $\alpha \in A$, such that $x_{\alpha'} \in U$ for every $\alpha' \geq \alpha$.

Example B.2. If $A = \mathbb{N}$, then a net in $\mathbb{X}$ is just a sequence with values in $\mathbb{X}$.

Žitković [Ž10] introduces the notation $\mathrm{Fin}(A)$, which denotes all non-empty finite subsets of a given set A. If X is a subset of a vector space, then $\mathrm{conv}(X)$ denotes the convex hull of X. Žitković then gives the following definition.

Definition B.3. Let $\{x_\alpha\}_{\alpha\in A}$ be a net in a topological vector space $\mathbb{X}$. A net $\{y_\beta\}_{\beta\in B}$ is called a *subnet of convex combinations of* $\{x_\alpha\}_{\alpha\in A}$ if there exists a map $D : B \to \mathrm{Fin}(A)$, such that

1. for every $\beta \in B$ we have $y_\beta \in \mathrm{conv}\{x_\alpha : \alpha \in D(\beta)\}$, and
2. for every $\alpha \in A$ there exists $\beta \in B$ such that $\alpha' \geq \alpha$ for all $\alpha' \in \bigcup_{\beta'\geq\beta} D(\beta')$.

Lemma B.4. *Let $\{y_\beta\}_{\beta\in B}$ be a subnet of convex combinations of $\{x_\alpha\}_{\alpha\in A}$, and let $\{z_\gamma\}_{\gamma\in C}$ be a subnet of convex combinations of $\{y_\beta\}_{\beta\in B}$. Then $\{z_\gamma\}_{\gamma\in C}$ is a subnet of convex combinations of $\{x_\alpha\}_{\alpha\in A}$.*

Proof. Let $D_B : B \to \mathrm{Fin}(A)$ and $D_C : C \to \mathrm{Fin}(B)$ be two maps as described in Definition B.3, D_B for $\{y_\beta\}_{\beta\in B}$ and D_C for $\{z_\gamma\}_{\gamma\in C}$. Define

$$D : C \to \mathrm{Fin}(A), \qquad D(\gamma) := \bigcup_{\beta\in D_C(\gamma)} D_B(\beta).$$

Then we have for all $\gamma \in C$ that

$$z_\gamma \in \mathrm{conv}\{y_\beta : \beta \in D_C(\gamma)\} \subseteq \mathrm{conv}\Big\{x_\alpha : \alpha \in \bigcup_{\beta\in D_C(\gamma)} D_B(\beta)\Big\} = \mathrm{conv}\{x_\alpha : \alpha \in D(\gamma)\},$$

and therefore condition 1. of Definition B.3 is satisfied. As for condition 2., let $\alpha \in A$. Then there exists $\beta \in B$, such that $\alpha' \geq \alpha$ for all $\alpha' \in \bigcup_{\beta'\geq\beta} D_B(\beta')$. For this β, there exists $\gamma \in C$, such that $\beta' \geq \beta$ for all $\beta' \in \bigcup_{\gamma'\geq\gamma} D_C(\gamma')$. Hence, $\alpha' \geq \alpha$ for all

$$\alpha' \in \bigcup_{\gamma'\geq\gamma} D(\gamma') = \bigcup_{\gamma'\geq\gamma}\ \bigcup_{\beta'\in D_C(\gamma')} D_B(\beta') \subseteq \bigcup_{\beta'\geq\beta} D_B(\beta').$$

□

One of the main results in [Ž10] is the following Lemma.

Lemma B.5. *A closed and convex subset $\mathbb{Y}$ of a topological vector space $\mathbb{X}$ is convexly compact if and only if for any net $\{x_\alpha\}_{\alpha\in A}$ in $\mathbb{Y}$ there exists a subnet of convex combinations $\{y_\beta\}_{\beta\in B}$, such that $\{y_\beta\}_{\beta\in B}$ converges to some $y \in \mathbb{Y}$.*

We will use this insight to prove a weak version of Tychonoff's theorem for convexly compact sets.

Proposition B.6. *For $n \in \mathbb{N}$ let $\mathbb{Y}_n$ be a convexly compact subset of the metric vector space $\mathbb{X}_n$. Then $\prod_{n\in\mathbb{N}} \mathbb{Y}_n$ is a convexly compact subset of $\prod_{n\in\mathbb{N}} \mathbb{X}_n$, equipped with the product topology.*

Proof. Let $\{x_\alpha : \alpha \in A\} = \{(x_\alpha(n))_{n\in\mathbb{N}} : \alpha \in A\}$ be a net in $\prod_{n\in\mathbb{N}} \mathbb{Y}_n$. Then $\{x_\alpha(0) : \alpha \in A\}$ is a net in $\mathbb{Y}_0$. By Lemma B.5 there exists a subnet of convex combinations $\{y^0_\beta : \beta \in B_0\}$ of $\{x_\alpha : \alpha \in A\}$, such that $\{y^0_\beta(0) : \beta \in B_0\}$ converges to some $y(0) \in \mathbb{Y}_0$. We can now inductively construct for every $k \geq 1$ a subnet of convex combinations $\{y^k_\beta : \beta \in B_k\}$ of $\{y^{k-1}_\beta : \beta \in B_{k-1}\}$, such that $\{y^k_\beta(k) : \beta \in B_k\}$ converges to some $y(k) \in \mathbb{Y}_k$. According to Lemma B.4, every $\{y^k_\beta : \beta \in B_k\}$ is a subnet of convex combinations of $\{x_\alpha : \alpha \in A\}$. We denote the corresponding maps from B_k to $\mathrm{Fin}(A)$ by D_k. Note that by construction $\{y^k_\beta(\ell) : \beta \in B_k\}$ converges to $y(\ell)$ for all $0 \leq \ell \leq k$. Now consider the directed set $\mathbb{N} \times A$ with the partial order $(k, \alpha) \leq (k', \alpha')$ if $k \leq k'$ and $\alpha \leq \alpha'$. Write d_ℓ for the distance on $\mathbb{Y}_\ell$. We define for $(k, \alpha) \in \mathbb{N} \times A$ the set of "admissible indices" as

$$C(k,\alpha) := \left\{\beta \in B_k : \alpha' \geq \alpha \text{ for all } \alpha' \in D_k(\beta), d_\ell(y^k_\beta(\ell), y(\ell)) \leq \frac{1}{k} \text{ for } \ell = 1, \dots, k\right\}.$$

The condition on $D_k(\beta)$ guarantees that the subnet of convex combinations that we are about to construct satisfies condition 2. of Definition B.3. By construction of the $\{y^k_\beta : \beta \in B_k\}$, every $C(k, \alpha)$ is non-empty. For every $(k, \alpha) \in \mathbb{N} \times A$ choose $\beta(k, \alpha) \in C(k, \alpha)$. Note that here we explicitly apply the Axiom of Choice! Set $z_{(k,\alpha)} := y^k_{\beta(k,\alpha)}$ and $D((k, \alpha)) := D_k(\beta(k, \alpha))$. Then $\{z_{(k,\alpha)} : (k, \alpha) \in \mathbb{N} \times A\}$ is a subnet of convex combinations of $\{x_\alpha : \alpha \in A\}$, which converges to $(y(n))_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} \mathbb{Y}_n$ in the product topology. Now Lemma B.5 implies that $\prod_n \mathbb{Y}_n$ is convexly compact. □

Remark B.7. The proof is surprisingly technical considering that we are dealing with a countable product of metric spaces. In this case compactness is equivalent to sequential compactness, and therefore the proof of Tychonoff's theorem follows from a diagonal sequence argument. But so far there seems to be no characterization of convex compactness in terms of sequential compactness, and therefore we had to work with nets rather than with sequences.

C. Conditioning on null sets

In Chapter 2 we constructed a probability measure $\widetilde{Q}$ by conditioning P on the null set $\{T_0 = \mathfrak{T}\} = \bigcap_{a\in[0,\infty)}\{T_a \leq T_0\}$ using an extension theorem. It is important to point out that the choice of the approximating sequence of events, necessary for this construction, is highly relevant. Consider for example $\Omega = [0, 1]$, equipped with the Lebesgue measure P, and let us assume that we want to condition P on $\{0, 1\}$. Then for any $p \in [0, 1]$ the sequence of sets $([0,(1-p)/n] \cup [1-p/n, 1])_{n\in\mathbb{N}}$ has $\{0, 1\}$ as its intersection, and $P(\cdot|A_n)$ converges to $(1-p)\delta_0 + p\delta_1$, where δ_0 and δ_1 denote the Dirac masses in 0 and 1 respectively. In this case there is of course a "canonical candidate" for the conditioned measure: For $p = 1/2$ we obtain $1/2\delta_0 + 1/2\delta_1$ in the limit.

Below we give an example that is more relevant to the situation studied in Chapter 2. We remark that Knight [Kni69] illustrates the same point with another example, which, in our opinion, is slightly more involved than the one presented in the following.

Consider the continuous martingale $\widetilde{X}$, defined as

$$\widetilde{X}_t = X_t + (X_t - 1)\mathbf{1}_{\{\tau_{3/4} \geq t\}} + \left(\frac{1}{8} - \frac{X_t}{2}\right)\mathbf{1}_{\{\tau_{3/4} < t \leq \tau_{1/4}\}}.$$

The process $\widetilde{X}$ moves twice as much as X until X hits 3/4, then it moves half as much as X until X catches up, which occurs when X hits 1/4. With this understanding, it is clear that $\widetilde{X}$ hits zero exactly when X does. Therefore, we have that $\{\tau_0 = \mathfrak{T}\} = \bigcap_{a \in [0,\infty)} \{\widetilde{\tau}_a \leq \widetilde{\tau}_0\}$, where $\widetilde{\tau}_a$ is defined exactly as τ_a with X replaced by $\widetilde{X}$ in (2.1).

Now, it is easy to see that $P(\cdot | \widetilde{\tau}_a \leq \tau_0)$ defines a consistent family of probability measures on the filtration $(\mathcal{F}_{\tau_0 \wedge \widetilde{\tau}_a})_{a>1}$; namely the one defined through the Radon-Nikodym derivatives $\widetilde{X}_{\tau_a}$. Since $P(\widetilde{X}_{\tau_a} \neq X_{\tau_a}) > 0$ for $a > 1/4$, the induced measure differs from the one in Theorem 2.2.2. Therefore, although in the limit we condition on the same event, the induced probability measures strongly depend on the approximating sequence of events.

D. Pathwise Hoeffding inequality

In the construction of the pathwise Itô integral for typical price processes we needed the following result, a pathwise formulation of the Hoeffding inequality. This is a slight adaptation of [Vov12], Theorem A.1.

Lemma D.1 ([Vov12], Theorem A.1). *Let $(\tau_n)_{n\in\mathbb{N}}$ be a strictly increasing sequence of stopping times with $\tau_0 = 0$, such that for every $\omega \in \Omega$ we have $\tau_n(\omega) = \infty$ for all but finitely many $n \in \mathbb{N}$. Let for $n \in \mathbb{N}$ the function $h_n : \Omega \to \mathbb{R}^d$ be $\mathcal{F}_{\tau_n}$–measurable, and suppose that there exists a $\mathcal{F}_{\tau_n}$–measurable bounded function $b_n : \Omega \to \mathbb{R}$, such that*

$$\sup_{t\in[0,T]} |h_n(\omega)(\omega(\tau_{n+1} \wedge t) - \omega(\tau_n \wedge t))| \leq b_n(\omega) \tag{36}$$

for all $\omega \in \Omega$. Then for any $\lambda \in \mathbb{R}$ there exists a simple strategy $H^\lambda \in \mathcal{H}_1$, such that

$$1 + (H^\lambda \cdot \omega)_t \geq \exp\left(\lambda \sum_{n=0}^{\infty} h_n(\omega)(\omega(\tau_{n+1} \wedge t) - \omega(\tau_n \wedge t)) - \frac{\lambda^2}{2} \sum_{n=0}^{N_t(\omega)} b_n^2(\omega)\right)$$

for all $\omega \in \Omega$ and all $t \in [0,T]$, where $N_t(\omega) := \max\{n \in \mathbb{N} : \tau_n(\omega) \leq t\}$.

Proof. Let $\lambda \in \mathbb{R}$. The proof is based on the following deterministic inequality: if (36) is satisfied, then for all $\omega \in \Omega$ and all $t \in [0,T]$ we have that

$$\begin{aligned}
&\exp\left(\lambda h_n(\omega)(\omega(\tau_{n+1} \wedge t) - \omega(\tau_n \wedge t)) - \frac{\lambda^2}{2} b_n^2(\omega)\right) - 1 \\
&\qquad \leq \exp\left(-\frac{\lambda^2}{2} b_n^2(\omega)\right) \frac{e^{\lambda b_n(\omega)} - e^{-\lambda b_n(\omega)}}{2b_n(\omega)} h_n(\omega)(\omega(\tau_{n+1} \wedge t) - \omega(\tau_n \wedge t))
\end{aligned}$$

$$=: f_n(\omega)(\omega(\tau_{n+1} \wedge t) - \omega(\tau_n \wedge t)), \tag{37}$$

where we set $0/0 = 0$. This inequality is shown in (A.1) in [Vov12]. We define

$$H_t^\lambda(\omega) := \sum_{n=0}^{\infty} F_n(\omega) 1_{(\tau_n, \tau_{n+1}]}(t),$$

with F_n that have to be specified. We choose $F_0(\omega) := f_0(\omega)$, which is bounded and $\mathcal{F}_{\tau_0}$–measurable, and on $[0, \tau_1]$ we obtain from (37) that

$$1 + (H^\lambda \cdot \omega)_t \geq \exp\left(\lambda h_0(\omega)(\omega(\tau_1 \wedge t) - \omega(\tau_0 \wedge t)) - \frac{\lambda^2}{2} b_0^2(\omega)\right).$$

Observe also that $1 + (H^\lambda \cdot \omega)_{\tau_1} = 1 + f_0(\omega)(\omega(\tau_1) - \omega(\tau_0))$ is bounded, because by assumption $h_0(\omega)(\omega(\tau_1) - \omega(\tau_0))$ is bounded by the bounded random variable $b_0(\omega)$.

Assume now that F_k has been defined for $k = 0, \ldots, m-1$, that

$$1 + (H^\lambda \cdot \omega)_t \geq \exp\left(\lambda \sum_{n=0}^{\infty} h_n(\omega)(\omega(\tau_{n+1} \wedge t) - \omega(\tau_n \wedge t)) - \frac{\lambda^2}{2} \sum_{n=0}^{N_t(\omega)} b_n^2(\omega)\right)$$

for all $t \in [0, \tau_m]$, and that $1 + (H^\lambda \cdot \omega)_{\tau_m}$ is bounded. We define $F_m(\omega) := (1 + (H^\lambda \cdot \omega)_{\tau_m}) f_m(\omega)$, which is $\mathcal{F}_{\tau_m}$–measurable and bounded. From (37) we obtain for $t \in [\tau_m, \tau_{m+1}]$ that

$$\begin{aligned}
1 + (H^\lambda \cdot \omega)_t &= 1 + (H^\lambda \cdot \omega)_{\tau_m} + (1 + (H^\lambda \cdot \omega)_{\tau_m}) f_m(\omega)(\omega(\tau_{m+1} \wedge t) - \omega(\tau_m \wedge t)) \\
&\geq (1 + (H^\lambda \cdot \omega)_{\tau_m}) \exp\left(\lambda h_m(\omega)(\omega(\tau_{m+1} \wedge t) - \omega(\tau_m \wedge t)) - \frac{\lambda^2}{2} b_m^2(\omega)\right) \\
&\geq \exp\left(\lambda \sum_{n=0}^{\infty} h_n(\omega)(\omega(\tau_{n+1} \wedge t) - \omega(\tau_n \wedge t)) - \frac{\lambda^2}{2} \sum_{n=0}^{N_t(\omega)} b_n^2(\omega)\right),
\end{aligned}$$

where the last step follows from the induction hypothesis. From the first line of the last equation we also obtain that $1 + (H^\lambda \cdot \omega)_{\tau_{m+1}}$ is bounded, because $f_m(\omega)(\omega(\tau_{m+1}) - \omega(\tau_m))$ is bounded for the same reason for which $f_0(\omega)(\omega(\tau_1) - \omega(\tau_0))$ is bounded. □

E. Regularity for Schauder expansions with affine coefficients

Here we study the regularity of series of Schauder functions that have affine functions as coefficients. First let us establish an auxiliary result.

Lemma E.1. *Let $s < t$ and let $f : [s,t] \to \mathcal{L}(\mathbb{R}^d, \mathbb{R}^n)$ and $g : [s,t] \to \mathbb{R}^d$ be affine functions, $f(r) = a_1 + (r - s)b_1$, and $g(r) = a_2 + (r - s)b_2$. Then for all $r \in (s,t)$ and* *for all $h > 0$ with $r - h \in [s,t]$ and $r + h \in [s,t]$ we have*

$$|(fg)_{r-h,r} - (fg)_{r,r+h}| \lesssim |t-s|^{-2} h^2 \|f\|_\infty \|g\|_\infty. \tag{38}$$

Proof. We have

$$\begin{aligned}|(fg)_{r-h,r} - (fg)_{r,r+h}| &= |2f(r)g(r) - f(r-h)g(r-h) - f(r+h)g(r+h)| \\ &= |-h^2 b_1 b_2|.\end{aligned}$$

Now $f_{s,t} = b_1(t-s)$, and therefore $|b_1| \lesssim |t-s|^{-1}\|f\|_\infty$, and similarly for b_2. □

Now we are in a position to prove the regularity estimate.

Lemma E.2. *Let $\alpha \in (0,2)$ and let $(u_{pm}) \in \mathcal{A}^\alpha$. Then $\sum_{p,m} u_{pm}\varphi_{pm} \in \mathcal{C}^\alpha$, and*

$$\Big\|\sum_{p,m} u_{pm}\varphi_{pm}\Big\|_\alpha \lesssim \|(u_{pm})\|_{\mathcal{A}^\alpha}.$$

Proof. We need to examine the coefficients $2^{-q}\langle \chi_{qn}, \mathrm{d}(\sum_{pm} u_{pm}\varphi_{pm})\rangle$. First let us treat the case $(p,m) = (-1,0)$ and $(p,m) = (0,0)$. Here it suffices to estimate the uniform norm of $\sum_{pm} u_{pm}\varphi_{pm}$. For fixed p, the $(\varphi_{pm})_m$ have disjoint support, and therefore

$$\sum_p \Big\|\sum_m u_{pm}\varphi_{pm}\Big\|_\infty \lesssim \sum_p \max_m \|u_{pm}\|_\infty \le \sum_p 2^{-p\alpha}\|(u_{pm})\|_{\mathcal{A}^\alpha} \lesssim \|(u_{pm})\|_{\mathcal{A}^\alpha}.$$

Now let $q \ge 0$ and $1 \le n \le 2^q$ be given. If $p > q$, then $\varphi_{pm}(t^i_{qn}) = 0$ for $i = 0,1,2$ and for all m. Hence

$$2^{-q}\left\langle \chi_{qn}, \mathrm{d}\left(\sum_{p,m} u_{pm}\varphi_{pm}\right)\right\rangle = 2^{-q}\sum_{p\le q}\sum_m \langle \chi_{qn}, \mathrm{d}(u_{pm}\varphi_{pm})\rangle.$$

Now if $p < q$, then there is at most one m with $\langle \chi_{qn}, \mathrm{d}(u_{pm}\varphi_{pm})\rangle \ne 0$. For this m, the support of χ_{qn} is contained in $[t^0_{pm}, t^1_{pm}]$ or in $[t^1_{pm}, t^2_{pm}]$, and u_{pm} and φ_{pm} are affine on these intervals. Therefore, we can apply (38) to obtain

$$\begin{aligned}\sum_m |2^{-q}\langle \chi_{qn}, \mathrm{d}(u_{pm}\varphi_{pm})\rangle| &= \sum_m |(u_{pm}\varphi_{pm})_{t^0_{qn},t^1_{qn}} - (u_{pm}\varphi_{pm})_{t^1_{qn},t^2_{qn}}| \\ &\lesssim 2^{2p}2^{-2q}\|u_{pm}\|_\infty\|\varphi_{pm}\|_\infty \lesssim 2^{p(2-\alpha)-2q}\|(u_{pm})\|_{\mathcal{A}^\alpha}.\end{aligned}$$

For $p = q$ we have $\varphi_{qn}(t^0_{qn}) = \varphi_{qn}(t^2_{qn}) = 0$ and $\varphi_{qn}(t^1_{qn}) = 1/2$, and thus

$$\begin{aligned}\sum_m |2^{-q}\langle \chi_{qn}, \mathrm{d}(u_{qm}\varphi_{qm})\rangle| &= \Big|(u_{qn}\varphi_{qn})_{t^0_{qn},t^1_{qn}} - (u_{qn}\varphi_{qn})_{t^1_{qn},t^2_{qn}}\Big| \\ &= |u(t^1_{qn})| \lesssim 2^{-\alpha q}\|(u_{pm})\|_{\mathcal{A}^\alpha} = 2^{p(2-\alpha)-2q}\|(u_{pm})\|_{\mathcal{A}^\alpha}.\end{aligned}$$

We combine the estimates for $p < q$ and $p = q$, and obtain

$$2^{-q}\left|\left\langle \chi_{qn}, d\left(\sum_{pm} u_{pm}\varphi_{pm}\right)\right\rangle\right| \lesssim \sum_{p\leq q} 2^{p(2-\alpha)-2q}\|(u_{pm})\|_{\mathcal{A}^\alpha} \simeq 2^{-\alpha q}\|(u_{pm})\|_{\mathcal{A}^\alpha},$$

where we used that $\alpha < 2$ and therefore $\sum_{p\leq q} 2^{p(2-\alpha)} \simeq 2^{q(2-\alpha)}$. □

F. Different partitions of unity

Here we show that if (χ, ρ) and $(\widetilde{\chi}, \widetilde{\rho})$ are two dyadic partitions of unity, with paraproduct operators $\pi_<, \pi_>, \pi_\circ$, and $\widetilde{\pi}_<, \widetilde{\pi}_>, \widetilde{\pi}_\circ$ respectively, then $\pi_< - \widetilde{\pi}_<$, $\pi_> - \widetilde{\pi}_>$, and $\pi_\circ - \widetilde{\pi}_\circ$ are bounded bilinear operators from $C^\alpha \times C^\beta$ to $C^{\alpha+\beta}$ – regardless whether $\alpha + \beta > 0$ or not.

Lemma F.1. *Let $\alpha, \beta \in \mathbb{R}$. Let (χ, ρ) and $(\widetilde{\chi}, \widetilde{\rho})$ be two dyadic partitions of unity. Let $(\Delta_i)_{i\geq -1}$ be the Littlewood-Paley blocks corresponding to (χ, ρ), and $(\widetilde{\Delta}_i)_{i\geq -1}$ those corresponding to $(\widetilde{\chi}, \widetilde{\rho})$. Let*

$$\pi_<(u, v) := \sum_{i<j-N_1} \Delta_i u \Delta_j v \qquad \textit{and} \qquad \widetilde{\pi}_<(u, v) := \sum_{i<j-N_2} \widetilde{\Delta}_i u \widetilde{\Delta}_j v,$$
$$\pi_\circ(u, v) := \sum_{|i-j|\leq N_1} \Delta_i u \Delta_j v \qquad \textit{and} \qquad \widetilde{\pi}_\circ(u, v) = \sum_{|i-j|\leq N_2} \widetilde{\Delta}_i u \widetilde{\Delta}_j v,$$
$$\pi_>(u, v) := \sum_{j<i-N_1} \Delta_i u \Delta_j v \qquad \textit{and} \qquad \widetilde{\pi}_>(u, v) := \sum_{j<i-N_2} \widetilde{\Delta}_i u \widetilde{\Delta}_j v,$$

where N_1 is large enough so that $\Delta_i \Delta_j = 0$ for $|i - j| > N_1$, and similarly for N_2. Then $\pi_< - \widetilde{\pi}_<$, $\pi_\circ - \widetilde{\pi}_\circ$, and $\pi_> - \widetilde{\pi}_>$ are bounded bilinear operators from $C^\alpha \times \mathcal{C}^\beta$ to $C^{\alpha+\beta}$, such that for all $u \in C^\alpha$, $v \in C^\beta$

$$\|\pi_<(u, v) - \widetilde{\pi}_<(u, v)\|_{\alpha+\beta} \lesssim \|u\|_\alpha \|v\|_\beta,$$
$$\|\pi_\circ(u, v) - \widetilde{\pi}_\circ(u, v)\|_{\alpha+\beta} \lesssim \|u\|_\alpha \|v\|_\beta,$$
$$\|\pi_>(u, v) - \widetilde{\pi}_>(u, v)\|_{\alpha+\beta} \lesssim \|u\|_\alpha \|v\|_\beta.$$

Proof. The statement for $\pi_< - \widetilde{\pi}_<$ (and thus for $\pi_> - \widetilde{\pi}_>$) is shown in Bony [Bon81], Theorem 2.1. Let us prove the statement for $\pi_\circ - \widetilde{\pi}_\circ$:

Let $u \in C^\alpha$ and $v \in C^\beta$ both have Fourier transforms of compact support. As we argued before, it suffices to show the statement for such smooth u, v, and to extend the operator $\pi_\circ - \widetilde{\pi}_\circ$ by a continuity argument. We have

$$\begin{aligned}\pi_\circ(u, v) - \widetilde{\pi}_\circ(u, v) &= (uv - \pi_<(u, v) - \pi_>(u, v)) - (uv - \widetilde{\pi}_<(u, v) - \widetilde{\pi}_>(u, v)) \\ &= (\widetilde{\pi}_<(u, v) - \pi_<(u, v)) + (\widetilde{\pi}_>(u, v) - \pi_>(u, v)).\end{aligned}$$

Hence, the estimate for $\pi_\circ - \widetilde{\pi}_\circ$ follows from the estimates for $\pi_< - \widetilde{\pi}_<$ and $\pi_> - \widetilde{\pi}_>$. □

G. Paralinearization theorem

Here we prove the slightly modified version of the paralinearization theorem which was used in the proof of Lemma 5.3.18.

The following Lemma allows us to estimate the Hölder-Besov norm of a series of functions $\sum_j u_j$, where the Fourier transform of u_j is not necessarily of compact support.

Lemma G.1 (Lemma 2.88 of [BCD11])**.** *Let $\alpha > 0$ and let $(u_j)_{j\in\mathbb{N}}$ be a sequence of smooth functions such that*

$$\left(\sup_{|\eta|\in\{0,\lfloor\alpha\rfloor+1\}} 2^{j(\alpha-|\eta|)} \|\partial^\eta u_j\|_{L^\infty}\right)_{j\in\mathbb{N}} \in \ell^\infty(\mathbb{N}).$$

Then $u = \sum_j u_j \in C^\alpha$, and

$$\|u\|_{C^\alpha} \lesssim_\alpha \left\| \left(\sup_{|\eta|\in\{0,\lfloor\alpha\rfloor+1\}} 2^{j(\alpha-|\eta|)} \|\partial^\eta u_j\|_{L^\infty}\right)_{j\in\mathbb{N}} \right\|_{\ell^\infty(\mathbb{N})}.$$

The following lemma about the action of a smooth function on $u \in C^\alpha$ is of course well known.

Lemma G.2. *Let $\alpha \in (0,1)$. If $u \in C^\alpha(\mathbb{R}^d,\mathbb{R}^n)$ and $F \in C^1_b(\mathbb{R}^n,\mathbb{R})$, then*

$$\|F(u)\|_\alpha \lesssim \|F\|_{C^1_b}\|u\|_\alpha.$$

Proof. Since $C^\alpha(\mathbb{R}^d,\mathbb{R}^n)$ is the space of bounded α–Hölder continuous functions, see p. 99 of [BCD11], the result is easily obtained by a first order Taylor expansion. □

Now we are in a position to prove the paralinearization theorem.

Lemma G.3. *Let $\alpha \in (0,1/2)$. If $u \in C^\alpha(\mathbb{R}^d,\mathbb{R}^n)$, $v \in C^{2\alpha}(\mathbb{R}^d,\mathbb{R}^n)$, and $F \in C^2_b(\mathbb{R}^n,\mathbb{R})$ with $F(0) = 0$, then*

$$\|F(u+v) - \pi_<(\mathrm{D}F(u+v),u)\|_{2\alpha} \lesssim (\|F\|_{L^\infty} + \|\mathrm{D}^2F\|_{L^\infty})(\|u\|_\alpha + \|v\|_{2\alpha})(1+\|u\|_\alpha).$$

In particular, we have for general $F \in C^2_b$, not necessarily satisfying $F(0) = 0$, that

$$\|F(u+v) - \pi_<(F(u+v),u)\|_{2\alpha} \lesssim \|F\|_{C^2_b}(1+\|v\|_{2\alpha})(1+\|u\|_\alpha)^2.$$

Proof. We will apply Lemma G.1 to the series

$$F(u+v) - \pi_<(\mathrm{D}F(u+v),u) = \sum_j F_j,$$

where

$$F_j := (F(S_{j+1}(u+v)) - F(S_j(u+v))) - S_{j-1}\mathrm{D}F(u+v)\Delta_j u. \tag{39}$$

We apply a Taylor expansion up to second order to the first term of F_j and obtain

$$\begin{aligned}
F(S_{j+1}(u+v)) - F(S_j(u+v)) &= \sum_{k=1}^{d} \int_0^1 \partial_{x_k} F(S_j(u+v) + r\Delta_j(u+v))\mathrm{d}r \Delta_j(u^k+v^k) \\
&= \mathrm{D}F(S_j(u+v))\Delta_j u \\
&\quad + \sum_{k,\ell=1}^{d} \int_0^1 (1-r)\frac{\partial^2}{\partial x_\ell \partial x_k} F(S_j(u+v)+r\Delta_j(u+v))\mathrm{d}r\Delta_j(u^\ell+v^\ell)\Delta_j u^k \\
&\quad + \int_0^1 \mathrm{D}F(S_j(u+v)+r\Delta_j(u+v))\Delta_j v \mathrm{d}r. \qquad (40)
\end{aligned}$$

Combining (39) and (40), we see that

$$\begin{aligned}
\|F_j\|_{L^\infty} &\lesssim \|(\mathrm{D}F(S_j(u+v)) - S_{j-1}\mathrm{D}F(u+v))\Delta_j u\|_{L^\infty} \\
&\quad + \left\| \int_0^1 \mathrm{D}F(S_j(u+v)+r\Delta_j(u+v))\Delta_j v\mathrm{d}r \right\|_{L^\infty} \\
&\quad + \left\| \sum_{k,\ell=1}^{d} \int_0^1 (1-r)\frac{\partial^2}{\partial x_\ell \partial x_k} F(S_j(u+v)+r\Delta_j(u+v))\mathrm{d}r\Delta_j(u^\ell+v^\ell)\Delta_j u^k \right\|_{L^\infty} \\
&\lesssim \|(\mathrm{D}F(S_j(u+v)) - \mathrm{D}F(u+v))\|_{L^\infty} 2^{-j\alpha}\|u\|_\alpha \\
&\quad + \|(\mathrm{D}F(u+v) - S_{j-1}\mathrm{D}F(u+v))\|_{L^\infty} 2^{-j\alpha}\|u\|_\alpha \\
&\quad + \|\mathrm{D}F\|_{L^\infty} 2^{-j2\alpha}\|v\|_{2\alpha} + \|\mathrm{D}^2F\|_{L^\infty} 2^{-j2\alpha}(\|u\|_\alpha + \|v\|_\alpha)\|u\|_\alpha. \qquad (41)
\end{aligned}$$

Now a Taylor expansion with integral remainder applied to $\mathrm{D}F(S_j(u+v)) - \mathrm{D}F(u+v)$ yields

$$\begin{aligned}
\|\mathrm{D}F(S_j(u+v)) - \mathrm{D}F(u+v)\|_{L^\infty} &\lesssim \|\mathrm{D}^2F\|_{L^\infty} \sum_{k\geq j} \|\Delta_k(u+v)\|_{L^\infty} \\
&\lesssim \|\mathrm{D}^2F\|_{L^\infty} 2^{-j\alpha}(\|u\|_\alpha + \|v\|_\alpha). \qquad (42)
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\|\mathrm{D}F(u+v) - S_{j-1}\mathrm{D}F(u+v)\|_{L^\infty} &\lesssim \sum_{k\geq j-1} \|\Delta_k \mathrm{D}F(u+v)\|_{L^\infty} \lesssim 2^{-j\alpha}\|\mathrm{D}F(u+v)\|_\alpha \\
&\lesssim 2^{-j\alpha}(\|\mathrm{D}F\|_{L^\infty} + \|\mathrm{D}^2F\|_{L^\infty})(\|u\|_\alpha + \|v\|_\alpha), \qquad (43)
\end{aligned}$$

where we used Lemma G.2 in the last step. Combining (41), (42), and (43), we obtain

$$\|F_j\|_{L^\infty} \lesssim 2^{-2j\alpha}(\|\mathrm{D}F\|_{L^\infty} + \|\mathrm{D}^2F\|_{L^\infty})(\|u\|_\alpha + \|v\|_{2\alpha})(1+\|u\|_\alpha).$$

For the derivatives we do not take advantage of any cancellations, but just estimate every term on its own. We have for all $\eta \in \mathbb{N}^d$ with $|\eta| = 1$ that

$$
\begin{aligned}
\partial^\eta F_j &= \partial^\eta [F(S_{j+1}(u+v)) - F(S_j(u+v)) - S_{j-1}\mathrm{D}F(u+v)\Delta_j u] \\
&= \mathrm{D}F(S_{j+1}(u+v))\partial^\eta S_{j+1}(u+v) - \mathrm{D}F(S_j(u+v))\partial^\eta S_j(u+v) \\
&\qquad - \partial^\eta [S_{j-1}\mathrm{D}F(u+v)]\Delta_j u - S_{j-1}\mathrm{D}F(u+v)\partial^\eta \Delta_j u \\
&= a_j + b_j + c_j + d_j,
\end{aligned}
$$

where

$$
\begin{aligned}
a_j &:= [\mathrm{D}F(S_{j+1}(u+v)) - \mathrm{D}F(S_j(u+v))]\partial^\eta S_{j+1} u, \\
b_j &:= \mathrm{D}F(S_{j+1}(u+v))\partial^\eta S_{j+1} v - \mathrm{D}F(S_j(u+v))\partial^\eta S_j v, \\
c_j &:= [\mathrm{D}F(S_j(u+v)) - S_{j-1}\mathrm{D}F(u+v)]\partial^\eta \Delta_j u, \text{ and} \\
d_j &:= -\partial^\eta (S_{j-1}\mathrm{D}F(u+v))\Delta_j u.
\end{aligned}
$$

Using first a Taylor expansion for a_j, and then that $1 - \alpha > 0$, we obtain

$$
\begin{aligned}
\|a_j\|_{L^\infty} &\leq \int_0^1 \|\mathrm{D}^2 F(S_j(u+v) + r\Delta_j(u+v))\|_{L^\infty} \mathrm{d}r \|\Delta_j(u+v)\|_{L^\infty} \|S_{j+1}\partial^\eta u\|_{L^\infty} \\
&\lesssim \|\mathrm{D}^2 F\|_{L^\infty} 2^{-j\alpha}(\|u\|_\alpha + \|v\|_\alpha) 2^{j(1-\alpha)} \|\partial^\eta u\|_{\alpha-1} \\
&\lesssim 2^{j(1-2\alpha)} \|\mathrm{D}^2 F\|_{L^\infty} (\|u\|_\alpha + \|v\|_\alpha)\|u\|_\alpha.
\end{aligned} \tag{44}
$$

Since $1 - 2\alpha > 0$, the next term b_j can be easily estimated by

$$
\|b_j\|_{L^\infty} \lesssim \|\mathrm{D}F\|_{L^\infty} \|\partial^\eta S_{j+1} v\|_{L^\infty} + \|\mathrm{D}F\|_{L^\infty} \|\partial^\eta S_j v\|_{L^\infty} \lesssim 2^{j(1-2\alpha)} \|\mathrm{D}F\|_{L^\infty} \|v\|_{2\alpha}. \tag{45}
$$

Estimates (42) and (43) imply that

$$
\|c_j\|_{L^\infty} \lesssim 2^{-j\alpha}(\|\mathrm{D}F\|_{L^\infty} + \|\mathrm{D}^2 F\|_{L^\infty})(\|u\|_\alpha + \|v\|_\alpha) 2^{j(1-\alpha)} \|u\|_\alpha \tag{46}
$$

$$
= 2^{j(1-2\alpha)} (\|\mathrm{D}F\|_{L^\infty} + \|\mathrm{D}^2 F\|_{L^\infty})(\|u\|_\alpha + \|v\|_\alpha)\|u\|_\alpha. \tag{47}
$$

Finally, we have for the last term

$$
\begin{aligned}
\|d_j\|_{L^\infty} &\leq \|S_{j-1}(\partial^\eta \mathrm{D}F(u+v))\|_{L^\infty} \|\Delta_j u\|_{L^\infty} \lesssim 2^{j(1-\alpha)} \|\partial^\eta \mathrm{D}F(u+v)\|_{\alpha-1} 2^{-j\alpha} \|u\|_\alpha \\
&\lesssim 2^{j(1-2\alpha)} \|\mathrm{D}F(u+v)\|_\alpha \|u\|_\alpha,
\end{aligned}
$$

so that another application of Lemma G.2 yields

$$
\|d_j\|_{L^\infty} \lesssim 2^{j(1-2\alpha)} (\|\mathrm{D}F\|_{L^\infty} + \|\mathrm{D}^2 F\|_{L^\infty})(\|u\|_\alpha + \|v\|_\alpha)\|u\|_\alpha. \tag{48}
$$

Combining (44)–(48), we see that

$$
\|\partial^\eta F_j\|_{L^\infty} \lesssim 2^{j(1-2\alpha)} (\|\mathrm{D}F\|_{L^\infty} + \|\mathrm{D}^2 F\|_{L^\infty})(\|u\|_\alpha + \|v\|_{2\alpha})(1 + \|u\|_\alpha).
$$

The proof is completed by applying Lemma G.1 to $\sum_j F_j$. $\square$

Bibliography

[ADI06] Stefan Ankirchner, Steffen Dereich, and Peter Imkeller, *The Shannon information of filtrations and the additional logarithmic utility of insiders*, Ann. Probab. **34** (2006), no. 2, 743–778.

[ADI07] ______, *Enlargement of filtrations and continuous Girsanov-type embeddings*, Séminaire de Probabilités XL, Springer, 2007, pp. 389–410.

[AIP13] Andreas Andresen, Peter Imkeller, and Nicolas Perkowski, *Large Deviations for Hilbert-Space-Valued Wiener Processes: A Sequence Space Approach*, Malliavin Calculus and Stochastic Analysis, Springer, 2013, pp. 115–138.

[AIS98] Jürgen Amendinger, Peter Imkeller, and Martin Schweizer, *Additional logarithmic utility of an insider*, Stochastic Process. Appl. **75** (1998), no. 2, 263–286.

[ALP95] Marco Avellaneda, Arnon Levy, and Antonio Parás, *Pricing and hedging derivative securities in markets with uncertain volatilities*, Appl. Math. Finance **2** (1995), no. 2, 73–88.

[Ame00] Jürgen Amendinger, *Martingale representation theorems for initially enlarged filtrations*, Stochastic Process. Appl. **89** (2000), no. 1, 101–116.

[Ank05] Stefan Ankirchner, *Information and semimartingales*, Ph.D. thesis, Humboldt-Universität zu Berlin, 2005.

[Bac00] Louis Bachelier, *Théorie de la spéculation*, Ann. Sci. École Norm. Sup. (3) **17** (1900), 21–86.

[Bau02] Fabrice Baudoin, *Conditioned stochastic differential equations: theory, examples and application to finance*, Stochastic Process. Appl. **100** (2002), 109–145.

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

[Bec01] Dirk Becherer, *The numeraire portfolio for unbounded semimartingales*, Finance Stoch. **5** (2001), no. 3, 327–341.

[BGL94] Gérard Ben Arous, Mihai Grădinaru, and Michel Ledoux, *Hölder norms and the support theorem for diffusions*, Ann. Inst. H. Poincaré Probab. Statist. **30** (1994), no. 3, 415–436.

[BGN10] Zdzisław Brzeźniak, Massimiliano Gubinelli, and Misha Neklyudov, *Global evolution of random vortex filament equation*, arXiv preprint arXiv:1008.1086 (2010).

[BGR05] Hakima Bessaih, Massimiliano Gubinelli, and Francesco Russo, *The evolution of a random vortex filament*, Ann. Probab. **33** (2005), no. 5, 1825–1855.

[BHLP11] Mathias Beiglböck, Pierre Henry-Labordère, and Friedrich Penkner, *Model-independent Bounds for Option Prices: A Mass Transport Approach*, arXiv preprint arXiv:1106.5929 (2011).

[BL94] G. Ben Arous and M. Ledoux, *Grandes déviations de Freidlin-Wentzell en norme hölderienne*, Séminaire de Probabilités, XXVIII, Lecture Notes in Math., vol. 1583, Springer, Berlin, 1994, pp. 293–299.

[Bon81] Jean-Michel Bony, *Calcul symbolique et propagation des singularites pour les équations aux dérivées partielles non linéaires*, Ann. Sci. Éc. Norm. Supér. (4) **14** (1981), 209–246.

[BR92] Paolo Baldi and Bernard Roynette, *Some exact equivalents for the Brownian motion in Hölder norm*, Probab. Theory Related Fields **93** (1992), no. 4, 457–484.

[Bru79] Michel Bruneau, *Sur la p-variation d'une surmartingale continue*, Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78), Lecture Notes in Math., vol. 721, Springer, Berlin, 1979, pp. 227–232.

[BS73] Fischer Black and Myron Scholes, *The pricing of options and corporate liabilities*, The Journal of Political Economy (1973), 637–654.

[CF09] Michael Caruana and Peter Friz, *Partial differential equations driven by rough paths*, J. Differential Equations **247** (2009), no. 1, 140–173.

[CF10] Rama Cont and David-Antoine Fournié, *Change of variable formulas for non-anticipative functionals on path space*, J. Funct. Anal. **259** (2010), no. 4, 1043–1072.

[CFO11] Michael Caruana, Peter K. Friz, and Harald Oberhauser, *A (rough) pathwise approach to a class of non-linear stochastic partial differential equations*, Annales de l'Institut Henri Poincare (C) Non Linear Analysis **28** (2011), no. 1, 27–46.

[CFR12] Peter Carr, Travis Fisher, and Johannes Ruf, *On the hedging of options on exploding exchange rates*, arXiv preprint arXiv:1202.6188 (2012).

[CG06] Jean-Yves Chemin and Isabelle Gallagher, *On the global wellposedness of the 3-D Navier-Stokes equations with large initial data*, Ann. Sci. École Norm. Sup. (4) **39** (2006), no. 4, 679–698.

[Cie60] Zbigniew Ciesielski, *On the isomorphisms of the space H^α and m*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. (1960).

[CKR93] Zbigniew Ciesielski, Gérard Kerkyacharian, and Bernard Roynette, *Quelques espaces fonctionnels associés à des processus gaussiens*, Studia Math. **107** (1993), no. 2, 171–204.

[CM94] René A. Carmona and S. A. Molchanov, *Parabolic Anderson problem and intermittency*, Mem. Amer. Math. Soc. **108** (1994), no. 518.

[CQ02] Laure Coutin and Zhongmin Qian, *Stochastic analysis, rough path analysis and fractional Brownian motions.*, Probab. Theory Relat. Fields **122** (2002), no. 1, 108–140.

[Del69] C. Dellacherie, *Ensembles aléatoires. I*, Séminaire de Probabilités, III (Univ. Strasbourg, 1967/68), Springer, Berlin, 1969, pp. 97–114.

[DF12] Joscha Diehl and Peter Friz, *Backward stochastic differential equations with rough drivers*, Ann. Probab. **40** (2012), no. 4, 1715–1758.

[DGT12] A. Deya, M. Gubinelli, and S. Tindel, *Non-linear rough heat equations*, Probab. Theory Related Fields **153** (2012), no. 1-2, 97–147.

[DM78] Claude Dellacherie and Paul-André Meyer, *Probabilities and potential*, North-Holland Mathematics Studies, vol. 29, North-Holland Publishing Co., Amsterdam, 1978.

[DM82] ———, *Probabilities and potential. B*, North-Holland Mathematics Studies, vol. 72, North-Holland Publishing Co., Amsterdam, 1982, Theory of martingales, Translated from the French by J. P. Wilson.

[DM06] Laurent Denis and Claude Martini, *A theoretical framework for the pricing of contingent claims in the presence of model uncertainty*, Ann. Appl. Probab. **16** (2006), no. 2, 827–852.

[Doo57] Joseph L. Doob, *Conditional Brownian motion and the boundary limits of harmonic functions*, Bull. Soc. Math. France **85** (1957), 431–458.

[DOR13] Mark Davis, Jan Obłój, and Vimal Raval, *Arbitrage bounds for prices of weighted variance swaps*, Math. Finance (2013), to appear.

[DS94] Freddy Delbaen and Walter Schachermayer, *A general version of the fundamental theorem of asset pricing*, Math. Ann. **300** (1994), no. 1, 463–520.

[DS95a] ———, *Arbitrage possibilities in Bessel processes and their relations to local martingales*, Probab. Theory Related Fields **102** (1995), no. 3, 357–366.

[DS95b] ———, *The existence of absolutely continuous local martingale measures*, Ann. Appl. Probab. **5** (1995), no. 4, 926–945.

[DS95c] ______, *The no-arbitrage property under a change of numéraire*, Stochastics Stochastics Rep. **53** (1995), 213–226.

[DS06] ______, *The mathematics of arbitrage*, Springer Finance, Springer, Berlin, 2006.

[DS12] Yan Dolinsky and H. Mete Soner, *Robust Hedging and Martingale Optimal Transport in Continuous Time*, arXiv preprint arXiv:1208.4922 (2012).

[Dur10] Rick Durrett, *Probability: theory and examples*, fourth ed., Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2010.

[EK86] Stewart N. Ethier and Thomas G. Kurtz, *Markov processes: Characterization and convergence*, John Wiley & Sons, 1986.

[ENO99] M'hamed Eddahbi, Modeste N'zi, and Youssef Ouknine, *Grandes déviations des diffusions sur les espaces de Besov-Orlicz et application*, Stochastics Stochastics Rep. **65** (1999), no. 3-4, 299–315.

[FG06] Hans Föllmer and Anne Gundel, *Robust projections in the class of martingale measures*, Illinois J. Math. **50** (2006), no. 1-4, 439–472.

[FGGR12] Peter K Friz, Benjamin Gess, Archil Gulisashvili, and Sebastian Riedel, *Spatial rough path lifts of stochastic convolutions*, arXiv preprint arXiv:1211.0046 (2012).

[FH13] Peter Friz and Martin Hairer, *A short course on rough paths*, in preparation (2013).

[FI93] Hans Föllmer and Peter Imkeller, *Anticipation cancelled by a Girsanov transformation: a paradox on Wiener space*, Ann. Inst. Henri Poincaré Probab. Stat. **29** (1993), no. 4, 569–586.

[FO11] Peter Friz and Harald Oberhauser, *On the splitting-up method for rough (partial) differential equations*, J. Differential Equations **251** (2011), no. 2, 316–338.

[Föl72] Hans Föllmer, *The exit measure of a supermartingale*, Probab. Theory Related Fields **21** (1972), no. 2, 154–166.

[Föl79] ______, *Calcul d'Itô sans probabilités*, Séminaire de Probabilités XV **80** (1979), 143–150.

[FV10a] Peter Friz and Nicolas Victoir, *Differential equations driven by Gaussian signals*, Ann. Inst. Henri Poincaré Probab. Stat. **46** (2010), no. 2, 369–413.

[FV10b] ______, *Multidimensional stochastic processes as rough paths. Theory and applications*, Cambridge University Press, 2010.

[Gan91] Nina Gantert, *Einige große Abweichungen der Brownschen Bewegung*, Ph.D. thesis, Universität Bonn, 1991.

[Gan94] ———, *Self-similarity of Brownian motion and a large deviation principle for random fields on a binary tree*, Probab. Theory Related Fields **98** (1994), no. 1, 7–20.

[GIP12] Massimiliano Gubinelli, Peter Imkeller, and Nicolas Perkowski, *Paraproducts, rough paths and controlled distributions*, arXiv preprint arXiv:1210.2684 (2012).

[GLT06] Massimiliano Gubinelli, Antoine Lejay, and Samy Tindel, *Young integrals and SPDEs*, Potential Anal. **25** (2006), no. 4, 307–326.

[Gub04] Massimiliano Gubinelli, *Controlling rough paths*, J. Funct. Anal. **216** (2004), no. 1, 86–140.

[Gub10] ———, *Ramification of rough paths*, J. Differential Equations **248** (2010), no. 4, 693–721.

[Gub12] ———, *Rough solutions for the periodic Korteweg–de Vries equation*, Commun. Pure Appl. Anal. **11** (2012), no. 2, 709–733.

[Hai11] Martin Hairer, *Rough stochastic PDEs*, Comm. Pure Appl. Math. **64** (2011), no. 11, 1547–1585.

[Hai13a] ———, *A theory of regularity structures*, arXiv preprint arXiv:1303.5113 (2013).

[Hai13b] ———, *Solving the KPZ equation*, Ann. Math. **178** (2013), no. 2, 559–664.

[HMW12] Martin Hairer, Jan Maas, and Hendrik Weber, *Approximating rough stochastic PDEs*, arXiv preprint arXiv:1202.3094 (2012).

[HP81] J. Michael Harrison and Stanley R. Pliska, *Martingales and stochastic integrals in the theory of continuous trading*, Stochastic Process. Appl. **11** (1981), no. 3, 215–260.

[HW13] Martin Hairer and Hendrik Weber, *Rough Burgers-like equations with multiplicative noise*, Probab. Theory Related Fields **155** (2013), no. 1-2, 71–126.

[IP11] Peter Imkeller and Nicolas Perkowski, *The existence of dominating local martingale measures*, arXiv preprint arXiv:1111.3885 (2011).

[IPW01] Peter Imkeller, Monique Pontier, and Ferenc Weisz, *Free lunch and arbitrage possibilities in a financial market model with an insider*, Stochastic Process. Appl. **92** (2001), no. 1, 103–130.

[Jac79] Jean Jacod, *Calcul stochastique et problemes de martingales*, vol. 714, Springer, 1979.

[Jac85] ———, *Grossissement initial, hypothese (H) et theoreme de Girsanov*, Grossissements de filtrations: exemples et applications; Lecture Notes in Mathematics Vol. 1118, Springer, 1985, pp. 15–35.

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

[JPS10] Robert A. Jarrow, Philip Protter, and Kazuhiro Shimbo, *Asset price bubbles in incomplete markets*, Math. Finance **20** (2010), no. 2, 145–185.

[JS03] Jean Jacod and Albert N. Shiryaev, *Limit theorems for stochastic processes*, 2nd ed., Springer, 2003.

[Kam94] Anna Kamont, *Isomorphism of some anisotropic Besov and sequence spaces*, Studia Math. **110** (1994), no. 2, 169–189.

[Kar95] Rajeeva L. Karandikar, *On pathwise stochastic integration*, Stochastic Process. Appl. **57** (1995), no. 1, 11–18.

[Kar10a] Constantinos Kardaras, *Finitely additive probabilities and the fundamental theorem of asset pricing*, Contemporary quantitative finance. Essays in honour of Eckhard Platen., Springer, 2010, pp. 19–34.

[Kar10b] ———, *On the stochastic behavior of optional processes up to random times*, arXiv preprint arXiv:1007.1124 (2010).

[Kar12] ———, *Market viability via absence of arbitrage of the first kind*, Finance Stoch. **16** (2012), no. 4, 651–667.

[KK07] Ioannis Karatzas and Constantinos Kardaras, *The numéraire portfolio in semimartingale financial models*, Finance Stoch. **11** (2007), no. 4, 447–493.

[KKN11] Constantinos Kardaras, Dörte Kreher, and Ashkan Nikeghbali, *Strict local martingales and bubbles*, arXiv preprint arXiv:1108.4177v1 (2011), 1–34.

[Kni69] Frank B. Knight, *Brownian local times and taboo processes*, Trans. Amer. Math. Soc. **143** (1969), 173–185.

[KP11] Constantinos Kardaras and Eckhard Platen, *On the semimartingale property of discounted asset-price processes*, Stochastic Process. Appl. **121** (2011), no. 11, 2678–2691.

[KPZ86] Mehran Kardar, Giorgio Parisi, and Yi-Cheng Zhang, *Dynamic scaling of growing interfaces*, Physical Review Letters **56** (1986), no. 9, 889–892.

[KS88] Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, Springer, 1988.

[KS99] Dmitry Kramkov and Walter Schachermayer, *The asymptotic elasticity of utility functions and optimal investment in incomplete markets*, Ann. Appl. Probab. **9** (1999), no. 3, 904–950.

[Kun76] Hiroshi Kunita, *Absolute continuity of Markov processes*, Séminaire de Probabilités X, Springer, 1976, pp. 44–77.

[LCL07] Terry J. Lyons, Michael Caruana, and Thierry Lévy, *Differential equations driven by rough paths*, Lecture Notes in Mathematics, vol. 1908, Springer, Berlin, 2007.

[LQ96] Terry J. Lyons and Zhongmin Qian, *Calculus for multiplicative functionals, Itô's formula and differential equations*, Itô's stochastic calculus and probability theory, Springer, Tokyo, 1996, pp. 233–250.

[LQ97] Terry Lyons and Zhongmin Qian, *Flow equations on spaces of rough paths*, J. Funct. Anal. **149** (1997), no. 1, 135–159.

[LQ02] ———, *System control and rough paths*, Oxford University Press, 2002.

[LR09] Sergey V. Lototsky and Boris L. Rozovskii, *Stochastic partial differential equations driven by purely spatial noise*, SIAM J. Math. Anal. **41** (2009), no. 4, 1295–1322.

[LW00] Mark Loewenstein and Gregory A. Willard, *Local martingales, arbitrage, and viability - Free snacks and cheap thrills*, Econom. Theory **161** (2000), 135–161.

[Lyo95] Terry J. Lyons, *Uncertain volatility and the risk-free synthesis of derivatives*, Appl. Math. Finance **2** (1995), no. 2, 117–133.

[Lyo98] ———, *Differential equations driven by rough signals*, Rev. Mat. Iberoam. **14** (1998), no. 2, 215–310.

[LZ99] Terry Lyons and Ofer Zeitouni, *Conditional exponential moments for iterated Wiener integrals*, Ann. Probab. **27** (1999), no. 4, 1738–1749.

[LŽ07] Kasper Larsen and Gordan Žitković, *Stability of utility-maximization in incomplete markets.*, Stochastic Processes Appl. **117** (2007), no. 11, 1642–1662.

[McK63] Henry P. McKean, *Excursions of a non-singular diffusion*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **1** (1963), 230–239.

[Mer73] Robert C. Merton, *Theory of rational option pricing*, Bell J. Econom. and Management Sci. **4** (1973), 141–183.

[Mey72] Paul A. Meyer, *La mesure de H. Föllmer en théorie des surmartingales*, Séminaire de Probabilités VI, Springer, 1972, pp. 118–129.

[NT11] David Nualart and Samy Tindel, *A construction of the rough path above fractional Brownian motion using Volterra's representation*, The Annals of Probability **39** (2011), no. 3, 1061–1096.

[Par67] Kalyanapuram R. Parthasarathy, *Probability measures on metric spaces*, Probability and Mathematical Statistics, No. 3, Academic Press Inc., New York, 1967.

[Pit75] James W. Pitman, *One-dimensional Brownian motion and the three-dimensional Bessel process*, Adv. in Appl. Probab. **7** (1975), no. 3, 511–526.

[PP10] Soumik Pal and Philip Protter, *Analysis of continuous strict local martingales via h-transforms*, Stochastic Process. Appl. **120** (2010), no. 8, 1424–1443.

[PP13] Nicolas Perkowski and David J. Prömel, *Pathwise stochastic integrals for model free finance*, in preparation, 2013.

[PR12] Nicolas Perkowski and Johannes Ruf, *Conditioned martingales*, Electron. Commun. Probab. **17** (2012), no. 48, 12.

[PR13] ———, *Supermartingales as Radon-Nikodym Derivatives and Related Measure Extensions*, in preparation, 2013.

[Pro04] Philip E. Protter, *Stochastic integration and differential equations*, 2nd ed., Springer, 2004.

[PT11] Giovanni Peccati and Murad S. Taqqu, *Wiener chaos: Moments, cumulants and diagrams. A survey with computer implementation.*, Springer, 2011.

[PY81] Jim Pitman and Marc Yor, *Bessel Processes and Infinitely Divisible Laws*, Stochastic integrals, Proc. LMS Durham Symp. 1980, Lect. Notes Math. 851, Springer, apr 1981, pp. 285–370.

[Rok10] Dmitry B. Rokhlin, *On the existence of an equivalent supermartingale density for a fork-convex family of random processes*, Math. Notes **87** (2010), no. 3–4, 556–563.

[Ros09] Mathieu Rosenbaum, *First order p-variations and Besov spaces*, Statist. Probab. Lett. **79** (2009), no. 1, 55–62.

[Roy93] Bernard Roynette, *Mouvement Brownien et espaces de Besov*, Stochastics Stochastics Rep. (1993), 221–260.

[Ruf13] Johannes Ruf, *Hedging under arbitrage*, Math. Finance **23** (2013), no. 2, 297–317.

[RY99] Daniel Revuz and Marc Yor, *Continuous martingales and Brownian motion*, 3rd ed., Springer, 1999.

[RY09] Bernard Roynette and Marc Yor, *Penalising Brownian Paths*, Springer, Berlin, 2009.

[Sch95] Martin Schweizer, *On the minimal martingale measure and the Föllmer-Schweizer decomposition*, Stoch. Anal. Appl. **13** (1995), 573–599.

[Son13] Shiqi Song, *An alternative proof of a result of Takaoka*, arXiv preprint arXiv:1306.1062 (2013).

[ST87] H.-J. Schmeisser and H. Triebel, *Topics in Fourier analysis and function spaces*, Mathematik und ihre Anwendungen in Physik und Technik [Mathematics and its Applications in Physics and Technology], vol. 42, Akademische Verlagsgesellschaft Geest & Portig K.-G., Leipzig, 1987.

[SV06] Daniel W. Stroock and S. R. Srinivasa Varadhan, *Multidimensional diffusion processes*, Classics in Mathematics, Springer, Berlin, 2006, Reprint of the 1997 edition.

[Tak13] Koichiro Takaoka, *On the condition of no unbounded profit with bounded risk*, Finance Stoch. (2013), to appear.

[Tei11] Josef Teichmann, *Another approach to some rough and stochastic partial differential equations*, Stoch. Dyn. **11** (2011), no. 2-3, 535–550.

[Tri06] Hans Triebel, *Theory of function spaces. III*, Monographs in Mathematics, vol. 100, Birkhäuser Verlag, Basel, 2006.

[Unt10a] Jérémie Unterberger, *A rough path over multidimensional fractional Brownian motion with arbitrary Hurst index by Fourier normal ordering*, Stochastic Processes and their Applications **120** (2010), no. 8, 1444–1472.

[Unt10b] ______, *Hölder-Continuous Rough Paths by Fourier Normal Ordering*, Comm. Math. Phys. **298** (2010), no. 1, 1–36.

[Vov08] Vladimir Vovk, *Continuous-time trading and the emergence of volatility*, Electron. Commun. Probab. **13** (2008), 319–324.

[Vov11] ______, *Ito calculus without probability in idealized financial markets*, arXiv preprint arXiv:1108.0799 (2011).

[Vov12] ______, *Continuous-time trading and the emergence of probability*, Finance Stoch. **16** (2012), no. 4, 561–609.

[Ž02] Gordan Žitković, *A filtered version of the bipolar theorem of Brannath and Schachermayer.*, J. Theor. Probab. **15** (2002), no. 1, 41–61.

[Ž10] ———, *Convex compactness and its applications*, Math. Financ. Econ. **3** (2010), no. 1, 1–12.

[Wil74] David Williams, *Path Decomposition and Continuity of Local Time for One-Dimensional Diffusions, I*, Proc. Lond. Math. Soc. (3) **28** (1974), no. 4, 738—768.

[Wue80] M. Wuermli, *Lokalzeiten für Martingale*, Master's thesis, Universität Bonn, 1980, supervised by Hans Föllmer.

[Yan80] Jia-An Yan, *Caractérisation d'une classe d'ensembles convexes de L^1 ou H^1*, Séminaire de Probabilités XIV, Springer, 1980, pp. 220–222.

[YM78] Marc Yor and Paul A. Meyer, *Sur l'extension d'un théorème de Doob à un noyau σ-fini d'après G. Mokobodzki*, Séminaire de Probabilités, XII (Univ. Strasbourg, Strasbourg, 1976/1977), Lecture Notes in Math., vol. 649, Springer, Berlin, 1978, pp. 482–488.

[Yoe85] Chantha Yoeurp, *Théorème de Girsanov généralisé et grossissement d'une filtration*, Grossissements de filtrations: exemples et applications; Lecture Notes in Mathematics Vol. 1118, Springer, 1985, pp. 172–196.

[You36] Laurence C. Young, *An inequality of the Hölder type, connected with Stieltjes integration*, Acta Math. **67** (1936), no. 1, 251–282.

List of Symbols

Δ_j a Littlewood-Paley block 134

Δ_p a dyadic Schauder block 89

$\mathcal{E}(X)$ the stochastic exponential of X 41

f_{pm} the m–th coefficient of f in dyadic generation number p 88

$f_{s,t}$ the increment $f(t) - f(s)$ 87

$\mathcal{F}u$ the Fourier transform of u 133

$\varphi(\mathrm{D})$ the pseudo-differential operator associated to φ; $\varphi(\mathrm{D})u = \mathcal{F}^{-1}(\varphi\mathcal{F})$ 134

φ_{pm} a rescaled Schauder function 88

$H \cdot S$ the stochastic integral of H against S 17

$\mathcal{H}_\lambda$ the set of λ–admissible strategies 17

$\mathcal{H}_{\lambda,s}$ the set of simple λ–admissible strategies 17, 73

$\mathcal{K}_1$ the set of terminal wealths under 1–admissible strategies 17

L the Lévy area 97

L^0 the space of random variables 17

$\pi_<$ the paraproduct: below diagonal part of the product 92, 136

$\pi_>$ the "opposite paraproduct": above diagonal part of the product 136

$\pi_\circ$ the diagonal part of the product 136

$\pi_\lessgtr$ the sum of $\pi_<$ and $\pi_>$ 137

$R(f, v, w)$ the "commutator" $\pi_\circ(\pi_<(f, v), w) - f\pi_\circ(v, w)$ 140

$\overline{R}(f, g, v, w)$ the "commutator" $\pi_\circ(\pi_<(f, v), \pi_<(g, w)) - fg\pi_\circ(v, w)$ 143

ρ_j the rescaled function from the partition of unity; $\rho_j = \rho(2^{-j}\cdot)$ and $\rho_{-1} = \chi$ 134

$\mathcal{S}$ the Schwartz space 133

$\mathcal{S}'$ the space of tempered distributions 133

$\mathcal{S}'(\mathbb{R}^d, \mathbb{R}^n)$ the space of tempered distributions on $\mathbb{R}^d$ with values in $\mathbb{R}^n$ 133

S_i the sum of the Littlewood-Paley blocks up to $i - 1$ 134

S_p the sum of the dyadic Schauder blocks up to p 89

$S^{\tau-}$ the process S stopped at $\tau-$ 19

S the symmetric part of the integral 97

$\mathbb{T}^d$ the d–dimensional torus $[-\pi, \pi]^d$ 131

t^i_{pm} the dyadic points 87

$\widehat{u}$ the Fourier transform of u 133

$\mathcal{W}_1$ the set of wealth processes under 1–admissible strategies 17

$\mathcal{W}_{1,s}$ the set of wealth processes under 1–admissible simple strategies 17

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 05.09.2013 Nicolas Simon Perkowski