Calibration of Non-Semimartingale Models - an Adjoint Approach

Dissertation zur Erlangung des Grades des
Doktors der Naturwissenschaften
der Fakultät Mathematik und Informatik
der Universität des Saarlandes

von

Matthias Thiel

Saarbrücken, 2023

Tag des Kolloquiums:	27.10.2023
Dekan:	Prof. Dr. Jürgen Steimle

Mitglieder des Prüfungsausschusses:

Vorsitzender:	Prof. Dr. Michael Bildhauer
1. Berichterstatter:	Prof. Dr. Christian Bender
2. Berichterstatter:	Prof. Dr. Andeas Neuenkirch
Protokollführer:	Dr. Johannes Hoffmann

Abstract

In this thesis we consider the calibration of a stochastic volatility model, where the volatility driving noise is given by a continuous process of finite p-variation, for $p \in (1, 2)$. This includes fractional Brownian motion with Hurst parameter $H \in (0.5, 1)$ as a relevant example for a driving noise of the volatility. The calibration will be done by the "optimize then discretize" approach, meaning that we first analyze the gradient of the cost function corresponding to our calibration problem in continuous time. We establish a new representation of the gradient, containing the solution of an anticipating backward stochastic differential equation, which we call the adjoint equation. The advantage of the adjoint equation lies in the fact that the dimension of the equation does not depend on the number of parameters of the model. This suggests that discretizing this equation and using it in a gradient-based Monte-Carlo optimization algorithm will significantly speed up the calibration in comparison to other methods, such as finite differences. We derive a suitable discretization scheme for the adjoint equation and establish the corresponding convergence rate. These theoretical results will then be used in a numerical case study, calibrating a fractional Heston-type model to observed option prices.

Zusammenfassung

In dieser Arbeit befassen wir uns mit der Kalibrierung eines stochastischen Volatilitätsmodells, dessen Volatilität von einem stetigen Prozess mit endlicher p-Variation, $p \in (1; 2)$, getrieben wird. Dies ermöglicht es, fraktionale Brownsche Bewegungen mit Hurst Parameter $H \in (0,5; 1)$ als relevante Beispiele für einen treibenden Prozess der Volatilität zu betrachten. Dabei gehen wir nach dem Ansatz "Optimieren-Dann-Diskretisieren" vor, d.h. wir analysieren zuerst den Gradienten der Kostenfunktion des zugrundeliegenden Kalibrierungsproblems in stetiger Zeit. Wir leiten eine neue Darstellung dieses Gradienten her, welche die Lösung eines Endwertproblems für eine antizipierende stochastische Differentialgleichung enthält. Diese Gleichung bezeichnen wir als die adjungierte Gleichung. Der Vorteil dieser Gleichung liegt darin, dass ihre Dimension unabhängig von der Anzahl der Parameter des Modells ist. Dies führt dazu, dass die Diskretisierung dieser Gleichung und ihre Verwendung in einem gradientenbasierten Monte-Carlo-Optimierungsalgorithmus die Laufzeit der Kalibrierung im Vergleich zu anderen Methoden, wie z.B. der Finite-Differenzen-Methode, signifikant verringert. Zusätzlich entwickeln wir ein passendes Diskretisierungsschema für die adjungierte Gleichung und bestimmen die zugehörige Konvergenzrate. Diese Resultate werden dann in einer numerischen Fallstudie verwendet, um ein fraktionales Heston-Modell an beobachtete Call-Options-Preise zu kalibrieren.

Acknowledgements

First and foremost, I want to thank my supervisor Prof. Dr. Christian Bender. Not only for giving me the opportunity to write this thesis and work at the Stochastics group of Saarland university, but for creating such an encouraging and pleasant working environment. It was a great pleasure and honor to work for and with him.

I also would like to thank Prof. Dr. Andreas Neuenkirch that he kindly agreed to review this thesis.

Furthermore, I want to thank all my current and former colleagues at the Stochastics group for their motivating words, support and all the fun conversations we had through these years. In particular I want to thank Steffen Meyer, Patrick Kern, Ulrike Mayer, Christian Gärtner, Thuan Nguyen, Frederik Herzberg and Christine Kilz.

Finally, I want to express my deepest gratitude towards my family, friends and especially my wife Saskia. Without your patience, support and encouragement during the last years, this thesis would not have been possible.

Contents

Introduction

In the area of financial engineering, continuous stochastic volatility models enjoy great popularity among practitioners. The term ”stochastic volatility” refers to the assumption that the volatility of an asset is itself given by a stochastic process. These models are able to replicate realistic traits of observed implied volatilities, such as volatility smiles or the leverage effect. One of the most famous models of this kind is the Heston model, introduced in Heston [1993]. In this model the volatility of the asset follows a Cox-Ingersoll-Ross process and the empirically observed negative correlation between the volatility and the asset returns (see, e.g. French et al. [1987]) is incorporated in this model by a parameter which directly influences the correlation between the driving Brownian motions. The Heston model is especially advantageous for practitioners, because of its semi analytic pricing formula for European options. The asset and volatility in this model and most of the famous stochastic volatility models in the 20th century, e.g. Hull and White [1987], Chesney and Scott [1989], Stein and Stein [2015], are modeled by stochastic differential equations governed by Brownian motion noises. Researchers and practitioners also consider another class of stochastic volatility models, where the driving noise of the volatility is given by a fractional Brownian motion with Hurst parameter $H \in (0,1)$ (see Mandelbrot and Ness [1968]), which is not a semimartingale for $H \neq 0.5$. These models are called fractional volatility models for $H > 0.5$, see e.g. Comte and Renault [1998], Chronopoulou and Viens [2012], Bezborodov et al. [2019], Mishura and Yurchenko-Tytarenko [2020], Lépinette and Mehrdoust [2016] or, inspired by the empirical findings of Gatheral et al. [2018] (preprint available since 2014), rough volatility models for $H < 0.5$, see Bayer et al. [2016], Fukasawa [2017], El Euch and Rosenbaum [2019]. For $H > 0.5$ the increments of the fBM are positively correlated and their autocorrelation function decays very slowly, such that these models are often refered to as long memory models. The term rough volatility stems from the roughness of the paths of the fractional Brownian motion for $H < 0.5$. In this thesis we consider a fractional stochastic volatility model in the ”long memory” case, but instead of working only with a fBM with Hurst parameter $H > 0.5$ as governing noise for our volatility, we consider a whole class of processes which contains the fBM, namely continuous processes of finite p-variation for $p \in (1,2)$. L.C. Young introduced in Young [1936] a generalization of the Riemann Stieljes integral based on the finite p-variation of integrand and integrator, which enables us to consider the underlying stochastic differential equation for the volatility pathwise. For the asset dynamics in our model we keep the Brownian motion as governing noise and allow the two processes to be correlated. This way we are able to avoid arbitrage problems, which arise when using the fractional Brownian motion in the asset dynamics, see e.g. Rogers [1997].

When it comes to stochastic volatility models, calibrating the parameters of the model to fit observed option prices is a key task for practitioners. The literature on calibration of fractional volatility models to observed option prices is rather scarce and only specific models are considered, see Mehrdoust and Fallah [2022], Mrázek et al. [2016]. Inspired by results from the calibration of standard stochastic volatility models, see Käbe et al. [2009], we aim to calibrate our fractional stochastic volatility model to observed option prices using a gradient-based Monte-Carlo optimization algorithm. A drawback of Monte-Carlo methods is their relatively slow convergence of order $\frac{1}{2}$, but combining it with a fast gradient-based local optimization algorithm can still produce satisfactory results. However, to use gradient-based optimization routines, one needs to calculate the gradient of the corresponding cost function with respect to the model parameters. This is usually done by finite differences, but especially in the Monte-Carlo setting, this approach becomes costly and leads to instabilities. The authors in Käbe et al. [2009] tackle this problem by using adjoint methods in a discretized setting to efficiently calculate this gradient, and show that it significantly speeds up the calibration of the Heston model, compared to the finite differences approach. Adjoint techniques are well known to speed up the numerical calculation of sensitivities, by solving corresponding backwards equations and found use in multiple disciplines like optimal design Giles and Pierce [2000], meteorology Charpentier and Ghemires [2000] or in financial modelling, to efficiently calculate option Greeks Giles and Glasserman [2006]. The paper Käbe et al. [2009] inspired us to look at adjoint techniques in the context of calibrating a fractional volatility model to observed option prices. Conceptually we will follow a different approach as the atuhors in Käbe et al. [2009], by optimizing the underlying cost function in the continuous setting, establishing a continuous representation of the gradient, which we then discretize to apply this representation in practice.

As we already mentioned, the stochastic differential equations underlying our model dynamics are of different nature. So we will first analyze the model equation itself and find sufficient conditions on the coefficients to ensure the existence and uniqueness of the solution and also the Frechét differentiability of the solution mapping with respect to the parameter, this will be the content of Chapter 2. We analyze the two equations successively. Starting with the fractional volatility SDE, we first give an introduction to p-variation spaces and the Young integral, as it is not a commonly known topic. Then using results from Nguyen et al. [2018] and Nguyen et al. [2020] on time dependent, multidimensional Young differential equations in the deterministic setting, we establish the required properties of the solution to the volatility equation. The existence, uniqueness and differentiability results concerning the asset dynamics SDE can be derived by standard results on Itô SDEs.

In Chapter 3 we introduce the cost function corresponding to our calibration problem. We find two ways of calculating the gradient of the cost function. One using the Frechét derivative of the model solution mapping, which solves an inhomogenous linear forward SDE. We call this equation the sensitivity equation. The second representation of this gradient will be our main result. By considering the forward integral by Russo and Vallois Russo and Vallois [1993a], which generalizes the pathwise Young and the Itô integral in our setting, we will be able to find a variation of constants formula for the explicit solution of the sensitivity equation. Furthermore using the forward integral and this explicit solution, we establish the second representation of the gradient, containing the solution to an anticipating backward stochastic differential equation, which we call the adjoint equation.

Having established the representations of the gradient of our cost function, we need to find suitable discretizations of the solutions to the underlying equations, to use our results in practice. This will be the content of Chapter 4. For the forward equations, namely the model dynamics equation and the linear sensitivity equation, we choose first order Euler schemes and prove that the discretizations converge to the solutions in any L^l-space, uniformly in time, establishing the corresponding convergence rate. For the discretization of the backwards adjoint equation, we derive a suitable backward Euler scheme and prove its convergence to the solution of the adjoint equation, again deriving the convergence rate. These schemes are then used to prove the convergence of the discretized cost function and its discretized gradient, inheriting the convergence rates we previously found.

Chapter 1 provides a summary over all the results we establish in this thesis. Since the introduction of preliminaries and the calculations needed to prove the given results are very extensive, it seems beneficial to start with such a chapter to deliver the essence of this thesis to the interested reader. We first give a a detailed description of the mathematical problems treated in this thesis and then summarize the results of Chapters 2 to 4 in the Sections 1.2 to 1.4. After that, these results will be applied in a numerical case study, where we first translate our discretization results to the Monte-Carlo setting and then efficiently calibrate a fractional Heston-type model to observed option prices using standard gradient-based optimization algorithms, contained in the Matlab optimization toolbox. We will show that the adjoint approach leads to a significant improvement of the computational time, compared to the use of the sensitivity equation or finite differences. At the end of this chapter, we give a detailed overview over the existing literature in the related fields of research.

Chapter 1

Overview on main results, numerical example and literature review

1.1 Problem formulation and setting

The goal of this thesis is to find an adjoint representation of the gradient of a cost function, evaluated at values of an asset whose dynamics are driven by fractional stochastic volatility model. This adjoint representation will then be used to efficiently calibrate the model to a given set of observed call option prices, using a gradient-based optimization algorithm. By "fractional" we mean that our volatility process will be modeled by a driving process w, which paths are almost surely continuous and of bounded p-variation for $p \in (1,2)$ (see Subsection 2.1.1, for the definition of the p-variation norm). This excludes the standard Brownian motion as driving process for the volatility as used in usual stochastic volatility models, like the Heston model. On the other hand the process driving the asset dynamics in our model will be a standard Brownian motion B. Since we want our model to be able to capture realistic smile behaviors of implied volatilities, we want the driving process w and B to be correlated. Now we introduce our model setting.

Let T be a positive constant and $n_1, m_1, n_2, m_2, d \in \mathbb{N} = \{1, 2, \dots\}$. Let $(\Omega, \mathcal{F}, \mathbb{F}, P)$ be a filtered probability space (satisfying the usual conditions) carrying an m_1-dimensional stochastic process $(w_t)_{t\in[0,T]}$, which paths are almost surely continuous and have finite p-variation for $p \in (1,2)$ and a m_2-dimensional standard Brownian motion $(B_t)_{t\in[0,T]}$, both adapted to the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$, possibly dependent. Furthermore let $\mathcal{U}$ be an open, convex and bounded subset of $\mathbb{R}^d$, which will be our parameter set. We consider the parameter dependent system of stochastic differential equations

$$\xi_t^u = \xi_0(u) + \int_0^t b(r, \xi_r^u, u)\, dr + \sum_{j=1}^{m_1} \int_0^t \sigma^j(r, \xi_r^u, u)\, dw_r^j, \tag{1.1}$$

$$x_t^u = x_0(u) + \int_0^t \hat{b}(r, x_r^u, \xi_r^u, u)\, dr + \sum_{j=1}^{m_2} \int_0^t \hat{\sigma}^j(r, x_r^u, \xi_r^u, u)\, dB_r^j, \tag{1.2}$$

where

$$\begin{aligned}
&\xi_0 : \mathcal{U} \to \mathbb{R}^{n_1}, \\
&b : [0,T] \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_1}, \\
&\sigma = (\sigma^1, \ldots, \sigma^{m_1}) : [0,T] \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_1 \times m_1}
\end{aligned}$$

and

$$\begin{aligned}
&x_0 : \mathcal{U} \to \mathbb{R}^{n_2}, \\
&\hat{b} : [0,T] \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_2}, \\
&\hat{\sigma} = (\hat{\sigma}^1, \ldots, \hat{\sigma}^{m_2}) : [0,T] \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_2 \times m_2},
\end{aligned}$$

denoted in matrix form by

$$\begin{aligned}
\mathcal{X}_t^u = \begin{pmatrix} \xi_t^u \\ x_t^u \end{pmatrix} = \begin{pmatrix} \xi_0(u) \\ x_0(u) \end{pmatrix} + \int_0^t \begin{pmatrix} b(r, \xi_r^u, u) \\ \hat{b}(r, x_r^u, \xi_r^u, u) \end{pmatrix} dr + \sum_{j=1}^{m_1} \int_0^t \begin{pmatrix} \sigma^j(r, \xi_r^u, u) \\ 0 \end{pmatrix} dw_r^j \\
+ \sum_{j=1}^{m_2} \int_0^t \begin{pmatrix} 0 \\ \hat{\sigma}^j(r, x_r^u, \xi_r^u, u) \end{pmatrix} dB_r^j.
\end{aligned} \tag{1.3}$$

Here the stochastic integral with respect to w is given by a pathwise Young integral (for details, see Subsection 2.1.1) and the stochastic integral with respect to B is a standard Itô integral. We denote by $|\cdot|$ the Frobenius norm on $\mathbb{R}^{n \times m}$, $n, m \in \mathbb{N}$ and define $\|x\|_{\infty,0,T} = \sup_{t \in [0,T]} |x_t|$ for every continuous function taking values in $\mathbb{R}^{n \times m}$. Under specific conditions on the coefficient functions and the initial value functions, we are able to formulate an existence and uniqueness result for equation (1.3) and get for every parameter $u \in \mathcal{U}$ a unique solution $\mathcal{X}^u$, which is an element of $L^l_{\mathbb{F}}(\Omega, C[0,T], \mathbb{R}^{n_1+n_2})$ for every $l \geq 1$, where

$$\begin{aligned}
&L^l_{\mathbb{F}}(\Omega, C[0,T], \mathbb{R}^{n_1+n_2}) \\
&:= \{x : \Omega \times [0,T] \to \mathbb{R}^{n_1+n_2} |\ x \text{ is } \mathbb{F}\text{-adapted process with almost surely continuous paths} \\
&\qquad \text{such that } \mathrm{E}[\|x\|^l_{\infty,0,T}] < \infty\}.
\end{aligned}$$

Moreover, we can prove that the solution mapping $u \mapsto \mathcal{X}^u$ is Fréchet differentiable as a map from $\mathbb{R}^d$ to any $L^l_{\mathbb{F}}(\Omega, C[0,T], \mathbb{R}^{n_1+n_2})$ for $l \geq 1$. The Fréchet derivative is given by a process $\mathcal{Y}^u \in L^l_{\mathbb{F}}(\Omega, C[0,T], \mathbb{R}^{(n_1+n_2)\times d})$, which is the unique solution of a corresponding system of inhomogenous linear stochastic differential equations. Since we want to calibrate our model to e.g.

prices of European call options, we define a general cost function

$$J(u) = \frac{1}{2}\sum_{\mu=1}^{M} \mathrm{E}\left[g_\mu(\mathcal{X}^u_{T_\mu})\right]^2,$$

where $0 < T_1 \leq \ldots, T_M = T$ is a sequence of times (e.g. the maturities of the observed options) and the functions

$$g_\mu : \mathbb{R}^{n_1+n_2} \to \mathbb{R}$$

satisfy conditions which ensure the integrability of the composition and the differentiability of the cost function $J : \mathcal{U} \to \mathbb{R}$. The goal is to minimize this cost function over all parameters $u \in \mathcal{U}$. To do this we follow the "optimize then discretize" approach and first calculate the gradient of J. By the results we mentioned, we could already calculate this gradient, using the chain rule for Fréchet derivatives and get

$$\nabla J(u) = \sum_{\mu=1}^{M} \mathrm{E}\left[g_\mu(\mathcal{X}^u_{T_\mu})\right] \mathrm{E}\left[g'(\mathcal{X}^u_{T_\mu})\mathcal{Y}^u_{T_\mu}\right]. \tag{1.4}$$

To approximate this term, we discretize the underlying equations using first order Euler schemes and show that the discretized gradient converges to ∇J for a sequence of partitions converging in mesh to zero, where we also give the convergence rate. Then using the Monte-Carlo approach we will be able to estimate the gradient. Focusing on the computational side, the computation of the discretized gradient boils down to numerically evaluating the values of the Euler scheme for $\mathcal{Y}^u$ on a partition $(t_i)_{i=0,\ldots,n}$ of $[0, T]$ for every Monte-Carlo path. Since $\mathcal{Y}^u$ takes values in $\mathbb{R}^{(n_1+n_2)\times d}$ this leads to very high computational costs, especially if the number of parameters d is very high, e.g. when the parameters are time dependent. The main goal of this thesis lies in the reduction of these computational costs by establishing a new representation of the gradient ∇J which does not involve the process $\mathcal{Y}^u$, but a process Λ^u taking values in $\mathbb{R}^{n_1+n_2}$, which solves an anticipating backward stochastic differential equation. Expressing the gradient of a cost function, utilizing the solution of a backward SDE, can be seen as a stochastic analogon to the adjoint sensitivity method in the ODE case, which has many applications in various fields of research (see Subsection 1.6, for a short overview). For that reason we call our anticipating backward SDE, the adjoint equation. The derivation of the adjoint equation will be done by expressing $\mathcal{Y}^u_{T_\mu}$ (which is the solution to an inhomogenous linear system of SDEs) by a variation of constants formula and reformulating (1.4). To establish this variation of constants formula and to deal with the $\mathcal{F}_T$ measurable (hence anticipating) random variables $\mathrm{E}\left[g_\mu(\mathcal{X}^u_{T_\mu})\right] g'(\mathcal{X}^u_{T_\mu})$, we will make use of the forward integral, introduced by Russo and Vallois [1993a]. This stochastic integral generalizes the Young and the Itô integral and allows for anticipating integrands. We will then be able to introduce the adjoint process Λ^u as the explicit solution to the adjoint equation.

This process Λ^u can then be approximated using a suitable discretization scheme running backwards in time, where we again establish the corresponding convergence rate. The numerical estimation of this new gradient now boils down to calculating the values of the discretization scheme for Λ^u on a partition $(t_i)_{i=0,\dots,n}$ of $[0,T]$ for every Monte-Carlo path. Since Λ^u takes values in $\mathbb{R}^{(n_1+n_2)}$ (as opposed to $\mathbb{R}^{(n_1+n_2)\times d}$) this reduces the computing time in comparison to the first mentioned approach. Especially in the case where the parameters are time dependent, this reduction is significant. In Subsection 1.5.1, we show the applicability of our theoretical results to calibrate a fractional Heston-type model to observed option prices using gradient-based optimization algorithms, contained in the Matlab optimization toolbox.

1.2 The model dynamics equation and its differentiability with respect to the parameter

The form of the model dynamics equation (1.1) has a key property, namely that the first equation (1.1) does not contain the solution process of the second one (1.2). This makes it possible to split the analysis of the given equations. We start with the treatment of the first equation, which is driven by a continuous process of finite p-variation. This allows us to interpret the dw_t integral as a pathwise Young integral and we call such an equation a stochastic Young differential equation. The properties of functions of finite p-variation, an introduction to the Young integral, existence and uniqueness of equation (1.1) and the Fréchet differentiability of the corresponding solution mapping are contained in Section 2.1. The results we establish in this section heavily rely on the results of Nguyen et al. [2018], where the authors consider non-autonomous Young differential equations in the deterministic setting. The second equation is a standard stochastic differential equation in the Itô sense. The equation has random coefficients since $\hat{b}$ and $\hat{\sigma}$ depend on the process ξ^u. The analysis of the properties of this equation are the topic of Section 2.2. This kind of equation often appears in stochastic control theory, so we cite and use in this section the results of Yong [2019] and Yong and Zhou [1999]. Note that there are various existing results on the existence and uniqueness of equations that are similar to (1.1), respectively (1.2). There are also related results on the Fréchet differentiability of the solution mappings, which could be adapted to our setting, but impose stronger conditions in the time domain of the coefficients. We will focus on this matter in Subsection 1.6. Now we summarize the important results of Chapter 2. Suppose we are in the setting introduced in Section 1.1. Before we state the results, we summarize all the conditions we impose on the coefficient functions, the initial value functions and the driving process w, for the rest of this chapter. Additionally to its path properties, we assume that the process w satisfies the following integrability condition

- Exponential moment condition: There exists $K > 0$ such that

$$E\left[e^{K\|w\|_{p,0,T}^2}\right] < \infty. \tag{1.5}$$

For the coefficients and the initial value function of equation (1.1), we assume:

(H_1) Let $\xi_0 : \mathcal{U} \to \mathbb{R}^{n_1}$ be continuously differentiable, such that ξ_0 and its Jacobian $D\xi_0$ are bounded by a constant L.

(H_2) Let $b : [0,T] \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_1}$ be a continuous function which satisfies:

- $b(t,x,u)$ is continuously differentiable with respect to x and u.
- There exists a constant L such that for all $x, y \in \mathbb{R}^{n_1}$, $u, v \in \mathcal{U}$ and every $t \in [0,T]$

$$|b(t,x,u)| \leq L$$

$$|b_x(t,x,u)| + |b_u(t,x,u)| \leq L$$

$$|b_x(t,x,u) - b_x(t,y,v)| + |b_u(t,x,u) - b_u(t,y,v)| \leq L(|x-y| + |u-v|).$$

(H_3) Let $\sigma := (\sigma^1, \dots, \sigma^{m_1}) : [0,T] \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_1 \times m_1}$ be a continuous function which satisfies:

- $\sigma(t,x,u)$ is twice continuously differentiable with respect to x and u.
- There exists a constant L, such that for all $x \in \mathbb{R}^n$, $u \in \mathcal{U}$, $t \in [0,T]$ and $j = 1, \dots, m_1$, $l = 1, \dots, n_1$, $k = 1, \dots, d$

$$\begin{aligned}
&|\sigma(t,x,u)| \leq L \\
&|\sigma_x^j(t,x,u)| + |\sigma_u^j(t,x,u)| \leq L \\
&\left|\frac{\partial}{\partial x}\sigma_{x_l}^j(t,x,u)\right| + \left|\frac{\partial}{\partial x}\sigma_{u_k}^j(t,x,u)\right| \leq L \\
&\left|\frac{\partial}{\partial u}\sigma_{x_l}^j(t,x,u)\right| + \left|\frac{\partial}{\partial u}\sigma_{u_k}^j(t,x,u)\right| \leq L,
\end{aligned}$$

where

$$\sigma_x^j(t,x,u) = \begin{pmatrix} \frac{\partial}{\partial x_1}\sigma^{1,j}(t,x,u) & \dots & \frac{\partial}{\partial x_{n_1}}\sigma^{1,j}(t,x,u) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1}\sigma^{n_1,j}(t,x,u) & \dots & \frac{\partial}{\partial x_{n_1}}\sigma^{n_1,j}(t,x,u) \end{pmatrix}$$

$$\sigma_u^j(t,x,u) = \begin{pmatrix} \frac{\partial}{\partial u_1}\sigma^{1,j}(t,x,u) & \dots & \frac{\partial}{\partial u_d}\sigma^{1,j}(t,x,u) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial u_1}\sigma^{n_1,j}(t,x,u) & \dots & \frac{\partial}{\partial u_d}\sigma^{n_1,j}(t,x,u) \end{pmatrix}$$

$$
\frac{\partial}{\partial x}\sigma^j_{x_l}(t,x,u) = \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_l}\sigma^{1,j}(t,x,u) & \dots & \frac{\partial^2}{\partial x_{n_1} \partial x_l}\sigma^{1,j}(t,x,u) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_1 \partial x_l}\sigma^{n_1,j}(t,x,u) & \dots & \frac{\partial^2}{\partial x_{n_1} \partial x_l}\sigma^{n_1,j}(t,x,u) \end{pmatrix}
$$

$$
\frac{\partial}{\partial u}\sigma^j_{u_k}(t,x,u) = \begin{pmatrix} \frac{\partial^2}{\partial u_1 \partial u_k}\sigma^{1,j}(t,x,u) & \dots & \frac{\partial^2}{\partial u_d \partial u_k}\sigma^{1,j}(t,x,u) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial u_1 \partial u_k}\sigma^{n_1,j}(t,x,u) & \dots & \frac{\partial^2}{\partial u_d \partial u_k}\sigma^{n_1,j}(t,x,u) \end{pmatrix}
$$

and analogously defined $\frac{\partial}{\partial x}\sigma^j_{u_k}$, $\frac{\partial}{\partial u}\sigma^j_{x_l}$.

– There exist constants L and $\beta \in [\frac{1}{2}, 1]$ such that for all $x, y \in \mathbb{R}^{n_1}$, $u, v \in \mathcal{U}$, $s \le t \in [0,T]$ and $j = 1, \dots, m_1$, $l = 1, \dots, n_1$, $k = 1, \dots, d$

$$
\begin{aligned}
&|\sigma(t,x,u) - \sigma(s,x,u)| \le L|t-s|^\beta \\
&|\sigma^j_x(t,x,u) - \sigma^j_x(s,x,u)| + |\sigma^j_u(t,x,u) - \sigma^j_u(s,x,u)| \le L|t-s|^\beta \\
&\left|\frac{\partial}{\partial x}\sigma^j_{x_l}(t,x,u) - \frac{\partial}{\partial x}\sigma^j_{x_l}(s,y,v)\right| + \left|\frac{\partial}{\partial u}\sigma^j_{x_l}(t,x,u) - \frac{\partial}{\partial u}\sigma^j_{x_l}(s,y,v)\right| \\
&+ \left|\frac{\partial}{\partial x}\sigma^j_{u_k}(t,x,u) - \frac{\partial}{\partial x}\sigma^j_{u_k}(s,y,v)\right| + \left|\frac{\partial}{\partial u}\sigma^j_{u_k}(t,x,u) - \frac{\partial}{\partial u}\sigma^j_{u_k}(s,y,v)\right| \\
&\le L\left(|t-s|^\beta + |x-y| + |u-v|\right).
\end{aligned}
$$

And for the initial value function and the coefficients of equation (1.2), we assume:

(B_1) The function $\hat{b} : [0,T] \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_2}$ is continuous with respect to the variables t, x, z and u and continuously differentiable with respect to x, z and u for all $t \in [0,T]$. Denote

$$
\hat{b}_x(t,x,z,u) = \left(\frac{\partial \hat{b}_i(t,x,z,u)}{\partial x_j}\right)_{1 \le i,j \le n_2}, \quad \hat{b}_z(t,x,z,u) = \left(\frac{\partial \hat{b}_i(t,x,z,u)}{\partial z_j}\right)_{1 \le i \le n_2, 1 \le j \le n_1}
$$

$$
\hat{b}_u(t,x,z,u) = \left(\frac{\partial \hat{b}_i(t,x,z,u)}{\partial u_j}\right)_{1 \le i \le n_2, 1 \le j \le d}.
$$

Furthermore there exists a constant $L > 0$ such that

$$
\sup_{t \in [0,T], x \in \mathbb{R}^{n_2}, z \in \mathbb{R}^{n_1}, u \in \mathcal{U}} |\hat{b}_x(t,x,z,u)| + |\hat{b}_z(t,x,z,u)| + |\hat{b}_u(t,x,z,u)| \le L.
$$

(B_2) The function $\hat{\sigma} = (\hat{\sigma}^1, \dots, \hat{\sigma}^k) : [0,T] \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_2 \times m_2}$ is continuous with respect to the variables t, x, z and u and continuously differentiable with respect to x, z and u for all $t \in [0,T]$. Denote for $j = 1, \dots, m_2$

$$\hat{\sigma}^j_x(t,x,z,u) = \left(\frac{\partial \hat{\sigma}^j_{i_1}(t,x,z,u)}{\partial x_{i_2}}\right)_{1\leq i_1,i_2\leq n_2}, \hat{\sigma}^j_z(t,x,z,u) = \left(\frac{\partial \hat{\sigma}^j_{i_1}(t,x,z,u)}{\partial z_{i_2}}\right)_{1\leq i_1\leq n_2,\ 1\leq i_2\leq n_1}$$

$$\hat{\sigma}^j_u(t,x,z,u) = \left(\frac{\partial \hat{\sigma}^j_{i_1}(t,x,z,u)}{\partial u_{i_2}}\right)_{1\leq i_1\leq n_2, 1\leq i_2\leq d}.$$

Furthermore there exists a constant $L > 0$ such that for $j = 1, \dots, m_2$

$$\sup_{t\in[0,T], x\in\mathbb{R}^{n_2}, \mathbb{R}^{n_1}, u\in\mathcal{U}} |\hat{\sigma}^j_x(t,x,z,u)| + |\hat{\sigma}^j_z(t,x,z,u)| + |\hat{\sigma}^j_u(t,x,z,u)| \leq L.$$

(B_3) Let $x_0 : \mathcal{U} \to \mathbb{R}^{n_2}$ be a continuously differentiable deterministic function, such that x_0 and its Jacobian Dx_0 are bounded by the constant L.

Note that the conditions, we impose on the coefficients are stronger than those needed to establish the following results, since they are needed to show the differentiability of the solution mapping and not just existence and uniqueness.

Theorem 1.1. *For every $u \in \mathcal{U}$ the equation (1.3) has a unique solution $\mathcal{X}^u$, which is an element of $L^l_{\mathbb{F}}(\Omega, C[0,T], \mathbb{R}^{n_1+n_2})$ for every $l \geq 1$ and*

$$\mathrm{E}\left[\|\mathcal{X}^u\|^l_{\infty,0,T}\right] \leq D_{\mathcal{X},l},$$

where $D_{\mathcal{X},l}$ is a positive constant independent of the parameter u.

Proof. See Corollary 2.36, Corollary 2.37, Lemma 2.41 and Remark 2.47. □

As one would expect the Fréchet derivative of the solution mapping of the equations (1.1) and (1.2) solve corresponding inhomogenous linear equations of respective kind. Define for $t \in [0,T]$ the $\mathbb{R}^{(n_1+n_2)\times d}$ valued system of linear equations

$$y^u_t = D\xi_0(u) + \int_0^t b_x(r,\xi^u_r,u)y^u_r + b_u(r,\xi^u_r,u)\,dr + \sum_{j=1}^{m_1}\int_0^t \sigma^j_x(r,\xi^u_r,u)y^u_r + \sigma^j_u(r,\xi^u_r,u)\,dw^j_r \quad (1.6)$$

$$\begin{aligned}\hat{y}^u_t &= Dx_0(u) + \int_0^t \hat{b}_x(r,x^u_r,\xi^u_r,u)\hat{y}^u_r + \hat{b}_z(r,x^u_r,\xi^u_r,u)y^u_r + b^u(r,x^u_r,\xi^u_r,u)\,dr \\ &\quad + \sum_{j=1}^{m_2}\int_0^t \hat{\sigma}^j_x(r,x^u_r,\xi^u_r,u)\hat{y}^u_r + \hat{\sigma}^j_z(r,x^u_r,\xi^u_r,u)y^u_r + \hat{\sigma}^j_u(r,x^u_r,\xi^u_r,u)\,dB^j_r \quad (1.7)\end{aligned}$$

and in matrix form

$$\mathcal{Y}^u_t = \begin{pmatrix} y^u_t \\ \hat{y}^u_t \end{pmatrix}$$

$$
\begin{aligned}
&= \begin{pmatrix} D\xi_0(u) \\ Dx_0(u) \end{pmatrix} + \int_0^t \begin{pmatrix} b_x(r,\xi_r^u,u) & 0 \\ \hat{b}_z(r,x_r^u,\xi_r^u,u) & \hat{b}_x(r,x_r^u,\xi_r^u,u) \end{pmatrix} \mathcal{Y}_r^u + \begin{pmatrix} b_u(r,\xi_r^u,u) \\ \hat{b}_u(r,x_r^u,\xi_r^u,u) \end{pmatrix} dr \\
&\quad + \sum_{j=1}^{m_1} \int_0^t \begin{pmatrix} \sigma_x^j(r,\xi_r^u,u) \\ 0 \end{pmatrix} \mathcal{Y}_r^u + \begin{pmatrix} \sigma_u^j(r,\xi_r^u,u) \\ 0 \end{pmatrix} dw_r^j \\
&\quad + \sum_{j=1}^{m_2} \int_0^t \begin{pmatrix} 0 \\ \hat{\sigma}_z^u(r,x_r^u,\xi_r^u,u) & \hat{\sigma}_x^j(r,x_r^u,\xi_r^u,u) \end{pmatrix} \mathcal{Y}_r^u + \begin{pmatrix} 0 \\ \hat{\sigma}_u^j(r,x_r^u,\xi_r^u,u) \end{pmatrix} dB_r^j. \qquad (1.8)
\end{aligned}
$$

To prove the Fréchet differentiability of the solution mapping $u \mapsto \mathcal{X}^u$, we first need to establish the existence and uniqueness of the inhomogenous linear equations (1.6) and (1.7). For equation (1.6) we use an existence and uniqueness result concerning vector valued homogenous linear YDEs from Nguyen et al. [2020] and establish an explicit solution to (1.6), stating a variation of constants formula. Uniqueness then follows easily. We get a solution process $y^u \in L^l_{\mathbb{F}}(\Omega, C[0,T], \mathbb{R}^{n_1 \times d})$ for every $l \geq 1$. The existence and uniqueness of the inhomogenous linear Itô SDE (1.6) follows by the same result used for the existence and uniqueness of equation (1.2), since the solution process y_r^u to (1.6) which appears linearly in (1.7) is a continuous $\mathcal{F}$-adapted process having moments of all orders, uniformly in t. Similar to the Young and also the ODE case we can formulate a variation of constants formula to get an explicit solution to (1.7), using the corresponding homogenous linear SDEs. These homogenous linear SDEs corresponding to the equations (1.6) and (1.7), will play a crucial role establishing our main goal, namely the formulation of the gradient of our cost function by the adjoint equation, and also in the approximation of the adjoint equation. Hence we will come back to these equations after we stated our differentiability result. In the Young case, we show that the derivative of the solution mapping $u \mapsto \xi^u$ is given by the solution to equation (1.7) generalizing the ideas of Han et al. [2012] to p-variation spaces, where related calculations were made for Hölder continuous paths, respectively using Hölder norms, for the special case of a fractional Brownian motion driver, in a stochastic control setting. In the Itô case the differentiability of the solution mapping $u \mapsto x^u$ follows by standard techniques from stochastic analysis.

Theorem 1.2. *The solution mapping $u \mapsto \mathcal{X}^u$ from $\mathcal{U}$ to $L^l_{\mathbb{F}}(\Omega, C[0,T], \mathbb{R}^{n_1+n_2})$ for $l \geq 1$ is Fréchet differentiable and the Fréchet derivative $D\mathcal{X}^u$ equals $\mathcal{Y}^u$, where $\mathcal{Y}^u$ is the unique solution to the SDE* (1.8)*. Furthermore, we have for every $l \geq 1$, that*

$$
\mathrm{E}\left[\|\mathcal{Y}^u\|^l_{\infty,0,T}\right] \leq D_{\mathcal{Y},l}
$$

for a positive constant $D_{\mathcal{Y},l}$, which is independent of u.

Proof. See Corollary 2.36, Corollary 2.37, Lemma 2.43 and Remark 2.47. □

The next intermediate goal, which will be achieved in the next section, is to establish a variation of constants formula for the equation (1.8). Since the processes used in the variation of constants formula of the the two respective equations (1.6) and (1.7) will be an important component in the calculations, we define them and state a result concerning their existence, uniqueness and boundedness. We define the $\mathbb{R}^{n_1 \times n_1}$ valued homogenous linear stochastic YDEs for a general initial time $s_0 \in [0, T]$ and we leave out the dependence of the involved processes on u for readability

$$\phi_t^{s_0} = I_{n_1} + \int_{s_0}^t b_x(r, \xi_r, u)\phi_r^{s_0}\, dr + \sum_{j=1}^m \int_{s_0}^t \sigma_x^j(r, \xi_r, u)\phi_r^{s_0}\, dw_r^j \tag{1.9}$$

and

$$\psi_t^{s_0} = I_n - \int_{s_0}^t \psi_t^{s_0} b_x(r, \xi_r, u)\, dr - \sum_{j=1}^m \int_{s_0}^t \psi_r^{s_0}\sigma_x^j(r, \xi_r, u)\, dw_r^j. \tag{1.10}$$

Note that the generalisation to an arbitrary initial time will only be important for the approximation of the adjoint equation later.

Lemma 1.3. *Both of the matrix valued YDEs* (1.9) *and* (1.10) *have a unique solution, which is an element of the space* $L^l_{\mathbb{F}}(\Omega, C[s_0, T], \mathbb{R}^{n_1 \times n_1})$ *for every* $l \geq 1$*. We have that*

$$\sup_{s_0 \in [0,T]} \max \left\{ \mathrm{E}\left[\|\psi^u\|^l_{\infty, s_0, T} \right], \mathrm{E}\left[\|\phi^u\|^l_{\infty, s_0, T} \right] \right\} \leq D_{\phi, l},$$

for every $l \geq 1$*, where the constant* $D_{\phi,l}$ *is independent of* u*. Furthermore, it holds for every* $u \in \mathcal{U}$ *and* $s_0 \in [0, T]$ *that* $\psi_t^{s_0,u} = (\phi^{s_0,u})_t^{-1}$ *for every* $t \in [s_0, T]$*,* P*-almost surely. The explicit solution to equation* (1.6) *is given by*

$$y_t^u = \phi_t D x_0(u) + \phi_t \int_0^t \phi_r^{-1} b_u^u(r)\, dr + \sum_{j=1}^m \phi_t \int_0^t \phi_r^{-1}\sigma_u^{u,j}(r)\, dw_r^j,$$

where ϕ *and* ϕ^{-1} *are the solutions to the equations* (1.9)*, respectively* (1.10) *with initial time* $s_0 = 0$*.*

Proof. Existence, uniqueness and boundedness follow by Corollary 2.36 and Corollary 2.37. The inverse relation of the processes follows by the pathwise application of Lemma 2.32 for almost all $\omega \in \Omega$. The variation of constants formula follows by the pathwise application of Lemma 2.33. □

Analogously, we state the similar results for the homogenous linear Itô SDEs. For a general initial time $s_0 \in [0, T]$, we define

$$\hat{\phi}_t^{s_0} = I_{n_2} + \int_{s_0}^t \hat{b}_x(r, x_r, \xi_r, u)\hat{\phi}_r^{s_0}\, dr + \sum_{j=1}^{m_2} \int_{s_0}^t \hat{\sigma}_x^j(r, x_r, \xi_r, u)\hat{\phi}_r^u\, dB_r^j \tag{1.11}$$

and

$$\hat{\psi}_t^{s_0} = I_{n_2} - \int_{s_0}^t \hat{\psi}_r^{s_0} \left(\hat{b}_x(r, x_r, \xi_r, u) - \sum_{j=1}^{m_2} (\hat{\sigma}_x^j(r, x_r, \xi_r, u))^2 \right) dr$$
$$- \sum_{j=1}^{m_2} \int_{s_0}^t \hat{\psi}_r^{s_0} \hat{\sigma}_x^j(r, x_r, \xi_r, u) \, dB_r^j. \tag{1.12}$$

Lemma 1.4. *Both of the matrix valued SDEs* (1.11) *and* (1.12) *have a unique solution, which is an element of the space* $L^l_{\mathbb{F}}(\Omega, C[s_0, T], \mathbb{R}^{n_2 \times n_2})$ *for every* $l \geq 1$. *We have that*

$$\sup_{s_0 \in [0,T]} \max \left\{ \mathrm{E}\left[\|\hat{\phi}^u\|^l_{\infty, s_0, T} \right], \mathrm{E}\left[\|\hat{\psi}^u\|^l_{\infty, s_0, T} \right] \right\} \leq D_{\hat{\phi}, l},$$

for every $l \geq 1$, *where the constant* $D_{\hat{\phi}, l}$ *is independent of* u. *Furthermore, it holds for every* $u \in \mathcal{U}$ *and* $s_0 \in [0, T]$ *that* $\hat{\psi}_t^{s_0, u} = (\hat{\phi}_t^{s_0, u})^{-1}$ *for every* $t \in [s_0, T]$, *P-almost surely. Setting* $s_0 = 0$ *and* $\hat{\phi}^0 = \phi$, *the solution to equation* (1.7) *is given by*

$$\hat{y}_t^u = \hat{\phi}_t D x_0(u) + \hat{\phi}_t \int_0^t \hat{\phi}_r^{-1} \left(\hat{b}_z^u(t) D\xi_r^u + \hat{b}_u^u(r) - \sum_{j=1}^{m_2} \hat{\sigma}_x^{u,j}(r) \left(\hat{\sigma}_z^{u,j}(r) D\xi_r^u + \hat{\sigma}_u^{u,j}(r) \right) \right) dr$$
$$+ \sum_{j=1}^{m_2} \hat{\phi}_t \int_0^t \hat{\phi}_r^{-1} \left(\hat{\sigma}_z^{u,j}(r) D\xi_r^u + \hat{\sigma}_u^{u,j}(r) \right) dB_r^j.$$

Proof. See Lemma 2.45. $\square$

The next section summarizes the results of Chapter 3, the main chapter of this thesis. We establish the representation of the gradient of our cost function, by the explicit solution to the adjoint equation, which is an anticipating backward stochastic differential equation.

1.3 The cost function and its gradient

We consider our cost function

$$J(u) = \frac{1}{2} \sum_{\mu=1}^M \mathrm{E}\left[g_\mu(\mathcal{X}_{T_\mu}^u) \right]^2,$$

where $0 < T_1 \leq \ldots, T_M = T$ is a sequence of times in $(0, T]$ and we assume for the rest of this chapter, that the functions

$$g_\mu : \mathbb{R}^{n_1 + n_2} \to \mathbb{R}$$

satisfy the following condition

(G) Let L be the constant used in the conditions on the coefficient functions b, σ, $\hat{b}$ and $\hat{\sigma}$. For every $\mu = 1, \ldots, M$, we have that $g_\mu : \mathbb{R}^{(n_1+n_2)} \to \mathbb{R}$ is continuously differentiable and we denote the derivative by

$$g_\mu'(z) = \left(\frac{\partial}{\partial z_1} g_\mu(z), \dots, \frac{\partial}{\partial z_{n_1+n_2}} g_\mu(z)\right) \in \mathbb{R}^{n_1+n_2}.$$

We assume for all $z, y \in \mathbb{R}^{n_1+n_2}$ that

$$|g_\mu'(z)| \leq L$$

and

$$|g_\mu'(z) - g_\mu'(y)| \leq L|z - y|.$$

Our goal is to calculate the gradient of our cost function. We will do this in two ways. One easy way to calculate the gradient would be to use the solution of the so called sensitivity equation $\mathcal{Y}$ (see (2.65)), which is the Fréchet differential $D\mathcal{X}^u$ of the solution mapping $u \mapsto \mathcal{X}^u$. Using the chain rule for Fréchet derivatives (see Ambrosetti and Prodi [1995] Proposition 1.1.4), we obtain for the gradient

$$\nabla J(u) = \sum_{\mu=1}^{M} E[g_\mu(\mathcal{X}_{T_\mu}^u)]\mathrm{E}[g_\mu'(\mathcal{X}_{T_\mu}^u)\mathcal{Y}_{T_\mu}^u].$$

Then we want to calculate an explicit solution to the $(n_1 + n_2) \times d$-dimensional system of linear SDEs (1.8), by establishing a variation of constants formula similar to Lemma 1.3, respectively Lemma 1.4, but for the whole system of equations. This explicit solution can then be used to get a second representation of the gradient, involving the solution to an anticipating backward stochastic differential equation, which we call the adjoint equation. In the course of these calculations, we will encounter several technical problems. First, one of the processes involved in the explicit solutions of our system of differential equations (1.8), will be the product of a process driven by Brownian motion and a process driven by w. Therefore we need an integration by parts rule which connects both of these processes. Remember that the involved integrals are of different type, one is a pathwise Young integral and the other is the standard Itô integral. Moreover, for the calculation of the adjoint equation, we would like to integrate over random variables which are $\mathcal{F}_T$ measurable and hence anticipating. This makes it impossible to use the Itô integral. Luckily all these problems can be solved by applying a stochastic integral which generalizes both, the pathwise Young and the Itô integral. The Section 3.1 is devoted to this generalization, called the forward integral, introduced by F. Russo and P. Vallois in the paper Russo and Vallois [1993a]. Here we will just give a short definition of the forward integral and comment on its properties. Let $(X_t)_{t\in[s_0,T]}$ be a continuous processes and $(Y_t)_{t\in[s_0,T]}$ be locally bounded, meaning that for every $t > s_0$, $\int_{s_0}^t Y_s \, ds < \infty$, P-almost surely. The forward integral is defined as the limit in ucp-sense of the ε-forward integral, if this limit exists. Precisely

$$\varepsilon - \text{forward integral} : I^{-}(\varepsilon, Y, dX)(t) = \int_{s_0}^{t} Y_s \frac{X_{(s+\varepsilon)\wedge T} - X_s}{\varepsilon}\, ds$$

and

$$\text{Forward-integral} : \int_{s_0}^{t} Y_s\, d^{-}X_s = \lim_{\varepsilon \searrow 0} I^{-}(\varepsilon, Y, dX)(t).$$

Note that a family of processes $(H_t^{\varepsilon})_{t\in[s_0,T]}$ converges to a process $(H_t)_{t\in[s_0,T]}$ in ucp-sense (uniform in probability), if

$$\lim_{\varepsilon\to 0} \sup_{t\in[s_0,T]} |H_t^{\varepsilon} - H_t| = 0$$

in probability. The key feature of this integral is that it coincides in our situation with the Riemann integral (see Theorem 3.5, cited from Russo and Vallois [2007], Proposition 1 7a)), the Itô integral (see Theorem 3.9, cited from Russo and Vallois [2007], Proposition 6) and the Young integral (see Theorem 3.7). The coincidence with the Young integral was proven in Russo and Vallois [2007] for Hölder continuous integrand and integrator, we generalized the results to the case of continuous integrand and integrator having finite p-respectively q-variation, such that $\frac{1}{p} + \frac{1}{q} > 1$. This enables us to state an integration by parts formula (Theorem 3.13), which is then used to establish the explicit solution to equation (1.8). To do this, we define the homogenous linear SDEs corresponding to the inhomogenous equation (1.8). To simplify the notation we first define the following functions for $r \in [0,T]$ and $u \in \mathcal{U}$:

$$\begin{aligned}
&b_x^u(r) = b_x(r, \xi_r^u, u), \quad b_u^u(r) = b_u(r, \xi_r^u, u),\\
&\sigma_x^{u,j}(r) = \sigma_x^j(r, \xi_r^u, u), \quad \sigma_u^{u,j}(r) = \sigma_u(r, \xi_r^u, u) \text{ for } j = 1, \dots, m_1,\\
&\hat{b}_x^u(r) = \hat{b}_x(r, x_r^u, \xi_r^u, u), \quad \hat{b}_z^u(r) = \hat{b}_z(r, x_r^u, \xi_r^u, u), \quad \hat{b}_u^u(r) = \hat{b}_u(r, x_r^u, \xi_r^u, u)\\
&\hat{\sigma}_x^{u,j}(r) = \hat{\sigma}_x^j(r, x_r^u, \xi_r^u, u), \quad \hat{\sigma}_z^{u,j}(r) = \hat{\sigma}_z^j(r, x_r^u, \xi_r^u, u), \quad \hat{\sigma}_u^{u,j}(r) = \hat{\sigma}_u^j(r, x_r^u, \xi_r^u, u) \text{ for } j = 1, \dots, m_2.
\end{aligned}$$

Let $s_0 \in [0,T]$ and leave out the dependence on the solution processes on u for readability, we define the $\mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}$-valued systems of equations for $t \in [s_0, T]$ by

$$\begin{aligned}
\Phi_t^{s_0} = I_{n_1+n_2} &+ \int_{s_0}^{t} \begin{pmatrix} b_x^u(r) & 0 \\ \hat{b}_z^u(r) & \hat{b}_x^u(r) \end{pmatrix} \Phi_r^{s_0}\, dr + \sum_{j=1}^{m_1} \int_{s_0}^{t} \begin{pmatrix} \sigma_x^{u,j}(r) & 0 \\ 0 & 0 \end{pmatrix} \Phi_r^{s_0}\, dw_r^j \\
&+ \sum_{j=1}^{m_2} \int_{s_0}^{t} \begin{pmatrix} 0 & 0 \\ \hat{\sigma}_z^{u,j}(r) & \hat{\sigma}_x^{u,j}(r) \end{pmatrix} \Phi_r^{s_0}\, dB_r^j, \qquad (1.13)
\end{aligned}$$

and

$$\Psi_t^{s_0} = I_{n_1+n_2} - \int_{s_0}^t \Psi_r^{s_0} \left[\begin{pmatrix} b_x^u(r) & 0 \\ \hat{b}_z^u(r) & \hat{b}_x^u(r) \end{pmatrix} - \sum_{j=1}^{m_2} \begin{pmatrix} 0 & 0 \\ \hat{\sigma}_z^u(r) & \hat{\sigma}_x^{u,j}(r) \end{pmatrix}^2 \right] dr$$
$$- \sum_{j=1}^{m_1} \int_{s_0}^t \Psi_r^{s_0} \begin{pmatrix} \sigma_x^{u,j}(r) & 0 \\ 0 & 0 \end{pmatrix} dw_r^j - \sum_{j=1}^{m_2} \int_{s_0}^t \Psi_r^{s_0} \begin{pmatrix} 0 & 0 \\ \hat{\sigma}_z^{u,j}(r) & \hat{\sigma}_x^{u,j}(r) \end{pmatrix} dB_r^j, \tag{1.14}$$

The following theorem is an important result, since it is the basis for the formulation of the adjoint equation.

Theorem 1.5. *For every $u \in \mathcal{U}$ and $s_0 \in [0,T]$ the equations* (1.13) *and* (1.14) *have a unique solution $\Phi^{s_0,u}$, respectively $\Psi^{s_0,u}$ in $L^l_{\mathbb{F}}(\Omega, C[s_0,T], \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)})$ for every $l \geq 1$, such that*

$$\sup_{s_0\in[0,T]} \max \left\{ \mathrm{E}\left[\|\Psi^{s_0,u}\|^l_{\infty,s_0,T} \right], \mathrm{E}\left[\|\Phi^{s_0,u}\|^l_{\infty,s_0,T} \right] \right\} \leq D_{\Phi,l},$$

where the positive constant $D_{\Phi,l}$ is independent of u. We have that $\Psi_t^{s_0,u} = (\Phi_t^{s_0,u})^{-1}$ for $t \in [s_0,T]$, P-almost surely. Furthermore the explicit solution to the equation (1.8) *for every $t \in [0,T]$ is given by the following variation of constants formula (here we set the initial time of the homogenous equations to $s_0 = 0$ and leave out the indexes s_0 and u)*

$$\mathcal{Y}_t^u = \Phi_t \begin{pmatrix} D\xi_0(u) \\ Dx_0(u) \end{pmatrix} + \Phi_t \int_0^t \Phi_r^{-1} \left[\begin{pmatrix} b_u^u(r) \\ \hat{b}_u^u(r) \end{pmatrix} - \sum_{j=1}^{m_2} \begin{pmatrix} 0 \\ \hat{\sigma}_x^{u,j}(r)\hat{\sigma}_u^{u,j}(r) \end{pmatrix} \right] dr$$
$$+ \sum_{j=1}^{m_1} \Phi_t \int_0^t \Phi_r^{-1} \begin{pmatrix} \sigma_u^{u,j}(r) \\ 0 \end{pmatrix} dw_r^j + \sum_{j=1}^{m_2} \Phi_t \int_0^t \Phi_r^{-1} \begin{pmatrix} 0 \\ \hat{\sigma}_u^{u,j}(r) \end{pmatrix} dB_r^j.$$

Proof. The statement follows from the arguments in Section 3.2 and Theorem 3.15, but here we want to give a small overview of the needed steps. Taking the form of the equations (1.13) and (1.14) into account, the solution processes need to be of the form

$$\Phi_t^{s_0} = \begin{pmatrix} \phi_t^{s_0} & 0 \\ \tilde{\phi}_t^{s_0} & \hat{\phi}_t^{s_0} \end{pmatrix}, \qquad \Psi_t^{s_0} = \begin{pmatrix} \psi_t^{s_0} & 0 \\ \tilde{\psi}_t^{s_0} & \hat{\psi}_t^{s_0} \end{pmatrix},$$

for every $t \in [s_0,T]$. Hence we get for both equations (1.13) and (1.14), three lower dimensional equations. The two equations on the diagonal are given by the equations (1.9) and (1.11), respectively (1.10) and (1.12), which we already analyzed in Lemma 1.3 and Lemma 1.4. There we got existence, uniqueness and boundedness of the solutions. Hence the only equations which need further examination, are

$$\tilde{\phi}_t^{s_0} = \int_{s_0}^t \hat{b}_x^u(r)\tilde{\phi}_r^{s_0} + \hat{b}_z^u(r)\phi_r^{s_0}\, dr + \sum_{j=1}^{m_2} \int_{s_0}^t \hat{\sigma}_x^{u,j}(r)\tilde{\phi}_r^{s_0} + \hat{\sigma}_z^{u,j}(r)\phi_r^{s_0} dB_r^j \tag{1.15}$$

and

$$\tilde{\psi}_t^{s_0} = -\int_{s_0}^t \tilde{\psi}_r^{s_0} \hat{b}_x^u(r) + \hat{\psi}_r^{s_0} \left[\hat{b}_z^u(r) - \sum_{j=1}^{m_2} \hat{\sigma}_x^{u,j}(r) \hat{\sigma}_z^{u,j}(r) \right] dr$$
$$- \sum_{j=1}^{m_1} \int_{s_0}^t \tilde{\psi}_r^{s_0} \sigma_x^{u,j}(r)\, dw_r^j - \sum_{j=1}^{m_2} \int_{s_0}^t \hat{\psi}_r^{s_0} \hat{\sigma}_z^{u,j}(r) dB_r^j \tag{1.16}$$

for $t \in [s_0, T]$, which both stem from the dependence of equation (1.2) on the solution of equation (1.1). The existence, uniqueness and boundedness of the solution to equation (1.15), again follows by standard results on stochastic analysis, since it is an inhomogenous linear Itô SDE. Equations of this kind were already considered, e.g. equation (1.7). We get the explicit solution by a variation of constants formula, which is given by

$$\tilde{\phi}_t^{s_0} = \hat{\phi}_t^{s_0} \int_{s_0}^t (\hat{\phi}_r^{s_0})^{-1} \left[\hat{b}_z^u(r) - \sum_{j=1}^{m_2} \hat{\sigma}_x^{u,j}(r) \hat{\sigma}_z^{u,j}(r) \right] \phi_r^{s_0}\, dr + \sum_{j=1}^{m_2} \hat{\phi}_t^{s_0} \int_{s_0}^t (\hat{\phi}_r^{s_0})^{-1} \hat{\sigma}_z^{u,j}(r) \phi_r^{s_0}\, dB_r^j.$$

The challenging part is the existence of a solution to equation (1.16), since the equation contains both Young and Itô integral terms. Fortunately we can get a candidate for a solution by considering the inverse relation between the solutions to the equations (1.13) and (1.14), which we hope to be satisfied. We already know that $\psi_t^{s_0} = (\phi_t^{s_0})^{-1}$ and $\hat{\psi}_t^{s_0} = (\hat{\phi}_t^{s_0})^{-1}$ for all $t \in [s_0, T]$, P-almost surely. Hence assuming this inverse relation and taking the explicit solution to equation (1.15) into account, we get the candidate

$$\tilde{\psi}_t^{s_0} = -(\hat{\phi}_t^{s_0})^{-1} \tilde{\phi}_t^{s_0} (\phi_t^{s_0})^{-1} \tag{1.17}$$
$$= \left(-\int_{s_0}^t (\hat{\phi}_r^{s_0})^{-1} \left[\hat{b}_z^u(r) - \sum_{j=1}^{m_2} \hat{\sigma}_x^{u,j}(r) \hat{\sigma}_z^{u,j}(r) \right] \phi_r^{s_0}\, dr - \sum_{j=1}^{m_2} \int_{s_0}^t \hat{\phi}_r^{-1} \hat{\sigma}_z^{u,j}(r) \phi_r^{s_0}\, dB_r^j \right) (\phi_t^{s_0})^{-1}. \tag{1.18}$$

This candidate for an explicit solution of equation (1.15) is now a product of a process driven by w_t, namely $(\phi_t^{s_0})^{-1}$ and a process driven by B, given by

$$-\int_{s_0}^t (\hat{\phi}_r^{s_0})^{-1} \left[\hat{b}_z^u(r) - \sum_{j=1}^{m_2} \hat{\sigma}_x^{u,j}(r) \hat{\sigma}_z^{u,j}(r) \right] \phi_r^{s_0}\, dr - \sum_{j=1}^{m_2} \int_{s_0}^t \hat{\phi}_r^{-1} \hat{\sigma}_z^{u,j}(r) \phi_r^{s_0}\, dB_r^j.$$

Here the introduction of the forward integral has its first application to our results. Using the integration by parts formula from Theorem 3.13, we can prove that (1.18) is indeed a solution to equation (1.16). Uniqueness then follows easily and the boundedness of the solution is a direct consequence of the representation (1.17), and the boundedness of the factors in L^l-norm for every $l \geq 1$, uniformly in t. Then we get directly, that $\Psi_t^{s_0,u} = (\Phi_t^{s_0,u})^{-1}$ for $t \in [s_0, T]$ P-almost surely and the variation of constants formula for the equation (1.12) follows again by an application of Theorem 3.13 on the product $(\Phi_t^0)^{-1}\mathcal{Y}_t$. $\square$

In the next theorem, we will state the second gradient representation by making use of the adjoint equation. This is the first main result of this thesis and uses the coincidence of the Itô and forward integral in our situation and a useful property of the forward integral, which directly follows by its definition. Let Z be any real valued random variable, e.g. an $\mathcal{F}_T$ measurable random variable, and let the process Y be forward integrable, then

$$Z\int_0^T Y_s d^- X_s = \int_0^T Z Y_s d^- X_s,$$

hence the forward integral allows for anticipating integrands. This property and the variation of constants formula from the last theorem, will be the main tool in proving the following result.

Theorem 1.6. *The gradient of the cost function is given by*

$$\begin{aligned}\nabla J(u) = \mathrm{E}\Bigg[\Lambda_0 \begin{pmatrix} D\xi_0(u) \\ Dx_0(u) \end{pmatrix} + \int_0^T \Lambda_r \Bigg[\begin{pmatrix} b_u^u(r) \\ \hat{b}_u^u(r) \end{pmatrix} - \sum_{j=1}^{m_2} \begin{pmatrix} 0 \\ \hat{\sigma}_x^{u,j}(r)\hat{\sigma}_u^{u,j}(r) \end{pmatrix} \Bigg] dr \\ + \sum_{j=1}^{m_1} \int_0^T \Lambda_r \begin{pmatrix} \sigma_u^{u,j}(r) \\ 0 \end{pmatrix} dw_r^j + \sum_{j=1}^{m_2} \int_0^T \Lambda_r \begin{pmatrix} 0 \\ \hat{\sigma}_u^{u,j}(r) \end{pmatrix} d^- B_r^j \Bigg],\end{aligned} \tag{1.19}$$

where the row vector

$$\Lambda_t = \sum_{T_\mu \geq t} E[g_\mu(\mathcal{X}_{T_\mu}^u)] g'_\mu(\mathcal{X}_{T_\mu}^u) \Phi_{T_\mu} \Phi_t^{-1} \text{ for } t \in [0,T]$$

is an element of $L^l(\Omega, C[0,T], \mathbb{R}^{n_1+n_2})$ and satisfies the anticipating BSDE

$$\begin{aligned}\Lambda_t = \sum_{T_\mu \geq t} E[g_\mu(\mathcal{X}_{T_\mu}^u)] g'_\mu(\mathcal{X}_{T_\mu}^u) + \int_t^T \Lambda_r \Bigg[\begin{pmatrix} b_x^u(r) & 0 \\ \hat{b}_z^u(r) & \hat{b}_x^u(r) \end{pmatrix} - \sum_{j=1}^{m_2} \begin{pmatrix} 0 & 0 \\ \hat{\sigma}_z^{u,j}(r) & \hat{\sigma}_x^{u,j}(r) \end{pmatrix}^2 \Bigg] dr \\ + \sum_{j=1}^{m_1} \int_t^T \Lambda_r \begin{pmatrix} \sigma_x^{u,j}(r) & 0 \\ 0 & 0 \end{pmatrix} dw_r^j + \sum_{j=1}^{m_2} \int_t^T \Lambda_r \begin{pmatrix} 0 & 0 \\ \hat{\sigma}_z^{u,j}(r) & \hat{\sigma}_x^{u,j}(r) \end{pmatrix} d^- B_r^j,\end{aligned} \tag{1.20}$$

for all $t \in [0,T]$. We call this equation the adjoint equation.

Proof. See Theorem 3.17. $\square$

Having established our main result, the rest of the thesis is concerned with its numerical approximation, to use the results in practice.

1.4 Approximation of the cost function and its gradient

For the discretization of the model dynamics equation (1.3) and the sensitivity equation (1.8), we choose the standard discretization by first order Euler schemes. We define the discretization schemes on $(\Omega, \mathcal{F}, \mathbb{F}, P)$ for a partition $\Pi^{\text{Euler}} = \Pi^{\text{E}} = (t_i)_{i=0,\dots,n}$ of $[0,T]$, which is not necessarily equidistant. For each $\omega \in \Omega$, $u \in \mathcal{U}$ and $i \in \{0,\dots,n-1\}$ we define the discrete Euler scheme $\mathcal{X}^n$ for the equation (1.3) by

$$\begin{aligned}
\mathcal{X}^n_{t_{i+1}}(\omega) &= \begin{pmatrix} \xi^n_{t_{i+1}}(\omega) \\ x^n_{t_{i+1}}(\omega) \end{pmatrix} \\
&= \begin{pmatrix} \xi^n_{t_i}(\omega) \\ x^n_{t_i}(\omega) \end{pmatrix} + \begin{pmatrix} b\left(t_i, \xi^n_{t_i}(\omega), u\right) \\ \hat{b}\left(r, x^n_{t_i}(\omega), \xi^n_{t_i}(\omega), u\right) \end{pmatrix} (t_{i+1} - t_i) \\
&\quad + \sum_{j=1}^{m_1} \begin{pmatrix} \sigma^j\left(t_i, \xi^n_{t_i}(\omega), u\right) \\ 0 \end{pmatrix} \left(w^j_{t_{i+1}}(\omega) - w^j_{t_i}(\omega)\right) \\
&\quad + \sum_{j=1}^{m_2} \begin{pmatrix} 0 \\ \hat{\sigma}^j\left(t_i, x^n_{t_i}(\omega), \xi^n_{t_i}(\omega), u\right) \end{pmatrix} \left(B^j_{t_{i+1}}(\omega) - B^j_{t_i}(\omega)\right)
\end{aligned}$$

and

$$\mathcal{X}^n_{t_0} = \mathcal{X}_0(u) = (\xi_0(u), x_0(u))^\top.$$

Here we left out the direct dependence of $\mathcal{X}^n$ on u for readability. We will consider the continuous interpolation

$$\begin{aligned}
\mathcal{X}^n_t(\omega) &= \begin{pmatrix} \xi^n_t(\omega) \\ x^n_t(\omega) \end{pmatrix} \\
&= \begin{pmatrix} \xi^n_{t_i}(\omega) \\ x^n_{t_i}(\omega) \end{pmatrix} + \begin{pmatrix} b\left(t_i, \xi^n_{t_i}(\omega), u\right) \\ \hat{b}\left(r, x^n_{t_i}(\omega), \xi^n_{t_i}(\omega), u\right) \end{pmatrix} (t - t_i) \\
&\quad + \sum_{j=1}^{m_1} \begin{pmatrix} \sigma^j\left(t_i, \xi^n_{t_i}(\omega), u\right) \\ 0 \end{pmatrix} \left(w^j_t(\omega) - w^j_{t_i}(\omega)\right) \\
&\quad + \sum_{j=1}^{m_2} \begin{pmatrix} 0 \\ \hat{\sigma}^j\left(t_i, x^{n,u}_{t_i}(\omega), \xi^n_{t_i}(\omega), u\right) \end{pmatrix} \left(B^j_t(\omega) - B^j_{t_i}(\omega)\right)
\end{aligned}$$

for $t \in [t_i, t_{i+1}]$ for every $i \in \{0,\dots,n-1\}$. Similarly for each $\omega \in \Omega$, $u \in \mathcal{U}$ and $i \in \{0,\dots,n-1\}$, we define the discrete Euler scheme $\mathcal{Y}^n$ for the equation (1.8) by

$$\begin{aligned}
\mathcal{Y}^n_{t_{i+1}} &= \begin{pmatrix} y^n_{t_{i+1}} \\ \hat{y}^n_{t_{i+1}} \end{pmatrix} \\
&= \mathcal{Y}^n_{t_i} + \left(\begin{pmatrix} b_x\left(t_i, \xi^n_{t_i}, u\right) & 0 \\ \hat{b}_z\left(t_i, x^n_{t_i}, \xi^n_{t_i}, u\right) & \hat{b}_x\left(t_i, x^n_{t_i}, \xi^n_{t_i}, u\right) \end{pmatrix} \mathcal{Y}^n_{t_i} + \begin{pmatrix} b_u\left(t_i, \xi^n_{t_i}, u\right) \\ \hat{b}_u\left(t_i, x^n_{t_i}, \xi^n_{t_i}, u\right) \end{pmatrix} \right) (t_{i+1} - t_i)
\end{aligned}$$

$$+\sum_{j=1}^{m_1}\left(\begin{pmatrix}\sigma_x^j\left(t_i,\xi_{t_i}^n,u\right) & 0\\ 0 & 0\end{pmatrix}\mathcal{Y}_{t_i}^n+\begin{pmatrix}\sigma_u^j(t_i,\xi_{t_i}^n,u)\\ 0\end{pmatrix}\right)\left(w_{t_{i+1}}^j-w_{t_i}^j\right)$$
$$+\sum_{j=1}^{m_2}\begin{pmatrix}0 & 0\\ \hat{\sigma}_z^j\left(t_i,x_{t_i}^n,\xi_{t_i}^n,u\right) & \hat{\sigma}_x^j\left(t_i,x_{t_i}^n,\xi_{t_i}^n,u\right)\end{pmatrix}\mathcal{Y}_{t_i}^n\left(B_{t_{i+1}}^j-B_{t_i}^j\right)$$
$$+\sum_{j=1}^{m_2}\begin{pmatrix}0\\ \hat{\sigma}_u^j\left(t_i,x_{t_i}^n,\xi_{t_i}^n,u\right)\end{pmatrix}\left(B_{t_{i+1}}^j-B_{t_i}^j\right) \tag{1.21}$$

with

$$\mathcal{Y}_{t_0}^n=\mathcal{Y}_0(u)=(D\xi_0(u),Dx_0(u))^\top$$

and its continuous interpolation defined as above. Again we leave out the direct dependencies of the processes on u and ω for readability. We will now be concerned with the convergence of $\mathcal{X}^{n,u}$ to $\mathcal{X}^u$ in $L^l(\Omega,C([0,T],\mathbb{R}^{n_1+n_2})$ and the convergence of $\mathcal{Y}^{n,u}$ to $\mathcal{Y}^u$ in $L^l(\Omega,C([0,T],\mathbb{R}^{(n_1+n_2)\times d})$ for every $u\in\mathcal{U}$ and $l\geq 2$. The restriction to $l\geq 2$ is just technical to facilitate the proofs and shorten the notation for discretization of the Itô SDEs and consequently the whole system. For $l\in(1,2]$ the results follow by the monotonicity of L^l-norms. To get the convergence results, we need to adapt and add some conditions on the coefficient functions, which we assume to be satisfied by the coefficient functions b, σ, $\hat{b}$, $\hat{\sigma}$ for the rest of this section, namely

(H_3*): The Hölder exponent β from condition (H_3) is an element of the interval $[\frac{1}{p},1]$, instead of $[\frac{1}{2},1]$.

(E_1): Let b, L be the function and constant from condition (H_2) and $\beta\in[\frac{1}{p},1]$ the same constant in condition (H_3^*). For every $x\in\mathbb{R}^{n_1}$, $u\in\mathcal{U}$ and $s\leq t\in[0,T]$, b satisfies

$$|b(t,x,u)-b(s,x,u)|+|b_x(t,x,u)-b_x(s,x,u)|+|b_u(t,x,u)-b_u(s,x,u)|\leq L|t-s|^\beta.$$

(E_2): Let $\hat{b}$, $\hat{\sigma}$ and L be the coefficient functions and the constant from condition (B_1), respectively (B_2). For all $x\in\mathbb{R}^{n_2}$, $y\in\mathbb{R}^{n_1}$, $u\in\mathcal{U}$ and $s\leq t\in[0,T]$, $\hat{b}$ and $\hat{\sigma}$ satisfy

$$|\hat{b}(t,x,y,u)-\hat{b}(s,x,y,u)|+|\hat{\sigma}(t,x,y,u)-\hat{\sigma}(s,x,y,u)|\leq L(1+|x|+|y|)(t-s)^{\frac{1}{2}}.$$

We will give different convergence results depending on the properties of the driving process w. The standard case is that w is a continuous process having paths of bounded p-variation for $p\in(1,2)$, where we do not assume any kind of Hölder condition on w. Because of this we cannot expect to get a convergence parameter which only depends on the mesh $|\Pi^{\mathrm{E}}|=\max_{i=0,\dots,n-1}|t_{i+1}-t_i|$ of the Euler partition, as it is in standard approximation schemes of Itô SDEs. We define two convergence parameters in this case. First, we define for all $\omega\in\Omega$, the parameter

$$\delta(\omega):=\max_{i=0,\dots,n-1}|t_{i+1}-t_i|+|w(\omega)|_{p,t_i,t_{i+1}}$$

for the pathwise convergence of the stochastic Young differential equations. Second, we define the L^l-convergence parameter for the stochastic Young differential equation

$$\delta_{1,l} := \mathrm{E}\left[\delta^l\right]^{\frac{1}{l}},$$

which is well defined, since w satisfies the exponential moment condition (1.5). The last convergence parameter we will use in the estimation of the convergence rate is defined by

$$\delta_2 := \max_{i=0,\ldots,n-1} |t_{i+1} - t_i|,$$

which is essential in the estimates for the Itô SDEs and for the whole systems under the following additional condition

(HA): Hölder assumption: Almost every path of the process w is Hölder continuous of order $H > \frac{1}{2}$ and the Hölder seminorm

$$|w|_{H-Hol,0,T} = \sup_{s,t\in\Delta([0,T])} \frac{|w_t - w_s|}{|t-s|^H}$$

has moments of all orders.

Again we can split the analysis of the convergence to the two respective cases of approximating the YDEs and the Itô SDEs. In both cases there is a vast amount of existing literature concerned with the convergence of Euler schemes and the corresponding convergence rate. Here we will only give the references which we used in our calculations but refer the reader to Subsection 1.6 for the discussion of similar results in the literature. Concerning the convergence of Euler schemes in the YDE case, namely ξ^n to the solution of equation (1.1) and y^n to the solution of (1.6), we first show boundedness of the Euler schemes, independent of n and the convergence to the solutions pathwise. With the exponential moment condition satisfied by w, we then get the L^l-convergence with respect to the parameter $\delta_{1,l}$. In these calculations we use ideas from Lejay [2010], where the author shows the convergence of Euler schemes (in the non-linear case) to solutions of YDEs in a time autonomous, deterministic setting. We will be especially careful with the ω-dependent constants in the estimates, to get the convergence results in L^l, by incorporating the greedy sequence ideas from Nguyen et al. [2018] and ideas for L^l-estimates from Hu et al. [2016]. Concerning the convergence analysis for equation (1.6), we found no suitable results in the literature, except for Chronopoulou and Tindel [2013], where the authors consider time autonomous equations in Hölder spaces driven by fractional Brownian motion. We were not able to replicate their ideas in our setting, so our proofs are very different from theirs. Hence, our results on the boundedness of y^n in Lemma 4.9, the corresponding Gronwall Lemma (Lemma 4.8) and the convergence result in Theorem 4.10 seem new and are of interest themselves. For the convergence of x^n to the solution of (1.2) and $\hat{y}^n$ to the solution of (1.7), we use standard techniques like the Burkholder-Davis-Gundy inequality and the Gronwall inequality. The ideas for the proofs are standard but taking the rather unusual setting of our model dynamics into account, we do the calculations rigorously. The standard reference for this topic is Kloeden and Platen [2011], which gives a comprehensive overview. The following two theorems provide the rate of the convergence of $\mathcal{X}^n$ to the solution of (1.3) and $\mathcal{Y}^n$ to the solution of equation (1.8).

Theorem 1.7. *We have for every $u \in \mathcal{U}$, that*

$$\mathrm{E}\left[\|\mathcal{X}^{n,u}\|_{\infty,0,T}^{l}\right] \leq D_{\mathcal{X}^n,l}$$

and

$$\mathrm{E}\left[\|\mathcal{X}^{u} - \mathcal{X}^{n^u}\|_{\infty,0,T}^{l}\right]^{\frac{1}{l}} \leq D_{K_{\mathcal{X}},l}\delta_{1,2l}^{(2-p)\wedge\frac{1}{2}},$$

for any $l \geq 2$, where the constants $D_{\mathcal{X}^n,l}$ and $D_{K_{\mathcal{X}},l}$ are independent of u and n. If w satisfies the condition (HA), then for any $l \geq 2$, there exists a constant $\tilde{D}_{K_{\mathcal{X}},l} > 0$ such that

$$\mathrm{E}\left[\|\mathcal{X}^{u} - \mathcal{X}^{n^u}\|_{\infty,0,T}^{l}\right]^{\frac{1}{l}} \leq \tilde{D}_{K_{\mathcal{X}},l}\delta_{2}^{(2H-1)\wedge\frac{1}{2}}.$$

Proof. See Theorem 4.7, together with the considerations of Subsection 4.1.3 for the rate under the condition (HA). □

Theorem 1.8. *We have for every $u \in \mathcal{U}$, that*

$$\mathrm{E}\left[\|\mathcal{Y}^{n,u}\|_{\infty,0,T}^{l}\right] \leq D_{\mathcal{Y}^n,l}$$

and

$$\mathrm{E}\left[\|\mathcal{Y}^{u} - \mathcal{Y}^{n,u}\|_{\infty,0,T}^{l}\right]^{\frac{1}{l}} \leq D_{K_{\mathcal{Y}},l}\delta_{1,4l}^{(2-p)\wedge\frac{1}{2}},$$

for any $l \geq 2$, where the constants $D_{\mathcal{Y}^n,l}$ and $D_{K_{\mathcal{Y}},l}$ are independent of u and n. If w satisfies the condition (HA), then for any $l \geq 2$, there exists a constant $\tilde{D}_{K_{\mathcal{Y}},l} > 0$ such that

$$\mathrm{E}\left[\|\mathcal{Y}^{u} - \mathcal{Y}^{n^u}\|_{\infty,0,T}^{l}\right]^{\frac{1}{l}} \leq \tilde{D}_{K_{\mathcal{Y}},l}\delta_{2}^{(2H-1)\wedge\frac{1}{2}}.$$

Proof. See Theorem 4.7, together with the considerations of Subsection 4.1.3 for the rate under the condition (HA). □

Now we come to the discretization of the explicit solution of the adjoint equation. Our anticipating backward adjoint equation is given by (1.20). Note that Λ_t is an $n_1 + n_2$-dimensional row vector. We define the approximation scheme on a partition $\Pi^{\mathrm{E}} = \{t_i\}_{i=0,\dots,n}$ of the interval

$[0, T]$ as

$$\Lambda_{t_i}^n = (\lambda_{t_i}^n, \hat{\lambda}_{t_i}^n) = \Lambda_{t_{i+1}}^n \left(I_{n_1+n_2} + \eta_{t_i,t_{i+1}}\right) + \sum_{T_\mu = t_i} E[g_\mu(\mathcal{X}_{T_\mu}^n)] g'_\mu(\mathcal{X}_{T_\mu}^n) \in \mathbb{R}^{(n_1+n_2)}, \tag{1.22}$$

where $\Lambda_{t_i}^n$ for $i = 0, \dots, n$ is a $n_1 + n_2$-dimensional row vector and

$$\eta_{t_i,t_{i+1}} = \begin{pmatrix} b_x(t_i, \xi_{t_i}^n, u) & 0 \\ \hat{b}_z(t_i, x_{t_i}^n, \xi_{t_i}^n, u) & \hat{b}_x(t_i, x_{t_i}^n, \xi_{t_i}^n) \end{pmatrix} (t_{i+1} - t_i) + \sum_{j=1}^{m_1} \begin{pmatrix} \sigma_x^j(t_i, \xi_{t_i}^n, u) & 0 \\ 0 & 0 \end{pmatrix} (w_{t_{i+1}}^j - w_{t_i}^j)$$
$$+ \sum_{j=1}^{m_2} \begin{pmatrix} 0 & 0 \\ \hat{\sigma}_z^j(t_i, x_{t_i}^n, \xi_{t_i}^n, u) & \hat{\sigma}_x^j(t_i, x_{t_i}^n, \xi_{t_i}^n, u) \end{pmatrix} (B_{t_{i+1}}^j - B_{t_i}^j) \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}$$

for all $i \in \{0, \dots, n-1\}$ and

$$\Lambda_T^n = \sum_{T_\mu = T} \mathrm{E}[g_\mu(\mathcal{X}_T^n)] g'_\mu(\mathcal{X}_T^n).$$

In Section 4.2, we give a short explanation how to derive this discretization scheme. We use the constant interpolation on the interval $[0, T]$, meaning that

$$\Lambda_t^n = \Lambda_{t_{i+1}}^n$$

for $t \in (t_i, t_{i+1}]$.

Analyzing the explicit solution of equation (1.20), given by

$$\Lambda_t = \sum_{T_\mu \geq t} E[g_\mu(\mathcal{X}_{T_\mu})]^\top g'_\mu(\mathcal{X}_{T_\mu}) \Phi_{T_\mu}^0 (\Phi_t^0)^{-1} \text{ for } t \in [0, T],$$

where Φ_t^0 respectively $(\Phi_t^0)^{-1} = \Psi_t^0$ are the unique solutions to equation (1.13) and (1.14) with initial time 0, we are able to find a connection between the convergence of Λ^n to Λ and the convergence of the forward Euler schemes corresponding to the equations (1.9), (1.15) and (1.11). Details of the derivation are given in Section 4.2. Utilizing this connection, we establish our second main result, given in the next theorem.

Theorem 1.9. *For all $u \in \mathcal{U}$ and $l \geq 2$, we have*

$$\sup_{t\in[0,T]} \mathrm{E}\left[|\Lambda_t - \Lambda_{t_{\underline{n}(t)}}^n|^l\right]^{\frac{1}{l}} \leq D_{K_\Lambda,l} \delta_{1,4l}^{(2-p)\wedge\frac{1}{2}},$$

where the constant $D_{K_\Lambda,l} > 0$ is independent of u and n. Under the assumption (HA), there exists a constant $\tilde{D}_{K_\Lambda,l} > 0$ such that

$$\sup_{t\in[0,T]} \mathrm{E}\left[|\Lambda_t - \Lambda_{t_{\underline{n}(t)}}^n|^l\right]^{\frac{1}{l}} \leq \tilde{D}_{K_\Lambda,l} \delta_2^{(2H-1)\wedge\frac{1}{2}}.$$

Proof. See Theorem 4.20. □

Having established these convergence results, we can use them for the approximation of the cost function and the two representations of its gradient. Our cost function is given by

$$J : \mathcal{U} \to \mathbb{R}, J(u) = \frac{1}{2} \sum_{\mu=1}^{M} \mathrm{E}[g_\mu(\mathcal{X}^u_{T_\mu})]^2.$$

We introduce the discretized cost function and prove its convergence to the cost function and that the same holds for the corresponding gradients. Let $\Pi^{\mathrm{E}} = (t_i)_{i=0,\dots,n}$ be a partition of the interval $[0, T]$ such that $(T_\mu)_{\mu=1,\dots,M} \subset \Pi^{\mathrm{E}}$, the discretized cost function and the discretized gradient are given by

$$\begin{aligned} J^n(u) &:= \frac{1}{2} \sum_{\mu=1}^{M} \mathrm{E}\left[g_\mu(\mathcal{X}^{n,u}_{T_\mu})\right]^2 \\ (\nabla J)^n(u) &:= \sum_{\mu=1}^{M} \mathrm{E}\left[g_\mu(\mathcal{X}^{n,u}_{T_\mu})\right] \mathrm{E}\left[g'_\mu(\mathcal{X}^{n,u}_{T_\mu})\mathcal{Y}^{n,u}_{T_\mu}\right] \end{aligned} \tag{1.23}$$

In the following corollary and lemma, we utilize the previous results on the convergence of the forward schemes corresponding to our model dynamics and the sensitivity equation.

Corollary 1.10. *There exists a constant $D_{K_J} > 0$, such that for every $u \in \mathcal{U}$, we have that*

$$|J(u) - J^n(u)| \le D_{K_J} \delta_{1,4}^{(2-p)\wedge\frac{1}{2}},$$

where the constant $D_{K_J} > 0$ is independent of u and n. Under the assumption (HA), there exists a constant $\tilde{D}_{K_J} > 0$ such that

$$|J(u) - J^n(u)| \le \tilde{D}_{K_J} \delta_2^{(2H-1)\wedge(\frac{1}{2})}.$$

Proof. See Corollary 4.21. □

Lemma 1.11. *There exists a constant $D_{K_{\nabla J}} > 0$, such that for every $u \in \mathcal{U}$, we have that*

$$|\nabla J(u) - (\nabla J)^n(u)| \le D_{K_{\nabla J}} \delta_{1,8}^{(2-p)\wedge\frac{1}{2}},$$

where the constant $D_{K_{\nabla J}} > 0$ is independent of u and n. Under the assumption (HA), we obtain

$$|\nabla J(u) - (\nabla J)^n(u)| \le \tilde{D}_{K_{\nabla J}} \delta_2^{(2H-1)\wedge\frac{1}{2}}.$$

Proof. See Lemma 4.22. □

Now we want to find a way to include the discretized solution of the adjoint equation given by (1.22) into the calculation of the gradient ∇J. We reformulate the representation of $(\nabla J)^n$ such that it contains Λ^n and get a discretized version of (1.19).

Lemma 1.12. *For every $u \in \mathcal{U}$, the discretized gradient $(\nabla J)^n(u)$ (see (1.23)) can be represented by*

$$(\nabla J)^n(u) = \mathrm{E}\left[\Lambda_0^n D\mathcal{X}_0^u + \sum_{i=0}^{n-1} \Lambda_{t_{i+1}}^n \eta_{t_i,t_{i+1}}^u\right],$$

where

$$\begin{aligned}\eta_{t_i,t_{i+1}}^u := &\begin{pmatrix} b_u(t_i, \xi_{t_i}^n, u) \\ \hat{b}_u(t_i, x_{t_i}^n, \xi_{t_i}^n, u) \end{pmatrix} (t_{i+1} - t_i) + \sum_{j=1}^{m_1} \begin{pmatrix} \sigma_u^j(t_i, \xi_{t_i}^n, u) \\ 0 \end{pmatrix} (w_{t_{i+1}}^j - w_{t_i}^j) \\ &+ \sum_{j=1}^{m_2} \begin{pmatrix} 0 \\ \hat{\sigma}_u^j(t_i, x_{t_i}^n, \xi_{t_i}^n, u) \end{pmatrix} (B_{t_{i+1}}^j - B_{t_i}^j)\end{aligned}$$

for all $i \in \{0, \ldots, n-1\}$.

Proof. See Lemma 4.23. $\square$

To show that our theoretical results can be used in practice, we consider a practical example in the next section.

1.5 Numerical experiment

We showed, that we can approximate the cost function and its gradient with respect to the parameter using the established discretization schemes. Since for the calculation of the discretized cost function and the discretized gradient we need to take expected values, we will use the Monte-Carlo method. A comprehensive introduction to Monte-Carlo methods is given by Glasserman [2013]. Using random number generators, we can simulate for every $i = 0, \ldots, n$ and $j = 1, \ldots, m_1$ realizations of i.i.d random variables $(w_{t_i}^{j,a})_{a=1,\ldots,A}$ and for every $i = 0, \ldots, n$ and $j = 1, \ldots, m_2$ realizations of i.i.d random variables $(B_{t_i}^{j,a})_{a=1,\ldots,A}$ such that $(w_{t_i}^{j,a})_{a=1,\ldots,A}$, respectively $(B_{t_i}^{j,a})_{a=1,\ldots,A}$ have the same distribution as our driving processes w^j respectively B_j at time t_i. Using these copies of w_{t_i} and B_{t_i} we get for every $u \in \mathcal{U}$ i.i.d copies of paths of the discretization schemes $\mathcal{X}^{n,u}$ and $\mathcal{Y}^{n,u}$, denoted by $\mathcal{X}^{n,u,a}$ and $\mathcal{Y}^{n,u,a}$. By the strong law of large numbers we then have, that the Monte-Carlo estimators

$$\frac{1}{A}\sum_{a=1}^{A} \mathcal{X}_{t_i}^{n,u,a} \quad \text{, respectively} \quad \frac{1}{A}\sum_{a=1}^{A} \mathcal{Y}_{t_i}^{n,u,a}$$

converge P-a.s. to the expected values

$$\mathrm{E}[\mathcal{X}_{t_i}^{n,u}] \text{ , respectively } \mathrm{E}[\mathcal{Y}_{t_i}^{n,u}],$$

for every $i = 0, \ldots, n$. This ideas directly translate to the approximation of the distretized cost function and the discretized gradient by the Monte-Carlo estimators

$$\begin{aligned} J^{n,A}(u) &= \frac{1}{2} \sum_{\mu=1}^{M} \left(\frac{1}{A} \sum_{a=1}^{A} g_\mu(\mathcal{X}_{T_\mu}^{n,u,a}) \right)^2 \\ (\nabla J)^{n,A}(u) &= \sum_{\mu=1}^{M} \left(\frac{1}{A} \sum_{a=1}^{A} g_\mu(\mathcal{X}_{T_\mu}^{n,u,a}) \right) \left(\frac{1}{A} \sum_{a=1}^{A} g'_\mu(\mathcal{X}_{T_\mu}^{n,u,a}) \mathcal{Y}_{T_\mu}^{n,u,a} \right). \end{aligned} \tag{1.24}$$

Using the central limit theorem, it is well established that the corresponding approximation error behaves asymptotically like $\mathcal{O}(A^{-\frac{1}{2}})$, see Glasserman [2013]. Hence, combining these considerations with our discretization results from Proposition 1.10 and Lemma 1.11, there exist constants D_1 and D_2, such that under the assumption (HA), we have

$$\mathrm{E}\left[\left| J(u) - J^{n,A}(u) \right| \right] \leq D_1 \left(\frac{1}{A} \right)^{\frac{1}{2}} + D_2 \delta_2^{(2H-1)\wedge\frac{1}{2}}$$

and constants D_3, D_4, such that

$$\mathrm{E}\left[\left| \nabla J(u) - (\nabla J)^{n,A}(u) \right| \right] \leq D_3 \left(\frac{1}{A} \right)^{\frac{1}{2}} + D_4 \delta_2^{(2H-1)\wedge\frac{1}{2}}. \tag{1.25}$$

In Lemma 1.12, we established another representation of $(\nabla J)^n$, utilizing the discretized adjoint equation (1.22). To use the adjoint equation numerically, we need to be able to use its Monte-Carlo paths to approximate the gradient ∇J. We define the Monte-Carlo paths of the discretized adjoint equation by

$$\Lambda_{t_i}^{n,a} = (\lambda_{t_i}^{n,a}, \hat{\lambda}_{t_i}^{n,a}) = \Lambda_{t_{i+1}}^{n,a} \left(I_{n_1+n_2} + \eta_{t_i,t_{i+1}}^a \right) + \sum_{T_\mu = t_i} \left(\frac{1}{A} \sum_{a=1}^{A} g_\mu(\mathcal{X}_{T_\mu}^{n,a}) \right) g'_\mu(\mathcal{X}_{T_\mu}^{n,a}) \in \mathbb{R}^{n_1+n_2},$$

where

$$\begin{aligned} \eta_{t_i,t_{i+1}}^a = &\begin{pmatrix} b_x(t_i, \xi_{t_i}^{n,a}, u) & 0 \\ \hat{b}_z(t_i, x_{t_i}^{n,a}, \xi_{t_i}^{n,a}, u) & \hat{b}_x(t_i, x_{t_i}^{n,a}, \xi_{t_i}^{n,a}) \end{pmatrix} (t_{i+1} - t_i) + \sum_{j=1}^{m_1} \begin{pmatrix} \sigma_x^j(t_i, \xi_{t_i}^{n,a}, u) & 0 \\ 0 & 0 \end{pmatrix} (w_{t_{i+1}}^{j,a} - w_{t_i}^{j,a}) \\ &+ \sum_{j=1}^{m_2} \begin{pmatrix} 0 & 0 \\ \hat{\sigma}_z^j(t_i, x_{t_i}^{n,a}, \xi_{t_i}^{n,a}, u) & \hat{\sigma}_x^j(t_i, x_{t_i}^{n,a}, \xi_{t_i}^{n,a}, u) \end{pmatrix} (B_{t_{i+1}}^{j,a} - B_{t_i}^{j,a}) \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)} \end{aligned}$$

for all $i \in \{0, \ldots, n-1\}$ and

$$\Lambda_T^{n,a} = \sum_{T_\mu = T} \left(\frac{1}{A} \sum_{a=1}^{A} g_\mu(\mathcal{X}_T^{n,a}) \right) g'_\mu(\mathcal{X}_T^{n,a}).$$

Now we could show that for every $u \in \mathcal{U}$ the estimator

$$\frac{1}{A} \sum_{a=1}^{A} \left(\Lambda_0^{n,a} D\mathcal{X}_0^u + \sum_{i=0}^{n-1} \Lambda_{t_{i+1}}^{n,a} \eta_{t_i,t_{i+1}}^{u,a} \right), \tag{1.26}$$

where

$$\begin{aligned}\eta_{t_i,t_{i+1}}^{u,a} := &\begin{pmatrix} b_u(t_i, \xi_{t_i}^{n,a}, u) \\ \hat{b}_u(t_i, x_{t_i}^{n,a}, \xi_{t_i}^{n,a}, u) \end{pmatrix} (t_{i+1} - t_i) + \sum_{j=1}^{m_1} \begin{pmatrix} \sigma_u(t_i, \xi_{t_i}^{n,a}, u) \\ 0 \end{pmatrix} (w_{t_{i+1}}^{j,a} - w_{t_i}^{j,a}) \\ &+ \sum_{j=1}^{m_2} \begin{pmatrix} 0 \\ \hat{\sigma}(t_i, x_{t_i}^{n,a}, \xi_{t_i}^{n,a}, u) \end{pmatrix} (B_{t_{i+1}}^{j,a} - B_{t_i}^{j,a})\end{aligned}$$

for all $i \in \{0, \ldots, n-1\}$, approximates the discrete gradient representation $(\nabla J)^n$ from Lemma 1.12. Instead we will prove that for every $u \in \mathcal{U}$, this estimator (1.26) is P-a.s. equal to the estimator

$$(\nabla J)^{n,A}(u) = \sum_{\mu=1}^{M} \left(\frac{1}{A} \sum_{a=1}^{A} g_\mu(\mathcal{X}_{T_\mu}^{n,u,a}) \right) \left(\frac{1}{A} \sum_{a=1}^{A} g'_\mu(\mathcal{X}_{T_\mu}^{n,u,a}) \mathcal{Y}_{T_\mu}^{n,u,a} \right).$$

This directly implies the convergence of the estimator (1.26) to the gradient $\nabla J(u)$ with respect to the number of subintervals n of the Euler partition Π^{E} and the number of Monte-Carlo samples A.

Corollary 1.13. *For every $u \in \mathcal{U}$ and we have*

$$(\nabla J)^{n,A}(u) = \frac{1}{A} \sum_{a=1}^{A} \left(\Lambda_{t_0}^{n,a} D\mathcal{X}_0^u + \sum_{i=n_0}^{n-1} \Lambda_{t_{i+1}}^{n,a} \eta_{t_i,t_{i+1}}^{u,a} \right).$$

Proof. We can repeat the same steps as in the proof of Lemma 4.23, where we exchange the processes $\mathcal{X}^n$, $\mathcal{Y}^n$, Λ^n by $\mathcal{X}^{n,a}$,$\mathcal{Y}^{n,a}$ and $\Lambda^{n,a}$, consider the sum

$$\sum_{i=0}^{n-1} \mathcal{Y}_{t_{i+1}}^{n,a}$$

instead of

$$\sum_{i=0}^{n-1} \mathcal{Y}_{t_{i+1}}^{n}$$

for every $a = 1, \ldots, A$ and take arithmetic means with respect to a instead of expected values. $\square$

This corollary states, that we get exactly the same result when realizing the random Monte-Carlo paths and calculate the gradient via the discrete sensitivity equation (1.24) or the adjoint method (1.26). This also shows that our "optimize then discretize" approach coincides with the "discretize then optimize" approach in this setting. If we had started with the Monte-Carlo paths of Euler discretization scheme $\mathcal{X}^{n,a}$ and calculated the derivatives of the recursions with respect to the parameter, we would have get the recursion corresponding to the Monte-Carlo paths of the discrete Euler scheme $\mathcal{Y}^{n,a}$. By Corollary 1.13, we would have obtained the discrete adjoint equation and the adjoint gradient representation directly. Taking these considerations into account, we can translate our calibration problem (P^n) (see (4.78)) to the discretized Monte-Carlo optimization problem

$$(\mathrm{P}^{n,A})\ \text{Find} \qquad \min_{u\in\mathcal{U}} J^{n,A}(u) = \min_{u\in\mathcal{U}} \frac{1}{2}\sum_{\mu=1}^{M}\left(\frac{1}{A}\sum_{a=1}^{A} g_\mu(\mathcal{X}^{n,a,u}_{T_\mu})\right)^2$$

subject to

$$\begin{aligned}
\mathcal{X}^{n,u,a}_{t_{i+1}} &= \begin{pmatrix}\xi^{n,u,a}_{t_{i+1}}\\ x^{n,u,a}_{t_{i+1}}\end{pmatrix} = \begin{pmatrix}\xi^{n,u,a}_{t_i}\\ x^{n,u,a}_{t_i}\end{pmatrix} + \begin{pmatrix} b(t_i,\xi^{n,u,a}_{t_i},u)\\ \hat{b}(t_i,x^{n,u,a}_{t_i},\xi^{n,u,a}_{t_i},u)\end{pmatrix}(t_{i+1}-t_i)\\
&+\sum_{j=1}^{m_1}\begin{pmatrix}\sigma^j(t_i,\xi^{n,u,a}_{t_i},u)\\ 0\end{pmatrix}(w^{j,a}_{t_{i+1}}-w^{j,a}_{t_i}) + \sum_{j=1}^{m_2}\begin{pmatrix}0\\ \hat{\sigma}^j(t_i,x^{n,u,a}_{t_i},\xi^{n,u,a}_{t_i},u)\end{pmatrix}(B^{j,a}_{t_{i+1}}-B^{j,a}_{t_i}),\\
\mathcal{X}^{n,u,a}_{t_{n_0}} &= \mathcal{X}_0,\ i=0,\dots,n-1,\ a=1,\dots,A.
\end{aligned}$$

For the computation we can just simulate realizations of the discretized Monte-Carlo paths $\mathcal{X}^{n,u,a}$ for all $a = 1,\dots,A$ once up to time $T_M = t_n = T$ and extract the values at times T_μ for all $\mu = 1,\dots,M$ to calculate the cost function. As suggested in Käbe et al. [2009], we store the increments of the processes w^j and B^j for every Monte-Carlo path to facilitate the computation. Since we are now able to approximate the value of the cost function, as well as its gradient with respect to the parameter, we are able to use smooth gradient-based optimization algorithms to find the minimum of the cost function. We summarize the techniques to calculate the needed gradient.

Finite Differences:
Probably the simplest way to calculate the gradient is to calculate the finite differences

$$\left(\frac{\partial J}{\partial u_i}\right)^{n,A}(u) = (\nabla J)^{n,A}(u)e_i \approx \frac{J^{n,A}(u+he_i)-J^{n,A}(u)}{h}e_i,$$

where e_i is the i-th unit column vector for $i = 1,\dots,d$. But especially in the Monte-Carlo framework this methods has multiple disadvantages. First of all it leads to a substantial computational cost, because of the d additional calculations of $J^{n,A}$ in every iteration of the optimization. Furthermore the right choice of the parameter h is essential.

Sensitivity equation:
The second method is to utilize the discretized sensitivity equation (1.21) in the discretized gradient (1.24). We generate A paths of the Euler schemes $\mathcal{X}^{n,u}$ and $\mathcal{Y}^{n,u}$ and get

$$(\nabla J)^{n,A}(u) = \sum_{\mu=1}^{M}\left(\sum_{a=1}^{A} g_\mu(\mathcal{X}_{T_\mu}^{n,a,u})\right)\left(\sum_{a=1}^{A} g'_\mu(\mathcal{X}_{T_\mu}^{n,a,u})\mathcal{Y}_{T_\mu}^{n,a,u}\right).$$

For the numerical computation we would need to solve the recursions for $X^{n,a} \in \mathbb{R}^{n_1+n_2}$ and the recursion for $Y^{n,a} \in \mathbb{R}^{(n_1+n_2)\times d}$ up to time T for each of the A Monte-Carlo paths. The computational cost grows linearly in the number of parameters.

Adjoint equation:
We showed that the Monte-Carlo estimator for the discretized gradient, given by (1.23), has a second representation,

$$(\nabla J)^{n,A}(u) = \frac{1}{A}\sum_{a=1}^{A}\left(\Lambda_{t_0}^{n,a} D\mathcal{X}_0^u + \sum_{i=n_0}^{n-1} \Lambda_{t_{i+1}}^{n,a}\eta_{t_i,t_{i+1}}^{u,a}\right),$$

utilizing the discrete adjoint equation (1.22). The adjoint method gives another possibility to calculate the gradient of the cost function with the advantage that instead of $(n_1+n_2)\times d$ forward solves of the recursion for Y, we have n_1+n_2 backward solves. So the calculation of the gradient does not depend on the number of parameters. Especially in the case of time dependent parameters this reduces the numerical effort substantially in comparison the other two mentioned methods.
Now we will use these methods on an short explicit example.

1.5.1 Case study: A fractional Heston-type model

To use our theoretical results from previous chapters to calibrate a model with volatility driven by process of finite p-variation for $p \in (1,2)$, we first need to define a model which suits our conditions and ensure the market we are trading in is arbitrage free to have the risk neutral pricing formula. There are several models which incorporate the long memory phenomenon of volatility, by using a fractional Brownian motion with Hurst parameter $H \in (0.5,1)$ as driving process for the volatility, i.e. Comte and Renault [1998], Chronopoulou and Viens [2012], Bezborodov et al. [2019], Mishura and Yurchenko-Tytarenko [2020], Lépinette and Mehrdoust [2016]. For our volatility process we would like to consider a fractional version of the Cox-Ingersoll-Ross model given by

$$v_t = v_0 + \int_0^t \kappa(\theta - v_r)\,dr + \int_0^t \zeta\sqrt{v_r}\,dB_r^H,$$

where B_t^H is a fractional Brownian motion with Hurst parameter $H \in (0.5,1)$. It was shown by Lépinette and Mehrdoust [2016] and also by Mishura and Yurchenko-Tytarenko [2020] that this equation has a unique positive solution, where the integral $\int_0^t \zeta\sqrt{v_r}\, dB_r^H$ is given by a pathwise Young integral. Furthermore in Lépinette and Mehrdoust [2016], the authors show that the process v_t is mean reverting to the parameter θ, hence the parameters can be interpreted similarly to the standard CIR model by

$$
\begin{aligned}
v_0 &= \text{volatility at time 0} \\
\theta &= \text{ long term mean} \\
\kappa &= \text{ rate of return to the long term mean} \\
\zeta &= \text{ volatility of volatility.}
\end{aligned}
$$

Another feature we want to incorporate is the correlation between the volatility process and the asset price process S_t. So fix $T > 0$ and let $(\Omega, \mathcal{F}, (\mathbb{F}_t)_{t\in[0,T]}, P)$ be a filtered probability space carrying two independent standard Brownian motions B_t^1, B_t^2 and a fractional Brownian motion B_t^H with Hurst parameter $H \in (0.5, 1)$, where $\mathbb{F}$ is the filtration generated by B_t^1 and B_t^2, and B_t^H has the integral representation (see Hu [2005])

$$
B_t^H = C_H \left(\int_0^t (t-u)^{H-\frac{1}{2}}\, dB_u^1 + \int_{-\infty}^0 (t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}}\, dB_u^1 \right),
$$

where

$$
C_H = \sqrt{\frac{2H\gamma(\frac{3}{2}-H)}{\gamma(H+\frac{1}{2})\gamma(2-2H)}}.
$$

By defining

$$
B_t = \rho B_r^1 + \sqrt{1-\rho^2} B_r^2,
$$

we obtain a standard Brownian motion B_t, which is correlated with B_1 in the following way

$$
Corr(B_t, B_t^1) = \rho
$$

for all $t \in [0, T]$. Note that ρ is not the correlation between B^H and the Brownian motion driving the asset price process B, but between B and the Brownian motion B_1 from the integral representation of B^H. This way we generate the desired correlation between the volatility process v and the asset price S of the following model, in a similar way as in Mishura and Yurchenko-Tytarenko [2020]. The fractional Heston-type model we would like to consider is given by the dynamics

$$
\begin{aligned}
v_t &= v_0 + \int_0^t \kappa(\theta - v_s)\, ds + \int_0^t \zeta\sqrt{v_s}\, dB_s^H \qquad (1.27) \\
S_t &= S_0 + \int_0^t (r-d)\, ds + \int_0^t \sqrt{v_s} S_s\, d(\rho B_s^1 + \sqrt{1-\rho^2} B_s^2),
\end{aligned}
$$

where we have the spot price S_0, the riskless rate r and the dividend yield d. We assume the market we are trading in, only consist of the asset S and a riskless bond e^{-rt} for $t \in [0, T]$. We would like to calibrate the model with respect to the parameters $u = (v_0, \kappa, \theta, \zeta, \rho)$ to a set of market observed European call option prices. The careful reader probably noticed some difficulties with the problem formulation above. Taking the previous sections into account, we are not able to use our framework on this problem because crucial conditions are not satisfied.

- The square root function in the integrand of the B^H integral is only Hölder continuous of order $\frac{1}{2}$. Our approximation results demand global Lipschitz continuity and differentiability in the state variable v_t. Furthermore the coefficients of the equation (1.27) need to be bounded.

- The disctretized volatility process v_t^n can become negative, since the increments of B^H are normally distributed.

- The coefficients of the equation S need to be continuously differentiable with bounded derivatives in all variables.

- The functions g_μ are not differentiable, because of the maximum function.

To account for all these problems, we take an approach similar to Käbe et al. [2009]. We adjust our dynamics by using a polynomial error function π_1 to smooth out the function $\pi(x) = \max(x, 0)$ at 0 and another function π_2 which smooths out the square root function at 0. The functions are given by

$$\pi_1(x) = \begin{cases} 0 & , x < -\varepsilon_1 \\ -\frac{1}{16(\varepsilon_1)^3}x^4 + \frac{3}{8\varepsilon_1}x^2 + \frac{1}{2}x + \frac{3\varepsilon_1}{16} & , -\varepsilon_1 \leq x \leq \varepsilon_1 \\ x & , x > \varepsilon_1 \end{cases}$$

for $x \in \mathbb{R}$ and an error parameter $\varepsilon_1 > 0$ which we choose to be 0.01 for all calculations. the second function is given by

$$\pi_2(x) = \begin{cases} 0 & , x < -\varepsilon_2 \\ -\frac{1}{256\varepsilon_2^{6,5}}(-15x^7 + 7\varepsilon_2 x^6 + 65\varepsilon_2^2 x^5 - 33\varepsilon_2^3 x^4 - 117\varepsilon_2^4 x^3 & \\ +77\varepsilon_2^5 x^2 + 195\varepsilon_2^6 x + 77\varepsilon_2^7) & , -\varepsilon_2 \leq x \leq \varepsilon_2 \\ \sqrt{x} & , x > \varepsilon_2 \end{cases}$$

for $x \in \mathbb{R}$ and an error parameter $\varepsilon_2 > 0$ which we choose to be 0.001 for all calculations For the upper bound of these 2 functions we construct theoretically a polynomial function which differentiably truncates the function at a truncation value Ξ. In practice we choose this truncation value so big, that it will not be needed, since v_t is mean reverting and for our moderate time horizon we do not expect the asset price in our model to explode. This is justified by our numerical findings. The dynamics of the adjusted fractional Heston-type model are given by

$$\begin{aligned} v_t &= v_0 + \int_0^t \kappa(\theta - \pi_1(v_s))\, ds + \int_0^t \zeta\pi_2(v_s)\, dB_s^H \\ S_t &= S_0 + \int_0^t (r-d)S_s\, ds + \int_0^t \pi_2(v_s)S_s\, d(\rho B_s^1 + \sqrt{1-\rho^2}B_s^2) \end{aligned} \tag{1.28}$$

and after a log transformation $\hat{S}_t = \log(S_t)$ in the asset equation, this yields

$$\begin{aligned} v_t &= v_0 + \int_0^t \kappa(\theta - \pi_1(v_s))\, ds + \int_0^t \zeta\pi_2(v_s)\, dB_s^H \\ \hat{S}_t &= \hat{S}_0 + \int_0^t (r-d) - \frac{1}{2}\pi_2(v_s)^2\, ds + \int_0^t \pi_2(v_s)\, d(\rho B_s^1 + \sqrt{1-\rho^2}B_s^2). \end{aligned}$$

Note that under these adjustments $\pi_2(v_r)$ is bounded and hence the SDE (1.28) has the explicit solution

$$S_t = S_0 e^{\left((r-d)t - \frac{1}{2}\int_0^t \pi_2(v_s)^2\, ds + \int_0^t \pi_2(v_s)\, dB_s\right)},$$

where $B_r = \rho B_r^1 + \sqrt{1-\rho^2}B_r^2$. For a dividend paying asset S_t the discounted price process $e^{-rt}S_t = e^{-rt}e^{\hat{S}_t}$ is then a martingale with respect to P and the price for a call option with maturity T_μ and Strike K_μ at time 0 in this model is given by

$$e^{-rT_\mu}\mathrm{E}\left[\pi\left(e^{\hat{S}^u_{T_\mu}} - K_\mu\right)\right]$$

by the risk neutral pricing formula. We approximate this value by

$$C_\mu^{mod}(u) = e^{-rT_\mu}\mathrm{E}\left[\pi_1\left(e^{\hat{S}^u_{T_\mu}} - K_\mu\right)\right]$$

and the cost function translates to

$$J(u) = \frac{1}{2}\sum_{\mu=1}^{M} \mathrm{E}\left[g_\mu\begin{pmatrix} v^u_{T_\mu} \\ \hat{S}^u_{T_\mu}\end{pmatrix}\right]^2 = \frac{1}{2}\sum_{\mu=1}^{M}\left(C_\mu^{mod}(u) - C_\mu^{obs}\right)^2,$$

where $g_\mu(x_1, x_2) = e^{-rT_\mu}\pi_1(e^{x_2} - K_\mu) - C_\mu^{obs}$. Putting this into the framework of Chapter 3, we obtain for a bounded, open and convex subset $\mathcal{U}$ of $\mathbb{R}^5$ the functions $b : [0,T] \times \mathbb{R} \times \mathcal{U} \to \mathbb{R}$ and $\sigma : [0,T] \times \mathbb{R} \times \mathcal{U} \to \mathbb{R}$ such that

$$\begin{aligned} b(t,x,u) &= u_2(u_3 - \pi_1(x)) & \sigma(t,x,u) &= \zeta\pi_2(x) \\ b_x(t,x,u) &= -u_2\pi_1'(x) & \sigma_x(t,x,u) &= \zeta\pi_2'(x) \\ b_u(t,x,u) &= (0, u_3 - \pi_1(x), u_2, 0, 0) & \sigma_u(t,x,u) &= (0,0,0,\pi_2(x),0) \end{aligned}$$

and the functions $\hat{b} : [0,T]\times\mathbb{R}\times\mathbb{R}\times\mathcal{U} \to \mathbb{R}$, $\hat{\sigma}^1 : [0,T]\times\mathbb{R}\times\mathbb{R}\times\mathcal{U} \to \mathbb{R}$ and $\hat{\sigma}^2 : [0,T]\times\mathbb{R}\times\mathbb{R}\times\mathcal{U} \to \mathbb{R}$ such that

$$\begin{array}{ll} \hat{b}(t,x,z,u) = (r-d) - \frac{1}{2}\pi_2(z) & \hat{\sigma}^1(t,x,z,u) = \pi_2(z)\rho \\ \hat{b}_x(t,x,z,u) = 0 & \hat{\sigma}^1_x(t,x,z,u) = 0 \\ \hat{b}_z(t,x,z,u) = -\frac{1}{2}\pi_2'(z) & \hat{\sigma}^1_z(t,x,z,u) = \pi_2'(z)\rho \\ \hat{b}_u(t,x,z,u) = (0,0,0,0,0) & \hat{\sigma}^1_u(t,x,z,u) = (0,0,0,0,\pi_2(z)) \end{array}.$$

$$\begin{array}{l} \hat{\sigma}^2(t,x,z,u) = \pi_2(z)\sqrt{(1-\rho^2)} \\ \hat{\sigma}^2_x(t,x,z,u) = 0 \\ \hat{\sigma}^2_z(t,x,z,u) = \pi_2'(z)\sqrt{1-\rho^2} \\ \hat{\sigma}^2_u(t,x,z,u) = \left(0,0,0,0,-\pi_2(z)\frac{\rho}{\sqrt{1-\rho^2}}\right) \end{array}.$$

Note that for the fractional Brownian motion with Hurst parameter $H > \frac{1}{2}$, we know that for every $H' < H$ (see Nualart [1995], p.274) condition (HA) is satisfied and by the Jain Monrad criterion (see Theorem 2.38) also the exponential moment condition (1.5) is satisfied for $p = \frac{1}{H'}$. We choose $H' = H - \epsilon$, for ϵ close to zero and $p = \frac{1}{H'}$. Furthermore, all the other conditions $(H_1), (H_2), (H_3^*), (B_1), (B_2), (B_3), (E_1), (E_2), (G)$ are also satisfied. Now we can make use of the results at the beginning of this section to calculate numerically the cost function and its gradient with the 3 different methods described. We use the gradients for a gradient-based optimization algorithm to calibrate our model to the data set of observed call option prices from Andersen and Brotherton-Ratcliffe [1998], that is shown in Table 1.1.

Table 1.1: Implied volatilities on the S&P 500 index with interest rate 0.06, continuous dividend yield 0.026 and spot price 590.

Mat/Str	501.5	531	560.5	590	619.5	649	678.5	708	767	826
0.175	0.190	0.168	0.133	0.113	0.102	0.097	0.120	0.142	0.169	0.200
0.425	0.177	0.155	0.138	0.125	0.109	0.103	0.100	0.114	0.130	0.150
0.695	0.172	0.157	0.144	0.133	0.118	0.104	0.100	0.101	0.108	0.124
0.940	0.171	0.159	0.149	0.137	0.127	0.113	0.106	0.103	0.100	0.110
1	0.171	0.159	0.150	0.138	0.128	0.115	0.107	0.103	0.099	0.108
1.5	0.169	0.160	0.151	0.142	0.133	0.124	0.119	0.113	0.107	0.102
2	0.169	0.161	0.153	0.145	0.137	0.130	0.126	0.119	0.115	0.111
3	0.168	0.161	0.155	0.149	0.143	0.137	0.133	0.128	0.124	0.123
4	0.168	0.162	0.157	0.152	0.148	0.143	0.139	0.135	0.130	0.128
5	0.168	0.164	0.159	0.154	0.151	0.148	0.144	0.140	0.136	0.132

We calibrate the model first for the 5 parameters $(v_0, \kappa, \theta, \zeta, \rho)$, where we choose a closed, convex subset of $\mathcal{U}$ given by

$$\hat{\mathcal{U}} := \{(v_0, \kappa, \theta, \zeta, \rho) \in \mathbb{R}^5 |\, v_0 \in [0.0001, 1],\ \kappa \in [0.0001, 2],\ \theta \in [0.0001, 2],\ \zeta \in [0.0001, 4],$$

$\rho \in [-0.99, 0.99]\}$

as our bounds for the calibration and choose different values for the Hurst parameter H. The calibration is done on two different layers, to speed up the process. We first use a small number of Monte-Carlo paths $A = 10000$ and a partition Π^{E} of $[0, 5]$ which contains the maturities of the observed option prices by equidistantly dividing $[0, 1]$ in subintervals of length 0.005 and the interval $[1, 5]$ equidistantly into subintervals of length 0.05. Hence we get $A = 10000$ and $n = 280$ with $|\Pi^{\mathrm{E}}| = 0.05$. Then we simulate the Brownian and fractional Brownian increments with the assumed correlation by using the Matlab function mvnrnd and store the increments for the whole optimization on this layer. Then we optimize our cost function, assuming the prices are normalized to $S_0 = 1$, using the Matlab fmincon function with the trust region reflective algorithm and a function tolerance of 10^{-6}, where the gradient of the cost function is calculated via the adjoint method, starting at the parameter values

$$v_0 = 0.1,\ \kappa = 1,\ \theta = 0.05\ , \zeta = 0.3,\ \rho = -0.7.$$

The found parameters are then used as starting value for the same procedure but on a finer layer, namely $M = 100000$ and $n = 560$ by cutting the length of the subintervals of the coarser layer in half. We repeat the optimization for the values $H \in \{0.6, 0.7, 0.8, 0.9\}$ 20 times. The results show the mean (μ) and standard deviation (Sd) of each parameter for each value of H, the mean over the 20 iterations of the average error over the 100 option prices avgErr $= \frac{1}{20}\sum_{j=1}^{20}\frac{1}{100}\sum_{\mu=1}^{100}|C_\mu^{mod,j}(u^*) - C_\mu^{obs}|$ and the average runtime of the optimization. We show in table 1.2 the results.

Table 1.2: Calibration results for different values of $H \in (\frac{1}{2}, 1)$, when calibrating our model with 5 parameters to the call option prices from table 1.1.

H	v_0		κ		θ		ζ		ρ	
	μ	Sd	μ	Sd	μ	Sd	μ	Sd	μ	Sd
0.6	0.014	0.0003	1.371	0.1043	0.018	0.0006	0.286	0.0196	-0.640	0.0291
0.7	0.015	0.0004	1.184	0.0486	0.017	0.0007	0.243	0.0099	-0.669	0.0265
0.8	0.015	0.0004	1.197	0.0736	0.016	0.0005	0.241	0.0090	-0.710	0.0202
0.9	0.015	0.0005	1.259	0.1033	0.0166	0.0008	0.232	0.0079	-0.871	0.0537

H	AvgErr	runtime in sec
0.6	0.00059	393.79
0.7	0.00083	373.78
0.8	0.00106	386.08
0.9	0.00128	486.70

To illustrate the convergence result (1.25), we consider only option prices with the maturities $1, 2, 3, 4, 5$ from table 1.1 to be able to consider equidistant partitions, where the mesh is cut in half in every step. We choose the number of subintervals of our partition successively as $n_i = 2^i \cdot 5$ for $i = 4, \ldots 10$ and $A = 100000$ with parameters $v_0 = 0.016$, $\kappa = 1$, $\theta = 0.02$, $\zeta = 0.3$, $\rho = -0.7$. If (1.25) holds, then by the triangle inequality, there also exist constants $\hat{D}_3$, $\hat{D}_4$ such that

$$\mathrm{E}\left[\left|(\nabla J)^{0.5n_i,A}(u) - (\nabla J)^{n_i,A}(u)\right|\right] \leq \hat{D}_3 \left(\frac{1}{A}\right)^{\frac{1}{2}} + \hat{D}_4 \left(\frac{5}{n_i}\right)^{(2H'-1)\wedge\frac{1}{2}}.$$

Hence, we calculate $(\nabla J)^{n_i,A}(u)$ for the same Monte-Carlo paths of the fractional and standard Brownian motion and repeat this 20 times. Then calculating

$$Err_i = \frac{1}{20}\sum_{j=1}^{20}(\nabla J)^{0.5n_i,A,j}(u) - (\nabla J)^{n_i,A,j}(u)$$

for $i = 5, \ldots, 9$. We do this with paths of two different fractional Brownian motions with the Hurst parameters $H = 0.65$ and $H = 0.8$. In figures 1.1 and 1.2, we show the log log plots for $x = \{n_5, \ldots, n_9\}$ and $y = \{Err_5, \ldots, Err_9\}$ and adding a reference line with the slopes -0.3 and -0.5 fitted to the last data point, illustrating our theoretical findings for the convergence rate.

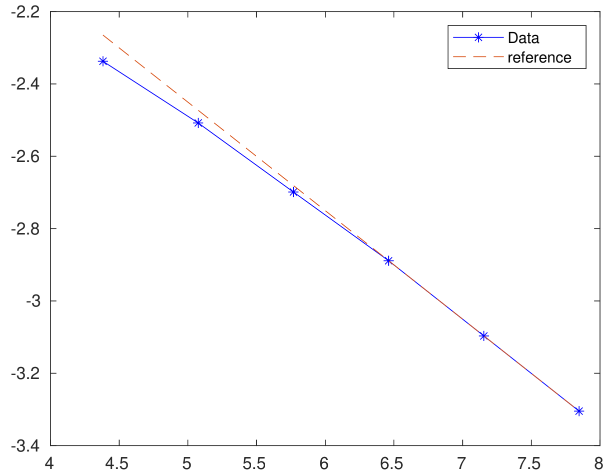

Figure 1.1: loglog plot with respect to n_i and the error Err_i for $i = 5, \ldots, 10$ and $H = 0.65$.

To compare the different methods of calculating the gradient, and to emphasize the advantages of the adjoint method, we introduce time dependent parameters $\kappa(t)$, $\theta(t)$, $\zeta(t)$ and $\rho(t)$ to our model, which we choose to be piecewise constant on intervals $(s_i, s_{i+1}]$, where $(s_i)_{i=0,\ldots,I}$ is a partition of $[0, T]$ such that $\{s_i\}_{i=1,\ldots,I} \subset \{T_\mu\}_{\mu=1,\ldots,10}$, where T_μ are the 10 maturities of the observed option prices. We show in table 1.3, the computing time of the gradient using the 3 methods described for a increasing numbers of parameters. We see that the computation time of the gradient calculation with the adjoint method stays almost constant for an increasing number of parameters, as expected. Using the sensitivity or the finite differences method for the calculation,

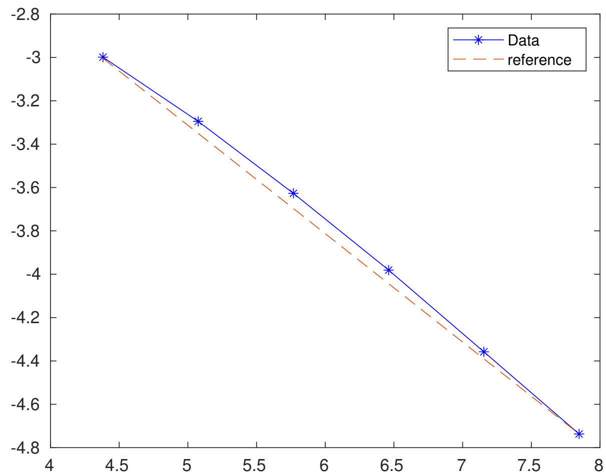

Figure 1.2: loglog plot with respect to n_i and the error Err_i for $i = 5, \ldots, 10$ and $H = 0.8$.

we see that the computation time increases linearily with the number of parameters. Hence our gradient calculation by using the adjoint model leads to a significant speed up of a Monte-Carlo-based calibration of our fractional stochastic volatility model. We emphasize that this numerical example is only supposed to serve as a proof of concept for the applicability of our theoretical results in practice. For sure, there are many ways to improve the performance of the method, e.g. by variance reduction techniques, parallelization and also the use of a more sophisticated optimization algorithm.

Table 1.3: Runtime (RT) of the calculation of the gradient of the cost function with the 3 different methods, for $A = 100000$, $n = 560$, and an increasing number of parameters.

Number of parameters	5	9	13	17	21	25	29	33	37	41
RT Adjoint	22.5	22.0	22.5	22.2	22.1	22.3	22.2	22.4	22.0	22.1
RT Sensitivity	26.1	30.1	35.1	39.8	44.7	49.0	53.2	57.5	61.8	66.0
RT Finite differences	51.4	85.8	120.2	156.4	191.1	226.9	261.1	293.4	327.9	362.7

1.6 Literature review

In this chapter we want to give an overview on existing literature which relates to our results. Starting this thesis we had a fractional stochastic volatility model in mind, where the driving process of the volatility (w in our notation) is a non semimartingale. We want to choose an appropriate path space which contains the fractional Brownian motion with Hurst parameter $H \in (\frac{1}{2}, 1)$, since it is the archtypical example of a long memory process used in financial modeling. Our first approach was to interpret our model as a special case of a mixed stochastic differential equation, where all equations are driven by both processes w and B. Such equations, where the driving process w is chosen as a fractional Brownian motion with Hurst parameter $H \in (\frac{1}{2}, 1)$ or more generally Hölder continuous processes with Hölder exponent $H' \in (\frac{1}{2}, 1)$ were considered by several authors in different settings, see Kubilius [2002], Guerra and Nualart [2008], Mishura [2008], Mishura and Shevchenko [2011b], Mishura and Shevchenko [2012], Shevchenko [2013]. Some of these results can not be applied to our situation, because they only consider the one-dimensional case (see Kubilius [2002]), or need the driving processes to be independent (see Guerra and Nualart [2008]), which we explicitly want to avoid. Another possibility to deal with existence and uniqueness comes from rough paths theory, introduced in Lyons [1998], which opens a whole new perspective on differential equations driven by rough processes, which even allows to consider fractional Brownian motion with Hurst parameter smaller than $\frac{1}{2}$. An existence and uniqueness result for such a mixed differential equation can be seen in Coutin and Qian [2002]. The drawback here is that again the driving processes need to be independent, which excludes this approach for our purposes. The results for existence and uniqueness with the weakest conditions is given in Shevchenko [2013] and these could be used to ensure existence and uniqueness for our system dynamics and also the linear equations, but under the assumption that w is Hölder continuous. Unfortunately we found no results concerning the differentiability of such equations with respect to a parameter, only results on continuity, see Mishura and Posashkova [2011]. Hence, we swapped from the more general setting of mixed stochastic differential equations to the setting presented in Section 1.1, which allows us to analyse the equations (1.1) and (1.2) successively. Since our focus are stochastic volatility models, where only the volatility process is driven by a non-semimartingale, this also seems the more natural approach. But this opens up the question which exact path properties we want from our driving process w and which solution spaces are appropriate for our volatility equation. If we first consider only a fractional Brownian motion with $H \in (\frac{1}{2}, 1)$ as driving process, this equation can be solved pathwise, because of the path properties of fractional Brownian motion having a.s. Hölder continuous paths with Hölder exponent in $(0, H)$. There seem to be essentially four ways of doing this pathwise approach, using Hölder norms (see Ruzmaikina [2000]), p-variation norms (see Lyons [1994], Dudley and Norvaiša [2010]), the fractional integration approach introduced by Zähle (see Zähle [1998], Nualart and Răşcanu [2002]), where the solutions are given in a Besov type space, or the rough paths approach we already mentioned (see e.g. Friz and Victoir [2010],Coutin and Qian [2002]). At the beginning of this thesis there was a recent publication Nguyen et al. [2018] on existence and uniqueness of time dependent, mutlidimensional differential equations driven by a continuous function of finite p-variation under similar conditions as in Nualart and Răşcanu [2002], which was perfect for our purposes. So we chose the results of Nguyen et al. [2018] together with Nguyen et al. [2020], considering linear Young differential equations, as basis for our pathwise considerations and developed all the properties like the continuity and differentiability with respect to the parameter of the solution mapping to equation (1.1) using their ideas of the greedy sequences. Concerning the Fréchet differentiability we used ideas from

Han et al. [2012], where they considered similar equations but using Hölder norms in a stochastic control setting. There were also result concerning the differentiability from rough paths theory, see Cass et al. [2013], where the Fréchet differentiability of the solution mapping with respect to the initial condition, for time independent coefficients was shown. By adding dimensions to the space variable, we could achieve the differentiability of our time and parameter dependent equation with respect to the parameter using their results, but under stronger conditions on the time variable as in our case. Concerning the Itô SDE (1.2) analysis, we use results from stochastic control theory given in Yong and Zhou [1999] and Yong [2019].

The main results in Chapter 3 incorporate the theory on Forward integration from Russo and Vallois, the references are given in this chapter. Adjoint methods have a long history. Originating in deterministic control theory, namely the Pontryagin maximum principle Boltyanski et al. [1960], it was then translated to the stochastic control setting in Bismut [1978]. An historical overview on results related to the maximum principle can be found in Yong and Zhou [1999]. Since then, adjoint methods to efficiently calculate gradients found many applications in various fields of research like meteorology Charpentier and Ghemires [2000], Lafore et al. [1998], optimal design Giles and Pierce [2000] or neural ODEs Chen et al. [2018], Zhuang et al. [2020]. Adjont methods in finance literature, were introduced in Giles and Glasserman [2006] to efficiently calculate option sensitivities, but found many applications related to finance, e.g. Capriotti [2011], Henrard [2013]. The key references, which motivated our work, were Käbe et al. [2009] and Käbe [2010], since we translate their adjoint approach for the calibration of the Heston stochastic volatility model, to the fractional stochastic volatility case. More recent publications, which are concerned with adjoint sensitivities for Stratonovich SDEs, are Li et al. [2019], Massaroli et al. [2021]. Concerning the adjoint sensitivities for SDEs driven by our specific model, we found no existing literature.

Coming to the first order Euler discretization results of Chapter 4, there are again different approaches to the topic, similar as for the existence and uniqueness result. We first focus on the discretization of the fractional SDE (1.1) driven by w. We choose to work in p-variation spaces considering this equation, but as it was for the existence and uniqueness result, every work considering the numerical approximation of fractional SDEs driven by a fractional Brownian motion with Hurst parameter $H > \frac{1}{2}$ could be of interest to us. The first articles concerning the discretization of fractional SDEs with fractional Brownian motion as driving noise by first order Euler schemes were Lin [1995], Nourdin [2005], Neuenkirch [2006], Nourdin and Neuenkirch [2007]. In the last of these references the authors prove that in the one-dimensional, time independent case the Euler scheme on a equidistant partition $(t_i)_{i=1,\dots,n}$ of $[0,1]$ convergences with a rate of n^{1-2H} pointwise a.s. to the solution of the equation. This rate is sharp in the sense that $n^{1-2H}|X_1^n - X_1|$ converges almost surely to a finite and nonzero limit. This result was then generalized by Mishura and Shevchenko [2008] to the multidimensional, time dependent case, showing that $n^{1-2H}\sup_{t\in[0,T]}|X_t^n - X_t|$ converges almost surely to a finite and nonzero limit. In Davie [2008] and Lejay [2010] the authors consider the first order Euler approximation of a time autonomous multidimensional differential equation in the deterministic setting with a continuous driving path of finite p-variation, where $p \in (1,2)$. The author in Lejay [2010] obtains the same rate of convergence as the previously mentioned works. Although the estimates are explicitly given only for a driftless equation, the setting is the closest to ours and we combine ideas for the estimates from Lejay [2010] with the greedy sequence ideas from Nguyen et al. [2018] to obtain our results. There are also results for higher order schemes, see e.g. Jamshidi and Kamrani [2021], Hu et al. [2016]. Or from the rough paths theory Deya et al. [2012], Araya et al. [2020], Friz and Victoir [2010], Davie [2008] for a fractional Brownian motion driver with Hurst parameter $H < \frac{1}{2}$. In this case, higher order schemes are needed as pointed out by an example in Deya et al. [2012]. Non of the aforementioned articles consider the convergence of discretization schemes of linear equations of the type (1.6). The only reference we found for similar linear equations was Chronopoulou and Tindel [2013], where the authors consider time autonomous equations in Hölder spaces driven by fractional Brownian motion. We already commented on their work in Section 1.4. For the convergence of the discretization schemes for the nonlinear and linear Itô equations (1.2) and (1.7), we use standard techniques like the Burkholder-Davis-Gundy inequality and the Gronwall inequality. The ideas for the proofs are standard, but taking the rather unusual setting of our model dynamics into account, we do the calculations rigorously. We refer the reader to Kloeden and Platen [2011] for an almost complete covering of the topic. Another approach comes from the theory of mixed differential equations we already mentioned, see Mishura and Shevchenko [2011a], Liu and Luo [2017] and Liu et al. [2020]. The authors in Mishura and Shevchenko [2011a] consider first order Euler schemes for a mixed differential equations driven by standard Brownian motion and a fractional Brownian motion with Hurst parameter $H > \frac{1}{2}$, but need the driving processes to be independent. In Liu and Luo [2017] the authors consider modified Euler schemes similar to Hu et al. [2016]. First order Euler schemes for mixed differential equations, where the drivers need not to be independent are considered in Liu et al. [2020]. The authors derive a convergence result of order $\mathcal{O}(\delta^{2H-1\wedge\frac{1}{2}})$ in probability in a Besov type space (which was introduced in Nualart and Răşcanu [2002]), where δ is the mesh of a partition of $[0,T]$. While their results could be used for the convergence of our Euler scheme X^n to the solution of equation (1.3), their conditions are not satisfied by the coefficients of our system of linear equations (1.8). We derive the same order of convergence but in L^l, $l \geq 1$, uniformly in time, for both the approximation of the non-linear model dynamics equation and the linear equation stemming from the Fréchet derivative with respect to the parameter of the model solution mapping. Considering the numerical example, we gave the corresponding references in Section 1.5.1.

Chapter 2

The model dynamics equation and its differentiability with respect to the parameter

The goal of this chapter is to introduce our model dynamics, state assumptions on the coefficients, which ensure the existence and uniqueness of the solution to our model dynamics equation and also the Fréchet differentiability of the corresponding solution mapping. We start by introducing the model dynamics. Let T be a positive constant and $n_1, m_1, n_2, m_2, d \in \mathbb{N} = \{1, 2, \dots\}$. Let $(\Omega, \mathcal{F}, \mathbb{F}, P)$ be a filtered probability space (satisfying the usual conditions) carrying an m_1-dimensional stochastic process $(w_t)_{t\in[0,T]}$, which paths are almost surely continuous and have finite p-variation for $p \in (1, 2)$ (see Subsection 2.1.1 for the definition of p-variation) and a m_2-dimensional standard Brownian motion $(B_t)_{t\in[0,T]}$, both adapted to the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$, possibly dependent. Furthermore let $\mathcal{U}$ be an open, convex and bounded subset of $\mathbb{R}^d$, which will be our parameter set. We consider the parameter dependent system of stochastic differential equations

$$\xi_t^u = \xi_0(u) + \int_0^t b(r, \xi_r^u, u)\, dr + \sum_{j=1}^{m_1} \int_0^t \sigma^j(r, \xi_r^u, u)\, dw_r^j, \tag{2.1}$$

$$x_t^u = x_0(u) + \int_0^t \hat{b}(r, x_r^u, \xi_r^u, u)\, dr + \sum_{j=1}^{m_2} \int_0^t \hat{\sigma}^j(r, x_r^u, \xi_r^u, u)\, dB_r^j, \tag{2.2}$$

where

$$\begin{aligned}
&\xi_0 : \mathcal{U} \to \mathbb{R}^{n_1},\\
&b : [0, T] \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_1},\\
&\sigma = (\sigma^1, \dots, \sigma^{m_1}) : [0, T] \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_1\times m_1}
\end{aligned}$$

and

$$\begin{aligned}
&x_0 : \mathcal{U} \to \mathbb{R}^{n_2}, \\
&\hat{b} : [0,T] \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_2}, \\
&\hat{\sigma} = (\hat{\sigma}^1, \dots, \hat{\sigma}^{m_2}) : [0,T] \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_2 \times m_2},
\end{aligned}$$

denoted in matrix form by

$$\begin{aligned}
\mathcal{X}_t^u = \begin{pmatrix} \xi_t^u \\ x_t^u \end{pmatrix} = \begin{pmatrix} \xi_0(u) \\ x_0(u) \end{pmatrix} + \int_0^t \begin{pmatrix} b(r, \xi_r^u, u) \\ \hat{b}(r, x_r^u, \xi_r^u, u) \end{pmatrix} dr + \sum_{j=1}^{m_1} \int_0^t \begin{pmatrix} \sigma^j(r, \xi_r^u, u) \\ 0 \end{pmatrix} dw_r^j \\
+ \sum_{j=1}^{m_2} \int_0^t \begin{pmatrix} 0 \\ \hat{\sigma}^j(r, x_r^u, \xi_r^u, u) \end{pmatrix} dB_r^j.
\end{aligned}$$

The form of the model dynamics, allows to consider the two equations (2.1) and (2.2) successively. We begin with the analysis of equation (2.1).

2.1 Parameter dependent Young differential equations

In this section we examine the stochastic (Young) differential equation given in (2.1). We cannot use the Itô calculus in this situation, because we only assume that the process w has paths that are almost surely continuous with finite p-variation for a given $p \in (1,2)$. A prominent example of such a process is the fractional Brownian motion with Hurst parameter $H > \frac{1}{2}$. Since the path properties of w allows to solve the equation (2.1) pathwise, we will develop the mathematical foundation of Young differential equations in the deterministic setting first. We start in Subsection 2.1.1 with the properties of functions with finite p-variation and an introduction of the Young integral.

2.1.1 Properties of p-variation and the Young integral

The definitions and properties of functions of finite p-variation in this subsection are mostly adopted from Dudley and Norvaiša [2010] and Friz and Victoir [2010], where the latter focus on continuous functions of finite p-variation. For a more detailed discussion on Young-integration, see Young [1936] and Dudley and Norvaiša [2010]. For the rest of this subsection let $n, m \in \mathbb{N}$, $T > 0$ a positive constant and $[s,t] \subset [0,T]$ intervals on the real line. We consider $(\mathbb{R}^{n \times m}, \|\cdot\|_F)$, where $\|\cdot\|_F$ is the Frobenius norm given by

$$\|x\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |x^{(i,j)}|^2 \right)^{\frac{1}{2}}.$$

For notational simplicity we write $|\cdot| = \|\cdot\|_F$ for the rest of the thesis. We define $\mathcal{P}([s,t]) := \{\Pi_k = (s = t_0 < t_1 < \cdots < t_k = t) | k \in \mathbb{N}\}$ as the set of all finite partitions of the interval $[s,t]$. For a partition Π_k we call $|\Pi_k| = \max_{i=0,\ldots,k-1}\{|t_{i+1} - t_i|\}$ the mesh of the partition and for $i = 0, \ldots, k-1$, we call $[t_i, t_{i+1}]$ the subintervals of the partition. If the number of subintervals of a partition does not need to be specified, we will omit the index k. Let $1 \le p < \infty$, the p-variation (distance, semi-norm) of a function $x : [s,t] \to \mathbb{R}^{n\times m}$ is then given by

$$|x|_{p,s,t} := \sup_{k\in\mathbb{N}, \Pi_k \in \mathcal{P}([s,t])} \left(\sum_{i=0}^{k-1} |x_{t_{i+1}} - x_{t_i}|^p \right)^{\frac{1}{p}}$$

and the corresponding p-variation norm

$$\|x\|_{p,s,t} := |x_s| + |x|_{p,s,t}.$$

Furthermore, we define the spaces

$$W^p([s,t], \mathbb{R}^{n\times m}) := \{x : [s,t] \to \mathbb{R}^{n\times m} |\, \|x\|_{p,s,t} < \infty\}$$

and

$$C^p([s,t], \mathbb{R}^{n\times m}) := \{x : [s,t] \to \mathbb{R}^{n\times m} |\, x \text{ is continuous and } \|x\|_{p,s,t} < \infty\}.$$

The uniform norm for a function x on $[s,t]$ is given by $\|x\|_{\infty,s,t} := \sup_{r\in[s,t]} |x_r|$. The obvious inequality

$$\|x\|_{\infty,s,t} \le |x_s| + |x|_{p,s,t} \tag{2.3}$$

shows that $C^p([s,t], \mathbb{R}^{n\times m})$ is a subspace of

$$C([s,t], \mathbb{R}^{n\times m}) := \{x : [s,t] \to \mathbb{R}^{n\times m} |\, x \text{ is continuous and } \|x\|_{\infty,s,t} < \infty\}$$

and all elements of $W^p([s,t], \mathbb{R}^{n\times m})$ are bounded with respect to the uniform norm. The following lemma states a well known fact.

Lemma 2.1. *Let $p \ge 1$, the spaces $W^p([0,T], \mathbb{R}^{n\times m})$ and $C^p([0,T], \mathbb{R}^{n\times m})$ equipped with the p-variation norm $\|\cdot\|_{p,0,T}$ are Banach spaces.*

Proof. The proof for the completeness of $W^p([0,T], \mathbb{R}^{n\times m})$ is given in Proposition 2.10 in Dudley and Norvaiša [1999]. Since $C^p([0,T], \mathbb{R}^{n\times m}) \subset W^p([0,T], \mathbb{R}^{n\times m})$ and by inequality (2.3) every Cauchy sequence in $C^p([0,T], \mathbb{R}^{n\times m})$ converges uniformly to a limit $x \in W^p([0,T], \mathbb{R}^{n\times m})$. Since the uniform limit of continuous functions is again continuous, we conclude $x \in C^p([0,T], \mathbb{R}^{n\times m})$, which yields the assertion. $\square$

Another embedding property is shown in the following lemma and is a easy implication of the inequality

$$\left(\sum_{i=0}^{k}|x_i|^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=0}^{k}|x_i|^q\right)^{\frac{1}{q}}$$

for every finite sequence of real numbers $(x_i)_{i=1,\dots,k}$ and $1 \leq q \leq p$.

Lemma 2.2. *Let $1 \leq q \leq p$. Is $x \in W^q([s,t],\mathbb{R}^{n\times m})$, then*

$$|x|_{p,s,t} \leq |x|_{q,s,t}$$

and $x \in W^p([s,t],\mathbb{R}^{n\times m})$. This also implies that $C^q([s,t],\mathbb{R}^{n\times m}) \subset C^p([s,t],\mathbb{R}^{n\times m})$.

In literature concerning functions of finite p-variation it is often beneficial to define the notion of a control function.

Definition 2.3. A continuous map φ taking values in the nonnegative real numbers, defined on the simplex $\Delta([s,t]) = \{(u,v) \in \mathbb{R}^2 \,|\, 0 \leq u \leq v \leq t\}$ is called a control function on $[s,t]$, if it satisfies the following conditions

- For all $r \in [s,t]$: $\varphi(r,r) = 0$.
- For all $u \leq r \leq v$ in $[s,t]$: $\varphi(u,r) + \varphi(r,v) \leq \varphi(u,v)$.

Lemma 2.4. *Let φ and ν be control functions on $[s,t]$, $C > 0$ and $x \geq 1$ be real constants, then $\varphi+\nu$, $\varphi\cdot\nu$, $C\varphi$ and φ^x also define control functions on $[s,t]$.*

Proof. For $\varphi+\nu$, $\varphi\cdot\nu$ and $C\varphi$ the only property of control functions that is not a direct consequence of the definition of a control function is the superadditivity of $\varphi\cdot\nu$. Let $u \leq r \leq r' \leq v$ in $[s,t]$, ν clearly satisfies $\nu(r,r') \leq \nu(u,v)$, which yields

$$\begin{aligned}&(\varphi\cdot\nu)(u,r) + (\varphi\cdot\nu)(r,v)\\ &\leq \varphi(u,r)\cdot\nu(u,v) + \varphi(r,v)\cdot\nu(u,v)\\ &\leq \varphi(u,v)\cdot\nu(u,v) = (\varphi\cdot\nu)(u,v),\end{aligned}$$

because of the superadditivity of φ. Also for φ^x only the superadditivity is not obvious. Since for all $a,b \geq 0$ and $x \geq 1$, we have

$$(a+b)^x \geq a^x + b^x,$$

it follows directly for $0 \leq u \leq r \leq v \leq T$ by superadditivity of φ that

$$\varphi(u,r)^x + \varphi(r,v)^x \leq (\varphi(u,r)+\varphi(r,v))^x \leq \varphi(u,v)^x.$$

□

Lemma 2.5. *Let φ and ν be control functions on $[s,t]$ and let $\alpha, \beta > 0$ such that $\alpha + \beta \geq 1$, then $\varphi^\alpha \cdot \nu^\beta$ is a control function on $[s,t]$.*

Proof. Compare Exercise 1.9 from Friz and Victoir [2010]. We only need to show superadditivity, the other properties of control functions are obviously satisfied. For $u \leq r \leq v \in [0,T]$, we have by the Hölder inequality

$$\begin{aligned}
&\varphi(u,r)^{\frac{\alpha}{\alpha+\beta}} \nu(u,r)^{\frac{\beta}{\alpha+\beta}} + \varphi(r,v)^{\frac{\alpha}{\alpha+\beta}} \nu(r,v)^{\frac{\beta}{\alpha+\beta}} \\
&\leq (\varphi(u,r) + \varphi(r,v))^{\frac{\alpha}{\alpha+\beta}} (\nu(u,r) + \nu(r,v))^{\frac{\beta}{\alpha+\beta}} \\
&\leq \varphi(u,v)^{\frac{\alpha}{\alpha+\beta}} + \nu(u,v)^{\frac{\beta}{\alpha+\beta}}.
\end{aligned}$$

The assertion follows by Lemma 2.4 with $x = \alpha + \beta$. □

Lemma 2.6. *Let $\varphi_1, \dots, \varphi_m$ be superadditive functions on $[s,t]$, $p \geq 1$, $C_1, \dots, C_k$ positive constants and $x : [s,t] \to \mathbb{R}^{n \times m}$ a function on $[s,t]$. The pointwise estimate*

$$|x_v - x_u| \leq \sum_{j=1}^m C_j \varphi_j(u,v)^{\frac{1}{p}} \text{ for all } u \leq v \text{ in } [s,t]$$

implies the p-variation estimate

$$|x|_{p,u,v} \leq \sum_{j=1}^m C_j \varphi_j(u,v)^{\frac{1}{p}} \text{ for all } u \leq v \text{ in } [s,t].$$

If φ_j is a control function on $[s,t]$ for all $j = 1, \dots, m$, then x is continuous on $[s,t]$.

Proof. By definition we have

$$|x|_{p,u,v} = \sup_{k \in \mathbb{N}, \Pi_k \in \mathcal{P}([u,v])} \left\{ \left(\sum_{j=0}^{k-1} |x_{t_{j+1}} - x_{t_j}|^p \right)^{\frac{1}{p}} \right\}.$$

Taking the assumption and the Minkowski inequality into account, we conclude

$$\begin{aligned}
|x|_{p,u,v} &\leq \sup_{k \in \mathbb{N}, \Pi_k \in \mathcal{P}([u,v])} \left\{ \left(\sum_{j=0}^{k-1} \left(\sum_{i=1}^m C_i \varphi_i(t_j, t_{j+1})^{\frac{1}{p}} \right)^p \right)^{\frac{1}{p}} \right\} \\
&\leq \sup_{k \in \mathbb{N}, \Pi_k \in \mathcal{P}([u,v])} \left\{ \sum_{i=1}^m \left(C_i^p \sum_{j=0}^{k-1} \varphi_i(t_j, t_{j+1}) \right)^{\frac{1}{p}} \right\} \\
&\leq \sup_{k \in \mathbb{N}, \Pi_k \in \mathcal{P}([u,v])} \left\{ \sum_{i=1}^m (C_i^p \varphi_i(u,v))^{\frac{1}{p}} \right\}
\end{aligned}$$

$$\leq \sum_{i=1}^{m} C_i \varphi_i(u,v)^{\frac{1}{p}}.$$

In the case, where all the φ_j are continuous for $j = 1, \ldots, m$, we have for $r \in [s,t]$

$$\lim_{u \nearrow r} |x_r - x_u| \leq \lim_{u \nearrow r} \sum_{i=1}^{m} C_i \varphi_i(u,r)^{\frac{1}{p}} = 0 \text{ and } \lim_{u \searrow r} |x_u - x_r| \leq \lim_{u \searrow r} \sum_{i=1}^{m} C_i \varphi_i(r,u)^{\frac{1}{p}} = 0.$$

□

The following proposition is an important basis for all the main results in this thesis and clarifies the connection between control functions and functions of finite p-variation.

Proposition 2.7. *Let $p \geq 1$ and $x : [s,t] \to \mathbb{R}^{n \times m}$ be a continuous function of finite p-variation, then*

$$\varphi(u,v) = |x|_{p,u,v}^{p}$$

defines a control function on $[s,t]$.

Proof. It is clear that $\varphi(u,u) = 0$ for all $u \in [s,t]$ and that $\varphi(u,v) \geq 0$ for all $u \leq v \in [s,t]$. To show superadditivity, let $u \leq r \leq v \in [s,t]$ and take arbitrary partitions $\Pi_k = (u = t_0, \ldots, t_k = r)$ of the interval $[u,r]$ and $\tilde{\Pi}_m = (r = \tilde{t}_0, \ldots, \tilde{t}_m = v)$ of the interval $[r,v]$. Then the sequence $\hat{\Pi}_{k+m} = (\hat{t}_0, \ldots, \hat{t}_{k+m}) := (t_0, \ldots, t_k = \tilde{t}_0, \ldots, \tilde{t}_m)$ is a partition of the interval $[u,v]$ and we have

$$\sum_{i=0}^{k-1} |x_{t_{i+1}} - x_{t_i}|^p + \sum_{i=0}^{m} |x_{\tilde{t}_{i+1}} - x_{\tilde{t}_i}|^p = \sum_{i=0}^{k+m} |x_{\hat{t}_{i+1}} - x_{\hat{t}_i}|^p \leq |x|_{p,u,v}^{p}.$$

Since Π_k and $\tilde{\Pi}_m$ are arbitrary partitions of $[u,r]$, respectively $[r,v]$, this yields the assertion by taking the supremum over all partitions of the two intevals. For the proof of continuity of φ, we refer the reader to Proposition 5.8.[1] of Friz and Victoir [2010]. □

The next three lemmas are just technical necessities for the proofs to come. We define for a function $x : [0,T] \to \mathbb{R}^{n \times m}$ and $s < t \in [0,T]$

$$\operatorname{Osc}(x,[s,t]) = \sup\{|x_v - x_u| | u < v \in [s,t]\}.$$

Lemma 2.8. *Let $x \in W^p([0,T], \mathbb{R}^{n \times m})$, $p \geq 1$. Let $p' > p$, we have*

$$|x|_{p',s,t} \leq \operatorname{Osc}(x,[s,t])^{1-\frac{p}{p'}} |x|_{p,s,t}^{\frac{p}{p'}}.$$

[1]The authors made a small error in the proof, which was corrected in the Errata to the book. It can be found on the web page of Peter Friz.

Proof. Let Π_k be an arbitrary partition of $[s,t]$, we have

$$\begin{aligned}\sum_{i=0}^{k-1} |x_{t_{i+1}} - x_{t_i}|^{p'} &\le \max_{i=0,\dots,k-1} |x_{t_{i+1}} - x_{t_i}|^{p'-p} \sum_{i=0}^{k-1} |x_{t_{i+1}} - x_{t_i}|^p \\ &\le \operatorname{Osc}(x,[s,t])^{p'-p} |x|_{p,s,t}^p.\end{aligned}$$

Taking the supremum over all partitions of $[s,t]$ and taking both sides to the power $\frac{1}{p'}$ yields the assertion. $\square$

Lemma 2.9. *Let $x \in W^p([s,t],\mathbb{R}^{n\times m})$, $p \ge 1$. If $s = t_0 < t_1 < \cdots < t_k = t$, then*

$$\sum_{i=0}^{k-1} |x|_{p,t_i,t_{i+1}}^p \le |x|_{p,s,t}^p \le k^{p-1} \sum_{i=0}^{k-1} |x|_{p,t_i,t_{i+1}}^p.$$

Proof. The first inequality follows from the superadditivity of $|x|_{p,u,v}^p$ and the second inequality can easily be seen by using the triangle inequality and the and Jensen inequality for convex functions. $\square$

Lemma 2.10. *Let $p \ge 1$, $B \in W^p([s,t],\mathbb{R}^{n\times n})$ and $x \in W^p([s,t],\mathbb{R}^{n\times m})$, then we have*

$$\|Bx\|_{p,s,t} = |B_s x_s| + |Bx|_{p,s,t} \le |B_s x_s| + \|B\|_{\infty,s,t}|x|_{p,s,t} + \|x\|_{\infty,s,t}|B|_{p,s,t} \le 2\|B\|_{p,s,t}\|x\|_{p,s,t}.$$

Proof. Using the definition of p-variation and the inequality (2.3), we obtain

$$\begin{aligned}\|Bx\|_{p,s,t} &= |B_s x_s| + \sup_{k\in\mathbb{N},\Pi_k([s,t])} \left(\sum_{i=0}^{k-1} |B_{t_{i+1}} x_{t_{i+1}} - B_{t_i} x_{t_i}|^p \right)^{\frac{1}{p}} \\ &= |B_s x_s| + \sup_{k\in\mathbb{N},\Pi_k([s,t])} \left(\sum_{i=0}^{k-1} |B_{t_{i+1}}(x_{t_{i+1}} - x_{t_i}) + (B_{t_{i+1}} - B_{t_i}) x_{t_i}|^p \right)^{\frac{1}{p}} \\ &\le |B_s x_s| + \|B\|_{\infty,s,t}|x|_{p,s,t} + \|x\|_{\infty,s,t}|B|_{p,s,t} \\ &\le |B_s x_s| + |B_s||x|_{p,s,t} + |B|_{p,s,t}|x|_{p,s,t} + |x_s||B|_{p,s,t} + |x|_{p,s,t}|B|_{p,s,t} \\ &\le 2\|B\|_{p,s,t}\|x\|_{p,s,t}.\end{aligned}$$

$\square$

In Young [1936], L.C. Young showed that it is possible do define an integral

$$\int_0^T x_r \, dw_r$$

as limit of Riemann-Stieltjes sums for partitions descending in mesh to 0 for an integrator w with unbounded total variation. He showed that it suffices for the existence of the integral, that

$x \in W^p([0,T],\mathbb{C})$ and $w \in W^q([0,T],\mathbb{C})$, where $p, q > 0$ and $\frac{1}{p} + \frac{1}{q} > 1$. We adapt the results of Dudley and Norvaiša [2010] to show the existence of the integral for functions $x : [0,T] \to \mathbb{R}^{n\times m}$ having finite p-variation for $p \geq 1$ and an integrator $w : [0,T] \to \mathbb{R}^{m\times d}$ (respectively $\mathbb{R}$) which is continuous and of finite q-variation for $q \geq 1$ and $\frac{1}{p} + \frac{1}{q} > 1$.

In Dudley and Norvaiša [2010], the authors define various generalizations of the classical Riemann-Stieltjes integral in a very general setting. The integral in Dudley and Norvaiša [2010] which is of interest to us is the full Stieltjes integral, which coincides in our setting with the classical Riemann-Stieltjes integral, because of the assumed continuity of the function w. So every result of Dudley and Norvaiša [2010], concerning the there defined integrals (RS), (RRS), (RYS) and (S) are applicable to our Integral. Therefore we will cite the results concerning the existence of the Riemann-Stieltjes integral and call the given integral Young-integral in the case where the p respectively q variation of the integrand and integrator satisfy the before mentioned condition $\frac{1}{p} + \frac{1}{q} > 1$. We will now summarize the results of Dudley and Norvaiša [2010] and adapt them to our specific setting.

Let X be the Banach space $(\mathbb{R}^{n\times m}, |\cdot|)$ and Y be either the Banach space $(\mathbb{R}^{m\times d}, |\cdot|)$ or $(\mathbb{R}, |\cdot|)$, all Banach spaces considered over $\mathbb{R}$. In the case where $Y = (\mathbb{R}^{n\times d}, |\cdot|)$ the standard matrix product defines a bilinear map from $X \times Y$ to $(\mathbb{R}^{n\times d}, |\cdot|)$ and $|xy| \leq |x||y|$ for all $x \in X$ and $y \in Y$. Furthermore for $Y = (\mathbb{R}, |\cdot|)$ the scalar multiplication defines a bilinear map from $X \times Y$ to $(\mathbb{R}^{n\times m}, |\cdot|)$ and $|xy| \leq |x||y|$ for all $x \in X$ and $y \in Y$. Therefore these spaces satisfy condition (1.14) of Dudley and Norvaiša [2010]. Let $\Pi_k = (t_i)_{i=0,\dots,k}$ be a partition of $[s,t]$ and $\Theta_k = (\tau_i)_{i=0,\dots,k-1}$ a sequence of times in $[s,t]$. We call (Π_k, Θ_k) a tagged partition of $[s,t]$ if

$$\theta_i \in [t_i, t_{i+1}] \text{ for all } i = 0, \dots, k-1.$$

We define the Riemann-Stieltjes sum of the functions $x : [s,t] \to X$ and $w : [s,t] \to Y$ on the tagged partition (Π_k, Θ_k) by

$$\mathrm{RS}(x, dw, (\Pi_k, \Theta_k)) = \sum_{i=0}^{k-1} x_{\theta_i}(w_{t_{i+1}} - w_{t_i}).$$

Now we are able to define the Riemann-Stieltjes integral.

Definition 2.11. Let $x : [s,t] \to X$ and $w : [s,t] \to Y$ be two functions. We say that the Riemann-Stieltjes integral (RS-integral)

$$\int_s^t x_r \, dw_r$$

exists with value $I \in \mathbb{R}^{n\times d}$ (respectively $\mathbb{R}^{n\times m}$), if for all $\varepsilon > 0$ there exists $\delta > 0$ such that for all tagged partitions (Π, Θ) of $[s,t]$ with $|\Pi| < \delta$, we have

$$|\mathrm{RS}(x, dw, (\Pi, \Theta)) - I| < \varepsilon.$$

For $s = t$ we set the integral to be 0.

The main tool in proving that the Riemann-Stieltjes sums converge, in the case where x is of finite q-variation and w is of finite p-variation for $\frac{1}{p} + \frac{1}{q} > 1$, is the so called Love-Young inequality. We first cite the Love-Young type inequality for Riemann-Stieltjes sums and then cite a result concerning the existence of the Riemann-Stieltjes integral in our setting.

Lemma 2.12 (Love-Young inequality for RS-sums)**.** *Let $k \in \mathbb{N}$ and (Π_k, Θ_k) be a tagged partition of the the interval $[s,t] \subset [0,T]$, where θ is an arbitrary element of $[s,t]$. Furthermore let $p, q \geq 1$ with $\frac{1}{p} + \frac{1}{q} > 1$, $w \in C^p([s,t], Y)$ and $x \in W^q([s,t], \mathbb{R}^{n \times m})$. We have*

$$\left| \sum_{i=0}^{k-1} x_{\theta_i}(w_{t_{i+1}} - w_{t_i}) - x_\theta(w_t - w_s) \right| \leq C_{p,q} |x|_{q,s,t} |w|_{p,s,t},$$

where $C_{p,q} = 1 + \zeta(\alpha)$ for $\zeta(x) = \sum_{i=1}^{\infty} \left(\frac{1}{i}\right)^x$ for $x > 1$ and $\alpha = \frac{1}{p} + \frac{1}{q}$. If additionally $\theta \in \Theta_k$, then $C_{p,q}$ reduces to $\zeta(\alpha)$. Moreover we have

$$\left| \sum_{i=0}^{k-1} x_{\theta_i}(w_{t_{i+1}} - w_{t_i}) \right| \leq \zeta(\alpha) \|x\|_{q,s,t} |w|_{p,s,t}.$$

Proof. See Corollary 3.87 in Dudley and Norvaiša [2010]. □

Theorem 2.13 (Young-Integral)**.** *For $1 \leq p,\ 1 \leq q$ such that $\alpha = \frac{1}{p} + \frac{1}{q} > 1$, $x \in W^q([s,t], \mathbb{R}^{n \times m})$ and $w \in C^p([s,t], Y)$ the Riemann-Stieltjes Integral $\int_s^t x_r \, dw_r$ exists and the inequality*

$$\left| \int_s^t x_r \, dw_r - x_\theta(w_t - w_s) \right| \leq C_{p,q} |x|_{q,s,t} |w|_{p,s,t} \tag{2.4}$$

holds for every $\theta \in [s,t]$, where $C_{p,q} = \zeta(\alpha)$ for $\zeta(x) = \sum_{i=1}^{\infty} \left(\frac{1}{i}\right)^x$ $(x > 1)$. Moreover, we have

$$\left| \int_s^t x \, dw \right| \leq C_{p,q} \|x\|_{q,s,t} |w|_{p,s,t}. \tag{2.5}$$

In this situation we call the integral the Young integral and inequality (2.5) *the Love-Young estimate.*

Proof. See Corollary 3.91 and Theorem 3.92 in Dudley and Norvaiša [2010]. □

Remark 2.14. In the case $x \in W^q([s,t], \mathbb{R}^{m\times d})$ and $w \in C^p([s,t], \mathbb{R}^{n\times m})$ we can also define the Young integral

$$\int_s^t (dw_r)x_r = \lim_{k\to\infty} \mathrm{RS}(dw, x, (\Pi_k, \Theta_k)) = \lim_{k\to\infty} \sum_{i=0}^{k} (w_{t_{i+1}} - w_{t_i})x_{\theta_i}$$

for every sequence of tagged partitions (Π_k, Θ_k) with $|\Pi_k| \to 0$ for $k \to \infty$. Then the inequalities in Lemma 2.12 and Theorem 2.13 also hold for the sums $\mathrm{RS}(dw, x, (\Pi_k, \Theta_k))$ and the given integral.

Now that we defined the Young integral, we list some of its properties. These are proven in Theorem 2.72 and 2.73 of Dudley and Norvaiša [2010].

Lemma 2.15. *Let $p, q \geq 1$ and $\frac{1}{p}+\frac{1}{q} > 1$, $x, x' \in W^q([s,t], \mathbb{R}^{n\times m})$, $w, w' \in C^p([s,t], Y)$, $C_1, C_2 \in \mathbb{R}$ and $u \in [s,t]$. In this situation all of the following Young integrals exist by the preceeding theorem and satisfy the following properties*

i) $\int_s^t C_1 x_r + C_2 x'_r \, dw_r = C_1 \int_s^t x_r \, dw_r + C_2 \int_s^t x'_r \, dw_r$.

ii) $\int_s^t x_r \, d(C_1 w_r + C_2 w'_r) = C_1 \int_s^t x_r \, dw_r + C_2 \int_s^t x_r \, dw'_r$.

iii) $\int_s^t x_r \, dw_r = \int_s^u x_r \, dw_r + \int_u^t x_r \, dw_r$.

These properties show that for $p, q \geq 1$ and $\frac{1}{p} + \frac{1}{q} > 1$ the integral operator

$$I_Y : W^q([s,t], \mathbb{R}^{n\times m}) \times C^p([s,t], Y) \to Z \quad (Z = \mathbb{R}^{n\times d} \text{ or } Z = \mathbb{R}^{n\times m} \text{ depending on } Y\ )$$
$$(x, w) \mapsto \int_s^t x_r \, dw_r$$

defines a bilinear map. The following lemma is devoted to the indefinite integral

$$I_Y(x, w)(u) = \int_s^u x_r \, dw_r \ \forall u \in [s,t]$$

and is an extension of Theorem 3.92 in Dudley and Norvaiša [2010], by proving its continuity on $[s,t]$ in our situation.

Lemma 2.16. *Let $1 \leq p,\ 1 \leq q$ such that $\alpha = \frac{1}{p}+\frac{1}{q} > 1$, $x \in W^q([s,t], \mathbb{R}^{n\times m})$ and $w \in C^p([s,t], Y)$. The indefinite integral $I_Y(x, w)$ exists and is an element of $C^p([s,t], Y)$. Furthermore, we have*

$$|I_Y(x, w)|_{p,s,t} = \left| \int_s^{\cdot} x_r \, dw_r \right|_{p,s,t} \leq C_{p,q} \|x\|_{q,s,t} |w|_{p,s,t}$$

for $C_{p,q} = \zeta(\alpha)$.

Proof. For every $r \in [s,t]$ the indefinite Integral $I_Y(x, w)(r)$ exists by Theorem 2.13. Let $u < v \in$

$[s,t]$, then we have by property iii) of Lemma 2.15 and the Young-Love estimate

$$|I_Y(x,w)(v) - I_Y(x,w)(u)| = \left| \int_u^v x_r \, dw_r \right| \leq C_{p,q} \|x\|_{q,u,v} |w|_{p,u,v} \leq C_{p,q} \|x\|_{q,s,t} |w|_{p,u,v}.$$

Since $\varphi(u,v) = |w|_{p,u,v}^p$ is a control function on $[s,t]$, we conclude the proof by applying Lemma 2.6. $\square$

We still need two results concerning the Young integral which will be especially helpful in dealing with linear Young differential equations, namely a substitution rule and an integration by parts formula. Both of these can also be found in Dudley and Norvaiša [2010] in different form. We will proof our version of the substitution rule and only cite the integration by parts formula.

Lemma 2.17 (Substitution rule)**.** *Let $x \in C^q([0,T], \mathbb{R}^{n\times n})$, $y \in C^q([0,T], \mathbb{R}^{n\times m})$ and $w \in C^p([0,T], \mathbb{R})$, such that $p, q \geq 1$ and $\frac{1}{p} + \frac{1}{q} > 1$, then we have for $[s,t] \subset [0,T]$*

$$\int_s^t (dI_Y(x,w)(r)) y_r = \int_s^t x_r y_r \, dw_r = \int_s^t x_r \, dI_Y(y,w)(r).$$

Proof. By Lemma 2.10 we know that xy is an element of $C^q([0,T], \mathbb{R}^{n\times m})$ and by Lemma 2.16, that $I_Y(y,w) \in C^p([0,T], \mathbb{R}^{n\times m})$. So by Theorem 2.13 and Remark 2.14 all integrals exist as limit of their Riemann Stieltjes sums for sequences of partitions of $[s,t]$ descending in mesh to 0. Let (Π_k, Θ_k) be a tagged partition of the interval $[s,t]$ and note that by Lemma 2.15 iii) the Young integral is additive in its limits. We only proof the first equality, the second follows by symmetry. For $X \in \mathbb{R}^{n\times m}$ and $u \in \{1,\dots,n\}$, $v \in \{1,\dots,m\}$ let $X_r^{(u,v)}$ be the component of X_r in the u-th row and v-th column, we have

$$\begin{aligned}
\left(\int_s^t (dI_Y(x,w)(r)) y_r \right)^{(u,v)} &= \lim_{k\to\infty} \sum_{i=0}^{k-1} \sum_{j=1}^{n} \left(\int_{t_i}^{t_{i+1}} x_r \, dw_r \right)^{(u,j)} y_{\theta_i}^{(j,v)} \\
&= \lim_{k\to\infty} \sum_{j=1}^{n} \sum_{i=0}^{k-1} \left(\int_{t_i}^{t_{i+1}} x_r^{(u,j)} \, dw_r \right) y_{\theta_i}^{(j,v)} \\
&= \lim_{k\to\infty} \sum_{j=1}^{n} \mathrm{RS}(dI_Y(x^{(u,j)}, w), y^{(j,v)}, (\Pi_k, \Theta_k))
\end{aligned}$$

and

$$\left(\int_s^t x_r y_r \, dw_r \right)^{(u,v)} = \lim_{k\to\infty} \sum_{i=0}^{k-1} \sum_{j=1}^{n} x_{\theta_i}^{(u,j)} y_{\theta_i}^{(j,v)} (w_{t_{i+1}} - w_{t_i})$$

$$= \lim_{k\to\infty} \sum_{j=1}^{n} \sum_{i=0}^{k-1} x_{\theta_i}^{(u,j)} y_{\theta_i}^{(j,v)} (w_{t_{i+1}} - w_{t_i})$$
$$= \lim_{k\to\infty} \sum_{j=1}^{n} \mathrm{RS}(x^{(u,j)} y^{(j,v)}, dw, (\Pi_k, \Theta_k)).$$

Notice that

$$\sum_{i=0}^{k-1} \left(\int_{t_i}^{t_{i+1}} x_r^{(u,j)} \, dw_r \right) y_{\theta_i}^{(j,v)} = \sum_{i=0}^{k-1} \left(\int_{t_i}^{t_{i+1}} x_r^{(u,j)} - x_{\theta_i}^{(u,j)} \, dw_r \right) y_{\theta_i}^{(j,v)} + \sum_{i=0}^{k-1} x_{\theta_i}^{(u,j)} y_{\theta_i}^{(j,v)} (w_{t_{i+1}} - w_{t_i}).$$

Therefore we have

$$\begin{aligned} &\sum_{j=1}^{n} |\mathrm{RS}(dI_Y(x,w)^{(u,j)}, y^{(j,v)}, (\Pi_k, \Theta_k)) - \mathrm{RS}(x^{(u,j)} y^{(j,v)}, dw, (\Pi_k, \Theta_k))| \\ &\le \sum_{j=1}^{n} \sum_{i=0}^{k-1} \left| \int_{t_i}^{t_{i+1}} x_r^{(u,j)} - x_{\theta_i}^{(u,j)} \, dw_r \right| \|y^{(j,v)}\|_{\infty,s,t} \\ &:= I_1. \end{aligned}$$

We choose $p' > p, q' > q$ such that $\frac{1}{p'} + \frac{1}{q'} > 1$ and obtain by using (2.4)

$$\begin{aligned} I_1 &\le \|y\|_{\infty,s,t} \sum_{j=1}^{n} \sum_{i=0}^{k-1} \left| \int_{t_i}^{t_{i+1}} x_r^{(u,j)} - x_{\theta_i}^{(u,j)} \, dw_r \right| \\ &\le C_{p,q} \|y\|_{\infty,s,t} \sum_{j=1}^{n} \sum_{i=0}^{k-1} |x^{(u,j)}|_{q',t_i,t_{i+1}} |w|_{p',t_i,t_{i+1}}. \end{aligned}$$

Taking Lemma 2.8 into account we can estimate

$$\begin{aligned} &\sum_{i=0}^{k-1} |x^{(u,j)}|_{q',t_i,t_{i+1}} |w|_{p',t_i,t_{i+1}} \\ &\le \max_{i=0,\dots,k-1} \left\{ \mathrm{Osc}(x^{(u,j)}, [t_i, t_{i+1}])^{1-\frac{q}{q'}} \cdot \mathrm{Osc}(w, [t_i, t_{i+1}])^{1-\frac{p}{p'}} \right\} \sum_{i=0}^{k-1} |x^{(u,j)}|_{q,t_i,t_{i+1}}^{\frac{q}{q'}} |w|_{p,t_i,t_{i+1}}^{\frac{p}{p'}} \\ &\le \max_{i=0,\dots,k-1} \left\{ \mathrm{Osc}(x^{(u,j)}, [t_i, t_{i+1}])^{1-\frac{q}{q'}} \cdot \mathrm{Osc}(w, [t_i, t_{i+1}])^{1-\frac{p}{p'}} \right\} |x^{(u,j)}|_{q,s,t}^{\frac{q}{q'}} |w|_{p,s,t}^{\frac{p}{p'}}, \end{aligned}$$

where we used Lemma 2.5 for the last estimate. Hence the term I_1 converges to 0 if $|\Pi_k| \to 0$, by the uniform continuity of $x^{(u,j)}$ for all $j = 1, \dots, n$ and w on $[0,T]$. □

Lemma 2.18 (Integration by parts)**.** *Let $x \in C^q([s,t], \mathbb{R}^{n\times n})$ and $y \in C^p([s,t], \mathbb{R}^{n\times m})$, such that*

$p, q \geq 1$ and $\frac{1}{p} + \frac{1}{q} > 1$, we have

$$x_t y_t - x_0 y_0 = \int_s^t x_r \, dy_r + \int_s^t (dx_r) y_r.$$

Proof. See Theorem 2.80 in Dudley and Norvaiša [2010]. □

Now that we established most of the properties of functions of finite p-variation and defined the Young-integral, we can concern ourselves with the notion of Young differential equations, which will be the topic of the following subsection.

2.1.2 Young differential equations

In this subsection we will cite very important results of Nguyen et al. [2018]. In this paper, the authors establish existence and uniqueness results for the non-autonomous ordinary Young differential equations of the form

$$x_t = x_0 + \int_0^t b(r, x_r) \, dr + \int_0^t \sigma(r, x_r) \, dw_r,$$

and therefore their results will play a crucial role in the remainder of this thesis. We will state and proof a version of their Gronwall-type lemma for a better understanding of the p-variation techniques involved. A problem one has to overcome to make the necessary estimates for the Gronwall-type lemma in p-variation spaces is to find partitions of the interval $[0, T]$ such that the p-variation of a given path is bounded by the same constant on every subinterval of the partition. The authors obtain these partitions by constructing a so called greedy sequence of times. The term greedy sequence of times was first introduced in Cass et al. [2013], here we present the construction of Nguyen et al. [2018] in a slightly modified way. We just exchange the specific control function $|w|_{p,s,t}^p$ used in the construction of a greedy sequence in Nguyen et al. [2018] by an arbitrary control function φ on $[0, T]$ and restrict ourselves on the time interval $[0, T]$. Our goal is to construct an increasing sequence of times $(\tau_i(\varphi))_{i=0,\ldots,k}$ with $\tau_k(\varphi) = T$, satisfying

$$\begin{aligned} |\tau_{i+1}(\varphi) - \tau_i(\varphi)| + \varphi^{\frac{1}{p}}(\tau_i(\varphi), \tau_{i+1}(\varphi)) &= \mu \text{ for } i = 0, \ldots, k-2 \\ |\tau_k(\varphi) - \tau_{k-1}(\varphi)| + \varphi^{\frac{1}{p}}(\tau_{k-1}(\varphi), \tau_k(\varphi)) &\leq \mu \end{aligned} \tag{2.6}$$

for $\mu > 0$, $p \geq 1$. We call such a sequence a greedy sequence of times. For the construction, we first define

$$\tau_0(\varphi) = 0, \;\; \tau_1(\varphi) = \sup_{0 \leq t \leq T} \left\{ t + \varphi^{\frac{1}{p}}(0, t) \leq \mu \right\}.$$

Notice that $\kappa(t) = t + \varphi^{\frac{1}{p}}(0, t)$ is continuous and strictly increasing with respect to t, with $\kappa(0) = 0$ and $\kappa(T) = T + \varphi^{\frac{1}{p}}(0, T)$. The intermediate value theorem ensures, that there exists a unique

$t > 0$ such that

$$t + \varphi^{\frac{1}{p}}(0, t) = \mu,$$

if $\mu < T + \varphi^{\frac{1}{p}}(0, T)$, else $\tau_1(\varphi) = T$. Hence, $\tau_1(\varphi)$ is well defined. We construct the rest of the sequence inductively. Assuming we have defined $\tau_j(\varphi)$ for an arbitrary $j \in \mathbb{N}$ with $\tau_j(\varphi) < T$, we construct $\tau_{j+1}(\varphi)$ in the following way. Using the same arguments as before, the supremum

$$\delta_j(\varphi) = \sup_{0 \le \delta \le T - \tau_j(\varphi)} \{\delta + \varphi^{\frac{1}{p}}(\tau_j(\varphi), \tau_j(\varphi) + \delta) \le \mu\}$$

is well defined. Hence we can set

$$\tau_{j+1}(\varphi) = \tau_j(\varphi) + \delta_j(\varphi).$$

This sequence satisfies the property (2.6). Now we prove that the number of times of the greedy sequence in an interval $[s, t] \subset [0, T]$ is finite. For $T > 0$ we introduce the notation

$$N(T, \varphi) := \inf\{k \in \mathbb{N} | \tau_k(\varphi) = T\}, \tag{2.7}$$

or more generally, for any $0 \le s < t \le T$, we define

$$\overline{N}(t, \varphi) = \sup_{k \in \mathbb{N}} \{\tau_k(\varphi) \le t\}$$

$$\underline{N}(t, \varphi) = \inf_{k \in \mathbb{N}} \{\tau_k(\varphi) \ge t\}$$

and

$$N(s, t, \varphi) = \overline{N}(t, \varphi) - \underline{N}(s, \varphi). \tag{2.8}$$

If $N(T, \varphi)$ is well defined, then our greedy sequence $(0, \tau_1, \ldots, \tau_{N(T,\varphi)} = T)$ defines a finite partition of the interval $[0, T]$. In the following, we write $\tau_i = \tau_i(\varphi)$, for notational simplicity.

Lemma 2.19. *Let $p' \ge p$ and φ be an arbitrary control function on $[0, T]$. The following estimate holds*

$$N(T, \varphi) \le \frac{2^{p'-1}}{\mu^{p'}} \left(T^{p'} + \varphi^{\frac{p'}{p}}(0, T) \right). \tag{2.9}$$

More generally

$$N(s, t, \varphi) \le \frac{2^{p'-1}}{\mu^{p'}} \left((t - s)^{p'} + \varphi^{\frac{p'}{p}}(s, t) \right). \tag{2.10}$$

Proof. See Nguyen et al. [2018], Lemma 2.6. Using Jensens inequality and the super additivity of the control function φ on $[0, T]$, we obtain for every $k \in \mathbb{N}$, such that $\tau_k < T$

$$k\mu^{p'} = \sum_{i=0}^{k-1} \mu^{p'} = \sum_{i=0}^{k-1} \left[|\tau_{i+1} - \tau_i| + \varphi^{\frac{1}{p}}(\tau_i, \tau_{i+1}) \right]^{p'}$$

$$\leq 2^{p'-1}\left[\sum_{i=0}^{k-1}|\tau_{i+1}-\tau_i|^{p'}+\left(\sum_{i=0}^{k-1}\varphi(\tau_i,\tau_{i+1})\right)^{\frac{p'}{p}}\right]$$

$$\leq 2^{p'-1}\left[\tau_k^{p'}+\varphi^{\frac{p'}{p}}(0,\tau_k)\right]. \tag{2.11}$$

Since left hand side of inequality (2.11) tends to infinity for $k \to \infty$, the right hand side has to be increasing to infinity in k as well. This implies that there exists $k \in \mathbb{N}$ such that $\tau_k = T$ by construction of our greedy sequence of times. Consequently, we obtain

$$N(T,w) \leq \frac{2^{p'-1}}{\mu^{p'}}\left(T^{p'}+\varphi^{\frac{p'}{p}}(0,T)\right).$$

Similarly, (2.10) holds. □

Using the greedy sequence of times the authors are able to formulate their Gronwall-type lemma. We adapt their ideas to our setting by considering a matrix valued function y in the space $W^q([0,T],\mathbb{R}^{n\times m})$ and using an arbitrary control function φ instead of $|w|^p_{p,s,t}$. Furthermore we simplify the needed condition to suit our purposes.

Lemma 2.20. *Let $1 \leq p \leq q$ be arbitrary and satisfy $\frac{1}{p}+\frac{1}{q} > 1$ and for $T > 0$. Assume that $y \in W^q([0,T],\mathbb{R}^{n\times m})$ and an arbitrary control function φ on $[0,T]$ satisfy the following condition: There exist constants $K_1, K_2 > 0$ such that for all $[s,t] \subset [0,T]$, which satisfy*

$$|t-s|+\varphi(s,t)^{\frac{1}{p}} \leq K_2,$$

we have

$$|y|_{q,s,t} \leq K_1 + |y_s|. \tag{2.12}$$

Then we get the estimate

$$|y|_{q,0,T} \leq (K_1+|y_0|)e^{2^pK_2^{-p}(T^p+\varphi(0,T))} \tag{2.13}$$

and

$$\|y\|_{\infty,0,T} \leq \|y\|_{q,0,T} \leq (K_1+2|y_0|)e^{2^pK_2^{-p}(T^p+\varphi(0,T))}.$$

If in line (2.12) the right side does only consists of the constant K_1, then the estimate simplifies to

$$|y|_{q,0,T} \leq K_1 2^{p-1}K_2^{-p}(T^p+\varphi(0,T)) \tag{2.14}$$

and

$$\|y\|_{\infty,0,T} \leq \|y\|_{q,0,T} \leq |y_0| + K_1 2^{p-1}K_2^{-p}(T^p+\varphi(0,T)). \tag{2.15}$$

Proof. Compare Nguyen et al. [2018], Lemma 3.3/ Remark 3.4/ Corollary 3.5. We construct a time sequence $0 = \tau_0 < \cdots < \tau_N = T$ on $[0,T]$ using the greedy sequence of times (explained before) with

$$(\tau_{i+1} - \tau_i) + \varphi^{\frac{1}{p}}(\tau_i, \tau_{i+1}) \leq K_2$$

for $i = 0, \dots, N-1$, where N is given by $N(T,\varphi)$ defined in (2.7). Then, by (2.12), we have

$$|y|_{q,s,t} \leq K_1 + |y_s| \tag{2.16}$$

for all $s,t \in [\tau_i, \tau_{i+1}], s \leq t$. This yields that

$$|y_{\tau_{i+1}}| \leq \|y\|_{\infty,\tau_i,\tau_{i+1}} \leq K_1 + 2|y_{\tau_i}|$$

for all $i = 0, \dots, N-1$. If $N = 1$, we have that (2.13) trivially holds. Now let $N \geq 2$ and fix $i \in \{0, \dots, N-1\}$ such that $\tau_i < t \leq \tau_{i+1}$. Inductively we get

$$\begin{aligned} K_1 + |y_{\tau_i}| &\leq K_1 + K_1 + 2|y_{\tau_{i-1}}| \\ &\leq 2(K_1 + |y_{\tau_{i-1}}|) \\ &\leq \dots \\ &\leq 2^{i-1}(K_1 + |y_{\tau_1}|) \\ &\leq 2^i(K_1 + |y_0|). \end{aligned}$$

Hence,

$$|y|_{q,\tau_i,\tau_{i+1}} \leq K_1 + |y_{\tau_i}| \leq 2^i(K_1 + |y_0|).$$

By Lemma 2.9, we obtain

$$\begin{aligned} |y|_{q,0,T} &\leq N^{\frac{q-1}{q}} \left(\sum_{i=0}^{N-1} |y|^q_{q,\tau_i,\tau_{i+1}} \right)^{\frac{1}{q}} \\ &\leq N^{\frac{q-1}{q}} (K_1 + |y_0|) \left(\sum_{i=0}^{N-1} 2^{iq} \right)^{\frac{1}{q}} \\ &\leq N(K_1 + |y_0|)2^N \\ &\leq (K_1 + |y_0|)e^{2N} \end{aligned} \tag{2.17}$$

Taking (2.9) with $p' = p$ into account, we obtain

$$|y|_{q,0,T} \leq (K_1 + |y_0|)e^{2^p K_2^{-p}(|T|^p + \varphi(0,T))}.$$

With the inequality $\|y\|_{\infty,0,T} \le \|y\|_{q,0,T} = |y_0| + |y|_{q,0,T}$ for all $s < t \in [0,T]$, we conclude

$$\|y\|_{\infty,0,T} \le \|y\|_{q,0,T} \le (K_1 + 2|y_0|)e^{2^p K_2^{-p}(T^p+\varphi(0,T))}.$$

Now suppose line (2.16) simplifies to

$$|y|_{q,s,t} \le K_1,$$

then we can directly use the decomposition (2.17) and get

$$|y|_{q,0,T} \le NK_1.$$

By (2.9) the assertion in (2.14) and (2.15) follows. □

We continue by citing the conditions for the coefficients of the YDE and the main theorem of their work concerning the existence and uniqueness of a solution to the given YDE.

(C_1) $\sigma(t,x) : [0,T]\times\mathbb{R}^n \to \mathbb{R}^{n\times m}$ is differentiable in x and there exist some constants $0 < \beta, \delta \le 1$, a control function $h(s,t)$ defined on $[0,T]$ and for every $N \ge 0$ there exists $M_N > 0$ such that the following properties hold:

$$(H_\sigma): \begin{cases} \text{(i) Lipschitz continuity} \\ |\sigma(t,x) - \sigma(t,y)| \le L_\sigma |x-y|, \ \forall x,y \in \mathbb{R}^n, \ \forall t \in [0,T] \\ \text{(ii) Local Hölder continuity} \\ |\partial_{x_i}\sigma(t,x) - \partial_{x_i}\sigma(t,y)| \le M_N|x-y|^\delta, \ \forall x,y \in \mathbb{R}^n, \ |x|,|y| \le N, \ \forall t \in [0,T] \\ \text{(iii) Generalized Hölder continuity in time} \\ |\sigma(t,x) - \sigma(s,x)| + |\partial_{x_i}\sigma(t,x) - \partial_{x_i}\sigma(s,x)| \le h(s,t)^\beta, \ \forall x \in \mathbb{R}^n, \ s,t \in [0,T], s < t. \end{cases}$$

(C_2) $b : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ and there exists $a > 0$ and $f \in L^{\frac{1}{1-\alpha}}([0,T],\mathbb{R}^n)$, where $\frac{1}{2} \le \alpha < 1$, and for every $N \ge 0$ exists $L_N > 0$ such that the following properties hold:

$$(H_b): \begin{cases} \text{(i) Local Lipschitz continuity} \\ |b(t,x) - b(t,y)| \le L_N|x-y|, \ \forall x,y \in \mathbb{R}^n, \ |x|,|y| \le N \ \forall t \in [0,T] \\ \text{(ii) Growth} \\ |b(t,x)| \le a|x| + f(t), \ \forall x \in \mathbb{R}^n, \ \forall t \in [0,T]. \end{cases}$$

(C_3) The parameters in (C_1) and (C_2) satisfy the inequalities $\delta > p-1, \ \beta > 1 - \frac{1}{p}, \ \delta\alpha > 1 - \frac{1}{p}$.

By the assumption $p \in (1,2)$ and the condition (C_3), one can choose consecutively constants

q_0, q such that

$$1 - \frac{1}{p} < \frac{1}{q_0} < \min\left\{\beta, \delta\alpha, \frac{\delta}{p}, \frac{1}{2}\right\} \tag{2.18}$$

$$\frac{1}{q_0 \delta} \leq \frac{1}{q} < \min\left\{\alpha, \frac{1}{p}\right\} \tag{2.19}$$

which gives an appropriate constant $q > 0$, such that the solution to the YDE x is given in the space $C^q([0,T], \mathbb{R}^n)$.

Theorem 2.21. *Let $w \in C^p([0,T], \mathbb{R}^m)$ for $p \in (1,2)$, T an arbitrary fixed positive number and $s_0 \in [0,T)$ be a abitrary initial time. Consider the YDE*

$$x_t = x_0 + \int_{s_0}^{t} b(r, x_r)\, dr + \int_{s_0}^{t} \sigma(r, x_r)\, dw_r,\ \ t \in [s_0, T],\ \ x_{s_0} \in \mathbb{R}^n,$$

with $x_0 \in \mathbb{R}^n$ being an arbitrary initial condition. Assume that the conditions (C_1), (C_2) and (C_3) hold. Then, this equation has a unique solution x in the space $C^q([s_0,T], \mathbb{R}^n)$, where q is chosen above satisfying (2.19). *Moreover, the solution is in $C^{p'}([s_0,T], \mathbb{R}^n)$, where $p' = \max\{p, \frac{1}{\alpha}\}$.*

Proof. Theorem 3.6 of Nguyen et al. [2018]. □

The last theorem gives us the existence and uniqueness of a solution to an YDE with general coefficients. But later we will have to calculate the explicit solution to an linear YDE which does not satisfy the conditions (C_1), (C_2) and (C_3). So we need another theorem to proof that such an solution exists, therefore we will use the results of Proposition 2.2 in Nguyen et al. [2020] concerning the existence and uniqueness of solutions to homogenous linear YDEs.

Theorem 2.22. *Let T be an arbitrary positive number and $s_0 \in [0,T]$. Let $w = (w^1, \ldots, w^m)^\top \in C^p([0,T], \mathbb{R}^m)$ for $p \in (1,2)$. Assume that $D \in C([s_0,T], \mathbb{R}^{n\times n})$, $E^j \in C^q([s_0,T], \mathbb{R}^{n\times n})$ for $j = 1, \ldots, m$ with $\frac{1}{p} + \frac{1}{q} > 1$ and $q > p$. Consider the YDE*

$$x_t = x_0 + \int_{s_0}^{t} D_r\, x_r\, dr + \sum_{j=1}^{m} \int_{s_0}^{t} E_r^j\, x_r\, dw_r^j,\ \ t \in [s_0, T],$$

with $x_0 \in \mathbb{R}^{n\times n}$ being an arbitrary initial condition. Then, this equation has a unique solution x in the space $C^p([s_0,T], \mathbb{R}^{n\times n})$.

Proof. See Proposition 2.2 in Nguyen et al. [2020]. Note that in Nguyen et al. [2020] the authors consider the corresponding vector valued YDE with w as element of $C^p([0,T], \mathbb{R})$. The adaption of the arguments in their proof to our setting are straightforward. □

2.1.3 Boundedness, continuity and Fréchet differentiability of the solution mapping

Now we consider the parameter dependent deterministic YDE of our interest given by

$$x_t^u = x_0(u) + \int_0^t b(r, x_r^u, u)dr + \int_0^t \sigma(r, x_r^u, u)dw_r,$$

where $w \in C^p([0,T], \mathbb{R}^m)$. In this subsection we generalize the approach of Han et al. [2012] to p-variation spaces, where related calculations were made for Hölder continuous paths, respectively using Hölder norms for the special case of a fractional Brownian motion driver in a stochastic control setting. But to ensure that all our results in the stochastic setting hold in L^l-sense for every $l \geq 1$, we assume also the boundedness of the coefficient functions in contrast to Han et al. [2012]. We define the set of parameters $\mathcal{U}$ as an open, bounded and convex subset of $\mathbb{R}^d$. We first need some assumptions about the coefficients and the initial value of the YDE

(H_1) Let $x_0 : \mathcal{U} \to \mathbb{R}^n$ be continuously differentiable, such that x_0 and its Jacobian Dx_0 are bounded by a constant L.

(H_2) Let $b : [0,T] \times \mathbb{R}^n \times \mathcal{U} \to \mathbb{R}^n$ be a continuous function which satisfies:

- $b(t,x,u)$ is continuously differentiable with respect to x and u.
- There exists a constant L such that for all $x, y \in \mathbb{R}^n$, $u, v \in \mathcal{U}$ and every $t \in [0,T]$

$$|b(t,x,u)| \leq L$$

$$|b_x(t,x,u)| + |b_u(t,x,u)| \leq L$$

$$|b_x(t,x,u) - b_x(t,y,v)| + |b_u(t,x,u) - b_u(t,y,v)| \leq L(|x-y| + |u-v|).$$

(H_3) Let $\sigma := (\sigma^1, \ldots, \sigma^m) : [0,T] \times \mathbb{R}^n \times \mathcal{U} \to \mathbb{R}^{n \times m}$ be a continuous function which satisfies:

- $\sigma(t,x,u)$ is twice continuously differentiable with respect to x and u.
- There exists a constant L, such that for all $x \in \mathbb{R}^n$, $u \in \mathcal{U}$, $t \in [0,T]$ and $j = 1, \ldots, m$, $l = 1, \ldots, n$, $k = 1, \ldots, d$

$$\begin{aligned}
&|\sigma(t,x,u) \leq L \\
&|\sigma_x^j(t,x,u)| + |\sigma_u^j(t,x,u)| \leq L \\
&\left| \frac{\partial}{\partial x} \sigma_{x_l}^j(t,x,u) \right| + \left| \frac{\partial}{\partial x} \sigma_{u_k}^j(t,x,u) \right| \leq L \\
&\left| \frac{\partial}{\partial u} \sigma_{x_l}^j(t,x,u) \right| + \left| \frac{\partial}{\partial u} \sigma_{u_k}^j(t,x,u) \right| \leq L,
\end{aligned}$$

where

$$\sigma_x^j(t,x,u) = \begin{pmatrix} \frac{\partial}{\partial x_1}\sigma^{1,j}(t,x,u) & \dots & \frac{\partial}{\partial x_n}\sigma^{1,j}(t,x,u) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1}\sigma^{n,j}(t,x,u) & \dots & \frac{\partial}{\partial x_n}\sigma^{n,j}(t,x,u) \end{pmatrix}$$

$$\sigma_u^j(t,x,u) = \begin{pmatrix} \frac{\partial}{\partial u_1}\sigma^{1,j}(t,x,u) & \dots & \frac{\partial}{\partial u_d}\sigma^{1,j}(t,x,u) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial u_1}\sigma^{n,j}(t,x,u) & \dots & \frac{\partial}{\partial u_d}\sigma^{n,j}(t,x,u) \end{pmatrix}$$

$$\frac{\partial}{\partial x}\sigma_{x_l}^j(t,x,u) = \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_l}\sigma^{1,j}(t,x,u) & \dots & \frac{\partial^2}{\partial x_n \partial x_l}\sigma^{1,j}(t,x,u) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_1 \partial x_l}\sigma^{n,j}(t,x,u) & \dots & \frac{\partial^2}{\partial x_n \partial x_l}\sigma^{n,j}(t,x,u) \end{pmatrix}$$

$$\frac{\partial}{\partial u}\sigma_{u_k}^j(t,x,u) = \begin{pmatrix} \frac{\partial^2}{\partial u_1 \partial u_k}\sigma^{1,j}(t,x,u) & \dots & \frac{\partial^2}{\partial u_d \partial u_k}\sigma^{1,j}(t,x,u) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial u_1 \partial u_k}\sigma^{n,j}(t,x,u) & \dots & \frac{\partial^2}{\partial u_d \partial u_k}\sigma^{n,j}(t,x,u) \end{pmatrix}$$

and analogously defined $\frac{\partial}{\partial x}\sigma_{u_k}^j$, $\frac{\partial}{\partial u}\sigma_{x_l}^j$.

– There exist constants L and $\beta \in [\frac{1}{2}, 1]$ such that for all $x, y \in \mathbb{R}^n$, $u, v \in \mathcal{U}$, $s \le t \in [0,T]$ and $j = 1, \dots, m$, $l = 1, \dots, n$, $k = 1, \dots, d$

$$\begin{aligned}
&|\sigma(t,x,u) - \sigma(s,x,u)| \le L|t-s|^\beta \\
&|\sigma_x^j(t,x,u) - \sigma_x^j(s,x,u)| + |\sigma_u^j(t,x,u) - \sigma_u^j(s,x,u)| \le L|t-s|^\beta \\
&\left|\frac{\partial}{\partial x}\sigma_{x_l}^j(t,x,u) - \frac{\partial}{\partial x}\sigma_{x_l}^j(s,y,v)\right| + \left|\frac{\partial}{\partial u}\sigma_{x_l}^j(t,x,u) - \frac{\partial}{\partial u}\sigma_{x_l}^j(s,y,v)\right| \\
&+ \left|\frac{\partial}{\partial x}\sigma_{u_k}^j(t,x,u) - \frac{\partial}{\partial x}\sigma_{u_k}^j(s,y,v)\right| + \left|\frac{\partial}{\partial u}\sigma_{u_k}^j(t,x,u) - \frac{\partial}{\partial u}\sigma_{u_k}^j(s,y,v)\right| \\
&\le L\Big(|t-s|^\beta + |x-y| + |u-v|\Big).
\end{aligned}$$

Remark 2.23. Note that by the boundedness of the partial derivatives of σ^j with respect to x and u for every $j = 1, \dots, m$, we clearly get Lipschitz continuity of the functions σ, σ_x^j and σ_u^j with respect to x and u. The Lipschitz constant will then depend on the dimensions m, n and d, but for notational simplicity, we will just choose L big enough such all these conditions are satisfied.

Note that if the coefficients b and σ satisfy the conditions (H_2) and (H_3) then for a given u the coefficients $b(t,x,u) = b(t,x)$ and $\sigma(t,x,u) = \sigma(t,x)$ obviously satisfy the conditions (C_1),

(C_2) and (C_3). Taking a look at the parameters involved in the conditions $(C_1), (C_2)$ and (C_3), we can then set $\delta = 1$ and because of the boundedness of the coefficients, we formally set $\alpha = 1$ in (2.18), (2.19). Since $\beta \geq \frac{1}{2}$ the inequalities (2.18) and (2.19) simplify to

$$1 - \frac{1}{p} < \frac{1}{q_0} < \frac{1}{2} \tag{2.20}$$

$$\frac{1}{q_0} \leq \frac{1}{q} < \frac{1}{p}. \tag{2.21}$$

These inequalities are satisfied, if $\frac{1}{p} + \frac{1}{q_0} > 1$, $q_0 > 2$ and $q_0 \geq q > p$. In the next corollary we use the existence results of Nguyen et al. [2018] to formulate an existence result for our parameter dependent YDE, which is a direct implication of the aforementioned arguments.

Corollary 2.24. *Let $w \in C^p([0,T], \mathbb{R}^m)$ for $p \in (1,2)$, T an arbitrary fixed positive number. Consider the parameter dependent YDE*

$$x_t^u = x_0(u) + \int_0^t b(r, x_r^u, u)\, dr + \int_0^t \sigma(r, x_r^u, u)\, dw_r,$$

with $x_0 : \mathcal{U} \to \mathbb{R}^n$ being an arbitrary initial condition and $u \in \mathcal{U}$. Assume that the conditions (H_1), (H_2) and (H_3) hold. Then, this equation has a unique solution x^u in the space $C^q([0,T], \mathbb{R}^n)$, where q is chosen according to the inequalities (2.20) *and* (2.21)*. Moreover, the solution is an element of $C^p([0,T], \mathbb{R}^n)$.*

Now we will come to the boundedness of the solution x^u which is independent of the given parameter, but first we will establish further properties of the coefficient σ, which are necessary for the proofs to come. This is a version of Lemma 3.1 in Nguyen et al. [2018] for the case of parameter dependent coefficients. To proof this lemma will will need an auxiliary result which is the well know mean value theorem for vector valued functions.

Lemma 2.25 (Mean value theorem)**.** *Let $U \subset \mathbb{R}^n$ be open and $f : U \to \mathbb{R}^m$ a continuously differentiable function. Let $x \in U$ and $h \in \mathbb{R}^n$, such that the whole line segment $x + th \in U$ for all $t \in [0,1]$. Then we have*

$$f(x+h) - f(x) = \int_0^1 Df(x + \lambda h) h \, d\lambda,$$

where Df is given by the Jacobi matrix of f.

Proof. See Forster [2008], §6, Theorem 5. □

Lemma 2.26. *Let $p \in (1,2)$, $q \in (2, \frac{p}{p-1})$, $T > 0$ and $s_0 \in [0,T]$. Assume that $(H_1) - (H_3)$ are satisfied.*

i) For every $u \in \mathcal{U}$, $x \in C^p([s_0,T],\mathbb{R}^n)$, we have that $\sigma(\cdot, x_\cdot, u) \in C^q([s_0,T],\mathbb{R}^{n\times m})$ and

$$|\sigma(\cdot, x_\cdot, u)|_{q,s,t} \le L\left(|t-s|^\beta + |x|_{p,s,t}\right)$$

for every $s,t \in [s_0,T]$.

ii) For every $u,v \in \mathcal{U}$, $x,y \in C^p([s_0,T],\mathbb{R}^n)$, $j=1,\dots,m$ and $s,t\in[s_0,T]$ we have that

$$\begin{aligned}&|\sigma^j(t,x_t,v) - \sigma^j(t,y_t,u) - \sigma^j(s,x_s,v) + \sigma^j(s,y_s,u)|\\ &\le L\bigg((|x_t-y_t| + |v-u|)\left(|t-s|^\beta + |x_t-x_s| + |y_t-y_s|\right) + |x_t-y_t-x_s+y_s|\bigg).\end{aligned}$$

iii) For every $u,v \in \mathcal{U}$, $x,y \in C^p([s_0,T],\mathbb{R}^n)$, $j=1,\dots,m$ and $s,t\in[s_0,T]$ we have that

$$\begin{aligned}&|\sigma^j(\cdot, x_\cdot, u) - \sigma^j(\cdot, y_\cdot, v)|_{q,s,t}\\ &\le L(\|x-y\|_{\infty,s,t} + (v-u))\left(|t-s|^\beta + |x|_{q,s,t} + |y|_{q,s,t}\right) + L|x-y|_{q,s,t}.\end{aligned}$$

Note that by conditions $(H_1)-(H_3)$, we could exchange in i) the function σ by σ^j_x respectively σ^j_u for every $j=1,\dots,m$ such that $\sigma^j_x(\cdot,x_\cdot,u) \in C^q([s_0,T],\mathbb{R}^{n\times n})$ and $\sigma^j_u(\cdot,x_\cdot,u) \in C^q([s_0,T],\mathbb{R}^{n\times d})$. In ii) and iii) we can exchange the functions σ^j by $\sigma^j_{x_i}$ for $i=1,\dots,n$, respectively $\sigma^j_{u_k}$ for $k=1,\dots,d$.

Proof. (i): By the space Lipschitz and time Hölder condition of σ and since $x \in C^p([s_0,T],\mathbb{R}^n)$, we can estimate

$$\begin{aligned}|\sigma(\cdot,x_\cdot,u)|_{q,s,t} &\le \sup_{k\in\mathbb{N},\Pi_k\in\mathcal{P}([s,t])}\left(\sum_{i=0}^{k-1}|\sigma(t_{i+1},x_{t_{i+1}},u)-\sigma(t_i,x_{t_i},u)|^q\right)^{\frac{1}{q}}\\ &\le L\sup_{k\in\mathbb{N},\Pi_k\in\mathcal{P}([s,t])}\left(\sum_{i=0}^{k-1}\left(|t_{i+1}-t_i|^\beta + |x_{t_{i+1}}-x_{t_i}|\right)^q\right)^{\frac{1}{q}}\\ &\le L\sup_{k\in\mathbb{N},\Pi_k\in\mathcal{P}([s,t])}\left(\sum_{i=0}^{k-1}|t_{i+1}-t_i|^{q\beta}\right)^{\frac{1}{q}}\\ &\quad + L\sup_{k\in\mathbb{N},\Pi_k\in\mathcal{P}([s,t])}\left(\sum_{i=0}^{k-1}|x_{t_{i+1}}-x_{t_i}|^q\right)^{\frac{1}{q}}\\ &\le L\left(|t-s|^\beta + |x|_{p,s,t}\right).\end{aligned}$$

where we used Lemma 2.2 and the inequalities $p<q\in(2,\frac{p}{p-1})$, $q\beta\ge 1$. We can conclude that $\sigma(\cdot,x_\cdot,u)\in C^q([s_0,T],\mathbb{R}^{n\times m})$.
(ii): is similar to the proof of Lemma 7.1 in Nualart and Răşcanu [2002]. Since $\mathcal{U}$ is convex, we

can use Lemma 2.25 and get

$$
\begin{aligned}
&|\sigma^j(t, x_t, v) - \sigma^j(t, y_t, u) - \sigma^j(s, x_s, v) - \sigma^j(s, y_s, u)| \\
&= \bigg| \int_0^1 \sigma^j_x(t, y_t + \lambda(x_t - y_t), u + \lambda(v-u))(x_t - y_t) + \sigma^j_u(t, y_t + \lambda(x_t - y_t), u + \lambda(v-u))(v-u) \\
&\quad - \sigma^j_x(s, y_s + \lambda(x_s - y_s), u + \lambda(v-u))(x_s - y_s) + \sigma^j_u(s, y_s + \lambda(x_s - y_s), u + \lambda(v-u))(v-u) \bigg|.
\end{aligned}
$$

By adding the term $\sigma^j_x(s, y_s + \lambda(x_s - y_s), u + \lambda(v-u))(x_t - y_t)$, we obtain

$$
\begin{aligned}
&= \bigg| \int_0^1 \left[\sigma^j_x(t, y_t + \lambda(x_t - y_t), u + \lambda(u-v)) - \sigma^j_x(s, y_s + \lambda(x_s - y_s), u + \lambda(u-v))\right](x_t - y_t) \\
&\quad + \sigma^j_x(s, y_s + \lambda(x_s - y_s), u + \lambda(v-u))(x_t - y_t - x_s + y_s) \\
&\quad + \left[\sigma^j_u(t, y_t + \lambda(x_t - y_t), u + \lambda(u-v)) - \sigma^j_u(s, y_s + \lambda(x_s - y_s), u + \lambda(u-v))\right](v-u)\, d\lambda \bigg| \\
&\leq L(|x_t - y_t| + |v-u|) \Big(|t-s|^\beta + |x_t - x_s| + |y_t - y_s|\Big) + L|x_t - y_t - x_s + y_s|.
\end{aligned}
$$

(iii): Using (ii) and the fact that $\varphi(s,t) = |t-s|^x$ for $x \geq 1$ is a control function on $[0,T]$, we obtain

$$
\begin{aligned}
&|\sigma^j(\cdot, x_\cdot, v) - \sigma^j(\cdot, y_\cdot, u)|_{q,s,t} \\
&\leq \sup_{k\in\mathbb{N}, \Pi_k \in \mathcal{P}([s,t])} \left(\sum_{i=0}^{k-1} |\sigma^j(t_{i+1}, x_{t_{i+1}}, v) - \sigma^j(t_{i+1}, y_{t_{i+1}}, u) - \sigma(t_i, x_{t_i}, v) + \sigma(t_i, y_{t_i}, u)|^q\right)^{\frac{1}{q}} \\
&\leq L(\|x-y\|_{\infty,s,t} + (v-u)) \Bigg[\sup_{k\in\mathbb{N}, \Pi_k \in \mathcal{P}([s,t])} \left(\sum_{i=0}^{k-1} |t_{i+1} - t_i|^{q\beta}\right)^{\frac{1}{q}} \\
&\quad + \sup_{k\in\mathbb{N}, \Pi_k \in \mathcal{P}([s,t])} \left(\sum_{i=0}^{k-1} |x_{t_{i+1}} - x_{t_i}|^q\right)^{\frac{1}{q}} + \sup_{k\in\mathbb{N}, \Pi_k \in \mathcal{P}([s,t])} \left(\sum_{i=0}^{k-1} |y_{t_{i+1}} - y_{t_i}|^q\right)^{\frac{1}{q}} \Bigg] \\
&\quad + \sup_{k\in\mathbb{N}, \Pi_k \in \mathcal{P}([s,t])} \left(\sum_{i=0}^{k-1} |x_{t_{i+1}} - y_{t_{i+1}} - x_{t_i} + y_{t_i}|^q\right)^{\frac{1}{q}} \\
&= L(\|x-y\|_{\infty,s,t} + (v-u))(|t-s|^\beta + |x|_{q,s,t} + |y|_{q,s,t}) + L|x-y|_{q,s,t}.
\end{aligned}
$$

The assertion follows, since $q > p$. $\square$

By Corollary 2.24 there exists a unique solution $x^u_\cdot$ corresponding to the YDE

$$
x^u_t = x_0(u) + \int_0^t b(r, x^u_r, u)\, dr + \int_0^t \sigma(r, x^u_r, u)\, dw_r
$$

$$= x_0(u) + \int_0^t b(r, x_r^u, u)\, dr + \sum_{j=1}^m \int_0^t \sigma^j(r, x_r^u, u)\, dw_r^j \tag{2.22}$$

for every $u \in \mathcal{U}$. We denote $b^u(\,\cdot\,) := b(\,\cdot\,, x_\cdot^u, u)$ and $\sigma^u(\,\cdot\,) := \sigma(\,\cdot\,, x_\cdot^u, u)$, respectively $\sigma^{j,u}(\,\cdot\,)$ for all $t \in [0, T]$.

Lemma 2.27. *Assume that x_0, σ and b satisfy the conditions (H_1), (H_2) and (H_3). Let $p \in (1, 2)$, $q > 2$ such that $\frac{1}{p} + \frac{1}{q} > 1$. Denote by x^u the solution of the equation (2.22) on $[0, T]$ and let $w \in C^p([0, T], \mathbb{R}^m)$. Define the constant*

$$C_1 := 2 \max\{L, LC_{p,q}, LC_{p,q}T^\beta, 1\}, \tag{2.23}$$

where L is the constant from conditions $(H_1) - (H_3)$ and $C_{p,q} = \zeta(\frac{1}{p} + \frac{1}{q_0}) \geq 1$ the constant from the Love-Young estimate (2.5). Then we have

$$|x^u|_{p,0,T} \leq 2^{2p-1} C_1^p (T^p + |w|_{p,0,T}^p) \tag{2.24}$$

and

$$\|x^u\|_{\infty,0,T} \leq \|x^u\|_{p,0,T} \leq L + 2^{2p-1} C_1^p (T^p + |w|_{p,0,T}^p). \tag{2.25}$$

Proof. Let $[s, t] \subset [0, T]$ and we drop the index u for the direct dependence of the solution process x on u for readability. We take a look at the term

$$|x|_{p,s,t} \leq \left| \int_0^\cdot b(r, x_r, u)\, dr \right|_{p,s,t} + \left| \int_0^\cdot \sigma(r, x_r, u)\, dw_r \right|_{p,s,t}.$$

The first integral can easily be estimated using the definition of the p-variation and the boundedness of b

$$\begin{aligned}
\left| \int_0^\cdot b(r, x_r, u) dr \right|_{p,s,t} &= \sup_{k \in \mathbb{N}, \Pi_k \in \mathcal{P}([s,t])} \left(\sum_{i=0}^{k-1} \left| \int_{t_i}^{t_{i+1}} b^u(r)\, dr \right|^p \right)^{\frac{1}{p}} \\
&\leq \sup_{k \in \mathbb{N}, \Pi_k \in \mathcal{P}([s,t])} \left(\sum_{i=0}^{k-1} \left(\int_{t_i}^{t_{i+1}} L\, dr \right)^p \right)^{\frac{1}{p}} \\
&\leq L(t - s).
\end{aligned}$$

The second integral can be controlled using (2.5). Taking Lemma 2.26 into account, we know that $\sigma(\cdot, x_\cdot, u) \in C^q([0, T], \mathbb{R}^{n \times m})$ for $q \in (2, \frac{p}{p-1})$, which yields

$$\begin{aligned}
\left| \int_0^\cdot \sigma^u(r) dw_r \right|_{p,s,t} &\leq C_{p,q} \|\sigma^u\|_{q,s,t} |w|_{p,s,t} \\
&\leq C_{p,q} (|\sigma^u(s)| + |\sigma^u|_{q,s,t}) |w|_{p,s,t}
\end{aligned}$$

$$\leq C_{p,q}(L + L(|t-s|^\beta + |x|_{p,s,t}))|w|_{p,s,t}.$$

Collecting the previous results, it follows

$$\begin{aligned}
|x|_{p,s,t} &\leq L(t-s) + LC_{p,q}|w|_{p,s,t}\left(1 + T^\beta + |x|_{p,s,t}\right) \\
&\leq C_1((t-s) + |w|_{p,s,t}) + |w|_{p,s,t}|x|_{p,s,t}) \\
&\leq C_1(1 + |x|_{p,s,t})((t-s) + |w|_{p,s,t})
\end{aligned} \tag{2.26}$$

for the constant $C_1 := 2\max\{L, LC_{p,q}, LC_{p,q}T^\beta, 1\}$. Now for every $[s,t] \in [0,T]$ such that

$$(t-s) + |w|_{p,s,t} \leq \frac{1}{2C_1}$$

we have

$$|x|_{p,s,t} \leq 1.$$

Hence the condition of Lemma 2.20 with constant bound and $\varphi(s,t) = |w|^p_{p,s,t}$ are satisfied, which yields the inequalities (2.24) and (2.25). □

Remark 2.28. In the proof of the last lemma we established for every $u \in \mathcal{U}$, that

$$|x^u|_{p,s,t} \leq 1$$

for every $s \leq t \in [0,T]$ such that $|t-s| + |w|_{p,s,t} \leq \frac{1}{2C_1}$, where C_1 is the constant defined in (2.23) and x^u is the unique solution to equation (2.22).

We have proven that the solution to our parameter dependent YDE is bounded independently of the parameter. Now we show the continuity of the solution mapping with respect to the parameter in the p-variation and consequently the uniform norm.

Lemma 2.29. *In the situation of Lemma 2.27, we have for every $\bar{u} \in \mathbb{R}^d$ such that $u + \bar{u} \in \mathcal{U}$, that*

$$\begin{aligned}
|x^{u+\bar{u}} - x^u|_{p,0,T} &\leq (1+L)|\bar{u}|e^{2^{3p}(3C_1m)^p\left(T^p + |w|^p_{p,0,T}\right)} \\
\|x^{u+\bar{u}} - x^u\|_{\infty,0,T} &\leq \|x^{u+\bar{u}} - x^u\|_{p,0,T} \\
&\leq (1+2L)|\bar{u}|e^{2^{3p}(3C_1m)^p\left(T^p + |w|^p_{p,0,T}\right)},
\end{aligned} \tag{2.27}$$

where the constant C_1 is defined by (2.23)*. This implies*

$$\lim_{|\bar{u}|\to 0} \|x^{u+\bar{u}} - x^u\|_{\infty,0,T} \leq \lim_{|\bar{u}|\to 0} \|x^{u+\bar{u}} - x^u\|_{p,0,T} = 0$$

for every $u \in \mathcal{U}$.

Proof. Set $v := u + \bar{u} \in \mathcal{U}$ and $\gamma^v_\cdot := x^v_\cdot - x^u_\cdot$. We obtain for $[s,t] \subset [0,T]$

$$|\gamma^v|_{p,s,t} \leq \left| \int_0^\cdot b^v(r) - b^u(r)\, dr \right|_{p,s,t} + \sum_{j=1}^m \left| \int_0^\cdot \sigma^{v,j}(r) - \sigma^{u,j}(r)\, dw^j_r \right|_{p,s,t} = I_1 + I_2.$$

The first term can be estimated using the Lipschitz continuity of b, which yields

$$\begin{aligned}
I_1 &= \sup_{k\in\mathbb{N},\Pi_k\in\mathcal{P}([s,t])} \left(\sum_{i=0}^{k-1} \left| \int_{t_i}^{t_{i+1}} b^v(r) - b^u(r)\, dr \right|^p \right)^{\frac{1}{p}} \\
&\leq L \sup_{k\in\mathbb{N},\Pi_k\in\mathcal{P}([s,t])} \left(\sum_{i=0}^{k-1} \left(\int_{t_i}^{t_{i+1}} |\gamma^v_r| + |\bar{u}|\, dr \right)^p \right)^{\frac{1}{p}} \\
&\leq L \sup_{k\in\mathbb{N},\Pi_k\in\mathcal{P}([s,t])} \left(\sum_{i=0}^{k-1} (t_{i+1} - t_i)^p (\|\gamma^v\|_{\infty,s,t} + |\bar{u}|)^p \right)^{\frac{1}{p}} \\
&\leq L(\|\gamma^v\|_{\infty,s,t} + |\bar{u}|)(t-s) \\
&\leq L(|\gamma^v_s| + |\gamma^v|_{p,s,t} + |\bar{u}|)(t-s).
\end{aligned}$$

Again we use (2.5) and Lemma 2.26 for the estimation of I_2. Let $q > 2$ and such that $\frac{1}{p} + \frac{1}{q} > 1$, we obtain

$$I_2 \leq C_{p,q} \sum_{j=1}^m (\|\sigma^{v,j} - \sigma^{u,j}\|_{\infty,s,t} + |\sigma^{v,j} - \sigma^{u,j}|_{q_0,s,t}) |w^j|_{p,s,t},$$

where for all $j = 1, \ldots, m$

$$\|\sigma^{v,j} - \sigma^{u,j}\|_{\infty,s,t} \leq L(\|\gamma^v\|_{\infty,s,t} + |\bar{u}|).$$

To estimate the q-variation term we use Lemma 2.26 iii) and get

$$|\sigma^{v,j} - \sigma^{u,j}|_{q,s,t} \leq L(\|\gamma^v\|_{\infty,s,t} + |\bar{u}|) \left(|t-s|^\beta + |x^v|_{p,s,t} + |x^u|_{p,s,t} \right) + L|\gamma^v|_{p,s,t}.$$

Now we are able to estimate the Young integral term I_2 by combining the last 3 inequalities, which yields

$$\begin{aligned}
I_2 &\leq LC_{p,q} \left[(|\gamma^v|_{\infty,s,t} + |\bar{u}|) \left(1 + T^\beta + |x^v|_{p,s,t} + |x^u|_{p,s,t} \right) + |\gamma^v|_{p,s,t} \right] \sum_{j=1}^m |w^j|_{p,s,t} \\
&\leq m2LC_{p,q} (|\gamma^v_s| + |\gamma^v|_{p,s,t} + |\bar{u}|) \left(1 + T^\beta + |x^v|_{p,s,t} + |x^u|_{p,s,t} \right) |w|_{p,s,t},
\end{aligned}$$

since $\sum_{j=1}^m |w^j|_{p,s,t} \leq m|w|_{p,s,t}$. Putting all terms together, we obtain

$$|\gamma^v|_{p,s,t} \leq L(|\gamma^v_s| + |\gamma^v|_{p,s,t} + |\bar{u}|)(t-s)$$

$$
\begin{aligned}
&+ m2LC_{p,q}(|\gamma^v_s| + |\gamma^v|_{p,s,t} + |\bar{u}|)(1 + T^\beta + |x^v|_{p,s,t} + |x^u|_{p,s,t})|w|_{p,s,t} \\
&\leq 2C_1 m(|\gamma^v_s| + |\gamma^v|_{p,s,t} + |\bar{u}|)(1 + |x^v|_{p,s,t} + |x^u|_{p,s,t})\left((t-s) + |w|_{p,s,t}\right),
\end{aligned}
$$

where C_1 is defined by (2.23). Now let $[s,t] \subset [0,T]$, such that

$$
(t-s) + |w|_{p,s,t} \leq \frac{1}{12C_1 m} < \frac{1}{2C_1}.
$$

Then we know by Remark 2.28 that

$$
\max\{|x^u|_{p,s,t}, |x^v|_{p,s,t}\} \leq 1,
$$

which yields

$$
|\gamma^v|_{p,s,t} \leq |\gamma^v_s| + |\bar{u}|.
$$

By our Gronwall-type lemma 2.20 with $\varphi(s,t) = |w|^p_{p,s,t}$, we obtain the estimates

$$
|\gamma^v|_{p,0,T} \leq (|\bar{u}| + |\gamma^v_0|)e^{2^{3p}(3C_1 m)^p\left(T^p + |w|^p_{p,0,T}\right)}
$$

and

$$
\|\gamma^v\|_{\infty,0,T} \leq \|\gamma^v\|_{p,0,T} \leq (|\bar{u}| + 2|\gamma^v_0|)e^{2^{3p}(3C_1 m)^p\left(T^p + |w|^p_{p,0,T}\right)}.
$$

We have $\gamma^v_0 = x_0(v) - x_0(u)$ and by condition (H_1), we know that the function $x_0 : \mathcal{U} \to \mathbb{R}^n$ is Lipschitz continuous, such that

$$
\|\gamma^v\|_{p,0,T} \leq (1 + 2L)|\bar{u}|e^{2^{3p}(3C_1 m)^p\left(T^p + |w|^p_{p,0,T}\right)}.
$$

Hence, the assertion follows. □

The rest of this subsection is devoted to prove the Fréchet differentiability of the solution mapping $u \mapsto x^u_\cdot$ from $\mathcal{U}$ to $C^p([0,T],\mathbb{R}^n)$. For $t \in [0,T]$ and $u \in \mathcal{U}$, we use the compact notation

$$
b^u_x(t) := b_x(t, x^u_t, u), \quad \sigma^{u,j}_x(t) := \sigma^j_x(t, x^u_t, u), \quad b^u_u(t) := b_u(t, x^u_t, u), \quad \sigma^{u,j}_u(t) := \sigma^j_u(t, x^u_t, u).
$$

for $j = 1, \ldots, m$.

Lemma 2.30. *In the situation of Lemma 2.27, we have*

$$
\lim_{|\bar{u}|\to 0} \left\| \frac{x^{u+\bar{u}} - x^u - y^u\bar{u}}{|\bar{u}|} \right\|_{\infty,0,T} \leq \lim_{|\bar{u}|\to 0} \left\| \frac{x^{u+\bar{u}} - x^u - y^u\bar{u}}{|\bar{u}|} \right\|_{p,0,T} = 0,
$$

where $y^u_t \in C^p([0,T],\mathbb{R}^{n\times d})$ is the matrix valued solution to the linear YDE

$$y^u_t = Dx_0(u) + \int_0^t b^u_x(r)y^u_r + b^u_u(r)\,dr + \sum_{j=1}^m \int_0^t \sigma^{u,j}_x(r)y^u_r + \sigma^{u,j}_u(r)\,dw^j_r. \tag{2.28}$$

Before we can proof Lemma 2.30, we first need to take a look at the inhomogenous linear YDE (2.28) and clarify that the solution y^u_t exits and is unique. Similar to the ODE case, we give an explicit solution for y^u by using solutions to the corresponding matrix-valued homogenous linear YDEs. So we will define the needed matrix valued YDEs, give the explicit solution to equation (2.28) and then establish its boundedness in $C^p([0,T],\mathbb{R}^{n\times d})$. Hence we get a bounded linear operator from $\mathbb{R}^d$ to $C^p([0,T],\mathbb{R}^n)$, and by Lemma 2.30 this is the Fréchet derivative of the solution mapping $u\mapsto x^u$. We define for every $u\in\mathcal{U}$ and the corresponding solution x^u to equation (2.22), the matrix valued homogenous linear YDEs with initial time s_0 for $0\le s_0\le T$ as

$$\begin{aligned}\phi^{s_0}_t &= I_n + \int_{s_0}^t b_x(r,x^u_r,u)\phi^{s_0}_r\,dr + \sum_{j=1}^m \int_{s_0}^t \sigma^j_x(r,x^u_r,u)\phi^{s_0}_r\,dw^j_r\\ &= I_n + \int_{s_0}^t b^u_x(r)\phi^{s_0}_r\,dr + \sum_{j=1}^m \int_{s_0}^t \sigma^{u,j}_x(r)\phi^{s_0}_r\,dw^j_r\end{aligned} \tag{2.29}$$

and

$$\begin{aligned}\psi^{s_0}_t &= I_n - \int_{s_0}^t \psi^{s_0}_t b_x(r,x^u_r,u)\,dr - \sum_{j=1}^m \int_{s_0}^t \psi^{s_0}_r\sigma^j_x(r,x^u_r,u)\,dw^j_r\\ &= I_n - \int_{s_0}^t \psi^{s_0}_t b^u_x(r)\,dr - \sum_{j=1}^m \int_{s_0}^t \psi^{s_0}_r\sigma^{u,j}_x(r)\,dw^j_r\end{aligned} \tag{2.30}$$

Here we left out the index u for the dependence of the solution processes on u for readability. Note that the introduction of an arbitrary initial time $s_0\in[0,T]$ will only be important later in Section 4.2. Now we use Theorem 2.22 to show that both YDEs have a solution in $C^p([s_0,T],\mathbb{R}^{n\times n})$.

Lemma 2.31. *Let $p\in(1,2)$ and $w\in C^p([0,T],\mathbb{R}^m)$. Consider the matrix valued linear YDEs (2.29) and (2.30), where the coefficient functions b and σ satisfy conditions (H_2) and (H_3). Let $x_0:\mathcal{U}\to\mathbb{R}^n$ be a function satisfying condition (H_1), then we know that for each $u\in\mathcal{U}$ there exists a solution x^u to the YDE (2.22) in the space $C^p([s_0,T],\mathbb{R}^n)$ and both of the matrix valued YDEs have a unique solution in the space $C^p([s_0,T],\mathbb{R}^{n\times n})$. Furthermore they are bounded independently of u by the estimate*

$$\begin{aligned}\max\{|\psi^{s_0}|_{p,s_0,T}, |\phi^{s_0}|_{p,s_0,T}\} &\le \sqrt{n}e^{2^{4p}(C_1m)^p\left(T^p+|w|^p_{p,0,T}\right)}\\ \max\{\|\psi^u\|_{\infty,s_0,T}, \|\phi^u\|_{\infty,s_0,T}\} &\le \max\{\|\psi^u\|_{p,s_0,T}, \|\phi^u\|_{p,s_0,T}\}\end{aligned}$$

$$\leq 2\sqrt{n}e^{2^{4p}(C_1 m)^p\left(T^p+|w|^p_{p,0,T}\right)},$$

where the constant C_1 is given by (2.23).

Proof. For readability we leave out the direct dependence of x, ϕ and ψ on u and s_0. First we can express equation (2.30) by its matrix transposed

$$\psi_t^\top = I_n - \int_{s_0}^t b_x^u(r)^\top \psi_r^\top \, dr - \sum_{j=1}^m \int_{s_0}^t \sigma_x^{u,j}(r)^\top \psi_r^\top \, dw_r^j$$

and that for $X \in C^p([s_0,T],\mathbb{R}^{n\times n})$, we have

$$|X| = |X^\top|, \quad \|X\|_{\infty,s_0,T} = \|X^\top\|_{\infty,s_0,T}, \quad \|X\|_{p,s_0,T} = \|X^\top\|_{p,s_0,T}.$$

So we will show existence and uniqueness of equation (2.29) and also estimate its p-variation norm. The results follow directly for ψ by the aforementioned arguments. Theorem 2.22 states that the unique solution to equation (2.29) exists, if $\|b_x^u\|_{\infty,s_0,T}$ and $\|\sigma_x^u\|_{q,s_0,T}$ are finite for $q > p$ and $\frac{1}{p} + \frac{1}{q} > 1$. Choose $q \in (2, \frac{p}{p-1})$. The boundedness of $\|b_x^u\|_{\infty,s_0,t}$ is a direct consequence of condition (H_2) and for $\sigma_x^{u,j}$, we have by Lemma 2.26 and (H_3), that

$$\|\sigma_x^j(\cdot, x_\cdot, u)\|_{q,s_0,T} \leq L\left(1 + |T - s_0|^\beta + |x|_{p,s_0,T}\right). \tag{2.31}$$

Since $x \in C^p([0,T],\mathbb{R}^n)$ the boundedness of $\|\sigma_x^j\|_{q,s_0,T}$ follows. Hence by Theorem 2.22, there exists a unique solution $\phi \in C^p([s_0,T],\mathbb{R}^{n\times n})$. Now we want to find an upper bound of the p-variation norm of ϕ and ψ which is independent of u. We have for $[s,t] \subset [s_0,T]$

$$|\phi|_{p,s,t} \leq \left|\int_{s_0}^{\cdot} b_x^u(r)\phi_r \, dr\right|_{p,s,t} + \sum_{j=1}^m \left|\int_{s_0}^{\cdot} \sigma_x^{u,j}(r)\phi_r \, dw_r^j\right|_{p,s,t} = I_1 + I_2.$$

With condition (H_2) it is clear that

$$\begin{aligned}
I_1 &= \sup_{k\in\mathbb{N},\Pi_k\in\mathcal{P}([s,t])} \left(\sum_{i=0}^{k-1} \left|\int_{t_i}^{t_{i+1}} b_x^u(r)\phi_r \, dr\right|^p\right)^{\frac{1}{p}} \\
&\leq L \sup_{k\in\mathbb{N},\Pi_k\in\mathcal{P}([s,t])} \left(\sum_{i=0}^{k-1} \left(\int_{t_i}^{t_{i+1}} |\phi_r| \, dr\right)^p\right)^{\frac{1}{p}} \\
&\leq L \sup_{k\in\mathbb{N},\Pi_k\in\mathcal{P}([s,t])} \left(\sum_{i=0}^{k-1} (t_{i+1} - t_i)^p \|\phi\|_{\infty,s,t}^p\right)^{\frac{1}{p}} \\
&\leq L\|\phi\|_{\infty,s,t}(t-s) \\
&\leq L\|\phi\|_{p,s,t}(t-s).
\end{aligned}$$

The second integral can be controlled using (2.5). Taking (2.31) into account, we know that $\sigma_x^j(\cdot, x_\cdot^u, u) \in C^q([s_0, T], \mathbb{R}^{n\times m})$ for $q > 2$, $\frac{1}{p} + \frac{1}{q} > 1$, which yields

$$I_2 \leq \sum_{j=1}^{m} \left| \int_{s_0}^{\cdot} \sigma_x^{u,j}(r)\phi_r \, dw_r^j \right|_{p,s,t} \leq C_{p,q} \sum_{j=1}^{m} |w^j|_{p,s,t} \|\sigma_x^{u,j}\phi\|_{q,s,t}.$$

By Lemma 2.10 and the same arguments used to estimate (2.31), we obtain for $j = 1, \dots, m$

$$\begin{aligned}
\|\sigma_x^{u,j}\phi\|_{q,s,t} &\leq 2\|\sigma_x^{u,j}\|_{q,s,t}\|\phi\|_{q,s,t} \\
&\leq 2L\left(1 + T^\beta + |x|_{q,s,t}\right)\|\phi\|_{q,s,t} \\
&\leq 2L\left(1 + T^\beta + |x|_{p,s,t}\right)\|\phi\|_{p,s,t}.
\end{aligned}$$

Collecting the previous results, it follows

$$\begin{aligned}
|\phi|_{p,s,t} &\leq L\|\phi\|_{p,s,t}(t-s) + m2LC_{p,q}\left(1 + T^\beta + |x|_{p,s,t}\right)\|\phi\|_{p,s,t}|w|_{p,s,t} \\
&\leq 2C_1 m\|\phi\|_{p,s,t}(1 + |x|_{p,s,t})\left(|t-s| + |w|_{p,s,t}\right) \\
&\leq 2C_1 m(|\phi_s| + |\phi|_{p,s,t})(1 + |x|_{p,s,t})\left(|t-s| + |w|_{p,s,t}\right) \qquad (2.32)
\end{aligned}$$

for every $[s,t] \subset [s_0, T]$, where C_1 is defined in (2.23). We prolong ϕ to the interval $[0,T]$, by setting $\phi_t = I_n$ for $t \in [0, s_0]$. Then $\phi \in C^p([0,T], \mathbb{R}^{n\times n})$ and the estimate (2.32) holds for all $[s,t] \subset [0,T]$. Now let $[s,t] \subset [0,T]$ such that

$$(t-s) + |w|_{p,s,t} \leq \frac{1}{8C_1 m} \leq \frac{1}{2C_1 m}.$$

Then, we know by Remark 2.28 that

$$|x|_{p,s,t} \leq 1,$$

which yields

$$|\phi|_{p,s,t} \leq |\phi_s|.$$

By our Gronwall-type lemma 2.20, with $\varphi(s,t) = |w|_{p,s,t}^p$ and $K_1 = 0$, we obtain the estimate

$$|\phi|_{p,s_0,T} = |\phi|_{p,0,T} \leq \sqrt{n} e^{2^{4p}(C_1 m)^p\left(T^p + |w|_{p,0,T}^p\right)},$$

which implies

$$\begin{aligned}
|\phi|_{\infty,0,T} &\leq \|\phi\|_{p,0,T} \\
&\leq 2\sqrt{n} e^{2^{4p}(C_1 m)^p\left(T^p + |w|_{p,0,T}^p\right)}.
\end{aligned}$$

□

Using the integration by parts formula for Young integrals we can establish the relationship between the processes ϕ and ψ.

Lemma 2.32. *Let ϕ^{s_0} and ψ^{s_0} be defined by* (2.29) *and* (2.30)*, then $\psi_t^{s_0} = (\phi_t^{s_0})^{-1}$ for all $t \in [s_0, T]$.*

Proof. First we drop the index s_0 for readability. We prove that $\psi_t \phi_t = I_n$ for all $t \in [s_0, T]$ using the integration by parts formula from Lemma 2.18 and the substitution rule from Lemma 2.17, we have

$$\begin{aligned} d(\psi_t \phi_t) &= \psi_t \, d\phi_t + (d\psi_t)\, \phi_t \\ &= \psi_t b_x^u(t) \phi_t \, dt + \sum_{j=1}^m \psi_t \sigma_x^{u,j}(t) \phi_t \, dw_t^j - \psi_t b_x^u(t) \phi_t \, dt - \sum_{j=1}^m \psi_t \sigma_x^{u,j}(t) \phi_t \, dw_t^j \\ &= 0. \end{aligned}$$

Noting that $\psi_{s_0} \phi_{s_0} = I_n$, we must have $\psi_t = \phi_t^{-1}$ for all $t \in [s_0, T]$. □

Now we give the explicit solution to equation (2.28) using the previous results.

Lemma 2.33. *Let $T > 0$ and $p \in (1, 2)$. Let $\phi = \phi^0$ and $\phi = \psi^0$ be defined by* (2.29) *and* (2.30) *for $s_0 = 0$. Then the solution to equation* (2.28) *on $[0, T]$ for $u \in \mathcal{U}$ is given by*

$$y_t^u = \phi_t D x_0(u) + \phi_t \int_0^t \phi_r^{-1} b_u^u(r) \, dr + \sum_{j=1}^m \phi_t \int_0^t \phi_r^{-1} \sigma_u^{u,j}(r) \, dw_r^j. \tag{2.33}$$

Furthermore this solution is unique and $y_t \in C^p([0,T], \mathbb{R}^{n \times d})$. Hence for every $u \in \mathcal{U}$ the solution process y_t^u defines a bounded linear operator $Dx_t^u := y_t^u$ from $\mathbb{R}^d$ to the space $C^p([0,T], \mathbb{R}^n)$.

Proof. Again we drop the index u on the function y for simplicity. Define the function

$$\gamma_t = D x_0(u) + \int_0^t \phi_r^{-1} b_u^u(r) \, dr + \sum_{j=1}^m \int_0^t \phi_r^{-1} \sigma_u^{u,j}(r) \, dw_r^j.$$

Since $\phi^{-1} \in C^p([0,T], \mathbb{R}^{n \times n})$ by Lemma 2.31 and $\sigma_u^{u,j} \in C^q([0,T], \mathbb{R}^{n \times d})$ for $q \in (2, \frac{p}{p-1})$ by Lemma 2.26, it is easy to see that γ is an element of $C^p([0,T], \mathbb{R}^{n \times d})$. We know that $\phi \in C^p([0,T], \mathbb{R}^{n \times n})$, hence, by Lemma 2.18 we get

$$\begin{aligned} d(\phi_t \gamma_t) &= \phi_t \, d\gamma_t + (d\phi_t)\, \gamma_t \\ &= b_u^u(t) \, dt + \sum_{j=1}^m \sigma_u^{u,j}(t) \, dw_t^j + b_x^u(t) \phi_t \gamma_t \, dt + \sum_{j=1}^m \sigma_x^{u,j}(t) \phi_t \gamma_t \, dw_t^j \\ &= b_x^u(t) \phi_t \gamma_t + b_u^u(t) \, dt + \sum_{j=1}^m \sigma_x^{u,j}(t) \phi_t \gamma_t + \sigma_u^{u,j}(t) \, dw_t^j. \end{aligned}$$

Since $\phi_0\gamma_0 = \gamma_0 = Dx_0(u)$, we conclude that $y_t = \phi_t\gamma_t \in C^p([0,T],\mathbb{R}^{n\times d})$ is a solution to the YDE (2.28). Now assume for a given parameter $u \in \mathcal{U}$ there are two solutions y^1 and y^2 to equation 2.28 on $[0,T]$ with $y_0^1 = y_0^2 = Dx_0$ then $z_t = y_t^1 - y_t^2$ satisfies the homogenous linear equation

$$z_t = \int_0^t b_x^u(r) z_r \, dr + \sum_{j=1}^m \int_0^t \sigma_x^{u,j}(r) z_r \, dw_r^j.$$

By Theorem 2.22, this equation has a unique solution, which then has to be 0. Hence the equation (2.28) has a unique solution in the space $C^p([0,T],\mathbb{R}^{n\times d})$, which is given by (2.33). The operator $Dx_\cdot^u : \mathbb{R}^d \to C^p([0,T],\mathbb{R}^n)$ is obviously linear. Notice that for a matrix valued function $X \in C^p([0,T],\mathbb{R}^{n\times d})$ and a vector $v \in \mathbb{R}^d$, we have for all $s \le t \in [0,T]$

$$\|Xv\|_{p,s,t} \le |X_s||v| + |X|_{p,s,t}|v| = \|X\|_{p,s,t}|v|,$$

because of the submultiplicativity of the Frobenius norm. Hence, we obtain for the operator norm of Dx^u

$$\|Dx_\cdot^u\| = \sup_{|z|=1} \|Dx_\cdot^u z\|_{p,0,T} \le \|y^u\|_{p,0,T} < \infty.$$

□

We have shown that a unique solution to equation (2.28) exists in the space $C^p([0,T],\mathbb{R}^{n\times d})$. In the next lemma we will give an estimate for the p-variation norm and the uniform norm of the solution which is independent of the parameter u.

Lemma 2.34. *Let $p \in (1,2)$ and $w \in C^p([0,T],\mathbb{R}^m)$. Consider the matrix valued linear YDE* (2.28)*, where the coefficient functions b and σ satisfy conditions (H_2) and (H_3). Let $x_0 : \mathcal{U} \to \mathbb{R}^n$ be a function satisfying condition (H_1), then we know that for each $u \in \mathcal{U}$ there exists a solution x^u to the YDE* (2.22) *in the space $C^p([0,T],\mathbb{R}^n)$ and the matrix valued YDE* (2.28) *has a unique solution in the space $C^p([0,T],\mathbb{R}^{n\times d})$. This solution is bounded independently of u by the estimate*

$$\begin{aligned} |y^u|_{p,0,T} &\le (1+L)e^{2^{4p}(C_1 m)^p\left(T^p + |w|_{p,0,T}^p\right)} \\ \|y^u\|_{\infty,0,T} \le \|y^u\|_{p,0,T} &\le (1+2L)e^{2^{4p}(C_1 m)^p\left(T^p + |w|_{p,0,T}^p\right)}, \end{aligned}$$

where the constant C_1 is given by (2.23)*.*

Proof. Again we leave out the index u for the processes x and y for simplicity. We have for $s \le t \in [0,T]$, that

$$\begin{aligned} |y|_{p,s,t} &\le \left| \int_0^\cdot b_x^u(r) y_r + b_u^u(r) \, dr \right|_{p,s,t} + \sum_{j=1}^m \left| \int_0^\cdot \sigma_x^{u,j}(r) y_r + \sigma_u^{u,j}(r) \, dw_r^j \right|_{p,s,t} \\ &= I_1 + I_2. \end{aligned}$$

I_1 can easily be estimated by the assumed conditions on the coefficient b, we have

$$I_1 \leq L(1 + \|y\|_{\infty,s,t})|t-s|.$$

For the estimation of I_2 keep in mind that $\sigma_x^{j,u} \in C^q([0,T], \mathbb{R}^{n\times n})$, $\sigma_u^{u,j} C^q([0,T], \mathbb{R}^{n\times d})$ by Lemma 2.26 for $q \in (2, \frac{p}{p-1})$ and that $y \in C^p([0,T], \mathbb{R}^{n\times d})$ by Lemma 2.33. Hence by Lemma 2.10, we know that $\sigma_x^{u,j}(\cdot)y_\cdot + \sigma_u^{u,j}(\cdot) \in C^q([0,T], \mathbb{R}^{n\times d})$ for every $j = 1, \dots, m$ and by (2.5), we obtain

$$\begin{aligned}
I_2 &\leq C_{p,q} \sum_{j=1}^m \|\sigma_x^{u,j}(\cdot)y_\cdot + \sigma_u^{u,j}(\cdot)\|_{q,s,t} |w^j|_{p,s,t} \\
&\leq C_{p,q} \sum_{j=1}^m \left(2\|\sigma_x^{u,j}(\cdot)\|_{q,s,t}\|y\|_{p,s,t} + \|\sigma_u^{u,j}(\cdot)\|_{q,s,t}\right) |w^j|_{p,s,t}
\end{aligned}$$

Using Lemma 2.26 i) on $\sigma_x^{u,j}$ and $\sigma_u^{u,j}$ for $j = 1, \dots, m$, we get the estimate

$$\begin{aligned}
I_2 &\leq 2LC_{p,q}m(1 + T^\beta + |x|_{p,s,t})(1 + \|y\|_{p,s,t})|w|_{p,s,t} \\
&\leq 2C_1 m(1 + |x|_{p,s,t})(1 + \|y\|_{p,s,t})|w|_{p,s,t}.
\end{aligned}$$

Collecting the terms, this yields

$$|y|_{p,s,t} \leq 2C_1 m(1 + |x|_{p,s,t})(1 + |y_s| + |y|_{p,s,t})(|t-s| + |w|_{p,s,t}). \tag{2.34}$$

for every $[s,t] \subset [0,T]$. Now let $[s,t] \subset [0,T]$ such that

$$(t-s) + |w|_{p,s,t} \leq \frac{1}{8C_1 m} \leq \frac{1}{2C_1 m}.$$

Then, we know by Remark 2.28 that

$$|x|_{p,s,t} \leq 1,$$

which yields

$$|y|_{p,s,t} \leq 1 + |y_s|.$$

By our Gronwall-type lemma 2.20, with $\varphi(s,t) = |w|_{p,s,t}^p$ and $K_1 = 1$, we obtain the estimate

$$\begin{aligned}
|y|_{p,0,T} &\leq (1 + |y_0|)e^{2^{4p}(C_1 m)^p\left(T^p + |w|_{p,0,T}^p\right)} \\
&\leq (1 + |Dx_0(u)|)e^{2^{4p}(C_1 m)^p\left(T^p + |w|_{p,0,T}^p\right)} \\
&\leq (1 + L)e^{2^{4p}(C_1 m)^p\left(T^p + |w|_{p,0,T}^p\right)},
\end{aligned}$$

which implies

$$|y|_{\infty,0,T} \leq \|y\|_{p,0,T} \leq (1+2L)e^{2^{4p}(C_1 m)^p\left(T^p+|w|^p_{p,0,T}\right)}.$$

$\square$

Now we are able to proof Lemma 2.30.

Proof of Lemma 2.30. In this proof we use C for a positive constant, which can have different values at different occasions. To minimize the notational effort, we set $m=1$ for the proof, and leave out the indices for the dependence of y on u. Let $u \in \mathcal{U}$ and $\bar{u} \in \mathbb{R}^d$ such that $u+\bar{u} \in \mathcal{U}$ ($\mathcal{U}$ is open and convex), set

$$\begin{aligned}
v &= u + \bar{u} \\
\gamma^v_\cdot &= x^v_\cdot - x^u_\cdot \\
\eta^v_\cdot &= \frac{1}{|\bar{u}|}(\gamma^v_\cdot - y_\cdot \bar{u}),
\end{aligned}$$

then we can write for $0 \leq t \leq T$:

$$\begin{aligned}
\eta^v_t = &\frac{\gamma^v_0 - Dx_0(u)\bar{u}}{|\bar{u}|} + \frac{1}{|\bar{u}|}\int_0^t [b^v(r) - b^u(r) - b^u_x(r)y_r\bar{u} - b^u_u(r)\bar{u}]\, dr \\
&+ \frac{1}{|\bar{u}|}\int_0^t [\sigma^v(r) - \sigma^u(r) - \sigma^u_x(r)y_r\bar{u} - \sigma^u_u(r)\bar{u}]\, dw_r.
\end{aligned}$$

Hence, we obtain for $[s,t] \subset [0,T]$

$$|\eta^v|_{p,s,t} \leq I_1 + I_2,$$

where

$$I_1 = \left| \frac{1}{|\bar{u}|}\int_0^\cdot b^v(r) - b^u(r) - b^u_x(r)y_r\bar{u} - b^u_u(r)\bar{u}\, dr \right|_{p,s,t}$$

and

$$I_2 = \left| \frac{1}{|\bar{u}|}\int_0^\cdot \sigma^v(r) - \sigma^u(r) - \sigma^u_x(r)y_r\bar{u} - \sigma^u_u(r)\bar{u}\, dw_r \right|_{p,s,t}.$$

The integral I_1 can be estimated using the mean value theorem

$$\begin{aligned}
I_1 &= \left| \frac{1}{|\bar{u}|}\int_0^\cdot b^v(r) - b^u(r) - b^u_x(r)y_r\bar{u} - b^u_u(r)\bar{u}\, dr \right|_{p,s,t} \\
&= \sup_{k\in\mathbb{N},\Pi_k\in\mathcal{P}([s,t])} \left(\sum_{i=0}^{k-1} \left| \frac{1}{|\bar{u}|}\int_{t_i}^{t_{i+1}} b^v(r) - b^u(r) - b^u_x(r)y_r\bar{u} - b^u_u(r)\bar{u}\, dr \right|^p \right)^{\frac{1}{p}} \\
&= \sup_{k\in\mathbb{N},\Pi_k\in\mathcal{P}([s,t])} \left(\sum_{i=0}^{k-1} \left| \frac{1}{|\bar{u}|}\int_{t_i}^{t_{i+1}} \int_0^1 b_x(r, x^u_r + \lambda\gamma^v_r, u+\lambda\bar{u})\gamma^v_r + b_u(r, x^u_r + \lambda\gamma^v_r, u+\lambda\bar{u})\bar{u}\, d\lambda \right.\right.
\end{aligned}$$

$$- b_x^u(r)y_r\bar{u} - b_u^u(r)\bar{u})\,dr\Big|^p\Big)^{\frac{1}{p}}.$$

Adding the term $b_x(r, x_r^u + \lambda\gamma_r^v, u + \lambda\bar{u})y_r\bar{u}$ yields

$$\begin{aligned}
&= \sup_{k\in\mathbb{N},\Pi_k\in\mathcal{P}([s,t])} \Bigg(\sum_{i=0}^{k-1}\Bigg|\int_{t_i}^{t_{i+1}}\int_0^1 b_x(r, x_r^u + \lambda\gamma_r^v, u + \lambda\bar{u})\eta^v(r) \\
&\quad + (b_x(r, x_r^u + \lambda\gamma_r^v, u + \lambda\bar{u}) - b_x^u(r))\frac{y_r\bar{u}}{|\bar{u}|} \\
&\quad + (b_u(r, x_r^u + \lambda\gamma_r^v, u + \lambda\bar{u}) - b_u^u(r))\frac{\bar{u}}{|\bar{u}|}\,d\lambda\,dr\Bigg|^p\Bigg)^{\frac{1}{p}}.
\end{aligned}$$

By the Lipschitz continuity of b_x and b_u in the state and parameter variables and the boundedness of b_x, it follows

$$\begin{aligned}
&\le C \sup_{k\in\mathbb{N},\Pi_k\in\mathcal{P}([s,t])} \Bigg(\sum_{i=0}^{k-1}\Big([\|\eta^v\|_{\infty,s,t} + (\|\gamma^v\|_{\infty,s,t} + |\bar{u}|)\|y\|_{\infty,s,t} + (\|\gamma^v\|_{\infty,s,t} + |\bar{u}|)]\,(t_{i+1} - t_i)\Big)^p\Bigg)^{\frac{1}{p}} \\
&= C\,[\|\eta^v\|_{\infty,s,t} + (\|\gamma^v\|_{\infty,s,t} + |\bar{u}|)(\|y\|_{\infty,s,t} + 1)]\,(t - s).
\end{aligned}$$

The norm $\|\gamma^v\|_{\infty,s,t}$ can be estimated by $C|\bar{u}|$ according to Lemma 2.29 and $\|y\|_{\infty,s,t} \le C$ by Lemma 2.34, we conclude

$$I_1 \le C(t - s)(\|\eta^v\|_{\infty,s,t} + |\bar{u}|).$$

The estimation of I_2 is more extensive. It is based on the inequality (2.5) and multiple applications of the mean value theorem. Taking Lemma 2.26 into account, we choose q such that $q > 2$, $\frac{1}{p}+\frac{1}{q} \ge 1$ and show that for $\tilde{\sigma} := (\sigma^v - \sigma^u - \sigma_x^u y\bar{u} - \sigma_u^u\bar{u})/|\bar{u}|$, we have $\|\tilde{\sigma}\|_{q,0,T} = |\tilde{\sigma}(0)| + |\tilde{\sigma}|_{q,0,T} < \infty$. The first term can easily be estimated by

$$\begin{aligned}
|\tilde{\sigma}(0)| &\le \left|\frac{\sigma^v(0) - \sigma^u(0) - \sigma_x^u(0)y_0\bar{u} - \sigma_u^u(0)\bar{u}}{|\bar{u}|}\right| \\
&\le C\left(\frac{|x_0(v) - x_0(u)| + |\bar{u}|}{|\bar{u}|} + |Dx_0(u)| + 1\right).
\end{aligned}$$

By Condition (H_1), $x_0(\cdot)$ is differentiable with bounded differential (i.e. Lipschitz), which yields $|\tilde{\sigma}(0)| < C$, where C is independent of $\bar{u}$ and u. For the second term we calculate

$$|\tilde{\sigma}|_{q,0,T} = \frac{1}{|\bar{u}|}|\sigma^v - \sigma^u|_{q,0,T} + \frac{1}{|\bar{u}|}|\sigma_x^u y\bar{u} + \sigma_u^u\bar{u}|_{q,0,T} =: J_1 + J_2.$$

Similar to the proof of Lemma 2.29, we can estimate J_1 by

$$J_1 \le \frac{C}{|\bar{u}|}(|\gamma_0^v| + |\gamma^v|_{p,0,T} + |\bar{u}|)$$

and with the inequality (2.27) and the Lipschitz coninuity of $x_0(\cdot)$ we have

$$J_1 \leq \frac{C}{|\bar{u}|}(C|\bar{u}|) < \infty.$$

To calculate J_2, we note that by the submultiplicativity of the Frobenius norm we have

$$J_2 \leq |\sigma_x^u y + \sigma_u^u|_{q,0,T}.$$

The boundedness of this term was established in the proof of Lemma 2.34. We conclude that $|\tilde{\sigma}|_{q,0,T} < \infty$ and therefore $\|\tilde{\sigma}\|_{q,0,T} < \infty$. Now we are able to apply the inequality (2.5) to the term I_2, we obtain

$$I_2 \leq C\|\tilde{\sigma}\|_{q,s,t}|w|_{p,s,t}.$$

Hence,

$$I_2 \leq C(\|\tilde{\sigma}\|_{\infty,s,t} + |\tilde{\sigma}|_{q,s,t})|w|_{p,s,t}.$$

The calculation of $\|\tilde{\sigma}\|_{\infty,s,t}$ can be carried out analogously to the the estimation of I_1. Since the necessary properties of b and σ are the same, we only have to omit the $(t-s)$ term in estimate of I_1, which yields

$$\|\tilde{\sigma}\|_{\infty,s,t} \leq C(\|\eta^v\|_{\infty,s,t} + |\bar{u}|). \tag{2.35}$$

For the estimation of $|\tilde{\sigma}|_{q,s,t}$, we take a look at the difference

$$\begin{aligned}
&|\tilde{\sigma}(t_{i+1}) - \tilde{\sigma}(t_i)| \\
&= \frac{1}{|\bar{u}|}\left|\sigma^v(t_{i+1}) - \sigma^u(t_{i+1}) - \sigma_x^u(t_{i+1})y_{t_{i+1}}\bar{u} - \sigma_u^u(t_{i+1})\bar{u} - \sigma^v(t_i) + \sigma^u(t_i) + \sigma_x^u(t_i)y_{t_i}\bar{u} + \sigma_u^u(t_i)\bar{u}\right|.
\end{aligned}$$

Using the mean value theorem on $\sigma^v(t_{i+1}) - \sigma^u(t_{i+1})$, respectively $\sigma^v(t_i) - \sigma^u(t_i)$ and similar calculations as for I_1 lead to

$$\begin{aligned}
= \Bigg| \int_0^1 & \sigma_x(t_{i+1}, x_{t_{i+1}}^u + \lambda\gamma_{t_{i+1}}^v, u + \lambda\bar{u})\eta_{t_{i+1}}^v + (\sigma_x(t_{i+1}, x_{t_{i+1}}^u + \lambda\gamma_{t_{i+1}}^v, u + \lambda\bar{u}) - \sigma_x^u(t_{i+1}))\frac{y_{t_{i+1}}\bar{u}}{|\bar{u}|} \\
& - \sigma_x(t_i, x_{t_i}^u + \lambda\gamma_{t_i}^v, u + \lambda\bar{u})\eta_{t_i}^v - (\sigma_x(t_i, x_{t_i}^u + \lambda\gamma_{t_i}^v, u + \lambda\bar{u}) - \sigma_x^u(t_i))\frac{y_{t_i}\bar{u}}{|\bar{u}|} \\
& + \left[\sigma_u(t_{i+1}, x_{t_{i+1}}^u + \lambda\gamma_{t_{i+1}}^v, u + \lambda\bar{u}) - \sigma_u^u(t_{i+1}) - \sigma_u(t_i, x_{t_i}^u + \lambda\gamma_{t_i}^v, u + \lambda\bar{u}) + \sigma_u^u(t_i)\right]\frac{\bar{u}}{|\bar{u}|}\,d\lambda\Bigg|.
\end{aligned}$$

By adding the terms $\sigma_x(t_i, x_{t_i}^u + \lambda\gamma_{t_i}^v, u + \lambda\bar{u})\eta_{t_{i+1}}^v$ and $\sigma_x(t_i, x_{t_i}^u + \lambda\gamma_{t_i}^v, u + \lambda\bar{u}) - \sigma_x^u(t_i))\frac{y_{t_{i+1}}\bar{u}}{|\bar{u}|}$, we can write

$$|\tilde{\sigma}(t_{i+1}) - \tilde{\sigma}(t_i)| \leq I_{21} + I_{22} + I_{23} + I_{24} + I_{25},$$

where

$$
\begin{aligned}
I_{21} &= \left| \int_0^1 (\sigma_x(t_{i+1}, x^u_{t_{i+1}} + \lambda\gamma^v_{t_{i+1}}, u + \lambda\bar{u}) - \sigma_x(t_i, x^u_{t_i} + \lambda\gamma^v_{t_i}, u + \lambda\bar{u}))\eta^v_{t_{i+1}} \, d\lambda \right|, \\
I_{22} &= \left| \int_0^1 \sigma_x(t_i, x^u_{t_i} + \lambda\gamma^v_{t_i}, u + \lambda\bar{u})(\eta^v_{t_{i+1}} - \eta^v_{t_i}) \, d\lambda \right|, \\
I_{23} &= \bigg| \int_0^1 (\sigma_x(t_{i+1}, x^u_{t_{i+1}} + \lambda\gamma^v_{t_{i+1}}, u + \lambda\bar{u}) - \sigma^u_x(t_{i+1}) \\
&\qquad - \sigma_x(t_i, x^u_{t_i} + \lambda\gamma^v_{t_i}, u + \lambda\bar{u}) + \sigma^u_x(t_i))y_{t_{i+1}} \, d\lambda \bigg|, \\
I_{24} &= \left| \int_0^1 (\sigma_x(t_i, x^u_{t_i} + \lambda\gamma^v_{t_i}, u + \lambda\bar{u}) - \sigma^u_x(t_i))(y_{t_{i+1}} - y_{t_i}) \, d\lambda \right|, \\
I_{25} &= \left| \int_0^1 \sigma_u(t_{i+1}, x^u_{t_{i+1}} + \lambda\gamma^v_{t_{i+1}}, u + \lambda\bar{u}) - \sigma^u_u(t_{i+1}) - \sigma_u(t_i, x^u_{t_i} + \lambda\gamma^v_{t_i}, u + \lambda\bar{u}) + \sigma^u_u(t_i) \, d\lambda \right|.
\end{aligned}
$$

With the condition (H_3) in mind, the integrals I_{21}, I_{22} and I_{24} can easily be handled

$$
I_{21} \le C\|\eta^v\|_{\infty,s,t}(|t_{i+1} - t_i|^\beta + |x^v_{t_{i+1}} - x^v_{t_i}| + |x^u_{t_{i+1}} - x^u_{t_i}|), \tag{2.36}
$$

$$
I_{22} \le C|\eta^v_{t_{i+1}} - \eta^v_{t_i}|, \tag{2.37}
$$

$$
I_{24} \le C|y_{t_{i+1}} - y_{t_i}|(\|\gamma^v\|_{\infty,s,t} + |\bar{u}|). \tag{2.38}
$$

Since the conditions on σ_x and σ_u are identical, we carry out the estimation of I_{23} and adapt the results to the Integral I_{25}. We have

$$
\begin{aligned}
I_{23} = \bigg| \int_0^1 &(\sigma_x(t_{i+1}, x^u_{t_{i+1}} + \lambda\gamma^v_{t_{i+1}}, u + \lambda\bar{u}) - \sigma^u_x(t_{i+1}) \\
&- \sigma_x(t_i, x^u_{t_i} + \lambda\gamma^v_{t_i}, u + \lambda\bar{u}) + \sigma^u_x(t_i))y_{t_{i+1}} \, d\lambda \bigg|,
\end{aligned}
$$

which yields

$$
\begin{aligned}
&\le \|y\|_{\infty,s,t} \bigg| \int_0^1 (\sigma_x(t_{i+1}, x^u_{t_{i+1}} + \lambda\gamma^v_{t_{i+1}}, u + \lambda\bar{u}) - \sigma^u_x(t_{i+1}) \\
&\quad - \sigma_x(t_i, x^u_{t_i} + \lambda\gamma^v_{t_i}, u + \lambda\bar{u}) + \sigma^u_x(t_i)) \, d\lambda \bigg| \\
&\le \|y\|_{\infty,s,t} \sum_{l=1}^n \int_0^1 \bigg| \sigma_{x_l}(t_{i+1}, x^u_{t_{i+1}} + \lambda\gamma^v_{t_{i+1}}, u + \lambda\bar{u}) - \sigma_{x_l}(t_{i+1}, x^u_{t_{i+1}}, u) \\
&\quad - \sigma_{x_l}(t_i, x^u_{t_i} + \lambda\gamma^v_{t_i}, u + \lambda\bar{u}) + \sigma_{x_l}(t_i, x^u_{t_i}, u) \bigg| \, d\lambda,
\end{aligned}
$$

where σ_{x_l} is the l-th column vector of the matrix σ_x. Now we apply Lemma 2.26 ii) on σ_{x_l} for every $l = 1, \dots, n$ and obtain

$$\begin{aligned} I_{23} &\le C\|y\|_{\infty,s,t}\Big[\Big(|t_{i+1}-t_i|^\beta + |x^v_{t_{i+1}} - x^v_{t_i}| + |x^u_{t_{i+1}} - x^u_{t_i}|\Big) \|\gamma^v\|_{\infty,s,t} \\ &\quad + |\gamma^v_{t_{i+1}} - \gamma^v_{t_i}| + \Big(|t_{i+1}-t_i|^\beta + |x^v_{t_{i+1}} - x^v_{t_i}| + |x^u_{t_{i+1}} - x^u_{t_i}|\Big) |\bar{u}|\Big] \\ &\le C\|y\|_{\infty,s,t}\Big[\Big(|t_{i+1}-t_i|^\beta + |x^v_{t_{i+1}} - x^v_{t_i}| + |x^u_{t_{i+1}} - x^u_{t_i}|\Big) (\|\gamma^v\|_{\infty,s,t} + |\bar{u}|) \\ &\quad + |\gamma^v_{t_{i+1}} - \gamma^v_{t_i}|\Big]. \end{aligned} \tag{2.39}$$

As mentioned before the estimation of I_{25} is completely analogous and yields

$$I_{25} \le C\Big[\Big(|t_{i+1}-t_i|^\beta + |x^v_{t_{i+1}} - x^v_{t_i}| + |x^u_{t_{i+1}} - x^u_{t_i}|\Big) (\|\gamma^v\|_{\infty,s,t} + |\bar{u}|) + |\gamma^v_{t_{i+1}} - \gamma^v_{t_i}|\Big]. \tag{2.40}$$

Collecting all the terms (2.36), (2.37), (2.38), (2.39) and (2.40), we are finally able to estimate the difference $|\tilde{\sigma}(t_{i+1}) - \tilde{\sigma}(t_i)|$ and in conclusion the q-variation of $\tilde{\sigma}$. We have

$$\begin{aligned} &|\tilde{\sigma}(t_{i+1}) - \tilde{\sigma}(t_i)| \\ &\le C\Big[\|\eta^v\|_{\infty,s,t}\Big(|t_{i+1}-t_i|^\beta + |x^v_{t_{i+1}} - x^v_{t_i}| + |x^u_{t_{i+1}} - x^u_{t_i}|\Big) + |\eta^v_{t_{i+1}} - \eta^v_{t_i}| + |y_{t_{i+1}} - y_{t_i}|(\|\gamma^v\|_{\infty,s,t} + |\bar{u}|) \\ &\quad + (\|y\|_{\infty,s,t} + 1)\Big(\Big(|t_{i+1}-t_i|^\beta + |x^v_{t_{i+1}} - x^v_{t_i}| + |x^u_{t_{i+1}} - x^u_{t_i}|\Big)(\|\gamma^v\|_{\infty,s,t} + |\bar{u}|) + |\gamma^v_{t_{i+1}} - \gamma^v_{t_i}|\Big)\Big], \end{aligned}$$

which yields (since $q > 2$, hence $q > p$ and $q\beta > 1$),

$$\begin{aligned} |\tilde{\sigma}|_{q,s,t} &= \sup_{k\in\mathbb{N},\Pi_k\in\mathcal{P}([s,t])} \left(\sum_{i=1}^{n-1} |\tilde{\sigma}(t_{i+1}) - \tilde{\sigma}(t_i)|^q\right)^{\frac{1}{q}} \\ &\le C\Big[\|\eta^v\|_{\infty,s,t}\Big(|t-s|^\beta + |x^v|_{p,s,t} + |x^u|_{p,s,t}\Big) + |\eta^v|_{p,s,t} + |y|_{p,s,t}(\|\gamma^v\|_{\infty,s,t} + |\bar{u}|) \\ &\quad + (\|y\|_{\infty,s,t} + 1)\Big(\Big(|t-s|^\beta + |x^v|_{p,s,t} + |x^u|_{p,s,t}\Big)(\|\gamma^v\|_{\infty,s,t} + |\bar{u}|) + |\gamma^v|_{p,s,t}\Big)\Big]. \end{aligned}$$

By Lemma 2.27, we know that $|x^v|_{p,s,t}$ and $|x^u|_{p,s,t}$ are bounded by a constant that is independent of the parameter. From Lemma 2.33 it is clear that $\|y\|_{\infty,s,t}$, $|y|_{p,s,t}$ are bounded by a constant and from Lemma 2.29 that $\|\gamma^v\|_{\infty,s,t}$ and $|\gamma^v|_{p,s,t}$ are bounded by $C|\bar{u}|$. We get

$$|\tilde{\sigma}|_{p,s,t} \le C(\|\eta^v\|_{\infty,s,t} + |\eta^v|_{p,s,t} + |\bar{u}|) \le C(|\eta^v_s| + |\eta^v|_{p,s,t} + |\bar{u}|). \tag{2.41}$$

Combining the results (2.35) and (2.41), we are able to estimate the integral I_2, we have

$$I_2 \le C(|\eta^v_s| + |\eta^v|_{p,s,t} + |\bar{u}|)|w|_{p,s,t},$$

where C is independent of $\bar{u}$ and u. Adding the estimate of I_1, we obtain

$$
\begin{aligned}
|\eta^v|_{p,s,t} \le I_1 + I_2 &\le C[(\|\eta^v\|_{\infty,s,t} + |\bar{u}|)(t-s) + (|\eta^v_s| + |\eta^v|_{p,s,t} + |\bar{u}|)|w|_{p,s,t}] \\
&\le C(|\eta^v_s| + |\eta^v|_{p,s,t} + |\bar{u}|)(|t-s| + |w|_{p,s,t}).
\end{aligned}
$$

For $[s,t] \subset [0,T]$ with

$$
|t-s| + |w|_{p,s,t} \le \frac{1}{2C}
$$

this yields

$$
|\eta^v|_{p,s,t} \le |\bar{u}| + |\eta^v_s|.
$$

Again we can use the Gronwall-type lemma 2.20 and get

$$
|\eta^v|_{p,0,T} \le (|\bar{u}| + |\eta^v_0|)\, e^{C(T^p + |w|^p_{p,0,T})}.
$$

Notice that η^v_0 is given by

$$
\eta^v_0 = \frac{x_0(v) - x_0(u) - Dx_0(u)\bar{u}}{|\bar{u}|}.
$$

Since x_0 is continuously differentiable, it is totally differentiable, which yields

$$
|\eta^v_0| \to 0 \text{ for } |\bar{u}| \to 0
$$

and therefore

$$
|\eta^v|_{p,0,T} \to 0, \text{ for } |\bar{u}| \to 0.
$$

We conclude

$$
\|\eta^v\|_{\infty,0,T} \le |\eta^v_0| + |\eta^v|_{p,0,T} \to 0, \text{ for } |\bar{u}| \to 0.
$$

□

The next theorem is the main result of this section and summarizes the previous results.

Theorem 2.35. *For a open bounded and convex subset $\mathcal{U} \subset \mathbb{R}^d$ the solution mapping $x_\cdot : \mathcal{U} \to C^p([0,T],\mathbb{R}^n)$ for equation (2.22) is Fréchet differentiable with differential*

$$
Dx^u_t = \phi_t Dx_0(u) + \phi_t \int_0^t \phi_r^{-1} b^u_u(r)\, dr + \sum_{j=1}^m \phi_t \int_0^t \phi_r^{-1} \sigma^{u,j}_u(r)\, dw^j_r. \tag{2.42}
$$

which is the explicit solution to the matrix valued linear equation (2.28) on $[0,T]$.

2.1.4 Stochastic setting

In this subsection we want to translate the previous results on ordinary Young differential equations into the stochastic setting described at the beginning of this chapter. For an introduction on the basic elements of stochastic processes, we refer the reader to standard textbooks like Kallenberg [2002]. For the rest of this thesis, we call a process $(X_t)_{t\in[0,T]}$ continuous and/or of bounded p-variation, if for almost every $\omega \in \Omega$

$$t \mapsto X_t(\omega) \text{ is continuous and/or } \|X(\omega)\|_{p,0,T} < \infty.$$

The first problem when dealing with stochastic process, which are of bounded p-variation is the measurability of the p-variation norm on an interval $[s,t]$ of a process. This problem arises, since we are taking the supremum over the uncountable set of all partitions of the interval $[s,t]$. Now let $X_{t\in[0,T]}$ be a continuous stochastic process of bounded p-variation defined on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$, adapted to $\mathbb{F}$. It is an easy implication of Blumenthal and Getoor [1960] (Lemma 2.2) that for a continuous process of bounded p-variation on $[s,t] \subset \mathbb{R}$

$$\|X(\omega)\|_{p,s,t} = \sup_{k\in\mathbb{N},\Pi_k\in\mathcal{P}_{\mathbb{Q}}([s,t])} \left(\sum_{i=0}^{k-1} |X_{t_{i+1}}(\omega) - X_{t_i}(\omega)|^p \right)^{\frac{1}{p}} \quad P-a.s.,$$

where $P_{\mathbb{Q}}([s,t])$ is the set of finite partitions $s \leq t_0 < \ldots, t_k \leq t$ of the interval $[s,t]$ such that $t_0, \ldots, t_k \in \mathbb{Q}$. Hence, for almost every $\omega \in \Omega$, $\|X(\omega)\|_{p,s,t}$ is the supremum of countable many $\mathcal{F}_t$-measurable random variables, which implies that it is a $\mathcal{F}_t$ measurable random variable itself. We will summarize the results of this section to establish all properties of the solution process to equation (2.1), which are crucial in the following sections. The following corollary is a direct implication of the results in Subsection 2.1.3.

Corollary 2.36. *Suppose we are in the setting introduced at the beginning of this chapter. If the coefficient functions b and σ and the initial value function $x_0 : \mathcal{U} \to \mathbb{R}^{n_1}$ satisfy the conditions (H_1), (H_2) and (H_3), then for every $u \in \mathcal{U}$ equation (2.1), given by*

$$\xi_t^u = \xi_0(u) + \int_0^t b(r, \xi_r^u, u)\, dr + \sum_{j=1}^{m_1} \int_0^t \sigma^j(r, \xi_r^u, u)\, dw_r^j,$$

the inhomogenous linear equation

$$y_t^u = D\xi_0(u) + \int_0^t b_x(r, \xi_r^u, u) y_r^u + b_u(r, \xi_r^u, u)\, dr + \sum_{j=1}^{m_1} \int_0^t \sigma_x^j(r, \xi_r^u, u) y_r^u + \sigma_u^j(r, \xi_r^u, u)\, dw_r^j$$

and the matrix valued homogenous linear equations for an initial time $s_0 \in [0,T]$

$$\phi_t^{s_0} = I_{n_1} + \int_{s_0}^t b_x(r, \xi_r^u, u) \phi_r^{s_0}\, dr + \sum_{j=1}^{m_1} \int_{s_0}^t \sigma_x^j(r, \xi_r^u, u) \phi_r^{s_0}\, dw_r^j$$

$$\psi_t^{s_0} = I_{n_1} - \int_{s_0}^t \psi_r^{s_0} b_x(r, \xi_r^u, u)\, dr - \sum_{j=1}^{m_1} \int_{s_0}^t \psi_r^{s_0} \sigma_x^j(r, \xi_r^u, u)\, dw_r^j$$

have unique pathwise solutions, such that $\xi^u_\cdot(\omega) \in C^p([0,T],\mathbb{R}^{n_1})$, $y^u_\cdot(\omega) \in C^p([0,T],\mathbb{R}^{n_1\times d})$, $\phi^{s_0}_\cdot(\omega) \in C^p([s_0,T],\mathbb{R}^{n_1\times n_1})$ *and* $\psi^{s_0}_\cdot(\omega) \in C^p([s_0,T],\mathbb{R}^{n_1\times n_1})$ *for almost every* $\omega \in \Omega$. *Furthermore, we get for almost every* $\omega \in \Omega$ *the bounds*

$$\begin{aligned}
\|\xi^u(\omega)\|_{\infty,0,T} &\leq \|\xi^u(\omega)\|_{p,0,T} \\
&\leq L + 2^{2p-1}C_1^p(T^p + |w(\omega)|^p_{p,0,T}). =: C_\xi(\omega) && (2.43) \\
\|\phi^u(\omega)\|_{\infty,s_0,T} &\leq \|\phi^u(\omega)\|_{p,s_0,T} \\
&\leq 2\sqrt{n}e^{2^{4p}(C_1m_1)^p\left(T^p+|w(\omega)|^p_{p,0,T}\right)} =: C_\phi(\omega) && (2.44) \\
\|y^u(\omega)\|_{\infty,0,T} &\leq \|y^u(\omega)\|_{p,0,T} \\
&\leq (1+2L)e^{2^{4p}(C_1m_1)^p\left(T^p+|w(\omega)|^p_{p,0,T}\right)} := C_y(\omega), && (2.45)
\end{aligned}$$

where the bound for $\|\psi^{s_0}(\omega)\|_{\infty,s_0,T}$ *and* $\|\psi^{s_0}(\omega)\|_{p,s_0,T}$ *is also the right hand side of* (2.44). *Here the constants* C_1 *is defined in* (2.23). *Let* $u \in \mathcal{U}$ *and* $\bar{u} \in \mathbb{R}^d$, *such that* $u + \bar{u} \in \mathcal{U}$, *we have for almost every* $\omega \in \Omega$

$$\begin{aligned}
\|\xi^{u+\bar{u}}(\omega) - \xi^u(\omega)\|_{\infty,0,T} &\leq \|\xi^{u+\bar{u}}(\omega) - \xi^u(\omega)\|_{p,0,T} \\
&\leq (1+2L)|\bar{u}|e^{2^{3p}(3C_1m_1)^p(T^p+|w(\omega)|^p_{p,0,T})}. && (2.46)
\end{aligned}$$

Moreover, the limits

$$\lim_{|\bar{u}|\to 0} \|\xi^{u+\bar{u}}(\omega) - \xi^u(\omega)\|_{\infty,0,T} \leq \lim_{|\bar{u}|\to 0} \|\xi^{u+\bar{u}}(\omega) - \xi^u(\omega)\|_{p,0,T} = 0 \tag{2.47}$$

$$\lim_{|\bar{u}|\to 0} \left\| \frac{\xi^{u+\bar{u}}(\omega) - \xi^u(\omega) - y^u(\omega)\bar{u}}{|\bar{u}|} \right\|_{\infty,0,T} \leq \lim_{|\bar{u}|\to 0} \left\| \frac{\xi^{u+\bar{u}}(\omega) - \xi^u(\omega) - y^u(\omega)\bar{u}}{|\bar{u}|} \right\|_{p,0,T} = 0$$

hold for almost every $\omega \in \Omega$ *and all* $u \in \mathcal{U}$.

Proof. Let $N \in \mathcal{F}$ be a P null set, such that $w_\cdot(\omega)$ is a continuous path of finite p-variation for all $\omega \in N^c$. Then we have for all $\omega \in N^c$ and $u \in \mathcal{U}$ that there exists a unique solution $\xi^u_\cdot(\omega) \in C^p([0,T],\mathbb{R}^{n_1})$ to the YDE

$$\xi^u_t(\omega) = \xi_0(u) + \int_0^t b(r,\xi^u_r(\omega),u)\,dr + \sum_{j=1}^{m_1}\int_0^t \sigma^j(r,\xi^u_r(\omega),u)\,dw^j_r(\omega),$$

by Corollary 2.24. Analogously, we obtain the solutions $y^u_\cdot(\omega) \in C^p([0,T],\mathbb{R}^{n_1\times d})$, $\phi^u_\cdot(\omega) \in C^p([s_0,T],\mathbb{R}^{n_1\times n_1})$ and $\psi^u_\cdot(\omega) \in C^p([s_0,T],\mathbb{R}^{n_1\times n_1})$ for every $\omega \in N^c$ by Lemma 2.31 and Lemma 2.33. The bounds (2.43), (2.44) and (2.45) are direct consequences of Lemma 2.27, 2.31 and 2.34. Estimate (2.46) follows from Lemma 2.29. The limits in the statement of the corollary hold for all $\omega \in N^c$ by Lemma 2.29 and Lemma 2.30, applied on the paths $\xi^u_\cdot(\omega)$ and $y^u_\cdot(\omega)$. □

Having established the pathwise bound and limits of the process $\xi^u_\cdot$ we are now interested if these bounds and limits also hold in L^l-sense for some $l \geq 1$, if we impose a suitable integrability condition on the driving process w. We give a positive answer in the following corollary.

Corollary 2.37. *In the situation of Corollary 2.36, if the process w satisfies the exponential moment condition*

$$E\left[e^{K|w|^2_{p,0,T}}\right] < \infty \tag{2.48}$$

for some constant $K > 0$, then for every $u \in \mathcal{U}$ and $l \geq 1$, we have

$$\begin{aligned}
\mathrm{E}\left[\|\xi^u\|^l_{\infty,0,T}\right] &\leq \mathrm{E}\left[\|\xi^u\|^l_{p,0,T}\right] \leq D_{\xi,l} \\
\mathrm{E}\left[\|\phi^{s_0}\|^l_{\infty,s_0,T}\right] &\leq \mathrm{E}\left[\|\phi^{s_0}\|^l_{p,s_0,T}\right] \leq D_{\phi,l} \\
\mathrm{E}\left[\|\psi^{s_0}\|^l_{\infty,s_0,T}\right] &\leq \mathrm{E}\left[\|\psi^{s_0}\|^l_{p,s_0,T}\right] \leq D_{\psi,l} \\
\mathrm{E}\left[\|y^u\|^l_{\infty,0,T}\right] &\leq \mathrm{E}\left[\|y^u\|^l_{p,0,T}\right] \leq D_{y,l}
\end{aligned} \tag{2.49}$$

for constants $D_{\xi,l}$, $D_{\phi,l}$, $D_{\psi,l}$, $D_{y,l}$ that are independent of u, where $D_{\phi,l}$, $D_{\psi,l}$ are also independent of s_0. Let $u \in \mathcal{U}$ and $\bar{u} \in \mathbb{R}^d$, such that $u + \bar{u} \in \mathcal{U}$, we have

$$\begin{aligned}
\mathrm{E}\left[\|\xi^{u+\bar{u}}(\omega) - \xi^u(\omega)\|^l_{\infty,0,T}\right] &\leq \mathrm{E}\left[\|\xi^{u+\bar{u}}(\omega) - \xi^u(\omega)\|_{p,0,T}\right] \\
&\leq C|\bar{u}|^l.
\end{aligned} \tag{2.50}$$

for a constant $C > 0$ which is independent of u. Moreover, the limits

$$\begin{aligned}
\lim_{|\bar{u}|\to 0} \mathrm{E}\left[\|\xi^{u+\bar{u}} - \xi^u\|^l_{\infty,0,T}\right] &\leq \lim_{|\bar{u}|\to 0} \mathrm{E}\left[\|\xi^{u+\bar{u}} - \xi^u\|^l_{p,0,T}\right] = 0 \\
\lim_{|\bar{u}|\to 0} \mathrm{E}\left[\left\|\frac{\xi^{u+\bar{u}} - \xi^u - y^u\bar{u}}{|\bar{u}|}\right\|^l_{\infty,0,T}\right] &\leq \lim_{|\bar{u}|\to 0} \mathrm{E}\left[\left\|\frac{\xi^{u+\bar{u}} - \xi^u - y^u\bar{u}}{|\bar{u}|}\right\|^l_{p,0,T}\right] = 0.
\end{aligned} \tag{2.51}$$

hold for all $u \in \mathcal{U}$ and $l \geq 1$.

Proof. The L^l-bounds of the processes are a direct consequence of the pathwise bounds in Corollary 2.36 and the exponential moment condition (2.48). Similarily (2.50) holds. Since the limit (2.47) holds almost surely according to Corollary 2.36, the limit holds in L^l-sense by the dominated convergence theorem, because of the pathwise bound (2.46) and the exponential moment condition. Again referring to Corollary 2.36, we know that the limit

$$\lim_{|\bar{u}|\to 0} \left\|\frac{\xi^{u+\bar{u}}(\omega) - \xi^u(\omega) - y^u(\omega)\bar{u}}{|\bar{u}|}\right\|_{\infty,0,T} \leq \lim_{|\bar{u}|\to 0} \left\|\frac{\xi^{u+\bar{u}}(\omega) - \xi^u(\omega) - y^u(\omega)\bar{u}}{|\bar{u}|}\right\|^l_{p,0,T} = 0$$

holds for almost every $\omega \in \Omega$. Using the pathwise bounds (2.45), (2.46), the exponential moment condition and the dominated convergence theorem show, that the limit (2.51) holds. □

To use the results of Corollary 2.37 in practice, we need to identify processes which satisfy this exponential moment condition. There is a important result in Jain and Monrad [1983] concerning Gaussian processes.

Theorem 2.38. *Let $(x_t)_{t\in[0,T]}$ be a separable Gaussian process of bounded p-variation for $p \in (1,2)$. Then there exists $C > 0$ such that*

$$E\left[e^{C\|x\|^2_{p,0,T}}\right] < \infty.$$

Proof. See Theorem 2.3 in Jain and Monrad [1983]. □

One prominent example of such a Gaussian process is the fractional Brownian motion with Hurst index $\frac{1}{2} < H < 1$, which paths are almost surely continuous and of bounded p-variation for $p > 1/H$. This process will be the main subject of our numerical experiments. By the last corollary, we know that the solution mapping $u \mapsto \xi^u_\cdot$ from $\mathcal{U}$ to $L^l_{\mathbb{F}}(\Omega, C([0,T], \mathbb{R}^{n_1})$, for an arbitrary $l \geq 1$ is Fréchet differentiable. That is the key property we need in the remainder of this thesis.

2.2 Parameter dependent SDEs with Brownian driver

Having established the important results concerning equation (2.1), we now focus on equation 2.2. But first we repeat the setting and add the needed assumptions on the coefficients and on the driving process w, which we found in the last section. As in the previous section we start by stating the setting and goal for this section, including the results we already obtained. Let T be a positive constant and $n_1, m_1, n_2, m_2, d \in \mathbb{N} = \{1, 2, \dots\}$. Let $(\Omega, \mathcal{F}, \mathbb{F}, P)$ be a filtered probability space, carrying an m_1-dimensional process w and a m_2-dimensional Brownian motion, both adapted to $\mathbb{F}$. We assume w to be a continuous, bounded p-variation process for $p \in (1,2)$ and w satisfies the exponential moment condition (2.48). Let $\mathcal{U}$ be a bounded, convex and open subset of $\mathbb{R}^d$ and $T > 0$ be a positive constant. We are interested in the system of parameter dependent stochastic differential equations given by

$$\xi^u_t = \xi_0(u) + \int_0^t b(r, \xi^u_r, u)\, dr + \int_0^t \sigma(r, \xi^u_r, u)\, dw_r \tag{2.52}$$

$$x^u_t = x_0(u) + \int_0^t \hat{b}(r, x^u_r, \xi^u_r, u)\, dr + \int_0^t \hat{\sigma}(r, x^u_r, \xi^u_r, u)\, dBr \tag{2.53}$$

for $t \in [0,T]$, where we assume

i) $\xi_0 : \mathcal{U} \to \mathbb{R}^{n_1}$, $b : [0,T] \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_1}$ and $\sigma : [0,T] \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_1 \times m_1}$ satisfy conditions $(H_1) - (H_3)$ and therefore, for every $u \in \mathcal{U}$ the process $\xi^u \in L^l_{\mathbb{F}}(\Omega, C([0,T]), \mathbb{R}^{n_1})$ is the unique solution to the given equation (2.52) for $l \geq 1$ by Corollary 2.36 and Corollary 2.37. Furthermore we know that the solution mapping $u \mapsto \xi^u$ from $\mathcal{U}$ to $L^l_{\mathbb{F}}(\Omega, C([0,T]), \mathbb{R}^{n_1})$ is Fréchet differentiable under these conditions, for every $l \geq 1$.

ii) $x_0 : \mathcal{U} \to \mathbb{R}^{n_2}$, $\hat{b} : [0,T] \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_2}$ and $\hat{\sigma} : [0,T] \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_2 \times m_2}$ are deterministic functions.

The goal is now to establish conditions on $\hat{b}$ and $\hat{\sigma}$ such that the second equation in our system admits a unique solution $x^u \in L^l_{\mathbb{F}}(\Omega, C([0,T]), \mathbb{R}^{n_2})$ for every $u \in \mathcal{U}$, $l \geq 1$ and the solution mapping $\mathcal{U} \to L^l_{\mathbb{F}}(\Omega, C([0,T]), \mathbb{R}^{n_2}), u \mapsto x^u$ is also Fréchet differentiable. Let the aforementioned assumptions hold for the rest of this section.

2.2.1 Existence and uniqueness

In this section we are going to examine the existence and uniqueness of the parameter dependent SDE (2.53). The solution to stochastic differential equations with a Brownian driver have extensively been studied and many results have been established using the rich theory of Itô calculus. Since most of these results are well known we omit here an introductory chapter on the subject and refer the reader to the corresponding literature. To name some literature concerning stochastic differential equation we suggset Øksendal [2014], Karatzas and Shreve [1991],Protter [2005], Ikeda and Watanabe [2014] and for controlled SDEs Pham [2009], Yong and Zhou [1999] and Yong [2019]. We will establish the most important theorems and proofs using the notation and ideas from Yong and Zhou [1999] and Yong [2019].

For any $u \in \mathcal{U}$, equation (2.53) is a time inhomogenous stochastic differential equation with random coefficients. We will first consider a more general equation.

$$\begin{aligned} x_t(\omega) &= x_0(\omega) + \int_{s_0}^t f(r, x_r(\omega), \omega)\, dr + \int_{s_0}^t g(r, x_r(\omega), \omega)\, dB_r(\omega) \\ &= x_0(\omega) + \int_{s_0}^t f(r, x_r(\omega), \omega)\, dr + \sum_{j=1}^{m_2} \int_{s_0}^t g^j(r, x_r(\omega), \omega)\, dB^j_r(\omega). \end{aligned} \tag{2.54}$$

In the following we will define the notion of a unique solution to equation (2.54) and give conditions under which such an solution will exists. Here we use the ideas of Yong [2019] on SDEs with random coefficients, explained in Chapter 1.3.

Definition 2.39. Let $T > 0$, $s_0 \in [0,T]$, $f : [s_0, T] \times \mathbb{R}^{n_2} \times \Omega \to \mathbb{R}^{n_2}$ and $g : [s_0, T] \times \mathbb{R}^{n_2} \times \Omega \to \mathbb{R}^{n_2 \times m_2}$ be given on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$ together with an $\mathbb{F}$-adapted m_2-dimensional Brownian motion and a random variable $x_0 : \Omega \to \mathbb{R}^{n_2}$ which is $\mathcal{F}_{s_0}$ measurable. An $\mathbb{F}$-adapted continuous process x_t, $t \in [s_0, T]$ is called a solution to (2.54), if

i) $\int_{s_0}^t |\hat{b}(r, x_r(\omega), \omega)| + |\hat{\sigma}(r, x_r(\omega), \omega)|^2\, dr < \infty$, $t \in [s_0, T]$, P-a.e. $\omega \in \Omega$.

ii) $x_t(\omega) = x_0(\omega) + \int_{s_0}^t \hat{b}(r, x_r(\omega), \omega)\, dr + \int_{s_0}^t \hat{\sigma}(r, x_r(\omega), \omega)\, dB_s$, $t \in [s_0, T]$, P-a.e. $\omega \in \Omega$.

Under the following conditions on the coefficient functions, we can state an uniqueness and existence result for the solution of the SDE (2.54).

(B) Let $T \in (0,\infty)$. The maps $f : [0,T] \times \mathbb{R}^{n_2} \times \Omega \to \mathbb{R}^{n_2}$ and $g^j : [0,T] \times \mathbb{R}^{n_2} \times \Omega \to \mathbb{R}^{n_2}$ be jointly measurable functions for every $j = 1,\dots,m_2$, where we equip $[0,T]$ and $\mathbb{R}^{n_2}$ with the corresponding Borel sigma algebras. For any $x \in \mathbb{R}^{n_2}$ the processes $f(\cdot,x,\cdot)$ and $g^j(\cdot,x,\cdot)$ are $\mathbb{F}$-progressively measurable. Furthermore there exists a constant $L > 0$, such that for all $t \in [0,T],\ x,y \in \mathbb{R}^{n_2},\ j = 1,\dots,m_2$, for some $l \geq 1$

$$\begin{cases} |f(t,x,\omega) - f(t,y,\omega)| \leq L|x-y| \ P-a.s. \\ |g^j(t,x,\omega) - g^j(t,y,\omega)| \leq L|x-y| \ P-a.s. \\ \mathrm{E}\left[\left(\int_{s_0}^T |f(r,0)|\,dr\right)^l + \sum_{j=1}^{m_2}\left(\int_{s_0}^T |g^j(r,0)|^2\,dr\right)^{\frac{l}{2}}\right] < \infty \end{cases}$$

Theorem 2.40. *Let (B) hold. Then for any $\mathcal{F}_{s_0}$ measurable random variable $x_0 \in L^l(\Omega,\mathbb{R}^{n_2})$, equation (2.54) admits a unique solution x such that $x \in L^l_{\mathbb{F}}(\Omega, C([s_0,T]),\mathbb{R}^{n_2})$ and we have*

$$\mathrm{E}[\|x\|^l_{\infty,s_0,T}] \leq C\left(\mathrm{E}\left[|x_0|^l\right] + \mathrm{E}\left[\int_{s_0}^T |f(r,0)|^l\,dr\right] + \sum_{j=1}^{m_2}\mathrm{E}\left[\int_{s_0}^T |g^j(r,0)|^2\,dr^{\frac{l}{2}}\right]\right). \qquad (2.55)$$

The constant C only depends on T, l and the constant L from condition (B).

Proof. See Theorem 1.25 in Yong [2019]. □

By Theorem 2.40, we know that there exists a solution to the SDE (2.54), which is an element of $L^l_{\mathbb{F}}(\Omega, C([s_0,T]),\mathbb{R}^{n_2})$, if condition (B) is satisfied for a given $l \geq 1$ and the initial value is $\mathcal{F}_{s_0}$ measurable and an element of $L^l(\Omega,\mathbb{R}^{n_2})$. If condition (B) is satisfied for all $l \geq 1$ and x_0 has moments of all orders, we get a solution which also has moments of all orders. Now we come back to our equation of interest (2.53). We will use the previous result to show that for every $u \in \mathcal{U}$ there exists a unique solution x^u to the equation (2.53), but since we need Fréchet differentiability of the map $u \mapsto x^u_\cdot$ from $\mathcal{U}$ to $L^l_{\mathbb{F}}(\Omega, C[0,T],\mathbb{R}^{n_2})$ for every $l \geq 1$, we need to assume further conditions on the coefficient functions $\hat{b}$ and $\hat{\sigma}$ and the initial value x_0.

(B_1) The function $\hat{b} : [0,T] \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_2}$ is continuous with respect to the variables t, x, z and u and continuously differentiable with respect to x, z and u for all $t \in [0,T]$. Denote

$$\hat{b}_x(t,x,z,u) = \left(\frac{\partial \hat{b}_i(t,x,z,u)}{\partial x_j}\right)_{1\leq i,j\leq n_2}, \qquad \hat{b}_z(t,x,z,u) = \left(\frac{\partial \hat{b}_i(t,x,z,u)}{\partial z_j}\right)_{1\leq i\leq n_2, 1\leq j\leq n_1}$$

$$\hat{b}_u(t,x,z,u) = \left(\frac{\partial \hat{b}_i(t,x,z,u)}{\partial u_j}\right)_{1\leq i\leq n_2, 1\leq j\leq d}.$$

Furthermore there exists a constant $L > 0$ such that

$$\sup_{t\in[0,T],x\in\mathbb{R}^{n_2},z\in\mathbb{R}^{n_1},u\in\mathcal{U}} |\hat{b}_x(t,x,z,u)| + |\hat{b}_z(t,x,z,u)| + |\hat{b}_u(t,x,z,u)| \leq L.$$

(B_2) The function $\hat{\sigma} = (\hat{\sigma}^1, \ldots, \hat{\sigma}^k) : [0,T] \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_2 \times m_2}$ is continuous with respect to the variables t, x, z and u and continuously differentiable with respect to x, z and u for all $t \in [0,T]$. Denote for $j = 1, \ldots, m_2$

$$\hat{\sigma}^j_x(t,x,z,u) = \left(\frac{\partial \hat{\sigma}^j_{i_1}(t,x,z,u)}{\partial x_{i_2}}\right)_{1\leq i_1,i_2\leq n_2}, \quad \hat{\sigma}^j_z(t,x,z,u) = \left(\frac{\partial \hat{\sigma}^j_{i_1}(t,x,z,u)}{\partial z_{i_2}}\right)_{1\leq i_1\leq n_2,\ 1\leq i_2\leq n_1}$$

$$\hat{\sigma}^j_u(t,x,z,u) = \left(\frac{\partial \hat{\sigma}^j_{i_1}(t,x,z,u)}{\partial u_{i_2}}\right)_{1\leq i_1\leq n_2, 1\leq i_2\leq d}.$$

Furthermore there exists a constant $L > 0$ such that for $j = 1, \ldots, m_2$

$$\sup_{t\in[0,T],x\in\mathbb{R}^{n_2},\mathbb{R}^{n_1},u\in\mathcal{U}} |\hat{\sigma}^j_x(t,x,z,u)| + |\hat{\sigma}^j_z(t,x,z,u)| + |\hat{\sigma}^j_u(t,x,z,u)| \leq L.$$

(B_3) Let $x_0 : \mathcal{U} \to \mathbb{R}^{n_2}$ be a continuously differentiable deterministic function, such that x_0 and its Jacobian Dx_0 are bounded by the constant L.

Now we first show that under the conditions (B_1), (B_2) and (B_3) equation (2.53) has a unique solution which is bounded intependently of the parameter u.

Lemma 2.41. *Let $T \in (0,\infty)$, $\hat{b}$, $\hat{\sigma}$ and $x_0(u)$ satisfy conditions (B_1), (B_2) and (B_3). Then for any $u \in \mathcal{U}$, equation (2.53) admits a unique solution x^u such that for all $l \geq 1$, $\mathrm{E}[\|x\|^l_{\infty,0,T}]$ is bounded by a constant $D_{x,l}$ which does not depend on u.*

Proof. Assume that $\hat{b}$, $\hat{\sigma}$ and x_0 satisfy conditions (B_1), (B_2) and (B_3). Fix $u \in \mathcal{U}$, then the coefficient functions

$$f : [0,T] \times \mathbb{R}^{n_2} \times \Omega \to \mathbb{R}^{n_2}, (t,x,\omega) \mapsto \hat{b}(t,x,\xi^u_t(\omega),u)$$

and

$$g^j : [0,T] \times \mathbb{R}^{n_2} \times \Omega \to \mathbb{R}^{n_2}, (t,x,\omega) \mapsto \hat{\sigma}^j(t,x,\xi^u_t(\omega),u)$$

satisfy condition (B) for all $l \geq 1$. The Lipschitz continuity in the x variable of f and g^j is a direct consequence of (B_1) and (B_2), where the Lipschitz constants do not depend on u. Since for every $u \in \mathcal{U}$ the process ξ^u_t is $\mathbb{F}$-adapted and has almost surely continuous paths, the process $\hat{b} : \Omega \times [0,T] \to \mathbb{R}^{n_2}, (\omega,t) \mapsto \hat{b}(t,0,\xi^u_t(\omega),u)$ is $\mathbb{F}$-progressively measurable and we have for all $t \in [0,T]$ and $u \in \mathcal{U}$

$$|\hat{b}(t,0,\xi^u_t,u)| \leq C(|\xi^u_t| + |u|) + \hat{b}(t,0,0,0) \text{ P-a.s..}$$

Since $\hat{b}$ is continuous in t on the compact interval $[0,T]$ and $\mathcal{U}$ is bounded, there exists a constant K_1 such that

$$E\left[\int_0^T |\hat{b}(r,0,\xi_r^u,u)|^l \, dr\right] \leq K_1 \mathrm{E}\left[\int_0^T |\xi_r^u|^l + 1 \, dr\right] \leq K_1 T(D_{\xi,l}+1) < K_2$$

by (2.49), where $K_2 > 0$ is a constant that is independent of u. Analogously, we can estimate

$$E\left[\int_0^T |\hat{\sigma}^j(r,0,\xi_r^u,u)|^2 \, dr^{\frac{l}{2}}\right] \leq K_3,$$

where $K_3 > 0$ is a constant that is independent of u. Since (B) is satisfied on $[0,T]$ and $x_0(u)$ is deterministic, Theorem 2.40 implies that there exists a unique solution to (2.53) and we have for $l \geq 1$ the estimate

$$\mathrm{E}[\|x^u\|_{\infty,0,T}^l] \leq C\left(|x_0(u)|^l + \mathrm{E}\left[\int_0^T |\hat{b}(r,0,\xi_r^u,u)|^l \, dr\right] + \sum_{j=1}^{m_2} \mathrm{E}\left[\left(\int_0^T |\hat{\sigma}^j(r,0,\xi_r^u,u)|^2 \, dr\right)^{\frac{l}{2}}\right]\right).$$

Taking the previous considerations and condition (B_3) into account, this yields

$$\mathrm{E}[\|x^u\|_{\infty,0,T}^l] \leq C\left(|x_0(u)|^l + K_2 + m_2 K_3\right) \leq C\left(L^l + K_2 + m_2 K_3\right) := D_{x,l},$$

where the constants C, K_2 and K_3 do not depend on the parameter u. $\square$

2.2.2 Fréchet differentiability of the solution mapping

We now repeat the same steps as in the previous section to show the Fréchet differentiability of the map $u \mapsto x_\cdot^u$ from $\mathcal{U}$ to $L^l_{\mathbb{F}}(\Omega, C[s_0,T], \mathbb{R}^{n_2})$.

Lemma 2.42. *Suppose we are in the situation of Lemma 2.41, we have for every $u \in \mathcal{U}$, that*

$$\lim_{|\bar{u}|\to 0} \mathrm{E}[\|x^{u+\bar{u}} - x^u\|_{\infty,0,T}^l] = 0$$

for an arbitrary $l \geq 1$.

Proof. We use C as a constant which can vary over the course of the proof. Let $u \in \mathcal{U}$ and set $v = u + \bar{u}$, $\gamma_\cdot^v = x_\cdot^v - x_\cdot^u$ and $\nu_\cdot^v = \xi_\cdot^v - \xi_\cdot^u$, where $v \in \mathcal{U}$ for a $\bar{u} \in \mathbb{R}^d$ close enough to u, since $\mathcal{U}$ is open and convex. Furthermore set for an arbitrary $u \in \mathcal{U}$, $\hat{b}^u(t) := \hat{b}(t, x_t^u, \xi_t^u, u)$ and $\hat{\sigma}^u(t) := \hat{\sigma}(t, x_t^u, \xi_t^u, u)$, where the j-th column vector of $\hat{\sigma}^u(t)$ is denoted by $\hat{\sigma}^{u,j}(r)$. We have

$$\gamma_t^v = \int_0^t \hat{b}^v(r) - \hat{b}^u(r) \, dr + \sum_{j=1}^{m_2} \int_0^t \hat{\sigma}^{v,j}(r) - \hat{\sigma}^{u,j}(r) \, dB_r^j.$$

With the mean value theorem (Lemma 2.25), we get

$$\gamma_t^v = x_0(v)-x_0(u)+\int_0^t \tilde{b}_x^u(r)\gamma_r^v+\tilde{b}_z^u(r)\nu_r^v+\tilde{b}_u^u(r)\bar{u}\,dr+\sum_{j=1}^{m_2}\int_0^t \tilde{\sigma}_x^{u,j}(r)\gamma_r^v+\tilde{\sigma}_z^{u,j}(r)\nu_r^v+\tilde{\sigma}_u^{u,j}(r)\bar{u}\,dB_r^j, \tag{2.56}$$

where

$$\begin{aligned}
\tilde{b}_x^u(r) &= \int_0^1 \hat{b}_x(r, x_r^u+\lambda\gamma_r^v, \xi_r^u+\lambda\nu_r^v, u+\lambda\bar{u})\,d\lambda\\
\tilde{b}_z^u(r) &= \int_0^1 \hat{b}_z(r, x_r^u+\lambda\gamma_r^v, \xi_r^u+\lambda\nu_r^v, u+\lambda\bar{u})\,d\lambda\\
\tilde{b}_u^u(r) &= \int_0^1 \hat{b}_u(r, x_r^u+\lambda\gamma_r^v, \xi_r^u+\lambda\nu_r^v, u+\lambda\bar{u})\,d\lambda\\
\tilde{\sigma}_x^{u,j}(r) &= \int_0^1 \hat{\sigma}_x^j(r, x_r^u+\lambda\gamma_r^v, \xi_r^u+\lambda\nu_r^v, u+\lambda\bar{u})\,d\lambda\\
\tilde{\sigma}_z^{u,j}(r) &= \int_0^1 \hat{\sigma}_z^j(r, x_r^u+\lambda\gamma_r^v, \xi_r^u+\lambda\nu_r^v, u+\lambda\bar{u})\,d\lambda\\
\tilde{\sigma}_x^{u,j}(r) &= \int_0^1 \hat{\sigma}_u^j(r, x_r^u+\lambda\gamma_r^v, \xi_r^u+\lambda\nu_r^v, u+\lambda\bar{u})\,d\lambda.
\end{aligned}$$

We know that $\hat{b}$ and $\hat{\sigma}$ satisfy the condition (B_1) and (B_2) and furthermore that $\xi^u, \xi^v \in L^l_{\mathbb{F}}(\Omega, C([0,T]), \mathbb{R}^{n_1})$ (by Corollary 2.37), $x^u, x^v \in L^l_{\mathbb{F}}(\Omega, C([0,T]), \mathbb{R}^{n_1})$ (by Lemma 2.41) and consequently $\gamma^v, \nu^v \in L^l_{\mathbb{F}}(\Omega, C([0,T]), \mathbb{R}^{n_1})$ for every $l \geq 1$. This implies that the coefficient functions of the SDE (2.56), defined by

$$f(r,x,\omega) := \tilde{b}_x^u(r)x+\tilde{b}_z^u(r)\nu_r^v+\tilde{b}_u^u(r)\bar{u}$$

and

$$g^j(r,x,\omega) := \tilde{\sigma}_x^{u,j}(r)x+\tilde{\sigma}_z^{u,j}(r)\nu_r^v+\tilde{\sigma}_u^{u,j}(r)\bar{u}$$

satisfy condition (B) on $[0,T]$, which yields using Theorem 2.40

$$\begin{aligned}
\mathrm{E}[\|\gamma^v\|^l_{\infty,0,T}] \leq C\Bigg(&\mathrm{E}\left[|x_0(v)-x_0(u)|^l\right]+\mathrm{E}\left[\int_0^T |\tilde{b}_z^u(r)\nu_r^v+\tilde{b}_u^u(r)\bar{u}|^l\,dr\right]\\
&+\sum_{j=1}^{m_2}\mathrm{E}\left[\int_0^T |\tilde{\sigma}_z^{u,j}(r)\nu_r^v+\tilde{\sigma}_u^{u,j}(r)\bar{u}|^2\,dr^{\frac{l}{2}}\right]\Bigg).
\end{aligned}$$

Using the results of the previous section, we know that $\mathrm{E}[\|\nu^v\|^l_{\infty,0,T}] \leq C|\bar{u}|^l$ (by (2.50)). Since $\hat{b}_z$, $\hat{b}_u$, $\hat{\sigma}_z^j$ and $\hat{\sigma}_u^j$ are bounded, we get

$$\leq C\left(\mathrm{E}\left[|x_0(v)-x_0(u)|^l\right]+\mathrm{E}\left[\int_0^T \|\nu^v\|^l_{\infty,0,T}+|\bar{u}|^l\,dr\right]+\sum_{j=1}^{m_2}\mathrm{E}\left[\left(\int_0^T \|\nu^v\|^2_{\infty,0,T}+|\bar{u}|^2\,dr\right)^{\frac{l}{2}}\right]\right)$$

$$\leq C\left(\mathrm{E}\left[|x_0(v)-x_0(u)|^l\right]+|\bar{u}|^l\right).$$

By condition (B_3) the function $x_0 : \mathcal{U} \to \mathbb{R}^{n_2}$ is Lipschitz continuous, which yields

$$\mathrm{E}[\|\gamma^v\|^l_{\infty,s_0,T}] \leq C|\bar{u}|^l$$

for a constant $C > 0$, which is independent of u. Hence, we conclude

$$\lim_{\bar{u}\to 0} \mathrm{E}[\|x^{u+\bar{u}} - x^u\|^l_{\infty,0,T}] = 0$$

for every $u \in \mathcal{U}$. □

For notational simplicity, we define for $u \in \mathcal{U}$ and $r \in [0,T]$, $\hat{b}^u_x(r) := \hat{b}_x(r, x^u_r, \xi^u_r, u)$ and repectively $\hat{b}^u_z$, $\hat{b}^u_u$, $\hat{\sigma}^{u,j}_x$, $\hat{\sigma}^{u,j}_z$, $\hat{\sigma}^{u,j}_u$.

Lemma 2.43. *Suppose we are in the situation of Lemma 2.41, we have*

$$\lim_{|\bar{u}|\to 0} \mathrm{E}\left[\left\|\frac{x^{u+\bar{u}} - x^u - \hat{y}\bar{u}}{|\bar{u}|}\right\|^l_{\infty,0,T}\right] = 0$$

for an arbitrary $l \geq 1$, where $\hat{y}_t \in \mathbb{R}^{n_2\times d}$ is the solution to the inhomogenous linear SDE

$$\begin{aligned}\hat{y}^u_t = Dx_0(u) &+ \int_0^t \hat{b}^u_x(r)\hat{y}^u_r + \hat{b}^u_z(r)D\xi^u_r + b^u_u(r)\,dr \\ &+ \sum_{j=1}^{m_2}\int_0^t \hat{\sigma}^{u,j}_x(r)\hat{y}^u_r + \hat{\sigma}^{u,j}_z(r)D\xi^u_r + \hat{\sigma}^{u,j}_u(r)\,dB^j_r.\end{aligned} \tag{2.57}$$

Contrary to the previous section, we can also use Theorem 2.40 to get the unique solution to the matrix valued linear equation (2.57) and do not need to establish the explicit solution first. But the explicit solution will be needed later in the thesis. So we will first state an existence result for equation (2.57) and establish the boundedness of the solution, independent of the parameter, and then introduce the homogenous linear equations to give the explicit solution.

Lemma 2.44. *Let $\hat{b}$ and $\hat{\sigma}$ satisfy condition (B_1) and (B_2). Then for every $u \in \mathcal{U}$, the inhomogenous linear matrix equations (2.57) has a unique solution $\hat{y}^u$, such that $\hat{y}^u \in L^l_{\mathbb{F}}(\Omega, C([0,T]), \mathbb{R}^{n_2\times d})$ for every $l \geq 1$. Furthermore for every $l \geq 1$, $\mathrm{E}\left[\|\hat{y}^u\|^l_{\infty,0,T}\right]$ is bounded independently of u, by a constant $D_{y,l}$. Hence the solution process y_t defines a bounded linear operator $Dx^u := y^u$ from $\mathbb{R}^d$ to the space $L^l_{\mathbb{F}}(\Omega, C([0,T]), \mathbb{R}^{n_2})$ for every $l \geq 1$ and $u \in \mathcal{U}$.*

Proof. Take a look at the columns of the matrix equation (2.57) given by

$$\hat{y}^{u,i}_t = (Dx_0(u))^i + \int_0^t \hat{b}^u_x(r)\hat{y}^{u,i}_r + (\hat{b}^u_u(r))^i\,dr + \sum_{j=1}^{m_2}\int_0^t \hat{\sigma}^{u,j}_x(r)\hat{y}^{u,i}_r + (\hat{\sigma}^{u,j}_u(r))^i\,dB^j_r \tag{2.58}$$

for $i = 1, \dots, d$. Then for a given $u \in \mathcal{U}$ and every $i = 1, \dots, d$ the coefficient functions

$$f_i : [0,T] \times \mathbb{R}^{n_2} \times \Omega \to \mathbb{R}^{n_2}, (t, y, \omega), \mapsto \hat{b}_x(t, x_t^u(\omega), \xi_t^u(\omega), u)y + (\hat{b}_u(t, x_t^u(\omega), \xi_t^u(\omega), u))^i$$

and

$$g_i^j : [0,T] \times \mathbb{R}^{n_2} \times \Omega \to \mathbb{R}^{n_2}, (t, y, \omega), \mapsto \hat{\sigma}_x^j(t, x_t^u(\omega), \xi_t^u(\omega), u)y + (\hat{\sigma}_u^j(t, x_t^u(\omega), \xi_t^u(\omega), u))^i$$

satisfy condition (B) for every $l \geq 1$ on $[0,T]$, where the Lipschitz constants with respect to y do not depend on u. By Theorem 2.40, equation (2.58) has a unique solution for all $i = 1, \dots, d$ and we have for all $l \geq 1$ by (2.55) and the boundedness of the partial derivatives of the coefficients and $Dx_0(u)$

$$\begin{aligned}
&\mathrm{E}[\|\hat{y}^{u,i}\|_{\infty,0,T}^l] \\
&\leq C\left(\mathrm{E}\left[|(Dx_0(u))^i|^l\right] + \mathrm{E}\left[\int_0^T |(\hat{b}_u(r, x_r^u, \xi_r^u, u))^i|^l\, dr\right] + \sum_{j=1}^{m_2} \mathrm{E}\left[\int_0^T |(\hat{\sigma}_u^j(r, x_r^u, \xi_r^u, u))^i|^2\, dr^{\frac{l}{2}}\right]\right) \\
&\leq C,
\end{aligned}$$

where C is independent of u and i. We conclude that equation (2.57) has a unique solution, such that

$$\mathrm{E}\left[\|\hat{y}^u\|_{\infty,0,T}^l\right] \leq D_{y,l},$$

where $D_{y,l}$ is a constant independent of u. □

Now we come to the explicit solution of equation (2.57). As a inhomogenous linear Itô-SDE, we can calculate the explicit solution to the equation using the solutions to the matrix valued homogenous linear SDEs, defined for $t \in [s_0, T]$ and $u \in \mathcal{U}$ by

$$\hat{\phi}_t^{s_0} = I_{n_2} + \int_{s_0}^t \hat{b}_x^u(r)\hat{\phi}_r^{s_0}\, dr + \sum_{j=1}^{m_2} \int_{s_0}^t \hat{\sigma}_x^{u,j}(r)\hat{\phi}_r^{s_0}\, dB_r^j \tag{2.59}$$

and

$$\hat{\psi}_t^{s_0} = I_{n_2} - \int_{s_0}^t \hat{\psi}_r^{s_0}\left(\hat{b}_x^u(r) - \sum_{j=1}^{m_2}(\hat{\sigma}_x^{u,j}(r))^2\right) dr - \sum_{j=1}^{m_2} \int_{s_0}^t \hat{\psi}_r^{s_0}\hat{\sigma}_x^{u,j}(r)\, dB_r^j. \tag{2.60}$$

In the following lemma we give an existence and uniqueness result for the two SDEs, state a well known fact on the relation of the processes $\hat{\phi}$ and $\hat{\psi}$ and give the explicit solution to equation (2.57).

Lemma 2.45. *Let $u \in \mathcal{U}$ and $\hat{b}$, $\hat{\sigma}$ satisfy conditions (B_1) and (B_2). Then the equations* (2.59) *and* (2.60) *have a unique solution $\hat{\phi}^u \in L_{\mathbb{F}}^l(\Omega, C([s_0,T]), \mathbb{R}^{n_2 \times n_2})$ and $\hat{\psi}^u \in L_{\mathbb{F}}^l(\Omega, C([s_0,T]), \mathbb{R}^{n_2 \times n_2})$*

for $l \geq 1$ and there exists a constant $D_{\hat{\phi},l}$, independent of u and s_0, such that

$$\max\left\{\mathrm{E}\left[\|\hat{\phi}^u\|^l_{\infty,s_0,T}\right], \mathrm{E}\left[\|\hat{\psi}^u\|^l_{\infty,s_0,T}\right]\right\} \leq D_{\hat{\phi},l}.$$

Moreover, $\hat{\psi}^{s_0}_t = (\hat{\phi}^{s_0}_t)^{-1}$ for all $t \in [s_0, T]$, P-almost surely. Setting $s_0 = 0$ and $\hat{\phi}^0 = \phi$, the solution to equation (2.57) is given by

$$\begin{aligned}\hat{y}^u_t &= \hat{\phi}_t D x_0(u) + \hat{\phi}_t \int_0^t \hat{\phi}_r^{-1} \left(\hat{b}^u_z(t) D\xi^u_r + \hat{b}^u_u(r) - \sum_{j=1}^{m_2} \hat{\sigma}^{u,j}_x(r) \left(\hat{\sigma}^{u,j}_z(r) D\xi^u_r + \hat{\sigma}^{u,j}_u(r)\right)\right) dr \\ &\quad + \sum_{j=1}^{m_2} \hat{\phi}_t \int_0^t \hat{\phi}_r^{-1} \left(\hat{\sigma}^{u,j}_z(r) D\xi^u_r + \hat{\sigma}^{u,j}_u(r)\right) dB^j_r.\end{aligned}$$

Proof. The existence, uniqueness and boundedness of the solutions of the SDEs (2.59) and (2.60) can be proved completely analogous to the proof of Lemma 2.44. For the relation of the process and the explicit solution to equation (2.57), we refer the reader to Yong and Zhou [1999], Chapter 6.3. □

Now we come to the proof of Lemma 2.43.

Proof of Lemma 2.43. In this proof we use C for a positive constant, which can have different values at different occasions. To minimize the notational effort, we set $m = 1$ for the proof and leave out the explicit dependencies of $\hat{y}$ on u. Set

$$\begin{aligned} v &= u + \bar{u} \\ \gamma^v_\cdot &= x^v_\cdot - x^u_\cdot \\ \nu^v_\cdot &= \xi^v_\cdot - \xi^u_\cdot \\ \zeta^v_\cdot &= \frac{1}{|\bar{u}|}\gamma^v_\cdot, \end{aligned}$$

where $v \in \mathcal{U}$ for a $\bar{u} \in \mathbb{R}^d$ close enough to u, since $\mathcal{U}$ is open and convex. We have

$$\zeta^v_t = \frac{x_0(v) - x_0(u)}{\bar{u}} + \frac{1}{\bar{u}} \int_0^t \hat{b}(r, x^v_r, \xi^v_r, v) - \hat{b}(r, x^u_r, \xi^u_r, u)\, dr + \frac{1}{\bar{u}} \int_0^t \hat{\sigma}(r, x^v_r, \xi^v_r, v) - \hat{\sigma}(r, x^u_r, \xi^u_r, u)\, dB_r.$$

By using the mean value theorem (Lemma 2.25) similar to the proof of Lemma 2.42, we can write for $0 \leq t \leq T$:

$$\zeta^v_t = \zeta^v_0 + \int_0^t \tilde{b}^u_x(r)\zeta^v_r + \tilde{b}^u_z(r)\frac{\nu^v}{|\bar{u}|} + \tilde{b}^u_u(r)\frac{\bar{u}}{|\bar{u}|}\, dr + \int_0^t \tilde{\sigma}^u_x(r)\zeta^v_r + \tilde{\sigma}^u_z(r)\frac{\nu^v_r}{|\bar{u}|} + \tilde{\sigma}^u_u(r)\frac{\bar{u}}{|\bar{u}|}\, dB_r,$$

where

$$
\begin{aligned}
\zeta_0^v &= \frac{\gamma_0^v}{\bar{u}} \\
\tilde{b}_x^u(r) &= \int_0^1 \hat{b}_x(r, x_r^u + \lambda\gamma_r^v, \xi_r^u + \lambda\nu_r^v, u + \lambda\bar{u})\, d\lambda \\
\tilde{b}_z^u(r) &= \int_0^1 \hat{b}_z(r, x_r^u + \lambda\gamma_r^v, \xi_r^u + \lambda\nu_r^v, u + \lambda\bar{u})\, d\lambda \\
\tilde{b}_u^u(r) &= \int_0^1 \hat{b}_u(r, x_r^u + \lambda\gamma_r^v, \xi_r^u + \lambda\nu_r^v, u + \lambda\bar{u})\, d\lambda \\
\tilde{\sigma}_x^u(r) &= \int_0^1 \hat{\sigma}_x(r, x_r^u + \lambda\gamma_r^v, \xi_r^u + \lambda\nu_r^v, u + \lambda\bar{u})\, d\lambda \\
\tilde{\sigma}_z^u(r) &= \int_0^1 \hat{\sigma}_z(r, x_r^u + \lambda\gamma_r^v, \xi_r^u + \lambda\nu_r^v, u + \lambda\bar{u})\, d\lambda \\
\tilde{\sigma}_u^u(r) &= \int_0^1 \hat{\sigma}_u(r, x_r^u + \lambda\gamma_r^v, \xi_r^u + \lambda\nu_r^v, u + \lambda\bar{u})\, d\lambda.
\end{aligned}
$$

By setting $\eta_\cdot^v = \zeta_\cdot^v - \frac{\hat{y}_\cdot \bar{u}}{|\bar{u}|}$, we obtain

$$
\begin{aligned}
\eta_t^v = {}& \frac{\gamma_0^v - Dx_0(u)\bar{u}}{\bar{u}} + \int_0^t \tilde{b}_x^u(r)\eta_r^v + (\tilde{b}_x^u(r) - \hat{b}_x^u(r))\frac{\hat{y}_r\bar{u}}{|\bar{u}|} + \tilde{b}_z^u(r)\left(\frac{\nu_r^v - D\xi_r^u\bar{u}}{|\bar{u}|}\right) \\
& + (\tilde{b}_z^u(r) - \hat{b}_z^u(r))\frac{D\xi_r^u\bar{u}}{|\bar{u}|} + (\tilde{b}_u^u(r) - \hat{b}_u^u(r))\frac{\bar{u}}{|\bar{u}|}\, dr \\
& + \int_0^t \tilde{\sigma}_x^u(r)\eta_r^v + (\tilde{\sigma}_x^u(r) - \hat{\sigma}_x^u(r))\frac{\hat{y}_r}{|\bar{u}|} + \tilde{\sigma}_z^u(r)\left(\frac{\nu_r^v - D\xi_r^u\bar{u}}{|\bar{u}|}\right) \\
& + (\tilde{\sigma}_z^u(r) - \hat{\sigma}_z^u(r))\frac{D\xi_r^u\bar{u}}{|\bar{u}|} + (\tilde{\sigma}_u^u(r) - \hat{\sigma}_u^u(r))\frac{\bar{u}}{|\bar{u}|}\, dB_r.
\end{aligned} \tag{2.61}
$$

The partial derivatives of $\hat{b}$ and $\hat{\sigma}$ are bounded by Conditions $(B1)$ and $(B2)$, $x^u \in L^l_{\mathbb{F}}(\Omega, C([0,T], \mathbb{R}^{n_2})$ by Lemma 2.41, $\xi^u \in L^l_{\mathbb{F}}(\Omega, C([0,T], \mathbb{R}^{n_1})$,

$$
E[\|\hat{y}\|^l_{\infty,0,T}] \leq C, \quad E[\|D\xi^u\|^l_{\infty,0,T}] \leq C, \quad \mathrm{E}[\|\gamma^v\|^l_{\infty,0,T}] + \mathrm{E}[\|\nu^v\|^l_{\infty,0,T}] \leq C|\bar{u}|^l \tag{2.62}
$$

by Lemma 2.44, Lemma 2.42 and Corollary 2.37 for every $l \geq 1$. This implies that the coefficient functions of the SDE (2.61) satisfy condition (B) on $[s_0, T]$. Since

$$
\eta_0^v = \frac{x_0(v) - x_0(u) - Dx_0(u)\bar{u}}{\bar{u}}
$$

is a non-random vector in $\mathbb{R}^{n_2}$, applying Theorem 2.40, yields

$$
\mathrm{E}\left[\|\eta^v\|^l_{\infty,0,T}\right] \leq C|\eta_0^v|^l + C\mathrm{E}\Bigg[\int_0^T \left|(\tilde{b}_x^u(r) - \hat{b}_x^u(r))\hat{y}_r\right|^l + \left|\tilde{b}_z^u(r)\left(\frac{\nu_r^v - D\xi_r^u\bar{u}}{|\bar{u}|}\right)\right|^l
$$

$$+\left|(\tilde b^u_z(r)-\hat b^u_z(r))D\xi^u_r\right|^l+\left|(\tilde b^u_u(r)-\hat b^u_u(r))\right|^l\,dr\Bigg]$$
$$+C\mathrm{E}\Bigg[\Bigg(\int_0^T|(\tilde\sigma^u_x(r)-\hat\sigma^u_x(r))\hat y_r|^2+\left|\tilde\sigma^u_z(r)\left(\frac{\nu^v_r-D\xi^u_r\bar u}{|\bar u|}\right)\right|^2$$
$$+|(\tilde\sigma^u_z(r)-\hat\sigma^u_z(r))D\xi^u_r|^2+|(\tilde\sigma^u_u(r)-\hat\sigma^u_u(r))|^2\,dr\Bigg)^{\frac{l}{2}}\Bigg].$$

We know that for every $l\geq 1$

$$\lim_{|\bar u|\to 0}\mathrm{E}\left[\left\|\frac{\nu^v-D\xi^u\bar u}{|\bar u|}\right\|^l_{\infty,0,T}\right]=0$$

by Corollary 2.37 and by condition (B_3), we know that $x_0:\mathcal{U}\to\mathbb{R}^{n_2}$ is totally differentiable, which yields

$$\lim_{|\bar u|\to 0}\left|\frac{x_0(v)-x_0(u)-Dx_0(u)\bar u}{|\bar u|}\right|=0.$$

Furthermore the partial derivatives of the coefficient functions $\hat b$ and $\hat\sigma$ are bounded by a positive constant and continuous with respect to the variables x, z and u. This implies that for every $t\in[0,T]$ and $l\geq 1$

$$\lim_{|\bar u|\to 0}\Bigg(|\tilde b^u_x(t)-\hat b^u_x(t)|^l+|\tilde b^u_z(t)-\hat b^u_z(t)|^l+|\tilde b^u_u(t)-\hat b^u_u(t)|^l$$
$$|\tilde\sigma^u_x(t)-\hat\sigma^u_x(t)|^l+|\tilde\sigma^u_z(t)-\hat\sigma^u_z(t)|^l+|\tilde\sigma^u_u(t)-\hat\sigma^u_u(t)|^l\Bigg)=0$$

in probability. Taking the estimates (2.62) into account and applying the dominated convergence theorem we conclude

$$\lim_{\varepsilon\to 0}\mathrm{E}\left[\left\|\frac{x^v-x^u-\hat y\bar u}{|\bar u|}\right\|^l_{\infty,0,T}\right]=0.$$

□

Similar to the last section we are now able to formulate the main result of this section as a direct consequence of Lemma 2.43, Lemma 2.45 and Lemma 2.44.

Theorem 2.46. *For a open, bounded and convex subset $\mathcal{U}\subset\mathbb{R}^d$ the solution mapping $x.:\mathcal{U}\to L^l_{\mathbb{F}}(\Omega,C([0,T],\mathbb{R}^{n_2})$ for $l\geq 1$ corresponding to equation* (2.53)*, where the coefficient function $\hat b$, $\hat\sigma$ and the initial value x_0 satisfy condition (B_1), (B_2) and (B_3), is Fréchet differentiable with Fréchet differential*

$$Dx^u_t=\hat\phi_tDx_0(u)+\hat\phi_t\int_0^t\hat\phi_r^{-1}\left(\hat b^u_z(t)D\xi^u_r+\hat b^u_u(r)-\sum_{j=1}^{m_2}\hat\sigma^{u,j}_x(r)\left(\hat\sigma^{u,j}_z(r)D\xi^u_r+\hat\sigma^{u,j}_u(r)\right)\right)dr$$

$$+\sum_{j=1}^{m_2} \hat{\phi}_t \int_0^t \hat{\phi}_r^{-1} \left(\hat{\sigma}_z^{u,j}(r) D\xi_r^u + \hat{\sigma}_u^{u,j}(r)\right) dB_r^j, \tag{2.63}$$

which is the explicit solution to equation (2.57).

2.3 Fréchet differentiability of the model dynamics equation with respect to the parameter

In this section we summarize the results and conditions of the last two sections, to obtain the necessary preliminaries for our main result, which is stated in the next chapter. Let $T > 0$, $(w_t)_{t\in[0,T]}$ be an m_1-dimensional continuous, bounded p-variation stochastic process for $p \in (1, 2)$, which satisfies the exponential moment condition (2.48), and $(B_t)_{t\in[0,T]}$ be a m_2-dimensional standard Brownian motion, both defined on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$, both adapted to the filtration $\mathbb{F}$. Let $\mathcal{U}$ be a bounded, convex and open subset of $\mathbb{R}^d$. For every $u \in \mathcal{U}$, we consider the following system of parameter dependent stochastic differential equations

$$\mathcal{X}_t^u = \begin{pmatrix} \xi_t^u \\ x_t^u \end{pmatrix} = \begin{pmatrix} \xi_0(u) \\ x_0(u) \end{pmatrix} + \int_0^t \begin{pmatrix} b(r, \xi_r^u, u) \\ \hat{b}(r, x_r^u, \xi_r^u, u) \end{pmatrix} dr + \sum_{j=1}^{m_1} \int_0^t \begin{pmatrix} \sigma^j(r, \xi_r^u, u) \\ 0 \end{pmatrix} dw_r^j$$
$$+\sum_{j=1}^{m_2} \int_0^t \begin{pmatrix} 0 \\ \hat{\sigma}^j(r, x_r^u, \xi_r^u, u) \end{pmatrix} dB_r^j \tag{2.64}$$

for $t \in [0, T]$ and

i) assume that $\xi_0 : \mathcal{U} \to \mathbb{R}^{n_1}$, $b : [0, T] \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_1}$, $\sigma = (\sigma^1, \ldots, \sigma^{m_1}) : [0, T] \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_1 \times m_1}$ satisfy the conditions $(H_1) - (H_3)$. By Section 2.1, we know that the unique solution ξ^u exists in the space $L^l_{\mathbb{F}}(\Omega, C[0, T], \mathbb{R}^{n_1})$ (for all $l \geq 1$) and the solution mapping $\mathcal{U} \to L^l_{\mathbb{F}}(\Omega, C[0, T], \mathbb{R}^{n_1}), u \mapsto \xi^u$ is Fréchet differentiable with Fréchet differential $D\xi_\cdot^u$ defined in (2.42). The process $D\xi_\cdot^u$ is the unique solution to the linear stochastic Young differential equation

$$y_t^u = Dx_0(u) + \int_0^t b_x^u(r) y_r^u + b_u^u(r)\, dr + \sum_{j=1}^{m_1} \int_0^t \sigma_x^{u,j}(r) y_r^u + \sigma_u^{u,j}(r)\, dw_r^j.$$

ii) assume that $x_0 : \mathcal{U} \to \mathbb{R}^{n_2}$, $\hat{b} : [0, T] \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_2}$, $\hat{\sigma} = (\hat{\sigma}^1, \ldots, \hat{\sigma}^{m_2}) : [0, T] \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_2 \times m_2}$ satisfy the conditions $(B_1), (B_2)$ and (B_3). By Chapter 2.2, we know that the unique solution $x_\cdot^u$ exists, and is an element of $L^l_{\mathbb{F}}(\Omega, C[0, T], \mathbb{R}^{n_2})$ for every $l \geq 1$. Furthermore, the solution mapping $\mathcal{U} \to L^l_{\mathbb{F}}(\Omega, C[0, T], \mathbb{R}^{n_2}), u \mapsto x^u$ is Fréchet differentiable with Fréchet differential $Dx_\cdot^u$ defined in (2.63). The process $Dx_\cdot^u$ is the unique solution to the linear stochastic differential equation

$$\hat{y}^u_t = Dx_0(u) + \int_0^t \hat{b}^u_x(r)\hat{y}^u_r + \hat{b}^u_z(r)D\xi^u_r + \hat{b}^u_u(r)\,dr$$
$$+ \sum_{j=1}^{m_2}\int_0^t \hat{\sigma}^{u,j}_x(r)\hat{y}^u_r + \hat{\sigma}^{u,j}_z(r)D\xi^u_r + \hat{\sigma}^{u,j}_u(r)\,dB^j_r.$$

Hence, we know that the map $u \mapsto \mathcal{X}^u_\cdot$ from $\mathcal{U}$ to $L^l_{\mathbb{F}}(\Omega, C[0,T], \mathbb{R}^{(n_1+n_2)})$ is Fréchet differentiable at any $u \in \mathcal{U}$ and we have

$$D\mathcal{X}_t(u) = \mathcal{Y}_t(u)$$

for every $t \in [0,T]$, where $\mathcal{Y}_\cdot(u) \in L^l_{\mathbb{F}}(\Omega, C[0,T], \mathbb{R}^{(n_1+n_2)\times d})$ is the unique solution to the system of linear stochastic differential equations

$$\begin{aligned}
\mathcal{Y}^u_t &= \begin{pmatrix} y^u_t \\ \hat{y}^u_t \end{pmatrix} \\
&= \begin{pmatrix} D\xi_0(u) \\ Dx_0(u) \end{pmatrix} + \int_0^t \begin{pmatrix} b^u_x(r)y^u_r \\ \hat{b}^u_z(r)y^u_r + \hat{b}^u_x(r)\hat{y}^u_r \end{pmatrix} + \begin{pmatrix} b^u_u(r) \\ \hat{b}^u_u(r) \end{pmatrix} dr \\
&\quad + \sum_{j=1}^{m_1}\int_0^t \begin{pmatrix} \sigma^{u,j}_x(r)y^u_r \\ 0 \end{pmatrix} dw^j_r + \sum_{j=1}^{m_1}\int_0^t \begin{pmatrix} \sigma^{u,j}_u(r) \\ 0 \end{pmatrix} dw^j_r \\
&\quad + \sum_{j=1}^{m_2}\int_0^t \begin{pmatrix} 0 \\ \hat{\sigma}^u_z(r)y^u_r + \hat{\sigma}^{u,j}_x(r)\hat{y}^u_r \end{pmatrix} dB^j_r + \sum_{j=1}^{m_2}\int_0^t \begin{pmatrix} 0 \\ \hat{\sigma}^{u,j}_u(r) \end{pmatrix} dB^j_r \\
&= \begin{pmatrix} D\xi_0(u) \\ Dx_0(u) \end{pmatrix} + \int_0^t \begin{pmatrix} b^u_x(r) & 0 \\ \hat{b}^u_z(r) & \hat{b}^u_x(r) \end{pmatrix} \mathcal{Y}^u_r + \begin{pmatrix} b^u_u(r) \\ \hat{b}^u_u(r) \end{pmatrix} dr \\
&\quad + \sum_{j=1}^{m_1}\int_0^t \begin{pmatrix} \sigma^{u,j}_x(r) & 0 \\ 0 & 0 \end{pmatrix} \mathcal{Y}^u_r\,dw^j_r + \sum_{j=1}^{m_1}\int_0^t \begin{pmatrix} \sigma^{u,j}_u(r) \\ 0 \end{pmatrix} \bar{u}\,dw^j_r \\
&\quad + \sum_{j=1}^{m_2}\int_0^t \begin{pmatrix} 0 & 0 \\ \hat{\sigma}^u_z(r) & \hat{\sigma}^{u,j}_x(r) \end{pmatrix} \mathcal{Y}^u_r\,dB^j_r + \sum_{j=1}^{m_2}\int_0^t \begin{pmatrix} 0 \\ \hat{\sigma}^{u,j}_u(r) \end{pmatrix} dB^j_r.
\end{aligned} \tag{2.65}$$

Remark 2.47. In the previous chapters, we examined the components of the processes $\mathcal{X}$ and $\mathcal{Y}$ given by ξ, x and $y, \hat{y}$. Here we summarize the results concerning the boundedness of the components to find upper bounds for $\mathcal{X}$ and $\mathcal{Y}$. For every $u \in \mathcal{U}$, we have by Corollary 2.37, Lemma 2.41 and Lemma 2.33 for every $l \geq 1$

$$\begin{aligned}
\mathrm{E}\left[\|\mathcal{X}^u\|^l_{\infty,0,T}\right] &\leq 2^{l-1}\left(\mathrm{E}\left[\|\xi^u\|^l_{\infty,0,T}\right] + \mathrm{E}\left[\|x^u\|^l_{\infty,0,T}\right]\right) \\
&\leq 2^{l-1}\left(D_{\xi,l} + D_{x,l}\right) \\
&\leq D_{\mathcal{X},l}
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}\left[\|\mathcal{Y}^u\|_{\infty,0,T}^l\right] &\leq 2^{l-1}\left(\mathrm{E}\left[\|y^u\|_{\infty,0,T}^l\right] + \mathrm{E}\left[\|\hat{y}^u\|_{\infty,0,T}^l\right]\right) \\
&\leq 2^{l-1}\left(D_{y,l} + D_{\hat{y},l}\right) \\
&\leq D_{\mathcal{Y},l}.
\end{aligned}$$

Having established the Fréchet differentiability of the solution mapping to our model dynamics equation, we will now introduce the cost function for our calibration problem and focus on the representation of its gradient.

Chapter 3

The cost function and its gradient

We assume for the rest of this chapter that we are in the situation of Section 2.3 and all the stated conditions in i) and ii) are satisfied. Since we want to calibrate our financial model to e.g. call option prices, we introduce a cost function. Let $M > 0$ be a positive constant (e.g. the number of options we are observing in the market) and $T_1 \leq \cdots \leq T_M = T$ a set of times (e.g. the corresponding maturities) on the interval $(0, T]$. Define the cost function

$$J : \mathcal{U} \to \mathbb{R},\ u \mapsto \frac{1}{2} \sum_{\mu=1}^{M} \mathrm{E}[g_\mu(\mathcal{X}^u_{T_\mu})]^2, \tag{3.1}$$

where we assume the following condition on the functions g_μ

(G) Let L be the constant used in the conditions on the coefficient functions b, σ, $\hat{b}$ and $\hat{\sigma}$. For every $\mu = 1, \ldots, M$, we assume that $g_\mu : \mathbb{R}^{(n_1+n_2)} \to \mathbb{R}$ is a continuously differentiable function and we denote the derivative

$$g'_\mu(z) = \left(\frac{\partial}{\partial z_1} g_\mu(z), \ldots, \frac{\partial}{\partial z_{n_1+n_2}} g_\mu(z) \right) \in \mathbb{R}^{(n_1+n_2)}.$$

We assume for all $z, y \in \mathbb{R}^{n_1+n_2}$ that

$$|g'_\mu(z)| \leq L$$

and

$$|g'_\mu(z) - g'_\mu(y)| \leq L|z - y|.$$

This condition ensures that $\mathrm{E}[|g_\mu(z)|] < \infty$ for every $z \in L(\Omega, \mathbb{R}^{n_1+n_2})$, since

$$\mathrm{E}[|g_\mu(z)|] \leq L\mathrm{E}[|z|] + |g_\mu(0)|. \tag{3.2}$$

Our goal in this chapter is to calculate the gradient of the cost function in two different ways. First, we will use the chain rule for Fréchet differentials to make use of equation (2.65), which we call the sensitivity equation. Then we will establish our main result, namely an adjoint representation for the gradient, which is given by the explicit solution to an anticipating backwards stochastic differential equation. The key ingredient for the adjoint representation is the explicit solution to the $(n_1 + n_2) \times d$-dimensional system of linear SDEs (2.65), which we obtain by establishing a variation of constants formula similar to Lemma 2.33, respectively Lemma 2.45, but for the whole system of equations. Finding this explicit solution will be the first main goal of this chapter. In the course of the calculations for the variation of constant formula and the adjoint representation, we will encounter several technical problems. First, one of processes involved in the solutions of our system of differential equations (2.65) will be the product of a process driven by Brownian motion and a process driven by w. Therefore, we need an integration by parts rule which connects both of these processes. Remember that the involved integrals are of different type, one is a pathwise Young integral and the other is the standard Itô integral. Furthermore, for the calculation of the adjoint equation, we will have to deal with stochastic integrals with Brownian motion integrator, whose integrands are anticipating, such that we will not be able to use the Itô integral. Luckily, all these problems can be solved by applying a stochastic integral which generalizes both, the pathwise Young and the Itô integral. The next section is devoted to this generalization, called the forward integral by Russo and Vallois.

3.1 Forward Integration

In this section we will define the forward integral and constitute all the properties which will be necessary for our calculations. The forward integral, together with the backward and symmetric integral, was first introduced by F. Russo and P. Vallois in the paper Russo and Vallois [1993a] and further developed in the following years in Russo and Vallois [1993b], Russo and Vallois [1995], Russo and Vallois [1996] and Russo and Vallois [2000]. A good summary of the aforementioned papers can be found in the lecture notes Russo and Vallois [2007]. We will formulate all the definitions and results for scalar valued processes and then generalize the important results to the multidimensional case. Let $T > 0$ and $s_0 \in [0, T]$, as convention for this subsection we prolongate every real function f on $[s_0, T]$ by setting $f(t) = f(s_0)$ for $t \in (-\infty, s_0)$ and $f(t) = f(T)$ for $t \in [T, \infty)$. Let $(X_t)_{t\in[s_0,T]}$ be a continuous processes and $(Y_t)_{t\in[s_0,T]}$ be locally bounded, meaning that for every $t > s_0$, $\int_{s_0}^t Y_s \, ds < \infty$ P-almost surely. First we need to define the type of convergence we are interested in.

Definition 3.1. A family of processes $(H_t^\varepsilon)_{t\in[s_0,T]}$ converges to a process $(H_t)_{t\in[s_0,T]}$ in ucp-sense (uniform in probability), if

$$\lim_{\varepsilon\to 0} \sup_{t\in[s_0,T]} |H_t^\varepsilon - H_t| = 0$$

in probability.

Definition 3.2. The forward integral, backward integral and the generalized covariation are defined as the limit in ucp-sense of the ε-forward integral, ε-backward integral respectively the ε-covariation, if these limits exist. Precisely

$$\begin{aligned}
\varepsilon - \text{forward integral} : I^-(\varepsilon, Y, dX)(t) &= \int_{s_0}^t Y_s \frac{X_{s+\varepsilon} - X_s}{\varepsilon} \, ds \\
\varepsilon - \text{backward integral} : I^+(\varepsilon, Y, dX)(t) &= \int_{s_0}^t Y_s \frac{X_s - X_{s-\varepsilon}}{\varepsilon} \, ds \\
\varepsilon - \text{covariation} : C(\varepsilon, X, Y)(t) &= \int_{s_0}^t \frac{(X_{s+\varepsilon} - X_s)(Y_{(s+\varepsilon)} - Y_s)}{\varepsilon} \, ds
\end{aligned}$$

and

$$\begin{aligned}
&\text{Forward-integral} : \int_{s_0}^t Y_s \, d^- X_s = \lim_{\varepsilon \searrow 0} I^-(\varepsilon, Y, dX)(t) \\
&\text{Backward-integral} : \int_{s_0}^t Y_s \, d^+ X_s = \lim_{\varepsilon \searrow 0} I^+(\varepsilon, Y, dX)(t) \\
&\text{Generalized covariation} : [X, Y]_t = \lim_{\varepsilon \searrow 0} C(\varepsilon, X, Y)(t).
\end{aligned}$$

- We say that a process X is a finite quadratic variation process, if the generalized covariation $[X] = [X, X]$ exists. In this case we call $[X]$ the quadratic variation of X.
- X is a zero quadratic variation process, if $[X] = 0$.
- A vector $((X_s^1, \ldots, X_s^n))_{s \in [s_0, T]}$ of continuous processes is said to have all its mutual covariations if $[X^i, X^j]$ exists for all $i, j = 1, \ldots, n$.

Remark 3.3. Let for some $n \in \mathbb{N}$, $(I_j)_{1 \leq j \leq n}$ be a sequence such that I_j is either the forward, backward integral or the generalized covarition of some processes, by convention for the rest of this subsection, an identity of the form $\sum_{j=1}^n I_j = 0$ means that, if we assume $n-1$ of the involed limits exist, then the n-th limit exists and and the identity holds true.

We will first establish all of the properties of the generalized covariation which will be needed in our calculations.

Proposition 3.4. *i) For continuous processes $(X_t)_{t \in [s_0, T]}$ and $(Y_t)_{t \in [s_0, T]}$ the operations $(X, Y) \to \int_{s_0}^t X \, d^- Y_s$ and $(X, Y) \to [X, Y]$ are bilinear.*

ii) $[X]$ is an increasing process, if it exists.

iii) $[X, Y|_t = \int_{s_0}^t X_s \, d^+ Y_s - \int_{s_0}^t X_s \, d^- Y_s$.

iv) If $(X_t)_{t \in [s_0, T]}$ and $(Y_t)_{t \in [s_0, T]}$ are finite quadratic variation processes, we have

$$|[X, Y]| \leq ([X][Y])^{\frac{1}{2}}.$$

v) If $(X_t)_{t\in[s_0,T]}$ is a finite quadratic variation process and $(Y_t)_{t\in[s_0,T]}$ a zero qudratic variation process, then $[X,Y]=0$.

vi) If $(Y_t)_{t\in[s_0,T]}$ is a continuous process and the paths of the process $(X_t)_{t\in[s_0,T]}$ are almost surely continuously differentiable, then

$$\int_{s_0}^{t} Y_r\, d^- X_s = \int_{s_0}^{t} Y_r X_s'\, ds.$$

vii) Let $(X_t)_{t\in[s_0,T]}$ and $(Y_t)_{t\in[s_0,T]}$ be $\mathbb{F}$-local martingales on a filtered probability space $(\Omega,\mathcal{F},\mathbb{F},P)$, H and $H^{'}$ be progressively measurable processes such that

$$\int_{s_0}^{t} H_r\, d\langle X,X\rangle_r < \infty \text{ and } \int_{s_0}^{t} H_r^{'}\, d\langle Y,Y\rangle_r < \infty$$

then

$$\left[\int_{s_0}^{t} H_r\, dX_r, \int_{s_0}^{t} H_r^{'}\, dY_r\right] = \int_{s_0}^{t} H_r H_r^{'}\, d[X,Y]_r.$$

Here $\langle X,X\rangle_t$ is the standard quadratic variation, defined as the limit in probability

$$\lim_{|\Pi_k|\to 0} \sum_{j=0}^{k-1} (X_{t_{i+1}} - X_{t_i})^2,$$

where $(\Pi_k)_{k\in\mathbb{N}}$ is a sequence of partitions of the interval $[s_0,t]$ converging to zero in mesh.

viii) Let $(M_t)_{t\in[s_0,T]}$ be an $\mathbb{F}$-local martingale, then $\langle M,M\rangle = [M,M]$.

ix) Let $(M_t)_{t\in[s_0,T]}$ be an continuous $\mathbb{F}$-local martingale, $(Y_t)_{t\in[s_0,T]}$ a càdlàg and $\mathbb{F}$-adapted process. If M and Y are independent then $[M,Y]=0$.

Proof. Properties *i*), *ii*) and *iii*) directly follow from the definition of the integral and the generalized covariation. The proof of *iv*), *vii*), *viii*) and *ix*) are given in Russo and Vallois [2007]. *v*) is a direct consequence of *iv*). *vi*) can easily be seen by applying the dominated convergence theorem. □

The following results show that the forward integral generalizes the Riemann Stieltjes, the Young and the Itô integral.

Theorem 3.5. *Let $(X_t)_{t\in[s_0,T]}$ be a continuous, bounded 1-variation process and $(Y_t)_{t\in[s_0,T]}$ be a continuous process, then*

$$\int_{s_0}^{t} Y_s\, d^+ X_s = \int_{s_0}^{t} Y_s\, d^- X_s = \int_{s_0}^{t} Y_s\, dX_s,$$

where the integral on the right hand side of the equation is a Riemann-Stieltjes integral.

Proof. See Russo and Vallois [2007], Proposition 1 7a). Here the authors prove a version of this statement with less restrictive conditions on the processes X and Y for the Lebesgue-Stieltjes integral. Applying the proof under the assumption of continuity of both processes yields our statement. $\square$

In case of the Young integral, we only found results in the literature in the case where integrand and integrator are Hölder continuous (see Russo and Vallois [2007]). We will generalize these results to continuous integrand and integrator having finite p-respectively q-variation, such that $\frac{1}{p} + \frac{1}{q} > 1$. To proof that the forward integral coincides with the Young integral, we first need a preliminary result.

Lemma 3.6. *Let $p' > p \geq 1$ and X be a continuous finite p-variation process, define for $\varepsilon > 0$ and $t \in [s_0, T]$*

$$X_t^{\varepsilon -} = \frac{1}{\varepsilon} \int_{s_0}^{t} X_{r+\varepsilon} - X_r \, dr$$

and

$$X_t^{\varepsilon +} = \frac{1}{\varepsilon} \int_{s_0}^{t} X_r - X_{r-\varepsilon} \, dr. \tag{3.3}$$

Then

$$\lim_{\varepsilon \to 0} |X^{\varepsilon -} - X|_{p', s_0, T} = \lim_{\varepsilon \to 0} |X^{\varepsilon +} - X|_{p', s_0, T} = 0.$$

Proof. We prove the statement for $X^{\varepsilon -}$, the proof for $X^{\varepsilon +}$ is completely analogue. For any $s_0 \leq t \leq T$, we define $Z_t^\varepsilon = X_t^{\varepsilon -} - X_t$. We have

$$X_t^{\varepsilon -} = \frac{1}{\varepsilon} \int_{s_0}^{t} X_{r+\varepsilon} - X_r \, dr = \frac{1}{\varepsilon} \int_{t}^{t+\varepsilon} X_r \, dr - \frac{1}{\varepsilon} \int_{s_0}^{s_0+\varepsilon} X_r \, dr$$

and

$$\begin{aligned} Z_t^\varepsilon - Z_s^\varepsilon &= \frac{1}{\varepsilon} \int_{t}^{t+\varepsilon} X_r - X_t \, dr - \frac{1}{\varepsilon} \int_{s}^{s+\varepsilon} X_r - X_s \, dr, \\ &= \frac{1}{\varepsilon} \int_{0}^{\varepsilon} X_{t+r} - X_t - (X_{s+r} - X_s) \, dr, \end{aligned}$$

where $s_0 \leq s \leq t \leq T$. We have using the Jensen inequality for $s_0 \leq s \leq t \leq T$

$$\begin{aligned} |Z_t^\varepsilon - Z_s^\varepsilon| &\leq \frac{1}{\varepsilon} \int_{0}^{\varepsilon} |X_{t+r} - X_t - (X_{s+r} - X_s)| \, dr \\ &\leq \left(\frac{1}{\varepsilon} \int_{0}^{\varepsilon} |X_{t+r} - X_t - (X_{s+r} - X_s)|^{p'-p} |X_{t+r} - X_t - (X_{s+r} - X_s)|^{p} \, dr \right)^{\frac{1}{p'}} \\ &\leq 2^{1-\frac{p}{p'}} \sup_{r \in [s_0, T]} \operatorname{Osc}(X, [r, r+\varepsilon])^{1-\frac{p}{p'}} \left(2^{p-1} \frac{1}{\varepsilon} \int_{0}^{\varepsilon} |X_{t+r} - X_{s+r}|^{p} + |X_t - X_s|^{p} \, dr \right)^{\frac{1}{p'}} \end{aligned}$$

$$\leq 2^{1-\frac{1}{p'}} \sup_{r\in[s_0,T]} \mathrm{Osc}(X,[r,r+\varepsilon])^{1-\frac{p}{p'}} \left(\frac{1}{\varepsilon}\int_0^\varepsilon |X_{(\cdot+r)}|^p_{p,s,t}\, dr + |X|^p_{p,s,t}\right)^{\frac{1}{p'}}.$$

It is easy to see that

$$\varphi(s,t) = \frac{1}{\varepsilon}\int_0^\varepsilon |X_{(\cdot+r)}|^p_{p,s,t}\, dr + |X|^p_{p,s,t}$$

is superadditive on $\Delta([s_0,T])$. By Lemma 2.6, this yields

$$\begin{aligned}
|Z^\varepsilon|_{p',s,t} &\leq 2^{1-\frac{1}{p'}} \sup_{r\in[s_0,T]} \mathrm{Osc}(X,[r,r+\varepsilon])^{1-\frac{p}{p'}} \left(\frac{1}{\varepsilon}\int_0^\varepsilon |X_{(\cdot+r)}|^p_{p,s,t}\, dr + |X|^p_{p,s,t}\right)^{\frac{1}{p'}} \\
&\leq 2^{1-\frac{1}{p'}} \left(2|X|^p_{p,s_0,T}\right)^{\frac{1}{p'}} \sup_{r\in[s_0,T]} \mathrm{Osc}(X,[r,r+\varepsilon])^{1-\frac{p}{p'}} \\
&\leq C \sup_{r\in[s_0,T]} \mathrm{Osc}(X,[r,r+\varepsilon])^{1-\frac{p}{p'}}.
\end{aligned}$$

Since X is uniformly continuous on $[s_0,T]$ and $X_t = X_T$ for $t \geq T$, this yields

$$\lim_{\varepsilon\to 0} |X^{\varepsilon-} - X|_{p',s_0,T} = 0.$$

□

Theorem 3.7. *Let $(X_t)_{t\in[s_0,T]}, (Y_t)_{t\in[s_0,T]}$ be two real valued processes such that the paths of X are almost surely continuous and of finite p-variation for $p \geq 1$ the paths of Y are almost surely continuous and of finite q-variation for $q \geq 1$ such that $\frac{1}{p}+\frac{1}{q} > 1$. Then*

$$\int_{s_0}^t Y_s\, d^+X_s = \int_{s_0}^t Y_s\, d^-X_s = \int_{s_0}^t Y_s\, d^{(y)}X_s.$$

Proof. For $t \in [s_0,T]$ define the process

$$X_t^{\varepsilon-} = \frac{1}{\varepsilon}\int_{s_0}^t X_{r+\varepsilon} - X_r\, dr$$

and the process

$$Z_t = \int_{s_0}^t Y_s\, d^{(y)}X_s - \int_{s_0}^t Y_s \frac{X_{s+\varepsilon}-X_s}{\varepsilon}\, ds.$$

Notice that for almost every $\omega \in \Omega$ the paths $X_t^{\varepsilon-}(\omega)$ are continuously differentiable, such that the Young integral

$$\int_{s_0}^t Y_s\, d^{(y)}X_s^{\varepsilon-}$$

coincides with the standard Riemann-Stieltjes integral

$$\int_{s_0}^{t} Y_s \, dX_s^{\varepsilon-}$$

and we have the relation

$$\int_{s_0}^{t} Y_s \frac{X_{s+\varepsilon} - X_s}{\varepsilon} \, ds = \int_{s_0}^{t} Y_s (X_s^{\varepsilon-})' \, ds = \int_{s_0}^{t} Y_s \, dX_s^{\varepsilon-} = \int_{s_0}^{t} Y_s \, d^{(y)} X_s^{\varepsilon-} \; P\text{-a.s.}$$

by the standard formula for Riemann-Stieltjes integrals with differentiable integrator. From our results on Young integrals it follows that for $p' > p$ such that $\frac{1}{q} + \frac{1}{p'} > 1$

$$\|Z_t\|_{\infty,s_0,T} = \left\| \int_{s_0}^{t} Y_s \, d^{(y)}(X_s - X_s^{\varepsilon-}) \right\|_{\infty,s_0,T} \leq C_{L.Y.} \|Y\|_{q,s_0,T} |X - X^{\varepsilon-}|_{p',s_0,T}.$$

By Lemma 3.6, we conclude that Z_t converges uniformly P-almost surely to 0 for $\varepsilon \to 0$, which implies that

$$\int_{s_0}^{t} Y_s \frac{X_{s+\varepsilon} - X_s}{\varepsilon} \, ds \to \int_{s_0}^{t} Y_s \, d^{(y)} X_s$$

uniformly in probability, for $\varepsilon \to 0$. The assertion for the backward integral can be proved analogously by using $X^{\varepsilon+}$ (see (3.3)) instead of $X^{\varepsilon-}$. $\square$

Using the last theorem we can establish a result concerning the generalized covariation of a continuous, bounded p-variation process.

Corollary 3.8. *Every continuous, bounded p-variation process $(X_t)_{t\in[s_0,T]}$ for $p \in [1,2)$ is a zero quadratic variation process.*

Proof. The statement follows directly by Proposition 3.4 *iii*) and Theorem 3.7. $\square$

Theorem 3.9. *Let $(X_t)_{t\in[s_0,T]}$ be a continuous $\mathbb{F}$-local martingale on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$ and $(Y_t)_{t\in[s_0,T]}$ be a progressively measurable, locally bounded, càdlàg process. Then the forward integral*

$$\int_{s_0}^{t} Y_s \, d^- X_s$$

coincides with the Itô-integral

$$\int_{s_0}^{t} Y_s \, dX_s.$$

Proof. See Russo and Vallois [2007], Proposition 6. $\square$

Remark 3.10. For an $\mathcal{F}_T$ measurable random variable Z and a progressively measurable, locally bounded, càdlàg process $(X_t)_{t\in[s_0,T]}$ we have

$$Z\int_{s_0}^{T} X_s\,dB_s = \int_{s_0}^{T} ZX_s\,d^-B_s$$

as the multiplication with a random variable does not change the existence of the forward integral as limit uniform in probability. This shows that the forward integral is suitable for anticipating integrands.

Since we are in a multidimensional setting in this thesis, we need to generalize the previous results correspondingly. The definition of the forward integral and the generalized covariation in the multidimensional case is straightforward. Let $(X_t)_{t\in[s_0,T]}$ be a continuous $m\times n$-dimensional process and $(Y_t)_{t\in[s_0,T]}$ be a continuous $n\times k$-dimensional process, then for $t\in[s_0,T]$

$$\int_{s_0}^{t} X_s\,d^-Y_s$$

is given by the ucp-limit of

$$\int_{s_0}^{t} X_s\frac{Y_{s+\varepsilon}-Y_s}{\varepsilon}\,ds = \left(\sum_{l=1}^{n}\frac{1}{\varepsilon}\int_{s_0}^{t} X_s^{i,l}\left(Y_{s+\varepsilon}^{l,j}-Y_s^{l,j}\right)\,ds\right)_{1\le i\le m,1\le j\le k}$$

for $\varepsilon\to 0$ if this limit exists and

$$[X,Y]_t$$

is given by the ucp-limit of

$$\frac{1}{\varepsilon}\int_{s_0}^{t}(X_{s+\varepsilon}-X_s)(Y_{s+\varepsilon}-Y_s)\,ds = \left(\sum_{l=1}^{n}\frac{1}{\varepsilon}\int_{s_0}^{t}\left(X_{s+\varepsilon}^{i,l}-X_s^{i,l}\right)\left(Y_{s+\varepsilon}^{l,j}-Y_s^{l,j}\right)\,ds\right)_{1\le i\le m,1\le j\le k}$$

for $\varepsilon\to 0$ if this limit exists. If both of the limits exists we can write

$$\int_{s_0}^{t} X_s\,d^-Y_s = \left(\sum_{l=1}^{n}\int_{s_0}^{t} X_s^{i,l}\,d^-Y_s^{l,j}\right)_{1\le i\le m,1\le j\le k}\in\mathbb{R}^{m\times k}$$

and

$$[X,Y]_t = \left(\sum_{l=1}^{n}\left[X^{i,l},Y^{l,j}\right]_t\right)_{1\le i\le m,1\le j\le k}\in\mathbb{R}^{m\times k}.$$

The multidimensional versions of the Riemann-Stieltjes, Young and Itô integral are analogously defined, such that the results in Theorems 3.5, 3.7 and 3.9 hold in the multidimensional case. The main result concerning the forward integral, which we need for our calculations, is the integration by parts formula. We first state and prove the formula for $\mathbb{R}$-valued processes.

Proposition 3.11 (Integration by parts)**.** *Let $(X_t)_{t\in[s_0,T]}$ and $(Y_t)_{t\in[s_0,T]}$ be continuous, $\mathbb{R}$-valued processes. Then*

$$X_tY_t = X_0Y_0 + \int_{s_0}^t X_s\, d^-Y_s + \int_{s_0}^t Y_s\, d^-X_s + [X,Y]_t.$$

Proof. We have

$$\begin{aligned}
&\frac{1}{\varepsilon}\int_{s_0}^t X_s(Y_{s+\varepsilon}-Y_s)\, ds + \frac{1}{\varepsilon}\int_{s_0}^t Y_s(X_{s+\varepsilon}-X_s)\, ds \\
&= \frac{1}{\varepsilon}\int_{s_0}^t (X_{s+\varepsilon}Y_{s+\varepsilon}-X_sY_s)\, ds - \frac{1}{\varepsilon}\int_{s_0}^t (X_{s+\varepsilon}-X_s)(Y_{s+\varepsilon}-Y_s)\, ds.
\end{aligned}$$

The first integral can be rewritten as

$$\begin{aligned}
\frac{1}{\varepsilon}\int_{s_0}^t (X_{s+\varepsilon}Y_{s+\varepsilon}-X_sY_s)\, ds &= \frac{1}{\varepsilon}\int_{s_0+\varepsilon}^{t+\varepsilon} X_sY_s\, ds - \int_{s_0}^t X_sY_s\, ds \\
&= \frac{1}{\varepsilon}\int_{t}^{t+\varepsilon} X_sY_s\, ds - \int_{s_0}^{s_0+\varepsilon} X_sY_s\, ds.
\end{aligned}$$

Then by the fundamental theorem of calculus, this integral converges uniformly P-a.s. to $X_tY_t - X_{s_0}Y_{s_0}$. Keeping Remark 3.3 in mind, the assertion follows. $\square$

In the multidimensional case we need to pay attention on the order of multiplication of the given matrix-valued processes. For the multidimensional integration by parts formula, we introduce another form of the forward integral for a continuous $m\times n$-dimensional process $(X_t)_{t\in[s_0,T]}$ and a continuous $n\times k$-dimensional process $(Y_t)_{t\in[s_0,T]}$, given by

$$\int_{s_0}^t (d^-X_s)Y_s$$

as the ucp-limit of

$$\begin{aligned}
\frac{1}{\varepsilon}\int_{s_0}^t (X_{s+\varepsilon}-X_s)Y_s\, ds &= \left(\sum_{l=1}^n \frac{1}{\varepsilon}\int_{s_0}^t \left(X_{s+\varepsilon}^{i,l}-X_s^{i,l}\right)Y_s^{l,j}\, ds\right)_{1\le i\le m, 1\le j\le k} \in \mathbb{R}^{m\times k} \\
&= \left(\sum_{l=1}^n \frac{1}{\varepsilon}\int_{s_0}^t Y_s^{l,j}\left(X_{s+\varepsilon}^{i,l}-X_s^{i,l}\right)\, ds\right)_{1\le i\le m, 1\le j\le k} \in \mathbb{R}^{m\times k},
\end{aligned}$$

for $\varepsilon\to 0$. If this limit exists, we have

$$\int_{s_0}^t (d^-X_s)Y_s = \left(\sum_{l=1}^n \int_{s_0}^t Y_s^{l,j}\, d^-X^{i,l}\right)_{1\le i\le m, 1\le j\le k} \in \mathbb{R}^{m\times k}.$$

Now we are able to state and prove the multidimensional integration by parts formula.

Theorem 3.12. *Let $(X_t)_{t\in[s_0,T]}$ be a continuous $m\times n$-dimensional process and $(Y_t)_{t\in[s_0,T]}$ be a continuous $n\times k$-dimensional process, then for $t\in[s_0,T]$ we have*

$$X_tY_t = X_{s_0}Y_{s_0} + \int_{s_0}^t X_s\, d^-Y_s + \int_{s_0}^t (d^-X_s)Y_s + [X,Y]_t.$$

Proof. Keeping Remark 3.3 in mind, let $i\in\{1,\dots,n\}$ and $j\in\{1,\dots,k\}$, we have the equality

$$\begin{aligned}&\left(\int_{s_0}^t X_s\, d^-Y_s\right)^{i,j} + \left(\int_{s_0}^t (d^-X_s)Y_s\, ds\right)^{i,j}\\ &= \sum_{l=1}^n \int_{s_0}^t X_s^{i,l}\, d^-Y_s^{l,j} + \sum_{l=1}^n \int_{s_0}^t Y_s^{l,j}\, d^-X_s^{i,l}.\end{aligned}$$

By the 1-dimensional integration by parts formula from Theorem 3.11 this equals

$$\begin{aligned}&= \sum_{l=1}^n \left(X_t^{i,l}Y_t^{l,j} - X_{s_0}^{i,l}Y_{s_0}^{l,j} - \left[X^{i,l},Y^{l,j}\right]\right)\\ &= (X_tY_t)^{i,j} - (X_{s_0}Y_{s_0})^{i,j} - ([X,Y]_t)^{i,j}.\end{aligned}$$

This proves the assertion. $\square$

The next theorem is the main result of this subsection and will be used multiple times in the remainder of this thesis. But first we introduce for $l\geq 1$ and $p\geq 1$ the space

$$\begin{aligned}&L^l_{\mathbb{F}}(\Omega, C^p[s_0,T],\mathbb{R}^{n_1+n_2})\\ &:= \{x:\Omega\times[s_0,T]\to\mathbb{R}^{n_1+n_2} \mid x \text{ is } \mathbb{F}\text{-adapted continuous, bounded } p\text{-variation process}\\ &\qquad \text{such that } \mathrm{E}[\|x\|^l_{p,s_0,T}]<\infty\}.\end{aligned}$$

Theorem 3.13. *Let $(\Omega,\mathcal{F},\mathbb{F},P)$ be a filtered probability space, carrying an m_1-dimensional continuous, bounded p-variation process $(w_t)_{t\in[0,T]}$ for $p\in(1,2)$ and an m_2-dimensional standard Brownian motion $(B_t)_{t\in[0,T]}$, both adapted to the filtration $\mathbb{F}$. Furthermore let $A\in L^l_{\mathbb{F}}(\Omega,C[s_0,T],\mathbb{R}^{m\times n})$, $C^i\in L^l_{\mathbb{F}}(\Omega,C^q[s_0,T],\mathbb{R}^{m\times n})$, $D^j\in L^l_{\mathbb{F}}(\Omega,C[s_0,T],\mathbb{R}^{m\times n})$, $\hat{A}\in L^l_{\mathbb{F}}(\Omega,C[s_0,T],\mathbb{R}^{n\times k})$, $\hat{C}^i\in L^l_{\mathbb{F}}(\Omega,C^q[s_0,T],\mathbb{R}^{n\times k})$, $\hat{D}^j\in L^l_{\mathbb{F}}(\Omega,C[s_0,T],\mathbb{R}^{n\times k})$ for every $l\geq 1$, $i=1,\dots,m_1$, $j=1,\dots,m_2$ and $q\geq p$ such that $\frac{1}{p}+\frac{1}{q}>1$. Define the processes $(X_t)_{t\in[s_0,T]}$ and $(Y_t)_{t\in[s_0,T]}$ by*

$$X_t = X_{s_0} + \int_{s_0}^t A_s\, ds + \sum_{j=1}^{m_1}\int_{s_0}^t C_s^j\, dw_s^j + \sum_{j=1}^{m_2}\int_{s_0}^t D_s^j\, dB_s^j$$

and

$$Y_t = Y_{s_0} + \int_{s_0}^t \hat{A}_s\, ds + \sum_{j=1}^{m_1}\int_{s_0}^t \hat{C}_s^j\, dw_s^j + \sum_{j=1}^{m_2}\int_{s_0}^t \hat{D}_s^j\, dB_s^j.$$

Then we have

$$
\begin{aligned}
X_t Y_t = X_{s_0} Y_{s_0} &+ \int_{s_0}^{t} X_s \hat{A}_s \, ds + \sum_{j=1}^{m_1} \int_{s_0}^{t} X_s \hat{C}_s^j \, dw_s^j + \sum_{j=1}^{m_2} X_s \hat{D}_s^j \, dB_s^j \\
&+ \int_{s_0}^{t} A_s Y_s \, ds + \sum_{j=1}^{m_1} \int_{s_0}^{t} C_s^j Y_s \, dw_s^j + \sum_{j=1}^{m_2} D_s^j Y_s \, dB_s^j \\
&+ \sum_{j=1}^{m_2} \int_{s_0}^{t} D_s^j \hat{D}_s^j \, ds
\end{aligned}
$$

for all $t \in [s_0, T]$.

Proof. The processes $(X_t)_{t\in[s_0,T]}$ and $(Y_t)_{t\in[s_0,T]}$ are $\mathbb{F}$-adapted, continuous and of bounded q_0-variation for every $q_0 > 2$, such that $\frac{1}{p} + \frac{1}{q_0} > 1$. To see this, take a look at the 3 integral processes

- $$(\omega, t) \mapsto \int_{s_0}^{t} A_s(\omega) \, ds.$$

  Since A_s is a continuous process, the integral process is of bounded 1-variation and therefore of bounded q_0-variation.

- $$(\omega, t) \mapsto \int_{s_0}^{t} \hat{C}_s^j(\omega) \, dw_s^j(\omega).$$

  For every $j = 1, \ldots, m_1$, C^j is a continuous process with bounded q-variation, where $\frac{1}{p} + \frac{1}{q} > 1$. By Lemma 2.16 the integral process is of bounded p-variation and again, since $q_0 > 2 > p$ of bounded q_0-variation.

- $$(\omega, t) \mapsto \int_{s_0}^{t} D_s^j(\omega) \, dB_s^j(\omega).$$

  Since for every $j = 1, \ldots, m_2$, $D^j \in L^2_{\mathbb{F}}(C([s_0, T], \mathbb{R}^{m\times n})$ every component of the matrix valued integral process is an $\mathbb{F}$-martingale. By Corollary 12.7 of Dudley and Norvaiša [2010], every semimartingale (and hence every martingale) has bounded p-variation for $p > 2$ on a bounded interval. Hence the integral process is of bounded q_0-variation.

The same arguments can be used for the bounded q_0-variation of Y. By the continuity of X and Y we can use Theorem 3.12 and obtain

$$X_t Y_t = X_{s_0} Y_{s_0} + \int_{s_0}^{t} X_s \, d^- Y_s + \int_{s_0}^{t} (d^- X_s) Y_s + [X, Y]_t. \tag{3.4}$$

Now take a look at the first forward integral in the previous equation, we have by the biliniarity of the forward integral that

$$\begin{aligned}\int_{s_0}^t X_s\, d^- Y_s &= \int_{s_0}^t X_s\, d^- \left(\int_{s_0}^s A_r\, dr\right) + \sum_{j=1}^{m_1} \int_{s_0}^t X_s\, d^- \left(\int_{s_0}^s C_r^j\, dw_r^j\right) + \sum_{j=1}^{m_2} \int_{s_0}^t X_s\, d^- \left(\int_{s_0}^s D_r^j\, dB_r^j\right)\\ &= I_1 + I_2 + I_3.\end{aligned}$$

For the integral I_1 we know that the (i,j)-th component of the matrix valued process

$$(\omega, s) \mapsto \int_{s_0}^s A_r(\omega)\, dr$$

is P-a.s. continuously differentiable with differential $A_s^{i,j}$, by Proposition 3.4 vi), we have

$$I_1 = \int_{s_0}^t X_s A_s\, ds.$$

The processes

$$(\omega, s) \mapsto \int_{s_0}^s C_r^j(\omega)\, dw_r^j(\omega)$$

are continuous, bounded p-variation processes for every $j = 1, \dots, m_1$ by Lemma 2.16, since $q \geq p$ and $\frac{1}{p} + \frac{1}{q} > 1$. By Theorem 3.7, we have

$$I_2 = \sum_{j=1}^{m_1} \int_{s_0}^t X_s\, d^{(y)} \left(\int_{s_0}^s C_r^j\, dw_r^j\right).$$

By our substitution rule from Lemma 2.17, since X is a continuous process with bounded q_0-variation for $\frac{1}{p} + \frac{1}{q_0} > 1$, we obtain

$$I_2 = \sum_{j=1}^{m_1} \int_{s_0}^t X_s C_s^j\, dw_s^j.$$

For the Itô integral, note that the (i,j)-th component of the matrix valued process

$$(\omega, t) \mapsto \int_{s_0}^t D_s^j(\omega)\, dB_s^j(\omega)$$

is a $\mathbb{F}$-martingale such that by Theorem 3.9, the forward integral coincides with the Itô integral, such that

$$I_3 = \sum_{j=1}^{m_2} \int_{s_0}^t X_s\, d \left(\int_{s_0}^s D_r^j\, dB_r^j\right).$$

Our substitution rule from Lemma 2.17 can also be formulated for the Itô integral (see Karatzas and Shreve [1991], Corollary 3.2.20), this yields

$$I_3 = \sum_{j=1}^{m_2} \int_{s_0}^{t} X_s D_s^j \, dB_r^j.$$

Collecting the terms, we obtain

$$\int_{s_0}^{t} X_s \, d^- Y_s = \int_{s_0}^{t} X_s A_s \, ds + \sum_{j=1}^{m_1} \int_{s_0}^{t} X_s C_s^j \, dw_s^j + \sum_{j=1}^{m_2} \int_{s_0}^{t} X_s D_s^j \, dB_s^j.$$

The calculations for the integral

$$\int_{s_0}^{t} (d^- X_s) Y_s$$

are completely analogous and we get

$$\int_{s_0}^{t} (d^- X_s) Y_s = \int_{s_0}^{t} \hat{A}_s Y_s \, ds + \sum_{j=1}^{m_1} \int_{s_0}^{t} \hat{C}_s^j Y_s \, dw_s^j + \sum_{j=1}^{m_2} \int_{s_0}^{t} \hat{D}_s^j Y_s \, dB_s^j.$$

Since both forward integrals in (3.4) exist, so does the generalized covariation $[X, Y]_t$ for $t \in [s_0, T]$. Take a look at the (i, j)-th component

$$[X, Y]_t^{i,j} = \sum_{l=1}^{n} \left[X^{i,l}, Y^{l,j} \right]_t,$$

where

$$X_t^{i,l} = X_0^{i,l} + \int_{s_0}^{t} A_s^{i,l} \, ds + \sum_{h=1}^{m_1} \int_{s_0}^{t} \left(C_s^h \right)^{i,l} \, dw_s^h + \sum_{h=1}^{m_2} \int_{s_0}^{t} \left(D_s^h \right)^{i,l} \, dB_s^h \tag{3.5}$$

and

$$Y_t^{l,j} = Y_0^{l,j} + \int_{s_0}^{t} \hat{A}_s^{l,j} \, ds + \sum_{h=1}^{m_1} \int_{s_0}^{t} \left(\hat{C}_s^h \right)^{l,j} \, dw_s^h + \sum_{h=1}^{m_2} \int_{s_0}^{t} \left(\hat{D}_s^h \right)^{l,j} \, dB_s^h. \tag{3.6}$$

Notice that the Riemann-Stieltjes and Young integral processes are continuous and of bounded p-variation for $p \in [1, 2)$. This implies that they are zero quadratic variation processes by Corollary 3.8. The quadratic variation of the Itô integral process can be calculated by Proposition 3.4 *vii*) and *viii*). $D^{i,l}$ is a.s. continuous and $\mathbb{F}$-adapted and therefore it has a progressively measurable version. B^h is a $\mathbb{F}$-martingale with $\langle B^h, B^h \rangle_t = t$ for all $h = 1, \ldots, m_2$ and $t \in [s_0, T]$. We get

$$\left[\int_{s_0}^{t} \left(D_s^h \right)^{i,l} \, dB_s^h, \int_{s_0}^{t} \left(D_s^h \right)^{i,l} \, dB_s^h \right] = \int_{s_0}^{t} \left(\left(D_s^h \right)^{i,l} \right)^2 \, ds$$

and analogously

$$\left[\int_{s_0}^t \left(\hat{D}_s^h\right)^{l,j} dB_s^h, \int_{s_0}^t \left(\hat{D}_s^h\right)^{l,j} dB_s^h\right] = \int_{s_0}^t \left(\left(\hat{D}_s^h\right)^{l,j}\right)^2 ds,$$

therefore are the Itô integral processes, finite quadratic variation processes. Now taking the bilinearity of the generalized covariation into account we need to calculate all the combinations of covariations $[I, J]$, where I is one of the $m_1 + m_2 + 1$ integral processes given in (3.5) and J is one of the $m_1 + m_2 + 1$ integral processes given in (3.6). Taking Proposition 3.4 $v)$ into account, the only covariations that do not vanish are those where I and J are Itô integral processes. For $h, h' \in \{1, \dots, m_2\}$ such that $h \neq h'$, the Brownian motions B^h and $B^{h'}$ are independent and so $[B^h, B^{h'}] = 0$ by Proposition 3.4 $ix)$. Then by Proposition 3.4 $viii)$, we have

$$\left[\int_{s_0}^t \left(D_s^h\right)^{i,l} dB_s^h, \int_{s_0}^t \left(\hat{D}_s^h\right)^{l,j} dB_s^{h'}\right] = 0.$$

So the only covariations that are not equal to zero are given by

$$\left[\int_{s_0}^t \left(D_s^h\right)^{i,l} dB_s^h, \int_{s_0}^t \left(\hat{D}_s^h\right)^{l,j} dB_s^h\right] = \int_{s_0}^t \left(D_s^h\right)^{i,l} \left(\hat{D}_s^h\right)^{l,j} ds$$

for all $h = 1, \dots, m_2$. This yields

$$[X, Y]_t^{i,j} = \sum_{l=1}^{n} \sum_{h=1}^{m_2} \int_{s_0}^t \left(D_s^h\right)^{i,l} \left(\hat{D}_s^h\right)^{l,j} ds,$$

and consequently

$$[X, Y]_t = \sum_{j=1}^{m_2} \int_{s_0}^t D_s^j \hat{D}_s^j \, ds.$$

This concludes the proof. □

3.2 Explicit solution to our system of linear differential equations

Now we have all the necessary results to formulate our variation of constants formula. We define the homogenous matrix valued stochastic differential equation with initial time $s_0 \in [0, T]$

$$\begin{aligned} \Phi_t^{s_0} = I_{n_1+n_2} &+ \int_{s_0}^t \begin{pmatrix} b_x^u(r) & 0 \\ \hat{b}_z^u(r) & \hat{b}_x^u(r) \end{pmatrix} \Phi_r^{s_0} \, dr + \sum_{j=1}^{m_1} \int_{s_0}^t \begin{pmatrix} \sigma_x^{u,j}(r) & 0 \\ 0 & 0 \end{pmatrix} \Phi_r^{s_0} \, dw_r^j \\ &+ \sum_{j=1}^{m_2} \int_{s_0}^t \begin{pmatrix} 0 & 0 \\ \hat{\sigma}_z^{u,j}(r) & \hat{\sigma}_x^{u,j}(r) \end{pmatrix} \Phi_r^{s_0} \, dB_r^j, \end{aligned} \tag{3.7}$$

with $\Phi_t^{s_0} \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}$ for $t \in [s_0, T]$. The solution to this equation is given by the matrix

$$\Phi_t^{s_0} = \begin{pmatrix} \phi_t^{s_0} & 0 \\ \tilde{\phi}_t^{s_0} & \hat{\phi}_t^{s_0} \end{pmatrix},$$

where for every $l \geq 1$

i) $\phi_\cdot^{s_0} \in L^l_{\mathbb{F}}(\Omega, C^p([s_0, T]), \mathbb{R}^{n_1\times n_1})$ is the unique solution to the homogenous linear SDE

$$\phi_t^{s_0} = I_{n_1} + \int_{s_0}^t b_x^u(r)\phi_r^{s_0}\, dr + \sum_{j=1}^{m_1} \int_{s_0}^t \sigma_x^{u,j}(r)\phi_r^{s_0}\, dw_r^j, \tag{3.8}$$

which exists by Corollary 2.36 and Corollary 2.37.

ii) $\hat{\phi}_\cdot^{s_0} \in L^l_{\mathbb{F}}(\Omega, C([s_0, T]), \mathbb{R}^{n_2\times n_2})$ is the unique solution to the homogenous linear SDE

$$\hat{\phi}_t^{s_0} = I_{n_2} + \int_{s_0}^t \hat{b}_x^u(r)\hat{\phi}_r^{s_0}\, dr + \sum_{j=1}^{m_2} \int_{s_0}^t \hat{\sigma}_x^{u,j}(r)\hat{\phi}_r^{s_0}\, dB_r^j,$$

which exists by Theorem 2.45.

iii) $\tilde{\phi}_\cdot^{s_0} \in L^l_{\mathbb{F}}(\Omega, C([s_0, T]), \mathbb{R}^{n_2\times n_1})$ is the unique solution to the inhomogenous linear SDE

$$\tilde{\phi}_t^{s_0} = \int_{s_0}^t \hat{b}_x^u(r)\tilde{\phi}_r + \hat{b}_z^u(r)\phi_r\, dr + \sum_{j=1}^{m_2} \int_{s_0}^t \hat{\sigma}_x^{u,j}(r)\tilde{\phi}_r + \hat{\sigma}_z^{u,j}(r)\phi_r\, dB_r^j,$$

which exists by Theorem 2.40. Hence, $\tilde{\phi}$ can be expressed using the solutions ϕ and $\hat{\phi}$ of the homogenous linear SDEs similar to Lemma 2.45, by

$$\tilde{\phi}_t = \hat{\phi}_t \int_{s_0}^t \hat{\phi}_r^{-1} \left[\hat{b}_z^u(r) - \sum_{j=1}^{m_2} \hat{\sigma}_x^{u,j}(r)\hat{\sigma}_z^{u,j}(r) \right] \phi_r\, dr + \sum_{j=1}^{m_2} \hat{\phi}_t \int_{s_0}^t \hat{\phi}_r^{-1}\hat{\sigma}_z^{u,j}(r)\phi_r\, dB_r^j.$$

Similar to the proof of Lemma 2.44, we can find a bound $D_{\tilde{\varphi},l}$ for $\mathrm{E}\left[\|\tilde{\phi}\|^l_{\infty,s_0,T}\right]$ for every $l \geq 1$ which is independent of the parameter u and s_0.

Hence equation (3.7) has a unique solution Φ^u, which is an element of $L^l_{\mathbb{F}}(\Omega, C([s_0, T]), \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)})$ for every $l \geq 1$ and there exists a constant $D_{\Phi,l}$ independent of u, such that

$$\sup_{s_0\in[0,T]} \mathrm{E}\left[\|\Phi^u\|^l_{\infty,s_0,T}\right] \leq D_{\Phi,l}.$$

For the desired explicit solution of equation (2.65), we need the inverse matrix to Φ_t. Therefore we define the matrix valued stochastic differential equation

$$\begin{aligned}\Psi_t^{s_0} = I_{n_1+n_2} - \int_{s_0}^t \Psi_r^{s_0} \left[\begin{pmatrix} b_x^u(r) & 0 \\ \hat{b}_z^u(r) & \hat{b}_x^u(r) \end{pmatrix} - \sum_{j=1}^{m_2} \begin{pmatrix} 0 & 0 \\ \hat{\sigma}_z^u(r) & \hat{\sigma}_x^{u,j}(r) \end{pmatrix}^2 \right] dr \\ - \sum_{j=1}^{m_1} \int_{s_0}^t \Psi_r^{s_0} \begin{pmatrix} \sigma_x^{u,j}(r) & 0 \\ 0 & 0 \end{pmatrix} dw_r^j - \sum_{j=1}^{m_2} \int_{s_0}^t \Psi_r^{s_0} \begin{pmatrix} 0 & 0 \\ \hat{\sigma}_z^{u,j}(r) & \hat{\sigma}_x^{u,j}(r) \end{pmatrix} dB_r^j, \qquad (3.9)\end{aligned}$$

with $\Psi_t^{s_0} \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}$ for $t \in [s_0, T]$. The solution to this equation is given by the matrix

$$\Psi_t^{s_0} = \begin{pmatrix} \psi_t^{s_0} & 0 \\ \tilde{\psi}_t^{s_0} & \hat{\psi}_t^{s_0} \end{pmatrix},$$

where

i) $\psi_\cdot^{s_0} \in L_{\mathbb{F}}^l(\Omega, C^p([s_0,T]), \mathbb{R}^{n_1\times n_1})$ is the unique solution to the homogenous linear SDE

$$\psi_t^{s_0} = I_{n_1} - \int_{s_0}^t \psi_r^{s_0} b_x^u(r)\, dr - \sum_{j=1}^{m_1} \int_{s_0}^t \psi_r^{s_0} \sigma_x^{u,j}(r)\, dw_r^j, \qquad (3.10)$$

which exists by Corollary 2.36 and Corollary 2.37. Furthermore, by Lemma 2.32, we have $\psi_t^{s_0} = (\phi_t^{s_0})^{-1}$ for $t \in [s_0, T]$, P-almost surely.

ii) $\hat{\psi}_\cdot^{s_0} \in L_{\mathbb{F}}^l(\Omega, C([s_0,T]), \mathbb{R}^{n_2\times n_2})$ is the unique solution to the homogenous linear SDE

$$\hat{\psi}_t^{s_0} = I_{n_2} - \int_{s_0}^t \hat{\psi}_r^{s_0} \left[\hat{b}_x^u(r) - \sum_{j=1}^{m_2} (\hat{\sigma}_x^{u,j}(r))^2 \right] dt - \sum_{j=1}^{m_2} \int_{s_0}^t \hat{\psi}_r^{s_0} \hat{\sigma}_x^{u,j}(r) dB_r^j, \qquad (3.11)$$

which exists by Theorem 2.45. Furthermore, we have $\hat{\psi}_t^{s_0} = (\hat{\phi}_t^{s_0})^{-1}$ for $t \in [s_0, T]$, P-almost surely.

iii) $\tilde{\psi}_\cdot^{s_0} \in L_{\mathbb{F}}^l(\Omega, C([s_0,T]), \mathbb{R}^{n_2\times n_1})$ is the solution to the inhomogenous linear SDE

$$\begin{aligned}\tilde{\psi}_t^{s_0} = - \int_{s_0}^t \tilde{\psi}_r^{s_0} \hat{b}_x^u(r) + \hat{\psi}_r^{s_0} \left[\hat{b}_z^u(r) - \sum_{j=1}^{m_2} \hat{\sigma}_x^{u,j}(r) \hat{\sigma}_z^{u,j}(r) \right] dr \\ - \sum_{j=1}^{m_1} \int_{s_0}^t \tilde{\psi}_r^{s_0} \sigma_x^{u,j}(r)\, dw_r^j - \sum_{j=1}^{m_2} \int_{s_0}^t \hat{\psi}_r^{s_0} \hat{\sigma}_z^{u,j}(r) dB_r^j. \qquad (3.12)\end{aligned}$$

We will prove this in the following Proposition. We defined equation (3.9) analogously to the variation of constants formulas, we established in Lemma 2.33 and Lemma 2.33. So we hope that the relation $\Psi_t^{s_0} = (\Phi_t^{s_0})^{-1}$ for $t \in [s_0, T]$ holds P-almost surely. Since we already found that $\psi_t^{s_0} = (\phi_t^{s_0})^{-1}$ for $t \in [s_0, T]$, P-almost surely and $\hat{\psi}_t^{s_0} = (\hat{\phi}_t^{s_0})^{-1}$ for $t \in [s_0, T]$, P-almost surely, we get a clear candidate for the solution of equation (3.12), which would ensure that $\Psi_t \Phi_t = I_{n_1+n_2}$ for $t \in [s_0, T]$, P-almost surely. It is given by

$$\begin{aligned}\tilde{\psi}_t^{s_0} &= -(\hat{\phi}_t^{s_0})^{-1} \tilde{\phi}_t^{s_0} (\phi_t^{s_0})^{-1} \\ &= \Bigg(- \int_{s_0}^t (\hat{\phi}_r^{s_0})^{-1} \left[\hat{b}_z^u(r) - \sum_{j=1}^{m_2} \hat{\sigma}_x^{u,j}(r) \hat{\sigma}_z^{u,j}(r) \right] \phi_r^{s_0} \, dr \\ &\quad - \sum_{j=1}^{m_2} \int_{s_0}^t (\hat{\phi}_r^{s_0})^{-1} \hat{\sigma}_z^{u,j}(r) \phi_r^{s_0} \, dB_r^j \Bigg) (\phi_t^{s_0})^{-1}.\end{aligned}$$

Since we have a clear candidate for the solution of equation (3.12), we just need some kind of Itô formula to prove the assertion. Note that the candidate is a product of a process driven by Brownian motion and a process of finite p-variation, with $p \in (1, 2)$, such that the standard Itô rule cannot be used. Therefore we will use Theorem 3.13.

Proposition 3.14. *The solution to equation* (3.12) *is given by*

$$\tilde{\psi}_t^{s_0} = \left(- \int_{s_0}^t (\hat{\phi}_r^{s_0})^{-1} \left[\hat{b}_z^u(r) - \sum_{j=1}^{m_2} \hat{\sigma}_x^{u,j}(r) \hat{\sigma}_z^{u,j}(r) \right] \phi_r^{s_0} \, dr - \sum_{j=1}^{m_2} \int_{s_0}^t (\hat{\phi}_r^{s_0})^{-1} \hat{\sigma}_z^{u,j}(r) \phi_r^{s_0} \, dB_r^j \right) (\phi_t^{s_0})^{-1},$$

where $(\hat{\phi}^{s_0}) - 1$, ϕ^{s_0}, $(\phi^{s_0})^{-1}$ *are defined by* (3.11), (3.8), (3.10).

Proof. For readability we leave out the dependence of the processes on s_0. Define for $t \in [s_0, T]$

$$\rho_t = - \int_{s_0}^t \hat{\phi}_r^{-1} \left[\hat{b}_z^u(r) - \sum_{j=1}^{m_2} \hat{\sigma}_x^{u,j}(r) \hat{\sigma}_z^{u,j}(r) \right] \phi_r \, dr - \sum_{j=1}^{m_2} \int_{s_0}^t \hat{\phi}_r^{-1} \hat{\sigma}_z^{u,j}(r) \phi_r \, dB_r^j,$$

such that $\tilde{\psi}_t = \rho_t \phi_t^{-1}$, where $\rho_t \in \mathbb{R}^{n_2 \times n_1}$ and $\phi_t^{-1} \in \mathbb{R}^{n_1 \times n_1}$. We have

$$\phi_t^{-1} = I_{n_1} - \int_{s_0}^t \phi_r^{-1} b_x^u(r) \, dr - \sum_{j=1}^{m_1} \int_{s_0}^t \phi_r^{-1} \sigma_x^{u,j}(r) \, dw_r^j.$$

To use Theorem 3.13, we define the processes

$$\begin{aligned}A_t &= -\hat{\phi}_t^{-1} \left[\hat{b}_z^u(t) - \sum_{j=1}^{m_2} \hat{\sigma}_x^{u,j}(t) \hat{\sigma}_z^{u,j}(t) \right] \phi_t \in \mathbb{R}^{n_2 \times n_1} \\ C_t^j &= 0 \in \mathbb{R}^{n_2 \times n_1}, \; j = 1, \dots, m_1 \\ D_t^j &= -\hat{\phi}_t^{-1} \hat{\sigma}_z^{u,j}(t) \phi_t \in \mathbb{R}^{n_2 \times n_1}, \; j = 1, \dots, m_2 \\ \hat{A}_t &= -\phi_t^{-1} b_x^u(t) \in \mathbb{R}^{n_1 \times n_1}\end{aligned}$$

$$\hat{C}_t^j = -\phi_t^{-1}\sigma_x^{u,j}(t) \in \mathbb{R}^{n_1\times n_1},\ j=1,\dots,m_1$$
$$\hat{D}_t^j = 0 \in \mathbb{R}^{n_1\times n_1},\ j=1,\dots,m_2.$$

Then we have $A, D^j \in L^l_{\mathbb{F}}(\Omega, C[s_0,T], \mathbb{R}^{n_2\times n_1})$ for $j=1,\dots,m_2$, $\hat{A}, \hat{D}^j \in L^l_{\mathbb{F}}(\Omega, C[s_0,T], \mathbb{R}^{n_1\times n_1})$ for $j=1,\dots,m_2$, $C^j \in L^l_{\mathbb{F}}(\Omega, C^p[s_0,T], \mathbb{R}^{n_2\times n_1})$ for $j=1,\dots,m_1$ and $\hat{C}^j \in L^l_{\mathbb{F}}(\Omega, C^p[s_0,T], \mathbb{R}^{n_1\times n_1})$ for $j=1,\dots,m_1$. So all the conditions of Theorem 3.13 are satisfied and we have

$$\begin{aligned}\rho_t\phi_t^{-1} = &-\int_{s_0}^t \rho_r\phi_r^{-1}b_x^u(r)\,dr - \sum_{j=1}^m \int_{s_0}^t \rho_r\phi_r^{-1}\sigma_x^{u,j}(r)\,dw_r^j\\ &-\int_{s_0}^t \hat{\phi}_r^{-1}\left[\hat{b}_z^u(r) - \sum_{j=1}^{m_2}\hat{\sigma}_x^{u,j}(r)\hat{\sigma}_z^{u,j}(r)\right]dr - \sum_{j=1}^{m_2}\int_{s_0}^t \hat{\phi}_r^{-1}\hat{\sigma}_z^{u,j}(r)\,dB_r^j,\end{aligned}$$

which concludes the proof. □

The solution to equation (3.12) is also unique. Since equation (3.11) has a unique solution, suppose $\tilde{\psi}^1$ and $\tilde{\psi}^2$ are two solutions to equation (3.12), then $z = \tilde{\psi}^1 - \tilde{\psi}^2$ satisfies the equation

$$z_t = -\int_{s_0}^t z_r\hat{b}_x^u(r)\,dr - \sum_{j=1}^{m_1}\int_{s_0}^t z_r\sigma_x^{u,j}(r)\,dw_r^j,$$

which has a unique solution by Theorem 2.22, which then has to be 0 for all $t \in [s_0,T]$. The boundedness of $\mathrm{E}\left[\|\tilde{\psi}^{s_0}\|^l_{\infty,s_0,T}\right]$, independent of the parameter and s_0 for every $l \geq 1$, follows by the representation

$$\tilde{\psi}_t^{s_0} = -(\hat{\phi}_t^{s_0})^{-1}\tilde{\phi}_t^{s_0}(\phi_t^{s_0})^{-1}$$

and the boundedness of all moments (uniform in time) of the processes on the right hand side, independent of u and s_0. This yields that equation (3.9) has a unique solution Ψ^{s_0} which is an element of $L^l_{\mathbb{F}}(\Omega, C([s_0,T]), \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)})$ for every $l \geq 1$ and there exists a constant $D_{\Psi,l}$ independent of u, such that

$$\sup_{s_0\in[0,T]} \mathrm{E}\left[\|\Psi^u\|^l_{\infty,s_0,T}\right] \leq D_{\Psi,l}.$$

It is easy to see that $\Psi_t^{s_0} = (\Phi_t^{s_0})^{-1}$ for $t \in [s_0,T]$, P-almost surely. Similar to the ODE or Itô-SDE case we show that it is possible to express the solution to the inhomogenous linear SDE (2.65) using the matrix valued processes $\Phi = \Phi^0$ and $\Phi^{-1} = (\Phi^0)^{-1}$ with initial time $s_0 = 0$.

Theorem 3.15. *Let Φ_t resp. Φ_t^{-1} be the solutions to the homogenous linear SDE* (3.7) *resp.* (3.9) *with initial time $s_0 = 0$. The unique solution to the inhomogenous linear SDE* (2.65) *for a given $u \in \mathcal{U}$ is given by*

$$\mathcal{Y}_t^u = \Phi_t\begin{pmatrix} D\xi_0(u)\\ Dx_0(u)\end{pmatrix} + \Phi_t\int_0^t \Phi_r^{-1}\Big[\begin{pmatrix} b_u^u(r)\\ \hat{b}_u^u(r)\end{pmatrix} - \sum_{j=1}^{m_2}\begin{pmatrix} 0\\ \hat{\sigma}_x^{u,j}(r)\hat{\sigma}_u^{u,j}(r)\end{pmatrix}\Big]\,dr$$

$$+\sum_{j=1}^{m_1}\Phi_t\int_0^t \Phi_r^{-1}\begin{pmatrix}\sigma_u^{u,j}(r)\\ 0\end{pmatrix} dw_r^j+\sum_{j=1}^{m_2}\Phi_t\int_0^t \Phi_r^{-1}\begin{pmatrix}0\\ \hat\sigma_u^{u,j}(r)\end{pmatrix} dB_r^j.$$

Proof. We proof the statement by applying Theorem 3.13 on the product $\Phi_t^{-1}\mathcal{Y}_t$. For this define the processes

$$\begin{aligned}
A_t &= -\Phi_t^{-1}\left[\begin{pmatrix}b_x^u(t) & 0\\ \hat b_z^u(t) & \hat b_x^u(t)\end{pmatrix}-\sum_{j=1}^{m_2}\begin{pmatrix}0 & 0\\ \hat\sigma_z^u(t) & \hat\sigma_x^{u,j}(t)\end{pmatrix}^2\right]\\
C_t^j &= -\Phi_t^{-1}\begin{pmatrix}\sigma_x^{u,j}(t) & 0\\ 0 & 0\end{pmatrix},\ j=1,\dots,m_1\\
D_t^j &= -\Phi_t^{-1}\begin{pmatrix}0 & 0\\ \hat\sigma_z^{u,j}(t) & \hat\sigma_x^{u,j}(t)\end{pmatrix},\ j=1,\dots,m_2\\
\hat A_t &= \begin{pmatrix}b_x^u(t) & 0\\ \hat b_z^u(t) & \hat b_x^u(t)\end{pmatrix}\mathcal{Y}_t^{u,\bar u}+\begin{pmatrix}b_u^u(t)\\ \hat b_u^u(t)\end{pmatrix}\bar u\\
\hat C_t^j &= \begin{pmatrix}\sigma_x^{u,j}(t) & 0\\ 0 & 0\end{pmatrix}\mathcal{Y}_t^{u,\bar u}+\begin{pmatrix}\sigma_u^{u,j}(t)\\ 0\end{pmatrix}\bar u,\ j=1,\dots,m_1\\
\hat D_t^j &= \begin{pmatrix}0 & 0\\ \hat\sigma_z^u(t) & \hat\sigma_x^{u,j}(t)\end{pmatrix}\mathcal{Y}_t^{u,\bar u}+\begin{pmatrix}0\\ \hat\sigma_u^{u,j}(t)\end{pmatrix}\bar u,\ j=1,\dots,m_2.
\end{aligned}$$

Then we have $A,D^j\in L^l_{\mathbb{F}}(\Omega,C[0,T],\mathbb{R}^{n_2\times n_1})$ for $j=1,\dots,m_2$, $\hat A,\hat D^j\in L^l_{\mathbb{F}}(\Omega,C[0,T],\mathbb{R}^{n_1\times n_1})$ for $j=1,\dots,m_2$, for $q>2$ such that $\frac{1}{p}+\frac{1}{q}>1$, $C^j\in L^l_{\mathbb{F}}(\Omega,C^q[0,T],\mathbb{R}^{n_2\times n_1})$ for $j=1,\dots,m_1$ and $\hat C^j\in L^l_{\mathbb{F}}(\Omega,C^q[0,T],\mathbb{R}^{n_1\times n_1})$ for $j=1,\dots,m_1$. So all the conditions of Theorem 3.13 are satisfied and we have

$$\begin{aligned}
\Phi_t^{-1}\mathcal{Y}_t &= \begin{pmatrix}D\xi_0(u)\\ Dx_0(u)\end{pmatrix}+\int_0^t\Phi_r^{-1}\begin{pmatrix}b_x^u(r) & 0\\ \hat b_z^u(r) & \hat b_x^u(r)\end{pmatrix}\mathcal{Y}_r\,dr+\int_0^t\Phi_r^{-1}\begin{pmatrix}b_u^u(r)\\ \hat b_u^u(r)\end{pmatrix}dr\\
&+\sum_{j=1}^{m_1}\int_0^t\Phi_r^{-1}\begin{pmatrix}\sigma_x^{u,j}(r) & 0\\ 0 & 0\end{pmatrix}\mathcal{Y}_r\,dw_r^j+\sum_{j=1}^{m_1}\int_0^t\Phi_r^{-1}\begin{pmatrix}\sigma_u^{u,j}(r)\\ 0\end{pmatrix}dw_r^j\\
&+\sum_{j=1}^{m_2}\int_0^t\Phi_r^{-1}\begin{pmatrix}0 & 0\\ \hat\sigma_z^u(r) & \hat\sigma_x^{u,j}(r)\end{pmatrix}\mathcal{Y}_r\,dB_r^j+\sum_{j=1}^{m_2}\int_0^t\Phi_r^{-1}\begin{pmatrix}0\\ \hat\sigma_u^{u,j}(r)\end{pmatrix}dB_r^j\\
&-\int_0^t\Phi_r^{-1}\left[\begin{pmatrix}b_x^u(r) & 0\\ \hat b_z^u(r) & \hat b_x^u(r)\end{pmatrix}-\sum_{j=1}^{m_2}\begin{pmatrix}0 & 0\\ \hat\sigma_z^u(r) & \hat\sigma_x^{u,j}(r)\end{pmatrix}^2\right]\mathcal{Y}_r\,dr\\
&-\sum_{j=1}^{m_1}\int_0^t\Phi_r^{-1}\begin{pmatrix}\sigma_x^{u,j}(r) & 0\\ 0 & 0\end{pmatrix}\mathcal{Y}_r\,dw_r^j-\sum_{j=1}^{m_2}\int_0^t\Phi_t^{-1}\begin{pmatrix}0 & 0\\ \hat\sigma_z^{u,j}(r) & \hat\sigma_x^{u,j}(r)\end{pmatrix}\mathcal{Y}_r\,dB_r^j\\
&-\sum_{j=1}^{m_2}\int_0^t\Phi_r^{-1}\begin{pmatrix}0 & 0\\ \hat\sigma_z^{u,j}(r) & \hat\sigma_x^{u,j}(r)\end{pmatrix}^2\mathcal{Y}_r\,dr-\sum_{j=1}^{m_2}\int_0^t\Phi_r^{-1}\begin{pmatrix}0\\ \hat\sigma_x^{u,j}(r)\hat\sigma_u^{u,j}(r)\end{pmatrix}dr
\end{aligned}$$

$$= \begin{pmatrix} D\xi_0(u) \\ Dx_0(u) \end{pmatrix} + \int_0^t \Phi_t^{-1} \left[\begin{pmatrix} b_u^u(r) \\ \hat{b}_u^u(r) \end{pmatrix} - \sum_{j=1}^{m_2} \begin{pmatrix} 0 \\ \hat{\sigma}_x^{u,j}(r)\hat{\sigma}_u^{u,j}(r) \end{pmatrix} \right] dr$$
$$+ \sum_{j=1}^{m_1} \int_0^t \Phi_t^{-1} \begin{pmatrix} \sigma_u^{u,j}(r) \\ 0 \end{pmatrix} dw_r^j + \sum_{j=1}^{m_2} \int_0^t \Phi_t^{-1} \begin{pmatrix} 0 \\ \hat{\sigma}_u^{u,j}(r) \end{pmatrix} dB_r^j.$$

Multiplying both sides of the last equation with Φ_t yields the assertion. □

Hence the map $u \mapsto \mathcal{X}_\cdot^u$ from $\mathbb{R}^d \to L_{\mathbb{F}}^l(\Omega, C[0,T], \mathbb{R}^{n_1+n_2})$ is Fréchet differentiable with derivative

$$D\mathcal{X}_t^u = \Phi_t \begin{pmatrix} D\xi_0(u) \\ Dx_0(u) \end{pmatrix} + \Phi_t \int_0^t \Phi_t^{-1} \left[\begin{pmatrix} b_u^u(r) \\ \hat{b}_u^u(r) \end{pmatrix} - \sum_{j=1}^{m_2} \begin{pmatrix} 0 \\ \hat{\sigma}_x^{u,j}(r)\hat{\sigma}_u^{u,j}(r) \end{pmatrix} \right] dr$$
$$+ \sum_{j=1}^{m_1} \Phi_t \int_0^t \Phi_t^{-1} \begin{pmatrix} \sigma_u^{u,j}(r) \\ 0 \end{pmatrix} dw_r^j + \sum_{j=1}^{m_2} \Phi_t \int_0^t \Phi_t^{-1} \begin{pmatrix} 0 \\ \hat{\sigma}_u^{u,j}(r) \end{pmatrix} dB_r^j \tag{3.13}$$

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

3.3 The gradient of the cost function and the adjoint equation

Now we come the the analysis of our cost function

$$J : \mathcal{U} \to \mathbb{R}, \ (u) \mapsto \frac{1}{2} \sum_{\mu=1}^{M} \mathrm{E}[g_\mu(\mathcal{X}_{T_\mu}^u)]^2$$

and its gradient. We first establish a standard result, which ensures that we can use the chain rule of Fréchet derivatives.

Lemma 3.16. *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable map with bounded derivative. The map*

$$f : L^2(\Omega, \mathbb{R}^{(n_1+n_2)}) \to \mathbb{R}, \ z \mapsto \frac{1}{2} E[g(z)]^2$$

is Fréchet differentiable for every $z \in L^2(\Omega, \mathbb{R}^{(n_1+n_2)})$ and we have

$$Df(z)h = \mathrm{E}[g(z)]\mathrm{E}[g'(z)h]$$

for every $h \in L^2(\Omega, \mathbb{R}^{(n_1+n_2)})$, where $g'(z)$ is an $n_1 + n_2$-dimensional row vector.

Proof. We need to show that

$$\lim_{\|h\|_{L^2} \to 0} \frac{|f(z+h) - f(z) - Df(z)h|}{\|h\|_{L^2}} = 0.$$

We take a look at the term

$$|f(z+h)-f(z)-Df(z)h| = \left|\frac{1}{2}\mathrm{E}[g(z+h)]^2 - \frac{1}{2}\mathrm{E}[g(z)]^2 - Df(z)h\right|.$$

For simplicity we leave out the index μ for the rest of the proof. With the Taylor expansion for functions between Banach spaces (see e.g. Ambrosetti and Prodi [1995], Chapter 1.4) applied on g, we obtain

$$g(z+h) = g(z) + g'(z)h + \int_0^1 (g'(z+rh) - g'(z))\, dr\, h = g(z) + g'(z)h + r(h)h.$$

By the continuity and boundedness of g', the Hölder inequality and the dominated convergence theorem, we obtain

$$\begin{aligned}|f(z+h)-f(z)-Df(z)h| &= \left|\frac{1}{2}\mathrm{E}[g(z)+g'(z)h+r(h)h]^2 - \frac{1}{2}|\mathrm{E}[g(z)]|^2 - df(z)h\right| \\ &= \left|\frac{1}{2}\mathrm{E}[g'(z)h]^2 + \frac{1}{2}\mathrm{E}[r(h)h]^2 + \mathrm{E}[g(z)]\mathrm{E}[r(h)h] + \mathrm{E}[g'(z)h]\mathrm{E}[r(h)h]\right| \\ &= o(\|h\|_{L^2}),\end{aligned}$$

which yields the desired limit. □

One easy way to calculate the gradient of the cost function would be to use the solution of the sensitivity equation $\mathcal{Y}$ (see (2.65)), which is the Fréchet differential of the solution mapping $u \mapsto \mathcal{X}^u$. Using the Chain rule for Fréchet derivatives (see Ambrosetti and Prodi [1995] Proposition 1.1.4), we obtain the gradient

$$\nabla J(u) = \sum_{\mu=1}^{M} E[g_\mu(\mathcal{X}^u_{T_\mu})]\mathrm{E}[g'_\mu(\mathcal{X}^u_{T_\mu})\mathcal{Y}^u_{T_\mu}]. \tag{3.14}$$

To calculate this gradient numerically we will discretize (using first order Euler schemes) the underlying equations and use the Monte-Carlo method for the estimation of the expected value. Focusing on the computational side, the computation of the discretized gradient boils down to numerically evaluating the values of the Euler scheme for $\mathcal{Y}^u$ on a partition $(t_i)_{i=0,\dots,n}$ of $[0,T]$ for every Monte-Carlo path. Since $\mathcal{Y}^u$ takes values in $\mathbb{R}^{(n_1+n_2)\times d}$ this leads to very high computational costs, especially if the number of parameters d is big, e.g. when the parameters are time dependent. The main goal of this thesis is to establish another representation of this gradient that does not involve the the process $\mathcal{Y}$, and whose numerical approximation is way cheaper. We introduce this representation in the next theorem, which is the first main result of this thesis. It involves the explicit solution of an anticipating backwards stochastic differential equation, namely the adjoint equation.

Theorem 3.17. *Let $M, T > 0$ and $(g_\mu)_{\mu=1,\dots,M}$ be a sequence of functions satisfying condition (G). Furthermore let $T_1 \le \dots \le T_M = T$ be a sequence of times in $(0,T]$, $u \in \mathcal{U}$ and let $\mathcal{X}^u_t$ be the unique solution to equation (2.64). The cost function*

$$J : \mathcal{U} \to \mathbb{R},\ u \mapsto \frac{1}{2}\sum_{\mu=1}^{M} \mathrm{E}[g_\mu(\mathcal{X}^u_{T_\mu})]^2$$

is totally differentiable and its gradient is given by

$$\begin{aligned}\nabla J(u) = \mathrm{E}\Bigg[\Lambda_0 \begin{pmatrix} D\xi_0(u) \\ Dx_0(u)\end{pmatrix} + \int_0^T \Lambda_r \Bigg[\begin{pmatrix} b^u_u(r) \\ \hat{b}^u_u(r)\end{pmatrix} - \sum_{j=1}^{m_2}\begin{pmatrix} 0 \\ \hat{\sigma}^{u,j}_x(r)\hat{\sigma}^{u,j}_u(r)\end{pmatrix}\Bigg] dr \\ + \sum_{j=1}^{m_1}\int_0^T \Lambda_r \begin{pmatrix} \sigma^{u,j}_u(r) \\ 0\end{pmatrix} dw^j_r + \sum_{j=1}^{m_2}\int_0^T \Lambda_r \begin{pmatrix} 0 \\ \hat{\sigma}^{u,j}_u(r)\end{pmatrix} d^- B^j_r\Bigg], \qquad (3.15)\end{aligned}$$

where the row vector

$$\Lambda_t = \sum_{T_\mu \ge t} E[g_\mu(\mathcal{X}^u_{T_\mu})] g'_\mu(\mathcal{X}^u_{T_\mu}) \Phi_{T_\mu}\Phi_t^{-1} \text{ for } t \in [0,T]$$

is an element of $L^l(\Omega, C[0,T], \mathbb{R}^{n_1+n_2})$ and satisfies the anticipating BSDE

$$\begin{aligned}\Lambda_t = \sum_{T_\mu \ge t} E[g_\mu(\mathcal{X}^u_{T_\mu})] g'_\mu(\mathcal{X}^u_{T_\mu}) + \int_t^T \Lambda_r \Bigg[\begin{pmatrix} b^u_x(r) & 0 \\ \hat{b}^u_z(r) & \hat{b}^u_x(r)\end{pmatrix} - \sum_{j=1}^{m_2}\begin{pmatrix} 0 & 0 \\ \hat{\sigma}^{u,j}_z(r) & \hat{\sigma}^{u,j}_x(r)\end{pmatrix}^2\Bigg] dr \\ + \sum_{j=1}^{m_1}\int_t^T \Lambda_r \begin{pmatrix} \sigma^{u,j}_x(r) & 0 \\ 0 & 0\end{pmatrix} dw^j_r + \sum_{j=1}^{m_2}\int_t^T \Lambda_r \begin{pmatrix} 0 & 0 \\ \hat{\sigma}^{u,j}_z(r) & \hat{\sigma}^{u,j}_x(r)\end{pmatrix} d^- B^j_r, \qquad (3.16)\end{aligned}$$

for all $t \in [0,T]$, which we call the adjoint equation.

Proof. By the definition of J and Lemma 3.16, we can use the chain rule for Fréchet differentials, which yields

$$\nabla J(u) = \sum_{\mu=1}^{M} E[g_\mu(\mathcal{X}^u_{T_\mu})] \mathrm{E}[g'_\mu(\mathcal{X}^u_{T_\mu}) D\mathcal{X}^u_{T_\mu}].$$

Using (3.13), we get

$$\begin{aligned}\nabla J(u) = \mathrm{E}\Bigg[\sum_{\mu=1}^{M} E[g_\mu(\mathcal{X}^u_{T_\mu})] g'_\mu(\mathcal{X}^u_{T_\mu}) \Phi_{T_\mu} \begin{pmatrix} D\xi_0(u) \\ Dx_0(u)\end{pmatrix} \\ + \sum_{\mu=1}^{M}\int_0^{T_\mu} E[g_\mu(\mathcal{X}^u_{T_\mu})] g'_\mu(\mathcal{X}^u_{T_\mu}) \Phi_{T_\mu}\Phi_r^{-1}\Bigg[\begin{pmatrix} b^u_u(r) \\ \hat{b}^u_u(r)\end{pmatrix} - \sum_{j=1}^{m_2}\begin{pmatrix} 0 \\ \hat{\sigma}^{u,j}_x(r)\hat{\sigma}^{u,j}_u(r)\end{pmatrix}\Bigg] dr\end{aligned}$$

$$
\begin{aligned}
&+\sum_{\mu=1}^{M}\sum_{j=1}^{m_1}\int_0^{T_\mu} E[g_\mu(\mathcal{X}^u_{T_\mu})]g'_\mu(\mathcal{X}^u_{T_\mu})\Phi_{T_\mu}\Phi_r^{-1}\begin{pmatrix}\sigma_u^{u,j}(r)\\ 0\end{pmatrix}dw_r^j\\
&+\sum_{\mu=1}^{M}\sum_{j=1}^{m_2}\int_0^{T_\mu} E[g_\mu(\mathcal{X}^u_{T_\mu})]g'_\mu(\mathcal{X}^u_{T_\mu})\Phi_{T_\mu}\Phi_r^{-1}\begin{pmatrix}0\\ \hat\sigma_u^{u,j}(r)\end{pmatrix}d^-B_r^j\Bigg].
\end{aligned}
$$

Notice that the last integral in the upper equation is no longer an Itô integral, since the intergrand is not adapted to the filtration $\mathbb{F}$. Instead we use the forward integral by Russo and Vallois (see Remark 3.10). Interchanging sums and integrals, we get

$$
\begin{aligned}
\nabla J(u) = \mathrm{E}\Bigg[&\sum_{\mu=1}^{M} E[g_\mu(\mathcal{X}^u_{T_\mu})]g'_\mu(\mathcal{X}^u_{T_\mu})\Phi_{T_\mu}\begin{pmatrix}D\xi_0(u)\\ Dx_0(u)\end{pmatrix}\\
&+\int_0^T\sum_{T_\mu\ge r} E[g_\mu(\mathcal{X}^u_{T_\mu})]g'_\mu(\mathcal{X}^u_{T_\mu})\Phi_{T_\mu}\Phi_r^{-1}\left[\begin{pmatrix}b_u^u(r)\\ \hat b_u^u(r)\end{pmatrix}-\sum_{j=1}^{m_2}\begin{pmatrix}0\\ \hat\sigma_x^{u,j}(r)\hat\sigma_u^{u,j}(r)\end{pmatrix}\right]dr\\
&+\sum_{j=1}^{m_1}\int_0^T\sum_{T_\mu\ge r} E[g_\mu(\mathcal{X}^u_{T_\mu})]g'_\mu(\mathcal{X}^u_{T_\mu})\Phi_{T_\mu}\Phi_r^{-1}\begin{pmatrix}\sigma_u^{u,j}(r)\\ 0\end{pmatrix}dw_r^j\\
&+\sum_{j=1}^{m_2}\int_0^T\sum_{T_\mu\ge r} E[g_\mu(\mathcal{X}^u_{T_\mu})]^\top g'_\mu(\mathcal{X}^u_{T_\mu})\Phi_{T_\mu}\Phi_r^{-1}\begin{pmatrix}0\\ \hat\sigma_u^{u,j}(r)\end{pmatrix}d^-B_r^j\Bigg].
\end{aligned}\tag{3.17}
$$

Now we define the process $\Lambda_\cdot \in L^l(\Omega, C[0,T], \mathbb{R}^{n_1+n_2})$ by

$$
\Lambda_t = \sum_{T_\mu\ge t} E[g_\mu(\mathcal{X}^u_{T_\mu})]g'_\mu(\mathcal{X}^u_{T_\mu})\Phi_{T_\mu}\Phi_t^{-1} \text{ for } t\in[0,T].
$$

By the definition of Φ^{-1} (see (3.9)) and the equality

$$
\begin{aligned}
\Phi_t^{-1} = \Phi_{T_\mu}^{-1} &+ \int_t^{T_\mu}\Phi_r^{-1}\left[\begin{pmatrix}b_x^u(r) & 0\\ \hat b_z^u(r) & \hat b_x^u(r)\end{pmatrix}-\sum_{j=1}^{m_2}\begin{pmatrix}0 & 0\\ \hat\sigma_z^u(r) & \hat\sigma_x^{u,j}(r)\end{pmatrix}^2\right]dr\\
&+\sum_{j=1}^{m_1}\int_t^{T_\mu}\Phi_r^{-1}\begin{pmatrix}\sigma_x^{u,j}(r) & 0\\ 0 & 0\end{pmatrix}dw_r^j+\sum_{j=1}^{m_2}\int_t^{T_\mu}\Phi_r^{-1}\begin{pmatrix}0 & 0\\ \hat\sigma_z^{u,j}(r) & \hat\sigma_x^{u,j}(r)\end{pmatrix}dB_r^j
\end{aligned}
$$

for all $t\le T_\mu$, we can argue that for every $t\in[0,T]$, we have

$$
\begin{aligned}
&\sum_{T_\mu\ge t} E[g_\mu(\mathcal{X}^u_{T_\mu})]g'_\mu(\mathcal{X}^u_{T_\mu})\Phi_{T_\mu}\Phi_t^{-1}\\
&=\sum_{T_\mu\ge t} E[g_\mu(\mathcal{X}^u_{T_\mu})]g'_\mu(\mathcal{X}^u_{T_\mu})\\
&\quad+\int_t^T\sum_{T_\mu\ge r} E[g_\mu(\mathcal{X}^u_{T_\mu})]g'_\mu(\mathcal{X}^u_{T_\mu})\Phi_{T_\mu}\Phi_r^{-1}\left[\begin{pmatrix}b_x^u(r) & 0\\ \hat b_z^u(r) & \hat b_x^u(r)\end{pmatrix}-\sum_{j=1}^{m_2}\begin{pmatrix}0 & 0\\ \hat\sigma_z^u(r) & \hat\sigma_x^{u,j}(r)\end{pmatrix}^2\right]dr
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{j=1}^{m_1}\int_t^T \sum_{T_\mu\geq r} E[g_\mu(\mathcal{X}^u_{T_\mu})]g'_\mu(\mathcal{X}^u_{T_\mu})\Phi_{T_\mu}\Phi_r^{-1}\begin{pmatrix}\sigma_x^{u,j}(r) & 0\\ 0 & 0\end{pmatrix} dw_r^j\\
&+\sum_{j=1}^{m_2}\int_t^T \sum_{T_\mu\geq r} E[g_\mu(\mathcal{X}^u_{T_\mu})]g'_\mu(\mathcal{X}^u_{T_\mu})\Phi_{T_\mu}\Phi_r^{-1}\begin{pmatrix}0 & 0\\ \hat{\sigma}_z^{u,j}(r) & \hat{\sigma}_x^{u,j}(r)\end{pmatrix} d^-B_r^j,
\end{aligned}
$$

by again interchanging sums and integrals. Hence, we see that Λ satisfies the anticipating BSDE

$$
\begin{aligned}
\Lambda_t = &\sum_{T_\mu\geq t} E[g_\mu(\mathcal{X}^u_{T_\mu})]g'_\mu(\mathcal{X}^u_{T_\mu}) + \int_t^T \Lambda_r\left[\begin{pmatrix}b_x^u(r) & 0\\ \hat{b}_z^u(r) & \hat{b}_x^u(r)\end{pmatrix} - \sum_{j=1}^{m_2}\begin{pmatrix}0 & 0\\ \hat{\sigma}_z^{u,j}(r) & \hat{\sigma}_x^{u,j}(r)\end{pmatrix}^2\right] dr\\
&+\sum_{j=1}^{m_1}\int_t^T \Lambda_r\begin{pmatrix}\sigma_x^{u,j}(r) & 0\\ 0 & 0\end{pmatrix} dw_r^j + \sum_{j=1}^{m_2}\int_t^T \Lambda_r\begin{pmatrix}0 & 0\\ \hat{\sigma}_z^{u,j}(r) & \hat{\sigma}_x^{u,j}(r)\end{pmatrix} d^-B_r^j.
\end{aligned}
$$

Using the definition of Λ in (3.17) yields the assertion. □

If we now take a look at the computational side, the backwards equation is $\mathbb{R}^{n_1+n_2}$ valued, in comparison to the $\mathbb{R}^{(n_1+n_2)\times d}$ valued process $\mathcal{Y}$, hence the numerical computation of λ will be significantly faster if we have a high amount of parameters. We have shown that our cost function is totally differentiable and established two representations of the gradient. The following chapter will now focus on the approximation of the cost function and its gradient to use the theoretical results in practice.

Chapter 4

Approximation of the cost function and its gradient

In this chapter we discuss the numerical approximation of the solution to the model dynamics equation $\mathcal{X}^u$ given in (2.64), the sensitivity equation $\mathcal{Y}^u$ given in (2.65), the solution Λ of equation (3.16) and consequently the cost function (3.1), together with the sensitivity and adjoint representation of its gradient. Since we need additional assumptions on the coefficients to get the corresponding convergence rates, we first state the standing assumptions for this chapter. Let for the rest of this chapter $(\Omega, \mathcal{F}, \mathbb{F}, P)$ be a filtered probability space (satisfying the usual assumptions) carrying a m_1-dimensional continuous, bounded p-variation ($p \in (1,2)$) process w, which satisfies the exponential moment condition (2.48) and a m_2-dimensional standard Brownian motion B, both adapted to the filtration $\mathbb{F}$. Let $\mathcal{U}$ be a bounded, open and convex subset of $\mathbb{R}^d$ and $T > 0$ be a positive constants. Furthermore the functions

$$\xi_0 : \mathcal{U} \to \mathbb{R}^{n_1}, \quad b : [0,T] \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_1}, \quad \sigma : [0,T] \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_1 \times m_1}$$

satisfy the conditions $(H_1), (H_2), (H_3^*)$, where (H_3^*) is the same condition as (H_3) but with the Hölder exponent β is now chosen to be in $[\frac{1}{p}, 1]$ instead of $[\frac{1}{2}, 1]$, and the time Hölder condition

(E_1) Let b, L be the function and constant from condition (H_2) and $\beta \in [\frac{1}{p}, 1]$ be the same constant as in condition (H_3^*). For every $x \in \mathbb{R}^{n_1}$, $u \in \mathcal{U}$ and $s \leq t \in [0,T]$, b satisfies

$$|b(t,x,u) - b(s,x,u)| + |b_x(t,x,u) - b_x(s,x,u)| + |b_u(t,x,u) - b_u(s,x,u)| \leq L|t-s|^\beta.$$

The functions

$$x_0 : \mathcal{U} \to \mathbb{R}^{n_2}, \quad \hat{b} : [0,T] \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_2}, \quad \hat{\sigma} : [0,T] \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_1} \times \mathcal{U} \to \mathbb{R}^{n_2 \times m_2}$$

satisfy the conditions $(B_1), (B_2), (B_3)$ and the time Hölder condition

(E_2): Let $\hat{b}$, $\hat{\sigma}$ and L be the coefficient functions and the constant from condition (B_1) respectively (B_2). For all $x \in \mathbb{R}^{n_2}$, $y \in \mathbb{R}^{n_1}$, $u \in \mathcal{U}$ and $s \leq t \in [0,T]$, $\hat{b}$ and $\hat{\sigma}$ satisfy

$$|\hat{b}(t,x,y,u) - \hat{b}(s,x,y,u)| + |\hat{\sigma}(t,x,y,u) - \hat{\sigma}(s,x,y,u)| \leq L(1+|x|+|y|)(t-s)^{\frac{1}{2}}.$$

Under these conditions all the results from the previous chapters hold. Since we do not assume any kind of Hölder condition on w, we cannot expect to get a convergence parameter which only depends on the mesh $|\Pi^{\mathrm{E}}| = \max_{i=0,\ldots,n-1} |t_{i+1} - t_i|$ of the Euler partition, as it is in standard approximation schemes of Itô SDEs. We define three convergence parameters. First, for all $\omega \in \Omega$, we define the parameter

$$\delta(\omega) := \max_{i=0,\ldots,n-1} |t_{i+1} - t_i| + |w(\omega)|_{p,t_i,t_{i+1}} \tag{4.1}$$

for the pathwise convergence of the stochastic Young differential equations. Second, we define the L^l-convergence parameter for the stochastic Young differential equation

$$\delta_{1,l} := \mathrm{E}\left[\delta^l\right]^{\frac{1}{l}}, \tag{4.2}$$

which is well defined, since w satisfies the exponential moment condition (2.48). And the last convergence parameter, which we will use in the estimation of the convergence rate of the Itô stochastic differential equation, is defined by

$$\delta_2 := \max_{i=0,\ldots,n-1} |t_{i+1} - t_i|.$$

In the following remark, we summarize some of the results of the previous chapters to facilitate the notation for the proofs to come.

Remark 4.1. Taking Corollary 2.36, Corollary 2.37, Lemma 2.41, Lemma 2.44 and Remark 2.47 into account, we know that there exist random variables C_ξ and C_y which are independent of u, such that for almost every $\omega \in \Omega$

$$\begin{aligned} \|\xi^u(\omega)\|_{\infty,0,T} &\leq \|\xi^u(\omega)\|_{p,0,T} \leq C_\xi(\omega) \\ \|y^u(\omega)\|_{\infty,0,T} &\leq \|y^u(\omega)\|_{p,0,T} \leq C_y(\omega) \end{aligned}$$

and for every $l \geq 1$ there exist positive constants $D_{\xi,l}$, $D_{x,l}$, $D_{y,l}$, $D_{\hat{y},l}$, $D_{\mathcal{X},l}$, $D_{\mathcal{Y},l}$ which are independent of u such that

$$\begin{aligned} \mathrm{E}\left[\|\xi^u\|^l_{\infty,0,T}\right] &\leq \mathrm{E}\left[\|\xi^u\|^l_{p,0,T}\right] \leq D_{\xi,l} \\ \mathrm{E}\left[\|y^u\|^l_{\infty,0,T}\right] &\leq \mathrm{E}\left[\|y^u\|^l_{p,0,T}\right] \leq D_{y,l} \end{aligned}$$

$$\begin{aligned}
\mathrm{E}\left[\|x^u\|^l_{\infty,0,T}\right] &\le D_{x,l}\\
\mathrm{E}\left[\|\hat{y}^u\|^l_{\infty,0,T}\right] &\le D_{\hat{y},l}\\
\mathrm{E}\left[\|\mathcal{X}^u\|^l_{\infty,0,T}\right] &\le D_{\mathcal{X},l}\\
\mathrm{E}\left[\|\mathcal{Y}^u\|^l_{\infty,0,T}\right] &\le D_{\mathcal{Y},l}.
\end{aligned}$$

For the remainder of this chapter, we postulate the aforementioned notations and assumptions.

4.1 Convergence of the Euler schemes for the forward equations

We recall the forward equations which are of interest to us. Namely, the model dynamics equation

$$\begin{aligned}
\mathcal{X}^u_t = \begin{pmatrix}\xi^u_t\\ x^u_t\end{pmatrix} &= \begin{pmatrix}\xi_0(u)\\ x_0(u)\end{pmatrix} + \int_0^t \begin{pmatrix} b\left(r,\xi^u_r,u\right)\\ \hat{b}\left(r,x^u_r,\xi^u_r,u\right)\end{pmatrix} dr + \sum_{j=1}^{m_1}\int_0^t \begin{pmatrix}\sigma^j\left(r,\xi^u_r,u\right)\\ 0\end{pmatrix} dw^j_r\\
&\quad + \sum_{j=1}^{m_2}\int_0^t \begin{pmatrix} 0\\ \hat{\sigma}^j\left(r,x^u_r,\xi^u_r,u\right)\end{pmatrix} dB^j_r \qquad (4.3)
\end{aligned}$$

and the sensitivity equation

$$\begin{aligned}
\mathcal{Y}^u_t &= \begin{pmatrix} y^u_t\\ \hat{y}^u_t\end{pmatrix}\\
&= \begin{pmatrix} D\xi_0(u)\\ Dx_0(u)\end{pmatrix} + \int_0^t \begin{pmatrix} b^u_x(r) & 0\\ \hat{b}^u_z(r) & \hat{b}^u_x(r)\end{pmatrix}\mathcal{Y}^u_r + \begin{pmatrix} b^u_u(r)\\ \hat{b}^u_u(r)\end{pmatrix} dr\\
&\quad + \sum_{j=1}^{m_1}\int_0^t \begin{pmatrix}\sigma^{u,j}_x(r) & 0\\ 0 & 0\end{pmatrix}\mathcal{Y}^u_r + \begin{pmatrix}\sigma^{u,j}_u(r)\\ 0\end{pmatrix} dw^j_r\\
&\quad + \sum_{j=1}^{m_2}\int_0^t \begin{pmatrix} 0 & 0\\ \hat{\sigma}^{u,j}_z(r) & \hat{\sigma}^{u,j}_x(r)\end{pmatrix}\mathcal{Y}^u_r + \begin{pmatrix} 0\\ \hat{\sigma}^{u,j}_u(r)\end{pmatrix} dB^j_r, \qquad (4.4)
\end{aligned}$$

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

$$\begin{aligned}
&b^u_x(t) = b_x\left(t,\xi^u_t,u\right),\ b^u_u(t) = b_u\left(t,\xi^u_t,u\right)\\
&\sigma^{u,j}_x(t) = \sigma^j_x\left(t,\xi^u_t,u\right),\ \sigma^{u,j}_u(t) = \sigma^j_u\left(t,\xi^u_t,u\right)\ \text{ for } j=1,\dots,m_1\\
&\hat{b}^u_x(t) = \hat{b}_x\left(t,x^u_t,\xi^u_t,u\right),\ \hat{b}^u_z(t) = \hat{b}_z\left(t,x^u_t,\xi^u_t,u\right),\ \hat{b}^u_u(t) = \hat{b}_u\left(t,x^u_t,\xi^u_t,u\right)\\
&\hat{\sigma}^{u,j}_x(t) = \hat{\sigma}^j_x\left(t,x^u_t,\xi^u_t,u\right),\ \hat{\sigma}^{u,j}_z(t) = \hat{\sigma}^j_z\left(t,x^u_t,\xi^u_t,u\right),\ \hat{\sigma}^{u,j}_u(t) = \hat{\sigma}^j_u\left(t,x^u_t,\xi^u_t,u\right)\ \text{ for } j=1,\dots,m_2.
\end{aligned}$$

We define their respective discrete Euler schemes on $(\Omega,\mathcal{F},\mathbb{F},P)$ for a partition $\Pi^{\text{Euler}} = \Pi^{\text{E}} = (t_i)_{i=0,\dots,n}$ of $[0,T]$. For each $\omega\in\Omega$, $u\in\mathcal{U}$ and $i\in\{0,\dots,n-1\}$, we define the discrete Euler scheme $\mathcal{X}^n$ for the equation (4.3) by

$$
\begin{aligned}
\mathcal{X}^n_{t_{i+1}}(\omega) &= \begin{pmatrix} \xi^n_{t_{i+1}}(\omega) \\ x^n_{t_{i+1}}(\omega) \end{pmatrix} \\
&= \begin{pmatrix} \xi^n_{t_i}(\omega) \\ x^n_{t_i}(\omega) \end{pmatrix} + \begin{pmatrix} b\left(t_i, \xi^n_{t_i}(\omega), u\right) \\ \hat{b}\left(t_i, x^n_{t_i}(\omega), \xi^n_{t_i}(\omega), u\right) \end{pmatrix} (t_{i+1} - t_i) \\
&\quad + \sum_{j=1}^{m_1} \begin{pmatrix} \sigma^j\left(t_i, \xi^n_{t_i}(\omega), u\right) \\ 0 \end{pmatrix} \left(w^j_{t_{i+1}}(\omega) - w^j_{t_i}(\omega)\right) \\
&\quad + \sum_{j=1}^{m_2} \begin{pmatrix} 0 \\ \hat{\sigma}^j\left(t_i, x^n_{t_i}(\omega), \xi^n_{t_i}(\omega), u\right) \end{pmatrix} \left(B^j_{t_{i+1}}(\omega) - B^j_{t_i}(\omega)\right)
\end{aligned}
$$

and

$$
\mathcal{X}^n_{t_0} = \mathcal{X}^u_0 = (\xi_0(u), x_0(u))^\top.
$$

Here we left out the direct dependence of the discrete processes ξ^n, x^n and $\mathcal{X}^n$ on u for readability. We consider the continuous interpolation

$$
\begin{aligned}
\mathcal{X}^n_t(\omega) &= \begin{pmatrix} \xi^n_t(\omega) \\ x^n_t(\omega) \end{pmatrix} \\
&= \begin{pmatrix} \xi^n_{t_i}(\omega) \\ x^n_{t_i}(\omega) \end{pmatrix} + \begin{pmatrix} b\left(t_i, \xi^n_{t_i}(\omega), u\right) \\ \hat{b}\left(t_i, x^n_{t_i}(\omega), \xi^n_{t_i}(\omega), u\right) \end{pmatrix} (t - t_i) \\
&\quad + \sum_{j=1}^{m_1} \begin{pmatrix} \sigma^j\left(t_i, \xi^n_{t_i}(\omega), u\right) \\ 0 \end{pmatrix} \left(w^j_t(\omega) - w^j_{t_i}(\omega)\right) \\
&\quad + \sum_{j=1}^{m_2} \begin{pmatrix} 0 \\ \hat{\sigma}^j\left(t_i, x^{n,u}_{t_i}(\omega), \xi^n_{t_i}(\omega), u\right) \end{pmatrix} \left(B^j_t(\omega) - B^j_{t_i}(\omega)\right)
\end{aligned} \tag{4.5}
$$

for $t \in [t_i, t_{i+1}]$ for every $i \in \{0, \dots, n-1\}$. Similarly for each $\omega \in \Omega$, $u \in \mathcal{U}$ and $i \in \{0, \dots, n-1\}$, we define the discrete Euler scheme $\mathcal{Y}^n$ for the equation (4.4) by

$$
\begin{aligned}
\mathcal{Y}^n_{t_{i+1}} &= \begin{pmatrix} y^n_{t_{i+1}} \\ \hat{y}^n_{t_{i+1}} \end{pmatrix} \\
&= \mathcal{Y}^n_{t_i} + \left(\begin{pmatrix} b_x\left(t_i, \xi^n_{t_i}, u\right) & 0 \\ \hat{b}_z\left(t_i, x^n_{t_i}, \xi^n_{t_i}, u\right) & \hat{b}_x\left(t_i, x^n_{t_i}, \xi^n_{t_i}, u\right) \end{pmatrix} \mathcal{Y}^n_{t_i} + \begin{pmatrix} b_u\left(t_i, \xi^n_{t_i}, u\right) \\ \hat{b}_u\left(t_i, x^n_{t_i}, \xi^n_{t_i}, u\right) \end{pmatrix} \right) (t_{i+1} - t_i) \\
&\quad + \sum_{j=1}^{m_1} \left(\begin{pmatrix} \sigma^j_x\left(t_i, \xi^n_{t_i}, u\right) & 0 \\ 0 & 0 \end{pmatrix} \mathcal{Y}^n_{t_i} + \begin{pmatrix} \sigma^j_u(t_i, \xi^n_{t_i}, u) \\ 0 \end{pmatrix} \right) \left(w^j_{t_{i+1}} - w^j_{t_i}\right) \\
&\quad + \sum_{j=1}^{m_2} \begin{pmatrix} 0 & 0 \\ \hat{\sigma}^j_z\left(t_i, x^n_{t_i}, \xi^n_{t_i}, u\right) & \hat{\sigma}^j_x\left(t_i, x^n_{t_i}, \xi^n_{t_i}, u\right) \end{pmatrix} \mathcal{Y}^n_{t_i} \left(B^j_{t_{i+1}} - B^j_{t_i}\right)
\end{aligned}
$$

$$+\sum_{j=1}^{m_2}\begin{pmatrix}0\\ \hat{\sigma}_u^j\left(t_i,x_{t_i}^n,\xi_{t_i}^n,u\right)\end{pmatrix}\left(B_{t_{i+1}}^j-B_{t_i}^j\right),$$

with

$$\mathcal{Y}_{t_0}^n(\omega)=\mathcal{Y}_0^u=(D\xi_0(u),Dx_0(u))^\top$$

and its continuous interpolation

$$\begin{aligned}
\mathcal{Y}_t^n&=\begin{pmatrix}y_t^n\\ \hat{y}_t^n\end{pmatrix}\\
&=\mathcal{Y}_{t_i}^n+\left(\begin{pmatrix}b_x\left(t_i,\xi_{t_i}^n,u\right) & 0\\ \hat{b}_z\left(t_i,x_{t_i}^n,\xi_{t_i}^n,u\right) & \hat{b}_x\left(t_i,x_{t_i}^n,\xi_{t_i}^n,u\right)\end{pmatrix}\mathcal{Y}_{t_i}^n+\begin{pmatrix}b_u\left(t_i,\xi_{t_i}^n,u\right)\\ \hat{b}_u\left(t_i,x_{t_i}^n,\xi_{t_i}^n,u\right)\end{pmatrix}\right)(t-t_i)\\
&\quad+\sum_{j=1}^{m_1}\left(\begin{pmatrix}\sigma_x^j\left(t_i,\xi_{t_i}^n,u\right) & 0\\ 0 & 0\end{pmatrix}\mathcal{Y}_{t_i}^n+\begin{pmatrix}\sigma_u^j\left(t_i,\xi_{t_i}^n,u\right)\\ 0\end{pmatrix}\right)\left(w_t^j-w_{t_i}^j\right)\\
&\quad+\sum_{j=1}^{m_2}\begin{pmatrix}0 & 0\\ \hat{\sigma}_z^j\left(t_i,x_{t_i}^n,\xi_{t_i}^n,u\right) & \hat{\sigma}_x^j\left(t_i,x_{t_i}^n,\xi_{t_i}^n,u\right)\end{pmatrix}\mathcal{Y}_{t_i}^n\left(B_t^j-B_{t_i}^j\right)\\
&\quad+\sum_{j=1}^{m_2}\begin{pmatrix}0\\ \hat{\sigma}_u^j\left(t_i,x_{t_i}^n,\xi_{t_i}^n,u\right)\end{pmatrix}\left(B_t^j-B_{t_i}^j\right)
\end{aligned}\tag{4.6}$$

for $t\in[t_i,t_{i+1}]$ for every $i\in\{0,\dots,n-1\}$. Again we leave out the direct dependencies of the processes on u and ω for readability. The goal of this section will be to establish the strong convergence

$$\lim_{n\to\infty}\mathrm{E}\left[\|\mathcal{X}-\mathcal{X}^n\|_{\infty,0,T}^l\right]^{\frac{1}{l}}=0,\quad \lim_{n\to\infty}\mathrm{E}\left[\|\mathcal{Y}-\mathcal{Y}^n\|_{\infty,0,T}^l\right]^{\frac{1}{l}}=0$$

for a sequence of partitions $(\Pi^n)_{n\in\mathbb{N}}$ with $|\Pi^n|\to 0$ for every $l\geq 2$ and find the corresponding convergence rate. In the Itô SDE case, and consequently for the whole system, we will state the boundedness and convergence results in the theorems only for $l\geq 2$, to shorten the proofs. The estimates for $l\in[1,2)$, then follow by the monotonicity of L^l-norms.

4.1.1 Convergence of the Euler scheme for model dynamics equation

Because of the different nature of the stochastic differential equations involved here, we split the convergence analysis of the Euler schemes into two parts. First we examine for a given parameter $u\in\mathcal{U}$ the convergence rate at which $\xi^{n,u}$ in (4.5) converges to ξ^u in (4.3) P-a.s. in the uniform norm and in $L^l(\Omega,C([0,T]),\mathbb{R}^{n_1})$ for $l\geq 1$. The convergence of Euler schemes of a differential equation driven by a process of finite p-variation have been studied by Lejay [2010], Davie [2008], Friz and Victoir [2010] and for the special case of fractional Brownian motion by Mishura [2008], Nourdin [2005], Nourdin and Neuenkirch [2007]. We will basically use the same idea for the calculation of the convergence rate as Lejay [2010], but since the author only considers the deterministic, time autonomous case, we adapt his results to our framework. We will especially be very careful with the ω-dependent constants in the estimates, to get the convergence results in L^l. First, we show that the p-variation of the continuous interpolation $\xi^{n,u}$ is bounded independently of the number n of subintervals of the Partition Π^{E} and the parameter u.

Lemma 4.2. *Let $u \in \mathcal{U}$, we have for almost every $\omega \in \Omega$*

$$|\xi^{n,u}(\omega)|_{p,0,T} \leq 2^{3p-1} C_1^p \left(T^p + |w(\omega)|_{p,0,T}^p\right)$$

and

$$\|\xi^{n,u}(\omega)\|_{\infty,0,T} \leq L + 2^{3p-1} C_1^p \left(T^p + |w(\omega)|_{p,0,T}^p\right) := C_{\xi^n}(\omega),$$

where the constant C_1 is given by (2.23)*. For $l \geq 1$, we have $\xi^{n,u} \in L^l_{\mathbb{F}}(\Omega, C([0,T|), \mathbb{R}^{n_1})$ with*

$$\mathrm{E}\left[\|\xi^{n,u}\|_{\infty,0,T}^l\right] \leq \mathrm{E}\left[C_{\xi^n}^l\right] := D_{\xi^n,l} < \infty.$$

Furthermore, setting $\overline{n}(t) = \sup\{i \in \mathbb{N} | t_i \in \Pi^E$ and $t_i \leq t\}$ for every $t \in [0,T]$, we have for almost every $\omega \in \Omega$

$$\|\xi^{n,u}_{\cdot}(\omega) - \xi^{n,u}_{t_{\overline{n}(\cdot)}}(\omega)\|_{\infty,0,T} \leq \max_{i=0,\dots,n-1} |\xi^{n,u}(\omega)|_{p,t_i,t_{i+1}} \leq L\delta(\omega) \tag{4.7}$$

and consequently

$$E\left[\|\xi^{n,u}_{\cdot} - \xi^{n,u}_{t_{\overline{n}(\cdot)}}\|_{\infty,0,T}^l\right] \leq L^l \delta_{1,l}^l, \tag{4.8}$$

where δ_1 is defined in (4.2)*.*

Proof. Let $\mathcal{A} \subset \Omega$, such that $P(\mathcal{A}) = 0$ and $w_{\cdot}(\omega)$ is continuous and of bounded p-variation ($p \in (1,2)$) for every $\omega \in \mathcal{A}^c$. First, we show that for a given $u \in \mathcal{U}$ and for every $\omega \in \mathcal{A}^c$ the paths $\xi^{n,u}_{\cdot}(\omega)$ are elements of $C^p([0,T], \mathbb{R}^{n_1})$. Let $\omega \in \mathcal{A}^c$ and $u \in \mathcal{U}$ be arbitrary, for notational simplicity we leave out the direct dependence of the involved processes on ω and u. We have that ξ^n satisfies the equation

$$\xi^n_t = \xi^n_0 + \int_0^t b\left(t_{\overline{n}(r)}, \xi^n_{t_{\overline{n}(r)}}, u\right) d_r + \int_0^t \sigma\left(t_{\overline{n}(r)}, \xi^n_{t_{\overline{n}(r)}}, u\right) dw_r.$$

By the conditions on b and σ and since w is continuous, the continuity of ξ^n follows directly. By Lemma 2.9, we can estimate

$$|\xi^n|_{p,0,T}^p \leq n^{p-1} \sum_{i=0}^{n-1} |\xi^n|_{p,t_i,t_{i+1}}^p.$$

For a given $i \in \{0, \dots, n-1\}$ and $s < t \in [t_i, t_{i+1}]$, we have

$$|\xi^n_t - \xi^n_s| \leq \left|b\left(t_i, \xi^n_{t_i}, u\right)(t-s) + \sigma\left(t_i, \xi^n_{t_i}, u\right)(w_t - w_s)\right| \leq L\left(|t-s| + |w|_{p,s,t}\right), \tag{4.9}$$

which yields by Lemma 2.6

$$|\xi^n|_{p,t_i,t_{i+1}} \le L(|t_{i+1} - t_i| + |w|_{p,t_i,t_{i+1}}). \tag{4.10}$$

Hence,

$$\begin{aligned}
|\xi^n|^p_{p,0,T} &\le n^{p-1} \sum_{i=0}^{n-1} L^p(|t_{i+1} - t_i| + |w|_{p,t_i,t_{i+1}})^p \\
&\le (2n)^{p-1} L^p \sum_{i=0}^{n-1} (t_{i+1} - t_i)^p + |w|^p_{p,t_i,t_{i+1}} \\
&\le (2n)^{p-1} L^p (T^p + |w|^p_{p,0,T}).
\end{aligned}$$

Since $w \in C^p\left([0,T], \mathbb{R}^{m_1}\right)$, we have $\xi^n \in C^p\left([0,T], \mathbb{R}^{n_1}\right)$. By Lemma 2.26, we know that $\sigma\left(\cdot, \xi^n_\cdot, u\right)$ is an element of $C^q\left([0,T], \mathbb{R}^{n_1 \times m_1}\right)$ for $q \in (2, \frac{p}{p-1})$. Notice that $\sigma\left(\cdot, \xi^n_\cdot, u\right)$ coincides on the partition points of Π^{E} with the function $\sigma\left(t_{\overline{n}(\cdot)}, \xi^n_{t_{\overline{n}(\cdot)}}, u\right)$, such that we can use the inequalities from Lemma 2.12. Our goal is to find an upper bound of ξ^n, which is independent of u and n. Let t_i be a partition point of Π^{E} for $i \in \{0, \dots, n-1\}$ and $s < t \in [t_i, t_{i+1}]$. We have by (4.9) that

$$|\xi^n_t - \xi^n_s| \le L(|t - s| + |w|_{p,s,t}). \tag{4.11}$$

Now let $0 \le t_{l-1} < s < t_l < t_{l+1} < \cdots < t_{l+m} = t_k < t \le t_{k+1} \le T$ for $m \ge 0$ and $t_{l-1}, \dots, t_{k+1} \in \Pi^{\mathrm{E}}$, we estimate

$$|\xi^n_t - \xi^n_s| \le |\xi^n_t - \xi^n_{t_k}| + |\xi^n_{t_k} - \xi^n_{t_l}| + |\xi^n_{t_l} - \xi^n_s|,$$

where the second term vanishes for $m = 0$. Using (4.11), we obtain

$$|\xi^n_t - \xi^n_s| \le L(|t - t_k| + |w|_{p,t_k,t}) + |\xi^n_{t_k} - \xi^n_{t_l}| + L(|t_l - s| + |w|_{p,s,t_l}).$$

For $m \ge 1$ the term $|\xi^n_{t_k} - \xi^n_{t_l}|$ can be decomposed by

$$|\xi^n_{t_k} - \xi^n_{t_l}| \le \left| \sum_{i=l}^{k-1} b\left(t_i, \xi^n_{t_i}, u\right)(t_{i+1} - t_i) \right| + \left| \sum_{i=l}^{k-1} \sigma\left(t_i, \xi^n_{t_i}, u\right)(w_{t_{i+1}} - w_{t_i}) \right| \le S_1 + S_2.$$

The sum S_1 can easily be estimated by the boundedness of b and the superadditivity of $\varphi(s,t) = |t - s|$ on $\Delta([0,T])$

$$S_1 \le \sum_{i=l}^{k-1} \left\| b\left(t_{\overline{n}(\cdot)}, \xi^n_{t_{\overline{n}(\cdot)}}, u\right) \right\|_{\infty,t_l,t_k} (t_{i+1} - t_i) \le L|t_k - t_l|.$$

Since $\sigma(\cdot, \xi^n_\cdot, u) \in C^q([0,T], \mathbb{R}^{n_1})$ for $q \in (2, \frac{p}{p-1})$, we can apply Lemma 2.12 and obtain

$$S_2 \leq C_{p,q} \left\| \sigma(\cdot, \xi^n_\cdot, u) \right\|_{q,t_l,t_k} |w|_{p,t_l,t_k}.$$

This yields by Lemma 2.26 i) and condition (H_3^*), that

$$\begin{aligned} S_2 &\leq C_{p,q} \left(L + L \left(T^\beta + |\xi^n|_{p,t_l,t_k} \right) \right) |w|_{p,t_l,t_k} \\ &\leq C_1 (1 + |\xi^n|_{p,t_l,t_k}) |w|_{p,t_l,t_k}, \end{aligned}$$

where $C_1 := 2 \max \left\{ L, C_{p,q} L, C_{p,q} L T^\beta, 1 \right\}$ analogue to (2.23). Hence,

$$|\xi^n_{t_k} - \xi^n_{t_l}| \leq C_1 (1 + |\xi^n|_{p,t_l,t_k})(|t_k - t_l| + |w|_{p,t_l,t_k}).$$

Putting all terms together, we obtain

$$\begin{aligned} |\xi^n_t - \xi^n_s| &\leq L(|t - t_k| + |w|_{p,t_k,t}) + L(|t_l - s| + |w|_{p,s,t_l}) + C_1(1 + |\xi^n|_{p,t_l,t_k})(|t_k - t_l| + |w|_{p,t_l,t_k}) \\ &\leq 2L(|t - s| + |w|_{p,s,t}) + C_1(1 + |\xi^n|_{p,s,t})(|t - s| + |w|_{p,s,t}) \\ &\leq 2C_1(1 + |\xi^n|_{p,s,t})(|t - s| + |w|_{p,s,t}). \end{aligned} \tag{4.12}$$

With (4.11) and (4.12), we get for every $[r, v] \subset [s, t] \subset [0, T]$

$$|\xi^n_v - \xi^n_r| \leq 2C_1(1 + |\xi^n|_{p,s,t})(|v - r| + |w|_{p,r,v}).$$

By Lemma 2.6, this yields

$$|\xi^n|_{p,s,t} \leq 2C_1(1 + |\xi^n|_{p,s,t})(|t - s| + |w|_{p,s,t}).$$

Now we have for every interval $[s, t] \in [0, T]$ which satisfies

$$|t - s| + |w|_{p,s,t} \leq \frac{1}{4C_1}$$

that

$$|\xi^n|_{p,s,t} \leq 1.$$

By our Gronwall-type lemma 2.20, this yields

$$|\xi^n|_{p,0,T} = \leq 2^{3p-1} C_1^p \left(T^p + |w|^p_{p,0,T} \right).$$

We conclude

$$\begin{aligned}\|\xi^n\|_{\infty,0,T} &\le |\xi_0(u)| + 2^{3p-1}C_1^p\left(T^p + |w|_{p,0,T}^p\right)\\ &\le L + 2^{3p-1}C_1^p\left(T^p + |w|_{p,0,T}^p\right).\end{aligned}$$

Since ω was arbitrary in $\mathcal{A}^c$ and u was arbitrary in $\mathcal{U}$, the inequalities for the p-variation and uniform norm of ξ^n hold P-almost surely and for all $u \in \mathcal{U}$. The $\mathbb{F}$-adaptedness of ξ^n is a direct implication of its definition and the $\mathbb{F}$-adaptedness of w. Since w satisfies the exponential moment condition (2.48), we have

$$\mathrm{E}\left[\|\xi^n\|_{\infty,0,T}^l\right] \le \mathrm{E}\left[\left(L + 2^{3p-1}C_1^p\left(T^p + |w|_{p,0,T}^p\right)\right)^l\right] := D_{\xi^n,l} < \infty.$$

Since $|\xi_t^n - \xi_{t_i}^n| \le |\xi^n|_{p,t_i,t_{i+1}}$ for $t \in [t_i, t_{i+1}]$ and $i \in \{0, \dots, n-1\}$, we have by (4.10) P-a.s.

$$\|\xi_\cdot^n - \xi_{t_{\overline{n}(\cdot)}}^n\|_{\infty,0,T}^l \le L^l\delta^l,$$

and consequently

$$E\left[\|\xi_\cdot^n - \xi_{t_{\overline{n}(\cdot)}}^n\|_{\infty,0,T}^l\right] \le L^l\delta_{1,l}^l.$$

□

Remark 4.3. Notice that in the situation of Lemma 4.2, we have for every $u \in \mathcal{U}$, $\omega \in \mathcal{A}^c$ and every interval $[s,t] \in [0,T]$ which satisfies

$$|t-s| + |w(\omega)|_{p,s,t} \le \frac{1}{4C_1}$$

that

$$|\xi^n(\omega)|_{p,s,t} \le 1.$$

The constant C_1 is defined in (2.23).

In the next theorem we give the convergence rate of the Euler scheme for the stochastic Young differential equation.

Theorem 4.4. *Let $u \in \mathcal{U}$, we have for almost every $\omega \in \Omega$*

$$\|\xi^u(\omega) - \xi^{n,u}(\omega)\|_{\infty,0,T} \le |\xi^u(\omega) - \xi^{n,u}(\omega)|_{p,0,T} \le K_\xi(\omega)\delta(\omega)^{2-p}, \tag{4.13}$$

where the random variable K_ξ has moments of all orders and is independent of n and u. Furthermore, we have for $l \ge 1$

$$E\left[\|\xi^u - \xi^{n,u}\|_{\infty,0,T}^l\right]^{\frac{1}{l}} \le \mathrm{E}\left[K_\xi^{2l}\right]^{\frac{1}{2l}}\delta_{1,2l}^{2-p} =: D_{K_\xi,2l}\delta_{1,2l}^{2-p}.$$

Proof. Let $\mathcal{A} \subset \Omega$, such that $P(\mathcal{A}) = 0$ and $w.(\omega)$ is continuous and of bounded p-variation for every $\omega \in \mathcal{A}^c$. Let $\omega \in \mathcal{A}^c$ and $u \in \mathcal{U}$ be arbitrary, for notational simplicity leave out the direct dependence of the involved processes on ω and u. Let $s \leq t \in [t_i, t_{i+1}]$ for some $i \in \{0, \dots, n-1\}$ and define $\gamma_t = \xi_t - \xi_t^n$ for all $t \in [0, T]$. We have

$$\xi_t^n - \xi_s^n = b\left(t_i, \xi_{t_i}^n, u\right)(t-s) + \sigma\left(t_i, \xi_{t_i}^n, u\right)(w_t - w_s)$$

and

$$\begin{aligned}
\xi_t - \xi_s &= \int_s^t b(r, \xi_r, u)\, dr + \int_s^t \sigma(r, \xi_r, u)\, dw_r \\
&= \int_s^t b(r, \xi_r, u) - b(s, \xi_s, u)\, dr + b(s, \xi_s, u)(t-s) \\
&\quad + \int_s^t \sigma(r, \xi_r, u) - \sigma(s, \xi_s, u)\, dw_r + \sigma(s, \xi_s, u)(w_t - w_s).
\end{aligned}$$

This yields

$$\begin{aligned}
\gamma_t - \gamma_s &= \xi_t - \xi_s - (\xi_t^n - \xi_s^n) \\
&= \left(b(s, \xi_s, u) - b\left(t_i, \xi_{t_i}^n, u\right)\right)(t-s) + \left(\sigma(s, \xi_s, u) - \sigma\left(t_i, \xi_{t_i}^n, u\right)\right)(w_t - w_s) \\
&\quad + \int_s^t b(r, \xi_r, u) - b(s, \xi_s, u)\, dr + \int_s^t \sigma(r, \xi_r, u) - \sigma(s, \xi_s, u)\, dw_r \\
&= \left(b(s, \xi_s, u) - b\left(s, \xi_s^n, u\right)\right)(t-s) + \left(b\left(s, \xi_s^n, u\right) - b\left(t_i, \xi_{t_i}^n, u\right)\right)(t-s) \\
&\quad + \left(\sigma(s, \xi_s, u) - \sigma\left(s, \xi_s^n, u\right)\right)(w_t - w_s) + \left(\sigma\left(s, \xi_s^n, u\right) - \sigma\left(t_i, \xi_{t_i}^n, u\right)\right)(w_t - w_s) \\
&\quad + \int_s^t b(r, \xi_r, u) - b(s, \xi_s, u)\, dr + \int_s^t \sigma(r, \xi_r, u) - \sigma(s, \xi_s, u)\, dw_r.
\end{aligned}$$

Using the conditions (H_2), (H_3^*), (E_1), the Love-Young estimate for $q \in (2, \frac{p}{p-1})$ and Lemma 2.26, we obtain

$$\begin{aligned}
|\gamma_t - \gamma_s| &\leq L|\gamma_s|(|t-s| + |w|_{p,s,t}) + L\left(|s-t_i|^\beta + |\xi^n|_{p,t_i,s}\right)(|t-s| + |w|_{p,s,t}) \\
&\quad + L(|t-s|^\beta + |\xi|_{p,s,t})|t-s| + C_{p,q}|\sigma(\cdot, \xi_\cdot, u)|_{q,s,t}|w|_{p,s,t} \\
&\leq L|\gamma_s|(|t-s| + |w|_{p,s,t}) + L\left(|t_{i+1}-t_i|^\beta + |\xi^n|_{p,t_i,t_{i+1}}\right)(|t-s| + |w|_{p,s,t}) \\
&\quad + L(|t-s|^\beta + |\xi|_{p,s,t})|t-s| + C_{p,q}L\left(|t-s|^\beta + |\xi|_{p,s,t}\right)|w|_{p,s,t} \\
&\leq L|\gamma_s|(|t-s| + |w|_{p,s,t}) + C_1\left(|t_{i+1}-t_i|^\beta + |\xi|_{p,t_i,t_{i+1}} + |\xi^n|_{p,t_i,t_{i+1}}\right)(|t-s| + |w|_{p,s,t}),
\end{aligned}$$

where the constant C_1 is given in (2.23). This can be estimated by (2.26) and (4.10)

$|\gamma_t - \gamma_s|$

$$
\begin{aligned}
&\le L|\gamma_s|(|t-s|+|w|_{p,s,t})+C_1\bigg((|t_{i+1}-t_i|+|w|_{p,t_i,t_{i+1}})^\beta+C_1(1+|\xi|_{p,t_i,t_{i+1}})(|t_{i+1}-t_i|+|w|_{p,t_i,t_{i+1}})\\
&\quad+L(|t_{i+1}-t_i|+|w|_{p,t_i,t_{i+1}})\bigg)(|t-s|+|w|_{p,s,t})\\
&\le L|\gamma_s|(|t-s|+|w|_{p,s,t})+C_1^2\left(\delta^\beta+(1+|\xi|_{p,0,T})\delta+\delta\right)(|t-s|+|w|_{p,s,t})\\
&\le L(|\gamma_s|+D_1(\delta))(|t-s|+|w|_{p,s,t}),
\end{aligned}
\tag{4.14}
$$

where $D_1(\delta)=C_1^2\left(\delta^\beta+(1+C_\xi)\delta+\delta\right)$. Now let $0\le t_{l-1}\le s<t_l<\cdots<t_{l+m}=t_k<t\le t_{k+1}\le T$ with $m\ge 0$. Then we have

$$
|\gamma_t-\gamma_s|\le|\gamma_t-\gamma_{t_k}|+|\gamma_{t_k}-\gamma_{t_l}|+|\gamma_{t_l}-\gamma_s|. \tag{4.15}
$$

The first and third term can be estimated by the previous considerations, which yields

$$
\begin{aligned}
|\gamma_t-\gamma_{t_k}|&\le L(|\gamma_{t_k}|+D_1(\delta))(|t-t_k|+|w|_{p,t_k,t})\\
&\le L(\|\gamma\|_{\infty,s,t}+D_1(\delta))(|t-s|+|w|_{p,s,t}) &(4.16)\\
|\gamma_{t_l}-\gamma_s|&\le L(|\gamma_s|+D_1(\delta))(|t_l-s|+|w|_{p,s,t_l})\\
&\le L(\|\gamma\|_{\infty,s,t}+D_1(\delta))(|t-s|+|w|_{p,s,t}). &(4.17)
\end{aligned}
$$

For $m\ge 1$ the second term in (4.15) does not vanish and we can estimate

$$
\begin{aligned}
&|\gamma_{t_k}-\gamma_{t_l}|\\
&\le\left|\sum_{i=l}^{k-1}\left(b(t_i,\xi_{t_i},u)-b\left(t_i,\xi^n_{t_i},u\right)\right)(t_{i+1}-t_i)\right|\\
&\quad+\sum_{j=1}^{m_1}\left|\sum_{i=l}^{k-1}\left(\sigma^j(t_i,\xi_{t_i},u)-\sigma^j\left(t_i,\xi^n_{t_i},u\right)\right)(w^j_{t_{i+1}}-w^j_{t_i})\right|\\
&\quad+\left|\sum_{i=l}^{k-1}\left(\int_{t_i}^{t_{i+1}}b(r,\xi_r,u)-b(t_i,\xi_{t_i},u)\,dr+\int_{t_i}^{t_{i+1}}\sigma(r,\xi_r,u)-\sigma(t_i,\xi_{t_i},u)\,dw_r\right)\right|\\
&=I_1+I_2+I_3.
\end{aligned}
$$

The term I_1 can easily be estimated by the Lipschitz continuity of b

$$
\begin{aligned}
I_1&\le\sum_{i=l}^{k-1}\left|b(t_i,\xi_{t_i},u)-b\left(t_i,\xi^n_{t_i},u\right)\right||t_{i+1}-t_i|\\
&\le L\sum_{i=l}^{k-1}|\gamma_{t_i}||t_{i+1}-t_i|\\
&\le L\|\gamma\|_{\infty,t_l,t_k}|t_k-t_l|.
\end{aligned}
\tag{4.18}
$$

We know that $\sigma^j(\cdot, \xi_\cdot, u)$ and $\sigma^j(\cdot, \xi_\cdot^n, u)$ are elements of $C^q([0,T], \mathbb{R}^{n_1 \times m_1})$ by Lemma 2.26 for every $j = 1, \dots, m_1$. Hence, we can use Lemma 2.12 for every $j = 1, \dots, m_1$ with $q \in (2, \frac{p}{p-1})$, which yields

$$\begin{aligned} I_2 &\leq \sum_{j=1}^{m_1} C_{p,q} \bigg(\left|\sigma^j(t_l, \xi_{t_l}, u) - \sigma^j\left(t_l, \xi_{t_l}^n, u\right)\right| + \left|\sigma^j(\cdot, \xi_\cdot, u) - \sigma^j(\cdot, \xi_\cdot^n, u)\right|_{q,t_l,t_k} \bigg) |w^j|_{p,t_l,t_k} \\ &\leq C_{p,q} \sum_{j=1}^{m_1} \Big(L|\gamma_{t_l}| + \left|\sigma^j(\cdot, \xi_\cdot, u) - \sigma^j(\cdot, \xi_\cdot^n, u)\right|_{q,t_l,t_k} \Big) |w^j|_{p,t_l,t_k}. \end{aligned} \tag{4.19}$$

By Lemma 2.26 iii), we have

$$\begin{aligned} &|\sigma^j(\cdot, \xi_\cdot, u) - \sigma^j(\cdot, \xi_\cdot^n, u)|_{q,t_l,t_k} \\ &\leq L\|\gamma\|_{\infty,t_l,t_k} \Big(|t_k - t_l|^\beta + |\xi^n|_{p,t_l,t_k} + |\xi|_{p,t_l,t_k} \Big) + L|\gamma|_{p,t_l,t_k}. \end{aligned} \tag{4.20}$$

Inserting (4.20) into (4.19) yields

$$\begin{aligned} I_2 &\leq C_{p,q} \sum_{j=1}^{m_1} \Big(L|\gamma_{t_l}| + L\|\gamma\|_{\infty,t_l,t_k} \Big(|t_k - t_l|^\beta + |\xi^n|_{p,t_l,t_k} + |\xi|_{p,t_l,t_k} \Big) + L|\gamma|_{p,t_l,t_k} \Big) |w^j|_{p,t_l,t_k} \\ &\leq m_1 C_1 (\|\gamma\|_{\infty,t_l,t_k} + |\gamma|_{p,t_l,t_k})(1 + |\xi^n|_{p,t_l,t_k} + |\xi|_{p,t_l,t_k})|w|_{p,t_l,t_k}, \end{aligned} \tag{4.21}$$

where we used that $\sum_{j=1}^{m_1} |w^j|_{p,s,t} \leq m_1 |w|_{p,s,t}$. The estimation of I_3 will be carried out with the Love-Young estimate, Lemma 2.26 i) and the Lipschitz and Hölder condition of the coefficient function b

$$\begin{aligned} I_3 &\leq \sum_{i=l}^{k-1} \int_{t_i}^{t_{i+1}} |b(r, \xi_r, u) - b(t_i, \xi_{t_i}, u)| \, dr + \sum_{i=l}^{k-1} \left| \int_{t_i}^{t_{i+1}} \sigma(r, \xi_r, u) - \sigma(t_i, \xi_{t_i}, u) \, dw_r \right| \\ &\leq L \sum_{i=l}^{k-1} \Big(|t_{i+1} - t_i|^\beta + \|\xi_\cdot - \xi_{t_i}\|_{\infty,t_i,t_{i+1}} \Big) (t_{i+1} - t_i) \\ &\quad + C_{p,q} L \sum_{i=l}^{k-1} \Big(|t_{i+1} - t_i|^\beta + |\xi|_{p,t_i,t_{i+1}} \Big) |w|_{p,t_i,t_{i+1}} \\ &\leq C_{p,q} L \sum_{i=l}^{k-1} \Big(|t_{i+1} - t_i|^\beta + |\xi|_{p,t_i,t_{i+1}} \Big) (|t_{i+1} - t_i| + |w|_{p,t_i,t_{i+1}}) \\ &\leq C_{p,q} L \sum_{i=l}^{k-1} |t_{i+1} - t_i|^\beta (|t_{i+1} - t_i| + |w|_{p,t_i,t_{i+1}}) + C_{p,q} L \sum_{i=l}^{k-1} |\xi|_{p,t_i,t_{i+1}} (|t_{i+1} - t_i| + |w|_{p,t_i,t_{i+1}}) \\ &:= I_{31} + I_{32}, \end{aligned}$$

where we used that $C_{p,q} \geq 1$. Using the Jensen inequality, we get

$$\begin{aligned}
I_{31} &\leq 2^{1-\frac{1}{p}} C_{p,q} L \sum_{i=l}^{k-1} |t_{i+1}-t_i|^\beta (|t_{i+1}-t_i|^p + |w|^p_{p,t_i,t_{i+1}})^{\frac{1}{p}} \\
&\leq 2^{1-\frac{1}{p}} C_{p,q} L \sum_{i=l}^{k-1} |t_{i+1}-t_i|^\beta (|t_{i+1}-t_i|^p + |w|^p_{p,t_i,t_{i+1}})^{1-\beta} (|t_{i+1}-t_i|^p + |w|^p_{p,t_i,t_{i+1}})^{\frac{1}{p}+\beta-1}.
\end{aligned}$$

The function $\varphi_1(s,t) = |t-s|^\beta (|t-s|^p + |w|^p_{p,s,t})^{1-\beta}$ is superadditive and increasing on $[0,T]$ by Lemma 2.4 and Lemma 2.5, which yields

$$\begin{aligned}
I_{31} &\leq 2^{1-\frac{1}{p}} C_{p,q} L \sum_{i=l}^{k-1} \varphi_1(t_i,t_{i+1}) \delta^{1+(\beta-1)p} \\
&\leq 2^{1-\frac{1}{p}} C_{p,q} L \varphi_1(t_l,t_k) \delta^{1+(\beta-1)p} \\
&\leq C_1 \varphi_1(0,T)^{1-\frac{1}{p}} \delta^{1+(\beta-1)p} \varphi_1(t_l,t_k)^{\frac{1}{p}}, \qquad (4.22)
\end{aligned}$$

where C_1 is defined by (2.23). The p-variation of ξ on an interval $[t_i,t_{i+1}]$ can be estimated with (2.26),

$$|\xi|_{p,t_i,t_{i+1}} \leq C_1(1 + |\xi|_{p,t_i,t_{i+1}})((t_{i+1}-t_i) + |w|_{p,t_i,t_{i+1}}).$$

We obtain

$$\begin{aligned}
I_{32} &\leq C_1 C_{p,q} L \sum_{i=l}^{k-1} (1+|\xi|_{p,t_i,t_{i+1}})(|t_{i+1}-t_i| + |w|_{p,t_i,t_{i+1}})^2 \\
&\leq 2^{2-\frac{2}{p}} C_1 C_{p,q} L (1+|\xi|_{p,0,T}) \sum_{i=l}^{k-1} (|t_{i+1}-t_i|^p + |w|^p_{p,t_i,t_{i+1}})^{\frac{2}{p}} \\
&\leq C_1^2 (1+|\xi|_{p,0,T}) \sum_{i=l}^{k-1} (|t_{i+1}-t_i|^p + |w|^p_{p,t_i,t_{i+1}})^{\frac{2}{p}}.
\end{aligned}$$

Define the control function $\varphi_2(s,t) = |t-s|^p + |w|^p_{p,s,t}$ on $[0,T]$ which yields

$$\begin{aligned}
I_{32} &\leq C_1^2 (1+|\xi|_{p,0,T}) \sum_{i=l}^{k-1} \varphi_2(t_i,t_{i+1})^{\frac{2}{p}} \\
&\leq C_1^2 (1+|\xi|_{p,0,T}) \sum_{i=l}^{k-1} \varphi_2(t_i,t_{i+1}) \varphi_2(t_i,t_{i+1})^{\frac{2}{p}-1} \\
&\leq C_1^2 (1+|\xi|_{p,0,T}) \sum_{i=l}^{k-1} \varphi_2(t_i,t_{i+1}) \left(|t_{i+1}-t_i|^p + |w|^p_{p,t_i,t_{i+1}}\right)^{\frac{2}{p}-1} \\
&\leq C_1^2 (1+|\xi|_{p,0,T}) \varphi_2(t_l,t_k) \delta^{2-p}
\end{aligned}$$

$$\leq C_1^2(1+|\xi|_{p,0,T})\varphi_2(0,T)^{1-\frac{1}{p}}\delta^{2-p}\varphi_2(t_l,t_k)^{\frac{1}{p}}. \tag{4.23}$$

Taking (4.22) and (4.23) into account, we have

$$I_3 \leq C_1^2\Big(\varphi_1(0,T)^{1-\frac{1}{p}}\delta^{1+(\beta-1)p}\varphi_1(t_l,t_k)^{\frac{1}{p}} + (1+|\xi|_{p,0,T})\varphi_2(0,T)^{1-\frac{1}{p}}\delta^{2-p}\varphi_2(t_l,t_k)^{\frac{1}{p}}\Big)$$

and since $\beta \in [\frac{1}{p}, 1]$ implies $2-p \leq 1+(\beta-1)p$, this yields

$$\begin{aligned} I_3 &\leq C_1^2\Big(\varphi_1(0,T)^{1-\frac{1}{p}}\delta^{2-p}\delta^{\beta p-1}\varphi_1(t_l,t_k)^{\frac{1}{p}} + (1+|\xi|_{p,0,T})\varphi_2(0,T)^{1-\frac{1}{p}}\delta^{2-p}\varphi_2(t_l,t_k)^{\frac{1}{p}}\Big) \\ &\leq \delta^{2-p}K_1(\omega)\left(\varphi_1(t_l,t_k)^{\frac{1}{p}} + \varphi_2(t_l,t_k)^{\frac{1}{p}}\right), \end{aligned} \tag{4.24}$$

where the random variables

$$\begin{aligned} \delta &= \max_{i=0,\dots,n-1}|t_{i+1}-t_i| + |w|_{p,t_i,t_{i+1}} \leq 1+T+|w|_{p,0,T} := C_w \qquad (4.25) \\ K_1(\omega) &= C_1^2(1+C_\xi)\left(\varphi_1(0,T)^{1-\frac{1}{p}}C_w^{\beta p-1} + \varphi_2(0,T)^{1-\frac{1}{p}}\right) \end{aligned}$$

have moments of all orders by the definitions of φ_1 and φ_2, the exponential moment condition (2.48) and Remark 4.1. Putting all the terms (4.18), (4.21) and (4.24) together, we obtain

$$\begin{aligned} &|\gamma_{t_k}-\gamma_{t_l}| \\ &\leq L\|\gamma\|_{\infty,t_l,t_k}|t_k-t_l| + m_1C_1(\|\gamma\|_{\infty,t_l,t_k}+|\gamma|_{p,t_l,t_k})(1+|\xi^n|_{p,t_l,t_k}+|\xi|_{p,t_l,t_k})|w|_{p,t_l,t_k} \\ &\quad + \delta^{2-p}K_1(\omega)\left(\varphi_1(t_l,t_k)^{\frac{1}{p}} + \varphi_2(t_l,t_k)^{\frac{1}{p}}\right) \\ &\leq m_1C_1(\|\gamma\|_{\infty,s,t}+|\gamma|_{p,s,t})(1+|\xi^n|_{p,s,t}+|\xi|_{p,s,t})(|t-s|+|w|_{p,s,t}) + A(s,t)^{\frac{1}{p}}, \end{aligned}$$

where

$$A(s,t) = 2^{p-1}\delta^{p(2-p)}K_1^p(\omega)(\varphi_1(s,t)+\varphi_2(s,t))$$

is a control function on $[0,T]$. Together with (4.16) and (4.17), this yields

$$\begin{aligned} &|\gamma_t-\gamma_s| \\ &\leq L(\|\gamma\|_{\infty,s,t}+D_1(\delta))(|t-s|+|w|_{p,s,t}) + L(\|\gamma\|_{\infty,s,t}+D_1(\delta))(|t-s|+|w|_{p,s,t}) \\ &\quad + m_1C_1(\|\gamma\|_{\infty,s,t}+|\gamma|_{p,s,t})(1+|\xi^n|_{p,s,t}+|\xi|_{p,s,t})(|t-s|+|w|_{p,s,t}) + A(s,t)^{\frac{1}{p}} \\ &\leq m_12C_1(D_1(\delta)+\|\gamma\|_{\infty,s,t}+|\gamma|_{p,s,t})(1+|\xi^n|_{p,s,t}+|\xi|_{p,s,t})(|t-s|+|w|_{p,s,t}) + A(s,t)^{\frac{1}{p}}. \end{aligned} \tag{4.26}$$

So we know by (4.14) and (4.26) that for every $[r,v] \subset [s,t] \subset [0,T]$

$$|\gamma_v-\gamma_r| \leq m_12C_1(D_1(\delta)+\|\gamma\|_{\infty,r,v}+|\gamma|_{p,r,v})(1+|\xi^n|_{p,r,v}+|\xi|_{p,r,v})(|v-r|+|w|_{p,r,v}) + A(r,v)^{\frac{1}{p}}$$

$$\le m_1 4C_1(D_1(\delta) + |\gamma_s| + |\gamma|_{p,s,t})(1 + |\xi^n|_{p,s,t} + |\xi|_{p,s,t})(|v-r| + |w|_{p,r,v}) + A(r,v)^{\frac{1}{p}},$$

since $\|\gamma\|_{\infty,s,t} \le |\gamma_s| + |\gamma|_{p,s,t}$. With Lemma 2.6, this yields

$$\begin{aligned} &|\gamma|_{p,s,t} \\ &\le m_1 4C_1(D_1(\delta) + |\gamma_s| + |\gamma|_{p,s,t})(1 + |\xi^n|_{p,s,t} + |\xi|_{p,s,t})(|t-s| + |w|_{p,s,t}) + A(s,t)^{\frac{1}{p}} \end{aligned} \tag{4.27}$$

for every $[s,t] \subset [0,T]$. Taking Remarks 2.28 and 4.3 into account, we can argue that for every $[s,t] \subset [0,T]$ such that

$$|t-s| + |w|_{p,s,t} \le \frac{1}{24C_1 m_1} \le \frac{1}{4C_1} \le \frac{1}{2C_1},$$

we have

$$|\xi^n|_{p,s,t} \le 1, \quad |\xi|_{p,s,t} \le 1$$

and by (4.27)

$$|\gamma|_{p,s,t} \le D_1(\delta) + 2A(0,T)^{\frac{1}{p}} + |\gamma_s|.$$

Now we can use Lemma 2.20 and since $\gamma_0 = \xi_0^n - \xi_0 = 0$, we obtain the estimate

$$\begin{aligned} |\gamma|_{p,0,T} &\le \left(D_1(\delta) + 2A(0,T)^{\frac{1}{p}} + |\gamma_0|\right) e^{2^p (24 m_1 C_1)^p (T^p + |w|_{p,0,T}^p)} \\ &\le \left(D_1(\delta) + 2A(0,T)^{\frac{1}{p}}\right) e^{7^{2p} (m_1 C_1)^p (T^p + |w|_{p,0,T}^p)} \end{aligned}$$

and therefore

$$\|\gamma\|_{\infty,0,T} \le \|\gamma\|_{p,0,T} \le \left(D_1(\delta) + 2A(0,T)^{\frac{1}{p}}\right) e^{7^{2p} (m_1 C_1)^p (T^p + |w|_{p,0,T}^p)}.$$

Now we examine $D_1(\delta)$ and $A(0,T)$ to get the convergence rate. Keeping (4.25) in mind and note that for $\beta \in [\frac{1}{p}, 1]$, we have $0 < 2-p \le 1 + (\beta-1)p \le \beta \le 1$. Hence,

$$\begin{aligned} D_1(\delta) &= C_1^2 \left(\delta^\beta + (1 + C_\xi)\delta + \delta\right) \\ &\le C_1^2 \left(C_w^{\beta-2+p} + (1 + C_\xi) C_w^{p-1} + C_w^{p-1}\right) \delta^{2-p} \\ &\le 3C_1^2 C_w (1 + C_\xi) \delta^{2-p} \end{aligned}$$

and

$$\begin{aligned} A(0,T)^{\frac{1}{p}} &= 2^{\frac{1}{p}} \delta^{2-p} K_1(\omega)(\varphi_1(0,T) + \varphi_2(0,T))^{\frac{1}{p}} \\ &= \delta^{2-p} K_2(\omega), \end{aligned}$$

where

$$2^{\frac{1}{p}} K_1(\omega)(\varphi_1(0,T) + \varphi_2(0,T))^{\frac{1}{p}}.$$

Collecting all the terms, we obtain

$$\begin{aligned}
&D_1(\delta) + 2A(0,T)^{\frac{1}{p}} \\
&\leq 3C_1^2 C_w \left(1 + C_\xi\right) \delta^{2-p} + 2\delta^{2-p} K_2(\omega) \\
&\leq (3C_1^2 C_w \left(1 + C_\xi\right) + 2K_2(\omega))\delta^{2-p}.
\end{aligned}$$

Hence there exists a constant $K_3(\omega)$ having moments of all orders which is independent of u and n such that

$$D_1(\delta) + 2A(0,T)^{\frac{1}{p}} \leq K_3(\omega)\delta^{2-p}.$$

This yields

$$\begin{aligned}
|\gamma|_{p,0,T} &\leq \delta^{2-p} K_3(\omega) e^{7^{2p}(m_1 C_1)^p (T^p + |w|^p_{p,0,T})} \\
&\leq K_\xi(\omega)\delta^{2-p},
\end{aligned}$$

where

$$K_\xi(\omega) := K_3(\omega) e^{7^{2p}(C_1 m1)^p (T^p + |w|^p_{p,0,T})}.$$

Meaning that for a given $\omega \in \mathcal{A}^c$ and $u \in \mathcal{U}$, we have

$$\begin{aligned}
\|\xi^u(\omega) - \xi^{n,u}(\omega)\|_{\infty,0,T} &\leq \|\xi^u(\omega) - \xi^{n,u}(\omega)\|_{p,0,T} \\
&= |\xi^u(\omega) - \xi^{n,u}(\omega)|_{p,0,T} \\
&= K_\xi(\omega)\delta(\omega)^{2-p}.
\end{aligned}$$

For the convergence rate in $L^l(\Omega, C([0,T]), \mathbb{R}^{n_1})$, we need to be careful, because δ depends on ω. Since w satisfies the exponential moment condition (2.48), we have that $K_\xi \in L^l(\Omega, \mathbb{R})$ and $\delta \in L^l(\Omega, \mathbb{R})$ for every $l \geq 1$. Now for a given $l \geq 1$ we have $\delta_{1,l} = \mathrm{E}[\delta^l]^{\frac{1}{l}}$ and we obtain with the Hölder inequality

$$\begin{aligned}
E\left[\|\xi^u - \xi^{n,u}\|^l_{\infty,0,T}\right]^{\frac{1}{l}} &\leq \mathrm{E}\left[\left(K_\xi \delta^{2-p}\right)^l\right]^{\frac{1}{l}} \\
&\leq \mathrm{E}\left[K_\xi^{2l}\right]^{\frac{1}{2l}} \mathrm{E}\left[\delta^{2l}\right]^{\frac{1}{2l}(2-p)} \\
&\leq \mathrm{E}\left[K_\xi^{2l}\right]^{\frac{1}{2l}} \delta_{1,2l}^{2-p}.
\end{aligned}$$

Note that K_ξ only depends on T, β, p, q, L and m_1. □

That is the same order of convergence as in Lejay [2010] in the autonomous case, by a slight different definition of δ. We now consider the solution of the second equation x^u in (4.3) and its Euler approximation scheme $x^{n,u}$ in (4.5). A comprehensive introduction to the numerical approximation of solutions to Itô SDEs is given in Kloeden and Platen [2011]. We use the previous results to show the strong convergence of x^n, consequently the convergence of $\|\mathcal{X} - \mathcal{X}^n\|_{\infty,0,T}$ in L^l-sense for $l \geq 2$. First, we show that x^n is bounded independently of the parameter and the number of subintervals of the Euler partition.

Lemma 4.5. *Let $u \in \mathcal{U}$ and $l \geq 2$, then there exists a positive constant $D_{x^n,l}$, independent of u, such that*

$$\mathrm{E}\left[\|x^n\|^l_{\infty,0,T}\right] \leq D_{x^n,l}.$$

Let

$$C_2 := 2\max\left\{L, L\sup\mathcal{U}, \max_{t\in[0,T]} \hat{b}(t,0,0,0)\right\},$$

we obtain for $\delta_2 = \max\limits_{i=0,\dots,n-1} |t_{i+1} - t_i|$ and $l \geq 2$, that

$$\mathrm{E}\left[\|x^n_\cdot - x^n_{t_{\overline{n}(\cdot)}}\|^l_{\infty,0,T}\right] \leq C(1 + D_{x^n,l} + D_{\xi^n,l})\delta_2^{\frac{l}{2}}, \tag{4.28}$$

where the constant $C > 0$ only depends on T, l, m_2 and C_2.

Proof. We omit the direct dependence of the involved processes on u for notational simplicity. The coefficient functions $\hat{b}$ and $\hat{\sigma}$ satisfy a linear growth condition by the Conditions (B_1) and (B_2). We proof this for $\hat{b}$, but the calculations for $\hat{\sigma}$ are completely analogous. Since $\hat{b}$ is continuous and continuously differentiable in its last 3 variables with bounded first derivatives, we obtain

$$\hat{b}(t,x,y,u) - \hat{b}(t,0,0,0) \leq |\hat{b}(t,x,y,u) - \hat{b}(t,0,0,0)| \leq L(|x| + |y| + |u|),$$

which yields the estimate

$$|\hat{b}(t,x,y,u)| \leq C_2(1 + |x| + |y|),$$

by the boundedness of $\mathcal{U}$ and the time continuity of $\hat{b}$ on the compact set $[0,T]$. For the rest of this proof we will use C as a constant which only depends on C_2, l, m_2 and T and can vary from line to line. By the definition of x^n, since $x_0(u)$ is constant, we get inductively that x^n is continuous and $\mathbb{F}$-adapted by the adaptedness and continuity of ξ^n and B. Furthermore by the linear growth condition of $\hat{\sigma}$, we get inductively that $x^n_{t_i}$ has moments of all orders for all $i = 0, \dots, n$ and the process

$$(\omega, s) \mapsto \int_0^s \hat{\sigma}(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u)\, dB_r$$

is a m_2-dimensional vector of $\mathbb{F}$-martingales. We have for $t \in [0,T]$ and $l \geq 2$ by the Jensen inequality

$$\begin{aligned}
\mathrm{E}[\|x^n\|_{\infty,0,t}^l] &\leq C\Bigg(\mathrm{E}\left[|x_0^n|^l\right] + \mathrm{E}\left[\sup_{s\in[0,t]}\left|\int_0^s \hat{b}\left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right)\, dr\right|^l\right] \\
&\quad + \mathrm{E}\left[\sup_{s\in[0,t]}\left|\int_0^s \hat{\sigma}\left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right)\, dB_r\right|^l\right]\Bigg) \\
&= C\left(\mathrm{E}\left[|x_0^n|^l\right] + I_1 + I_2\right).
\end{aligned}$$

We can estimate I_1 by the linear growth condition of $\hat{b}$ and Lemma 4.2 and Fubinis theorem

$$\begin{aligned}
I_1 &\leq \mathrm{E}\left[\int_0^t \left|\hat{b}\left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right)\right|^l\, dr\right] \\
&\leq C\mathrm{E}\left[\int_0^t \left(1 + \left|x^n_{t_{\overline{n}(r)}}\right| + \left|\xi^n_{t_{\overline{n}(r)}}\right|\right)^l\, dr\right] \\
&\leq C\mathrm{E}\left[\left(1 + \|\xi^n\|_{\infty,0,T}\right)^l\right] + C\int_0^t \mathrm{E}\left[\|x^n\|_{\infty,0,r}^l\right]\, dr \\
&\leq C(1 + D_{\xi^n,l}) + C\int_0^t \mathrm{E}\left[\|x^n\|_{\infty,0,r}^l\right]\, dr.
\end{aligned}$$

Similar calculations after the use of the Burkholder-Davis-Gundy inequality (Theorem 3.28/Remark 3.30 in Karatzas and Shreve [1991]) yield

$$\begin{aligned}
I_2 &\leq C\mathrm{E}\left[\left(\int_0^t \left|\hat{\sigma}\left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right)\right|^2\, dr\right)^{\frac{l}{2}}\right] \\
&\leq C\mathrm{E}\left[\int_0^t \left(1 + \left|\xi^n_{t_{\overline{n}(r)}}\right| + \left|x^n_{t_{\overline{n}(r)}}\right|\right)^l\, dr\right] \\
&\leq C\mathrm{E}\left[\left(1 + \|\xi^n\|_{\infty,0,T}\right)^l\right] + C\int_0^t \mathrm{E}\left[\|x^n\|_{\infty,0,r}^l\right]\, dr \\
&\leq C(1 + D_{\xi^n,l}) + C\int_0^t \mathrm{E}\left[\|x^n\|_{\infty,0,r}^l\right]\, dr.
\end{aligned}$$

Putting all terms together, we get

$$\mathrm{E}[\|x^n\|_{\infty,0,t}^l] \leq C|x_0(u)|^l + C(1 + D_{\xi^n,l}) + C\int_0^t \mathrm{E}\left[\|x^n\|_{\infty,0,r}^l\right]\, dr.$$

We conclude by the Gronwall inequality (Lemma 6.2 in Hale [2009]) and Condition (B_3)

$$\begin{aligned}
E\left[\|x^n\|_{\infty,0,T}^l\right] &\leq C(|x_0(u)|^l + 1 + D_{\xi^n,l})e^C \\
&\leq C(L^l + 1 + D_{\xi^n,l})e^C \\
&\leq C(1 + D_{\xi^n,l})e^C := D_{x^n,l}
\end{aligned}$$

for all $l \geq 2$. Now let $t \in [t_i, t_{i+1}]$ for $i \in \{0, \dots, n-1\}$ and $l \geq 2$, we have

$$
\begin{aligned}
&\mathrm{E}\left[|x_t^n - x_{t_i}^n|^l\right] \\
&\leq C\mathrm{E}\left[\left|\int_{t_i}^t \hat{b}\left(t_i, x_{t_i}^n, \xi_{t_i}^n, u\right)\, dr\right|^l + \left|\int_{t_i}^t \hat{\sigma}\left(t_i, x_{t_i}^n, \xi_{t_i}^n, u\right)\, dB_r\right|^l\right] \\
&\leq C\mathrm{E}\left[(t-t_i)^{l-1}\int_{t_i}^t \left|\hat{b}\left(t_i, x_{t_i}^n, \xi_{t_i}^n, u\right)\right|^l\, dr\right] + C\mathrm{E}\left[\left(\int_{t_i}^t \left|\hat{\sigma}\left(t_i, x_{t_i}^n, \xi_{t_i}^n, u\right)\right|^2\, dr\right)^{\frac{l}{2}}\right] \\
&\leq C(t-t_i)^{l-1}\mathrm{E}\left[\int_{t_i}^t (1 + \|x^n\|_{\infty,0,T} + \|\xi^n\|_{\infty,0,T}|)^l\, dr\right] \\
&\quad + C\mathrm{E}\left[\left((1 + \|x^n\|_{\infty,0,T} + \|\xi^n\|_{\infty,0,T})^2\,(t-t_i)\right)^{\frac{l}{2}}\right] \\
&\leq C\mathrm{E}\left[(1 + \|x^n\|_{\infty,0,T} + \|\xi^n\|_{\infty,0,T})^l\right]\delta_2^{\frac{l}{2}} \\
&\leq C(1 + D_{x^n,l} + D_{\xi^n,l})\delta_2^{\frac{l}{2}}.
\end{aligned}
$$

This yields the estimate (4.28) for $l \geq 2$. □

Having established the boundedness of the Euler scheme, we can focus on the convergence rate, which will be stated in the following theorem.

Theorem 4.6. *Let $u \in \mathcal{U}$, we have*

$$
\mathrm{E}\left[\|x^u - x^{n,u}\|_{\infty,0,T}^l\right]^{\frac{1}{l}} \leq D_{K_x,l}\delta_{1,2l}^{(2-p)\wedge\frac{1}{2}}
$$

for any $l \geq 2$, where the constant $D_{K_x,l}$ is independent of u and n.

Proof. We use C as a generic constants which has different values over the course of the proof, but only depends on T, l and m_2. We have for $t \in [0,T]$

$$
\begin{aligned}
x_t - x_t^n &= x_0(u) + \int_0^t \hat{b}(r, x_r, \xi_r, u)\, dr + \int_0^t \hat{\sigma}(r, x_r, \xi_r, u)\, dB_r \\
&\quad - x_0^n - \int_0^t \hat{b}\left(t_{\overline{n}(r)}, x_{t_{\overline{n}(r)}}^n, \xi_{t_{\overline{n}(r)}}^n, u\right)\, dr - \int_0^t \hat{\sigma}\left(t_{\overline{n}(r)}, x_{t_{\overline{n}(r)}}^n, \xi_{t_{\overline{n}(r)}}^n, u\right)\, dB_r \\
&= x_0(u) - x_0^n + \int_0^t \hat{b}(r, x_r, \xi_r, u) - \hat{b}\left(t_{\overline{n}(r)}, x_r^n, \xi_r^n, u\right)\, dr \\
&\quad + \int_0^t \hat{\sigma}(r, x_r, \xi_r, u) - \hat{\sigma}\left(t_{\overline{n}(r)}, x_r^n, \xi_r^n, u\right)\, dB_r \\
&\quad + \int_0^t \hat{b}\left(t_{\overline{n}(r)}, x_r^n, \xi_r^n, u\right) - \hat{b}\left(t_{\overline{n}(r)}, x_{t_{\overline{n}(r)}}^n, \xi_{t_{\overline{n}(r)}}^n, u\right)\, dr \\
&\quad + \int_0^t \hat{\sigma}\left(t_{\overline{n}(r)}, x_r^n, \xi_r^n, u\right) - \hat{\sigma}\left(t_{\overline{n}(r)}, x_{t_{\overline{n}(r)}}^n, \xi_{t_{\overline{n}(r)}}^n, u\right)\, dB_r,
\end{aligned}
$$

which yields for $l \geq 2$, since $x_0^n = x_0(u)$

$$\begin{aligned}
\mathrm{E}\left[\|x - x^n\|_{\infty,0,t}^l\right] \leq C\Bigg(& \mathrm{E}\left[\sup_{0\leq s\leq t}\left|\int_0^s \hat{b}(r, x_r, \xi_r, u) - \hat{b}\left(t_{\overline{n}(r)}, x_r^n, \xi_r^n, u\right)\, dr\right|^l\right] \\
&+ \mathrm{E}\left[\sup_{0\leq s\leq t}\left|\int_0^s \hat{\sigma}(r, x_r, \xi_r, u) - \hat{\sigma}\left(t_{\overline{n}(r)}, x_r^n, \xi_r^n, u\right)\, dB_r\right|^l\right] \\
&+ \mathrm{E}\left[\sup_{0\leq s\leq t}\left|\int_0^s \hat{b}\left(t_{\overline{n}(r)}, x_r^n, \xi_r^n, u\right) - \hat{b}\left(t_{\overline{n}(r)}, x_{t_{\overline{n}(r)}}^n, \xi_{t_{\overline{n}(r)}}^n, u\right)\, dr\right|^l\right] \\
&+ \mathrm{E}\left[\sup_{0\leq s\leq t}\left|\int_0^s \hat{\sigma}\left(t_{\overline{n}(r)}, x_r^n, \xi_r^n, u\right) - \hat{\sigma}\left(t_{\overline{n}(r)}, x_{t_{\overline{n}(r)}}^n, \xi_{t_{\overline{n}(r)}}^n, u\right)\, dB_r\right|^l\right]\Bigg) \\
= C(I_1 + I_2 + I_3 + I_4).&
\end{aligned}$$

We define

$$\delta_2 := \max_{i=0,\ldots,n-1} |t_{i+1} - t_i|$$

and estimate the term I_1 by conditions (B_1), (B_2), (E_2) and Theorem 4.4

$$\begin{aligned}
I_1 &\leq C\mathrm{E}\left[\int_0^t \left|\hat{b}(r, x_r, \xi_r, u) - \hat{b}\left(t_{\overline{n}(r)}, x_r^n, \xi_r^n, u\right)\right|^l\, dr\right] \\
&\leq C\mathrm{E}\left[\int_0^t (1 + |x_r| + |\xi_r|)^l \delta_2^{\frac{l}{2}} + |x_r - x_r^n|^l + |\xi_r - \xi_r^n|^l\, dr\right] \\
&\leq C(1 + D_{x,l} + D_{\xi,l})\delta_2^{\frac{l}{2}} + C\mathrm{E}\left[\|\xi - \xi^n\|_{\infty,0,T}^l\right] + C\int_0^t \mathrm{E}\left[\|x - x^n\|_{\infty,0,r}^l\right]\, dr \\
&\leq C(1 + D_{x,l} + D_{\xi,l})\delta_2^{\frac{l}{2}} + CD_{K_\xi,2l}^l \delta_{1,2l}^{l(2-p)} + C\int_0^t \mathrm{E}\left[\|x - x^n\|_{\infty,0,r}^l\right]\, dr.
\end{aligned}$$

Similar arguments for the estimation of I_2, after using the Burkholder-Davis-Gundy inequality, yield

$$\begin{aligned}
I_2 &\leq C\mathrm{E}\left[\left(\int_0^t \left|\hat{\sigma}(r, x_r, \xi_r, u) - \hat{\sigma}\left(t_{\overline{n}(r)}, x_r^n, \xi_r^n, u\right)\right|^2\, dr\right)^{\frac{l}{2}}\right] \\
&\leq C\mathrm{E}\left[\int_0^t (1 + |x_r| + |\xi_r|)^l \delta_2^{\frac{l}{2}} + |x_r - x_r^n|^l + |\xi_r - \xi_r^n|^l\, dr\right] \\
&\leq C(1 + D_{x,l} + D_{\xi,l})\delta_2^{\frac{l}{2}} + C\mathrm{E}\left[\|\xi - \xi^n\|_{\infty,0,T}^l\right] + C\int_0^t \mathrm{E}\left[\|x - x^n\|_{\infty,0,r}^l\right]\, dr \\
&\leq C(1 + D_{x,l} + D_{\xi,l})\delta_2^{\frac{l}{2}} + CD_{K_\xi,2l}^l \delta_{1,2l}^{l(2-p)} + C\int_0^t \mathrm{E}\left[\|x - x^n\|_{\infty,0,r}^l\right]\, dr.
\end{aligned}$$

Furthermore, we obtain

$$
\begin{aligned}
I_3 &\leq C\mathrm{E}\left[\int_0^t \left|\hat{b}\left(t_{\overline{n}(r)}, x_r^n, \xi_r^n, u\right) - \hat{b}\left(t_{\overline{n}(r)}, x_{t_{\overline{n}(r)}}^n, \xi_{t_{\overline{n}(r)}}^n, u\right)\right|^l \, dr\right] \\
&\leq C\mathrm{E}\left[\int_0^t |x_r^n - x_{t_{\overline{n}(r)}}^n|^l + |\xi_r^n - \xi_{t_{\overline{n}(r)}}^n|^l \, dr\right],
\end{aligned}
$$

where

$$
\mathrm{E}\left[\|x_\cdot^n - x_{t_{\overline{n}(\cdot)}}^n\|_{\infty,0,T}^l\right] \leq C(1 + D_{x^n,l} + D_{\xi^n,l})\delta_2^{\frac{l}{2}}
$$

by (4.28) and

$$
\mathrm{E}\left[\|\xi_\cdot^n - \xi_{t_{\overline{n}(\cdot)}}^n\|_{\infty,0,T}^l \, dr\right] \leq L^l \delta_{1,l}^l
$$

by (4.8). This yields

$$
I_3 \leq C\delta_{1,l}^l + C(1 + D_{x^n,l} + D_{\xi^n,l})\delta_2^{\frac{l}{2}}.
$$

By similar calculation as for I_3 together with the Burkholder-Davis-Gundy inequality, we obtain

$$
I_4 \leq C\delta_{1,l}^l + C(1 + D_{x^n,l} + D_{\xi^n,l})\delta_2^{\frac{l}{2}}.
$$

Combining all the estimates, it follows

$$
\begin{aligned}
\mathrm{E}\left[\|x - x^n\|_{\infty,0,t}^l\right] \leq{}& C\delta_2^{\frac{l}{2}}(1 + D_{x,l} + D_{x^n,l} + D_{\xi,l} + D_{\xi^n,l}) + C\delta_{1,l}^l + CD_{K_\xi,2l}^l \delta_{1,2l}^{l(2-p)} \\
&+ C\int_0^t \mathrm{E}\left[\|x - x^n\|_{\infty,0,r}^l\right] \, dr.
\end{aligned}
$$

By the Gronwall inequality, we conclude

$$
\mathrm{E}\left[\|x - x^n\|_{\infty,0,T}^l\right] \leq \left(C\delta_{1,l}^l + C(1 + D_{1,l})\delta_2^{\frac{l}{2}} + CD_{K_\xi,2l}^l \delta_{1,2l}^{l(2-p)}\right) e^C,
$$

where $D_{1,l} := D_{x,l} + D_{x^n,l} + D_{\xi,l} + D_{\xi^n,l}$. Since $\delta_{1,l} \leq \delta_{1,2l}$ and

$$
\delta_{1,2l} \leq \mathrm{E}\left[(T + |w|_{p,0,T})^{2l}\right]^{\frac{1}{2l}} \leq \mathrm{E}\left[(1 + T + |w|_{p,0,T})^{2l}\right]^{\frac{1}{2l}} := D_{w,2l} \geq 1, \tag{4.29}
$$

we can estimate

$$
\begin{aligned}
&\mathrm{E}\left[\|x - x^n\|_{\infty,0,T}^l\right] \\
&\leq C\delta_{1,l}^l + C(1 + D_{1,l})\delta_2^{\frac{l}{2}} + CD_{K_\xi,2l}^l \delta_{1,2l}^{l(2-p)} \\
&\leq C\delta_{1,2l}^{l(2-p)}\left(1 + D_{w,2l}^{l(p-1)} + D_{K_\xi,2l}^l\right) + C(1 + D_{1,l})\delta_2^{\frac{l}{2}}.
\end{aligned} \tag{4.30}
$$

Now we can use the inequality $\delta_2 \leq \delta_{1,2l}$ to get a convergence rate in the parameter $\delta_{1,2l}$, but note that this is not ideal in the case where the process w is Hölder continuous for some Hölder exponent $H \in (\frac{1}{2}, 1)$. We will come back to this case in Subsection 4.1.3. We conclude

$$\begin{aligned}
\mathrm{E}\left[\|x - x^n\|_{\infty,0,t}^l\right]^{\frac{1}{l}} &\leq C D_{w,2l}^{\frac{1}{2}} \left(D_{w,2l}^{l(p-1)} + D_{1,l} + D_{K_\xi,2l}^l\right)^{\frac{1}{l}} \delta_{1,2l}^{(2-p)\wedge\frac{1}{2}} \\
&:= D_{K_x,l}\delta_{1,2l}^{(2-p)\wedge\frac{1}{2}}
\end{aligned}$$

for a constant $D_{K_x,l}$ which is independent of u and n. $\square$

We end this subsection with its main result, concerning the convergence rate of $\mathcal{X}^n$ to $\mathcal{X}$. The result is a direct consequence of Theorem 4.4 and Theorem 4.6.

Theorem 4.7. *With the notations and assumptions from the beginning of this chapter, we have for every $u \in \mathcal{U}$*

$$\mathrm{E}\left[\|\mathcal{X}^{n,u}\|_{\infty,0,T}^l\right] \leq D_{\mathcal{X}^n,l} \tag{4.31}$$

and

$$\mathrm{E}\left[\|\mathcal{X}^u - \mathcal{X}^{n^u}\|_{\infty,0,T}^l\right]^{\frac{1}{l}} \leq D_{K_{\mathcal{X}},l}\delta_{1,2l}^{(2-p)\wedge\frac{1}{2}}.$$

for any $l \geq 2$, where the constants $D_{\mathcal{X}^n,l}$ and $D_{K_{\mathcal{X}},l}$ are independent of u and n.

Proof. We have by the Jensen inequality

$$\mathrm{E}\left[\|\mathcal{X}^{n,u}\|_{\infty,0,T}^l\right] \leq 2^{l-1}\left(\mathrm{E}\left[\|\xi^{n,u}\|_{\infty,0,T}^l\right] + \mathrm{E}\left[\|x^{n,u}\|_{\infty,0,T}^l\right]\right).$$

By Lemma 4.2 and Lemma 4.5, this yields

$$\mathrm{E}\left[\|\mathcal{X}^{n,u}\|_{\infty,0,T}^l\right] \leq 2^{l-1}(D_{\xi^n,l} + D_{x^n,l}) := D_{\mathcal{X}^n,l}.$$

Using similar arguments, utilizing Theorem 4.4 and Theorem 4.6, we get

$$\begin{aligned}
\mathrm{E}\left[\|\mathcal{X}^u - \mathcal{X}^{n,u}\|_{\infty,0,T}^l\right]^{\frac{1}{l}} &\leq 2^{1-\frac{1}{l}}\left(D_{K_\xi,2l}\delta_{1,2l}^{2-p} + D_{K_x,l}\delta_{1,2l}^{(2-p)\wedge\frac{1}{2}}\right) \\
&\leq 2^{1-\frac{1}{l}}\left(D_{K_\xi,2l}D_{w,2l}^{\frac{1}{2}} + D_{K_x,l}\right)\delta_{1,2l}^{(2-p)\wedge\frac{1}{2}} \\
&= D_{K_{\mathcal{X}},l}\delta_{1,2l}^{(2-p)\wedge\frac{1}{2}}.
\end{aligned}$$

This proves the assertion. $\square$

The results in this subsection show that we can approximate the solution to the model dynamics equation with the corresponding first order Euler schemes. We continue with the approximation of the sensitivity equation (4.4).

4.1.2 Convergence of the Euler scheme for the sensitivity equation

Similar to the last subsection, we split the convergence analysis. First, we examine the convergence of y^n in (4.6) to y in (4.4) P-a.s. in uniform norm and in $L^l_{\mathbb{F}}(\Omega, C([0,T], \mathbb{R}^{n_1 \times d})$ for $l \geq 1$. We start by showing that the continuous interpolation of y^n is P-a.s. bounded, independently of the number of subintervals of the partition n and of the parameter u. In Lemma 4.2, we were able to estimate the p-variation of ξ^n on every interval $[\tau_i, \tau_{i+1}]$ of a given greedy sequence of times by a positive constant, because of the bounded coefficient functions b and σ. In the equation

$$\begin{aligned} y_t^{n,u} = D\xi_0(u) + \int_0^t b_x\left(t_{\overline{n}(r)}, \xi^{n,u}_{t_{\overline{n}(r)}}, u\right) y^{n,u}_{t_{\overline{n}(r)}} + b_u\left(t_{\overline{n}(r)}, \xi^{n,u}_{t_{\overline{n}(r)}}, u\right) dr \\ + \sum_{j=1}^{m_1} \int_0^t \sigma^j_x\left(t_{\overline{n}(r)}, \xi^{n,u}_{t_{\overline{n}(r)}}, u\right) y^{n,u}_{t_{\overline{n}(r)}} + \sigma^j_u\left(t_{\overline{n}(r)}, \xi^{n,u}_{t_{\overline{n}(r)}}, u\right) dw^j_r, \end{aligned} \tag{4.32}$$

which is satisfied by the continuous interpolation of the Euler scheme for the process y given in (4.6) for all $t \in [0,T]$, we will not be able to estimate the p-variation of y on any interval of a greedy sequence directly. To see this let t_i be a partition point of the Euler partition Π^{E} and $t_i < s < t_{i+1}$, then the p-variation of y on the interval $[s,t]$ for any $t > s$ will depend on the value of y at time t_i, because of the factor $y_{t_{\overline{n}(r)}}$ in the coefficients of equation (4.32). This will not allow us to estimate the p-variation of y on an arbitrary interval $[s,t] \subset [0,T]$ directly. We will have to restrict ourselves on the partition points of the Euler partition Π^{E}. This on the other hand yields the problem, that we need to utilize the greedy sequences of times to get the desired estimate, and the greedy sequence and Π^{E} need not to have common partition points apart from 0 and T. To work around this problem, we need preliminary results. We will construct a new partition $\Pi^{\mathrm{c}} = \{\theta_j\}_{j=0,\ldots,\mathcal{N}} \subset \Pi^{\mathrm{E}}$ on $[0,T]$, whose number of subintervals will not depend on n but the number of subintervals of the greedy sequences of times. Recall the two mentioned partitions we already have on $[0,T]$, namely the partition for the Euler scheme

$$\Pi^{\mathrm{E}} = \{t_i\}_{i=0,\ldots,n} \text{ with } 0 = t_0 < t_1 < \ldots, t_n = T$$

and the greedy sequence of times for a constant $0 < M$

$$\Pi^{\mathrm{g}} = \{\tau_i\}_{i=0,\ldots,N} \text{ with } 0 = \tau_0 < \tau_1 < \cdots < \tau_N = T,$$

which satisfies

$$\begin{aligned} |\tau_{i+1} - \tau_i| + |w|_{p,\tau_i,\tau_{i+1}} = M \text{ for all } i = 0, \ldots, N-1 \\ |\tau_N - \tau_{N-1}| + |w|_{p,\tau_{N-1},\tau_N} \leq M. \end{aligned}$$

Note that the number of subintervals in Π^{g} is bounded, as long as w is continuous and has finite p-variation on $[0,T]$, see Lemma 2.19. If $\tau_j \leq t_i \leq t_{i+1} \leq \tau_{j+1}$ for some $i \in \{0, \ldots, n-1\}$ and

$j \in \{0, \dots, N-1\}$, we can estimate

$$|t_{i+1} - t_i| + |w|_{p,t_i,t_{i+1}} \leq |\tau_{j+1} - \tau_j| + |w|_{p,\tau_j,\tau_{j+1}} \leq M.$$

But since we are not able to foretell the position of the greedy sequence partition points, there is a possibility that we have multiple partition points of the greedy sequence between two partition points of Π^{E}. Let $\tau_{j-1} < t_i < \tau_j < \dots < \tau_{j+m} < t_{i+1} < \tau_{j+m+1}$ and denote the number of subintervals of the greedy sequence between t_i and t_{i+1} by $m = N(t_i, t_{i+1})$ (see (2.8)). If we now want to estimate $|t_{i+1} - t_i| + |w|_{p,t_i,t_{i+1}}$, we have to take all the subintervals of the greedy sequence between τ_{j-1} and τ_{j+m+1} into account. We have by the triangle and Jensen inequality (compare Lemma 2.9)

$$\begin{aligned}
|t_{i+1} - t_i| + |w|_{p,t_i,t_{i+1}} &\leq |\tau_{j+m+1} - \tau_{j-1}| + |w|_{p,\tau_{j-1},\tau_{j+m+1}} \\
&\leq \sum_{i=0}^{m+1} |\tau_{j+i} - \tau_{j-1+i}| + \left((m+2)^{p-1} \sum_{i=0}^{m+1} |w|^p_{p,\tau_{j-1+i},\tau_{j+i}} \right)^{\frac{1}{p}} \\
&\leq \sum_{i=0}^{m+1} |\tau_{j+i} - \tau_{j-1+i}| + (m+2)^{1-\frac{1}{p}} \sum_{i=0}^{m+1} |w|_{p,\tau_{j-1+i},\tau_{j+i}} \\
&\leq (N(t_i, t_{i+1}) + 2)^{1-\frac{1}{p}} \sum_{i=0}^{N(t_i,t_{i+1})+1} \left(|\tau_{j+i} - \tau_{j-1+i}| + |w|_{p,\tau_{j-1+i},\tau_{j+i}} \right) \\
&\leq (N(t_i, t_{i+1}) + 2)^{2-\frac{1}{p}} M. \qquad (4.33)
\end{aligned}$$

To take these possibilities into account and have notational foundation for the calculations to come, we construct the subpartition Π^{c} of Π^{E} in the following way. First, we introduce the notation

$$\begin{aligned}
\underline{n} &: [0,T] \to \mathbb{N}, s \mapsto \min\{i \in \{0, \dots, n\} \,|\, t_i \in \Pi^{\mathrm{E}} \text{ and } t_i \geq s\} \qquad (4.34) \\
\overline{n} &: [0,T] \to \mathbb{N}, s \mapsto \max\{i \in \{0, \dots, n\} \,|\, t_i \in \Pi^{\mathrm{E}} \text{ and } t_i \leq s\}
\end{aligned}$$

and define the new partition

$$(\theta_j)_{j=0,\dots,\mathcal{N}} = \Pi^{\mathrm{c}} = \left\{ t \in \Pi^{\mathrm{E}} \,\middle|\, \exists \tau \in \Pi^{\mathrm{g}} \text{ such that } t = t_{\underline{n}(\tau)} \text{ or } t = t_{\overline{n}(\tau)} \right\}.$$

We give a graphical illustration of the partitions in figure 4.1.

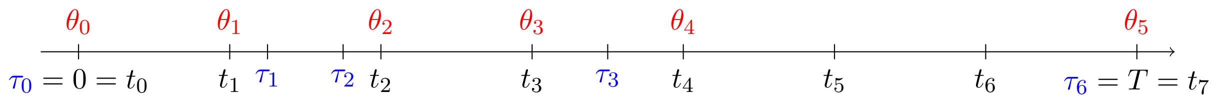

Figure 4.1: Graphical illustration of the construction of the partition Π^{c}.

Some properties of Π^{c} are given by

i) If $\tau = t$ for a $\tau \in \Pi^{\mathrm{g}}$ and $t \in \Pi^{\mathrm{E}}$, then there exists $\theta \in \Pi^{\mathrm{c}}$ such that $\theta = t = t_{\underline{n}(\tau)} = t_{\overline{n}(\tau)}$.

ii) $0, T \in \Pi^{\mathrm{c}}$, since $0 = \tau_0 = t_0 = \theta_0$ and $T = t_n = \tau_N = \theta_{\mathcal{N}}$.

iii) There can be multiple partition points $\tau \in \Pi^{\mathrm{greedy}}$ such that $\theta_j = t_{\overline{n}(\tau)}$ and $\theta_{j+1} = t_{\underline{n}(\tau)}$, e.g. τ_1, τ_2 in Figure 4.1.

For some further notation, we recall from Chapter 2.1

$$\begin{aligned}
&\underline{N} : [0,T] \to \mathbb{N}, s \mapsto \min\{i \in \{0,\dots,N\} \,|\, \tau_i \in \Pi^{\mathrm{g}} \text{ and } \tau_i \geq s\} \\
&\overline{N} : [0,T] \to \mathbb{N}, s \mapsto \max\{i \in \{0,\dots,N\} \,|\, \tau_i \in \Pi^{\mathrm{g}} \text{ and } \tau_i \leq s\} \\
&N : \Delta([0,T]) \to \mathbb{N}, (s,t) \mapsto \overline{N}(t) - \underline{N}(s)
\end{aligned}$$

and define the new functions

$$\begin{aligned}
&\underline{\mathcal{N}} : [0,T] \to \mathbb{N}, s \mapsto \min\{i \in \{0,\dots,\mathcal{N}\} \,|\, \theta_i \in \Pi^{\mathrm{c}} \text{ and } \theta_i \geq s\} \\
&\overline{\mathcal{N}} : [0,T] \to \mathbb{N}, s \mapsto \max\{i \in \{0,\dots,\mathcal{N}\} \,|\, \theta_i \in \Pi^{\mathrm{c}} \text{ and } \theta_i \leq s\} \\
&\mathcal{N} : \Delta([0,T]) \to \mathbb{N}, (s,t) \mapsto \overline{\mathcal{N}}(t) - \underline{\mathcal{N}}(s).
\end{aligned}$$

We have by construction $\mathcal{N}(s,t) \leq 2N(s,t) + 1$ for all $(s,t) \in \Delta([0,T])$. The next lemma is of Gronwall type and utilizes the partitions we just defined.

Lemma 4.8 (Gronwall type lemma on the Euler partition)**.** *Let $\Pi^E = \{t_i\}_{i=0,\dots,n}$ be a partition of $[0,T]$ and let $x \in W^p([0,T], \mathbb{R}^{d\times m})$, where $p \in (1,2)$. Furthermore let $w : [0,T] \to \mathbb{R}^m$ be a continuous function of finite p-variation, $K_1, a > 0$ be constants. If for every $t_i \in \Pi^E$, $i \in \{0,\dots,n-1\}$, we have*

$$|x|_{p,t_i,t_{i+1}} \leq a(K_1 + |x_{t_i}|)(|t_{i+1} - t_i| + |w|_{p,t_i,t_{i+1}}) \tag{4.35}$$

and there exists a constant $K_2 \leq \frac{1}{a}$ such that for $t_l, t_k \in \Pi^E$ with $0 \leq t_l < t_{l+1} < t_k \leq T$ and

$$|t_k - t_l| + |w|_{p,t_l,t_k} \leq K_2,$$

it holds that

$$|x|_{p,t_l,t_k} \leq K_1 + |x_{t_l}|, \tag{4.36}$$

then we obtain

$$|x|_{p,0,T} \leq (3K_1 + |x_0|) \left(2^p K_2^{-p} \left(T^p + |w|^p_{p,0,T}\right) + 1\right) \exp\left(2^p 3 K_2^{-p} \left(T^p + |w|^p_{p,0,T}\right) + 2\right).$$

Proof. We can construct our partitions Π^{g} and Π^{c} for the constant K_2 with $0 < K_2 < \frac{1}{a}$. The number of subintervals of the partitions Π^{g} and Π^{c} is given by N respectively $\mathcal{N}$. Now we consider the p-variation of x on the subintervals $[\theta_i, \theta_{i+1}]$ of the partition Π^{c} for $i \in \{0, \dots, \mathcal{N}-1\}$, for which there are two possibilities.

Case 1: There exists $\tau_l \in \Pi^{\mathrm{g}}$ and $i \in \{0, \dots, \mathcal{N}-1\}$ such that $\theta_i = t_{\overline{n}(\tau_l)}$ and $\theta_{i+1} = t_{\underline{n}(\tau_l)}$ (e.g. $[\theta_1, \theta_2]$, $[\theta_3, \theta_4]$ in figure 4.1). By construction it follows that there exists $j \in \{0, \dots, n\}$ such that $\theta_i = t_j$ and $\theta_{i+1} = t_{j+1}$. By Property iii) of the Partition Π^{c} there can be multiple partition points of Π^{g} in the interval $[t_j, t_{j+1}]$. We estimate using (4.33) and (4.35)

$$\begin{aligned}
|x|_{p,\theta_i,\theta_{i+1}} &= |x|_{p,t_j,t_{j+1}} \\
&\leq a(K_1 + |x_{t_j}|)(|t_{j+1} - t_j| + |w|_{p,t_j,t_{j+1}}) \\
&\leq a(K_1 + |x_{t_j}|)(|\tau_{\underline{N}(t_{j+1})} - \tau_{\overline{N}(t_j)}| + |w|_{p,\tau_{\overline{N}(t_j)},\tau_{\underline{N}(t_{j+1})}}) \\
&\leq (K_1 + |x_{\theta_i}|)(N(\theta_i, \theta_{i+1}) + 2)^{2-\frac{1}{p}}.
\end{aligned} \tag{4.37}$$

Case 2: There exists $\tau_j, \tau_{j+1} \in \Pi^{\mathrm{g}}$ and $i \in \{0, \dots, \mathcal{N}-1\}$ such that $\theta_i = t_{\underline{n}(\tau_j)}$ and $\theta_{i+1} = t_{\overline{n}(\tau_{j+1})}$ (e.g. $[\theta_i, \theta_{i+1}]$ for $i \in \{0, 2, 4\}$ in Figure 4.1). Then there exists a finite number $k - l = m \geq 1$ of subintervals of Π^{E} in the interval $[\theta_i, \theta_{i+1}]$. Let $\theta_i = t_l < t_{l+1}, \dots, t_{l+m} = t_k = \theta_{i+1}$, if $m = 1$ we have by (4.35)

$$\begin{aligned}
|x|_{p,\theta_i,\theta_{i+1}} &= |x|_{p,t_l,t_{l+1}} \\
&\leq a(K_1 + |x_{t_l}|)(|t_{l+1} - t_l| + |w|_{p,t_l,t_{l+1}}).
\end{aligned}$$

By assumption on the form of $[\theta_i, \theta_{i+1}]$, we have

$$|t_{l+1} - t_l| + |w|_{p,t_l,t_{l+1}} \leq |\tau_{j+1} - \tau_j| + |w|_{p,\tau_j,\tau_{j+1}} \leq K_2 \leq \frac{1}{a},$$

which yields

$$|x|_{p,\theta_i,\theta_{i+1}} = |x|_{p,t_l,t_{l+1}} \leq K_1 + |x_{\theta_i}|. \tag{4.38}$$

Now let $m \geq 2$, since

$$|t_k - t_l| + |w|_{p,t_l,t_k} \leq K_2,$$

we have by (4.36) that

$$|x|_{p,\theta_i,\theta_{i+1}} = |x|_{p,t_l,t_k} \leq K_1 + |x_{t_l}| = K_1 + |x_{\theta_i}|. \tag{4.39}$$

By taking (4.37), (4.38) and (4.39) into account, this yields

$$|x|_{p,\theta_i,\theta_{i+1}} \leq (N(\theta_i, \theta_{i+1}) + 2)^{2-\frac{1}{p}}(K_1 + |x_{\theta_i}|)$$

for every $i \in \{0, \dots, \mathcal{N}-1\}$. Now we show inductively that

$$|x_{\theta_i}| \le e^{2(N(0,\theta_i)+\mathcal{N}(0,\theta_i))}(2K_1 + |x_0|) \tag{4.40}$$

for every $i \in \{1, \dots, \mathcal{N}\}$. We have for all $i \in \{0, \dots, \mathcal{N}-1\}$ that

$$|x_{\theta_{i+1}}| \le |x_{\theta_i}| + |x|_{p,\theta_i,\theta_{i+1}} \le |x_{\theta_i}| + (N(\theta_i, \theta_{i+1}) + 2)^{2-\frac{1}{p}}(K_1 + |x_{\theta_i}|).$$

Hence, for $i = 1$

$$|x_{\theta_1}| \le 2(N(0, \theta_1) + 2)^{2-\frac{1}{p}}(K_1 + |x_0|).$$

Note that

$$2(x+2)^{2-\frac{1}{p}} \le 2(x+2)^{\frac{3}{2}} \le e^{2(x+1)}$$

for every $x \ge 0$. Hence

$$|x_{\theta_1}| \le e^{2(N(0,\theta_1)+1)}(K_1 + |x_0|).$$

Since $\mathcal{N}(0, \theta_1) = 1$ the statement follows for $i = 1$. Now assume (4.40) holds for some $i \in \{0, \dots, \mathcal{N}-1\}$, then

$$\begin{aligned}
|x_{\theta_{i+1}}| &\le |x_{\theta_i}| + (N(\theta_i, \theta_{i+1}) + 2)^{2-\frac{1}{p}}(K_1 + |x_{\theta_i}|) \\
&\le e^{2(N(0,\theta_i)+\mathcal{N}(0,\theta_i))}(2K_1 + |x_0|) + e^{2(N(\theta_i,\theta_{i+1})+1)}\frac{1}{2}\left(K_1 + e^{2(N(0,\theta_i)+\mathcal{N}(0,\theta_i))}(2K_1 + |x_0|)\right).
\end{aligned}$$

Note that every argument of the exponential functions in the above inequality is bigger or equal to two. Hence

$$\begin{aligned}
|x_{\theta_{i+1}}| &\le e^{2(N(0,\theta_i)+\mathcal{N}(0,\theta_i))}e^{2(N(\theta_i,\theta_{i+1})+1)}\left(\frac{1}{2}K_1 + \frac{1}{2}|x_0| + \frac{1}{2}K_1 + K_1 + \frac{1}{2}|x_0|\right) \\
&\le e^{2(N(0,\theta_{i+1})+\mathcal{N}(0,\theta_{i+1}))}(2K_1 + |x_0|)
\end{aligned}$$

and the statement holds for all $i \in \{0, \dots, \mathcal{N}-1\}$. Consequently for $0 \le i \le \mathcal{N}-1$, we have

$$\begin{aligned}
|x|_{p,\theta_i,\theta_{i+1}} &\le (N(\theta_i, \theta_{i+1}) + 2)^{2-\frac{1}{p}}(K_1 + |x_{\theta_i}|) \\
&\le e^{2(N(\theta_i,\theta_{i+1})+1)}\left(K_1 + e^{2(N(0,\theta_i)+\mathcal{N}(0,\theta_i))}(2K_1 + |x_0|)\right) \\
&\le e^{2(N(0,\theta_{i+1})+\mathcal{N}(0,\theta_{i+1}))}(3K_1 + |x_0|).
\end{aligned}$$

These considerations enable us to finish the proof. We have

$$
\begin{aligned}
|x|_{p,0,T} &\leq \left(\mathcal{N}(0,T)^{p-1} \sum_{i=0}^{\mathcal{N}(0,T)-1} |x|^p_{p,\theta_i,\theta_{i+1}} \right)^{\frac{1}{p}} \\
&\leq \mathcal{N}(0,T)^{1-\frac{1}{p}} (3K_1 + |x_0|) \left(\sum_{i=0}^{\mathcal{N}(0,T)-1} e^{2p(N(0,\theta_{i+1})+\mathcal{N}(0,\theta_{i+1}))} \right)^{\frac{1}{p}} \\
&\leq \mathcal{N}(0,T)(3K_1 + |x_0|) e^{2(N(0,T)+\mathcal{N}(0,T))}.
\end{aligned}
$$

Now keep in mind that $\mathcal{N}(0,T) \leq 2N(0,T)+1$ by construction of Π^{c}, this yields

$$
|x|_{p,0,T} \leq (2N(0,T)+1)(3K_1 + |x_0|) e^{6N(0,T)+2}.
$$

Taking Lemma 2.19 into account, we know that

$$
N(0,T) \leq 2^{p-1} K_2^{-p} \left(T^p + |w|^p_{p,0,T} \right),
$$

which yields

$$
|x|_{p,0,T} \leq (3K_1 + |x_0|) \left(2^p K_2^{-p} \left(T^p + |w|^p_{p,0,T} \right) + 1 \right) \exp\left(2^p 3 K_2^{-p} \left(T^p + |w|^p_{p,0,T} \right) + 2 \right).
$$

□

Now we can use these results and repeat the same steps as in the last subsection, but this time on the linear equations.

Lemma 4.9. *We have for a given $u \in \mathcal{U}$ and for almost every $\omega \in \Omega$*

$$
|y^n(\omega)|_{p,0,T} \leq (3+L) \left(2^{3p} (3C_1 m_1)^p \left(T^p + |w|^p_{p,0,T} \right) + 1 \right) \exp\left(2^{3p} 3 (3C_1 m_1)^p \left(T^p + |w(\omega)|^p_{p,0,T} \right) + 2 \right)
$$

and

$$
\begin{aligned}
&\|y^{n,u}(\omega)\|_{\infty,0,T} \\
&\leq L + (3+L) \left(2^{3p} (3C_1 m_1)^p \left(T^p + |w(\omega)|^p_{p,0,T} \right) + 1 \right) \exp\left(2^{3p} 3 (3C_1 m_1)^p \left(T^p + |w(\omega)|^p_{p,0,T} \right) + 2 \right) \\
&:= C_{y^n}(\omega).
\end{aligned}
$$

For $l \geq 1$, we have $y^{n,u} \in L^l_{\mathbb{F}}(\Omega, C([0,T]), \mathbb{R}^{n_1})$ with

$$
\mathrm{E}\left[\|y^{n,u}\|^l_{\infty,0,T} \right] \leq \mathrm{E}\left[C^l_{y^n} \right] := D_{y^n,l}.
$$

Furthermore, we have for almost every $\omega \in \Omega$, $u \in \mathcal{U}$ that

$$\|y^{n,u}(\omega) - y^{n,u}_{t_{\overline{n}(\cdot)}}(\omega)\|_{\infty,0,T} \leq \max_{i=0,\dots,n-1} |y^{n,u}|_{p,t_i,t_{i+1}} \leq Lm_1(1 + C_{y_n}(\omega))\delta(\omega) \tag{4.41}$$

and consequently

$$\mathrm{E}\left[\|y^{n,u}_{\cdot} - y^{n,u}_{t_{\overline{n}(\cdot)}}\|^l_{\infty,0,T}\right] \leq (Lm_1)^l \left(1 + D^{\frac{1}{2}}_{y^n,2l}\right) \delta^l_{1,2l}$$

for all $l \geq 1$.

Proof. Let $\mathcal{A} \subset \Omega$, such that $P(\mathcal{A}) = 0$ and $w.(\omega)$ is continuous and of bounded p-variation ($p \in (1,2)$) for every $\omega \in \mathcal{A}^c$. First we show that for a given $u \in \mathcal{U}$ and for all $\omega \in \mathcal{A}^c$ the paths $y^{n,u}_{\cdot}(\omega)$ are elements of $C^p([0,T], \mathbb{R}^{n_1 \times d})$. Let $\omega \in \mathcal{A}^c$, $u \in \mathcal{U}$ be arbitrary, for notational simplicity we leave out the direct dependence of the involved processes on ω and u. We have that y^n satisfies the equation

$$\begin{aligned} y^n_t = y^n_0 &+ \int_0^t b_x\left(t_{\overline{n}(r)}, \xi^n_{t_{\overline{n}(r)}}, u\right) y^n_{t_{\overline{n}(r)}} + b_u\left(t_{\overline{n}(r)}, \xi^n_{t_{\overline{n}(r)}}, u\right) d_r \\ &+ \sum_{j=1}^{m_1} \int_0^t \sigma^j_x\left(t_{\overline{n}(r)}, \xi^n_{t_{\overline{n}(r)}}, u\right) y^n_{t_{\overline{n}(r)}} + \sigma^j_u\left(t_{\overline{n}(r)}, \xi^n_{t_{\overline{n}(r)}}, u\right) dw^j_r. \end{aligned}$$

Since w is continuous, the continuity of y^n follows directly. By Lemma 2.9, we can estimate

$$|y^n|^p_{p,0,T} = \leq n^{p-1} \sum_{i=0}^{n-1} |y^n|^p_{p,t_i,t_{i+1}}. \tag{4.42}$$

For a given $i \in \{0, \dots, n-1\}$ we take a look at the interval $[t_i, t_{i+1}]$. We have

$$\begin{aligned} |y^n|_{p,t_i,t_{i+1}} &\leq \left|b_x(t_i, \xi^n_{t_i}, u) y^n_{t_i} + b_u(t_i, \xi^n_{t_i}, u)\right| (t_{i+1} - t_i) \\ &\quad + \sum_{j=1}^{m_1} \left|\sigma^j_x(t_i, \xi^n_{t_i}, u) y^n_{t_i} + \sigma^j_x(t_i, \xi^n_{t_i}, u)\right| |w^j|_{p,t_i,t_{i+1}} \\ &\leq Lm_1(1 + |y^n_{t_i}|)(|t_{i+1} - t_i| + |w|_{p,t_i,t_{i+1}}), \end{aligned} \tag{4.43}$$

where we used that $\sum_{j=1}^{m_1} |w^j|_{p,t_i,t_{i+1}} \leq m_1 |w|_{p,t_i,t_{i+1}}$. Since

$$|y^n_{t_i}| \leq |y^n_{t_{i-1}}| + |y^n|_{p,t_{i-1},t_i}$$

and

$$|y_0| = |D\xi_0(u)| \leq L$$

we get inductively by (4.43), since $w \in C^p([0,T], \mathbb{R}^{m_1})$, that $|y^n|_{p,t_i,t_{i+1}} < \infty$ for every $i \in \{0, \dots, n-1\}$ and by (4.42), we have $y^n \in C^p([0,T], \mathbb{R}^{n_1 \times d})$. But this bound of y^n depends on n. We know that $\sigma^j_x(\cdot, \xi^n_{\cdot}, u)$ is an element of $C^q([0,T], \mathbb{R}^{n_1 \times n_1})$ and $\sigma^j_u(\cdot, \xi^n_{\cdot}, u)$ is an element of

$C^q([0,T],\mathbb{R}^{n_1\times d})$ for $q\in(2,\frac{p}{p-1})$ and every $j=1,\dots,m_2$ by Lemma 2.26. Hence $\sigma_x^j(\cdot,\xi_\cdot^n,u)y_\cdot^n+\sigma_u^j(\cdot,\xi_\cdot^n,u)$ is an element of $C^q([0,T],\mathbb{R}^{n_1\times d})$ for $q\in(2,\frac{p}{p-1})$ and every $j=1,\dots,m_2$. Notice that this function does not appear in the equation (4.6), but will be used in the estimation of $|y^n|_{p,0,T}$. Now we want to find an upper bound for y^n which is independent of n and u. Let t_i be a partition point of Π^{E} for $i\in\{0,\dots,n-1\}$ and $s\le t\in[t_i,t_{i+1}]$, we estimate similar to (4.43) that

$$\begin{aligned}|y_t^n-y_s^n| &\le Lm_1(1+|y_{t_i}^n|)(|t-s|+|w|_{p,s,t})\\ &\le Lm_1(1+\|y^n\|_{\infty,t_i,t_{i+1}})(|t-s|+|w|_{p,s,t}).\end{aligned}\tag{4.44}$$

Now let $0\le t_l\le s<t_{l+1}<\dots<t_{l+m}=t_{k-1}<t\le t_k\le T$ for $m\ge 1$ and $t_l,\dots,t_k\in\Pi^{\mathrm{E}}$, we estimate

$$|y_t^n-y_s^n|\le|y_t^n-y_{t_{k-1}}^n|+|y_{t_{k-1}}^n-y_{t_{l+1}}^n|+|y_{t_{l+1}}^n-y_s^n|,$$

where the second term vanishes for $m=1$. By (4.44), we obtain

$$\begin{aligned}|y_t^n-y_s^n|\le{}& Lm_1(1+\|y^n\|_{\infty,t_{k-1},t_k})(|t-t_{k-1}|+|w|_{p,t_{k-1},t})+|y_{t_{k-1}}^n-y_{t_{l+1}}^n|\\ &+Lm_1(1+\|y^n\|_{\infty,t_l,t_{l+1}})(|t_{l+1}-s|+|w|_{p,s,t_{l+1}}).\end{aligned}$$

For $m\ge 2$ the term $|y_{t_k}^n-y_{t_l}^n|$ can be decomposed by

$$\begin{aligned}|y_{t_{k-1}}^n-y_{t_{l+1}}^n|\le{}&\left|\sum_{i=l+1}^{k-2}(b_x(t_i,\xi_{t_i}^n,u)y_{t_i}^n+b_u(t_i,\xi_{t_i}^n,u))(t_{i+1}-t_i)\right|\\ &+\sum_{j=1}^{m_1}\left|\sum_{i=l+1}^{k-2}(\sigma_x^j(t_i,\xi_{t_i}^n,u)y_{t_i}^n+\sigma_u^j(t_i,\xi_{t_i}^n,u))(w_{t_{i+1}}^j-w_{t_i}^j)\right|\\ ={}&S_1+S_2.\end{aligned}$$

The sum S_1 can easily be estimated by the boundedness of b_x, b_u and the superadditivity of $\varphi(s,t)=|t-s|$ on $\Delta([0,T])$ and

$$S_1\le L(1+\|y^n\|_{\infty,t_{l+1},t_{k-1}})(t_{k-1}-t_{l+1}).$$

Taking Lemma 2.12 into account, we obtain for $q\in(2,\frac{p}{p-1})$

$$S_2\le\sum_{j=1}^{m_1}C_{p,q}\|\sigma_x^j(\cdot,\xi_\cdot^n,u)y_\cdot^n+\sigma_u^j(\cdot,\xi_\cdot^n,u)\|_{q,t_{l+1},t_{k-1}}|w^j|_{p,t_{l+1},t_{k-1}}$$

and by Lemma 2.10 and Lemma 2.26 i) used on $\sigma_x^j(\cdot, \xi_\cdot^n, u)$ and $\sigma_u^j(\cdot, \xi_\cdot^n, u)$, we get

$$\begin{aligned}
S_2 &\leq \sum_{j=1}^{m_1} C_{p,q} \left(2\|\sigma_x^j(\cdot, \xi_\cdot^n, u)\|_{q,t_{l+1},t_{k-1}} \|y^n\|_{q,t_{l+1},t_{k-1}} + \|\sigma_u^j(\cdot, \xi_\cdot^n, u)\|_{q,t_{l+1},t_{k-1}}\right) |w^j|_{p,t_{l+1},t_{k-1}} \\
&\leq 2m_1 L C_{p,q}(1 + \|y^n\|_{p,t_{l+1},t_{k-1}}) \left(1 + |t_{k-1} - t_{l+1}|^\beta + |\xi^n|_{p,t_{l+1},t_{k-1}}\right) |w|_{p,t_{l+1},t_{k-1}}.
\end{aligned}$$

We obtain

$$\begin{aligned}
&|y_{t_{k-1}}^n - y_{t_{l+1}}^n| \\
&\leq 2C_1 m_1 (1 + \|y^n\|_{p,t_{l+1},t_{k-1}})(1 + |\xi^n|_{p,t_{l+1},t_{k-1}})(|t_{k-1} - t_{l+1}| + |w|_{p,t_{l+1},t_{k-1}}),
\end{aligned}$$

where C_1 is defined in (2.23). Collecting all the terms, leads to

$$\begin{aligned}
&|y_t^n - y_s^n| \\
&\leq L m_1 (1 + \|y^n\|_{\infty,t_{k-1},t_k})(|t - t_{k-1}| + |w|_{p,t_{k-1},t}) + L m_1 (1 + \|y^n\|_{\infty,t_l,t_{l+1}})(|t_{l+1} - s| + |w|_{p,s,t_{l+1}}) \\
&\quad + 2C_1 m_1 (1 + \|y^n\|_{p,t_{l+1},t_{k-1}})(1 + |\xi^n|_{p,t_{l+1},t_{k-1}})(|t_{k-1} - t_{l+1}| + |w|_{p,t_{l+1},t_{k-1}}).
\end{aligned}$$

Since $t_l \leq s < t_{l+1} < \cdots < t_{l+m} = t_{k-1} < t \leq t_k$ this can be estimated by

$$\begin{aligned}
|y_t^n - y_s^n| &\leq L m_1 (1 + \|y^n\|_{p,t_l,t_k})(|t - s| + |w|_{p,s,t}) + L m_1 (1 + \|y^n\|_{p,t_l,t_k})(|t - s| + |w|_{p,s,t}) \\
&\quad + 2C_1 m_1 (1 + \|y^n\|_{p,t_l,t_k})(1 + |\xi^n|_{p,t_l,t_k})(|t - s| + |w|_{p,s,t}) \\
&\leq 3C_1 m_1 (1 + |y_{t_l}^n| + |y^n|_{p,t_l,t_k})(1 + |\xi^n|_{p,t_l,t_k})(|t - s| + |w|_{p,s,t}). \qquad (4.45)
\end{aligned}$$

Taking (4.44) and (4.45) into account, we have for all $l < k \in \{0, \ldots, n-1\}$ and $t_l \leq s < t \leq t_k$ that

$$|y_t^n - y_s^n| \leq 3C_1 m_1 (1 + |y_{t_l}^n| + |y^n|_{p,t_l,t_k})(1 + |\xi^n|_{p,t_l,t_k})(|t - s| + |w|_{p,s,t}).$$

By Lemma 2.6 this yields

$$|y^n|_{p,t_l,t_k} \leq 3C_1 m_1 (1 + |y_{t_l}^n| + |y^n|_{p,t_l,t_k})(1 + |\xi^n|_{p,t_l,t_k})(|t_k - t_l| + |w|_{p,t_l,t_k}).$$

Now we have for every interval $[t_l, t_k] \in [0, T]$ which satisfies

$$|t_k - t_l| + |w|_{p,t_l,t_k} \leq \frac{1}{12C_1 m_1}$$

that

$$|\xi^n|_{p,t_l,t_k} \leq 1,$$

by Remark 4.3 and therefore

$$|y^n|_{p,t_l,t_k} \leq 1 + |y^n_{t_l}|. \tag{4.46}$$

Hence by (4.43) and (4.46) we can use Lemma 4.8 with $a = Lm_1$, $K_1 = 1$, $K_2 = \frac{1}{12C_1m_1} \leq \frac{1}{Lm_1} = \frac{1}{a}$ and obtain the estimate

$$|y^n|_{p,0,T} \leq (3 + |y^n_0|) \left(2^{3p}(3C_1m_1)^p \left(T^p + |w|^p_{p,0,T}\right) + 1\right) \exp\left(2^{3p}3(3C_1m_1)^p \left(T^p + |w|^p_{p,0,T}\right) + 2\right).$$

By condition (H_1), we have $|y^n_0| = |D\xi_0(u)| \leq L$, this implies

$$|y^n|_{p,0,T} \leq (3 + L) \left(2^{3p}(3C_1m_1)^p \left(T^p + |w|^p_{p,0,T}\right) + 1\right) \exp\left(2^{3p}3(3C_1m_1)^p \left(T^p + |w|^p_{p,0,T}\right) + 2\right)$$

and consequently

$$\begin{aligned}|y^n|_{\infty,0,T} &\leq (3 + 2L) \left(2^{3p}(3C_1m_1)^p \left(T^p + |w|^p_{p,0,T}\right) + 1\right) \exp\left(2^{3p}3(3C_1m_1)^p \left(T^p + |w|^p_{p,0,T}\right) + 2\right)\\ &= C_{y^n}.\end{aligned}$$

Since ω was arbitrary in $\mathcal{A}^c$, the inequalities for the p-variation and uniform norm of y^n hold P-almost surely and for all $u \in \mathcal{U}$. The $\mathbb{F}$-adaptedness of y^n is a direct implication of its definition and the $\mathbb{F}$-adaptedness of ξ^n and w. Since w satisfies the exponential moment condition (2.48), we get

$$\mathrm{E}\left[\|y^n\|^l_{\infty,0,T}\right] \leq \mathrm{E}\left[C^l_{y^n}\right] := D_{y^n,l} < \infty.$$

Let $t \in [t_i, t_{i+1}]$ for some $i \in \{0, \ldots, n-1\}$, it follows by (4.43)

$$|y^n_t - y^n_{t_i}| \leq |y^n|_{p,t_i,t_{i+1}} \leq Lm_1(1 + |y^n_{t_i}|)(|t_{i+1} - t_i| + |w|_{p,t_i,t_{i+1}}) \leq Lm_1(1 + \|y^n\|_{\infty,0,T})\delta.$$

Hence, we have P-almost surely

$$\|y^n_\cdot - y^n_{t_{\overline{n}(\cdot)}}\|^l_{\infty,0,T} \leq \max_{i=0,\ldots,n-1} |y^n|^l_{p,t_i,t_{i+1}} \leq (Lm_1)^l(1 + C_{y^n})^l\delta^l,$$

where δ is defined in (4.1). Consequently by the Hölder inequality

$$\begin{aligned}E\left[\|y^n_\cdot - y^n_{t_{\overline{n}(\cdot)}}\|^l_{\infty,0,T}\right] &\leq (Lm_1)^l\mathrm{E}\left[(1 + C_{y^n})^{2l}\right]^{\frac{1}{2}} \mathrm{E}\left[\delta^{2l}\right]^{\frac{1}{2}}\\ &\leq (Lm_1)^l \left(1 + D^{\frac{1}{2}}_{y^n,2l}\right) \delta^l_{1,2l}.\end{aligned}$$

□

In the next theorem, we establish the convergence rate of the Euler Scheme y^n corresponding to the solution y of the pathwise linear stochastic Young differential equation from (4.4).

Theorem 4.10. *We have for a given $u \in \mathcal{U}$ and for almost every $\omega \in \Omega$*

$$\|y^u(\omega) - y^{n,u}(\omega)\|_{\infty,0,T} \leq K_y(\omega)\delta(\omega)^{2-p},$$

where the random variable K_y has moments of all orders and is independent of n and u. Furthermore, we have for all $l \geq 1$, that

$$E\left[\|y^u - y^{n,u}\|^l_{\infty,0,T}\right]^{\frac{1}{l}} \leq D_{K_y,2l}\delta^{2-p}_{1,2l}.$$

Proof. Let $\mathcal{A} \subset \Omega$, such that $P(\mathcal{A}) = 0$ and $w.(\omega)$ is continuous and of bounded p-variation for every $\omega \in \mathcal{A}^c$. Let $\omega \in \mathcal{A}^c$, $u \in \mathcal{U}$ be arbitrary, for notational simplicity leave out the direct dependence of the involved processes on ω and u. Let $s \leq t \in [t_i, t_{i+1}]$ for some $i \in \{0, \dots, n-1\}$ and define $\gamma_t = y_t - y^n_t$ for all $t \in [0,T]$, we have

$$\begin{aligned} y^n_t - y^n_s = &\left(b_x(t_i, \xi^n_{t_i}, u)y^n_{t_i} + b_u(t_i, \xi^n_{t_i}, u)\right)(t-s) \\ &+ \sum_{j=1}^{m_1} \left(\sigma^j_x(t_i, \xi^n_{t_i}, u)y^n_{t_i} + \sigma^j_u(t_i, \xi^n_{t_i}, u)\right)(w^j_t - w^j_s) \end{aligned}$$

and

$$\begin{aligned} y_t - y_s = &(b_x(s, \xi_s, u)y_s + b_u(s, \xi_s, u))(t-s) \\ &+ \sum_{j=1}^{m_1} \left(\sigma^j_x(s, \xi_s, u)y_s + \sigma^j_u(s, \xi_s, u)\right)\left(w^j_t - w^j_s\right) \\ &+ \int_s^t b_x(r, \xi_r, u)y_r + b_u(r, \xi_r, u) - b_x(s, \xi_s, u)y_s - b_u(s, \xi_s, u)\, dr \\ &+ \sum_{j=1}^{m_1} \int_s^t \sigma^j_x(r, \xi_r, u)y_r + \sigma^j_u(r, \xi_r, u) - \sigma^j_x(s, \xi_s, u)y_s - \sigma^j_u(s, \xi_s, u)\, dw^j_r. \end{aligned}$$

To simplify the notation we define $b^\xi_x(r) = b_x(r, \xi_r, u)$ and analogously $b^\xi_u(r)$, $\sigma^{\xi,j}_x(r)$ and $\sigma^{\xi,j}_u(r)$ for all $r \in [0,T]$ and $j = 1, \dots, m_1$. Furthermore define $b^n_x(r) := b_x(r, \xi^n_r, u)$ and analogously $b^n_u(r)$, $\sigma^{n,j}_x(r)$ and $\sigma^{n,j}_u(r)$ for all $r \in [0,T]$ and $j = 1, \dots, m_1$. We get

$$\begin{aligned} |\gamma_t - \gamma_s| \leq &\left|\left(b^\xi_x(s)y_s + b^\xi_u(s) - b^n_x(s)y^n_s - b^n_u(s)\right)(t-s)\right| \\ &+ \left|\left(b^n_x(s)y^n_s + b^n_u(s) - b^n_x(t_i)y^n_{t_i} - b^n_u(t_i)\right)(t-s)\right| \\ &+ \sum_{j=1}^{m_1} \left|\left(\sigma^{\xi,j}_x(s)y_s + \sigma^{\xi,j}_u(s) - \sigma^{n,j}_x(s)y^n_s - \sigma^{n,j}_u(s)\right)\left(w^j_t - w^j_s\right)\right| \end{aligned}$$

$$
\begin{aligned}
&+\sum_{j=1}^{m_1}\left|\left(\sigma_x^{n,j}(s)y_s^n+\sigma_u^{n,j}(s)-\sigma_x^{n,j}(t_i)y_{t_i}^n-\sigma_u^{n,j}(t_i)\right)\left(w_t^j-w_s^j\right)\right|\\
&+\left|\int_s^t b_x^\xi(r)y_r+b_u^\xi(r)-b_x^\xi(s)y_s-b_u^\xi(s)\,dr\right|\\
&+\sum_{j=1}^{m_1}\left|\int_s^t \sigma_x^{\xi,j}(r)y_r+\sigma_u^{\xi,j}(r)-\sigma_x^{\xi,j}(s)y_s-\sigma_u^{\xi,j}(s)\,dw_r^j\right|\\
=\;& I_1+I_2+I_3+I_4+I_5+I_6,
\end{aligned}
$$

where I_l stands for the term in the l-th line. Using conditions (H_2), (H_3^*) and (E_1), we obtain

$$
\begin{aligned}
I_1 &\le \left|(b_x^\xi(s)-b_x^n(s))y_s+b_x^n(s)(y_s-y_s^n)+(b_u^\xi(s)-b_u^n(s))\right||t-s|\\
&\le \left(L|\xi_s-\xi_s^n|(1+|y_s|)+L|y_s-y_s^n|\right)|t-s|\\
&\le \left(L\|\xi-\xi^n\|_{\infty,0,T}(1+\|y\|_{\infty,0,T})+L|\gamma_s|\right)|t-s|
\end{aligned}
$$

and

$$
\begin{aligned}
I_2 &\le \left|(b_x^n(s)-b_x^n(t_i))y_s^n+b_x^n(t_i)(y_s^n-y_{t_i}^n)+(b_u^n(s)-b_u^n(t_i))\right||t-s|\\
&\le \left(L(|s-t_i|^\beta+|\xi_s^n-\xi_{t_i}^n|)(1+|y_s^n|)+L|y_s^n-y_{t_i}^n|\right)|t-s|\\
&\le \left(L(|t_{i+1}-t_i|^\beta+|\xi^n|_{p,t_i,t_{i+1}})(1+\|y^n\|_{\infty,0,T})+L|y^n|_{p,t_i,t_{i+1}}\right)|t-s|.
\end{aligned}
$$

The estimates for I_3 and I_4 are completely analogue to I_1 and I_2, where we just use $\sum_{j=1}^{m_1}|w^j|_{p,s,t}\le m_1|w|_{p,s,t}$, which yields

$$
\begin{aligned}
I_3 &\le m_1\left(L|\xi_s-\xi_s^n|(1+|y_s|)+L|y_s-y_s^n|\right)|w|_{p,s,t}\\
&\le m_1\left(L\|\xi-\xi^n\|_{\infty,0,T}(1+\|y\|_{\infty,0,T})+L|\gamma_s|\right)|w|_{p,s,t}
\end{aligned}
$$

and

$$
\begin{aligned}
I_4 &\le m_1\left(L(|s-t_i|^\beta+|\xi_s^n-\xi_{t_i}^n|)(1+|y_s^n|)+L|y_s^n-y_{t_i}^n|\right)|w|_{p,s,t}\\
&\le m_1\left(L(|t_{i+1}-t_i|^\beta+|\xi^n|_{p,t_i,t_{i+1}})(1+\|y^n\|_{\infty,0,T})+L|y^n|_{p,t_i,t_{i+1}}\right)|w|_{p,s,t}.
\end{aligned}
$$

The estimation of I_5 is again similar to the previous calculation

$$
\begin{aligned}
I_5 &\le \int_s^t \left|(b_x^\xi(r)-b^\xi(s))y_r+b^\xi(s)(y_r-y_s)+(b_u^\xi(r)-b_u^\xi(s))\right|dr\\
&\le (L(|t-s|^\beta+|\xi|_{p,s,t})(1+\|y\|_{\infty,s,t})+L|y|_{p,s,t})|t-s|\\
&\le (L(|t_{i+1}-t_i|^\beta+|\xi|_{p,t_i,t_{i+1}})(1+\|y\|_{\infty,0,T})+L|y|_{p,t_i,t_{i+1}})|t-s|.
\end{aligned}
$$

For the estimation of I_6 we use the Love young estimate for $q \in (2, \frac{p}{p-1})$, Lemma 2.10 and Lemma 2.26 and obtain

$$\begin{aligned}
I_6 &\leq \sum_{j=1}^{m_1} C_{p,q} |\sigma_x^{\xi,j}(\cdot) y_\cdot + \sigma_u^{\xi,j}(\cdot)|_{q,s,t} |w^j|_{p,s,t} \\
&\leq \sum_{j=1}^{m_1} C_{p,q} \left(|\sigma_x^{\xi,j}|_{q,s,t} \|y\|_{\infty,s,t} + \|\sigma_x^{\xi,j}\|_{\infty,s,t} |y|_{q,s,t} + |\sigma_u^{\xi,j}|_{q,s,t} \right) |w^j|_{p,s,t} \\
&\leq m_1 (L(|t-s|^\beta + |\xi|_{p,s,t})(1 + \|y\|_{\infty,s,t}) + L|y|_{p,s,t}) |w|_{p,s,t} \\
&\leq m_1 (L(|t_{i+1} - t_i|^\beta + |\xi|_{p,t_i,t_{i+1}})(1 + \|y\|_{\infty,0,T}) + L|y|_{p,t_i,t_{i+1}}) |w|_{p,s,t}.
\end{aligned}$$

Collecting the terms and using the estimates from Lemma 4.9 and Remark 4.1, this yields

$$\begin{aligned}
|\gamma_t - \gamma_s| \leq C_{p,q} L m_1 \Big(& |\gamma_s| + \|\xi - \xi^n\|_{\infty,t_i,t_{i+1}} (1 + \|y\|_{\infty,0,T}) \\
& + (|t_{i+1} - t_i|^\beta + |\xi^n|_{p,t_i,t_{i+1}})(1 + \|y^n\|_{\infty,0,T}) + |y^n|_{p,t_i,t_{i+1}} \\
& + (|t_{i+1} - t_i|^\beta + |\xi|_{p,t_i,t_{i+1}})(1 + \|y\|_{\infty,0,T}) + |y|_{p,t_i,t_{i+1}} \Big) (|t-s| + |w|_{p,s,t}) \\
\leq C_{p,q} L m_1 \Big(& |\gamma_s| + \|\xi - \xi^n\|_{\infty,t_i,t_{i+1}} (1 + C_y) \\
& + (|t_{i+1} - t_i|^\beta + |\xi^n|_{p,t_i,t_{i+1}})(1 + C_{y^n}) + |y^n|_{p,t_i,t_{i+1}} \\
& + (|t_{i+1} - t_i|^\beta + |\xi|_{p,t_i,t_{i+1}})(1 + C_y) + |y|_{p,t_i,t_{i+1}} \Big) (|t-s| + |w|_{p,s,t}) \\
\leq C_{p,q} L m_1 (|t-s| + |w|_{p,s,t}) \Big(& |\gamma_s| + |y|_{p,t_i,t_{i+1}} + |y^n|_{p,t_i,t_{i+1}} \\
& + (\|\xi - \xi^n\|_{\infty,t_i,t_{i+1}} + 2|t_{i+1} - t_i|^\beta + |\xi|_{p,t_i,t_{i+1}} + |\xi^n|_{p,t_i,t_{i+1}})(1 + C_{y^n} + C_y) \Big). \quad (4.47)
\end{aligned}$$

We know from (2.26), (2.34), (4.7), (4.13) and (4.41)

$$\begin{aligned}
|\xi|_{p,t_i,t_{i+1}} &\leq C_1 (1 + C_\xi) \delta \\
|y|_{p,t_i,t_{i+1}} &\leq 2 C_1 m_1 (1 + C_y)(1 + C_\xi) \delta \\
|\xi^n|_{p,t_i,t_{i+1}} &\leq L \delta \\
\|\xi - \xi^n\|_{\infty,t_i,t_{i+1}} &\leq K_\xi \delta^{2-p} \\
|y^n|_{p,t_i,t_{i+1}} &\leq L m_1 (1 + C_{y^n}) \delta,
\end{aligned}$$

where C_1 is defined in (2.23). Furthermore we can estimate

$$|t_{i+1} - t_i|^\beta \leq (|t_{i+1} - t_i| + |w|_{p,t_i,t_{i+1}})^\beta \leq \delta^\beta.$$

Inserting all the estimates in (4.47) yields

$$
\begin{aligned}
|\gamma_t-\gamma_s| &\le C_{p,q}Lm_1\bigg(|\gamma_s| + 2C_1m_1(1+C_y)(1+C_\xi)\delta + Lm_1(1+C_{y^n})\delta \\
&\quad + (K_\xi\delta^{2-p} + 2\delta^\beta + C_1(1+C_\xi)\delta + L\delta)(1+C_{y^n}+C_y)\bigg)(|t-s|+|w|_{p,s,t}) \\
&\le C_{p,q}Lm_1\bigg(|\gamma_s| + \delta 2C_1m_1((1+C_y)(1+C_\xi)+1+C_{y^n}) \\
&\quad + (\delta^{2-p}+\delta^\beta+2\delta)2C_1K_\xi(1+C_\xi)(1+C_{y^n}+C_y)\bigg)(|t-s|+|w|_{p,s,t}) \\
&\le C_{p,q}Lm_1\bigg(|\gamma_s| + D_1(\omega)(\delta^{2-p}+\delta^\beta+3\delta)\bigg)(|t-s|+|w|_{p,s,t}),
\end{aligned}
$$

where

$$D_1(\omega) := 2C_1m_1K_\xi(1+C_\xi)(2+C_{y^n}+C_y).$$

Now we take a closer look at the term

$$\delta^{2-p}+\delta^\beta+3\delta.$$

We can estimate since $\beta \ge \frac{1}{p}$

$$\delta^{2-p}+\delta^\beta+3\delta \le \delta^{2-p}\left(1+\delta^{\beta+p-2}+3\delta^{p-1}\right)$$

and since $\delta \le (1+T+|w|_{p,0,T}) := C_w(\omega)$, where C_w has finite moments of all orders, we get

$$\delta^{2-p}+3\delta+\delta^\beta \le (1+4C_w(\omega))\,\delta^{2-p}.$$

We conclude

$$|\gamma_t-\gamma_s| \le C_1m_1(\|\gamma\|_{\infty,s,t} + D_2(\omega)\delta^{2-p})(|t-s|+|w|_{p,s,t}),$$

where

$$D_2(\omega) := D_1(\omega)\left(1+4C_w(\omega)\right).$$

Now let $t_0 \le t_{l-1} \le s < t_l < \cdots < t_{l+m} = t_k < t \le t_{k+1} \le T$ with $m \ge 0$. Then we have

$$|\gamma_t-\gamma_s| \le |\gamma_t-\gamma_{t_k}| + |\gamma_{t_k}-\gamma_{t_l}| + |\gamma_{t_l}-\gamma_s|. \tag{4.48}$$

The first and third term can be estimated by the previous considerations, which yields

$$
\begin{aligned}
|\gamma_t-\gamma_{t_k}| &\le C_1m_1(\|\gamma\|_{\infty,t_k,t} + D_2(\omega)\delta^{2-p})(|t-t_k|+|w|_{p,t_k,t}) \\
&\le C_1m_1(|\gamma\|_{\infty,s,t} + D_2(\omega)\delta^{2-p})(|t-s|+|w|_{p,s,t}) \\
|\gamma_{t_l}-\gamma_s| &\le C_1m_1(\|\gamma\|_{\infty,s,t_l} + D_2(\omega)\delta^{2-p})(|t_l-s|+|w|_{p,s,t_l})
\end{aligned} \tag{4.49}
$$

$$\leq C_1 m_1(\|\gamma\|_{\infty,s,t} + D_2(\omega)\delta^{2-p})(|t-s| + |w|_{p,s,t}). \tag{4.50}$$

For $m \geq 1$ the second term in (4.48) does not vanish and we get

$$\begin{aligned}
&|y_{t_k} - y^n_{t_k} - y_{t_l} + y^n_{t_l}| \\
&\leq \left|\sum_{i=l}^{k-1} \left(b^\xi_x(t_i)y_{t_i} + b^\xi_u(t_i) - b^n_x(t_i)y^n_{t_i} - b^n_u(t_i)\right)(t_{i+1} - t_i)\right| \\
&\quad + \sum_{j=1}^{m_1}\left|\sum_{i=l}^{k-1}\left(\sigma^{\xi,j}_x(t_i)y_{t_i} + \sigma^{\xi,j}_u(t_i) - \sigma^{n,j}_x(t_i)y^n_{t_i} - \sigma^{n,j}_u(t_i)\right)(w^j_{t_{i+1}} - w^j_{t_i})\right| \\
&\quad + \left|\sum_{i=l}^{k-1}\left(\int_{t_i}^{t_{i+1}} b^\xi_x(r)y_r + b^\xi_u(r) - b^\xi_x(t_i)y_{t_i} - b^\xi_u(t_i)\,dr\right.\right. \\
&\quad \left.\left. + \sum_{j=1}^{m_1}\int_{t_i}^{t_{i+1}} \sigma^{\xi,j}_x(r)y_r + \sigma^{\xi,j}_u(r) - \sigma^{\xi,j}_x(t_i)y_{t_i} - \sigma^{\xi,j}_u(t_i)\,dw^j_r\right)\right| \\
&= I_1 + I_2 + I_3.
\end{aligned}$$

To estimate the term I_1 we use the condition (H_2) and take a look at the term

$$\begin{aligned}
&\left|b^\xi_x(t_i)y_{t_i} + b^\xi_u(t_i) - b^n_x(t_i)y^n_{t_i} - b^n_u(t_i)\right| \\
&\leq \left|b^\xi_x(t_i) - b^n_x(t_i)\right||y_{t_i}| + |b^n_x(t_i)|\,|\gamma_{t_i}| + \left|b^\xi_u(t_i) - b^n_u(t_i)\right| \\
&\leq L\|\xi - \xi^n\|_{\infty,0,T}(1 + \|y\|_{\infty,0,T}) + L\|\gamma\|_{\infty,t_l,t_k}.
\end{aligned}$$

Taking (4.13) and Remark 4.1 into account, this yields

$$\leq L(K_\xi(1 + C_y)\delta^{2-p} + \|\gamma\|_{\infty,t_l,t_k})$$

for all $i = l, \ldots, k-1$. Inserting that into I_1, we obtain

$$\begin{aligned}
I_1 &\leq L\sum_{i=l}^{k-1}\left(K_\xi(1 + C_y)\delta^{2-p} + \|\gamma\|_{\infty,t_l,t_k}\right)(t_{i+1} - t_i) \\
&\leq L\left(K_\xi(1 + C_y)\delta^{2-p} + \|\gamma\|_{\infty,t_l,t_k}\right)(t_k - t_l).
\end{aligned}\tag{4.51}$$

Using Lemma 2.12 similar to the last lemma with $q \in (2, \frac{p}{p-1})$ and Lemma 2.10, yields

$$\begin{aligned}
I_2 &\leq C_{p,q}\sum_{j=1}^{m_1}\|\sigma^{\xi,j}_x(\cdot)y_\cdot + \sigma^{\xi,j}_u(\cdot) - \sigma^{n,j}_x(\cdot)y^n_\cdot - \sigma^{n,j}_u(\cdot)\|_{q,t_l,t_k}|w|_{p,t_l,t_k} \\
&= C_{p,q}\sum_{j=1}^{m_1}\|(\sigma^{\xi,j}_x(\cdot) - \sigma^{n,j}_x(\cdot))y_\cdot + \sigma^{n,j}_x(\cdot)(y_\cdot - y^n_\cdot) + \sigma^{\xi,j}_u(\cdot) - \sigma^{n,j}_u(\cdot)\|_{q,t_l,t_k}|w^j|_{p,t_l,t_k}
\end{aligned}$$

$$
\begin{aligned}
&\le C_{p,q} \sum_{j=1}^{m_1} \Big(2\|\sigma_x^{\xi,j}(\cdot) - \sigma_x^{n,j}(\cdot)\|_{q,t_l,t_k} \|y\|_{q,t_l,t_k} + 2\|\sigma_x^{n,j}\|_{q,t_l,t_k} \|\gamma\|_{q,t_l,t_k} \\
&\quad + \|\sigma_u^{\xi,j}(\cdot) - \sigma_u^{n,j}(\cdot)\|_{q,t_l,t_k} \Big) |w^j|_{p,t_l,t_k}.
\end{aligned}
$$

Taking Lemma 2.26 i) and iii) into account to estimate the q-variation norm of the partial derivatives of the coefficient function σ , we obtain

$$
\begin{aligned}
&\le C_{p,q} \sum_{j=1}^{m_1} \Big(2 \left(L(1 + |t_k - t_l|^\beta + |\xi^n|_{q,t_l,t_k} + |\xi|_{q,t_l,t_k}) \|\xi - \xi^n\|_{\infty,0,T} + L|\xi_\cdot - \xi^n_\cdot|_{q,s,t} \right) (1 + \|y\|_{p,0,T}) \\
&\quad + 2L(1 + |t_k - t_l|^\beta + |\xi^n|_{q,t_l,t_k}) \|\gamma\|_{q,t_l,t_k} \Big) |w^j|_{p,t_l,t_k} \\
&\le 2LC_{p,q} m_1 \Big(\left((1 + T^\beta + |\xi^n|_{p,0,T} + |\xi|_{p,0,T}) \|\xi - \xi^n\|_{\infty,0,T} + |\xi_\cdot - \xi^n_\cdot|_{p,0,T} \right) (1 + \|y\|_{p,0,T}) \\
&\quad + (1 + |t_k - t_l|^\beta + |\xi^n|_{p,t_l,t_k}) \|\gamma\|_{p,t_l,t_k} \Big) |w|_{p,t_l,t_k}.
\end{aligned}
$$

By Remark 4.1 and (4.13), this yields

$$
\begin{aligned}
I_2 &\le 2C_1 m_1 \Big(\left((1 + C_{\xi^n} + C_\xi) K_\xi \delta^{2-p} + K_\xi \delta^{2-p} \right) (1 + C_y) \\
&\quad + (1 + |\xi^n|_{p,t_l,t_k}) \|\gamma\|_{p,t_l,t_k} \Big) |w|_{p,t_l,t_k} \\
&\le 2C_1 m_1 \left((2 + C_{\xi^n} + C_\xi)(1 + C_y) K_\xi \delta^{2-p} + (1 + |\xi^n|_{p,t_l,t_k}) \|\gamma\|_{p,t_l,t_k} \right) |w|_{p,t_l,t_k} \\
&= \left(D_3(\omega) \delta^{2-p} + 2C_1 m_1 (1 + |\xi^n|_{p,t_l,t_k}) \|\gamma\|_{p,t_l,t_k} \right) |w|_{p,t_l,t_k},
\end{aligned}
\tag{4.52}
$$

where

$$
D_3(\omega) := 2C_1 m_1 (2 + C_{\xi^n} + C_\xi)(1 + C_y) K_\xi.
$$

The estimation of I_3 will be carried out with the Love-Young inequality and the Lipschitz and Hölder condition of the coefficient function b

$$
\begin{aligned}
I_3 &\le \sum_{i=l}^{k-1} \int_{t_i}^{t_{i+1}} \left| b_x^\xi(r) y_r + b_u^\xi(r) - b_x^\xi(t_i) y_{t_i} - b_u^\xi(t_i) \right| dr \\
&\quad + \sum_{i=l}^{k-1} \left| \int_{t_i}^{t_{i+1}} \sigma_x^\xi(r) y_r + \sigma_u^\xi(r) - \sigma_x^\xi(t_i) y_{t_i} - \sigma_u^\xi(t_i)\, dw_r \right|.
\end{aligned}
\tag{4.53}
$$

We first estimate for $r \in [t_i, t_{i+1}]$, using (H_2) and (E_1) the term

$$
\begin{aligned}
&\left| b_x^\xi(r) y_r + b_u^\xi(r) - b_x^\xi(t_i) y_{t_i} - b_u^\xi(t_i) \right| \\
&\le \left| b_x^\xi(r) - b_x^\xi(t_i) \right| |y_r| + \left| b_x^\xi(t_i) \right| |y_r - y_{t_i}| + \left| b_u^\xi(r) - b_u^\xi(t_i) \right|
\end{aligned}
$$

$$
\begin{aligned}
&\le L(|r-t_i|^\beta + |\xi_r - \xi_{t_i}|)(1+\|y\|_{\infty,t_i,t_{i+1}}) + L|y|_{p,t_i,t_{i+1}} \\
&\le L(|t_{i+1}-t_i|^\beta + |\xi|_{p,t_i,t_{i+1}})(1+\|y\|_{\infty,0,T}) + L|y|_{p,t_i,t_{i+1}} \\
&\le L(1+C_y)(|t_{i+1}-t_i|^\beta + |\xi|_{p,t_i,t_{i+1}}) + L|y|_{p,t_i,t_{i+1}}.
\end{aligned}
$$

Inserting this into the integral in (4.53), we obtain for the first sum

$$
\begin{aligned}
&\sum_{i=l}^{k-1} \int_{t_i}^{t_{i+1}} \left| b_x^\xi(r) y_r + b_u^\xi(r) - b_x^\xi(t_i) y_{t_i} - b_u^\xi(t_i) \right| dr \\
&\le L(1+C_y) \sum_{i=l}^{k-1} (|t_{i+1}-t_i|^\beta + |\xi|_{p,t_i,t_{i+1}} + |y|_{p,t_i,t_{i+1}})(t_{i+1}-t_i).
\end{aligned}
$$

The second sum can be estimated using the Love-Young estimate and Lemma 2.10, we have for $q \in (2, \frac{p}{p-1})$

$$
\begin{aligned}
&\sum_{j=1}^{m_1} \sum_{i=l}^{k-1} \left| \int_{t_i}^{t_{i+1}} \sigma_x^{\xi,j}(r) y_r + \sigma_u^{\xi,j}(r) - \sigma_x^{\xi,j}(t_i) y_{t_i} - \sigma_u^{\xi,j}(t_i)\, dw_r^j \right| \\
&\le C_{p,q} \sum_{j=1}^{m_1} \sum_{i=l}^{k-1} \left| \sigma_x^{\xi,j}(\cdot) y_\cdot + \sigma_u^{\xi,j}(\cdot) \right|_{q,t_i,t_{i+1}} |w^j|_{p,t_i,t_{i+1}} \\
&\le C_{p,q} \sum_{j=1}^{m_1} \sum_{i=l}^{k-1} \left(|\sigma_x^{\xi,j}|_{q,t_i,t_{i+1}} \|y\|_{\infty,t_i,t_{i+1}} + |\sigma_x^{\xi,j}|_{\infty,t_i,t_{i+1}} |y|_{p,t_i,t_{i+1}} + |\sigma_u^{\xi,j}(\cdot)|_{q,t_i,t_{i+1}} \right) |w^j|_{p,t_i,t_{i+1}}.
\end{aligned}
$$

Using Lemma 2.26 i) for $\sigma_x^{\xi,j}$ and $\sigma_u^{\xi,j}$ for $j = 1, \dots, m_1$ implies

$$
\begin{aligned}
&\sum_{j=1}^{m_1} \sum_{i=l}^{k-1} \left| \int_{t_i}^{t_{i+1}} \sigma_x^{\xi,j}(r) y_r + \sigma_u^{\xi,j}(r) - \sigma_x^{\xi,j}(t_i) y_{t_i} - \sigma_u^{\xi,j}(t_i)\, dw_r^j \right| \\
&\le C_{p,q} \sum_{j=1}^{m_1} \sum_{i=l}^{k-1} \left(L(|t_{i+1}-t_i|^\beta + |\xi|_{p,t_i,t_{i+1}})(1+C_y) + L|y|_{p,t_i,t_{i+1}} \right) |w^j|_{p,t_i,t_{i+1}} \\
&\le LC_{p,q} m_1 (1+C_y) \sum_{i=l}^{k-1} \left(|t_{i+1}-t_i|^\beta + |\xi|_{p,t_i,t_{i+1}} + |y|_{p,t_i,t_{i+1}} \right) |w|_{p,t_i,t_{i+1}}.
\end{aligned}
$$

Combining the previous results leads to

$$
\begin{aligned}
I_3 &\le LC_{p,q} m_1 (1+C_y) \sum_{i=l}^{k-1} (|t_{i+1}-t_i|^\beta + |\xi|_{p,t_i,t_{i+1}} + |y|_{p,t_i,t_{i+1}})(|t_{i+1}-t_i| + |w|_{p,t_i,t_{i+1}}) \\
&\le m_1(1+C_y) \Bigg(LC_{p,q} \sum_{i=l}^{k-1} (|t_{i+1}-t_i|^\beta + |\xi|_{p,t_i,t_{i+1}})(|t_{i+1}-t_i| + |w|_{p,t_i,t_{i+1}})
\end{aligned}
$$

$$
\left. + \sum_{i=l}^{k-1} LC_{p,q} |y|_{p,t_i,t_{i+1}} (|t_{i+1} - t_i| + |w|_{p,t_i,t_{i+1}}) \right)
$$

$$
= m_1(1 + C_y)(I_{31} + I_{32}).
$$

The estimation of I_{31} is completely analogous to the estimation of I_3 in the proof of Theorem 4.4. We obtain

$$
I_{31} \leq \delta^{2-p} D_4(\omega) \left(\varphi_1(t_l, t_k)^{\frac{1}{p}} + \varphi_2(t_l, t_k)^{\frac{1}{p}} \right),
$$

where

$$
D_4(\omega) = C_1^2 (1 + C_\xi) \left(\varphi_1(0, T)^{1-\frac{1}{p}} C_w^{p\beta - 1} + \varphi_2(0, T)^{1-\frac{1}{p}} \right)
$$

and

$$
\varphi_1(s, t) = |t - s|^\beta (|t - s|^p + |w|_{p,s,t}^p)^{1-\beta}
$$

$$
\varphi_2(s, t) = |t - s|^p + |w|_{p,s,t}^p.
$$

For the estimation of I_{32} consider (2.34), which gives

$$
|y|_{p,t_i,t_{i+1}} \leq C_1 m_1 (1 + C_\xi)(1 + C_y)(|t_{i+1} - t_i| + |w|_{p,t_i,t_{i+1}}).
$$

This yields

$$
\begin{aligned}
I_{32} &\leq C_{p,q} L C_1 m_1 (1 + C_y)(1 + C_\xi) \sum_{i=l}^{k-1} (|t_{i+1} - t_i| + |w|_{p,t_i,t_{i+1}})^2 \\
&\leq C_1^2 m_1 (1 + C_y)(1 + C_\xi) \sum_{i=l}^{k-1} \varphi_2(t_i, t_{i+1})^{\frac{2}{p}} \\
&\leq C_1^2 m_1 (1 + C_y)(1 + C_\xi) \varphi_2(0, T)^{1-\frac{1}{p}} \delta^{2-p} \varphi_2(t_l, t_k)^{\frac{1}{p}}.
\end{aligned}
$$

Combining the estimates for I_{31} and I_{32}, we get

$$
\begin{aligned}
I_3 &\leq m_1(1 + C_y) \Bigg(\delta^{2-p} D_4(\omega) \left(\varphi_1(t_l, t_k)^{\frac{1}{p}} + \varphi_2(t_l, t_k)^{\frac{1}{p}} \right) \\
&\qquad + C_1^2 m_1 (1 + C_y)(1 + C_\xi) \varphi_2(0, T)^{1-\frac{1}{p}} \delta^{2-p} \varphi_2(t_l, t_k)^{\frac{1}{p}} \Bigg) \\
&\leq D_5(\omega) \delta^{2-p} \left(\varphi_1(t_l, t_k)^{\frac{1}{p}} + \varphi_2(t_l, t_k)^{\frac{1}{p}} \right),
\end{aligned} \tag{4.54}
$$

where

$$
D_5(\omega) := m_1(1 + C_y) \Bigg(D_4(\omega) + C_1^2 m_1 (1 + C_y)(1 + C_\xi) \varphi_2(0, T)^{1-\frac{1}{p}} \Bigg).
$$

Putting all the terms (4.51), (4.52) and (4.54) together, we obtain

$$\begin{aligned}|y_{t_k}-y^n_{t_k}-y_{t_l}+y^n_{t_l}| &\le L\left(K_\xi(1+C_y)\delta^{2-p}+\|\gamma\|_{\infty,t_l,t_k}\right)(t_k-t_l)\\&\quad+\left(D_3(\omega)\delta^{2-p}+2C_1m_1(1+|\xi^n|_{p,t_l,t_k})\|\gamma\|_{p,t_l,t_k}\right)|w|_{p,t_l,t_k}\\&\quad+D_5(\omega)\delta^{2-p}\left(\varphi_1(t_l,t_k)^{\frac{1}{p}}+\varphi_2(t_l,t_k)^{\frac{1}{p}}\right)\\&\le \delta^{2-p}D_6(\omega)\left(|t_k-t_l|+|w|_{p,t_l,t_k}+\varphi_1(t_l,t_k)^{\frac{1}{p}}+\varphi_2(t_l,t_k)^{\frac{1}{p}}\right)\\&\quad+2C_1m_1(1+|\xi^n|_{p,t_l,t_k})\|\gamma\|_{p,t_l,t_k}(|t_k-t_l|+|w|_{p,t_l,t_k}).\end{aligned}$$

Notice that $K_\xi\ge 1$ (seen in the proof of Theorem 4.4) such that

$$\begin{aligned}D_6(\omega)&:=2m_1^2C_1^2(1+C_y)^2(2+C_\xi+C_{\xi^n})K_\xi\left(1+\varphi_1(0,T)^{\frac{1}{p}}C_w^{p\beta-1}+\varphi_2(0,T)^{\frac{1}{p}}\right)\\&\ge\max\left\{LK_\xi(1+C_y),D_3(\omega),D_5(\omega)\right\}.\end{aligned}$$

Coming back to (4.48) combined with (4.49) and (4.50), this yields for all $[s,t]\subset[0,T]$

$$\begin{aligned}|\gamma_t-\gamma_s|&\le 2C_1m_1(\|\gamma\|_{\infty,s,t}+D_2(\omega)\delta^{2-p})(|t-s|+|w|_{p,s,t})\\&\quad+2C_1m_1\left(1+|\xi^n|_{p,s,t}\right)\|\gamma\|_{p,s,t}(|t-s|+|w|_{p,s,t})\\&\quad+D_6(\omega)\delta^{2-p}\left(\varphi_1(s,t)^{\frac{1}{p}}+\varphi_2(s,t)^{\frac{1}{p}}+|t-s|+|w|_{p,s,t}\right)\\&\le 4C_1m_1\left(1+|\xi^n|_{p,s,t}\right)\|\gamma\|_{p,s,t}(|t-s|+|w|_{p,s,t})\\&\quad+D_7(\omega)\delta^{2-p}\left(\varphi_1(s,t)^{\frac{1}{p}}+\varphi_2(s,t)^{\frac{1}{p}}+|t-s|+|w|_{p,s,t}\right),\end{aligned}$$

where

$$\begin{aligned}D_7(\omega)&:=8m_1^2C_1^2(1+C_y+C_{y^n})^2(2+C_\xi+C_{\xi^n})(1+4D_w(\omega))K_\xi\left(1+\varphi_1(0,T)^{\frac{1}{p}}C_w^{p\beta-1}+\varphi_2(0,T)^{\frac{1}{p}}\right)\\&\ge 2\max\{2C_1m_1D_2(\omega),D_6(\omega)\}.\end{aligned}$$

This means that for all $[r,v]\subset[s,t]\subset[0,T]$ we have

$$\begin{aligned}|\gamma_r-\gamma_v|&\le 4C_1m_1\left(1+|\xi^n|_{p,s,t}\right)\|\gamma\|_{p,s,t}(|v-r|+|w|_{p,r,v})\\&\quad+D_7(\omega)\delta^{2-p}\left(\varphi_1(r,v)^{\frac{1}{p}}+\varphi_2(r,v)^{\frac{1}{p}}+|v-r|+|w|_{p,r,v}\right).\end{aligned}$$

With Lemma 2.6 this yields for every $[s,t]\subset[0,T]$.

$$\begin{aligned}|\gamma|_{p,s,t}&\le 4C_1m_1\left(1+|\xi^n|_{p,s,t}\right)\|\gamma\|_{p,s,t}(|t-s|+|w|_{p,s,t})\\&\quad+D_7(\omega)\delta^{2-p}\left(\varphi_1(s,t)^{\frac{1}{p}}+\varphi_2(s,t)^{\frac{1}{p}}+|t-s|+|w|_{p,s,t}\right)\\&\le 4C_1m_1(1+|\xi^n|_{p,s,t})(|\gamma_s|+|\gamma|_{p,s,t})(|t-s|+|w|_{p,s,t})+D_8(\omega)\delta^{2-p},\end{aligned}$$

where

$$D_8(\omega) := D_7(\omega)\left(\varphi_1(0,T)^{\frac{1}{p}} + \varphi_2(0,T)^{\frac{1}{p}} + (T + |w|_{p,0,T})\right)$$

is P-a.s. bounded independently of u and n and has finite moments of all orders. Taking Remark 4.3 into account, we can argue that for every $[s,t] \subset [0,T]$ such that

$$|t-s| + |w|_{p,s,t} \leq \frac{1}{16C_1 m_1}$$

we have

$$|\xi^n|_{p,s,t} \leq 1$$

and

$$|\gamma|_{p,s,t} \leq 2D_8(\omega)\delta^{2-p} + |\gamma_s|.$$

Now we can use Lemma 2.20 and the fact that $\gamma_0 = 0$, to obtain the estimate

$$\begin{aligned}|\gamma|_{\infty,0,T} \leq |\gamma|_{p,0,T} &\leq \left(2D_8(\omega)\delta^{2-p} + |\gamma_0|\right) e^{2^{5p}(m_1C_1)^p(T^p + |w|^p_{p,0,T})} \\ &\leq 2D_8(\omega)\delta^{2-p} e^{2^{5p}(m_1C_1)^p(T^p + |w|^p_{p,0,T})}.\end{aligned}$$

Hence,

$$\|\gamma\|_{\infty,0,T} \leq K_y(\omega)\delta^{2-p},$$

where

$$K_y(\omega) := 2D_8(\omega)e^{2^{5p}(m_1C_1)^p(T^p + |w|^p_{p,0,T})}.$$

The random variable $K_y(\omega)$ is bounded independently of u and n and has finite moments of all orders, since w satisfies the exponential moment condition. The same holds for δ, such that we can use the Hölder inequality and obtain for $l \geq 1$ and $\delta_{1,l} := \mathrm{E}[\delta^l]^{\frac{1}{l}}$ the estimate

$$\mathrm{E}\left[\|\gamma\|^l_{\infty,0,T}\right] \leq \mathrm{E}\left[K_y^{2l}\right]^{\frac{1}{2l}} \delta^{2-p}_{1,2l} \leq D_{K_y,2l}\delta^{2-p}_{1,2l}.$$

□

Having established the convergence rate for the discretization of the linear stochastic Young differential equation from (4.4), we now consider the solution $\hat{y}$ in (4.4) and its Euler approximation scheme $\hat{y}^n$ from (4.6).

Lemma 4.11. *We have for a given $u \in \mathcal{U}$ and $l \geq 2$, that there exists a constant $D_{\hat{y}^n,l}$ independent of u and n, such that*

$$\mathrm{E}\left[\|\hat{y}^{n,u}\|^l_{\infty,0,T}\right] \leq D_{\hat{y}^n,l}.$$

Furthermore, we obtain for $\delta_2 = \max_{i=0,\dots,n-1} |t_{i+1} - t_i|$ *and* $l \geq 2$, *that*

$$\mathrm{E}\left[\|\hat{y}^{n,u}_{\cdot} - \hat{y}^{n,u}_{t_{\overline{n}(\cdot)}}\|^l_{\infty,0,T}\right] \leq C(1 + D_{\hat{y}^n,l} + D_{y^n,l})\delta_2^{\frac{l}{2}}. \tag{4.55}$$

The constant $C > 0$ *only depends on* T, l, m_2 *and* L.

Proof. We omit the direct dependence of the involved processes on u for notational simplicity. Furthermore we define $\hat{b}^n_x(r) := \hat{b}_x(t_i, x^n_r, \xi^n_r, u)$ for every $r \in [0,T]$ and analogously $\hat{b}^n_z$, $\hat{b}^n_u$, $\hat{\sigma}^{n,j}_x$, $\hat{\sigma}^{n,j}_z$ and $\hat{\sigma}^{n,j}_u$ for each $j = 1, \dots, m_2$ and let C be a generic constant that is only dependent on L, l, m_2 and T. It is easy to see that the process

$$(\omega, t) \mapsto \sum_{j=1}^{m_2} \int_0^t \hat{\sigma}^{n,j}_x(t_{\overline{n}(r)})\hat{y}^n_{t_{\overline{n}(r)}}(\omega) + \hat{\sigma}^{n,j}_z(t_{\overline{n}(r)})y^n_{t_{\overline{n}(r)}}(\omega) + \hat{\sigma}^{n,j}_u(t_{\overline{n}(r)})\, dB^j(\omega)$$

for $t \in [0,T]$ is well defined and an $n_2 \times d$-dimensional matrix of $\mathbb{F}$-martingales. We have for $t \in [0,T]$ and $l \geq 2$ by the Jensen inequality and condition (B_3), since $\hat{y}^n_0 = \hat{y}_0 = Dx_0(u)$

$$\begin{aligned}
\mathrm{E}[\|\hat{y}^n\|^l_{\infty,0,t}] &= \mathrm{E}[\|\hat{y}^n\|^l_{\infty,0,t}] \\
&\leq C\Bigg(\mathrm{E}\left[|\hat{y}^n_0|^l\right] + \mathrm{E}\left[\sup_{s\in[0,t]}\left|\int_0^s \hat{b}^n_x(t_{\overline{n}(r)})\hat{y}^n_{t_{\overline{n}(r)}} + \hat{b}^n_z(t_{\overline{n}(r)})y^n_{t_{\overline{n}(r)}} + \hat{b}^n_u(t_{\overline{n}(r)})\, dr\right|^l\right] \\
&\quad + \sum_{j=1}^{m_2}\mathrm{E}\left[\sup_{s\in[0,t]}\left|\int_0^s \hat{\sigma}^{n,j}_x(t_{\overline{n}(r)})\hat{y}^n_{t_{\overline{n}(r)}} + \hat{\sigma}^{n,j}_z(t_{\overline{n}(r)})y^n_{t_{\overline{n}(r)}} + \hat{\sigma}^{n,j}_u(t_{\overline{n}(r)})\, dB^j_r\right|^l\right]\Bigg) \\
&= C(1 + I_1 + I_2).
\end{aligned}$$

We can estimate I_1 by the boundedness of $\hat{b}_x$, $\hat{b}_z$ $\hat{b}_u$, Lemma 4.9 and Fubinis theorem

$$\begin{aligned}
I_1 &\leq \mathrm{E}\left[\int_0^t \left|\hat{b}^n_x(t_{\overline{n}(r)})\hat{y}^n_{t_{\overline{n}(r)}} + \hat{b}^n_z(t_{\overline{n}(r)})y^n_{t_{\overline{n}(r)}} + \hat{b}^n_u(t_{\overline{n}(r)})\right|^l\, dr\right] \\
&\leq C\mathrm{E}\left[\int_0^t \left(1 + |\hat{y}^n_{t_{\overline{n}(r)}}| + |y^n_{t_{\overline{n}(r)}}|\right)^l\, dr\right] \\
&\leq C(1 + D_{y^n,l}) + C\int_0^t \mathrm{E}\left[\|\hat{y}^n\|^l_{\infty,0,r}\right]\, dr.
\end{aligned}$$

Similar calculations after the use of the Burkholder-Davis-Gundy inequality yield

$$\begin{aligned}
I_2 &\leq C\mathrm{E}\left[\left(\int_0^t \left|\hat{\sigma}^n_x(t_{\overline{n}(r)})\hat{y}^n_{t_{\overline{n}(r)}} + \hat{\sigma}^n_z(t_{\overline{n}(r)})y^n_{t_{\overline{n}(r)}} + \hat{\sigma}^n_u(t_{\overline{n}(r)})\right|^2\, dr\right)^{\frac{l}{2}}\right] \\
&\leq C(1 + D_{y^n,l}) + C\int_0^t \mathrm{E}\left[\|\hat{y}^n\|^l_{\infty,0,r}\right]\, dr.
\end{aligned}$$

Hence, we get

$$\mathrm{E}[\|\hat{y}\|^l_{\infty,0,t}] \leq C(1 + D_{y^n,l}) + C\int_0^t \mathrm{E}\left[\|\hat{y}^n\|^l_{\infty,0,r}\right] dr$$

and we conclude by the Gronwall inequality

$$\mathrm{E}\left[\|\hat{y}^n\|^l_{\infty,0,T}\right] \leq C(1 + D_{y^n,l})e^C := D_{\hat{y},l}. \tag{4.56}$$

Now let $t \in [t_i, t_{i+1}]$ for $i \in \{0, \dots, n-1\}$ and $l \geq 2$, we have

$$\begin{aligned}
&\mathrm{E}\left[|\hat{y}^n_t - \hat{y}^n_{t_i}|^l\right] \\
&\leq C\mathrm{E}\left[\left|\int_{t_i}^t \left(\hat{b}_x(t_i, x^n_{t_i}, \xi^n_{t_i}, u)\hat{y}^n_{t_i} + \hat{b}_z(t_i, x^n_{t_i}, \xi^n_{t_i}, u)y^n_{t_i} + \hat{b}_u(t_i, x^n_{t_i}, \xi^n_{t_i}, u)\right) d_r\right|^l\right] \\
&\quad + C\sum_{j=1}^{m_2} \mathrm{E}\left[\left|\int_{t_i}^t \left(\hat{\sigma}^j_x(t_i, x^n_{t_i}, \xi^n_{t_i}, u)\hat{y}^n_{t_i} + \hat{\sigma}^j_z(t_i, x^n_{t_i}, \xi^n_{t_i}, u)y^n_{t_i} + \hat{\sigma}^j_u(t_i, x^n_{t_i}, \xi^n_{t_i}, u)\right) dB^j_r\right|^l\right].
\end{aligned}$$

By the boundedness of the coefficient functions, the estimate (4.56), Lemma 4.9 and by the Burkholder-Davis-Gundy inequality, we get

$$\begin{aligned}
\mathrm{E}\left[|\hat{y}^n_t - \hat{y}^n_{t_i}|^l\right] &\leq C(t - t_i)^l\mathrm{E}\left[\|\hat{y}^n\|^l_{\infty,0,T} + \|y^n\|^l_{\infty,0,T} + 1\right] \\
&\quad + C(t - t_i)^{\frac{l}{2}}\mathrm{E}\left[\|\hat{y}^n\|^l_{\infty,0,T} + \|y^n\|^l_{\infty,0,T} + 1\right] \\
&\leq C(1 + D_{\hat{y}^n,l} + D_{y^n,l})\delta_2^{\frac{l}{2}}.
\end{aligned}$$

This yields the estimate (4.55) for $l \geq 2$. □

Theorem 4.12. *We have for a given $u \in \mathcal{U}$ and $l \geq 2$ that*

$$\mathrm{E}\left[\|\hat{y}^u - \hat{y}^{n,u}\|^l_{\infty,0,T}\right]^{\frac{1}{l}} \leq D_{K_{\hat{y}},l}\delta_{1,4l}^{(2-p)\wedge\frac{1}{2}},$$

where the constant $D_{K_{\hat{y}},l}$ is independent of u and n.

Proof. We have for $t \in [0, T]$

$$\begin{aligned}
\hat{y}_t - \hat{y}^n_t = \hat{y}_0 &+ \int_0^t \hat{b}_x(r, x_r, \xi_r, u)\hat{y}_r + \hat{b}_z(r, x_r, \xi_r, u)y_r + \hat{b}_u(r, x_r, \xi_r, u)\, dr \\
&+ \sum_{j=1}^{m_2}\int_0^t \hat{\sigma}^j_x(r, x_r, \xi_r, u)\hat{y}_r + \hat{\sigma}^j_z(r, x_r, \xi_r, u)y_r + \hat{\sigma}^j_u(r, x_r, \xi_r, u)\, dB^j_r \\
&- \hat{y}^n_0 - \int_0^t \hat{b}_x\left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right)\hat{y}^n_{t_{\overline{n}(r)}} + \hat{b}_z\left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right)y^n_{t_{\overline{n}(r)}}
\end{aligned}$$

$$
\begin{aligned}
&+ \hat{b}_u \left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right) dr \\
&- \sum_{j=1}^{m_2} \int_0^t \hat{\sigma}^j_x \left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right) \hat{y}^n_{t_{\overline{n}(r)}} + \hat{\sigma}^j_z \left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right) y^n_{t_{\overline{n}(r)}} \\
&+ \hat{\sigma}^j_u \left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right) dB^j_r \\
=\,& \hat{y}_0 - \hat{y}^n_0 + \int_0^t \hat{b}_x(r, x_r, \xi_r, u)\hat{y}_r - \hat{b}_x \left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right) \hat{y}^n_{t_{\overline{n}(r)}}\, dr \\
&+ \int_0^t \hat{b}_z(r, x_r, \xi_r, u) y_r - \hat{b}_z \left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right) y^n_{t_{\overline{n}(r)}}\, dr \\
&+ \int_0^t \hat{b}_u(r, x_r, \xi_r, u) - \hat{b}_u \left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right) dr \\
&+ \sum_{j=1}^{m_2} \int_0^t \hat{\sigma}^j_x(r, x_r, \xi_r, u)\hat{y}_r - \hat{\sigma}^j_x \left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right) \hat{y}^n_{t_{\overline{n}(r)}}\, dB^j_r \\
&+ \sum_{j=1}^{m_2} \int_0^t \hat{\sigma}^j_z(r, x_r, \xi_r, u) y_r - \hat{\sigma}^j_z \left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right) y^n_{t_{\overline{n}(r)}}\, dB^j_r \\
&+ \sum_{j=1}^{m_2} \int_0^t \hat{\sigma}^j_u(r, x_r, \xi_r, u) - \hat{\sigma}^j_u \left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right) dB^j_r.
\end{aligned}
$$

Let $C > 0$ be a constant which only depends on T, l, m_2 and L, we have for $l \geq 2$, since $\hat{y}^n_0 = \hat{y}_0$

$$
\begin{aligned}
&\mathrm{E}\left[\|\hat{y} - \hat{y}^n\|^l_{\infty,0,t}\right] \\
&\leq C\Bigg(\mathrm{E}\left[\sup_{0\leq s\leq t}\left|\int_0^s \hat{b}_x(r, x_r, \xi_r, u)\hat{y}_r - \hat{b}_x \left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right) \hat{y}^n_{t_{\overline{n}(r)}}\, dr\right|^l\right] \\
&\quad + \mathrm{E}\left[\sup_{0\leq s\leq t}\left|\int_0^s \hat{b}_z(r, x_r, \xi_r, u) y_r - \hat{b}_z \left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right) y^n_{t_{\overline{n}(r)}}\, d_r\right|^l\right] \\
&\quad + \mathrm{E}\left[\sup_{0\leq s\leq t}\left|\int_0^s \hat{b}_u(r, x_r, \xi_r, u) - \hat{b}_u \left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right) dr\right|^l\right] \\
&\quad + \sum_{j=1}^{m_2} \mathrm{E}\left[\sup_{0\leq s\leq t}\left|\int_0^s \hat{\sigma}^j_x(r, x_r, \xi_r, u)\hat{y}_r - \hat{\sigma}^j_x \left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right) \hat{y}^n_{t_{\overline{n}(r)}}\, dB^j_r\right|^l\right] \\
&\quad + \sum_{j=1}^{m_2} \mathrm{E}\left[\sup_{0\leq s\leq t}\left|\int_0^s \hat{\sigma}^j_z(r, x_r, \xi_r, u) y_r - \hat{\sigma}^j_z \left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right) y^n_{t_{\overline{n}(r)}}\, dB^j_r\right|^l\right] \\
&\quad + \sum_{j=1}^{m_2} \mathrm{E}\left[\sup_{0\leq s\leq t}\left|\int_0^s \hat{\sigma}_u(r, x_r, \xi_r, u) - \hat{\sigma}_u \left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right) dB_r\right|^l\right]\Bigg) \\
&= C(I_1 + I_2 + I_3 + I_4 + I_5 + I_6).
\end{aligned}
$$

We estimate the term I_1 by conditions (B_1), (E_2), Lemma 4.11 and the Jensen inequality

$$
\begin{aligned}
I_1 \leq C\mathrm{E}\bigg[\int_0^t \bigg| & \hat{b}_x(r, x_r, \xi_r, u)(\hat{y}_r - \hat{y}_r^n) + \hat{b}_x(r, x_r, \xi_r, u)(\hat{y}_r^n - \hat{y}^n_{t_{\overline{n}(r)}}) \\
& + \left(\hat{b}_x(r, x_r, \xi_r, u) - \hat{b}_x\left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right)\right) \hat{y}^n_{t_{\overline{n}(r)}} \bigg|^l dr\bigg] \\
\leq & C\mathrm{E}\left[\int_0^t \|\hat{y} - \hat{y}^n\|^l_{\infty,0,r}\, dr\right] + C\mathrm{E}\left[\|\hat{y}^n_\cdot - \hat{y}^n_{t_{\overline{n}(\cdot)}}\|^l_{\infty,0,T}\right] \\
& + C\mathrm{E}\left[\int_0^t \left((1 + |x_r| + |\xi_r|)\delta_2^{\frac{1}{2}} + \|x_\cdot - x^n_{t_{\overline{n}(\cdot)}}\|_{\infty,0,T} + \|\xi_\cdot - \xi^n_{t_{\overline{n}(\cdot)}}\|_{\infty,0,T}\right)^l \|\hat{y}^n\|^l_{\infty,0,T}\, dr\right] \\
\leq & C\int_0^t \mathrm{E}\left[\|\hat{y} - \hat{y}^n\|^l_{\infty,0,r}\right] dr + C(1 + D_{\hat{y}^n,l} + D_{y^n,l})\delta_2^{\frac{l}{2}} \\
& + C\mathrm{E}\left[(1 + \|x\|_{\infty,0,T} + \|\xi\|_{\infty,0,T})^l \delta_2^{\frac{l}{2}} \|\hat{y}^n\|^l_{\infty,0,T}\right] \\
& + C\mathrm{E}\left[\left(\|x - x^n\|^l_{\infty,0,T} + \|x^n_\cdot - x^n_{t_{\overline{n}(\cdot)}}\|^l_{\infty,0,T}\right) \|\hat{y}^n\|^l_{\infty,0,T}\right] \\
& + C\mathrm{E}\left[\left(\|\xi - \xi^n\|^l_{\infty,0,T} + \|\xi^n_\cdot - \xi^n_{t_{\overline{n}(\cdot)}}\|^l_{\infty,0,T}\right) \|\hat{y}^n\|^l_{\infty,0,T}\right].
\end{aligned}
$$

Using the Hölder inequality and then the results from Remark 4.1, Lemma 4.2, Lemma 4.5, Theorem 4.4 and Theorem 4.6 we get

$$
\begin{aligned}
I_1 \leq & C\int_0^t \mathrm{E}\left[\|\hat{y} - \hat{y}^n\|^l_{\infty,0,r}\right] dr + C(1 + D_{\hat{y}^n,l} + D_{y^n,l})\delta_2^{\frac{l}{2}} \\
& + C\mathrm{E}\left[\|\hat{y}^n\|^{2l}_{\infty,0,T}\right]^{\frac{1}{2}} \bigg(\mathrm{E}\left[(1 + \|x\|_{\infty,0,T} + \|\xi\|_{\infty,0,T})^{2l}\right]^{\frac{1}{2}} \delta_2^{\frac{l}{2}} + \mathrm{E}\left[\|x - x^n\|^{2l}_{\infty,0,T}\right]^{\frac{1}{2}} \\
& + \mathrm{E}\left[\|x^n_\cdot - x^n_{t_{\overline{n}(\cdot)}}\|^{2l}_{\infty,0,T}\right]^{\frac{1}{2}} + \mathrm{E}\left[\|\xi - \xi^n\|^{2l}_{\infty,0,T}\right]^{\frac{1}{2}} + \mathrm{E}\left[\|\xi^n_\cdot - \xi^n_{t_{\overline{n}(\cdot)}}\|^{2l}_{\infty,0,T}\right]^{\frac{1}{2}}\bigg) \\
\leq & C\int_0^t \mathrm{E}\left[\|\hat{y} - \hat{y}^n\|^l_{\infty,0,r}\right] dr + C(1 + D_{\hat{y}^n,l} + D_{y^n,l})\delta_2^{\frac{l}{2}} \\
& + CD^{\frac{1}{2}}_{\hat{y}^n,2l}\bigg((1 + D_{x,2l} + D_{\xi,2l})^{\frac{1}{2}}\delta_2^{\frac{l}{2}} + D^l_{K_x,4l}\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} + (1 + D_{x^n,2l} + D_{\xi^n,2l})^{\frac{1}{2}}\delta_2^{\frac{l}{2}} \\
& + D^l_{K_\xi,4l}\delta_{1,4l}^{l(2-p)} + \delta^l_{1,2l}\bigg).
\end{aligned}
$$

Since

$$
\delta_2 \leq \delta_{1,2l} \leq \delta_{1,4l} \leq \mathrm{E}\left[(1 + T + |w|_{p,0,T})^{4l}\right]^{\frac{1}{4l}} := D_{w,4l} \geq 1 \tag{4.57}
$$

analogue to (4.29), we obtain

$$
\begin{aligned}
I_1 \leq & C\int_0^t \mathrm{E}\left[\|\hat{y} - \hat{y}^n\|^l_{\infty,0,r}\right] dr + C(1 + D_{\hat{y}^n,l} + D_{y^n,l})D^{\frac{l}{2}}_{w,4l}\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} \\
& + CD^{\frac{1}{2}}_{\hat{y}^n,2l}\bigg((1 + D_{x,2l} + D_{\xi,2l})^{\frac{1}{2}}D^{\frac{l}{2}}_{w,4l} + D^l_{K_x,4l} + (1 + D_{x^n,2l} + D_{\xi^n,2l})^{\frac{1}{2}}D^{\frac{l}{2}}_{w,4l}
\end{aligned}
$$

$$+ D^l_{K_\xi,4l} D^{\frac{l}{2}}_{w,4l} + D^l_{w,4l}\Big) \delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})}$$
$$\leq D_{1,l}\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} + C\int_0^t \mathrm{E}\left[\|\hat{y}-\hat{y}^n\|^l_{\infty,0,r}\right] dr,$$

where

$$D_{1,l} := C(1 + D_{\hat{y}^n,l} + D_{y^n,l}) D^{\frac{l}{2}}_{w,4l} + C D^{\frac{1}{2}}_{\hat{y}^n,2l}\Big((1 + D_{x,2l} + D_{\xi,2l})^{\frac{1}{2}} D^{\frac{l}{2}}_{w,4l} + D^l_{K_x,4l}$$
$$+ (1 + D_{x^n,2l} + D_{\xi^n,2l})^{\frac{1}{2}} D^{\frac{l}{2}}_{w,4l} + D^l_{K_\xi,4l} D^{\frac{l}{2}}_{w,4l} + D^l_{w,4l}\Big).$$

Analogously after using the Burkholder-Davis-Gundy and Jensen inequality, we can estimate the term I_4 and obtain

$$I_4 \leq D_{4,l}\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} + C\int_0^t \mathrm{E}\left[\|\hat{y}-\hat{y}^n\|^l_{\infty,0,r}\right] dr,$$

where $D_{1,l}$ and $D_{4,l}$ only differ in the constant C because of the Burkholder-Davis-Gundy inequality. Since $\hat{b}_x$, $\hat{b}_z$ and $\hat{b}_u$ share the same properties the estimation of I_2 and I_3 is very similar to the estimation of I_1, we just need to exchange the processes $\hat{y}$ and $\hat{y}^n$ by y and y^n for the estimation of I_2 and use the corresponding results. We get

$$I_2 \leq C\mathrm{E}\Bigg[\int_0^t \Big|\hat{b}_z(r, x_r, \xi_r, u)(y_r - y^n_r) + \hat{b}_z(r, x_r, \xi_r, u)(y^n_r - y^n_{t_{\overline{n}(r)}})$$
$$+ \left(\hat{b}_z(r, x_r, \xi_r, u) - \hat{b}_z\left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right)\right) y^n_{t_{\overline{n}(r)}}\Big|^l dr\Bigg]$$
$$\leq C\mathrm{E}\left[\|y - y^n\|^l_{\infty,0,T}\right] + C\mathrm{E}\left[\|y^n_\cdot - y^n_{t_{\overline{n}(\cdot)}}\|^l_{\infty,0,T}\right]$$
$$+ C\mathrm{E}\left[\int_0^t \left|\left(\hat{b}_z(r, x_r, \xi_r, u) - \hat{b}_z\left(t_{\overline{n}(r)}, x^n_{t_{\overline{n}(r)}}, \xi^n_{t_{\overline{n}(r)}}, u\right)\right) y^n_{t_{\overline{n}(r)}}\right|^l dr\right].$$

Using the Hölder inequality, Lemma 4.2, Lemma 4.5, Theorem 4.4 and Theorem 4.6 we get

$$I_2 \leq C D^l_{K_y,2l}\delta_{1,2l}^{l(2-p)} + C\left(1 + D^{\frac{1}{2}}_{y^n,2l}\right)\delta^l_{1,2l}$$
$$+ C D^{\frac{1}{2}}_{y^n,2l}\Big((1 + D_{x,2l} + D_{\xi,2l})^{\frac{1}{2}}\delta_2^{\frac{l}{2}} + D^l_{K_x,4l}\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} + (1 + D_{x^n,2l} + D_{\xi^n,2l})^{\frac{1}{2}}\delta_2^{\frac{l}{2}}$$
$$+ D^l_{K_\xi,4l}\delta_{1,4l}^{l(2-p)} + \delta^l_{1,2l}\Big).$$

Again using the inequality (4.57), we obtain

$$I_2 \leq \delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})}\left(C D^l_{K_y,2l} D^{\frac{l}{2}}_{w,4l} + C\left(1 + D^{\frac{1}{2}}_{y^n,2l}\right) D^l_{w,4l}\right)$$

$$
\begin{aligned}
&+ CD_{y^n,2l}^{\frac{1}{2}}\bigg((1 + D_{x,2l} + D_{\xi,2l})^{\frac{1}{2}} D_{w,4l}^{\frac{l}{2}} + D_{K_x,4l}^{l} + (1 + D_{x^n,2l} + D_{\xi^n,2l})^{\frac{1}{2}} D_{w,4l}^{\frac{l}{2}} \\
&+ D_{K_\xi,4l}^{l} D_{w,4l}^{\frac{l}{2}} + D_{w,4l}^{l}\bigg)\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} \\
\leq\ & D_{2,l}\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})},
\end{aligned}
$$

where

$$
\begin{aligned}
D_{2,l} := &\left(CD_{K_y,2l}^{l} D_{w,4l}^{\frac{l}{2}} + C\left(1 + D_{y^n,2l}^{\frac{1}{2}}\right) D_{w,4l}^{l}\right) \\
&+ CD_{y^n,2l}^{\frac{1}{2}}\bigg((1 + D_{x,2l} + D_{\xi,2l})^{\frac{1}{2}} D_{w,4l}^{\frac{l}{2}} + D_{K_x,4l}^{l} + (1 + D_{x^n,2l} + D_{\xi^n,2l})^{\frac{1}{2}} D_{w,4l}^{\frac{l}{2}} \\
&+ D_{K_\xi,4l}^{l} D_{w,4l}^{\frac{l}{2}} + D_{w,4l}^{l}\bigg).
\end{aligned}
$$

Analogously, we can estimate I_5 by

$$
I_5 \leq D_{5,l}\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})},
$$

where $D_{2,l}$ and $D_{5,l}$ only differ in the constant C because of the Burkholder-Davis-Gundy inequality. The term I_3 can be estimated using the condition (B_1), (E_2) and the Jensen inequality

$$
\begin{aligned}
I_3 \leq\ & C\mathrm{E}\left[\int_0^t \left|\hat{b}_u(r, x_r, \xi_r, u) - \hat{b}_u\left(t_{\overline{n}(r)}, x_{t_{\overline{n}(r)}}^n, \xi_{t_{\overline{n}(r)}}^n, u\right)\right|^l dr\right] \\
\leq\ & C\mathrm{E}\bigg[(1 + \|x\|_{\infty,0,T} + \|\xi\|_{\infty,0,T})^l \delta_2^{\frac{l}{2}} + \|x - x^n\|_{\infty,0,T}^l + \|x_\cdot^n - x_{t_{\overline{n}(\cdot)}}^n\|_{\infty,0,T}^l \\
&+ \|\xi - \xi^n\|_{\infty,0,T}^l + \|\xi_\cdot^n - \xi_{t_{\overline{n}(\cdot)}}^n\|_{\infty,0,T}^l\bigg].
\end{aligned}
$$

Again using the results of Lemma 4.2, Lemma 4.5, Theorem 4.4 and Theorem 4.6 this yields

$$
\begin{aligned}
I_3 \leq C\bigg(&(1 + D_{x,l} + D_{\xi,l})\delta_2^{\frac{l}{2}} + D_{K_x,2l}^{l}\delta_{1,2l}^{l((2-p)\wedge\frac{1}{2})} + (1 + D_{x^n,l} + D_{\xi^n,l})^{\frac{1}{2}}\delta_2^{\frac{l}{2}} \\
&+ D_{K_\xi,2l}^{l}\delta_{1,2l}^{l(2-p)} + \delta_{1,l}^{l}\bigg).
\end{aligned}
$$

Since

$$
\delta_2 \leq \delta_{1,l} \leq \delta_{2,l} \leq \mathrm{E}\left[(1 + T + |w|_{p,0,T})^{2l}\right]^{\frac{1}{2l}} := \tilde{D}_{w,2l} \geq 1,
$$

we obtain

$$
\begin{aligned}
I_3 \leq \delta_{1,2l}^{l((2-p)\wedge\frac{1}{2})} C\bigg(&(1 + D_{x,l} + D_{\xi,l}) D_{w,2l}^{\frac{l}{2}} + D_{K_x,2l}^{l} + (1 + D_{x^n,l} + D_{\xi^n,l})^{\frac{1}{2}} D_{w,2l}^{\frac{l}{2}} \\
&+ D_{K_\xi,2l}^{l} D_{w,2l}^{\frac{l}{2}} + D_{w,2l}^{l}\bigg)
\end{aligned}
$$

$$\leq D_{3,l}\delta_{1,2l}^{l((2-p)\wedge\frac{1}{2})},$$

where

$$D_{3,l} := C\bigg((1 + D_{x,l} + D_{\xi,l})D_{w,2l}^{\frac{l}{2}} + D_{K_x,2l}^{l} + (1 + D_{x^n,l} + D_{\xi^n,l})^{\frac{1}{2}}D_{w,2l}^{\frac{l}{2}}$$
$$+ D_{K_\xi,2l}^{l}D_{w,2l}^{\frac{l}{2}} + D_{w,2l}^{l}\bigg).$$

Analogously, we obtain

$$I_6 \leq D_{6,l}\delta_{1,2l}^{l((2-p)\wedge\frac{1}{2})},$$

where again $D_{3,l}$ and $D_{6,l}$ only differ in the constant C, because of the Burkholder-Davis-Gundy inequality. Combining all the estimates, it follows since $\delta_{1,l} \leq \delta_{1,2l}$

$$\begin{aligned}
\mathrm{E}\left[\|\hat{y} - \hat{y}^n\|_{\infty,0,t}^{l}|\right] &\leq C\int_0^t \mathrm{E}\left[\|\hat{y} - \hat{y}^n\|_{\infty,0,r}^{l}\right]\,dr + D_{1,l}\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} \\
&\quad + D_{4,l}\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} + D_{2,l}\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} + D_{5,l}\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} \\
&\quad + D_{3,l}\delta_{1,2l}^{l((2-p)\wedge\frac{1}{2})} + D_{6,l}\delta_{1,2l}^{l((2-p)\wedge\frac{1}{2})} \\
&\leq C\int_0^t \mathrm{E}\left[\|\hat{y} - \hat{y}^n\|_{\infty,0,r}^{l}\right]\,dr + D_l\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})},
\end{aligned}$$

where

$$D_l := \max\{D_{1,l},\ldots,D_{6,l}\}.$$

By the Gronwall inequality, we conclude

$$\mathrm{E}\left[\|\hat{y} - \hat{y}^n\|_{\infty,0,T}^{l}|\right] \leq D_l\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})}e^C.$$

This yields

$$\mathrm{E}\left[\|\hat{y} - \hat{y}^n\|_{\infty,0,T}^{l}\right]^{\frac{1}{l}} \leq D_{K_{\hat{y}},l}\delta_{1,4l}^{(2-p)\wedge\frac{1}{2}},$$

where

$$D_{K_{\hat{y}},l} := D_l^{\frac{1}{l}}e^C,$$

which concludes the proof. □

We end this subsection with its main result concerning the convergence rate at which $\mathcal{Y}^n$ converges to $\mathcal{Y}$ in the L^l-norm, uniform in time. The result is a direct consequence of Theorem 4.10 and Theorem 4.12.

Theorem 4.13. *With the notations and assumptions from the beginning of this chapter, we have* *that for every $l \geq 2$, there exists a constants $D_{\mathcal{Y}^n,l}$ and $D_{K_\mathcal{Y},l}$ independent of n and u such that*

$$\mathrm{E}\left[\|\mathcal{Y}^{n,u}\|_{\infty,0,T}^l\right] \leq D_{\mathcal{Y}^n,l}$$

and

$$\mathrm{E}\left[\|\mathcal{Y}^u - \mathcal{Y}^{n,u}\|_{\infty,0,T}^l\right]^{\frac{1}{l}} \leq D_{K_\mathcal{Y},l}\delta_{1,4l}^{(2-p)\wedge\frac{1}{2}}$$

for every $u \in \mathcal{U}$.

Proof. We have

$$\mathrm{E}\left[\|\mathcal{Y}^{n,u}\|_{\infty,0,T}^l\right] \leq 2^{l-1}\left(\mathrm{E}\left[\|y^{n,u}\|_{\infty,0,T}^l\right] + \mathrm{E}\left[\|\hat{y}^{n,u}\|_{\infty,0,T}^l\right]\right)$$

and by Lemma 4.9 and Lemma 4.11, this yields

$$\mathrm{E}\left[\|\mathcal{Y}^{n,u}\|_{\infty,0,T}^l\right] \leq 2^{l-1}\left(D_{y^n,l} + D_{\hat{y}^n,l}\right) := D_{\mathcal{Y}^n,l}.$$

By Theorem 4.10 and Theorem 4.12, we get

$$\begin{aligned} E\left[\|\mathcal{Y}^u - \mathcal{Y}^{n,u}\|_{\infty,0,T}^l\right]^{\frac{1}{l}} &\leq 2^{1-\frac{1}{l}}\left(D_{K_y,2l}\delta_{1,2l}^{2-p} + D_{K_{\hat{y}},l}\delta_{1,4l}^{(2-p)\wedge\frac{1}{2}}\right) \\ &\leq 2^{1-\frac{1}{l}}\left(D_{K_y,2l}D_{w,4l}^{\frac{1}{2}} + D_{K_{\hat{y}},l}\right)\delta_{1,4l}^{(2-p)\wedge\frac{1}{2}} \\ := D_{K_\mathcal{Y},l}\delta_{1,4l}^{(2-p)\wedge\frac{1}{2}}. \end{aligned}$$

□

In the proof of Theorem 4.6, we mentioned that an estimate was not ideal in the case where the driving process w is Hölder continuous. But since this case is important for practical purposes, we will devote the next subsection to this matter.

4.1.3 Convergence rates for Hölder continuous driving processes

For numerical experiments the convergence parameter $\delta_{1,l}$ is not very practical. Instead, we wish to use the convergence parameter δ_2 which only depends on the mesh of the underlying Euler partition. To accomplish that, we assume that w is Hölder continuous of order $H \in (\frac{1}{2}, 1)$, which implies that w is of bounded p-variation for $p \geq \frac{1}{H}$. We get

$$\delta(\omega)^{2-p} \leq \left(\delta_2 + C_H(\omega)\delta_2^H\right)^{2-\frac{1}{H}} \leq (T^{1-H} + C_H(\omega))\delta_2^{2H-1},$$

where $C_H(\omega)$ is the Hölder seminorm of the path $w(\omega)$. Consequently, if the process w is Hölder continuous of order H, the convergence rate for the convergence of the Young differential equation

ξ^n to ξ and y^n to y is equal to those in Mishura [2008] and Nourdin and Neuenkirch [2007]. In the case where $\mathrm{E}\left[C_H^l\right] < \infty$, we get for every $l \geq 1$

$$\delta_{1,l} \leq \mathrm{E}\left[(T^{1-H} + C_H)^l\right]^{\frac{1}{l}} \delta_2^H. \tag{4.58}$$

This moment condition is e.g. satisfied by the fractional Brownian motion with Hurst parameter $H' > \frac{1}{2}$, for every $H < H'$ (see Nualart [1995], p.274). In the proof of Theorem 4.6, we noted that the estimate $\delta_2 \leq \delta_{1,2l}$ was not ideal in the case were the process w is Hölder continuous. To see this, suppose we just use the inequality (4.58) in the statement of Theorem 4.6, then we get the convergence rate

$$\delta_2^{(2H-1)\wedge\frac{H}{2}}.$$

If we would not use the estimate $\delta_2 \leq \delta_{1,2l}$ and use instead (4.58) in (4.30) to estimate $\delta_{1,2l}$, we get

$$\begin{aligned} &\mathrm{E}\left[\|x - x^n\|_{\infty,0,t}^l\right] \\ &\leq C\mathrm{E}\left[(T^{1-H} + C_H)^{2l}\right]^{\frac{1}{2l}} \delta_2^{l(2H-1)} \left(D_{w,2l}^l + D_{K_\xi,2l}^l D_{w,2l}^l\right) + C(1 + D_{1,l})\delta_2^{\frac{l}{2}} \end{aligned}$$

and hence there exists a constant $\hat{D}_{K_x,l}$ independent of u and n, such that

$$\mathrm{E}\left[\|x - x^n\|_{\infty,0,t}^l\right]^{\frac{1}{l}} \leq \hat{D}_{K_x,l}\delta_2^{(2H-1)\wedge\frac{1}{2}}.$$

Hence, we get in the Hölder case the better convergence rate $(2H-1)\wedge\frac{1}{2} \geq (2H-1)\wedge\frac{H}{2}$, which can be seen as the worst of both cases, the rate $2H-1$ as standard for the convergence of the first order Euler Scheme in the YDE case and the standard rate $\frac{1}{2}$ for the convergence of the first order Euler scheme in the Brownian SDE case. With the above arguments, we get for every $u \in \mathcal{U}$ the modified estimates

$$E\left[\|\xi^u - \xi^{n,u}\|_{\infty,0,T}^l\right]^{\frac{1}{l}} \leq \hat{D}_{K_\xi,l}\delta_2^{2H-1} \tag{4.59}$$

for $l \geq 1$ and in the Brownian SDE case, respectively the whole system,

$$E\left[\|x^u - x^{n,u}\|_{\infty,0,T}^l\right]^{\frac{1}{l}} \leq \hat{D}_{K_x,l}\delta_2^{(2H-1)\wedge\frac{1}{2}} \tag{4.60}$$

and

$$E\left[\|\mathcal{X}^u - \mathcal{X}^{n,u}\|_{\infty,0,T}^l\right]^{\frac{1}{l}} \leq \hat{D}_{K_\mathcal{X},l}\delta_2^{(2H-1)\wedge\frac{1}{2}} \tag{4.61}$$

for $l \geq 2$. The constants $\hat{D}_{K_\xi,l}$, $\hat{D}_{K_x,l}$ and $\hat{D}_{\mathcal{X},l}$ are independent of u and n. These estimates now carry over to the proof of Theorem 4.12. Using (4.58), (4.59) and (4.60) in every estimation of the terms I_i for $i = 1, \dots, 6$, we get the estimate

$$\mathrm{E}\left[\|\hat{y} - \hat{y}^n\|^l_{\infty,0,T}\right]^{\frac{1}{l}} \leq \hat{D}_{K_{\hat{y}},l}\delta_2^{(2H-1)\wedge\frac{1}{2}},$$

where the constant $\hat{D}_{K_{\hat{y}},l}$ is independent of u and n. This also changes the rate in Theorem 4.13 and we conclude

$$E\left[\|\mathcal{Y}^u - \mathcal{Y}^{n,u}\|^l_{\infty,0,T}\right]^{\frac{1}{l}} \leq \hat{D}_{\mathcal{Y},l}\delta_2^{(2H-1)\wedge\frac{1}{2}}.$$

Our governing noise of the volatility process ξ in our practical example will be a fractional Brownian motion with Hurst parameter $H > \frac{1}{2}$, so that in the following sections, we will state the convergence results in the general case where w is just a continuous process of finite p-variation for $p \in (1, 2)$ and furthermore we will state the results under the following additional assumption

(HA): Hölder assumption: Almost every path of the process w is Hölder continuous of order $H > \frac{1}{2}$ and the Hölder seminorm

$$|w|_{H-Hol,0,T} = \sup_{s,t\in\Delta([0,T])} \frac{|w_t - w_s|}{|t-s|^H}$$

has moments of all orders.

4.2 Convergence of the approximating scheme for the backwards adjoint equation

In the last section, we established the convergence rate of the Euler schemes for the forward equations, we are interested in. In this section, we keep the assumptions and notations from the beginning of this chapter and add the assumption, that the sequence $(g_\mu)_{\mu=1,\dots,M}$ of functions $g_\mu : \mathbb{R}^{(n_1+n_2)} \to \mathbb{R}$ satisfies condition (G) given in the introduction to Chapter 3. We come to the approximation of the backwards adjoint equation given by

$$\begin{aligned}\Lambda_t = \sum_{T_\mu \geq t} E[g_\mu(\mathcal{X}^u_{T_\mu})]g'_\mu(\mathcal{X}^u_{T_\mu}) + \int_t^T \Lambda_r\left[\begin{pmatrix} b^u_x(r) & 0 \\ \hat{b}^u_z(r) & \hat{b}^u_x(r)\end{pmatrix} - \sum_{j=1}^{m_2}\begin{pmatrix} 0 & 0 \\ \hat{\sigma}^{u,j}_z(r) & \hat{\sigma}^{u,j}_x(r)\end{pmatrix}^2\right]dr \\ + \sum_{j=1}^{m_1}\int_t^T \Lambda_r \begin{pmatrix}\sigma^{u,j}_x(r) & 0 \\ 0 & 0\end{pmatrix} dw^j_r + \sum_{j=1}^{m_2}\int_t^T \Lambda_r \begin{pmatrix} 0 & 0 \\ \hat{\sigma}^{u,j}_z(r) & \hat{\sigma}^{u,j}_x(r)\end{pmatrix} d^- B^j_r,\end{aligned} \tag{4.62}$$

where $0 < T_1 \leq \cdots \leq T_M = T \in [0, T]$ and $t \in [0, T]$. Here $b^u_x(r) := b_x(r, \xi^u_r, u)$, $\hat{b}^u_x(r) := \hat{b}_x(r, x^u_r, \xi^u_r, u)$ and the other functions $\sigma^{u,j}_x$, $\hat{b}^u_z$, $\hat{\sigma}^{u,j}_x$ and $\hat{\sigma}^{u,j}_z$ are defined analogously. For readability we will leave out the index u in cases where no confusion might occur. We know by

Theorem 3.17 that this equation has the explicit solution given by

$$\Lambda_t = \sum_{T_\mu \geq t} E[g_\mu(\mathcal{X}_{T_\mu})] g'_\mu(\mathcal{X}_{T_\mu}) \Phi_{T_\mu} \Phi_t^{-1} \text{ for } t \in [0, T],$$

where Φ and $\Phi^{-1} = \Psi$ are the solutions to the homogenous linear matrix valued SDEs given in (3.7) and (3.9) with initial time 0. Note that Λ_t is an $n_1 + n_2$-dimensional row vector.

To derive a suitable discretization scheme for the approximation of the solution to the adjoint equation, we heuristically develop the ideas in one dimension. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ be a filtered probability space, satisfying the usual conditions, carrying a one dimension continuous process of finite p-variation w_t ($p \in (1, 2)$) and a standard Brownian motion B_t, both adapted to the filtration $\mathbb{F}$. Take a look at the anticipating backward equation

$$\lambda_t = \lambda_T + \int_t^T \lambda_r (b_r - \hat{\sigma}_r^2)\, dt + \int_t^T \lambda_r \sigma_r \, dw_r + \int_t^T \lambda_r \hat{\sigma}_r \, d^- B_r, \tag{4.63}$$

where λ_T is a constant terminal value, and the process b, σ and $\hat{\sigma}$ are $\mathbb{F}$-adapted processes satisfying conditions to ensure the existence of the above integrals. Now we take an equidistant partition $0 = t_0 < t_1 < \cdots < t_n = T$ of the interval $[0, T]$ and consider the first order Euler approximation of the above equation, taking the left endpoints of the integrands, because of the forward integral. By setting $\Delta_i = t_{i+1} - t_i$, $\Delta w_i = w_{t_{i+1}} - w_{t_i}$ and $\Delta B_i = B_{t_{i+1}} - B_{t_i}$, we obtain

$$\begin{aligned}\lambda_{t_i} &= \lambda_{t_{i+1}} + \lambda_{t_i} \left[(b_{t_i} - \hat{\sigma}_{t_i}^2)\Delta_i + \sigma_{t_i} \Delta w_i + \hat{\sigma}_{t_i} \Delta B_i \right] \\ &= \lambda_{t_{i+1}} + \lambda_{t_i} \eta_{t_i}\end{aligned}$$

and by rearranging the terms

$$\lambda_{t_{i+1}} = \lambda_{t_i}(1 - \eta_{t_i})$$

for every $i = 0, \ldots, n-1$. Since we have an terminal condition, we need a backwards scheme and therefor use the Taylor expansion of order 2 on the function $f(x) = (1-x)^{-1}$, such that

$$\begin{aligned}\lambda_{t_i} &= \lambda_{t_{i+1}}(1 + \eta_{t_i} + \eta_{t_i}^2 + R_i) \\ &= \lambda_{t_{i+1}}(1 + (b_{t_i} - \hat{\sigma}_{t_i}^2)\Delta_i + \sigma_{t_i}\Delta w_i + \hat{\sigma}_{t_i}\Delta B_i + \hat{\sigma}_{t_i}^2 \Delta B_i^2 + \tilde{R}_i)\end{aligned}$$

for $i = 1, \ldots, n-1$. Note that all the terms in $\tilde{R}_i$ have 0 quadratic variation, since w has bounded p-variation for $p \in (1, 2)$. Since the quadratic variation of B_t is t and $\hat{\sigma}$ is $\mathbb{F}$-adapted and square integrable, by choosing the mesh of our partition small enough, we get

$$\sum_{i=0}^{n-1} \hat{\sigma}_{t_i}^2 (\Delta B_i^2 - \Delta_i) \approx 0.$$

This suggests that

$$\lambda_{t_i} \approx \lambda_{t_{i+1}} \left[b_{t_i}\Delta_i + \sigma_{t_i}\Delta w_i + \hat{\sigma}_{t_i}\Delta B_i \right]$$

yields a suitable discretization for the equation (4.63). Now translating these ideas to the multidimenisonal case, we define the approximation scheme for the adjoint equation (4.62) on a partition $\Pi^{\mathrm{E}} = \{t_i\}_{i=0,\dots,n}$ of the interval $[0,T]$ by

$$\Lambda^n_{t_i} = (\lambda^n_{t_i}, \hat{\lambda}^n_{t_i}) = \Lambda^n_{t_{i+1}} \left(I_{n_1+n_2} + \eta_{t_i,t_{i+1}} \right) + \sum_{T_\mu = t_i} E[g_\mu(\mathcal{X}^n_{T_\mu})] g'_\mu(\mathcal{X}^n_{T_\mu}) \in \mathbb{R}^{(n_1+n_2)}, \tag{4.64}$$

where

$$\begin{aligned}
\eta_{t_i,t_{i+1}} = &\begin{pmatrix} b_x(t_i,\xi^n_{t_i},u) & 0 \\ \hat{b}_z(t_i,x^n_{t_i},\xi^n_{t_i},u) & \hat{b}_x(t_i,x^n_{t_i},\xi^n_{t_i},u) \end{pmatrix} (t_{i+1}-t_i) + \sum_{j=1}^{m_1} \begin{pmatrix} \sigma^j_x(t_i,\xi^n_{t_i},u) & 0 \\ 0 & 0 \end{pmatrix} (w^j_{t_{i+1}} - w^j_{t_i}) \\
&+ \sum_{j=1}^{m_2} \begin{pmatrix} 0 & 0 \\ \hat{\sigma}^j_z(t_i,x^n_{t_i},\xi^n_{t_i},u) & \hat{\sigma}^j_x(t_i,x^n_{t_i},\xi^n_{t_i},u) \end{pmatrix} (B^j_{t_{i+1}} - B^j_{t_i}) \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}
\end{aligned}$$

for all $i \in \{0,\dots,n-1\}$ and

$$\Lambda^n_T = \sum_{T_\mu = T} \mathrm{E}[g_\mu(\mathcal{X}^n_T)] g'_\mu(\mathcal{X}^n_T).$$

We use in this case the constant interpolation on the interval $[0,T]$, meaning that

$$\Lambda^n_t = \Lambda^n_{t_{i+1}}$$

for $t \in (t_i, t_{i+1}]$ and $i \in \{0,\dots,n-1\}$. Notice that by the definition of Λ^n, we have for $i \in \{0,\dots n-1\}$ that

$$\begin{aligned}
\Lambda^n_{t_i} = &\sum_{T_\mu = T} \mathrm{E}[g_\mu(\mathcal{X}^n_T)] g'_\mu(\mathcal{X}^n_T)(I_{n_1+n_2} + \eta_{t_{n-1},T}) \cdot \dots \cdot (I_{n_1+n_2} + \eta_{t_{i+1},t_{i+2}})(I_{n_1+n_2} + \eta_{t_i,t_{i+1}}) \\
&+ \sum_{T_\mu = t_{n-1}} \mathrm{E}[g_\mu(\mathcal{X}^n_{t_{n-1}})] g'_\mu(\mathcal{X}^n_{t_{n-1}})(I_{n_1+n_2} + \eta_{t_{n-2},t_{n-1}}) \cdot \dots \cdot (I_{n_1+n_2} + \eta_{t_i,t_{i+1}}) \\
&+ \dots + \sum_{T_\mu = t_{i+1}} \mathrm{E}[g_\mu(\mathcal{X}^n_{t_{i+1}})] g'_\mu(\mathcal{X}^n_{t_{i+1}})(I_{n_1+n_2} + \eta_{t_i,t_{i+1}}) \\
&+ \sum_{T_\mu = t_i} \mathrm{E}[g_\mu(\mathcal{X}^n_{t_i})] g'_\mu(\mathcal{X}^n_{t_i}).
\end{aligned} \tag{4.65}$$

The goal of this section is to show that for $|\Pi^{\mathrm{E}}| \to 0$ and $l \geq 2$, we have

$$\sup_{t\in[0,T]} \mathrm{E}\left[\left| \Lambda_t - \Lambda^n_{t_{\underline{n}(t)}} \right|^l \right] \to 0$$

(see (4.34), for the definition of $\underline{n}$) and to find the rate of this convergence. To show this, we take a closer look at the explicit solution of equation (3.17), given by

$$\Lambda_t = \sum_{T_\mu \geq t} E[g_\mu(\mathcal{X}_{T_\mu})]^\top g'_\mu(\mathcal{X}_{T_\mu}) \Phi_{T_\mu} \Phi_t^{-1} \text{ for } t \in [0,T].$$

We add the initial time 0 as index to the processes Φ and Φ^{-1}, such that $\Phi_{T_\mu} = \Phi^0_{T_\mu}$ and $\Phi_t^{-1} = (\Phi_t^0)^{-1}$ in the last equation. We know that Φ^0 and $(\Phi^0)^{-1}$ are elements of $L^l_{\mathbb{F}}(\Omega, C([0,T], \mathbb{R}^{n_1 \times n_2})$ for every $l \geq 1$, such that we have for $t \in [0,T]$ and $s \in [t,T]$ that

$$\begin{aligned}
\Phi_s^0(\Phi_t^0)^{-1} = &= I_{n_1+n_2} + \int_t^s \begin{pmatrix} b_x(r,\xi_r,u) & 0 \\ \hat{b}_z(r,x_r,\xi_r,u) & \hat{b}_x(r,x_r,\xi_r,u) \end{pmatrix} \Phi_r^{s_0}(\Phi_t^0)^{-1}\, dr \\
&+ \sum_{j=1}^{m_1} \int_t^s \begin{pmatrix} \sigma_x^j(r,\xi_r),u) & 0 \\ 0 & 0 \end{pmatrix} \Phi_r^{s_0}(\Phi_t^0)^{-1}\, dw_r^j \\
&+ \sum_{j=1}^{m_2} \int_t^s \begin{pmatrix} 0 & 0 \\ \hat{\sigma}_z^j(r,x_r,\xi_r,u) & \hat{\sigma}_x^j(r,x_r,\xi_r,u) \end{pmatrix} \Phi_r^{s_0}(\Phi_t^0)^{-1}\, dB_r^j,
\end{aligned}$$

such that $\Phi_s^0(\Phi_t^0)^{-1}$ satisfies the equation

$$\begin{aligned}
\Phi_s^t = I_{n_1+n_2} &+ \int_t^s \begin{pmatrix} b_x(r,\xi_r,u) & 0 \\ \hat{b}_z(r,x_r,\xi_r,u) & \hat{b}_x(r,x_r,\xi_r,u) \end{pmatrix} \Phi_r^t\, dr \\
&+ \sum_{j=1}^{m_1} \int_t^s \begin{pmatrix} \sigma_x^j(r,\xi_r),u) & 0 \\ 0 & 0 \end{pmatrix} \Phi_r^t\, dw_r^j \\
&+ \sum_{j=1}^{m_2} \int_t^s \begin{pmatrix} 0 & 0 \\ \hat{\sigma}_z^j(r,x_r,\xi_r,u) & \hat{\sigma}_x^j(r,x_r,\xi_r,u) \end{pmatrix} \Phi_r^t\, dB_r^j
\end{aligned} \tag{4.66}$$

for $s \in [t,T]$. In Section 3.2 we established that equation (4.66) has the unique solution

$$\Phi_s^t = \begin{pmatrix} \phi_s^t & 0 \\ \tilde{\phi}_s^t & \hat{\phi}_s^t \end{pmatrix},$$

where

$$\begin{aligned}
\phi_s^t &= I_{n_1} + \int_t^s b_x(r,x_r,\xi_r,u)\phi_r^t\, dr + \sum_{j=1}^{m_1} \int_t^s \sigma_x^j(r,x_r,\xi_r,u)\phi_r^t\, dw_r^j \\
\hat{\phi}_s^t &= I_{n_2} + \int_t^s \hat{b}_x(r,x_r,\xi_r,u)\hat{\phi}_r^t\, dt + \sum_{j=1}^{m_2} \int_t^s \hat{\sigma}_x^j(r,x_r,\xi_r,u)\hat{\phi}_r^t\, dB_r^j \\
\tilde{\phi}_s^t &= \int_t^s \hat{b}_z(r,x_r,\xi_r,u)\phi_r^t + \hat{b}_x(r,x_r,\xi_r,u)\tilde{\phi}_r^t\, dr
\end{aligned} \tag{4.67}$$

$$+\sum_{j=1}^{m_2}\int_t^s \hat{\sigma}_z^j(r,x_r,\xi_r,u)\phi_r^t + \hat{\sigma}_x^j(r,x_r,\xi_r,u)\tilde{\phi}_r^t\, dB_r^j \tag{4.68}$$

and that for every $t\in[0,T]$, Φ^t is an element of $L^l_{\mathbb{F}}(\Omega, C([t,T]), \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)})$ for every $l\geq 1$, such that

$$\sup_{t\in[0,T]} E\left[\|\Phi^t\|^l_{\infty,t,T}\right] \leq D_{\Phi,l}, \tag{4.69}$$

where $D_{\Phi,l}>0$ is a constant independent of u. This yields that

$$\Lambda_t = \sum_{T_\mu\geq t} E[g_\mu(\mathcal{X}_{T_\mu})]^\top g'_\mu(\mathcal{X}_{T_\mu})\Phi^t_{T_\mu} \text{ for } t\in[0,T].$$

Now let $\Pi^{\mathrm{E}}=\{t_i\}_{i=0,\dots,n}$ be a partition of the interval $[0,T]$ and for every $k\in\{0,\dots,n-1\}$, we define the Euler scheme

$$V_{t_i}^{t_k} = \begin{pmatrix} \varphi_{t_i}^{t_k} & 0 \\ \tilde{\varphi}_{t_i}^{t_k} & \hat{\varphi}_{t_i}^{t_k}\end{pmatrix} \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}$$

for $i\in\{k,\dots,n-1\}$, such that

$$\begin{aligned} V_{t_{i+1}}^{t_k} &= V_{t_i}^{t_k} + \begin{pmatrix} b_x(t_i,\xi^n_{t_i},u) & 0 \\ \hat{b}_z(t_i,x^n_{t_i},\xi^n_{t_i},u) & \hat{b}_x(t_i,x^n_{t_i},\xi^n_{t_i},u)\end{pmatrix} V_{t_i}^{t_k}(t_{i+1}-t_i) \\ &\quad + \sum_{j=1}^{m_1}\begin{pmatrix} \sigma_x^j(t_i,\xi^n_{t_i},u) & 0 \\ 0 & 0\end{pmatrix} V_{t_i}^{t_k}(w^j_{t_{i+1}}-w^j_{t_i}) \\ &\quad + \sum_{j=1}^{m_2}\begin{pmatrix} 0 & 0 \\ \hat{\sigma}_z^j(t_i,x^n_{t_i},\xi^n_{t_i},u) & \hat{\sigma}_x^j(t_i,x^n_{t_i},\xi^n_{t_i},u)\end{pmatrix} V_{t_i}^{t_k}(B^j_{t_{i+1}}-B^j_{t_i}) \\ &= (I_{n_1+n_2}+\eta_{t_i,t_{i+1}})V_{t_i}^{t_k} \end{aligned} \tag{4.70}$$

and

$$V_{t_k}^{t_k} = I_{n_1+n_2}.$$

Here we use the continuous interpolation, such that

$$V_r^{t_k} = (I_{n_1+n_2}+\eta_{t_i,r})V_{t_i}^{t_k} \tag{4.71}$$

for $r\in[t_i,t_{i+1})$, $i\in\{k,\dots,n-1\}$ and

$$V_r^{t_k} = I_{n_1+n_2}$$

for all $r\in[0,t_k]$. Notice that

$$V_{t_i}^{t_k} = (I_{n_1+n_2}+\eta_{t_{i-1},t_i})\cdot\dots\cdot(I_{n_1+n_2}+\eta_{t_k,t_{k+1}}),$$

which yields by (4.65)

$$\begin{aligned}\Lambda^n_{t_{\underline{n}(t)}} &= \sum_{T_\mu=T} \mathrm{E}[g_\mu(\mathcal{X}^n_T)]g'_\mu(\mathcal{X}^n_T)V_T^{t_{\underline{n}(t)}} + \sum_{T_\mu=t_{n-1}} \mathrm{E}[g_\mu(\mathcal{X}^n_{t_{n-1}})]g'_\mu(\mathcal{X}^n_{t_{n-1}})V_{t_{n-1}}^{t_{\underline{n}(t)}} \\ &\quad + \cdots + \sum_{T_\mu=t_{\underline{n}(t)+1}} \mathrm{E}[g_\mu(\mathcal{X}^n_{t_{i+1}})]g'_\mu(\mathcal{X}^n_{t_{i+1}})V_{t_{\underline{n}(t)+1}}^{t_{\underline{n}(t)}} + \sum_{T_\mu=t_{\underline{n}(t)}} \mathrm{E}[g_\mu(\mathcal{X}^n_{t_{\underline{n}(t)}})]g'_\mu(\mathcal{X}^n_{t_{\underline{n}(t)}}) \\ &= \sum_{T_\mu\geq t_{\underline{n}(t)}} \mathrm{E}[g_\mu(\mathcal{X}^n_{T_\mu})]g'_\mu(\mathcal{X}^n_{T_\mu})V_{T_\mu}^{t_{\underline{n}(t)}}.\end{aligned}$$

This connection between Λ^n and V allows us to proof the convergence of the backwards approximation scheme Λ^n to Λ by proving the convergence

$$\sup_{t\in[0,T]} \mathrm{E}\left[\|\Phi^t - V^{\underline{n}(t)}\|^l_{\infty,t_{\underline{n}(t)},T}\right] \to 0 \tag{4.72}$$

for $|\Pi^{\mathrm{E}}| \to 0$. To see this, take a look at

$$\sup_{t\in[0,T]} \mathrm{E}\left[\left|\Lambda_t - \Lambda^n_{\underline{n}(t)}\right|^l\right]. \tag{4.73}$$

Notice that

$$\begin{aligned}&\left|\Lambda_t - \Lambda^n_{\underline{n}(t)}\right| \\ &= \left|\sum_{T_\mu\geq t} E[g_\mu(\mathcal{X}_{T_\mu})]g'_\mu(\mathcal{X}^u_{T_\mu})\Phi^t_{T_\mu} - \sum_{T_\mu\geq t_{\underline{n}(t)}} \mathrm{E}[g_\mu(\mathcal{X}^n_{T_\mu})]g'_\mu(\mathcal{X}^n_{T_\mu})V_{T_\mu}^{t_{\underline{n}(t)}}\right| \\ &\leq \sum_{T_\mu\geq t_{\underline{n}(t)}} \left|\mathrm{E}[g_\mu(\mathcal{X}_{T_\mu})]g'_\mu(\mathcal{X}_{T_\mu}) - \mathrm{E}[g_\mu(\mathcal{X}^n_{T_\mu})]g'_\mu(\mathcal{X}^n_{T_\mu})\right| |\Phi^t_{T_\mu}| \\ &\quad + \sum_{T_\mu\geq t_{\underline{n}(t)}} \left|\mathrm{E}[g_\mu(\mathcal{X}^n_{T_\mu})]g'(\mathcal{X}^n_{T_\mu})\right| \left|\Phi^t_{T_\mu} - V_{T_\mu}^{t_{\underline{n}(t)}}\right| \\ &\leq \sum_{T_\mu\geq t_{\underline{n}(t)}} \left(\left|\mathrm{E}[g_\mu(\mathcal{X}_{T_\mu})] - \mathrm{E}[g_\mu(\mathcal{X}^n_{T_\mu})]\right| |g'_\mu(\mathcal{X}_{T_\mu})| + \left|\mathrm{E}[g_\mu(\mathcal{X}^n_{T_\mu})]\right| \left|g'_\mu(\mathcal{X}_{T_\mu}) - g'_\mu(\mathcal{X}^n_{T_\mu})\right|\right)|\Phi^t_{T_\mu}| \\ &\quad + \sum_{T_\mu\geq t_{\underline{n}(t)}} \left|\mathrm{E}[g_\mu(\mathcal{X}^n_{T_\mu})]g'_\mu(\mathcal{X}^n_{T_\mu})\right| \left|\Phi^t_{T_\mu} - V_{T_\mu}^{t_{\underline{n}(t)}}\right|.\end{aligned}$$

By condition (G), (3.2), Lemma 4.7 and (4.69) this yields

$$\begin{aligned}&\left|\Lambda_t - \Lambda^n_{\underline{n}(t)}\right| \\ &\leq \sum_{T_\mu\geq t_{\underline{n}(t)}} \left(L^2\mathrm{E}[|\mathcal{X}_{T_\mu} - \mathcal{X}^n_{T_\mu}|] + \left(L\mathrm{E}[|\mathcal{X}^n_{T_\mu}|] + g_\mu(0)\right) L\left|\mathcal{X}_{T_\mu} - \mathcal{X}^n_{T_\mu}\right|\right)|\Phi^t_{T_\mu}|\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{T_\mu \geq t_{\underline{n}(t)}} L\left(L\mathrm{E}[|\mathcal{X}^n_{T_\mu}|] + g_\mu(0)\right) \left|\Phi^t_{T_\mu} - V^{t_{\underline{n}(t)}}_{T_\mu}\right| \\
&\leq CE\left[\|\mathcal{X} - \mathcal{X}^n\|_{\infty,0,T}\right] \|\Phi^t\|_{\infty,t,T} + C(D_{\mathcal{X},1} + \max_{\mu=1,\ldots,M} g_\mu(0))\|\mathcal{X} - \mathcal{X}^n\|_{\infty,0,T}\|\Phi^t\|_{\infty,t,T} \\
&+ C(D_{\mathcal{X},1} + \max_{\mu=1,\ldots,M} g_\mu(0)) \left\|\Phi^t - V^{t_{\underline{n}(t)}}\right\|_{\infty,t_{\underline{n}(t)},T},
\end{aligned}
$$

where C depends on M and L. Inserting this inequality into (4.73) and taking (4.69), the Jensen and the Hölder inequality into account, this yields

$$
\begin{aligned}
&\sup_{t\in[0,T]} \mathrm{E}\left[\left|\Lambda_t - \Lambda^n_{t_{\underline{n}(t)}}\right|^l\right] \\
&\leq CD_{\Phi,l}\mathrm{E}[\|\mathcal{X} - \mathcal{X}^n\|_{\infty,s_0,T}]^l + C(D_{\mathcal{X},1} + \max_{\mu=1,\ldots,M} g_\mu(0))^l \mathrm{E}\left[\|\mathcal{X} - \mathcal{X}^n\|^{2l}_{\infty,0,T}\right]^{\frac{1}{2}} D^{\frac{1}{2}}_{\Phi,2l} \\
&+ C(D_{\mathcal{X},1} + \max_{\mu=1,\ldots,M} g_\mu(0))^l \sup_{t\in[0,T]} \mathrm{E}\left[\left\|\Phi^t - V^{t_{\underline{n}(t)}}\right\|^l_{\infty,t_{\underline{n}(t)},T}\right].
\end{aligned}
$$

By Theorem 4.7, we have for every $u \in \mathcal{U}$ and $l \geq 2$, where $\delta_{1,l}$ is defined in (4.2),that

$$
\mathrm{E}\left[\|\mathcal{X} - \mathcal{X}^n\|^l_{\infty,0,T}\right] \leq D^l_{K_{\mathcal{X}},l}\delta^{l(2-p)\wedge\frac{1}{2}}_{1,2l},
$$

which yields by the monotonicity of L^l-norms

$$
\begin{aligned}
&\sup_{t\in[0,T]} \mathrm{E}\left[\left|\Lambda_t - \Lambda^n_{t_{\underline{n}(t)}}\right|^l\right] \\
&\leq CD_{\Phi,l}D^l_{K_{\mathcal{X}},2}\delta^{l(2-p)\wedge\frac{1}{2}}_{1,4} + C(D_{\mathcal{X},1} + \max_{\mu=1,\ldots,M} g_\mu(0))^l D^{\frac{1}{2}}_{\Phi,2l} D^l_{K_{\mathcal{X}},2l}\delta^{l(2-p)\wedge\frac{1}{2}}_{1,4l} \\
&+ C(D_{\mathcal{X},1} + \max_{\mu=1,\ldots,M} g_\mu(0))^l \sup_{t\in[0,T]} \mathrm{E}\left[\|\Phi^t - V^{t_{\underline{n}(t)}}\|^l_{\infty,t_{\underline{n}(t)},T}\right].
\end{aligned}
$$

Hence, there exist constants $D_1, D_2 > 0$ independent of u and n such that

$$
\begin{aligned}
&\sup_{t\in[0,T]} \mathrm{E}\left[\left|\Lambda_t - \Lambda^n_{t_{\underline{n}(t)}}\right|^l\right] \\
&\leq D_1\delta^{l(2-p)\wedge\frac{1}{2}}_{1,4l} + D_2 \sup_{t\in[0,T]} \mathrm{E}\left[\|\Phi^t - V^{t_{\underline{n}(t)}}\|^l_{\infty,t_{\underline{n}(t)},T}\right].
\end{aligned}
\tag{4.74}
$$

Now we will establish (4.72) and examine the corresponding convergence rate. Note that we defined three matrix valued Euler approximations in (4.70). For $t \in [0,T]$ and $i \in \{\underline{n}(t), \ldots, n-1\}$, we defined

$$
\varphi^{t_{\underline{n}(t)}}_{t_{i+1}} = \varphi^{t_{\underline{n}(t)}}_{t_i} + b_x(t_j, \xi^n_{t_i}, u)\varphi^{t_{\underline{n}(t)}}_{t_i}(t_{i+1} - t_i) + \sum_{j=1}^{m_1} \sigma^j_x(t_i, \xi^n_{t_i}, u)\varphi^{t_{\underline{n}(t)}}_{t_i}(w^j_{t_{i+1}} - w^j_{t_i})
$$

$$
\begin{aligned}
\hat{\varphi}^{t_{\underline{n}(t)}}_{t_{i+1}} &= \hat{\varphi}^{t_{\underline{n}(t)}}_{t_i} + \hat{b}_x(t_i, x^n_{t_i}, \xi^n_{t_i}, u)\hat{\varphi}^{t_{\underline{n}(t)}}_{t_i}(t_{i+1} - t_i) + \sum_{j=1}^{m_2} \hat{\sigma}^j_x(t_i, x^n_{t_i}, \xi^n_{t_i}, u)\hat{\varphi}^{t_{\underline{n}(t)}}_{t_i}(B^j_{t_{i+1}} - B^j_{t_i}) \\
\tilde{\varphi}^{t_{\underline{n}(t)}}_{t_{i+1}} &= \tilde{\varphi}^{t_{\underline{n}(t)}}_{t_i} + \left[\hat{b}_z(t_i, x^n_{t_i}, \xi^n_{t_i}, u)\varphi^{t_{\underline{n}(t)}}_{t_i} + \hat{b}_x(t_i, x^n_{t_i}, \xi^n_{t_i}, u)\tilde{\varphi}^{t_{\underline{n}(t)}}_{t_i}\right](t_{i+1} - t_i) \\
&\quad + \sum_{j=1}^{m_2} \left[\hat{\sigma}^j_z(t_i, x^n_{t_i}, \xi^n_{t_i}, u)\varphi^{t_{\underline{n}(t)}}_{t_i} + \hat{\sigma}^j_x(t_i, x^n_{t_i}, \xi^n_{t_i}, u)\tilde{\varphi}^{t_{\underline{n}(t)}}_{t_i}\right]\left(B^j_{t_{i+1}} - B^j_{t_i}\right),
\end{aligned}
$$

where

$$
\varphi^{t_{\underline{n}(t)}}_{t_{\underline{n}(t)}} = I_{n_1}, \quad \hat{\varphi}^{t_{\underline{n}(t)}}_{t_{\underline{n}(t)}} = I_{n_2}, \quad \tilde{\varphi}^{t_{\underline{n}(t)}}_{t_{\underline{n}(t)}} = 0 \in \mathbb{R}^{n_2 \times n_1}
$$

and the continuous interpolation defined trough (4.71) in the similar way. We will show that for $|\Pi^{\mathrm{E}}| \to 0$

$$
\begin{aligned}
\sup_{t\in[0,T]} \mathrm{E}\left[\|\phi^t - \varphi^{t_{\underline{n}(t)}}\|^l_{\infty, t_{\underline{n}(t)}, T}\right] &\to 0 \\
\sup_{t\in[0,T]} \mathrm{E}\left[\|\hat{\phi}^t - \hat{\varphi}^{t_{\underline{n}(t)}}\|^l_{\infty, t_{\underline{n}(t)}, T}\right] &\to 0 \\
\sup_{t\in[0,T]} \mathrm{E}\left[\|\tilde{\phi}^t - \tilde{\varphi}^{t_{\underline{n}(t)}}\|^l_{\infty, t_{\underline{n}(t)}, T}\right] &\to 0
\end{aligned}
$$

successively. Conceptually we repeat the same steps from the last section, by showing the boundedness of the respective Euler schemes first and then prove their convergence.

Lemma 4.14. *We have for a given $u \in \mathcal{U}$, for all $t \in [0,T]$ and for almost every $\omega \in \Omega$*

$$
\begin{aligned}
&\|\varphi\|_{p, t_{\underline{n}(t)}, T} \\
&\leq \sqrt{n_1}\left(2^{3p}(3C_1 m_1)^p\left(T^p + |w|^p_{p,0,T}\right) + 1\right)\exp\left(2^{3p}3(3C_1 m_1)^p\left(T^p + |w|^p_{p,0,T}\right) + 2\right)
\end{aligned}
$$

and

$$
\begin{aligned}
&\|\varphi\|_{\infty, t_{\underline{n}(t)}, T} \\
&\leq 2\sqrt{n_1}\left(2^{3p}(3C_1 m_1)^p\left(T^p + |w|^p_{p,0,T}\right) + 1\right)\exp\left(2^{3p}3(3C_1 m_1)^p\left(T^p + |w|^p_{p,0,T}\right) + 2\right) \\
&:= C_\varphi(\omega).
\end{aligned}
$$

For $l \geq 1$, we have $\varphi^{u, t_{\underline{n}(t)}} \in L^l_{\mathbb{F}}(\Omega, C([0,T]), \mathbb{R}^{n_1 \times n_1})$ with

$$
\sup_{t\in[0,T]} \mathrm{E}\left[\|\varphi^{u, t_{\underline{n}(t)}}\|^l_{\infty, t_{\underline{n}(t)}, T}\right] \leq \mathrm{E}\left[C^l_\varphi\right] := D_{\varphi, l}.
$$

Furthermore, we have for almost every $\omega \in \Omega$, $u \in \mathcal{U}$ and $t \in [0,T]$

$$\|\varphi_{\cdot}^{u,t_{\underline{n}(t)}}(\omega) - \varphi_{t_{\overline{n}(\cdot)}}^{u,t_{\underline{n}(t)}}(\omega)\|_{\infty,t_{\underline{n}(t)},T} \leq \max_{i=\underline{n}(t),\ldots,n-1} |\varphi^{u,t_{\underline{n}(t)}}|_{p,t_i,t_{i+1}} \leq Lm_1 C_\varphi(\omega)\delta(\omega)$$

and consequently

$$\sup_{t\in[0,T]} \mathrm{E}\left[\|\varphi_{\cdot}^{u,t_{\underline{n}(t)}} - \varphi_{t_{\overline{n}(\cdot)}}^{u,t_{\underline{n}(t)}}\|_{\infty,t_{\underline{n}(t)},T}^{l}\right] \leq (Lm_1)^l D_{\varphi,2l}^{\frac{1}{2}} \delta_{1,2l}^{l}.$$

Proof. Note that φ is very similar to y^n in the last section, the difference is that φ does not contain the partial differentials of the coefficient functions with respect to the parameter u and we need to be careful with the initial time of φ which is now t instead of 0. Let $\mathcal{A} \subset \Omega$, such that $P(\mathcal{A}) = 0$ and $w_{\cdot}(\omega)$ is continuous and of bounded p-variation ($p \in (1,2)$) for every $\omega \in \mathcal{A}^c$. Let $\omega \in \mathcal{A}^c$, $u \in \mathcal{U}$ and $t \in [0,T]$ be arbitrary, for notational simplicity we leave out the direct dependence of the involved processes on ω, u and $t_{\underline{n}(t)}$. Note that since we prolonged φ to the interval $[0,T]$ by setting $\varphi_s = I_{n_1}$ for all $s \in [0, t_{\underline{n}(t)}]$, we get that $|\varphi|_{p,0,T} = |\varphi|_{p,t_{\underline{n}(t)},T}$. Adapting the estimates from Lemma 4.9 it is easy to see that the path $\varphi_{\cdot}$ is an element of $C^p([0,T],\mathbb{R}^{n_1\times n_1})$ and that $\sigma_x^j(\cdot,\xi_{\cdot}^n,u)\varphi_{\cdot}$ is an element of $C^q([0,T],\mathbb{R}^{n_1\times n_1})$ for $q \in (2,\frac{p}{p-1})$ and every $j = 1,\ldots,m_1$. Again repeating the steps from Lemma 4.9 on the Euler partition Π^{E}, we obtain for every $t_l, t_k \in \Pi^{\mathrm{E}}$ such that $t_l, t_k \in [t_{\underline{n}(t)}, T]$ that

$$|\varphi|_{p,t_l,t_k} \leq 3C_1 m_1(|\varphi_{t_l}| + |\varphi|_{p,t_l,t_k})(1 + |\xi^n|_{p,t_l,t_k})(|t_k - t_l| + |w|_{p,t_l,t_k}).$$

Note that this inequality then also holds for every $t_l, t_k \in \Pi^{\mathrm{E}}$ such that $t_l, t_k \in [0,T]$. Now we have for every interval $[t_l,t_k] \in [0,T]$ which satisfies

$$|t_k - t_l| + |w|_{p,t_l,t_k} \leq \frac{1}{12C_1 m_1}$$

that

$$|\xi^n|_{p,t_l,t_k} \leq 1$$

by Remark 4.3 and therefore

$$|\varphi|_{p,t_l,t_k} \leq |\varphi_{t_l}|.$$

Hence, we can use Lemma 4.8 with $a = Lm_1$, $K_1 = 1$, $K_2 = \frac{1}{12C_1m_1} \leq \frac{1}{Lm_1} = \frac{1}{a}$ and obtain the estimate

$$\begin{aligned}|\varphi|_{p,t_{\underline{n}(t)},T} &= |\varphi|_{p,0,T} \\ &\leq |\varphi_{t_{\underline{n}(t)}}| \left(2^{3p}(3C_1m_1)^p\left(T^p + |w|_{p,0,T}^p\right) + 1\right) \exp\left(2^{3p}3(3C_1m_1)^p\left(T^p + |w|_{p,0,T}^p\right) + 2\right).\end{aligned}$$

We have $|\varphi_{t_{\underline{n}(t)}}| = |I_{n_1}| = \sqrt{n_1}$ and consequently

$$\begin{aligned}
&\|\varphi\|_{\infty,t_{\underline{n}(t)},T} \\
&\leq |I_{n_1}| + |\varphi|_{p,t_{\underline{n}(t)},T} \\
&\leq 2\sqrt{n_1}\left(2^{3p}(3C_1m_1)^p\left(T^p + |w|_{p,0,T}^p\right) + 1\right)\exp\left(2^{3p}3(3C_1m_1)^p\left(T^p + |w|_{p,0,T}^p\right) + 2\right) \\
&:= C_\varphi.
\end{aligned}$$

Since ω was arbitrary in $\mathcal{A}^c$ and t was arbitrary in $[0,T]$, the inequalities for the p-variation and uniform norm of $\varphi^{t_{\underline{n}(t)}}$ hold P-almost surely and for all $u \in \mathcal{U}$ and $t \in [0,T]$. The $\mathbb{F}$-adaptedness of $\varphi^{t_{\underline{n}(t)}}$ is a direct implication of its definition and the $\mathbb{F}$-adaptedness of ξ^n and w. Since w satisfies the exponential moment condition (2.48), we get

$$\sup_{t\in[0,T]} \mathrm{E}\left[\|\varphi^{t_{\underline{n}(t)}}\|_{\infty,t_{\underline{n}(t)},T}^l\right] \leq \mathrm{E}\left[C_\varphi^l\right] := D_{\varphi,l} < \infty.$$

Let t again be arbitrary in $[0,T]$ and let $s \in [t_i, t_{i+1}]$ for some $i \in \{\underline{n}(t), \ldots, n-1\}$, then

$$|\varphi_s - \varphi_{t_i}| \leq |\varphi|_{p,t_i,t_{i+1}} \leq Lm_1|\varphi_{t_i}|(|t_{i+1} - t_i| + |w|_{p,t_i,t_{i+1}}) \leq Lm_1\|\varphi\|_{\infty,t_{\underline{n}(t)},T}\delta.$$

Hence, we have P-almost surely

$$\|\varphi_\cdot - \varphi_{t_{\overline{n}(\cdot)}}\|_{\infty,t_{\underline{n}(t)},T} \leq \max_{i=\underline{n}(t),\ldots,n-1} |\varphi|_{p,t_i,t_{i+1}} \leq Lm_1C_\varphi\delta,$$

where δ is defined in (4.1). Consequently by the Hölder inequality

$$\begin{aligned}
E\left[\|\varphi_\cdot - \varphi_{t_{\overline{n}(\cdot)}}\|_{\infty,t_{\underline{n}(t)},T}^l\right] &\leq (Lm_1)^l\mathrm{E}\left[C_\varphi^{2l}\right]^{\frac{1}{2}}\mathrm{E}\left[\delta^{2l}\right]^{\frac{1}{2}} \\
&\leq (Lm_1)^l\mathrm{E}\left[C_\varphi^{2l}\right]^{\frac{1}{2}}\delta_{1,2l}^l \\
&\leq (Lm_1)^l D_{\varphi,2l}^{\frac{1}{2}}\delta_{1,2l}^l.
\end{aligned}$$

Since the right hand side of the last inequality is independent of t and t was arbitrary in $[0,T]$, we conclude

$$\sup_{t\in[0,T]} \mathrm{E}\left[\|\varphi_\cdot - \varphi_{t_{\overline{n}(\cdot)}}\|_{\infty,t_{\underline{n}(t)},T}^l\right] \leq (Lm_1)^l D_{\varphi,2l}^{\frac{1}{2}}\delta_{1,2l}^l.$$

□

Theorem 4.15. *We have for a given $u \in \mathcal{U}$, for all $t \in [0,T]$ and for almost every $\omega \in \Omega$*

$$\|\phi^{t,u}(\omega) - \varphi^{t_{\underline{n}(t)},u}(\omega)\|_{\infty,t_{\underline{n}(t)},T} \leq K_\phi(\omega)\delta(\omega)^{2-p}, \tag{4.75}$$

where the random variable K_ϕ has moments of all orders and is independent of n, u and t. Furthermore, we have

$$\sup_{t\in[0,T]} E\left[\|\phi^{t,u}-\varphi^{t_{\underline{n}(t)},u}\|^l_{\infty,t_{\underline{n}(t)},T}\right]^{\frac{1}{l}} \leq \mathrm{E}\left[K_\phi^{2l}\right]^{\frac{1}{2l}} \delta_{1,2l}^{2-p} := D_{K_\phi,2l}\delta_{1,2l}^{2-p}$$

for all $l \geq 1$.

Proof. Let $\mathcal{A} \subset \Omega$, such that $P(\mathcal{A}) = 0$ and $w.(\omega)$ is continuous and of bounded p-variation for every $\omega \in \mathcal{A}^c$. Let $\omega \in \mathcal{A}^c$, $u \in \mathcal{U}$ and $t \in [0,T]$ be arbitrary and define $\gamma_s^{t,u} = \phi_s^{t,u} - \varphi_s^{t_{\underline{n}(t)},u}$ for all $s \in [0,T]$, for notational simplicity leave out the direct dependence of the involved processes on ω, u and t. Focusing on the interval $[t_{\underline{n}(t)},T]$, we can repeat the arguments form Theorem 4.4 analogously, where we just exchange in all the estimates for $I_1,\dots,I_6$ in the short time step case and $I_1,\dots,I_3$ in the multistep case the constants $(1+C_y)$ by C_ϕ and $(1+C_{y^n})$ by C_φ. We obtain with

$$D_1(\omega) := 8m_1^2C_1^2(C_\phi + C_{\varphi^n})^2(2+C_\xi+C_{\xi^n})(1+4D_w(\omega))K_\xi\left(1+\varphi_1(0,T)^{\frac{1}{p}}C_w^{p\beta-1}+\varphi_2(0,T)^{\frac{1}{p}}\right)$$

for every $[r,s] \subset [t_{\underline{n}(t)},T]$, that

$$\begin{aligned}|\gamma|_{p,r,s} &\leq 4C_1m_1\left(1+|\xi^n|_{p,r,s}\right)\|\gamma\|_{p,r,s}(|s-r|+|w|_{p,r,s}) \\ &\quad + D_1(\omega)\delta^{2-p}\left(\varphi_1(r,s)^{\frac{1}{p}}+\varphi_2(r,s)^{\frac{1}{p}}+|s-r|+|w|_{p,r,s}\right) \\ &\leq 4C_1m_1(1+|\xi^n|_{p,r,s})(|\gamma_r|+|\gamma|_{p,r,s})(|s-r|+|w|_{p,r,s})+D_2(\omega)\delta^{2-p},\end{aligned}$$

where

$$D_2(\omega) := D_1(\omega)\left(\varphi_1(0,T)^{\frac{1}{p}}+\varphi_2(0,T)^{\frac{1}{p}}+|T|+|w|_{p,0,T}\right)$$

is P-a.s. bounded independently of u and n and has finite moments of all orders. Now we set $\gamma_s = \gamma_{t_{\underline{n}(t)}}$ for all $s \in [0, t_{\underline{n}(t)}]$, such that, taking Remark 4.3 into account, we can argue that for every $[r,s] \subset [0,T]$ such that

$$|s-r|+|w|_{p,r,s} \leq \frac{1}{16C_1m_1}$$

we have

$$|\xi^n|_{p,r,s} \leq 1$$

and

$$|\gamma|_{p,r,s} \leq 2D_2(\omega)\delta^{2-p}+|\gamma_r|,$$

where D_2 is independent of t, n and u. Now we can use Lemma 2.20 and obtain the estimate

$$|\gamma|_{p,t_{\underline{n}(t)},T} \leq (2D_2(\omega)\delta^{2-p}+|\gamma_{t_{\underline{n}(t)}}|)e^{2^{5p}(m_1C_1)^p(T^p+|w|^p_{p,0,T})}$$

and therefore

$$\|\gamma\|_{\infty,t_{\underline{n}(t)},T} \leq (2D_2(\omega)\delta^{2-p} + 2|\gamma_{t_{\underline{n}(t)}}|)e^{2^{5p}(m_1C_1)^p(T^p+|w|^p_{p,0,T})}.$$

We have by (2.32) used on the interval $[t, t_{\underline{n}(t)}]$

$$\begin{aligned}|\gamma_{t_{\underline{n}(t)}}| &= |\phi_{t_{\underline{n}(t)}} - \varphi_{t_{\underline{n}(t)}}| \\ &= \|\phi - \varphi\|_{\infty,t,t_{\underline{n}(t)}} \\ &\leq |\phi|_{p,t,t_{\underline{n}(t)}} \\ &\leq 2C_1m_1C_\phi(1+C_\xi)\delta \\ &\leq D_2(\omega)\delta^{2-p}.\end{aligned}$$

Which yields

$$\begin{aligned}\|\gamma\|_{\infty,t_{\underline{n}(t)},T} &\leq 4D_2(\omega)\delta^{2-p}e^{2^{5p}(m_1C_1)^p(T^p+|w|^p_{p,0,T})} \\ &\leq K_\phi(\omega)\delta^{2-p},\end{aligned}$$

where

$$K_\phi(\omega) := 4D_2(\omega)e^{2^{5p}(m_1C_1)^p(T^p+|w|^p_{p,0,T})}.$$

Since $K_\phi(\omega)$ is P-a.s. bounded independently of u, n and t and has finite moments of all orders, we get the estimate (4.75). The same holds for δ, such that we can use the Hölder inequality and obtain for $l \geq 1$ and $\delta_{1,l} := \mathrm{E}[\delta^l]^{\frac{1}{l}}$ the estimate

$$\sup_{t\in[0,T]} \mathrm{E}\left[\|\gamma^t\|^l_{\infty,t_{\underline{n}(t)},T}\right]^{\frac{1}{l}} \leq \mathrm{E}\left[K_\phi^{2l}\right]^{\frac{1}{2l}} \delta_{1,2l}^{2-p} \leq D_{K_\phi,2l}\delta_{1,2l}^{2-p}.$$

$\square$

Now we come to the approximation of the solution to the Brownian motion driven SDEs (4.67) and (4.68). These are matrix valued linear SDEs which are very similar to the inhomogenous linear SDE (2.57), if we exchange $\hat{y}$ by $\hat{\varphi}$ and neglect the derivatives w.r.t. z and u respectively exchange $\hat{y}$ and y by $\tilde{\varphi}$ and φ and neglect the derivatives w.r.t. u. So in the following lemmas and theorems we will refer to the proof of Lemma 4.11 and Theorem 4.12, since the calculations are analogous.

Lemma 4.16. *We have for a given $u \in \mathcal{U}$ and $l \geq 2$ that*

$$\sup_{t\in[0,T]} \mathrm{E}\left[\|\hat{\varphi}^{u,t_{\underline{n}(t)}}\|^l_{\infty,t_{\underline{n}(t)},T}\right] \leq Cn^{\frac{l}{2}}_2 e^C := D_{\hat{\varphi},l}.$$

Furthermore, we obtain for $\delta_2 = \max_{i=0,\dots,n-1} |t_{i+1} - t_i|$, *that*

$$\sup_{t\in[0,T]} \mathrm{E}\left[\|\hat{\varphi}_{\cdot}^{u,t_{\underline{n}(t)}} - \hat{\varphi}_{t_{\overline{n}(\cdot)}}^{u,t_{\underline{n}(t)}}\|_{\infty,t_{\underline{n}(t)},T}^{l}\right] \leq C D_{\hat{\varphi},l}\delta_2^{\frac{l}{2}}. \tag{4.76}$$

The constant $C > 0$ *only depends on* T, l, m_2 *and* L.

Proof. Let $t \in [0, T]$ be arbitrary, we omit the direct dependence of the involved processes on u and t for notational simplicity. It is easy to see that $\hat{\varphi} \in L^l_{\mathbb{F}}(\Omega, C([t_{\underline{n}(t)}, T], \mathbb{R}^{n_2\times n_2})$ for every $l \geq 2$ and that the process

$$(\omega, s) \mapsto \sum_{j=1}^{m_2} \int_{t_{\underline{n}(t)}}^{s} \hat{\sigma}_x^{n,j}(t_{\overline{n}(v)})\hat{\varphi}_{t_{\overline{n}(v)}}(\omega)\, dB_v^j(\omega)$$

for $s \in [t_{\underline{n}(t)}, T]$ is well defined and an $n_2 \times n_2$-dimensional matrix of $\mathbb{F}$-martingales. Using the same arguments as in the proof of Lemma 4.11, we get for every $s \in [t_{\underline{n}(t)}, T]$

$$\mathrm{E}[\|\hat{\varphi}\|_{\infty,t_{\underline{n}(t)},s}^{l}] \leq C n_2^{\frac{l}{2}} + C \int_{t_{\underline{n}(t)}}^{s} \mathrm{E}\left[\|\hat{\varphi}\|_{\infty,t_{\underline{n}(t)},v}^{l}\right] dv$$

and we conclude by the Gronwall inequality

$$E\left[\|\hat{\varphi}\|_{\infty,t_{\underline{n}(t)},T}^{l}\right] \leq C n_2^{\frac{l}{2}} e^{C} := D_{\hat{\varphi},l}.$$

Since the right hand side of the last inequality does not depend on t, we get

$$\sup_{t\in[0,T]} \mathrm{E}\left[\|\hat{\varphi}^{u,t}\|_{\infty,t_{\underline{n}(t)},T}^{l}\right] \leq D_{\hat{\varphi},l}.$$

Now let $t \in [0, T]$ and $s \in [t_i, t_{i+1}]$ for $i \in \{\underline{n}(t), \dots, n-1\}$, we have

$$\begin{aligned} \mathrm{E}\left[|\hat{\varphi}_s - \hat{\varphi}_{t_i}|^l\right] \leq{}& C\mathrm{E}\Bigg[\Big|\int_{t_i}^{s} \hat{b}_x(t_i, x_{t_i}^n, \xi_{t_i}^n, u)\hat{\varphi}_{t_i}\, d_v\Big|^l\Bigg] \\ &+ C\sum_{j=1}^{m_2} \mathrm{E}\Bigg[\Big|\int_{t_i}^{s} \hat{\sigma}_x^j(t_i, x_{t_i}^n, \xi_{t_i}^n, u)\hat{\varphi}_{t_i}\, dB_v^j\Big|^l\Bigg]. \end{aligned}$$

By the boundedness of the coefficient functions and by the Burkholder-Davis-Gundy inequality, we get

$$\begin{aligned} \mathrm{E}\left[|\hat{\varphi}_s - \hat{\varphi}_{t_i}|^l\right] &\leq C(s-t_i)^l \mathrm{E}\Bigg[\|\hat{\varphi}\|_{\infty,t_{\underline{n}(t)},T}^{l}\Bigg] + C(s-t_i)^{\frac{l}{2}-1}\mathrm{E}\Bigg[\int_{t_i}^{s} \|\hat{\varphi}\|_{\infty,t_{\underline{n}(t)},T}^{l}\, dr\Bigg] \\ &\leq C D_{\hat{\varphi},l}\delta_2^{\frac{l}{2}}. \end{aligned}$$

This yields the estimate (4.76) for $l \geq 2$. $\square$

Theorem 4.17. *We have for a given $u \in \mathcal{U}$ and $l \geq 2$ that*

$$\sup_{t\in[0,T]} \mathrm{E}\left[\left\|\hat{\phi}^{u,t} - \hat{\varphi}^{u,t_{\underline{n}(t)}}\right\|^l_{\infty,t_{\underline{n}(t)},T}\right]^{\frac{1}{l}} \leq D_{K_{\hat{\phi}},l}\delta_{1,4l}^{(2-p)\wedge\frac{1}{2}}$$

for any $l \geq 2$, where the constant $D_{K_{\hat{\phi}},l}$ is independent of u and n.

Proof. Let $t \in [0,T]$ and $s \in [t_{\underline{n}(t)}, T]$ and consider the processes $\hat{\phi}^{u,t}$ and $\hat{\varphi}^{u,t_{\underline{n}(t)}}$ on $[t_{\underline{n}(t)}, T]$, where we omit the indexes u, t and $t_{\underline{n}(t)}$ for readability. Let C be a generic constant that is only dependent on L, l, m_2 and T. Since now we focus on the interval $[t_{\underline{n}(t)}, T]$ instead of $[0,T]$ as in Theorem 4.12, we need to take care of the term

$$I_0 = \mathrm{E}\left[|\hat{\phi}_{t_{\underline{n}(t)}} - \hat{\varphi}_{t_{\underline{n}(t)}}|^l\right] \leq \mathrm{E}\left[\|\hat{\phi} - \hat{\varphi}\|^l_{\infty,t,t_{\underline{n}(t)}}\right] = \mathrm{E}\left[\|\hat{\phi} - I_{n2}\|^l_{\infty,t,t_{\underline{n}(t)}}\right].$$

We have using the Jensen and Burkolder-Davis-Gundy inequality and condition (B_1) and (B_2)

$$\begin{aligned}
I_0 &\leq C\mathrm{E}\Bigg[\sup_{s\in[t,t_{\underline{n}(t)}]}\left|\int_t^s \hat{b}_x(r,x_r,\xi_r,u)\hat{\phi}_r\, d_r\right|^l\Bigg] \\
&\quad + C\sum_{j=1}^{m_2}\mathrm{E}\Bigg[\sup_{s\in[t,t_{\underline{n}(t)}]}\left|\int_t^s \hat{\sigma}^j_x(r,x_r,\xi_r,u)\hat{\phi}_r\, dB^j_r\right|^l\Bigg] \\
&\leq C(t_{\underline{n}(t)} - t)^{l-1}\mathrm{E}\left[\int_t^{t_{\underline{n}(t)}} |\hat{\phi}_r|^l\, d_r\right] + C(t_{\underline{n}(t)} - t)^{\frac{l}{2}-1}\mathrm{E}\left[\int_t^{t_{\underline{n}(t)}} |\hat{\phi}_r|^l\, dr\right] \\
&\leq CD_{\hat{\phi},l}\delta_2^{\frac{l}{2}} \\
&\leq D_{0,l}\delta_2^{\frac{l}{2}},
\end{aligned}$$

where

$$D_{0,l} := CD_{\hat{\phi},l}.$$

The other estimates are completely analogous to the estimates from the proof of Theorem 4.12. By adapting the corresponding constants

$$\begin{aligned}
D_{1,l} := CD_{\hat{\phi},l}D^{\frac{l}{2}}_{w,4l} + CD^{\frac{1}{2}}_{\hat{\varphi},2l}\Big(&(1 + D_{x,2l} + D_{\xi,2l})^{\frac{1}{2}}D^{\frac{l}{2}}_{w,4l} + D^l_{K_x,4l} \\
&+ (1 + D_{x^n,2l} + D_{\xi^n,2l})^{\frac{1}{2}}D^{\frac{l}{2}}_{w,4l} + D^l_{K_\xi,4l}D^{\frac{l}{2}}_{w,4l} + D^l_{w,4l}\Big)
\end{aligned}$$

and $D_{2,l}$ which only differs from $D_{1,l}$ in the constant C because of the Burkholder-Davis-Gundy inequality, we get that

$$\begin{aligned}
\mathrm{E}\left[\|\hat{\phi}-\hat{\varphi}\|^l_{\infty,t_{\underline{n}(t)},s}\right] &\le C\int_{t_{\underline{n}(t)}}^s \mathrm{E}\left[\|\hat{\phi}-\hat{\varphi}\|^l_{\infty,t_{\underline{n}(t)},v}\right]dv + D_{0,l}\delta_2^{\frac{l}{2}} + D_{1,l}\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} + D_2\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})}\\
&\le C\int_{t_{\underline{n}(t)}}^s \mathrm{E}\left[\|\hat{\phi}-\hat{\varphi}\|^l_{\infty,t_{\underline{n}(t)},v}\right]dv + D_l\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})},
\end{aligned}$$

where

$$D_l := \max\{D_{0,l}D_{w,4l}^{\frac{l}{2}}, D_{1,l}, D_{2,l}\}.$$

By the Gronwall inequality, we conclude

$$\begin{aligned}
\mathrm{E}\left[\|\hat{\phi}^t-\hat{\varphi}^{t_{\underline{n}(t)}}\|^l_{\infty,t_{\underline{n}(t)},T}\right] &\le D_l\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})}e^C\\
&\le D_{K_{\hat{\phi}},l}\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})}.
\end{aligned}$$

Since the right hand side of the last inequality does not depend on t, the assertion follows. $\square$

Lemma 4.18. *We have for a given $u\in\mathcal{U}$ and $l\ge 2$ that*

$$\sup_{t\in[0,T]}\mathrm{E}\left[\|\tilde{\varphi}^{u,t_{\underline{n}(t)}}\|^l_{\infty,t_{\underline{n}(t)},T}\right] \le C(1+D_{\tilde{\varphi},l})e^C := D_{\tilde{\varphi},l}.$$

Furthermore we obtain for $\delta_2 = \max_{i=0,\dots,n-1}|t_{i+1}-t_i|$, that

$$\sup_{t\in[0,T]}\mathrm{E}\left[\|\tilde{\varphi}^{u,t_{\underline{n}(t)}}_{\cdot} - \tilde{\varphi}^{u,t_{\underline{n}(t)}}_{t_{\overline{n}(\cdot)}}\|^l_{\infty,t_{\underline{n}(t)},T}\right] \le C(D_{\tilde{\varphi},l}+D_{\varphi,l})\delta_2^{\frac{l}{2}}. \tag{4.77}$$

The constant $C>0$ only depends on T, l, m_2 and L.

Proof. Let $t\in[0,T]$ be arbitrary, we omit the direct dependence of the involved processes on u and t for notational simplicity. It is easy to see that $\tilde{\varphi}\in L^l_{\mathbb{F}}(\Omega, C([t_{\underline{n}(t)},T]),\mathbb{R}^{n_2\times n_1})$ and that the process

$$(\omega,s)\mapsto \sum_{j=1}^{m_2}\int_{t_{\underline{n}(t)}}^s \hat{\sigma}_x^{n,j}(t_{\overline{n}(v)})\tilde{\varphi}_{t_{\overline{n}(v)}}(\omega) + \hat{\sigma}_z^{n,j}(t_{\overline{n}(v)})\varphi_{t_{\overline{n}(v)}}(\omega)\,dB^j_v(\omega)$$

for $s\in[t_{\underline{n}(t)},T]$ is well defined and an $n_2\times n_1$-dimensional matrix of $\mathbb{F}$-martingales. Using the same arguments as in the proof of Lemma 4.11 and the results from Lemma 4.14, we get for every $s\in[t_{\underline{n}(t)},T]$

$$\mathrm{E}\left[\|\tilde{\varphi}\|^l_{\infty,t_{\underline{n}(t)},s}\right] \le CD_{\varphi,l} + C\int_{t_{\underline{n}(t)}}^s \mathrm{E}\left[\|\tilde{\varphi}\|^l_{\infty,t_{\underline{n}(t)},v}\right]dv$$

and we conclude by the Gronwall inequality

$$E\left[\|\tilde{\varphi}\|^l_{\infty,t_{\underline{n}(t)},T}\right] \leq CD_{\varphi,l}e^C := D_{\tilde{\varphi},l}.$$

Since the right hand side of the last inequality does not depend on t, we get

$$\sup_{t\in[0,T]} \mathrm{E}\left[\|\tilde{\varphi}^{u,t}\|^l_{\infty,t_{\underline{n}(t)},T}\right] \leq D_{\tilde{\varphi},l}.$$

Now let $t \in [0,T]$ and $s \in [t_i, t_{i+1}]$ for $i \in \{\underline{n}(t), \dots, n-1\}$, we have

$$\begin{aligned}\mathrm{E}\left[|\tilde{\varphi}_s - \tilde{\varphi}_{t_i}|^l\right] \leq C\mathrm{E}&\left[\left|\int_{t_i}^s \hat{b}_x(t_i, x^n_{t_i}, \xi^n_{t_i}, u)\tilde{\varphi}_{t_i} + \hat{b}_z(t_i, x^n_{t_i}, \xi^n_{t_i}, u)\varphi_{t_i}\, d_v\right|^l\right]\\ &+ C\sum_{j=1}^{m_2}\mathrm{E}\left[\left|\int_{t_i}^s \hat{\sigma}^j_x(t_i, x^n_{t_i}, \xi^n_{t_i}, u)\tilde{\varphi}_{t_i} + \hat{\sigma}^j_z(t_i, x^n_{t_i}, \xi^n_{t_i}, u)\varphi_{t_i}\, dB^j_v\right|^l\right].\end{aligned}$$

By the boundedness of the coefficient functions, Lemma 4.14 and by the Burkholder-Davis-Gundy inequality, we get

$$\begin{aligned}\mathrm{E}\left[|\tilde{\varphi}_s - \tilde{\varphi}_{t_i}|^l\right] &\leq C(s-t_i)^l\mathrm{E}\left[\|\tilde{\varphi}\|^l_{\infty,t_{\underline{n}(t)},T} + \|\varphi\|^l_{\infty,t_{\underline{n}(t)},T}\right]\\ &\quad + C(s-t_i)^{\frac{l}{2}-1}\mathrm{E}\left[\int_{t_i}^s \|\tilde{\varphi}\|^l_{\infty,t_{\underline{n}(t)},T} + \|\varphi\|^l_{\infty,t_{\underline{n}(t)},T}\, dv\right]\\ &\leq C(D_{\tilde{\varphi},l} + D_{\varphi,l})\delta^{\frac{l}{2}}_2.\end{aligned}$$

This yields the estimate (4.77) for $l \geq 2$. □

Theorem 4.19. *We have for a given $u \in \mathcal{U}$ and $l \geq 2$ that*

$$\sup_{t\in[0,T]} \mathrm{E}\left[\left\|\tilde{\phi}^{u,t} - \tilde{\varphi}^{u,t_{\underline{n}(t)}}\right\|^l_{\infty,t_{\underline{n}(t)},T}\right]^{\frac{1}{l}} \leq D_{K_{\tilde{\phi}},l}\delta^{(2-p)\wedge\frac{1}{2}}_{1,4l}$$

for any $l \geq 2$, where the constant $D_{K_{\tilde{\phi}},l}$ is independent of u and n.

Proof. Let $t \in [0,T]$ and $s \in [t_{\underline{n}(t)}, T]$ and consider the processes $\tilde{\phi}^{u,t}$ and $\tilde{\varphi}^{u,t_{\underline{n}(t)}}$ on $[t_{\underline{n}(t)}, T]$, where we omit the indexes u, t and $t_{\underline{n}(t)}$ for readability. Let $C > 0$ be a constant which only depends on T, l, m_2 and L. Since now we focus on the interval $[t_{\underline{n}(t)}, T]$ instead of $[0,T]$ as in Theorem 4.12, we need to take care of the term

$$I_0 = \mathrm{E}\left[|\tilde{\phi}_{t_{\underline{n}(t)}} - \tilde{\varphi}_{t_{\underline{n}(t)}}|^l\right] \leq \mathrm{E}\left[\|\tilde{\phi} - \tilde{\varphi}\|^l_{\infty,t,t_{\underline{n}(t)}}\right] = \mathrm{E}\left[\|\tilde{\phi}\|^l_{\infty,t,t_{\underline{n}(t)}}\right].$$

We have using the Jensen and Burkolder-Davis-Gundy inequality and condition (B_1) and (B_2)

$$
\begin{aligned}
I_0 &\leq C\mathrm{E}\Big[\sup_{s\in[t,t_{\underline{n}(t)}]}\Big|\int_t^s \hat{b}_x(r,x_r,\xi_r,u)\tilde{\phi}_r + \hat{b}_z(r,x_r,\xi_r,u)\phi_r\, d_r\Big|^l\Big] \\
&\quad + C\sum_{j=1}^{m_2}\mathrm{E}\Big[\sup_{s\in[t,t_{\underline{n}(t)}]}\Big|\int_t^s \hat{\sigma}_x^j(r,x_r,\xi_r,u)\tilde{\phi}_r + \hat{\sigma}_z^j(r,x_r,\xi_r,u)\phi_r\, dB_r^j\Big|^l\Big] \\
&\leq C(t_{\underline{n}(t)}-t)^{l-1}\mathrm{E}\left[\int_t^{t_{\underline{n}(t)}}(|\tilde{\phi}_r|+|\phi_r|)^l\, d_r\right] + C(t_{\underline{n}(t)}-t)^{\frac{l}{2}-1}\mathrm{E}\left[\int_t^{t_{\underline{n}(t)}}(|\tilde{\phi}_r|+|\phi_r|)^l\, dr\right] \\
&\leq C(D_{\tilde{\phi},l}+D_{\phi,l})\delta_2^{\frac{l}{2}} \\
&\leq D_{0,l}\delta_2^{\frac{l}{2}},
\end{aligned}
$$

where

$$
D_{0,l} := C(D_{\tilde{\phi},l}+D_{\phi,l}).
$$

By repeating the arguments from the proof of Theorem 4.12 and the results from Lemma 4.14 and Theorem 4.15, we have for

$$
\begin{aligned}
D_{1,l} &:= C(D_{\tilde{\varphi},l}+D_{\varphi,l})D_{w,4l}^{\frac{l}{2}} + CD_{\tilde{\varphi},2l}^{\frac{1}{2}}\bigg((1+D_{x,2l}+D_{\xi,2l})^{\frac{1}{2}}D_{w,4l}^{\frac{l}{2}} + D_{K_x,4l}^l \\
&\qquad + (1+D_{x^n,2l}+D_{\xi^n,2l})^{\frac{1}{2}}D_{w,4l}^{\frac{l}{2}} + D_{K_\xi,4l}^l D_{w,4l}^{l\frac{1}{2}} + D_{w,4l}^l\bigg) \\
D_{2,l} &:= CD_{K_\phi,2l}^l D_{w,4l}^{\frac{l}{2}} + CD_{\varphi,2l}^{\frac{1}{2}}D_{w,4l}^l \\
&\qquad + CD_{\varphi,2l}^{\frac{1}{2}}\bigg((1+D_{x,2l}+D_{\xi,2l})^{\frac{1}{2}}D_{w,4l}^{\frac{l}{2}} + D_{K_x,4l}^l + (1+D_{x^n,2l}+D_{\xi^n,2l})^{\frac{1}{2}}D_{w,4l}^{\frac{l}{2}} \\
&\qquad + D_{K_\xi,4l}^l D_{w,4l}^{\frac{l}{2}} + D_{w,4l}^l\bigg)
\end{aligned}
$$

and constants $D_{3,l}$, $D_{4,l}$ such that $D_{1,l}$ and $D_{3,l}$, respectively $D_{2,l}$ and $D_{4,l}$ only differ in the constant C because of the Burkholder-Davis-Gundy inequality, that

$$
\begin{aligned}
\mathrm{E}\left[\|\tilde{\phi}-\tilde{\varphi}\|_{\infty,t_{\underline{n}(t)},s}^l\right] &\leq C\int_{t_{\underline{n}(t)}}^s \mathrm{E}\left[\|\tilde{\phi}-\tilde{\varphi}\|_{\infty,t_{\underline{n}(t)},s}^l\right] dv + D_{0,l}\delta_2^{\frac{l}{2}} + D_{1,l}\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} \\
&\quad + D_{3,l}\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} + D_{2,l}\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} + D_{4,l}\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} \\
&\leq C\int_{t_{\underline{n}(t)}}^s \mathrm{E}\left[\|\tilde{\phi}-\tilde{\varphi}\|_{\infty,t_{\underline{n}(t)},v}^l\right] dv + D_l\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})},
\end{aligned}
$$

where

$$
D_l := C\max\{D_{0,l}D_{w,4l}^{\frac{l}{2}}, D_{1,l}\ldots, D_{4,l}\}.
$$

By the Gronwall inequality, we conclude

$$\mathrm{E}\left[\|\tilde{\phi}^t - \tilde{\varphi}^{t_{\underline{n}(t)}}\|^l_{\infty,t_{\underline{n}(t)},T}\right] \le D_l \delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} e^C$$
$$:= D^l_{K_{\tilde{\phi}},l} \delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})}.$$

Since the right hand side of the last inequality does not depend on t, the assertion follows. $\square$

The following theorem is the second main result of this thesis, where we summarize all the results of this subsection.

Theorem 4.20. *For all $u \in \mathcal{U}$ and $l \ge 2$, we have*

$$\sup_{t\in[0,T]} \mathrm{E}\left[|\Lambda_t - \Lambda^n_{t_{\underline{n}(t)}}|^l\right]^{\frac{1}{l}} \le D_{K_\Lambda,l} \delta_{1,4l}^{(2-p)\wedge\frac{1}{2}}.$$

and under the assumption (HA), we get

$$\sup_{t\in[0,T]} \mathrm{E}\left[|\Lambda_t - \Lambda^n_{t_{\underline{n}(t)}}|^l\right]^{\frac{1}{l}} \le \tilde{D}_{K_\Lambda,l} \delta_2^{(2H-1)\wedge\frac{1}{2}},$$

for constants $D_{K_\Lambda,l}$ and $\tilde{D}_{K_\Lambda,l}$ independent of u and n.

Proof. Let $C > 0$ be a constant only dependent on $L, M, l, D_{\mathcal{X}^n,1}, D_{\Phi,l}, D^l_{K_{\mathcal{X}},l}$ and $\max_{\mu=1,\dots,M} g_\mu(0)$. By (4.74), we have

$$\begin{aligned}
&\sup_{t\in[0,T]} \mathrm{E}\left[\left|\Lambda_t - \Lambda^n_{t_{\underline{n}(t)}}\right|^l\right] \\
&\le C\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} + C \sup_{t\in[0,T]} \mathrm{E}\left[\|\Phi^t - V^{t_{\underline{n}(t)}}\|^l_{\infty,t_{\underline{n}(t)},T}\right] \\
&\le C\delta_{1,4l}^{l((2-p)\wedge\frac{1}{2})} + C \sup_{t\in[0,T]} \mathrm{E}\left[\|\phi^t - \varphi^{t_{\underline{n}(t)}}\|^l_{\infty,t_{\underline{n}(t)},T}\right] \\
&\quad + C \sup_{t\in[0,T]} \mathrm{E}\left[\|\hat{\phi}^t - \hat{\varphi}^{t_{\underline{n}(t)}}\|^l_{\infty,t_{\underline{n}(t)},T}\right] + C \sup_{t\in[0,T]} \mathrm{E}\left[\|\tilde{\phi}^t - \tilde{\varphi}^{t_{\underline{n}(t)}}\|^l_{\infty,t_{\underline{n}(t)},T}\right].
\end{aligned}$$

Taking Theorem 4.15, Theorem 4.17 and Theorem 4.19 into account, this yields

$$\begin{aligned}
&\sup_{t\in[0,T]} \mathrm{E}\left[\left|\Lambda_t - \Lambda^n_{t_{\underline{n}(t)}}\right|^l\right]^{\frac{1}{l}} \\
&\le C\delta_{1,4l}^{(2-p)\wedge\frac{1}{2}} + CD_{K_\phi,2l}\delta_{1,2l}^{2-p} + CD_{K_{\hat{\phi}},l}\delta_{1,4l}^{(2-p)\wedge\frac{1}{2}} + CD_{K_{\tilde{\phi}},l}\delta_{1,4l}^{(2-p)\wedge\frac{1}{2}} \\
&\le D_{K_\Lambda,l}\delta_{1,4l}^{(2-p)\wedge\frac{1}{2}},
\end{aligned}$$

where

$$D_{K_\Lambda,l} := C(1 + D_{K_\phi,2l} D_{w,4l}^{\frac{1}{2}} + D_{K_{\hat{\phi}},l} + D_{K_{\tilde{\phi}},l}).$$

Under the assumption (HA), we follow the same arguments as in Subsection 4.1.3 to proof the assertion. $\square$

4.3 Discretization of the cost function and its gradient

Let $t_0 = T_1 \leq \ldots, T_m = T$ be a sequence of times in $[0,T]$, and $(g_\mu)_{\mu=1,\ldots,M}$ be a sequence of functions satisfying condition (G). Our cost function is given by

$$J : \mathcal{U} \to \mathbb{R}, u \mapsto \frac{1}{2}\sum_{\mu=1}^{M} \mathrm{E}[g_\mu(\mathcal{X}^u_{T_\mu})]^2.$$

The problem we want to approximate numerically is given by

$$\text{(P) Find} \qquad \min_{u\in\mathcal{U}} J(u) = \min_{u\in\mathcal{U}} \frac{1}{2}\sum_{\mu=1}^{M} \mathrm{E}[g_\mu(\mathcal{X}^u_{T_\mu})]^2$$

subject to

$$\mathcal{X}^u_t = \begin{pmatrix} \xi^u_t \\ x^u_t \end{pmatrix} = \begin{pmatrix} \xi_0(u) \\ x_0(u) \end{pmatrix} + \int_0^t \begin{pmatrix} b(r,\xi^u_r,u) \\ \hat{b}(r,x^u_r,\xi^u_r,u) \end{pmatrix} dr + \sum_{j=1}^{m_1}\int_0^t \begin{pmatrix} \sigma^j(r,\xi^u_r,u) \\ 0 \end{pmatrix} dw^j_r$$
$$+ \sum_{j=1}^{m_2}\int_0^t \begin{pmatrix} 0 \\ \hat{\sigma}^j(r,x^u_r,\xi^u_r,u) \end{pmatrix} dB^j_r.$$

We introduce the discretized calibration problem and show that the discretized cost function converges to the cost function and the same holds for the corresponding gradients. Let $\Pi^{\mathrm{E}} = (t_i)_{i=0,\ldots,n}$ be partition of the interval $[0,T]$ such that $(T_\mu)_{\mu=1,\ldots,M} \subset \Pi^{\mathrm{E}} \setminus \{0\}$, with the notations of Section 4.1, the discretized calibration problem is given by

$$(\mathrm{P}^n) \text{ Find} \qquad \min_{u\in\mathcal{U}} J^n(u) = \min_{u\in\mathcal{U}} \frac{1}{2}\sum_{\mu=1}^{M} \mathrm{E}\left[g_\mu(\mathcal{X}^{n,u}_{T_\mu})\right]^2 \tag{4.78}$$

subject to

$$\mathcal{X}^n_{t_{i+1}} = \begin{pmatrix} \xi^n_{t_{i+1}} \\ x^n_{t_{i+1}} \end{pmatrix} = \begin{pmatrix} \xi^n_{t_i} \\ x^n_{t_i} \end{pmatrix} + \begin{pmatrix} b(t_i,\xi^n_{t_i},u) \\ \hat{b}(t_i,x^n_{t_i},\xi^n_{t_i},u) \end{pmatrix} (t_{i+1}-t_i) + \sum_{j=1}^{m_1} \begin{pmatrix} \sigma^j(t_i,\xi^n_{t_i},u) \\ 0 \end{pmatrix} (w^j_{t_{i+1}} - w^j_{t_i})$$
$$+ \sum_{j=1}^{m_2} \begin{pmatrix} 0 \\ \hat{\sigma}^j(t_i,x^n_{t_i},\xi^n_{t_i},u) \end{pmatrix} (B^j_{t_{i+1}} - B^j_{t_i}), \qquad i = 0,\ldots,n-1, \mathcal{X}^n_{t_0} = \mathcal{X}_0(u)$$

In the following corollary, we utilize the results from the previous section to show that J^n approximates J.

Corollary 4.21. *There exists a constant $D_{K_J} > 0$, such that for every $u \in \mathcal{U}$, we have that*

$$|J(u) - J^n(u)| \leq D_{K_J} \delta_{1,4}^{(2-p)\wedge\frac{1}{2}}.$$

Under the assumption (HA), there exists a constant $\tilde{D}_{K_J} > 0$ such that

$$|J(u) - J^n(u)| \leq \tilde{D}_{K_J} \delta_{2}^{(2H-1)\wedge\frac{1}{2}}.$$

Proof. Let $u \in \mathcal{U}$, since $(g_\mu)_{\mu=1,\dots,M}$ satisfies condition (G), we get

$$\begin{aligned}
&|J(u) - J^n(u)| \\
&= \frac{1}{2}\sum_{\mu=1}^{M} \left| \mathrm{E}\left[g_\mu(\mathcal{X}_{T_\mu}^u)\right]^2 - \mathrm{E}\left[g_\mu(\mathcal{X}_{T_\mu}^{n,u})\right]^2 \right| \\
&\leq \frac{1}{2}\sum_{\mu=1}^{M} \mathrm{E}\left[\left|g_\mu(\mathcal{X}_{T_\mu}^u) - g_\mu(\mathcal{X}_{T_\mu}^{n,u})\right|\right] \mathrm{E}\left[\left|g_\mu(\mathcal{X}_{T_\mu}^u) + g_\mu(\mathcal{X}_{T_\mu}^{n,u})\right|\right] \\
&\leq \frac{1}{2}ML\left(L\mathrm{E}\left[\|\mathcal{X}^u\|_{\infty,0,T}\right] + L\mathrm{E}\left[\|\mathcal{X}^{n,u}\|_{\infty,0,T}\right] + 2\max_{\mu=1,\dots,M}|g_\mu(0)|\right) \mathrm{E}\left[\|\mathcal{X}^u - \mathcal{X}^{n,u}\|_{\infty,0,T}\right].
\end{aligned}$$

By Theorem 4.7, Remark 2.47 and the monotonicity of L^l-norms, we get

$$\begin{aligned}
|J(u) - J^n(u)| &\leq \frac{1}{2}ML\left(LD_{\mathcal{X},1} + LD_{\mathcal{X}^n,1} + 2\max_{\mu=1,\dots,M}|g_\mu(0)|\right) D_{K_{\mathcal{X}},2}\delta_{1,4}^{(2-p)\wedge\frac{1}{2}} \\
&\leq D_{K_J}\delta_{1,4}^{(2-p)\wedge\frac{1}{2}}.
\end{aligned}$$

Under the assumption (HA), we have by (4.61) and the monotonicity of L^l-norms

$$E\left[\|\mathcal{X}^u - \mathcal{X}^{n,u}\|_{\infty,0,T}\right] \leq \hat{D}_{K_{\mathcal{X}},2}\delta_2^{(2H-1)\wedge\frac{1}{2}}.$$

Hence, we get for every $u \in \mathcal{U}$

$$|J(u) - J^n(u)| \leq \hat{D}_{K_J}\delta_2^{(2H-1)\wedge\frac{1}{2}}.$$

$\square$

For the gradient ∇J, we have two representations, one using the sensitivity equation (3.14) and the one established in Lemma 3.17, using the solution to the adjoint equation (3.16). In Subsection 4.1.2, we showed that we can approximate the sensitivity equation (4.4) by its corresponding discretization scheme (4.6). Now we will use this to show that we can approximate ∇J with the help of the Euler approximations of $\mathcal{X}$ and $\mathcal{Y}$.

Lemma 4.22. *For every $u \in \mathcal{U}$, we define the discretized gradient*

$$(\nabla J)^n(u) := \sum_{\mu=1}^{M} \mathrm{E}\left[g_\mu(\mathcal{X}^{n,u}_{T_\mu})\right] \mathrm{E}\left[g'_\mu(\mathcal{X}^{n,u}_{T_\mu})\mathcal{Y}^{n,u}_{T_\mu}\right].$$

Then there exist positive constants $D_{K_{\nabla J}}$ and $\tilde{D}_{K_{\nabla J}}$, independent of u and n, such that

$$|\nabla J(u) - (\nabla J)^n(u)| \leq D_{K_{\nabla J}} \delta_{1,8}^{(2-p)\wedge\frac{1}{2}}.$$

and under the assumption (HA), we get

$$|\nabla J(u) - (\nabla J)^n(u)| \leq \tilde{D}_{K_{\nabla J}} \delta_2^{(2H-1)\wedge\frac{1}{2}}.$$

Proof. By definition of $\nabla J(u)$ and $(\nabla J)^n(u)$, we have

$$\begin{aligned}
|\nabla J(u) - (\nabla J)^n(u)| &\leq \sum_{\mu=1}^{M} \mathrm{E}\left[\left|g_\mu(\mathcal{X}^u_{T_\mu}) - g_\mu(\mathcal{X}^{n,u}_{T_\mu})\right|\right] \mathrm{E}\left[\left|g'_\mu(\mathcal{X}^u_{T_\mu})\mathcal{Y}^u_{T_\mu}\right|\right] \\
&\quad + \sum_{\mu=1}^{M} \mathrm{E}\left[\left|g_\mu(\mathcal{X}^{n,u}_{T_\mu})\right|\right] \mathrm{E}\left[\left|g'_\mu(\mathcal{X}^u_{T_\mu})\mathcal{Y}^u_{T_\mu} - g'_\mu(\mathcal{X}^{n,u}_{T_\mu})\mathcal{Y}^{n,u}_{T_\mu}\right|\right] \\
&= S_1 + S_2.
\end{aligned}$$

By the Lipschitz continuity of g, the boundedness of g', Remark 2.47, Theorem 4.7 and the monotonicity of L^l-norms, we can estimate the first sum by

$$\begin{aligned}
S_1 &\leq ML^2 D_{\mathcal{Y},1} \mathrm{E}\left[\|\mathcal{X}^u - \mathcal{X}^{n,u}\|_{\infty,0,T}\right] \\
&\leq ML^2 D_{\mathcal{Y},1} D_{K_{\mathcal{X}},2} \delta_{1,4}^{(2-p)\wedge\frac{1}{2}}.
\end{aligned}$$

The sum S_2 can be decomposed by

$$S_2 \leq \sum_{\mu=1}^{M} \mathrm{E}\left[\left|g_\mu(\mathcal{X}^{n,u}_{T_\mu})\right|\right] \mathrm{E}\left[\left|g'_\mu(\mathcal{X}^u_{T_\mu}) - g'_\mu(\mathcal{X}^{n,u}_{T_\mu})\right| \left|\mathcal{Y}^u_{T_\mu}\right| + \left|g'_\mu(\mathcal{X}^{n,u}_{T_\mu})\right| \left|\mathcal{Y}^u_{T_\mu} - \mathcal{Y}^{n,u}_{T_\mu}\right|\right].$$

Using (3.2), the Lipschitz continuity of g', the boundedness of g' and (4.31) this yields

$$S_2 \leq M\left(LD_{\mathcal{X}^n,1} + \max_{\mu=1,\dots,M} g_\mu(0)\right) L\mathrm{E}\left[\|\mathcal{X}^u - \mathcal{X}^{n,u}\|_{\infty,0,T} \|\mathcal{Y}^u\|_{\infty,0,T} + \|\mathcal{Y}^u - \mathcal{Y}^{n,u}\|_{\infty,0,T}\right].$$

By the Hölder inequality, Theorem 4.7, Theorem 4.13 and again the monotonicity of L^l-norms, this yields

$$S_2 \le M \left(LD_{\mathcal{X}^n,1} + \max_{\mu=1,\dots,M} g_\mu(0) \right) L \bigg(\mathrm{E}\left[\|\mathcal{X}^u - \mathcal{X}^{n,u}\|^2_{\infty,0,T} \right]^{\frac{1}{2}} \mathrm{E}\left[\|\mathcal{Y}^u\|^2_{\infty,0,T} \right]^{\frac{1}{2}}$$

$$+ \mathrm{E}\left[\|\mathcal{Y}^u - \mathcal{Y}^{n,u}\|_{\infty,0,T} \right] \bigg)$$

$$\le ML \left(LD_{\mathcal{X}^n,1} + \max_{\mu=1,\dots,M} g_\mu(0) \right) \left(D_{K_\mathcal{X},2} \delta_{1,4}^{(2-p)\wedge\frac{1}{2}} D_{\mathcal{Y},2}^{\frac{1}{2}} + D_{K_\mathcal{Y},2} \delta_{1,8}^{(2-p)\wedge\frac{1}{2}} \right).$$

Hence, there exists a constant D_{K_J}, independent of u and n, such that

$$|\nabla J(u) - (\nabla J)^n(u)| \le D_{K_J} \delta_{1,8}^{(2-p)\wedge\frac{1}{2}},$$

since $\delta_{1,4} \le \delta_{1,8}$. Using the arguments from Subsection 4.1.3, the convergence rate under the assumption (HA) follows. □

We want to include the discretized adjoint equation given by (4.64) into the calculation of the gradient ∇J. But instead of discretizing the gradient given in (3.15), we find another representation of $(\nabla J)^n$, which contains Λ^n.

Lemma 4.23. *For every $u \in \mathcal{U}$, the discretized gradient $(\nabla J)^n(u)$, can be represented by*

$$(\nabla J)^n(u) = \mathrm{E}\left[\Lambda_0^n D\mathcal{X}_0^u + \sum_{i=0}^{n-1} \Lambda_{t_{i+1}}^n \eta_{t_i,t_{i+1}}^u \right],$$

where

$$\eta_{t_i,t_{i+1}}^u := \begin{pmatrix} b_u(t_i, \xi_{t_i}^n, u) \\ \hat{b}_u(t_i, x_{t_i}^n, \xi_{t_i}^n, u) \end{pmatrix} (t_{i+1} - t_i) + \sum_{j=1}^{m_1} \begin{pmatrix} \sigma_u^j(t_i, \xi_{t_i}^n, u) \\ 0 \end{pmatrix} (w_{t_{i+1}}^j - w_{t_i}^j)$$
$$+ \sum_{j=1}^{m_2} \begin{pmatrix} 0 \\ \hat{\sigma}_u^j(t_i, x_{t_i}^n, \xi_{t_i}^n, u) \end{pmatrix} (B_{t_{i+1}}^j - B_{t_i}^j)$$

for all $i \in \{0, \dots, n-1\}$. Here the discrete adjoint equation is given by

$$\Lambda_{t_i}^n = (\lambda_{t_i}^n, \hat{\lambda}_{t_i}^n) = \Lambda_{t_{i+1}}^n \left(I_{n_1+n_2} + \eta_{t_i,t_{i+1}} \right) + \sum_{T_\mu = t_i} E[g_\mu(\mathcal{X}_{T_\mu}^n)] g'_\mu(\mathcal{X}_{T_\mu}^n) \in \mathbb{R}^{n_1+n_2},$$

where

$$\eta_{t_i,t_{i+1}} = \begin{pmatrix} b_x(t_i, \xi_{t_i}^n, u) & 0 \\ \hat{b}_z(t_i, x_{t_i}^n, \xi_{t_i}^n, u) & \hat{b}_x(t_i, x_{t_i}^n, \xi_{t_i}^n, u) \end{pmatrix} (t_{i+1} - t_i) + \sum_{j=1}^{m_1} \begin{pmatrix} \sigma_x^j(t_i, \xi_{t_i}^n, u) & 0 \\ 0 & 0 \end{pmatrix} (w_{t_{i+1}}^j - w_{t_i}^j)$$
$$+ \sum_{j=1}^{m_2} \begin{pmatrix} 0 & 0 \\ \hat{\sigma}_z^j(t_i, x_{t_i}^n, \xi_{t_i}^n, u) & \hat{\sigma}_x^j(t_i, x_{t_i}^n, \xi_{t_i}^n, u) \end{pmatrix} (B_{t_{i+1}}^j - B_{t_i}^j) \in \mathbb{R}^{n_1+n_2}$$

for all $i \in \{0, \ldots, n-1\}$ *and*

$$\Lambda_T^n = \sum_{T_\mu = T} \mathrm{E}[g_\mu(\mathcal{X}_T^n)] g'_\mu(\mathcal{X}_T^n),$$

analogously to (4.64).

Proof. We use ideas similar to the proof of Theorem 5.1 in Käbe et al. [2009]. For $u \in \mathcal{U}$ we have the discretized gradient representation

$$(\nabla J)^n(u) := \sum_{\mu=1}^{M} \mathrm{E}\left[g_\mu(\mathcal{X}_{T_\mu}^{n,u})\right] \mathrm{E}\left[g'_\mu(\mathcal{X}_{T_\mu}^{n,u}) \mathcal{Y}_{T_\mu}^{n,u}\right],$$

where for all $i \in \{0, \ldots, n-1\}$, we have

$$\begin{aligned}
\mathcal{Y}_{t_{i+1}}^n &= \begin{pmatrix} y_{t_{i+1}}^n \\ \hat{y}_{t_{i+1}}^n \end{pmatrix} \\
&:= \mathcal{Y}_{t_i}^n + \left(\begin{pmatrix} b_x\left(t_i, \xi_{t_i}^n, u\right) & 0 \\ \hat{b}_z\left(t_i, x_{t_i}^n, \xi_{t_i}^n, u\right) & \hat{b}_x\left(t_i, x_{t_i}^n, \xi_{t_i}^n, u\right) \end{pmatrix} \mathcal{Y}_{t_i}^n + \begin{pmatrix} b_u\left(t_i, \xi_{t_i}^n, u\right) \\ \hat{b}_u\left(t_i, x_{t_i}^n, \xi_{t_i}^n, u\right) \end{pmatrix} \right) (t_{i+1} - t_i) \\
&\quad + \sum_{j=1}^{m_1} \left(\begin{pmatrix} \sigma_x^j\left(t_i, \xi_{t_i}^n, u\right) & 0 \\ 0 & 0 \end{pmatrix} \mathcal{Y}_{t_i}^n + \begin{pmatrix} \sigma_u^j(t_i, \xi_{t_i}^n, u) \\ 0 \end{pmatrix} \right) \left(w_{t_{i+1}}^j - w_{t_i}^j \right) \\
&\quad + \sum_{j=1}^{m_2} \left(\begin{pmatrix} 0 & 0 \\ \hat{\sigma}_z^j\left(t_i, x_{t_i}^n, \xi_{t_i}^n, u\right) & \hat{\sigma}_x^j\left(t_i, x_{t_i}^n, \xi_{t_i}^n, u\right) \end{pmatrix} \mathcal{Y}_{t_i}^n + \begin{pmatrix} 0 \\ \hat{\sigma}_u^j\left(t_i, x_{t_i}^n, \xi_{t_i}^n, u\right) \end{pmatrix} \right) \left(B_{t_{i+1}}^j - B_{t_i}^j \right) \\
&= \mathcal{Y}_{t_i}^n + \left(\mathcal{B}_x^n(t_i) \mathcal{Y}_{t_i}^n + \mathcal{B}_u^n(t_i)\right)(t_{i+1} - t_i) + \sum_{j=1}^{m_1} \left(\Sigma_x^{n,j}(t_i) \mathcal{Y}_{t_i}^n + \Sigma_u(t_i)^{n,j}\right) \left(w_{t_{i+1}}^j - w_{t_i}^j \right) \\
&\quad + \sum_{j=1}^{m_2} \left(\hat{\Sigma}_x^{n,j}(t_i) \mathcal{Y}_{t_i}^n + \hat{\Sigma}_u^{n,j}(t_i) \right) \left(B_{t_{i+1}}^j - B_{t_i}^j \right) \\
&= (I_{n_1+n_2} + \eta_{t_i,t_{i+1}}) \mathcal{Y}_{t_i}^n + \eta_{t_i,t_{i+1}}^u
\end{aligned}$$

with

$$\mathcal{Y}_0^n = \mathcal{Y}_0 = (D\xi_0(u), Dx_0(u))^\top.$$

We consider the sum

$$\sum_{i=0}^{n-1} \mathcal{Y}_{t_{i+1}}^n$$

and multiply each of the recursive equations with row vectors $\Lambda_{t_{i+1}}^n \in \mathbb{R}^{n_1+n_2}$. Furthermore we add the term $\Lambda_0^n \mathcal{Y}_0^n$ on both sides, which yields

$$\sum_{i=0}^{n-1} \Lambda_{t_{i+1}}^n \mathcal{Y}_{t_{i+1}}^n + \Lambda_0^n \mathcal{Y}_0^n = \sum_{i=0}^{n-1} \Lambda_{t_{i+1}}^n \left((I_{n_1+n_2} + \eta_{t_i,t_{i+1}}) \mathcal{Y}_{t_i}^n + \eta_{t_i,t_{i+1}}^u \right) + \Lambda_0^n \mathcal{Y}_0^n.$$

This is equivalent to

$$\sum_{i=0}^{n-1}\left(\Lambda_{t_i}^n - \Lambda_{t_{i+1}}^n\left(I_{n_1+n_2}+\eta_{t_i,t_{i+1}}\right)\right)\mathcal{Y}_{t_i}^n + \Lambda_{t_n}^n\mathcal{Y}_{t_n}^n = \Lambda_0^n\mathcal{Y}_0^n + \sum_{i=0}^{n-1}\Lambda_{t_{i+1}}^n\eta_{t_i,t_{i+1}}^u.$$

If we now choose $\Lambda_{t_i}^n$ for $i=0,\dots,n$ according to the statement of the Lemma, this yields

$$\begin{aligned}&\sum_{i=0}^{n-1}\sum_{T_\mu=t_i}E[g_\mu(\mathcal{X}_{T_\mu}^n)]g_\mu'(\mathcal{X}_{T_\mu}^n)\mathcal{Y}_{T_\mu}^n + \sum_{T_\mu=T}E[g_\mu(\mathcal{X}_T^n)]g_\mu'(\mathcal{X}_T^n)\mathcal{Y}_{T_M}^n\\ &= \Lambda_0^n(D\xi_0(u), Dx_0(u))^\top + \sum_{i=0}^{n-1}\Lambda_{t_{i+1}}^n\eta_{t_i,t_{i+1}}^u,\end{aligned}$$

which is equivalent to

$$\sum_{\mu=1}^{M}E[g_\mu(\mathcal{X}_{T_\mu}^n)]g_\mu'(\mathcal{X}_{T_\mu}^n)\mathcal{Y}_{T_\mu}^n = \Lambda_0^n D\mathcal{X}_0(u) + \sum_{i=0}^{n-1}\Lambda_{t_{i+1}}^n\eta_{t_i,t_{i+1}}^u.$$

Taking the expected value on both sides, yields

$$(\nabla J)^n = \mathrm{E}\left[\Lambda_0^n D\mathcal{X}_0 + \sum_{i=0}^{n-1}\Lambda_{t_{i+1}}^n\eta_{t_i,t_{i+1}}^u\right]$$

and hence, the assertion. □

We proved that we can approximate the gradient of our cost function by discretizing either the sensitivity equation or the adjoint equation. How these results can be used in practice is shown in Section 1.5.

Bibliography

A. Ambrosetti and G. Prodi. *A Primer of Nonlinear Analysis*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1995.

L.B.G. Andersen and R. Brotherton-Ratcliffe. The equity option volatility smile: an implicit finite-difference approach. *Journal of Computational Finance*, 1(2):5–37, 1998.

H. Araya, J.A. León, and S. Torres. Numerical Scheme for Stochastic Differential Equations Driven by Fractional Brownian Motion with 1/4< H<1/2. *Journal of Theoretical Probability*, 33(3):1211–1237, September 2020.

C. Bayer, P. Friz, and J. Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6): 887–904, 2016.

V. Bezborodov, L. Di Persio, and Y. Mishura. Option pricing with fractional stochastic volatility and discontinuous payoff function of polynomial growth. *Method. Comput. Appl. Prob.*, 21(1): 331–366, mar 2019.

J.-M. Bismut. An introductory approach to duality in optimal stochastic control. *SIAM Review*, 20(1):62–78, 1978.

R.M. Blumenthal and R. Getoor. Some theorems on stable processes. *Transactions of the American Mathematical Society*, 95:263–273, 1960.

V.G. Boltyanski, R.V. Gamkrelidze, E.F. Mishchenko, and L.S. Pontryagin. The maximum principle in the theory of optimal processes of control. *IFAC Proceedings Volumes*, 1(1):464–469, 1960.

L. Capriotti. Fast greeks by algorithmic differentiation. *The Journal of Computational Finance*, 14:3–35, 03 2011.

T. Cass, C. Litterer, and T. Lyons. Integrability and tail estimates for Gaussian rough differential equations. *The Annals of Probability*, 41(4):3026–3050, 2013.

I. Charpentier and M. Ghemires. Efficient adjoint derivatives: application to the meteorological model meso-nh. *Optimization Methods and Software*, 13(1):35–63, 2000.

T.Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud. Neural ordinary differential equations. 2018. URL `http://arxiv.org/abs/1806.07366`.

M. Chesney and L. Scott. Pricing european currency options: A comparison of the modified black- scholes model and a random variance model. *The Journal of Financial and Quantitative Analysis*, 24(3):267–284, 1989.

A. Chronopoulou and S. Tindel. On inference for fractional differential equations. *Statistical Inference for Stochastic Processes*, 16(1):29–61, April 2013.

A. Chronopoulou and F. Viens. Estimation and pricing under long-memory stochastic volatility. *Annals of Finance*, 8(2):379–403, 2012.

F. Comte and E. Renault. Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4):291–323, 1998.

L. Coutin and Z. Qian. Stochastic analysis, rough path analysis and fractional brownian motions. *Probability Theory and Related Fields*, 122:108–140, 2002.

A.M. Davie. Differential Equations Driven by Rough Paths: An Approach via Discrete Approximation. *Applied Mathematics Research eXpress*, 2008, 01 2008.

A. Deya, A. Neuenkirch, and S. Tindel. A Milstein-type scheme without Lévy area terms for SDEs driven by fractional Brownian motion. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 48(2):518 – 550, 2012.

R.M. Dudley and R. Norvaiša. *Product integrals, young integrals and p-variation*, pages 73–208. Springer Berlin Heidelberg, Berlin, Heidelberg, 1999.

R.M. Dudley and R. Norvaiša. *Concrete Functional Calculus*. Springer Monographs in Mathematics. Springer New York, 2010.

O. El Euch and M. Rosenbaum. The characteristic function of rough heston models. *Mathematical Finance*, 29(1):3–38, 2019.

O. Forster. *Analysis 2: Differentialrechnung im IRn, gewöhnliche Differentialgleichungen*. Analysis / Otto Forster. Vieweg+Teubner Verlag, 2008.

K.R. French, G.W. Schwert, and R.F. Stambaugh. Expected stock returns and volatility. *Journal of Financial Economics*, 19(1):3–29, 1987.

P. Friz and N. Victoir. *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*, volume 120 of *Cambridge studies in advanced mathematics*. Cambridge University Press, 2010.

M. Fukasawa. Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance*, 17(2):189–198, 2017.

J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6): 933–949, 2018.

M. Giles and P. Glasserman. Smoking adjoints: fast evaluation of greeks in monte carlo calculations. *Risk*, 19:88–92, 01 2006.

M.B. Giles and N.A. Pierce. An introduction to the adjoint approach to design. *Flow, Turbulence and Combustion*, 65:393–415, 2000.

P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Stochastic Modelling and Applied Probability. Springer New York, 2013.

J. Guerra and D. Nualart. Stochastic differential equations driven by fractional brownian motion and standard brownian motion. *Stochastic Analysis and Applications*, 26:1053 – 1075, 2008.

J.K. Hale. *Ordinary Differential Equations*. Dover Books on Mathematics Series. Dover Publications, 2009.

Y. Han, Y. Hu, and J. Song. Maximum principle for general controlled systems driven by fractional brownian motions. *Applied Mathematics & Optimization*, 67:279–322, 2012.

M. Henrard. Calibration in finance: Very fast greeks through algorithmic differentiation and implicit function. *Procedia Computer Science*, 18:1145–1154, 2013.

S. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343, 1993.

Y. Hu. Integral transformations and anticipative calculus for fractional brownian motions / yaozhong hu. *SERBIULA (sistema Librum 2.0)*, 175, 05 2005.

Y. Hu, Y. Liu, and D. Nualart. Rate of convergence ansasymptotic error distribution of euler approximation schemes for fractional diffusions. *The Annals of Applied Probability*, 26(2): 1147–1207, 2016.

J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2):281–300, 1987.

N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. ISSN. Elsevier Science, 2014.

N.C. Jain and D. Monrad. Gaussian Measures in B_p. *The Annals of Probability*, 11(1):46 – 57, 1983.

Nahid Jamshidi and Minoo Kamrani. Convergence of a numerical scheme associated to stochastic differential equations with fractional brownian motion. *Applied Numerical Mathematics*, 167: 108–118, 2021.

C. Käbe. *Feasibility and Efficiency of Monte Carlo based Calibration of Financial Market Models*. doctoralthesis, Universität Trier, 2010.

C. Käbe, J. Maruhn, and E. Sachs. Adjoint-based monte carlo calibration of financial market models. *Finance and Stochastics*, 13(3):351–379, September 2009.

O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics (113) (Book 113). Springer New York, 1991.

P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2011.

K. Kubilius. The existence and uniqueness of the solution of an integral equation driven by a p-semimartingale of special type. *Stochastic Processes and their Applications*, 98(2):289–315, 2002.

J. P. Lafore, J. Stein, N. Asencio, P. Bougeault, Vincent Ducrocq, J. Duron, C. Fischer, P. Héreil, P. Mascart, V. Masson, J.-P. Pinty, J.L. Redelsperger, E. Richard, and J. Vilà-Guerau de Arellano. The Meso-NH Atmospheric Simulation System. Part I: adiabatic formulation and control simulations. *Annales Geophysicae*, 16(1):90–109, 1998.

A. Lejay. Controlled differential equations as young integrals: A simple approach. *Journal of Differential Equations*, 249(8):1777–1798, 2010.

E. Lépinette and F. Mehrdoust. A fractional version of the heston model with hurst parameter $h \in (1/2, 1)$. *Econometrics: Multiple Equation Models eJournal*, 2016.

X. Li, T.-K.L. Wong, R.T.Q. Chen, and D.K. Duvenaud. Scalable gradients and variational inference for stochastic differential equations. In *Second Symposium on Advances in Approximate Bayesian Inference*, 2019. URL `https://openreview.net/forum?id=HyeL9yh4KH`.

S.J. Lin. Stochastic analysis of fractional brownian motions. *Stochastics and Stochastic Reports*, 55(1-2):121–140, 1995.

W. Liu and J. Luo. Modified euler approximation of stochastic differential equation driven by brownian motion and fractional brownian motion. *Communications in Statistics - Theory and Methods*, 46(15):7427–7443, 2017.

W. Liu, Y. Jiang, and Z. Li. Rate of convergence of euler approximation of time-dependent mixed sdes driven by brownian motions and fractional brownian motions. *AIMS Mathematics*, 5(3): 2163–2195, 2020.

T. Lyons. Differential equations driven by rough signals (i): An extension of an inequality of l. c. young. *Mathematical Research Letters*, 1:451–464, 1994.

T. Lyons. Differential equations driven by rough signals. *Revista Matemática Iberoamericana*, 14 (2):215–310, 1998.

B.B. Mandelbrot and J.W. Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM Review*, 10(4):422–437, 1968.

S. Massaroli, M. Poli, S. Peluchetti, J. Park, A. Yamashita, and H. Asama. Learning stochastic optimal policies via gradient descent. *IEEE Control Systems Letters*, 06 2021.

F. Mehrdoust and S. Fallah. On the calibration of fractional two-factor stochastic volatility model with non-lipschitz diffusions. *Communications in Statistics - Simulation and Computation*, 51 (11):6332–6351, 2022.

Y. Mishura. *Stochastic Calculus for Fractional Brownian Motion and Related Processes*. Number Nr. 1929 in Lecture Notes in Mathematics. Springer, 2008.

Y. Mishura and S.V. Posashkova. Stochastic differential equations driven by a wiener process and fractional brownian motion: Convergence in besov space with respect to a parameter. *Computers & Mathematics with Applications*, 62(3):1166–1180, 2011.

Y. Mishura and G. Shevchenko. The rate of convergence for euler approximations of solutions of stochastic differential equations driven by fractional brownian motion. *Stochastics*, 80(5): 489–511, 2008.

Y. Mishura and G. Shevchenko. Rate of convergence of euler approximations of solution to mixed stochastic differential equation involving brownian motion and fractional brownian motion. *Random Operators and Stochastic Equations*, 19, 11 2011a.

Y. Mishura and G.M. Shevchenko. Existence and uniqueness of the solution of stochastic differential equation involving wiener process and fractional brownian motion with hurst index h > 1/2. *Communications in Statistics - Theory and Methods*, 40(19-20):3492–3508, 2011b.

Y. Mishura and G.M. Shevchenko. Mixed stochastic differential equations with long-range dependence: Existence, uniqueness and convergence of solutions. *Computers & Mathematics with Applications*, 64(10):3217–3227, 2012.

Y. Mishura and A. Yurchenko-Tytarenko. Approximating expected value of an option with non-lipschitz payoff in fractional heston-type model, 2020.

M. Mrázek, J. Pospíšil, and T. Sobotka. On calibration of stochastic and fractional stochastic volatility models. *European Journal of Operational Research*, 254(3):1036–1046, 2016.

A. Neuenkirch. Optimal approximation of sde's with additive fractional noise. *Journal of Complexity*, 22(4):459–474, 2006.

D.C. Nguyen, H.D. Luu, and T.H. Phan. Nonautonomous young differential equations revisited. *Journal of Dynamics and Differential Equations*, 30, 12 2018.

D.C. Nguyen, H.D. Luu, and T.H. Phan. Lyapunov spectrum of nonautonomous linear young differential equations. *Journal of Dynamics and Differential Equations*, 32:1749 – 1777, 2020.

I. Nourdin. Schémas d'approximation associés à une équation différentielle dirigée par une fonction höldérienne ; cas du mouvement brownien fractionnaire. *Comptes Rendus Mathematique*, 340 (8):611–614, 2005.

I. Nourdin and A. Neuenkirch. Exact rate of convergence of some approximation schemes associated to sdes driven by a fractional brownian motion. *Journal of Theoretical Probability*, 20(4): 871–899, 2007.

D. Nualart. *The Malliavin Calculus and Related Topics*. Probability and its applications : a series of the applied probability trust. Springer-Verlag, 1995.

D. Nualart and A. Răşcanu. Differential equations driven by fractional brownian motion. *Collectanea Mathematica*, 53:55–81, 2002.

B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications (Universitext)*. Springer, 6th edition, 2014.

H. Pham. *Continuous-time Stochastic Control and Optimization with Financial Applications*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2009.

P.E. Protter. *Stochastic Integration and Differential Equations*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2005.

L.C.G. Rogers. Arbitrage with fractional brownian motion. *Mathematical Finance*, 7(1):95–105, 1997.

F. Russo and P. Vallois. Forward, backward and symmetric stochastic integration. *Probability Theory and Related Fields*, 97:403–421, 1993a.

F. Russo and P. Vallois. Non-causal stochastic integration for lad lag processes. *Stochastic Analysis and Related Topics (Oslo, 1992), Stochastics Monogr.*, 8, 01 1993b.

F. Russo and P. Vallois. The generalized covariation process and ito formula. *Stochastic Processes and their Applications*, 59(1):81–104, 1995.

F. Russo and P. Vallois. Ito formula for c 1 -functions of semimartingales. *Probability Theory and Related Fields*, 104:27–41, 01 1996.

F. Russo and P. Vallois. Stochastic calculus with respect to continuous finite quadratic variation processes. *Stochastics An International Journal of Probability and Stochastic Processes*, 70: 1–40, 07 2000.

F. Russo and P. Vallois. *Elements of Stochastic Calculus via Regularization*, pages 147–185. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

A.A. Ruzmaikina. Stieltjes integrals of hölder continuous functions with applications to fractional brownian motion. *Journal of Statistical Physics*, 100:1049–1069, 2000.

G. Shevchenko. Mixed stochastic delay differential equations. *Theory of Probability and Mathematical Statistics*, 89, 06 2013.

E.M. Stein and J.C. Stein. Stock Price Distributions with Stochastic Volatility: An Analytic Approach. *The Review of Financial Studies*, 4(4):727–752, 05 2015.

J. Yong. Stochastic optimal control — a concise introduction. *Mathematical Control & Related Fields*, 01 2019.

J. Yong and X.Y. Zhou. *Stochastic Controls: Hamiltonian Systems and HJB Equations*. Stochastic Modelling and Applied Probability. Springer New York, 1999.

L.C. Young. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Mathematica*, 67(none):251 – 282, 1936.

M. Zähle. Integration with respect to fractal functions and stochastic calculus. i. *Probability Theory and Related Fields*, 111:333–374, 1998.

J. Zhuang, N. Dvornek, X. Li, S. Tatikonda, X. Papademetris, and J. Duncan. Adaptive checkpoint adjoint method for gradient estimation in neural ODE. 119:11639–11649, 13–18 Jul 2020.