CHALMERS | GÖTEBORG UNIVERSITY

MASTER'S THESIS

On Weak Differentiability of Backward SDEs and Cross Hedging of Insurance Derivatives

ADAM ANDERSSON

Department of Mathematical Statistics
CHALMERS UNIVERSITY OF TECHNOLOGY
GÖTEBORG UNIVERSITY
Göteborg, Sweden 2008

Thesis for the Degree of Master of Science (30 credits)

On Weak Differentiability of Backward SDEs and Cross Hedging of Insurance Derivatives

Adam Andersson

CHALMERS | GÖTEBORG UNIVERSITY

Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE – 412 96 Göteborg, Sweden
Göteborg, September 2008

Abstract

This thesis deals with distributional differentiability of the solution (X,Y,Z) to a quadratic non-degenerate forward-backward SDE. The differentiability is considered with respect to the initial value of the solution X to the coupled forward SDE. It is proved that the solution process Y is weakly differentiable, and that the solution process Z can be represented using the distributional gradient of Y. This result is new in the way that it relaxes technical conditions imposed by previous authors in a significant way and in a way that is important e.g., in the applications described below. The proof makes use of Dirichlet space techniques to conclude that Y is a member of a local Sobolev space.

Our results are applied to derive new results in mathematical finance and insurance theory. When derivatives are written on non-tradeable underlying assets, such as weather, a strongly correlated tradeable asset price process is used instead of the non-tradeable one to partially hedge the risk of the derivative. This concept is known as cross hedging. Applications for non-differentiable European type pay off functions are given and explicit hedging strategies are derived using a distributional gradient.

Keywords: Backward stochastic differential equation; Distributional differentiability; Dirichlet space; Cross hedging; Insurance derivative; Weather derivative; Explicit hedging strategy.

Acknowledgements

I would very much like to thank my supervisor professor Patrik Albin at Chalmers for good advices and encouragement during the work as well as for his excellent courses in stochastic calculus and stochastic processes. I would also like to thank professor Boualem Djehiche at KTH for first giving me the suggestion to work with cross hedging. The subject was perfect for me. Most of all I would like to thank professor Peter Imkeller, doctor Stefan Ankirchner and PhD student Gonçalo Dos Reis at Humboldt University in Berlin for inviting me for a two days discussion on cross hedging and BSDEs in Berlin and especially to Stefan Ankirchner and Gonçalo Dos Reis for giving me fruitful feedback on my work via mail. I very much appreciate the way mathematics is done in Berlin. Thank you also PhD student Mattias Sundén and professor Stig Larsson for answering my questions in a clear and precise way. Finally, I would like to thank my wife Fereshteh and my daughter Alice for their support and patience.

Contents

Chapter 1

Introduction

Imagine that you are the owner of a Swedish ice cream factory. Then you are exposed to weather risk. A warm and sunny summer, people will spend their time on the beach, with an ice cream in their hand. A cold or rainy summer, they will complain over the Swedish weather or perhaps travel to a warmer place. They will anyway not eat ice cream to a great extent. It is clear that your economic gain, from selling ice cream, depends on the weather. So, how can you protect yourself against such risks? Insurance are often for material or economical losses, not for default income, due to bad weather. One way could be to write a financial contract, like an option, on some weather index. Accumulated average temperatures or sun hours during a summer could be suitable indices. Then the risk will be spread. In case of unfavorable summer weather, you get money according to the contract. Under favorable weather you get nothing. In either case you pay the premium for the contract. You expose yourself of a lower risk.

Say, that you choose to buy, or get short in, a European call option with strike price K, written on some (artificial) sunshine index. The index will have value X_T at time T of maturity. The random income at time T is given by

$$(X_T - K)^+ = \max(0, X_T - K) := F(X_T).$$

How shall the contract be priced? Now, we must turn the perspective to the seller of the derivative, having the long position. Her obligation to pay $F(X_T)$ at time T, implies a risk. If the underlying X were a tradable asset she would hedge the risk of the derivative. This would be done by investing in the underlying asset according to an optimal hedging strategy. The hedged risk would be considered when she sets the premium, by the machinery of Black-Scholes pricing theory.

It would be nice if sunshine was tradable, but it isn't. She can not buy herself a portfolio of sunshine. Pricing without hedging would imply a greater risk, and she would not last long unless she gets a high premium. A way to get around this, for her, could be to invest in a tradable asset, correlated to sunshine. In such a way the risk could partially be hedged by investing in the correlated asset. Assets possibly correlated to sunshine are for example heating oil futures or electricity futures. The concept is known as cross hedging, and the kind of derivatives, written on non-tradable underlyings, are called insurance derivatives.

Actors more likely to buy insurance derivatives are energy companies, sensitive to cold summers and warm winters. Chicago Mercantile Exchange was the first in 1997 to offer derivatives written on accumulated heating degree days (cHDD) and cooling degree days (cCDD). A heating degree day (HDD) and (CDD) is given by

$$\text{HDD} = \max(0, 18 - T) \qquad \text{and} \qquad \text{CDD} = \max(0, T - 18),$$

respectively, where T is the average temperature during a day. Statistics has shown that when the average temperature is 18 degrees Celsius the energy consumption is the lowest. When it is higher, energy is used for cooling and when it is lower, energy is used for heating. The cHDD index is given by

$$\text{cHDD} = \sum_{i=1}^{31} HDD_i$$

where HDD_i is the daily HDD's 31 days back in time. It can be seen as a moving average process. The cCDD index is defined analogously.

Historical weather data is today used to price cHDDs. The distribution of the outcome of the index is estimated and thereafter the distribution of the payoff. The derivative is priced by the expected payoff, discounted at the risk free rate [8]. This does not involve cross hedging or any hedging at all.

Hedging can be static. Then an investment is done at time zero, and no investment is done thereafter. Hedging can also be dynamic. Then the hedger invests according to a hedging strategy. The number of shares invested changes in time, as information is revealed, and new computations can be done. An approach to dynamic cross hedging and pricing is by setting up a stochastic control problem. An optimal strategy is sought that maximizes the expected utility of that investment. This can be done analytically by solving the Hamilton-Jacobi-Bellman partial differential equation [1] or by a stochastic approach using forward-backward stochastic differential equations (FBSDE) [2]. The former approach is limited to the case of one-dimensional assets. The latter approach can be applied with multiple-dimensions both for the tradable and the non-tradable assets. One drawback of the FBSDE approach is that the payoff function must be smooth. Hence, European put and call options can not be priced and hedged with that approach. In this thesis mathematical results are proved, that makes it possible to relax the smoothness property of the payoff function. With these results European put and call options can be priced and hedged, at least theoretically, when the non-tradable process satisfies a non-degeneracy condition. Suitable numerics must be used to implement this.

The mathematical results are about differentiability in the weak sense of FBSDEs. They rests on stochastic calculus, some distribution theory, measure theory and the theory of two specific Dirichlet spaces. Knowledge in stochastic calculus is assumed, of the reader, as well as some knowledge about measure theory and Lebesgue integration. Familiarity with Sobolev spaces and distribution theory makes the reading considerably easier.

The work is organized as follows:

Chapter 2 deals with BSDEs and FBSDEs. The first section introduces the subject. The second section contains more specific result needed in this thesis. The chapter ends with some history of the theory.

Chapter 3 first gives an introduction to weak derivatives and Sobolev spaces. Section two introduces Dirichlet spaces, essential for the proofs of Chapter 4. The chapter ends with a section about densities for stochastic differential equation and explains the concepts of non-degeneracy of SDEs.

Chapter 4 contains the mathematical results in this thesis. It starts with assumptions and a section with useful results. In Section 3 the main results are stated and proved.

Chapter 5 is about optimal cross hedging. The chapter presents the market model and solution approach of finding prices and cross hedging strategies of insurance derivatives. In the last section the results from Chapter 4 are applied to derive an explicit expression for the hedging strategies.

Chapter 6 contains a discussion and conclusion.

Chapter 2

Backward stochastic differential equations

Backward stochastic differentiable equations have proved to be useful in optimal stochastic control theory, mathematical finance and partial differential equations (PDEs). In finance the control processes usually are somehow related to investment strategies. The first section of this chapter introduces BSDEs and FBSDEs present the intimate connection between BSDEs and the martingale representation theorem. This connection is crucial for the understanding of BSDEs. Section 2 introduces BSDEs with random Lipschitz continuous generator. An existence and uniqueness result is also presented. All results presented in this chapter are well known. The last section presents some history of BSDEs.

2.1 Forward-backward stochastic differential equations

Let W_t be a d-dimensional Wiener process on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$. The filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ is the natural filtration of W_t completed by the $\mathbb{P}$-null sets of Ω. A forward backward stochastic differential equation (FBSDE) is a system of equations,

$$\begin{cases} dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \\ dY_t = -f(t, X_t, Z_t)dt + Z_t dW_t, \qquad t \in [0, T], \\ X_0 = x, \\ Y_T = g(X_T). \end{cases} \tag{2.1}$$

The coefficients $b : [0, T] \times \mathbb{R}^m \to \mathbb{R}^m$ and $\sigma : [0, T] \times \mathbb{R}^m \to \mathbb{R}^{m\times d}$ are supposed to be measurable and satisfy Lipschitz conditions and linear growth conditions in the space variable, i.e. $\exists C \geq 0$:

$$\begin{cases} |b(t, x) - b(t, \bar{x})| + |\sigma(t, x) - \sigma(t, \bar{x})| & \leq C|x - \bar{x}|, \\ |b(t, x)| + |\sigma(t, x)| & \leq C(1 + |x|) \end{cases} \tag{2.2}$$

$\forall(t, x, \bar{x}) \in [0, T]\times\mathbb{R}^m\times\mathbb{R}^m$. The equation for X is a forward Itô SDE. Notice that $\sigma(t, X_t)dW_t$ is a matrix multiplication. The norm $|\sigma|$ is the Frobenius norm

$$|\sigma| = \sqrt{\text{trace}(\sigma^*\sigma)}, \tag{2.3}$$

where σ^* is the transpose of σ. The equation for Y is known as a backward stochastic differential equation. The function $f : \Omega \times [0,T] \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}$ is called the *generator* of the FBSDE and $g : \mathbb{R}^m \to \mathbb{R}$ determines the terminal value. The equation for Y is also a forward SDE, but the "control" process Z controls it to satisfy the terminal value. It will be shown below that Y has a deterministic initial value. The process X is m-dimensional, the process Y is one-dimensional and the process Z is d-dimensional. The triple (X, Y, Z) is called the solution of FBSDE (2.1).

To be able to understand BSDEs recall the martingale representation theorem.

Theorem 2.1. *[20] **The martingale representation theorem.** Suppose that M_t is a square integrable martingale w.r.t. $\mathcal{F}_t$. Then there exist a unique, predictable, square integrable and d-dimensional process Z_t such that:*

$$M_t = \mathbb{E}[M_0] + \int_0^t Z_r dW_r$$

almost surely, for all $t \in [0, \infty)$.

The processes Y and Z are defined as follows. Define the martingale

$$M_t = \mathbb{E}\left[g(X_T) + \int_0^T f(r, X_r, Z_r)dr \middle| \mathcal{F}_t \right] \quad \text{for } t \in [0,T]. \tag{2.4}$$

Here, M_t is under suitable assumptions, presented later, square integrable. The martingale representation theorem therefore implies the existence of a unique predictable d-dimensional process Z_t such that:

$$M_t = M_0 + \int_0^t Z_r dW_r. \tag{2.5}$$

Let now,

$$Y_t = M_t - \int_0^t f(r, X_r, Z_r)dr.$$

It follows that:

$$\begin{cases} Y_t = M_0 - \displaystyle\int_0^t f(r, X_r, Z_r)dr + \int_0^t Z_r dW_r \\ M_0 = \mathbb{E}\left[g(X_T) + \displaystyle\int_0^T f(r, X_r, Z_r)dr \right] \end{cases} \tag{2.6}$$

as in (2.1). Equations (2.4) and (2.5) considered together is the third equation that determines (X, Y, Z). Notice that Y has a deterministic initial value M_0. Since $\int_0^t f(r, X_r, Z_r)dr$ is $\mathcal{F}_t$-measurable (2.6) and (2.4) gives that

$$Y_t = \mathbb{E}\left[g(X_T) + \int_t^T f(r, X_r, Z_r)dr \middle| \mathcal{F}_t \right]. \tag{2.7}$$

Further, $Y_T = M_0 - \int_0^T f(r, X_r, Z_r)dr + \int_0^T Z_r dW_r = g(X_T)$ from (2.6) and (2.7). Adding $0 = g(X_T) - M_0 + \int_0^T f(r, X_r, Z_r)dr - \int_0^T Z_r dW_r$ to (2.6) the most common form of a BSDE is obtained:

$$Y_t = g(X_T) + \int_t^T f(r, X_r, Z_r)dr - \int_t^T Z_r dW_r, \quad t \in [0, T]. \tag{2.8}$$

The process $\{X_t, Y_t, Z_t\}_{t\in[0,T]}$ is a Markov process. Conditioned on $X_t = x$, for $(t, x) \in [0, T] \times \mathbb{R}^m$, the FBSDE is given by:

$$\begin{cases} X_s^{t,x} = x + \int_t^s b(r, X_r^{t,x})dr + \int_t^s \sigma(r, X_r^{t,x})dW_r \\ Y_s^{t,x} = g(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x}, Z_r^{t,x})dr - \int_s^T Z_r^{t,x} dW_r, \quad s \in [t, T]. \end{cases} \tag{2.9}$$

Equation (2.9) will be denoted FBSDE(b, σ, g, f). The initial value is suppressed in the notation since it throughout this thesis will be an arbitrary vector $x\mathbb{R}^m$. Let ξ be a square integrable and $\mathcal{F}_T$-measurable random variable and $f : \Omega \times [0, T] \times \mathbb{R}^d \to \mathbb{R}$ be a generator function. The BSDE

$$Y_t = \xi + \int_t^T f(r, Z_r)dr - \int_t^T Z_r dW_r, \quad t \in [0, T]. \tag{2.10}$$

will be denoted BSDE(ξ, f). The process Z is determined analogously as for a FBSDE. When nothing is else is said the solutions to FBSDE(b, σ, g, f) and the solution to BSDE(ξ, f) will be denoted (X, Y, Z) and (Y, Z), respectively. A forward SDE with coefficients b and σ will be denoted SDE(b, σ).

2.2 BSDEs with random Lipschitz generators

Let $\{H_t\}_{t\in[0,T]}$ be an integrable, predictable and positive process. The process $\int_0^\cdot H_s dW_s$ is called a bounded mean oscillation (BMO) martingale if

$$\mathbb{E}\left[\int_\tau^T H_s^2 ds \middle| \mathcal{F}_\tau\right] \leq D \tag{2.11}$$

almost surely, for all stopping times $\tau \in [0, T]$ and some constant $D > 0$. The smallest $D > 0$ that satisfies (2.11) is called the BMO norm of $\int_0^\cdot H_s dW_s$ and will be denoted $\| \int_0^\cdot H_s dW_s\|_{\text{BMO}}$.

Let's introduce the following process spaces:

- $\mathcal{S}^p(\mathbb{R}^k)$ is the space of all k-dimensional predictable processes $\{Y_t\}_{t\in[0,T]}$ such that $\mathbb{E}[\sup_{t\in[0,T]} |Y_t|^p] < \infty$,
- $\mathcal{H}^p(\mathbb{R}^d)$ is the space of all d-dimensional predictable processes $\{Z_t\}_{t\in[0,T]}$ such that $\mathbb{E}[(\int_0^T |Z_t|^2 dt)^{p/2}] < \infty$,
- $\mathcal{S}^\infty(\mathbb{R}) = \cup_{p>2}\mathcal{S}^p(\mathbb{R})$, those processes are bounded for all t, $\mathbb{P}$-almost surely.
- $\mathcal{H}^\infty(\mathbb{R}^d) = \cup_{p>2}\mathcal{H}^p(\mathbb{R}^d)$.

Theorem 2.2. *[6] **(Existence and uniqueness for BSDEs with random Lipschitz condition)** Assume that BSDE(ξ, f) satisfies the random Lipschitz condition*

$$|f(t,z) - f(t,\widehat{z})| \leq H_t|z - \widehat{z}|, \qquad \forall (t,z,\widehat{z}) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d, \tag{2.12}$$

where H_t is integrable, predictable, non-negative and $\int_0^\cdot H_t dW_s$ is a BMO-martingale. Further assume that for some $p^* > 1$ it holds that

$$\mathbb{E}\left[|\xi|^{p^*} + \left(\int_0^T |f(s,0)|ds\right)^{p^*}\right] < \infty \tag{2.13}$$

and $|f(t,z)| \leq g(t) + H_t|z|$, $\forall (t,z) \in [0,T] \times \mathbb{R}^d$, where $g : [0,T] \times \Omega \to \mathbb{R}_+$ is a function satisfying

$$\mathbb{E}\left[\left(\int_0^T g(s)ds\right)^{p^*}\right] < \infty.$$

Then there exist a unique solution $(Y,Z) \in \mathcal{S}^p(\mathbb{R}) \times \mathcal{H}^p(\mathbb{R}^d)$ for $1 < p < p^*$.

2.3 History

In 1978 Bismut [3] introduced a linear BSDE, as the adjoint equation to the maximum principle, in optimal stochastic control theory. First, in 1990, Pardoux and Peng published a paper [21] were they proved the existence of an adapted solution, to a BSDE, in the case of Lipschitz continuous generator. After that, the subject grew rapidly. The awareness, of the possibility to use BSDEs in finance, increased. In 1997 Karoui, Peng and Quenez published an important and long reference paper on BSDEs [12], containing theory as well as applications in finance. It is still frequently referred to in papers published today. In 2000 Kobylanski published an important paper [13] on quadratic BSDEs and the connection to viscosity and Sobolev solutions to non linear parabolic PDEs. That paper too is still frequently referred to in papers dealing with quadratic BSDEs. A large amount of other papers has been published on the subject. There are a lot of variations, BSDEs driven by Levy processes, BSDEs with jumps or delays, numerics of BSDEs, etc.. Much of the motivation has come from mathematical finance. BSDEs are useful in utility maximization problems in incomplete markets, i.e., markets were there are risks not possible to hedge completely. So far there hasn't been written any books about the subject. However there are two conference texts namely [11] from 1997 and [16] from 1999. The paper in this line that that is fundamental for this thesis is [2] "Pricing and hedging of derivatives based on non-tradable underlyings" by Ankirchner, Imkeller and Dos Reis.

Chapter 3

Weak derivatives and Sobolev spaces

In the first section of this chapter so called weak or distributional derivatives will be introduced. Weak derivatives need not be functions but rather distributions. The function spaces that functions with weak partial derivatives live in, namely Sobolev spaces is also introduced. The main tool for proving that the solution Y of a FBSDE is a member of a Sobolev space is the concept of Dirichlet spaces. Those will be introduced in the second section together with a useful result. In the third section, the results of section two will be applied and extended to stochastic differential equations. Results for degenerate SDEs with random initial value will be presented together with an introduction to non-degenerate SDEs. All the results in this chapter are well-known.

3.1 The Sobolev space H^1_{loc}

A function $\psi : \mathbb{R}^m \to \mathbb{R}$ is said to have compact support if there is a compact set $K \subset \mathbb{R}^m$ such that $\psi = 0$ outside K. A continuously differentiable function with compact support is called a test function. If $g : \mathbb{R} \to \mathbb{R}$ is a square integrable continuously differentiable function, then integration by parts applies

$$\int_K g \frac{d\psi}{dx_i} dx = -\int_K \frac{dg}{dx_i} \psi dx. \tag{3.1}$$

The boundary term disappears since ψ vanishes on the boundary of K. This operation is obviously possible when g is continuously differentiable. If g is not continuously differentiable but there exist a function dg/dx_i in

$$L_2(K) = \left\{ f : K \to \mathbb{R} : \int_K |f(x)|^2 \, dx < \infty \right\}$$

such that (3.1) holds, then dg/dx_i is called a weak partial derivative. If this derivative exists, then it is unique in $L_2(K)$ by the Riesz representation theorem.

The space $L^2_{\mathrm{loc}}(\mathbb{R}^m)$ is the space of functions from $\mathbb{R}^m$ to $\mathbb{R}$ that are Lebesgue square integrable on every compact subset $K \subset \mathbb{R}^m$. Define the local Sobolev space $H^1_{\mathrm{loc}}(\mathbb{R}^m)$ by

$$H^1_{\mathrm{loc}}(\mathbb{R}^m) = \left\{ f \in L^2_{\mathrm{loc}}(\mathbb{R}^m) : \frac{\partial}{\partial x_i} f \in L^2_{\mathrm{loc}}(\mathbb{R}^m), 1 \le i \le m \right\}.$$

(see [15]). This space is of interest in this thesis as it will be proved that $Y \in H^1_{\text{loc}}(\mathbb{R}^m)$ where (X, Y, Z) is the solution to a certain quadratic FBSDE.

The non-differentiable functions considered in this text will be continuously differentiable almost everywhere. A subset of those functions are the locally Lipschitz continuous functions, and a subset of those functions in turn are the globally Lipschitz continuous functions.

Example 3.1. The weak derivative of the payoff function of a European call option $F(x) = \max(0, x - K)$, $K \in \mathbb{R}$, is the equivalence class of functions

$$\frac{dF}{dx}(x) = \begin{cases} 0, & x < K \\ C, & x = K \\ 1, & x > K \end{cases}$$

for arbitrary $C \in \mathbb{R}$. This makes better sense in $L_2(\mathbb{R})$ where dF/dx is unique rather than arbitrary since objects in L_2 are equivalence classes up to Lebesgue almost everywhere equality. On the other hand, if C is fixed then dF/dx is called a version.

3.2 Two Dirichlet spaces $\mathbf{d}$ and $\widetilde{\mathbf{d}}$

Let $h : \mathbb{R}^m \longrightarrow \mathbb{R}$ be a fixed, continuous and positive function satisfying $\int_{\mathbb{R}^m} h(x)dx = 1$ and $\int_{\mathbb{R}^m} |x|^2 h(x)dx < \infty$. The space $\mathbf{d}$ is defined by

$$\mathbf{d} = \left\{ f \in L_2(\mathbb{R}^m, h) : \frac{\partial}{\partial x_i} f \in L_2(\mathbb{R}^m, h), 1 \leq i \leq m \right\},$$

where

$$L_2(\mathbb{R}^m, h) = \left\{ f : \mathbb{R}^m \to \mathbb{R} : \left(\int_{\mathbb{R}^m} |f(x)|^2 h(x)dx \right)^{1/2} := \|f\|_{L_2(\mathbb{R}^m,h)} < \infty \right\}.$$

The derivative is considered in the weak sense. The space $\mathbf{d}$ equipped with the norm

$$\|f\|_{\mathbf{d}} = \left[\|f\|^2_{L_2(\mathbb{R}^m,h)} + \sum_{i=1}^m \|\frac{\partial}{\partial x_i} f\|^2_{L_2(\mathbb{R}^m,h)} \right]^{1/2}$$

is a so called classical Dirichlet space. This space is a Hilbert space and a subspace of the local Sobolev space $H^1_{\text{loc}}(\mathbb{R}^m)$.

Next, define an enlarged probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$, where $\widetilde{\Omega} = \Omega \times \mathbb{R}^m$, $\mathcal{F}$ is the product σ-algebra of $\mathcal{F}$ and the Borel σ-algebra of $\mathbb{R}^m$ and $\widetilde{\mathbb{P}}$ is the product measure $\mathbb{P} \times hdx$. The expected value on $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ will be denoted $\widetilde{\mathbb{E}}$. Let

$$L_2(\widetilde{\Omega}) = \left\{ f : \Omega \times \mathbb{R}^m \to \mathbb{R} : \|f\|_{L_2(\widetilde{\Omega})} = \widetilde{\mathbb{E}}[|f|^2]^{1/2} = \left(\int_{\mathbb{R}^m} \mathbb{E}[|f(x)|^2] h(x)dx \right)^{\frac{1}{2}} < \infty \right\},$$

and $\widetilde{D}_i$ be the space of functions $u : \Omega \times \mathbb{R}^m \to \mathbb{R}$ that has a version $\widetilde{u}$ such that $\varepsilon \mapsto \widetilde{u}(\omega, x + \varepsilon e_i)$ is locally absolutely continuous $\forall(\omega, x) \in \widetilde{\Omega}, \varepsilon \in \mathbb{R}$, $1 \leq i \leq m$. Here e_i is the i-th unit vector in $\mathbb{R}^m$. Locally absolutely continuous functions are continuously differentiable almost everywhere. Lipschitz continuous functions satisfies this property. Further define the operator ∇_i on $\widetilde{D}_i$ by

$$\nabla_i u(x,\omega) = \lim_{\varepsilon \to 0} \frac{\widetilde{u}(x+\varepsilon e_i,\omega) - \widetilde{u}(x,\omega)}{\varepsilon}$$

for $\widetilde{u}$ being a locally absolutely continuous version of u.

Now we are ready to define a second Dirichlet space $\widetilde{\mathbf{d}}$ on $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ by

$$\widetilde{\mathbf{d}} = \left\{ f \in L_2(\widetilde{\Omega}) \bigcap \left(\bigcap_{i=1}^{m} \widetilde{D}_i \right) : \nabla_i f \in L_2(\widetilde{\Omega}), 1 \leq i \leq m \right\},$$

The space $\widetilde{\mathbf{d}}$ equipped with the norm

$$\| \cdot \|_{\widetilde{\mathbf{d}}} = \left[\| \cdot \|^2_{L_2(\widetilde{\Omega})} + \sum_{i=1}^{m} \| \nabla_i \cdot \|^2_{L_2(\widetilde{\Omega})} \right]^{\frac{1}{2}}$$

is a so called general Dirichlet space.

In Chapter 4 $\widetilde{\mathbf{d}}$ will be used in the proof that the process Y of the solution (X, Y, Z) of a FBSDE belongs to $\mathbf{d}$. The following proposition connects the two spaces:

Proposition 3.2. *[5] If $u \in \widetilde{\boldsymbol{d}}$, then*

$$u(\cdot,\omega) \in \boldsymbol{d} \quad \text{and} \quad \frac{\partial}{\partial x_i} u(x,\omega) = \nabla_i u(x,\omega) \quad \widetilde{\mathbb{P}} - a.s., \quad 1 \leq i \leq m.$$

3.3 Densities and non-degeneracy of SDEs

This section extends the results from the previous section, to stochastic differential equations. It also explains important properties of certain SDEs. Consider for $t \in [0, T]$ the SDE(b, σ). The coefficients $b : [0, T] \times \mathbb{R}^m \longrightarrow \mathbb{R}^m$ and $\sigma : [0, T] \times \mathbb{R}^m \longrightarrow \mathbb{R}^{m \times d}$ satisfies global Lipschitz and linear growth conditions (2.2). The SDE is said to be non-degenerate if, for some constant $C > 0$,

$$\xi^* \sigma(t,x) \sigma^*(t,x) \xi \geq C|\xi|^2, \quad \forall (t,x,\xi) \in [0,T] \times \mathbb{R}^m \times \mathbb{R}^m \tag{3.2}$$

holds. Here, σ^* and ξ^* denotes the transpose of σ and ξ.

Theorem 3.3. *[4] Given assumption (3.2), X_t^x has a density for all $(t, x) \in (0, T] \times \mathbb{R}^m$.*

Example 3.4. Consider the SDEs:

$$dX_t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} dW_t, \quad d\widehat{X}_t = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} dW_t,$$

for a 2-dimensional Wiener process W_t and $X_0 = \widehat{X}_0 = 0$. The first has solution $X_t = (W_t^1, W_t^2)$ and the second $\widehat{X}_t = (W_t^1, W_t^1)$. It is clear that X_t will evolve freely in the entire $\mathbb{R}^2$-plane and that $\widehat{X}_t$ will evolve along the line $\mathcal{L} = \{(x,x) : x \in \mathbb{R}\} \subset \mathbb{R}^2$. X is non-degenerate and $\widehat{X}$ is degenerate. This is a trivial example, but given the non-degeneracy condition (3.2), the paths of the solutions to SDE(b, σ), will not be limited to any Lebesgue null-set of $\mathbb{R}^m$.

Example 3.5. Later in the thesis, non-degeneracy will be used to conclude that

$$\mathbb{E}\left[|\xi^1(X_T) - \xi^2(X_T)|\right] = 0, \tag{3.3}$$

where $\xi^1 : \mathbb{R}^m \to \mathbb{R}$ and $\xi^2 : \mathbb{R}^m \to \mathbb{R}$ satisfies $\xi^1(x) = \xi^2(x)$ except at Lebesgue null-sets of $\mathbb{R}^m$. Suppose that X and $\widehat{X}$ are the processes in the preceding example and that $\xi^1(x) = \xi^2(x)$ except on the line $\mathcal{L}$. Then (3.3) holds for X but not for $\widehat{X}$. The conclusion would be impossible for any degenerate SDE regardless of what null-set $\xi^1(x)$ and $\xi^2(x)$ differs on.

Chapter 4

Weak differentiability of quadratic non-degenerate FBSDEs

In this chapter the main mathematical contributions of this thesis will be presented. Results for classical differentiability of the solution process Y of a quadratic FBSDE, proved in [2], will be generalized. In the first section our technical assumptions will be presented and in the second section a collection of results that will be needed are listed. In the third section, weak differentiability of Y will be stated and proved when the coupled forward SDE is non-degenerate. A useful representation result is also proved. In the last section the same will be proved when the forward SDE is degenerate. In that case a slightly different FBSDE will be considered and finally proved to represent the weak gradient of Y.

4.1 Assumptions

Consider FBSDE(b, σ, g, f). The Coefficients b and σ are assumed to satisfy Lipschitz and linear growth condition (2.2) together with non-degeneracy condition (3.2). The terminal function $g : \mathbb{R}^m \to \mathbb{R}$ is assumed to be deterministic, bounded, measurable and Lipschitz continuous. The generator $f : [0,T]\times\mathbb{R}^m\times\mathbb{R}^d \to \mathbb{R}$ is assumed to be measurable, continuously differentiable in x and z and satisfy

$$\begin{cases} |f(t,x,z)| & \leq \quad C(1+|z|^2) & \text{a.s.}, \\ |f(t,x,z) - f(t,\bar{x},z)| & \leq \quad C(1+|z|)|x-\bar{x}| & \text{a.s.}, \\ |\nabla_z f(t,x,z)| & \leq \quad C(1+|z|) & \text{a.s.}, \\ |\nabla_x f(t,x,z) - \nabla_x f(t,\bar{x},\bar{z})| & \leq \quad C(1+|z|+|\bar{z}|)(|x-\bar{x}|+|z-\bar{z}|) & \text{a.s.}, \end{cases}$$

for some constant $C > 0$, $\forall (t,x,\bar{x},z,\bar{z}) \in [0,T]\times\mathbb{R}^m\times\mathbb{R}^m\times\mathbb{R}^d\times\mathbb{R}^d$. When these assumptions holds the FBSDE is said to satisfy assumption **(A)**. We call the FBSDE(b,σ,g,f) under assumption **(A)** quadratic, to distinguish it from generators globally Lipschitz continuous in z.

4.2 Some useful results

The following moment estimate will be the main tool when proving our main result:

Lemma 4.1. *[2](Moment estimate for BSDEs with random Lipschitz generator)* *Consider the BSDE(ξ, f). Suppose that condition (2.12) holds and that for all $\beta \geq 1$ we have $\int_0^T |f(s,0)|ds \in L^\beta(\mathbb{P})$. Let $p > 1$. Then there exist constants $q > 1$ and $C > 0$, depending only on p, T and the BMO-norm of $\int_0^\cdot H_t dt$ where $\{H_t\}_{t\in[0,T]}$ is the random Lipschitz bound, such that we have*

$$\mathbb{E}\left[\sup_{t\in[0,T]} |Y_t|^{2p}\right] + \mathbb{E}\left[(\int_0^T |Z_s|^2 ds)^p\right] \leq C\left(\mathbb{E}\left[|\xi|^{2pq} + (\int_0^T |f(s,0)|ds)^{2pq}\right]\right)^{\frac{1}{q}}.$$

The following three results will also be of great importance.

Lemma 4.2. *[17] Consider the FBSDE(b, σ, g, f). Given assumption (A), the process $\int_0^\cdot Z_r^{t,x} dW_r$ is a BMO-martingale. The BMO-norm only depends on the terminal value, the function $f(s, X_s^{t,x}, 0)$, and the duration $T - t$.*

Lemma 4.2 shows that if $Z_s^{t,x}$ is the random Lipschitz bound for a generator of a BSDE, then the moment estimate Lemma 4.1 can be applied. This will be used frequently in the proof of the main results of this thesis.

Proposition 4.3. *Under assumption (A) the FBSDE(b, σ, g, f) satisfies a random Lipschitz condition with BMO bound. Moreover, the solution (X, Y, Z) is unique with $(X, Y, Z) \in \mathcal{S}^\infty(\mathbb{R}^m) \times \mathcal{S}^\infty(\mathbb{R}) \times \mathcal{H}^\infty(\mathbb{R}^d)$.*

Proof. First, X satisfies the usual Itô conditions and is hence well defined and unique. Proposition 4.7 guaranties $X \in \mathcal{S}^p(\mathbb{R}^m)$ for all $p \geq 2$, i.e., $X \in \mathcal{S}^\infty(\mathbb{R}^m)$. Next, the generator is differentiable and hence by the mean value theorem and assumption $\nabla_z f(t, x, z) \leq C(1+|z|)$, $\exists \lambda \in [0, 1]$:

$$\begin{aligned} |f(t,x,z) - f(t,x,\widehat{z})| &\leq |\nabla_z f(t,x,\lambda z + (1-\lambda)\widehat{z})||z - \widehat{z}| \leq C(1 + |\lambda z + (1-\lambda)\widehat{z}|)|z - \widehat{z}| \\ &\leq C(1 + |z| + |\widehat{z}|)|z - \widehat{z}|. \end{aligned}$$

This implies that the generator satisfies a random Lipschitz condition with Lipschitz bound $C(1 + |Z_s^{t,x}| + |\widehat{Z}_s^{t,x}|)$. Lemma 4.2 implies that $\int_0^\cdot Z_r^{t,x} dW_r$ is a BMO martingale and hence the bound satisfies the assumption of Theorem 2.2. Moreover the boundedness assumption on g and $f(t, 0, 0)$ implies $(Y, Z) \in \mathcal{S}^\infty(\mathbb{R}) \times \mathcal{H}^\infty(\mathbb{R}^d)$ by Theorem 2.2. □

Lemma 4.4. *Consider the FBSDE(b, σ, g, f). Given assumption (A), the mapping $x \mapsto Y_s^{t,x}$ is Lipschitz continuous for all $t \in [0, T]$ and $s \in [t, T]$.*

Proof. The Lemma as Lemma 6.3 in [2] was stated under the stronger assumptions of Theorem 4.5. However, this was done for notational simplicity. The proof carries over to our setting without changes. □

Next follows two important results on classical differentiability for quadratic FBSDEs.

Theorem 4.5. *[2] Consider the FBSDE(b,σ,g,f). Assume (A), with the additional requirements that the terminal function g is twice continuously differentiable and b and σ are continuously differentiable in x with Lipschitz continuous first derivative. Then for every fixed $t \in [0,T]$ $X_s^{t,x}$ and $Y_s^{t,x}$ are continuous in s and continuously differentiable in x. Moreover, there exist a process $\nabla_x Z_s^{t,x} \in \mathcal{H}^2(\mathbb{R}^d)$ such that $(\nabla_x Y_s^{t,x}, \nabla_x Z_s^{t,x})$, for $s \in [t,T]$, is the solution to the BSDE:*

$$\begin{aligned}\nabla_x Y_s^{t,x} = \;& \nabla_x g(X_T^{t,x})\nabla_x X_T^{t,x} - \int_s^T \nabla_x Z_r^{t,x} dW_r \\ & + \int_s^T [\nabla_x f(r, X_r^{t,x}, Z_r^{t,x})\nabla_x X_r^{t,x} + \nabla_z f(r, X_r^{t,x}, Z_r^{t,x})\nabla_x Z_r^{t,x}]dr\end{aligned}$$

Theorem 4.6. *Let the assumptions of the previous theorem hold. Then for $\forall s \in [t,T]$ and $u(t,x) := Y_t^{t,x}$*

$$Z_s^{t,x} = \nabla_x u(s, X_s^{t,x})\sigma(s, X_s^{t,x})$$

for almost all $t \in [0,T]$, $\mathbb{P}$-almost surely.

Proof. The theorem was stated [2] with the extra assumption of the existence of a sequence $\{f^n\}_{n\geq 1}$, of generators, Lipschitz continuous in z, converging locally uniformly to f. The assumption is not needed since it is always possible to find such a sequence, when f is quadratic. □

The following estimate for classical SDEs will be needed.

Theorem 4.7. *[14] Consider the SDE(b,σ) with initial value $x : \Omega \to \mathbb{R}^m$. Assume that $b : \Omega \times [0,T] \times \mathbb{R}^m \to \mathbb{R}^m$ and $\sigma : \Omega \times [0,T] \times \mathbb{R}^m \to \mathbb{R}^{m\times d}$ are Lipschitz continuous in the space variable. Then for any $p \geq 2$, there exist a constant C, only depending on p, T and the Lipschitz bounds of b and σ, such that:*

$$\mathbb{E}\left[\sup_{r\in[0,T]} |X_r|^p\right] \leq C\left(\mathbb{E}\left[|x|^p\right] + \mathbb{E}\left[\int_0^T (|b(r,0)|^p + |\sigma(r,0)|^p)dr\right]\right).$$

Finally, the following inequality will be used frequently.

Lemma 4.8. *For $x_i, \ldots, x_k \in V$, where $(V, \|\cdot\|)$ is a normed vector space it holds that for $p \geq 0$, $\|x_1 + \ldots + x_k\|^p \leq k^p(\|x_1\|^p + \ldots + \|x_k\|^p)$.*

Proof. $\|x_1 + \ldots + x_k\|^p \leq (k \max(\|x_1\|, \ldots, \|x_k\|))^p \leq k^p(\|x_1\|^p + \ldots + \|x_k\|^p)$. □

4.3 Main result

In this Section weak differentiability of FBSDE(b,σ,g,f) will be stated and proved under assumption (A). The following FBSDE is known as the variational equation of FBSDE(b,σ,g,f).

$$\begin{cases} \Phi^{t,x}_{i,s} = & e_i + \int_t^s \nabla_x b(r, X^{t,x}_r)\Phi^{t,x}_{i,r}dr + \sum_{j=1}^d \int_t^s \nabla_x \sigma_j(r, X^{t,x}_r)\Phi^{t,x}_{i,r}dW^j_r \\ \Psi^{t,x}_{i,s} = & \nabla_x g(X^{t,x}_T)\Phi^{t,x}_{i,T} + \int_s^T (\nabla_x f(r, X^{t,x}_r, Z^{t,x}_r)\Phi^{t,x}_{i,r} \\ & +\nabla_z f(r, X^{t,x}_r, Z^{t,x}_r)\Gamma^{t,x}_{i,r})dr - \int_s^T \Gamma^{t,x}_{i,r}dW_r, \qquad i = 1,\dots,m. \end{cases} \tag{4.1}$$

for $s \in [t,T]$. It will be denoted ∇_iFBSDE(b,σ,g,f), $i = 1,\dots,m$, componentwise or ∇FBSDE(b,σ,g,f) otherwise. Here, $\nabla_x b$, $\nabla_x \sigma$ and $\nabla_x g$ are the gradients of b, σ and g in the weak sense. The index i denotes the i:th column of Φ, Ψ and Γ. Further, $\nabla_x \sigma_j$ denotes the gradient of the j'th row of σ. If g, σ, g, X, Y and Z were differentiable w.r.t. x then Φ, Ψ and Γ would be the gradients of X, Y and Z. The following Theorem states that $Y^{t,x}_s$ is weakly differentiable with respect to x and that $\Psi^{t,x}_s$ is its weak gradient.

Theorem 4.9. *Let assumption (A) hold. Then,*

(i) the function, $x \mapsto Y^{t,x}_s$ belongs to $H^1_{loc}(\mathbb{R}^m)$ $\mathbb{P}$-a.s., $\forall t \in [0,T], \forall s \in [t,T]$.

(ii) the weak gradient $\nabla_x Y^{t,x}_s = \Psi^{t,x}_s$, for almost all x $\mathbb{P}$-a.s., $\forall t \in [0,T], \forall s \in [t,T]$, where $(\Phi^{t,x}, \Psi^{t,x}, \Gamma^{t,x}) \in \mathcal{S}^\infty(\mathbb{R}^{m\times m}) \times \mathcal{S}^\infty(\mathbb{R}^{1\times m}) \times \mathcal{H}^\infty(\mathbb{R}^{d\times m})$ is the unique solution to ∇FBSDE(b,σ,g,f).

The proof will be divided into four steps. The main idea is to prove that $Y^{t,\cdot}_s \in L_2(\widetilde{\Omega})$ and $\nabla_x Y^{t,\cdot}_s \in L_2(\widetilde{\Omega})$, i.e. that $Y^{t,x}_s$ belongs to the Dirichlet space $\widetilde{\mathbf{d}}$, $\forall t \in [0,T], \forall s \in [t,T]$, and that Equation 4.1 has a well defined solution. After that, the result follows easily in the last step of the proof. The proof techniques are mainly those of [19] and [2]. The first of these papers [19] gives a similar proof for weak differentiability for BSDEs with Lipschitz continuous generator in x and z and $m = d$. The second of these papers [2] contains results and techniques for working with quadratic BSDEs and BSDEs satisfying a random Lipschitz condition.

Proof. **Step 1:** Let $\phi : \mathbb{R}^m \to \mathbb{R}$ be a infinitely continuously differentiable and nonnegative function with support in the unit ball and $\int_{\mathbb{R}^m} \phi(x)dx = 1$. Then the functions, known as mollifiers, defined by $\phi^n(x) = n^m\phi(nx), n \geq 1$ has the same properties for all n, only that the support is vanishing as $n \to \infty$. By convolutions, define the functions

$$\begin{cases} b^n(t,x) & = (b * \phi^n)(t,x) & = \int_{\mathbb{R}^m} b(t, x-\xi)\phi^n(\xi)d\xi, \\ \sigma^n(t,x) & = (\sigma * \phi^n)(t,x) & = \int_{\mathbb{R}^m} \sigma(t, x-\xi)\phi^n(\xi)d\xi, \\ g^n(x) & = (g * \phi^n)(x) & = \int_{\mathbb{R}^m} g(x-\xi)\phi^n(\xi)d\xi. \end{cases} \tag{4.2}$$

It is known that b^n, σ^n and g^n are in $\mathcal{C}^\infty$, $\forall n \in \mathbb{Z}_+$ and converges uniformly to b, σ and g and that $\nabla_x b^n$, $\nabla_x \sigma^n$ and $\nabla_x g^n$ converges dx-a.e. to $\nabla_x b$, $\nabla_x \sigma$ and $\nabla_x g$ as $n \to \infty$[9]. Consider for $(t,x,n) \in [0,T] \times \mathbb{R}^m \times \mathbb{Z}_+$ and $s \in [t,T]$ the sequence of FBSDE(b^n, σ^n, g^n, f) with corresponding solution $(X^{t,x,n}, Y^{t,x,n}, Z^{t,x,n})$.

First, consider the convergence for $X_s^{t,x,n}$. Let

$$\Delta X_s^{t,x,n} := X_s^{t,x,n} - X_s^{t,x}.$$

The difference satisfies

$$\Delta X_s^{t,x,n} = \int_t^s (b^n(r, X_r^{t,x,n}) - b(r, X_r^{t,x}))dr + \int_t^s (\sigma^n(r, X_r^{t,x,n}) - \sigma(r, X_r^{t,x}))dW_r.$$

By using Theorem 4.7 it follows that for all $p \geq 2$, $\exists C > 0$ such that

$$\begin{aligned} \mathbb{E}\left[\sup_{s\in[t,T]} |\Delta X_s^{t,x,n}|^p\right] &\leq C\mathbb{E}\left[\int_0^T [|b^n(t,0) - b(t,0)|^p + |\sigma^n(t,0) - \sigma(t,0)|^p]dt\right] \\ &\to 0 \quad \text{as } n \to \infty. \end{aligned} \tag{4.3}$$

Here the uniform constant C exists since it only depends on p, T and the Lipschitz bounds for b, b^n, σ and σ^n. The Lipschitz bounds of b^n and σ^n isn't higher than that of b and σ and p and T are fixed. Further the convergence in (4.3) is bounded. By assumption, $b(t,0)$ and $\sigma(t,0)$ are bounded for all $t \in [0,T]$. The same holds for $b^n(t,0)$ and $\sigma^n(t,0), \forall (n,t) \in \mathbb{Z}_+ \times [0,T]$. The coefficients b^n and σ^n converges uniformly to b and σ as $n \to \infty$.

Let

$$\begin{cases} \Delta Y_s^{t,x,n} & := Y_s^{t,x,n} - Y_s^{t,x}, \\ \Delta Z_s^{t,x,n} & := Z_s^{t,x,n} - Z_s^{t,x}, \\ \Delta g^{t,x,n} & := g^n(X_T^{t,x,n}) - g(X_T^{t,x}). \end{cases}$$

The difference process $(\Delta Y_s^{t,x,n}, \Delta Z_s^{t,x,n})$ satisfies the BSDE$(\Delta g(X_T^{t,x}), \widehat{f}^n)$ with generator

$$\begin{aligned} \widehat{f}^n &:= f(s, X_s^{t,x,n}, Z_s^{t,x,n}) - f(s, X_s^{t,x}, Z_s^{t,x}) \\ &= f(s, X_s^{t,x,n}, Z_s^{t,x,n}) - f(s, X_r^{t,x}, Z_s^{t,x,n}) + f(s, X_s^{t,x}, Z_s^{t,x,n}) - f(s, X_s^{t,x}, Z_s^{t,x}). \end{aligned}$$

Line integral transformations are used, to show that the generator satisfies a random Lipschitz condition. The chain rule gives,

$$\begin{aligned} f(s, X_s^{t,x,n}, Z_s^{t,x,n}) - f(s, X_s^{t,x}, Z_s^{t,x,n}) &= \int_0^1 \frac{d}{d\theta} f(s, X_s^{t,x} + \theta(X_s^{t,x,n} - X_s^{t,x}), Z_s^{t,x,n})d\theta \\ &= \int_0^1 \nabla_x f(s, X_s^{t,x} + \theta(X_s^{t,x,n} - X_s^{t,x}), Z_s^{t,x,n})d\theta \Delta X_s^{t,x,n} \\ &:= J_s^n \Delta X_s^{t,x,n} \end{aligned}$$

and similarly

$$
\begin{aligned}
f(s, X_s^{t,x}, Z_s^{t,x,n}) - f(s, X_s^{t,x}, Z_s^{t,x}) &= \int_0^1 \nabla_z f(s, X_s^{t,x}, Z_s^{t,x} + \theta(Z_s^{t,x,n} - Z_s^{t,x}))d\theta \Delta Z_s^{t,x,n} \\
&:= H_s^n \Delta Z_s^{t,x,n}.
\end{aligned}
$$

Hence

$$\widehat{f}^n(s,v) = H_s^n v + J_s^n \Delta X_s^{t,x,n}.$$

The generator clearly satisfies the random Lipschitz condition,

$$|\widehat{f}^n(s,v) - \widehat{f}^n(s,v')| \leq H_s^n |v - v'|$$

for all $(s,v,v') \in [t,T] \times \mathbb{R}^d \times \mathbb{R}^d$. Now, by assumption and the mean value theorem,

$$
\begin{aligned}
H_s^n &= \int_0^1 \nabla_z f(s, X_s^{t,x}, Z_s^{t,x} + \theta(Z_s^{t,x,n} - Z_s^{t,x}))d\theta \\
&\leq C \int_0^1 (1 + |Z_s^{t,x} + \theta(Z_s^{t,x,n} - Z_s^{t,x})|)d\theta \\
&= C(1 + |Z_s^{t,x} + \xi(Z_s^{t,x,n} - Z_s^{t,x})|) \\
&\leq C(1 + |Z_s^{t,x}| + |Z_s^{t,x,n}|)
\end{aligned}
\tag{4.4}
$$

for some $0 \leq \xi \leq 1$ and for each $n \in \mathbb{Z}_+$. Lemma 4.2 states that $\int_0^\cdot Z_r^{t,x} dW_r$ and $\int_0^\cdot Z_r^{t,x,n} dW_r$ are BMO-martingales $\forall n \in \mathbb{Z}_+$ and it follows from (4.4) that so is also $\int_0^\cdot H_r^n dW_r$, $\forall n \in \mathbb{Z}_+$. Next, a uniform bound for the BMO-norms of $\int_0^\cdot Z_r^{t,x,n} dW_r$, are sought for all $n \in \mathbb{Z}_+$. By Lemma 4.2 the BMO-norms only depends on $g^n(X_T^{t,x,n})$, $f(s, X_s^{t,x,n}, 0)$, $s \in [t,T]$, and $T - t$. The terminal function g is bounded, and the mollifiers ϕ^n integrates to one, hence the convolution (4.2) does not increase the bound. It follows that the $|g^n|$ are uniformly bounded. Next, by the assumption $|f(s,x,z)| \leq C(1 + |z|^2)$ it follows that $|f(s, X_s^{t,x,n}, 0)|$ is uniformly bounded. Finally, $T - t$ is the same for all n. It follows that the BMO-norms of $\| \int_0^\cdot Z_r^{t,x,n} dW_r \|_{BMO} < \beta$, for some $\beta \geq 0$, $\forall n \in \mathbb{Z}_+$. Therefore the same holds for $\int_0^\cdot H_r^n dW_r$ from the estimate (4.4).

Now, the moment estimate Lemma 4.1 will be applied on $\Delta Y^{t,x,n}$ and $\Delta Z^{t,x,n}$. The constants $C_n > 0$ and $q_n > 1$, appearing in every estimation of $\Delta Y^{t,x,n}$ and $\Delta Z^{t,x,n}$, $\forall n \in \mathbb{Z}_+$, only depends on p, $T - t$ and the BMO-norms. Since a uniform bound of the BMO-norms has been proved, it follows that q_n and C_n, are uniformly bounded by some constants $C > 0$ and $q > 1$. Lemma 4.1 therefore gives that, for any $p > 1$ and every $n \in \mathbb{Z}_+$ there exists constants $q > 1$ and $C > 0$ such that

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{r \in [t,T]} |\Delta Y_r^{t,x,n}|^{2p}\right] + \mathbb{E}\left[\left(\int_t^T |\Delta Z_r^{t,x,n}|^2 dr\right)^p\right] \\
&\leq C\left(\mathbb{E}\left[|\Delta g^{t,x,n}|^{2pq}\right] + \mathbb{E}\left[(\int_t^T |J_r^n \Delta X_r^{t,x,n}| dr)^{2pq}\right]\right)^{1/q}
\end{aligned}
\tag{4.5}
$$

First, by Lemma 4.8

$$\mathbb{E}\left[|\Delta g^{t,x,n}|^{2pq}\right] \leq C\mathbb{E}\left[|g^n(X_T^{t,x,n}) - g^n(X_T^{t,x})|^{2pq} + |g^n(X_T^{t,x}) - g(X_T^{t,x})|^{2pq}\right]$$
$$\rightarrow 0 \text{ as } n \rightarrow \infty.$$

The convergence is bounded and follows since $g^n \rightarrow g$, $X_s^{t,x,n} \rightarrow X_s^{t,x}$ from (4.3) and the boundedness of g and g^n. The differentiability of f in x and assumption $|f(t,x,z) - f(t,\hat{x},z)| \leq C(1+|z|)|x-\hat{x}|$ implies that

$$|\nabla_x f(t,x,z)| \leq C(1+|z|). \tag{4.6}$$

Second, by (4.6), Cauchy-Schwartz inequality and using the fact that

$$\int_t^T |f(r)|^p dr \leq (T-t) \sup_{r\in[0,T]} |f(r)|^p \tag{4.7}$$

it follows that

$$\begin{aligned} &\mathbb{E}\left[\left(\int_t^T |J_r^n \Delta X_r^{t,x,n}| dr\right)^{2pq}\right] \\ &\leq C\mathbb{E}\left[\left(\int_t^T (1+|Z_r^{t,x,n}|)^2 dr\right)^{2pq}\right]^{1/2} \mathbb{E}\left[\sup_{r\in[t,T]} |\Delta X_r^{t,x,n}|^{4pq}\right]^{1/2} \\ &\rightarrow 0 \text{ as } n\rightarrow\infty. \end{aligned} \tag{4.8}$$

The first factor of (4.8) is uniformly bounded, since it can be estimated by Lemma 4.1, with uniform constants C and q by the same arguments as above. The convergence to zero of the second factor follows from (4.3).

It has now been proved that $Y^{t,x,n} \rightarrow Y^{t,x}$ in $\mathcal{S}^\infty(\mathbb{R})$ and that $Z^{t,x,n} \rightarrow Z^{t,x}$ in $\mathcal{H}^\infty(\mathbb{R}^d)$. Recall from Section 3.2 that $h : \mathbb{R}^m \rightarrow \mathbb{R}$ is a, continuous positive function satisfying $\int_{\mathbb{R}^m} h(x)dx = 1$ and $\int_{\mathbb{R}^m} |x|^2 h(x)dx < \infty$. It follows that such a function must be bounded. Both $Y_s^{t,x,n}$ and $Y_s^{t,x}$ are in $S^\infty(\mathbb{R})$ by Proposition 4.3 and hence bounded, for almost all $s \in [t,T]$, almost surely. Now, by bounded convergence

$$\lim_{n\rightarrow\infty} \mathbb{E}\left[\int_{\mathbb{R}^m} |Y_s^{t,x,n} - Y_s^{t,x}|^{2p} h(x)dx\right] = 0.$$

Hence, $Y_s^{t,\cdot,n} \rightarrow Y_s^{t,\cdot}$ in $L_{2p}(\widetilde{\Omega}) \subset L_2(\widetilde{\Omega})$, $\forall s \in [t,T]$.

Step 2: The functions b, σ and g are all Lipschitz continuous. Hence, they are continuously differentiable almost everywhere, i.e., the weak partial derivatives equals the classical partial derivatives except at a set of Lebesgue measure zero.

The question of this step in the proof is to prove that the solution $(\Phi_s^{t,x}, \Psi_s^{t,x}, \Gamma_s^{t,x})$, to the variational equation ∇FBSDE(b,σ,g,f), is well defined, i.e., does not depend on Borel (dx a.e.) versions of the weak gradients $\nabla_x b, \nabla_x \sigma$ and $\nabla_x g$. For all $s \in [t,T]$, let $\nabla_x b^1 = \nabla_x b^2$, $\nabla_x \sigma^1 = \nabla_x \sigma^2$ and $\nabla_x g^1 = \nabla_x g^2$ except at a set $\mathcal{N} \subset [0,T]\times\mathbb{R}^m$ with Lebesgue measure zero.

Let $(\Phi^1_{i,s}, \Psi^1_{i,s}, \Gamma^1_{i,s})$ and $(\Phi^2_{i,s}, \Psi^2_{i,s}, \Gamma^2_{i,s})$, $i = 1, \dots, m$, be the solutions to ∇_iFBSDE(b, σ, g, f), $i = 1, \dots, m$, with versions $(\nabla_x b^1, \nabla_x \sigma^1, \nabla_x g^1)$ and $(\nabla_x b^2, \nabla_x \sigma^2, \nabla_x g^2)$ of the weak gradients respectively. The equation must be considered componentwise to be able to use moment estimate Lemma 4.1. The superscript ^{t,x} is omitted for notational simplicity. It will be proved that the solutions are identical.

The uniqueness of Φ^1_s and the identity $\Phi^1_s = \Phi^2_s$ in $\mathcal{S}^2(\mathbb{R}^{m\times m})$, has been proved under the non-degeneracy assumption (3.2) [19]. Denote the coefficients of $\Phi^{t,x}_{i,s}$,

$$\begin{cases} \alpha(r, \phi) = \nabla_x b(r, X^{t,x}_r)\phi, \\ \\ \beta(r, \phi) = \sum_{j=1}^d \nabla_x \sigma_j(r, X^{t,x}_r)\phi, \end{cases}$$

for $(\omega, r, \phi) \in \Omega \times [t, T] \times \mathbb{R}^{m\times m}$. It is known that the Frobenius norm (2.3) is sub-multiplicative, i.e., it satisfies $|AB| = |A||B|$. The coefficients therefore satisfies

$$\begin{aligned} |\alpha(t, \phi) - \alpha(t, \widehat{\phi})| &\le |\nabla_x b(r, X^{t,x}_r)||\phi - \widehat{\phi}| \\ \\ &\le m^2 C^2 |\phi - \widehat{\phi}|. \end{aligned}$$

and

$$\begin{aligned} |\beta(t, \phi) - \beta(t, \widehat{\phi})| &\le |\textstyle\sum_{j=i}^d \nabla_x \sigma_j(r, X^{t,x}_r)||\phi - \widehat{\phi}| \\ \\ &\le d m^2 C^2 |\phi - \widehat{\phi}|. \end{aligned}$$

The bounds follows since each element in the matrices $\nabla_x \sigma_j$ and $\nabla_x b$ are bounded by the common Lipschitz constant C of b and σ. The coefficients are hence Lipschitz continuous. Theorem 4.7 can be applied to conclude that $\Phi^{t,x}_{i,s} := \Phi^1_{i,s} = \Phi^2_{i,s}$, $i = 1, \dots, m$, are unique in $\mathcal{S}^\infty(\mathbb{R}^m)$. It remains to prove that $\Psi^1_{i,s} = \Psi^2_{i,s}$ and $\Gamma^1_{i,s} = \Gamma^2_{i,s}$ and that they are unique in $\mathcal{S}^\infty(\mathbb{R}^{1\times m})$ and $\mathcal{H}^\infty(\mathbb{R}^{d\times m})$ respectively, and hence well defined.

Let

$$\begin{cases} \Delta\Psi_{i,s} &:= \Psi^1_{i,s} - \Psi^2_{i,s}, \\ \\ \Delta\Gamma_{i,s} &:= \Gamma^1_{i,s} - \Gamma^2_{i,s}, \\ \\ \Delta\xi_i &:= (\nabla_x g^1(X^{t,x}_T) - \nabla_x g^2(X^{t,x}_T))\Phi^{t,x}_{i,T}. \end{cases}$$

The process $(\Delta\Psi_{i,s}, \Delta\Gamma_{i,s})$ satisfies $BSDE(\Delta\xi_i, \widehat{f})$ for

$$\widehat{f}(s, v) = \nabla_z f(s, X^{t,x}_s, Z^{t,x}_s)v.$$

It satisfies the random Lipschitz condition:

$$\begin{aligned} |\widehat{f}(s, v) - \widehat{f}(s, \widehat{v})| &\le \nabla_z f(s, X^{t,x}_s, Z^{t,x}_s)|v - \widehat{v}| \\ &\le C(1 + |Z^{t,x}_s|)|v - \widehat{v}| \end{aligned}$$

$\forall (s, v, \widehat{v}) \in [t, T] \times \mathbb{R}^{m\times m} \times \mathbb{R}^{m\times m}$, by assumption. From Lemma 4.2 $\int_0^\cdot Z^{t,x}_r dW_r$ is a BMO-martingale. It then follows from Lemma 4.1 and Cauchy Schwartz inequality that for any $p > 1$ there exist a $q > 1$ and $C > 0$ such that:

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{r\in[t,T]}|\Delta\Psi_{i,r}|^{2p}\right]+\mathbb{E}\left[\left(\int_t^T|\Delta\Gamma_{i,r}|^2dr\right)^p\right]\\
&\le C\mathbb{E}\left[|\nabla_x g^1(X_T^{t,x})-\nabla_x g^2(X_T^{t,x})|^{4pq}\right]^{1/(2q)}\mathbb{E}\left[|\Phi_{i,T}^{t,x}|^{4pq}\right]^{1/(2q)}\\
&=0.
\end{aligned}
$$

The first factor vanishes since $\nabla_x g^1(x)-\nabla_x g^2(x)=0$ for almost all $x\in\mathbb{R}^m$ and since $X_T^{t,x}$ has a density from Theorem 3.3. The second factor is finite since $\Phi_{i,T}^{t,x}$ has finite moments. Hence for all $s\in[t,T]$ and $i=1,\dots,m$, $\Psi_{i,s}^{t,x}:=\Psi_{i,s}^1=\Psi_{i,s}^2$ and $\Gamma_{i,s}^{t,x}:=\Gamma_{i,s}^1=\Gamma_{i,s}^2$.

To be able to conclude that the solution (Φ,Ψ,Γ) is unique in $\mathcal{S}^\infty(\mathbb{R}^{m\times m})\times\mathcal{S}^\infty(\mathbb{R}^{1\times m})\times\mathcal{H}^\infty(\mathbb{R}^{d\times m})$ and hence well defined, the condition (2.13) of Theorem 2.2 must be checked for $p>1$ and every component. First,

$$
\mathbb{E}\left[|\nabla_x g(X_T^{t,x})\Phi_{i,T}^{t,x}|^p\right]<\infty,\quad i=1,\dots,m,
$$

since g is Lipschitz continuous and hence has bounded partial derivatives and $\Phi_{i,T}^{t,x}$ has finite moments. Denote the generator of ∇_iFBSDE(b,σ,g,f), by $\overline{f}_i:\Omega\times[0,T]\times\mathbb{R}^{d\times m}\to\mathbb{R}$. It can be identified from (4.1). It satisfies

$$
\overline{f}_i(r,0)=\nabla_x f(r,X_r^{t,x},Z_r^{t,x})\Phi_{i,r}^{t,x}.
$$

Using Cauchy Schwartz inequality, (4.6) and (4.7) the second term of (2.13) can be estimated by

$$
\begin{aligned}
\mathbb{E}\left[\left(\int_0^T|\overline{f}_i(r,0)|dr\right)^p\right]&=\mathbb{E}\left[\left(\int_0^T|\nabla_x f(r,X_r^{t,x},Z_r^{t,x})\Phi_{i,r}^{t,x}|dr\right)^p\right]\\
&\le\mathbb{E}\left[\left(\int_0^T|\nabla_x f(r,X_r^{t,x},Z_r^{t,x})|^2dr\right)^p\right]^{1/2}\mathbb{E}\left[\left(\int_0^T|\Phi_{i,r}^{t,x}|^2dr\right)^p\right]^{1/2}\\
&\le\mathbb{E}\left[\left(\int_0^T|C(1+|Z_r^{t,x}|)|^2dr\right)^p\right]^{1/2}\mathbb{E}\left[T\sup_{r\in[t,T]}|\Phi_{i,r}^{t,x}|^{2p}\right]^{1/2}\\
&<\infty,\qquad i=1,\dots,m.
\end{aligned}
$$

The finiteness of the first factor follows since $Z^{t,x}\in\mathcal{H}^\infty(\mathbb{R}^d)$ from Proposition 4.3. The finiteness of the second factor follows from Theorem 4.7. It holds for any $p>1$ and hence $(\Psi^{t,x},\Gamma^{t,x})$ is unique in $\mathcal{S}^\infty(\mathbb{R}^{1\times m})\times\mathcal{H}^\infty(\mathbb{R}^{d\times m})$ from Theorem 2.2. It can be concluded that $(\Psi_s^{t,x},\Gamma_s^{t,x})$ are well defined processes.

Step 3: Again consider the approximating functions b^n, σ^n and g^n. Define, for $(t,x,n)\in[0,T]\times\mathbb{R}^m\times\mathbb{Z}_+$ and $s\in[t,T]$, an approximating sequence ∇FBSDE(b^n,σ^n,g^n,f) of variational equations with solution $(\Phi^{t,x,n},\Psi^{t,x,n},\Gamma^{t,x,n})$. It will be proved in this step that

$\Phi^{t,x,n} \to \Phi^{t,x}$ in $\mathcal{S}^\infty(\mathbb{R}^{d\times m})$, $\Psi_r^{t,\cdot,n} \to \Psi_r^{t,\cdot}$ in $L_2(\widetilde{\Omega})$ and that $\Gamma^{t,x,n} \to \Gamma^{t,x}$ in $\mathcal{H}^\infty(\mathbb{R}^{d\times m})$ as $n \to \infty$.

It has been proved [19] that, $\forall (t,x) \in [0,T] \times \mathbb{R}^m$,

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{s\in[t,T]} |\Phi_s^{t,x,n} - \Phi_s^{t,x}|^{2p}\right] = 0 \tag{4.9}$$

for $p = 1$. In this proof $p \geq 1$ is needed. The generalization of the proof [19] is a matter of notation and will not be presented.

Let $(\Phi_i^{t,x,n}, \Psi_i^{t,x,n}, \Gamma_i^{t,x,n})$ be the solutions to ∇_iFBSDE(b^n, σ^n, g^n, f), $i = 1, \ldots, m$. Denote the difference by

$$\begin{cases} \Delta\Psi_{i,s}^{t,x,n} & := \Psi_{i,s}^{t,x,n} - \Psi_{i,s}^{t,x}, \\ \Delta\Gamma_{i,s}^{t,x,n} & := \Gamma_{i,s}^{t,x,n} - \Gamma_{i,s}^{t,x}, \\ \Delta\xi_i^{t,x,n} & := \nabla_x g^n(X_T^{t,x,n})\Phi_{i,T}^{t,x,n} - \nabla_x g(X_T^{t,x})\Phi_{i,T}^{t,x} \end{cases}$$

The process $(\Delta\Psi_{i,s}^{t,x,n}, \Delta\Gamma_{i,s}^{t,x,n})$ satisfies $BSDE(\Delta\xi_i^{t,x,n}, \widehat{f}_i^n)$ for

$$\begin{aligned} \widehat{f}_i^n(r,v) \;:=\; & \nabla_x f(r, X_r^{t,x,n}, Z_r^{t,x,n})\Phi_{i,r}^{t,x,n} - \nabla_x f(r, X_r^{t,x}, Z_r^{t,x})\Phi_{i,r}^{t,x} \\ & +(\nabla_z f(r, X_r^{t,x,n}, Z_r^{t,x,n}) - \nabla_z f(r, X_r^{t,x}, Z_r^{t,x}))\Gamma_{i,r}^{t,x} \\ & +\nabla_z f(r, X_r^{t,x,n}, Z_r^{t,x,n})v. \end{aligned}$$

The generator $\widehat{f}_i^n$ satisfies a random Lipschitz condition

$$\begin{aligned} |\widehat{f}_i^n(t,v) - \widehat{f}_i^n(t,\widehat{v})| &= \nabla_z f(r, X_r^{t,x,n}, Z_r^{t,x,n})|v - \widehat{v}| \\ &\leq C(1 + |Z_r^{t,x,n}|)|v - \widehat{v}| \end{aligned}$$

by assumption. Recall from Lemma 4.2 that $\int_t^\cdot Z_r^{t,x,n} dB_r$ is a BMO-martingale and that its BMO-norms are uniformly bounded from step 1. Hence moment estimate Lemma 4.1 can be applied with uniform bounds for the constants C and q together with Lemma 4.8. For any $p > 1$ there exist constants $q > 1$ and $C > 0$ such that, $\forall n \in \mathbb{Z}_+$:

$$\begin{aligned} \mathbb{E}\left[\sup_{r\in[t,T]} |\Delta\Psi_{i,r}^{t,x,n}|^{2p}\right] &+ \mathbb{E}\left[\left(\int_t^T |\Delta\Gamma_{i,r}^{t,x,n}|^2 dr\right)^p\right] \\ &\leq C\left(\mathbb{E}\left[|\Delta\xi_i^{t,x,n}|^{2pq}\right] + I_i^{t,x,n} + J_i^{t,x,n}\right)^{1/q}, \end{aligned} \tag{4.10}$$

where

$$I_i^{t,x,n} := \mathbb{E}\left[\left(\int_t^T |(\nabla_z f(r, X_r^{t,x,n}, Z_r^{t,x,n}) - \nabla_z f(r, X_r^{t,x}, Z_r^{t,x}))\Gamma_{i,r}^{t,x}|dr\right)^{2pq}\right]$$

$$J_i^{t,x,n} := \mathbb{E}\left[\left(\int_t^T |\nabla_x f(r, X_r^{t,x,n}, Z_r^{t,x,n})\Phi_{i,r}^{t,x,n} - \nabla_x f(r, X_r^{t,x}, Z_r^{t,x})\Phi_{i,r}^{t,x}|dr\right)^{2pq}\right].$$

First, by Lemma 4.8 and Cauchy Schwartz inequality

$$\begin{aligned}
\mathbb{E}\left[|\xi_i^{t,x,n}|^{2pq}\right] &= \mathbb{E}\left[|\nabla_x g^n(X_T^{t,x,n})\Phi_{i,T}^{t,x,n} - \nabla_x g(X_T^{t,x})\Phi_{i,T}^{t,x}|^{2pq}\right] \\
&\leq C\left(\mathbb{E}\left[|\nabla_x g^n(X_T^{t,x,n})|^{4pq}\right]^{1/2} \mathbb{E}\left[|\Phi_{i,T}^{t,x,n} - \Phi_{i,T}^{t,x}|^{4pq}\right]^{1/2}\right. \\
&\quad + \mathbb{E}\left[|\nabla_x g^n(X_T^{t,x,n}) - \nabla_x g^n(X_T^{t,x})|^{4pq}\right]^{1/2} \mathbb{E}\left[|\Phi_{i,T}^{t,x}|^{4pq}\right]^{1/2} \\
&\quad \left. + \mathbb{E}\left[|\nabla_x g^n(X_T^{t,x}) - \nabla_x g(X_T^{t,x})|^{4pq}\right]^{1/2} \mathbb{E}\left[|\Phi_{i,T}^{t,x}|^{4pq}\right]^{1/2}\right) \\
&\to 0 \text{ as } n \to \infty.
\end{aligned}$$

The dominated convergence theorem applies since $\nabla_x g$ and $\nabla_x g^n$ are bounded by the Lipschitz constants of g and g^n, $\forall n \in \mathbb{Z}_+$, respectively and $\Phi_{i,T}^{t,x,n}$ and $\Phi_{i,T}^{t,x}$ have finite moments, $i = 1, \ldots, m$. The convergence to zero follows since $X^{t,x,n} \to X^{t,x}$ in $\mathcal{S}^\infty(\mathbb{R}^m)$, $\Phi_i^{t,x,n} \to \Phi_i^{t,x}$ in $\mathcal{S}^\infty(\mathbb{R}^{m\times m})$ from step 1 and $\nabla_x g^n \to \nabla_x g$ for almost all $x \in \mathbb{R}^m$. The final value $X_T^{t,x}$ has a density by Theorem 3.3 and will hence not attain its values at undefined point of $\nabla_x g$.

Next, lets estimate $I_i^{t,x,n}$ by Lemma 4.8 and Cauchy Schwartz inequality:

$$\begin{aligned}
I_i^{t,x,n} &\leq C\mathbb{E}\left[\left(\int_t^T |\Gamma_{i,r}^{t,x}|^2 dr\right)^{2pq}\right]^{1/2} \\
&\quad \times \left(\mathbb{E}\left[\left(\int_t^T |\nabla_x f(r, X_r^{t,x,n}, Z_r^{t,x,n}) - \nabla_x f(r, X_r^{t,x,n}, Z_r^{t,x})|^2 dr\right)^{2pq}\right]\right. \\
&\quad \left. + \mathbb{E}\left[\left(\int_t^T |\nabla_x f(r, X_r^{t,x,n}, Z_r^{t,x}) - \nabla_x f(r, X_r^{t,x}, Z_r^{t,x})|^2 dr\right)^{2pq}\right]\right)^{1/2} \\
&\to 0 \text{ as } n \to \infty.
\end{aligned}$$

The first factor is finite since $\Gamma_i^{t,x} \in \mathcal{H}^\infty(\mathbb{R}^d)$ from step 2. The second factor is by using (4.6) bounded by

$$C\mathbb{E}\left[\left(\int_t^T (1 + |Z_r^{t,x}| + |Z_r^{t,x,n}|)^2 dr\right)^{2pq}\right] < \infty.$$

The finiteness follows since $Z^{t,x}, Z^{t,x,n} \in \mathcal{H}^\infty(\mathbb{R}^d)$ from Proposition 4.3. Dominated convergence, the continuity of $\nabla_x f$ in x and z and the convergence $X^{t,x,n} \to X^{t,x}$ in $\mathcal{S}^\infty(\mathbb{R}^m)$ and $Z^{t,x,n} \to Z^{t,x}$ in $\mathcal{H}^\infty(\mathbb{R}^d)$ implies that $\lim_{n\to\infty} I_i^{t,x,n} = 0$.

Finally, $\lim_{n\to\infty} J_i^{t,x,n} = 0$, $i = 1, \ldots, m$, by similar use of Lemma 4.8, Cauchy Schwartz inequality, the assumptions, convergence results and dominated convergence. Hence, the columns of $\Psi^{t,x,n}$ and $\Gamma^{t,x,n}$ converges to the columns of $\Psi^{t,x}$ and $\Gamma^{t,x}$ in $\mathcal{S}^\infty(\mathbb{R})$ and $\mathcal{H}^\infty(\mathbb{R}^d)$ respectively. It follows that $\Psi^{t,x,n} \to \Psi^{t,x}$ in $\mathcal{S}^\infty(\mathbb{R}^m)$ and $\Gamma^{t,x,n} \to \Gamma^{t,x}$ in $\mathcal{H}^\infty(\mathbb{R}^{d\times m})$ as $n \to \infty$.

The function h, defined in section 3.2, is continuous with integral one and hence bounded. Moreover $\Psi_r^{t,x,n}$ and $\Psi_r^{t,x}$ are essentially bounded. It follows from the bounded convergence theorem that

$$\lim_{n\to\infty} \mathbb{E}\left[\int_{\mathbb{R}^m} \sup_{r\in[t,T]} |\Psi_r^{t,x,n} - \Psi_r^{t,x}|^{2p} h(x)dx\right] = 0$$

and in particular $\Psi_s^{t,\cdot,n} \to \Psi_s^{t,\cdot}$ in $L_{2p}(\widetilde{\Omega}) \subset L_2(\widetilde{\Omega})$, $s \in [t,T]$.

Step 4: Finally, lets put the results from step 1 and 3 together with Theorem 4.5:

$$\begin{cases} Y_s^{t,\cdot,n} & \to & Y_s^{t,\cdot} & \in L_2(\widetilde{\Omega}) & \forall s \in [t,T] \\ \Psi_s^{t,\cdot,n} & \to & \Psi_s^{t,\cdot} & \in L_2(\widetilde{\Omega}) & \forall s \in [t,T] \\ \frac{\partial}{\partial x_i} Y_s^{t,x,n} & = & \Psi_{i,s}^{t,x,n} & 1 \le i \le m & \forall s \in [t,T], \forall x \in \mathbb{R}^m. \end{cases}$$

Lemma 4.4 states that $x \mapsto Y_s^{t,x}$ is Lipschitz continuous, which in turn implies the weaker condition of absolute continuity of $\varepsilon \mapsto Y_s^{t,x+\varepsilon e_i}, 1 \le i \le m$. Hence $Y_s^{t,\cdot} \in (\cap_{i=1}^m \widetilde{D}_i)$ and $\nabla_i Y_s^{t,x}$ is well defined. The convergence then holds with respect to the Dirichlet-$\widetilde{\mathbf{d}}$ norm

$$\|\cdot\|_{\widetilde{\mathbf{d}}} = \left[\|\cdot\|_{L_2(\widetilde{\Omega})}^2 + \sum_{i=1}^m \|\nabla_i(\cdot)\|_{L_2(\widetilde{\Omega})}^2\right]^{\frac{1}{2}}.$$

Hence $Y_s^{t,\cdot} \in \widetilde{\mathbf{d}}$. Proposition 3.2 tells that $Y_s^{t,\cdot} \in \widetilde{\mathbf{d}} \implies Y_s^{t,x} \in \mathbf{d} \subset H_{\text{loc}}^1(\mathbb{R}^m)$. Hence (i) is proved. (ii) follows immediately.

□

Corollary 4.10. *Assume (A). Then for $u(t,x) := Y_t^{t,x}$ it holds that $u(t,\cdot) \in H_{loc}^1(\mathbb{R}^m)$ and*

$$Z_s^{t,x} = \nabla_x u(s, X_s^{t,x})\sigma(s, X_s^{t,x}) \tag{4.11}$$

for Lebesgue a.e. $s \in [t,T]$, $\mathbb{P}$-almost surely, where $\nabla_x u$ is the weak gradient of u.

Proof. First $u(t,\cdot) := Y_t^{t,\cdot} \in \mathbf{d} \subset H_{\text{loc}}^1(\mathbb{R}^m), \forall t \in [0,T]$ from Theorem 4.9. The corollary holds from Theorem 4.6 under assumption (A) and the additional requirement that the terminal function g is twice continuously differentiable and b and σ are continuously differentiable in x. The approximating coefficients in the proof of the previous theorem satisfies these conditions. Hence, for $n \in \mathbb{Z}_+$, Lebesgue a.a. $s \in [t,T]$ and $\mathbb{P}$-a.s.

$$Z_s^{t,x,n} = \nabla_x u^n(s, X_s^{t,x,n})\sigma^n(s, X_s^{t,x,n}).$$

Since $u(t,x) := Y_t^{t,x}$ and $u^n(t,x) := Y_t^{t,x,n}$ it holds, from results in the proof of the previous theorem, that $u^n(t,x) \to u(t,x)$ and $\nabla_x u^n(t,x) \to \nabla_x u(t,x)$ in $L_2(\widetilde{\Omega})$. Also, $X_s^{t,x,n} \to X_s^{t,x}$ and $\sigma^n \to \sigma$ as $n \to \infty$. Hence,

$$\nabla_x u^n(s, X_s^{t,x,n})\sigma^n(s, X_s^{t,x,n}) \to \nabla_x u(s, X_s^{t,x})\sigma(s, X_s^{t,x}) \text{ as } n \to \infty.$$

Moreover $Z^{t,x,n} \to Z^{t,x}$ in $\mathcal{H}^\infty(\mathbb{R}^d)$ as $n \to \infty$. Hence the result holds in the limit. It is easy to check, by using the Lipschitz condition on u in x and the linear growth condition on σ in x, that the right hand side of (4.11) is in $\mathcal{H}^\infty(\mathbb{R}^d)$. $\square$

Chapter 5

Application to insurance and finance: Optimal cross hedging

The market for financial derivatives has exploded the last 25 years. Most often the contracts are written on tradable underlyings such as stocks, grain or oil etc.. In that case the derivatives are priced by creating a replicating portfolio, containing shares of the underlyings such that the value of the portfolio equals that of the derivative. The fair price of the derivative is then the same as the cost to create the replicating portfolio. The purpose of buying the derivative can be either for hedging, speculation or arbitrage purposes [8].

There is an increasing market for derivatives written on indices such as temperature, rain, snow fall or economic loss or other non-tradable indices. The main purpose to write such contracts is for insurance or for insurance companies as an alternative to classical reinsurance. Risks can in such a way be moved to the financial market. Since the indices are non-tradable it's impossible to create replicating portfolios to price derivatives written on them. It is also by the same reason impossible to hedge the risk of the derivative directly. The way to tackle this is by cross hedging, i.e. to find a strongly correlated and tradable asset and use it for hedging. It is of course impossible to hedge all risk since the underlying and the correlated asset are not completely correlated. The market is said to be incomplete.

The approach often taken when pricing and hedging in incomplete markets is that of maximizing the utility of an investment by choosing an optimal hedging strategy. This is also the approach taken here. In the first section the assumptions and market model will be presented and also some examples. In the second the solution approach by solving a FBSDE is presented. So far nothing has been new but rather taken from [2] and [7]. In the fourth section the main results of this thesis is applied. An explicit expression for the optimal cross hedging strategy is derived, in terms of the weak gradient of a FBSDE. The gain of this is that, European put and call options or other derivatives with non-differentiable payoff functions can be written.

5.1 Assumptions and market model

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$, $T > 0$, be a filtered probability space with a d-dimensional Wiener process W_t. $\mathcal{F}_t$ is the natural filtration of W_t completed by the $\mathbb{P}$-null set of Ω. A financial derivative with maturity T and payoff function $F : \mathbb{R}^m \to \mathbb{R}$ is written on a m-dimensional non-tradable risk process X. For $x \in \mathbb{R}^m$ the dynamics of X is given by:

$$X_t = x + \int_0^t b(r, X_r)dr + \int_0^t \sigma(r, X_r)dW_r \quad t \in [0, T].$$

The coefficients $b : [0, T] \times \mathbb{R}^m \to \mathbb{R}^m$ and $\sigma : [0, T] \times \mathbb{R}^m \to \mathbb{R}^{m \times d}$ are assumed to satisfy a global Lipschitz and linear growth condition, i.e. there exist a $C > 0$ such that

$$\begin{cases} |b(t,x) - b(t,\bar{x})| + |\sigma(t,x) - \sigma(t,\bar{x})| & \leq C|x - \bar{x}|, \\ |b(t,x)| + |\sigma(t,x)| & \leq C(1 + |x|). \end{cases}$$

$\forall (t, x, \bar{x}) \in [0, T] \times \mathbb{R}^m \times \mathbb{R}^m$. Moreover $b(t, 0)$ and $\sigma(t, 0)$ are assumed bounded $\forall t \in [0, T]$ and σ is assumed to satisfy the non-degeneracy condition

$$\xi^* \sigma(t,x)\sigma^*(t,x)\xi \geq C|\xi|^2, \quad \forall (t, x, \xi) \in [0, T] \times \mathbb{R}^m \times \mathbb{R}^m \tag{5.1}$$

for some $C > 0$. The random income at time T of maturity of the derivative is $F(X_T)$. F is assumed to be bounded and Lipschitz continuous. The boundedness of F seems unnatural for many derivatives but the bound can be chosen arbitrarily high and doesn't imply any problems in practice.

Conditioned on the information $X_t = x$, for $x \in \mathbb{R}^m$, the non-tradable process will be denoted $X_s^{t,x}$, and satisfy:

$$X_s^{t,x} = x + \int_t^s b(r, X_r^{t,x})dr + \int_t^s \sigma(r, X_r^{t,x})dW_r, \quad s \in [t, T].$$

Since X is non-tradable it's impossible to hedge the risk associated with the derivative directly. Therefore a correlated and tradable asset price process is used to partially hedge the risk. The k-dimensional asset price process is given by:

$$S_t^i = s_0^i + \int_0^t S_r^i(\alpha_i(r, X_r)dr + \beta_i(r, X_r)dW_r), \quad i = 1, \ldots, k.$$

Here α_i and β_i denotes the i:th rows of the functions $\alpha : [0, T] \times \mathbb{R}^m \to \mathbb{R}^k$ and $\beta : [0, T] \times \mathbb{R}^m \to \mathbb{R}^{k \times d}$, $i = 1, \ldots, k$. The coefficient α is assumed to be bounded and β is assumed to satisfy the condition

$$\varepsilon I^{k \times k} \leq \beta(t,x)\beta^*(t,x) \leq K I^{k \times k} \tag{5.2}$$

for some $0 < \varepsilon < K$, $\forall (t, x) \in [0, T] \times \mathbb{R}^m$, where β^* is the transpose of β and $I^{k \times k}$ is the identity matrix in $\mathbb{R}^k$. It implies that $\beta(t,x)\beta^*(t,x)$ is invertible and bounded. Both α and β satisfy a global Lipschitz condition:

$$|\alpha(t,x) - \alpha(t,\widehat{x})| + |\beta(t,x) - \beta(t,\widehat{x})| \leq C|x - \widehat{x}|$$

$\forall (t, x, \widehat{x}) \in [0, T] \times \mathbb{R}^m \times \mathbb{R}^m$, $C > 0$. Moreover α and β are assumed to be continuously differentiable in x. Notice that X and S are driven by the same Wiener process. Their correlation is determined by σ and β.

To rule out arbitrage opportunities the assumption $d \geq k$ must hold, i.e. there must be more sources of uncertainty than number of tradable assets. When all the assumptions above holds assumption **(B)** will be said to be fulfilled.

The following example is taken from [2].

Example 5.1. A company producing kerosene (ke) from crude oil (co) is sensitive against sudden increases in the price of crude oil. It therefore invests in so called crack spreads. They are European options on the difference of the crude oil price and the kerosene price, i.e. derivatives with payoff function $F(X_T^{co}, X_T^{ke}) = [X_T^{co} - X_T^{ke} - K]^+$ where K is the strike price. T is the time to maturity. The market for trading kerosene is not liquid enough to warrant future contracts on it. Therefore some liquid and strongly correlated asset must be used to partially hedge the risk associated with the crack spread. Heating oil (ho) has this property and it is liquid. Here cross hedging applies.

In [2] the following model for the indices is presented:

$$\begin{aligned} dX_t^{ke} &= X_t^{ke}(b_1 dt + \gamma_2 dW_t^1 + \gamma_3 dW_t^2 + \gamma_4 dW_t^3) \\ dX_t^{co} &= X_t^{co}(b_2 dt + \gamma_1 dW_t^1) \\ dS_t^{ho} &= S_t^{ho}(b_3 dt + \beta_1 dW_t^1 + \beta_2 dW_t^2) \end{aligned}$$

for $b_1, b_2, b_3 \in \mathbb{R}, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \beta_1, \beta_2 \in \mathbb{R} \setminus \{0\}$ and $t \in [0, T]$. With the result of chapter 4 an explicit expression of the optimal cross hedging strategy can be obtained. This was not possible before since F is not continuously differentiable. See the last section below.

An *investment strategy* is a k-dimensional predictable process λ that satisfies $\int_0^t \lambda_r^i \frac{dS_r^i}{S_r^i} < \infty$, $i = 1, \dots, k$. λ_t^i is the value of the portfolio invested in the i:th asset at time $t \in [0, T]$. The total gain from investing according to λ in the time interval $[t, s]$ is $G_s^{\lambda,t} = \sum_{i=1}^k \int_t^s \lambda_r^i \frac{dS_r^i}{S_r^i}$. The gain conditioned on $X_t = x$, for $x \in \mathbb{R}^m$, is denoted $G_s^{\lambda,t,x}$ and is given by:

$$G_s^{\lambda,t,x} = \sum_{i=1}^k \int_t^s \lambda_s^i [\alpha_i(r, X_r^{t,x}) dr + \beta_i(r, X_r^{t,x}) dW_r].$$

Let $\mathcal{A}^{t,x}$ denote the space of all strategies λ satisfying $\mathbb{E}[\int_t^T |\lambda_r \beta(r, X_r^{t,x})|^2 dr] < \infty$ and such that the family $\{e^{-\eta G_\tau^{\lambda,t,x}} : \tau \in [t, T] \text{ is a stopping time}\}$ is uniformly integrable, $\eta > 0$. Strategies in $\mathcal{A}^{t,x}$ are called *admissible*.

It seems natural to seek for an optimal strategy that in some sense maximizes the gain of the investment. The approach here is to maximize the expected exponential utility. The exponential utility function is given by:

$$U(y) = -e^{-\eta y}.$$

The risk aversion coefficient $\eta > 0$, $y \in \mathbb{R}$. The maximal expected utility, conditioned on $X_t = x$, for risk level $x \in \mathbb{R}^m$ at time $t \in [0, T]$ and wealth $v \in \mathbb{R}$, of an investment without the derivative is

$$V^0(t, x, v) := \sup_{\lambda \in \mathcal{A}^{t,x}} \mathbb{E}\left[U(v + G_T^{\lambda,t,x})\right].$$

In terms of stochastic control theory V^0 is called the *value function* of a stochastic control problem. It has been shown in [7] that there exist an almost surely unique *optimal strategy* $\pi \in \mathcal{A}^{t,x}$ such that

$$V^0(t, x, v) = \mathbb{E}\left[U(v + G_T^{\pi,t,x})\right].$$

Remark 5.2. The exponential utility function punishes losses strongly and rewards gains moderately. Hence the strategy π is the strategy that minimizes risk for losses. π does of course depend on the risk aversion coefficient η. The smaller $\eta > 0$ is the more rewarded are high gains but more likely are losses. U is a concave utility function and all such are risk averse. On the other side convex utility functions are risk-seeking, i.e. very high gains are preferred even though they are unlikely to occur. A linear utility function maximizes the expected value and the concept of utility is gone.

If the portfolio contains the derivative $F(X_T)$ the conditional maximal expected utility and value function becomes

$$V^F(t,x,v) := \sup_{\lambda \in \mathcal{A}^{t,x}} \mathbb{E}\left[U(v + G_T^{\lambda,t,x} + F(X_T^{t,x}))\right].$$

Also in this case there exist an almost surely unique investment strategy $\widehat{\pi}$ that satisfies

$$V^F(t,x,v) = \mathbb{E}\left[U(v + G_T^{\widehat{\pi},t,x} + F(X_T^{t,x}))\right].$$

The difference of the two strategies

$$\Delta = \widehat{\pi} - \pi$$

is called the *derivative hedge* and is used to hedge the derivative. In a later subsection explicit expression for the derivative hedge will be derived via the distributional gradient of a quadratic FBSDE. It is a generalization of the classical Δ-hedge in the Black-Scholes model. In the case of a complete market, i.e. when $d = k$ and $S = R$, derivative hedge coincides with Black-Scholes Δ-hedge. Next, how shall the derivative be priced? This is solved by calculating the so called *indifference price* $p(t,x)$ at time t conditioned on $X_t = x$ and wealth v, given by:

$$V^F(t,x,v - p(t,x)) = V^0(t,x,v).$$

It is the price that makes the buyer of the derivative indifferent, in a utility point of view, to wether she should buy the derivative or not. It will later be seen that the derivative hedge can be expressed in terms of the distributional gradient of the indifference price $p(t,x)$. The mathematical problem to find the optimal investment strategy is an optimal stochastic control problem. It is often tackled by solving the so called Hamilton-Jacobi-Bellman PDE. The approach here is taken from [2] and [7] and uses FBSDEs.

5.2 Solution to the optimal cross hedging problem via a FBSDE

In this section a FBSDE will be used to find explicit expressions for the indifference price and optimal cross hedging strategies presented above. The results have been proved in [2] and [7]. Fix $(t,x) \in [0,T] \times \mathbb{R}^m$. Assumption (5.2) implies that $\beta(t,x)\beta^*(t,x)$ is invertible hence the mapping $\beta(t,x) : \mathbb{R}^k \to \mathbb{R}^d$ is one-to-one. Recall that a trading strategy λ_t is a k-dimensional process, corresponding to the value of a portfolio invested in the i:th asset at time t. In the solution approach here the d-dimensional image strategies given by $\lambda_t\beta(t,x)$ will be considered instead. Let

$$C(t,x) = \left\{ s\beta(t,x) : s \in \mathbb{R}^k \right\}.$$

be the constraint set for the image strategies. The matrix $\beta(t,x)$ is not necessarily onto. Hence $C(t,x)$ is in fact a constraint set. It is closed and convex. Let

$$\vartheta(t,x) = \beta^*(t,x)(\beta(t,x)\beta^*(t,x))^{-1}\alpha(t,x).$$

The process ϑ is bounded since α and β are bounded and $\beta\beta^* \leq KI^{k\times k}$ from assumption 5.2. Let $\mathrm{dist}(z,C) = \min\{|z-u| : u \in C\}$ be the distance of a vector $z \in \mathbb{R}^d$ to the closed and convex set C. Define the generator of a FBSDE by:

$$f : [0,T] \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}, (t,x,z) \mapsto z\vartheta(t,x) + \frac{1}{2\eta}|\vartheta(t,x)|^2 - \frac{\eta}{2}\mathrm{dist}^2(z + \frac{1}{\eta}\vartheta(t,x)), C(t,x)).$$

There exist a unique solution to the FBSDE

$$\widehat{Y}_s^{t,x} = F(X_T^{t,x}) - \int_s^T f(r, X_r^{t,x}, \widehat{Z}_r^{t,x})dr - \int_s^T \widehat{Z}_r^{t,x} dW_r,$$

$s \in [t,T]$, since f has quadratic growth in z. The value function, defined in the previous section, for an investment with the derivative is given by:

$$V^F(t,x,v) = -e^{-\eta(v - \widehat{Y}_t^{t,x})}.$$

Let $\prod_{C(t,x)}(z)$ denote the orthogonal projection of $z \in \mathbb{R}^d$ onto the subspace $C(t,x)$. Conditioned on $X_t = x$ the optimal cross hedging strategy is given by:

$$\widehat{\pi}_s\beta(s, R_s^{t,x}) = \prod_{C(t,x)} [\widehat{Z}_s^{t,x} + \frac{1}{\eta}\vartheta(s, X_s^{t,x})],$$

$s \in [t,T]$. Analogously, an investment without the derivative give rise to the FBSDE:

$$Y_s^{t,x} = -\int_s^T f(r, X_r^{t,x}, Z_r^{t,x})dr - \int_s^T Z_r^{t,x} dW_r,$$

with value function

$$V^0(t,x,v) = -e^{-\eta(v - Y_t^{t,x})}$$

and optimal strategy:

$$\pi_s\beta(s, R_s^{t,x}) = \prod_{C(t,x)} [Z_s^{t,x} + \frac{1}{\eta}\vartheta(s, X_s^{t,x})],$$

$s \in [t,T]$. The projection operator is linear hence the derivative hedge is given by

$$\Delta_s\beta(s, X_s^{t,x}) = \prod_{C_(t,x)} [\widehat{Z}_s^{t,x} - Z_s^{t,x}].$$

Recall the definition of the indifference price $p(t,x)$ and notice that,

$$V^F(t,x,v-p(t,x)) = -e^{-\eta(v-p(t,x)-\widehat{Y}_t^{t,x})} = -e^{-\eta(v-Y_t^{t,x})} = V^0(t,x,v)$$

which in turn implies that

$$p(t,x) = Y_t^{t,x} - \widehat{Y}_t^{t,x}.$$

5.3 Explicit hedging strategy using the weak price gradient

In this section the results on weak differentiability of FBSDE from Chapter 4 will be applied to the optimal cross hedging problem. The major advantage of the new results is that the payoff function no longer must be continuously differentiable. This implies that put and call options can be written on the underlying and explicit hedging strategies derived via the weak gradient of the indifference price. Another is that the coefficients of the non-tradable asset process no longer need to be differentiable but only Lipschitz continuous with linear growth.

Theorem 5.3. *Under assumption (B) the functions $\widehat{u}(t,x) := \widehat{Y}_t^{t,x}$ and $u(t,x) := Y_t^{t,x}$ are weakly differentiable with respect to x.*

Proof. Assumption (B) implies assumption (A) of chapter (4). The trickier parts of the proof concerns the generator and was made in [2]. Theorem 4.9(i) applies since (A) holds. □

Since $p(t,x) = Y_t^{t,x} - \widehat{Y}_t^{t,x}$ the following corollary holds.

Corollary 5.4. *Under assumption (B) the indifference price $p(t,x)$ is weakly differentiable with respect to x.*

Theorem 5.5. *Assume (B). Then the derivative hedge, for risk level $x \in \mathbb{R}^m$ at time $t \in [0,T]$, is given by*

$$\Delta(t,x) = -\prod_{C(t,x)}[\nabla_x p(t,x)\sigma(t,x)]\beta^*(t,x)(\beta(t,x)\beta^*(t,x))^{-1}$$

for $(t,x) \in [0,T] \times \mathbb{R}^m$, where $\nabla_x p$ is the price gradient considered in the weak sense and $\prod_{C(t,x)}(z)$ is the orthogonal projection of $z \in \mathbb{R}^d$ onto $C(t,x) := \{s\beta(t,x) : s \in \mathbb{R}^k\}$.

Proof. Recall that

$$\Delta_s\beta(s,X_s^{t,x}) = \prod_{C(t,x)}[\widehat{Z}_s^{t,x} - Z_s^{t,x}],$$

which implies

$$\Delta(t,x) = \prod_{C(t,x)}[\widehat{Z}_t^{t,x} - Z_t^{t,x}]\beta^*(t,x)(\beta(t,x)\beta^*(t,x))^{-1}.$$

Corollary 4.10 implies that $\widehat{Z}_t^{t,x} = \nabla_x\widehat{u}(t,x)\sigma(t,x)$ and $Z_t^{t,x} = \nabla_x u(t,x)\sigma(t,x)$ in the distributional sense. Hence,

$$\widehat{Z}_t^{t,x} - Z_t^{t,x} = (\nabla_x\widehat{u}(t,x) - \nabla_x u(t,x))\sigma(t,x) = \nabla_x(\underbrace{\widehat{Y}_t^{t,x} - Y_t^{t,x}}_{=-p(t,x)})\sigma(t,x)$$

and the result follows. □

Chapter 6

Conclusion and discussion

The results of this thesis has made it theoretically possible to compute the derivative hedge in the cross hedging problem, when the payoff function is Lipschitz continuous and bounded and under the restriction of a non-degenerate risk process. Also, the differentiability assumption on the coefficients of the SDE, for the underlying, has been relaxed. The price for this was that the strategies are expressed in term of the weak gradient of the solution to a backward stochastic differential equation instead of a classical gradient. What this means numerically must be further explored. The boundedness of the payoff function is not a problem in practice since the bound can be set arbitrarily high.

The FBSDE approach has the advantage that it can be used in multi dimensions, i.e. both the tradable and non-tradable assets can theoretically be in any finite number of dimensions. Numerically it is though a more complicated approach. An alternative and classical approach is to solve the Hamilton-Jacobi-Bellman (HJB) PDE. It is the main tool in optimal stochastic control theory. In [1] an explicit solution to the HJB-PDE was derived for the cross hedging problem when both the tradable and the non-tradable asset were of dimension one. After manipulations they showed that it could be solved via Feynman-Kac's formula and hence by simulations of SDEs. The approach holds when the underlying is either non-degenerate or a geometric Brownian motion. Numerically it is about simulation of SDEs, but it is limited to one dimension. Other less recent methods is mostly for pricing and does not give optimal hedging strategies in a dynamic way, if at all. For the financial results to be useful suitable numerics must be used to compute the solution to the variational equations (4.1).

Now, to the mathematical part of this thesis, chapter 4. Are the assumptions contained in (A) of chapter 4 necessary? The differentiability assumption on the generator f is to our judgment necessary. When trying to relax it, a joint non-degeneracy condition on the process (X, Z) would be needed to conclude that

$$\mathbb{E}\left[\int_0^T |\nabla_z f^1(r, X_r, Z_r) - \nabla_z f^2(r, X_r, Z_r)|^2 dr\right] = 0 \tag{6.1}$$

where $\nabla_z f^1$ and $\nabla_z f^2$ only differs on a null set. But Z_t is determined by $v(t, X_t)$, where v is a deterministic function. Hence (X, Z) will be limited to take values in a null set of $\mathbb{R}^{m+d}$, i.e., along a curve. The non-degeneracy condition of the forward SDE is probably not necessary. Great effort has been made by the author to relax it. To get around the degeneracy problem, a FBSDE with solution $(\widetilde{X}, \widetilde{Y}, \widetilde{Z})$, having random initial value with a density has been used. In such way corresponding step 1-3 of Theorem 4.9 has been proved for the new equation. For FBSDEs with Lipschitz continuous generator it is easy by using results in [5] to conclude that the solutions to the corresponding approximating variational equations with solutions

$(\widetilde{\Phi}^n, \widetilde{\Psi}^n, \widetilde{\Gamma}^n)$ are the weak gradients of $(\widetilde{X}^n, \widetilde{Y}^n, \widetilde{Z}^n)$. For quadratic generators the framework of [5], [4] or even [18] is not covered. Such a proof would involve Dirichlet forms or Malliavin calculus, if it is possible. One other possibility to expand the results would be to let the generator f depend on Y. Then the result could be used for regularity and representation results for PDEs, as in [19]. That would result in even lengthier proofs. Since the generator f in the cross hedging problem does not depend on Y the theory here is restricted to that case.

In this thesis the weak differentiability has been proved by weakening results on classical differentiability. This is in some sense unnatural since classical differentiability is something stronger and feels unnecessary to prove in order to later weaken it. An alternative approach would be to start with nothing and prove weak differentiability directly. Such an approach would contain Malliavin calculus instead of the closely related theory of Dirichlet forms and Dirichlet spaces use here. The Malliavin derivative is the infinite dimensional distributional derivative in ω for some probability space, in this case the Wiener space. In fact, the Malliavin derivative $\{D_t Y_t; 0 \leq t \leq T\}$ is a version of $\{Z_t; 0 \leq t \leq T\}$ [12]. By the Malliavin chain rule the connection between Y and Z from chapter 4 is obtained [2]. Further X is Malliavin differentiable and then $u(X)$ is Malliavin differentiable if u is Lipschitz continuous [18]. Hence if $u(t,x) := Y_t^{t,x}$ is Lipschitz continuous $u(s, X_s^{t,x})$ is Malliavin differentiable, which in turn implies that u is differentiable in the weak sense from the Malliavin chain rule. If it is possible to obtain the representation of the weak gradients, in terms of the solution to a variational equation, as in Theorem 4.9(ii) using Malliavin calculus and a direct approach remains to be explored.

Bibliography

[1] Ankirchner, S., Imkeller, P. and Popier, A. (2008). Optimal cross hedging of insurance derivatives. *Stochastic Analysis and Applications* **26** 679–709.

[2] Ankirchner, S., Imkeller, P. and Dos Reis, G. (2007). Pricing and hedging of derivatives based on non-tradable underlyings. arXiv:0712.3746v1 [math.PR].

[3] Bismut J.M. (1978). An introductory approach to duality in optimal stochastic control. *SIAM review* **20** 62–78.

[4] Bouleau, N. and Hirsch, F. (1991) *Dirichlet forms and analysis on Wiener space.* De Gruyter Studies in Mathematics **14**.

[5] Bouleau, N. and Hirsch, F. (1989) On the derivability with respect to the initial data of the solution of a stochastic differential equation with Lipschitz coefficients. Lecture Notes in Mathematics **1393** 39–57.

[6] Briand, P. and Confortola, F. (2007) BSDEs with stochastic Lipschitz condition and quadratic PDEs in Hilbert spaces. *Stochastic Processes and Their Applications* **118** 818–838.

[7] Hu, Y., Imkeller, P. and Müller, M. (2005). Utility maximization in incomplete markets. *The Annals of Applied Probability* **15** 1691–1712.

[8] Hull, J.C. (2006). *Options, Futures and Other Derivatives*, sixth edition. Prentice Hall.

[9] Hörmander, L. (1983). *The Analysis of Linear Partial Differential Operators 1.* Springer.

[10] Imkeller, P. (2008). *Malliavin's Calculus and Applications in Stochastic Control and Finance.* Non-published lecture notes for a course in Warschaw the spring 2008.

[11] Karoui, N.E. and Mazliak, L. (1997). Backward Stochastic Differential Equations. Longman.

[12] Karoui, N.E., Peng, S. and Quenez, M.C. (1997). Backward stochastic differential equations in finance. *Mathematical Finance* **7** 1–71.

[13] Kobylanski, M. (2000). Backward stochastic differential equations and partial differential equations with quadratic growth. *Annals of Probability* **28** 558–602.

[14] Krylov, N.V. (1980). *Controlled diffusion processes.* Springer.

[15] Larsson, S. and Thome, V. (2005). *Partial Differentiable Equations with Numerical Methods.* Springer.

[16] Ma, J. and Yong, J. (1999). *Forward-Backward Stochastic Differential Eqautions and Their Applications*. Springer.

[17] Morlais, M.A. (2008). Quadratic BSDEs driven by a continuous martingale and applications to the utility maximization problem. arXiv:math/0610749v3 [math.PR].

[18] Nualart, D. (2006). *The Malliavin Calculus and Related Topics, second edition.* Springer.

[19] N'Zi, M., Ouknine, Y. and Sulem, A. (2006). Regularity and representation of viscosity solutions of partial differential equations via backward stochastic differential equations. *Stochastic Processes and Their Applications* **116** 1319–1339.

[20] Øksendal, B. (2003). *Stochastic Differential Equations, an Introduction with Applications, sixth edition.* Springer.

[21] Pardoux, E. and Peng, S. (1990). Adapted solution of a backward stochastic differential equation. *Systems & Control Letters* **14** 55–61.