Approximating solutions of backward doubly stochastic differential equations with measurable coefficients using a time discretization scheme

BY

Cyrus Yeadon

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy in Mathematics of Loughborough University

March 2015

To Lauren, my parents, Uncle Fred, our little darling Alexa and the wonderful staff of the Elizabeth Garrett Anderson Wing at University College Hospital who saved her.

Acknowledgements

I would like to thank my supervisors, Professor Huaizhong Zhao and Dr Chunrong Feng and the EPSRC for the opportunity to continue my education in Mathematics. I would also like to thank my supervisors for gently steering me away me from the many rabbit holes I have found so inviting over the past four years.

I am also grateful for the support that I have received from my family and Lady Vivien Cockroft and would like to acknowledge the free software distributed by the nice people at Debian - you have all made the writing of this thesis much more pleasant.

Abstract

It has been shown that backward doubly stochastic differential equations (BDSDEs) provide a probabilistic representation for a certain class of nonlinear parabolic stochastic partial differential equations (SPDEs). It has also been shown that the solution of a BDSDE with Lipschitz coefficients can be approximated by first discretizing time and then calculating a sequence of conditional expectations. Given fixed points in time and space, this approximation has been shown to converge in mean square.

In this thesis, we investigate the approximation of solutions of BDSDEs with coefficients that are measurable in time and space using a time discretization scheme with a view towards applications to SPDEs. To achieve this, we require the underlying forward diffusion to have smooth coefficients and we consider convergence in a norm which includes a weighted spatial integral. This combination of smoother forward coefficients and weaker norm allows the use of an equivalence of norms result which is key to our approach. We additionally take a brief look at the approximation of solutions of a class of infinite horizon BDSDEs with a view towards approximating stationary solutions of SPDEs.

Whilst we remain agnostic with regards to the implementation of our discretization scheme, our scheme should be amenable to a Monte Carlo simulation based approach. If this is the case, we propose that in addition to being attractive from a performance perspective in higher dimensions, such an approach has a potential advantage when considering measurable coefficients. Specifically, since we only discretize time and effectively rely on simulations of the underlying forward diffusion to explore space, we are potentially less vulnerable to systematically overestimating or underestimating the effects of coefficients with spatial discontinuities than alternative approaches such as finite difference or finite element schemes that do discretize space.

Another advantage of the BDSDE approach is that it is possible to derive an upper bound on the error of our method for a fairly broad class of conditions in a single analysis. Furthermore, our conditions seem more general in some respects than is typically considered in the SPDE literature.

Contents

1 Introduction

As explained in the abstract:

> It has been shown that backward doubly stochastic differential equations (BDSDEs) provide a probabilistic representation for a certain class of nonlinear parabolic stochastic partial differential equations (SPDEs).
>
> In this thesis, we investigate the approximation of solutions of BDSDEs with coefficients that are measurable in time and space using a time discretization scheme with a view towards applications to SPDEs. To achieve this, we require the underlying forward diffusion to have smooth coefficients and we consider convergence in a norm which includes a weighted spatial integral. This combination of smoother forward coefficients and weaker norm allows the use of an equivalence of norms result which is key to our approach. We additionally take a brief look at the approximation of solutions of a class of infinite horizon BDSDEs with a view towards approximating stationary solutions of SPDEs.

The connection between BDSDEs and nonlinear parabolic SPDEs was established in [36] for BDSDEs with smooth coefficients and extended to the measurable coefficient case of this thesis in [5] and [50]. The approximation scheme we define for BDSDEs with measurable coefficients is based upon the approximation schemes of [8] and [49] for BSDEs with Lipschitz coefficients. We note that in [2], an approximation scheme for BDSDEs with Lipschitz coefficients also based upon [8] and [49] is defined. Whilst there is some overlap between the work of this thesis and [2], the two works were developed independently of each other and as our conditions are significantly weaker than those of [2], the overlap is not substantial.

SPDEs of parabolic type have a variety of applications including (see e.g. [12]): chemical reactions, neurophysiology, population genetics, turbulence and geophysical fluid dynamics. As is the case for PDEs, the exact solution of a SPDE is typically a difficult problem which motivates their approximation by numerical methods. As a consequence, the development of new numerical methods for SPDEs with weakened conditions will only help to broaden the applicability of SPDEs.

There are a variety of strategies for the numerical approximation of parabolic SPDEs of which we mention just a few: [1] considers weak convergence of finite difference and finite element schemes for linear SPDEs with L^2 coefficient and additive noise; [30], [18] and [19] consider finite difference schemes for nonlinear SPDEs with continuous coefficients; [21] constructs stochastic Taylor expansions for nonlinear SPDEs with smooth coefficients and additive noise; [32] considers a Monte Carlo scheme based upon the method of characteristics for linear SPDEs with smooth coefficients; [44] considers weak convergence of a finite difference scheme for the stochastic heat equation; [47] considers finite element schemes for nonlinear SPDEs with Lipschitz coefficients; [2] considers the approximation of BDSDEs with Lipschitz coefficients via time discretization which (as is observed in the abstract above) implicitly provides an approximation to a class of nonlinear parabolic SPDEs.

With the exception of [32] and [2], the equations considered in the above are essentially the stochastic heat equation with additional terms. It seems an advantage of probabilistic schemes such as the Monte Carlo scheme of [32] and the BDSDE approach leveraged in this thesis and in [2] that it is easier to consider more general drift and diffusion terms. We note, however, that both [32] and [2] require significantly more regularity on the coefficients of their respective SPDEs than we do.

The structure of this thesis is as follows. In Chapter 2 we provide some background material on standard notation and stochastic analysis. In Chapter 3 we provide a review of the literature on BSDEs, BDSDEs and their numerical approximation. In Chapter 4 we introduce the notation used in this thesis and define the problems that we wish to solve.

The main work of this thesis commences in Chapter 5. The approach taken in this thesis is to first approximate the BDSDEs that we wish to solve with BDSDEs with more regular coefficients. We then define a discretization scheme for these BDSDEs with more regular coefficents. To this end, in Chapter 5 we present the approximation of BDSDEs with measurable coefficients with BDSDEs with Lipschitz coefficents and BDSDEs with smooth coefficents. In Chapter 6 we then derive some results on the regularity of the solutions of BDSDEs with Lipschitz coefficients and BDSDEs with smooth coefficients. In Chapter 7 we define a discretization scheme for BDSDEs with Lipschitz coefficents and determine an upper bound for the error of the scheme using the regularity results of Chapter 6. In Chapter 8 we consider the problem of defining a discretization scheme for infinite horizon BDSDEs with contractive coefficients. Chapter 9 is a discussion of this thesis and potential future problems to consider and finally Chapter A in the appendix is a collection of useful results.

2

Background Material

This chapter contains background material on notation, function spaces and stochastic analysis. Please proceed to Chapter 3 if this material is already familiar. The material for this chapter is taken from [4], [16], [22], [23], [26], [33], [35], [37], [40], [41], [43] and [45]. Further details on the majority of the material can be found in [23], [37] and [40].

General Notation

Let d be a positive integer. For any vector $x \in \mathbb{R}^d$, $|x|$ will denote the standard Euclidean norm of x. For any $d \times d$ matrix A, $\|A\| := \sqrt{\mathrm{Tr} AA^T}$.

Let p and q be positive integers. $C^0(\mathbb{R}^p, \mathbb{R}^q)$ denotes the space of continuous functions $f : \mathbb{R}^p \to \mathbb{R}^q$. For $k \geq 1$, $C^k(\mathbb{R}^p, \mathbb{R}^q)$ consists of all functions in $C^0(\mathbb{R}^p, \mathbb{R}^q)$ whose derivatives of order less than or equal to k are continuous. For $k \geq 1$, $C_b^k(\mathbb{R}^p, \mathbb{R}^q)$ consists of all functions in $C^k(\mathbb{R}^p, \mathbb{R}^q)$ whose derivatives of order less than or equal to k are bounded. Note that this does not imply that the function itself is bounded. For $k \geq 1$, $C_0^k(\mathbb{R}^p, \mathbb{R}^q)$ consists of all functions in $C^k(\mathbb{R}^p, \mathbb{R}^q)$ whose support is a compact subset of $\mathbb{R}^p$.

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, p be a real number with $p \geq 1$ and d a positive integer. Then $L^p((\Omega, \mathcal{F}, \mu); \mathbb{R}^d)$ consists of all $\mathbb{R}^d$ valued Borel-measurable functions such that $\int_\Omega |f|^p d\mu < \infty$. If d_1 and d_2 are positive integers then $L^p(\mathbb{R}^{d_1}; \mathbb{R}^{d_2}) \equiv L^p((\mathbb{R}^{d_1}, \mathcal{B}(\mathbb{R}^{d_1}), l); \mathbb{R}^{d_2})$ where l denotes the Lebesgue measure. For a non-negative function $\rho \in L^1(\mathbb{R}^{d_1}; \mathbb{R})$, the ρ-weighted space $L^p_\rho(\mathbb{R}^{d_1}; \mathbb{R}^{d_2})$ consists of all $\mathbb{R}^{d_2}$-valued Borel-measurable functions such that $\int_{\mathbb{R}^{d_1}} |f(x)|^p \rho(x) dx < \infty$.

Filtrations, Martingales and Brownian Motion

> "A stochastic process is a mathematical model for the occurrence, at each moment after the initial time, of a random phenomenon." [23]

For the remainder of this chapter we will assume as given a complete probability space $(\Omega, \mathcal{F}, P)$.

Definition. *A family of σ-algebras $\{\mathcal{F}_t; t \geq 0\}$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for $s \in [0, t)$ is said to be a filtration of $\mathcal{F}$.*

Definition. *A stochastic process X on $(\Omega, \mathcal{F}, P)$ is a collection of $\mathbb{R}^d$-valued (with $d \geq 1$) random variables $\{X_t; t \geq 0\}$ on $(\Omega, \mathcal{F}, P)$. If $(\Omega, \mathcal{F}, P)$ is equipped with a filtration $\{\mathcal{F}_t; t \geq 0\}$ and X_t is an $\mathcal{F}_t$-measurable random variable for each t, then the process X is said to be adapted to $\{\mathcal{F}_t\}$ and we write $\{X_t, \mathcal{F}_t; t \geq 0\}$.*

Definition. *A real-valued, adapted process $\{M_t, \mathcal{F}_t; t \geq 0\}$ is called a martingale (respectively supermartingale, submartingale) with respect to the filtration $\{\mathcal{F}_t\}$ if for every $0 \leq s \leq t < \infty$,*

1. *$M_t \in L^1((\Omega, \mathcal{F}, P); \mathbb{R})$.*
2. *$E[M_t|\mathcal{F}_s] = M_s$ a.s. (respectively $\leq M_s$, $\geq M_s$).*

Definition. *Let X and Y be stochastic processes defined on $(\Omega, \mathcal{F}, P)$. They are said to be versions or modifications of each other if $X_t = Y_t$ a.s. for each $t \geq 0$. They are said to be indistinguishable if a.s. it holds that $X_t = Y_t$ for all $t \geq 0$.*

Definition. *A stochastic process X is said to be continuous (respectively left-continuous, right-continuous, cadlag) if it a.s. has sample paths which are continuous (respectively left-continuous, right-continuous, cadlag).*

Definition. *A filtration $\{\mathcal{F}_t; t \geq 0\}$ is said to satisfy the usual conditions if*

1. *$\mathcal{F}_0$ contains the P-null sets of $\mathcal{F}$.*
2. *$\mathcal{F}_t = \bigcap_{u>t} \mathcal{F}_u$ for all $t \geq 0$.*

Remark. It is easy to find an example of why the first of the "usual conditions" is desirable. For example, if $\{X_t, \mathcal{F}_t; t \geq 0\}$ is an adapted process, Y is a modification of X and $\mathcal{F}_0$ contains the P-null sets of $\mathcal{F}$ then Y is also adapted to $\{\mathcal{F}_t\}$. The following theorem provides an example of the benefit of additionally assuming the second of the usual conditions.

Theorem. *Suppose that $\{\mathcal{F}_t; t \geq 0\}$ satisfies the usual conditions and let $\{M_t, \mathcal{F}_t; t \geq 0\}$ be a martingale. Then there exists a unique modification of M which is cadlag.*

Definition. *Let X be a stochastic process. The natural filtration of X, denoted $\{\mathcal{F}_t^X; t \geq 0\}$, is defined for each $t \geq 0$ by $\mathcal{F}_t^X := \sigma(X_s; s \in [0,t])$. Obviously, X is adapted to $\{\mathcal{F}_t^X\}$. If the P-null sets of $\mathcal{F}$ are denoted by $\mathcal{N}$ then the augmented filtration of X is defined for each $t \geq 0$ by $\sigma(\mathcal{F}_t^X \cup \mathcal{N})$.*

Definition. *An adapted, continuous process $\{W_t, \mathcal{F}_t; t \geq 0\}$ taking values in $\mathbb{R}^d$ (with $d \geq 1$) is called a d-dimensional $\{\mathcal{F}_t\}$ standard Wiener process or standard Brownian motion if*

1. *$W_0 = 0$ a.s.*
2. *For $s \in [0,t)$, $W_t - W_s$ is independent of $\mathcal{F}_s$.*
3. *For $s \in [0,t)$, $W_t - W_s$ is a Gaussian random variable with mean zero and covariance matrix $(t-s)I_d$, where I_d denotes the $d \times d$ identity matrix.*

Remark. If $\{\mathcal{F}_t; t \geq 0\}$ is taken to be the augmented filtration of the Wiener process W, then W is sometimes called a Wiener process without specifying the filtration. The following theorem shows that it is not difficult to attain the usual conditions of a filtration.

Theorem. *The augmented filtration of the standard Wiener process satisfies the usual conditions.*

Definition. *A random variable $T : \Omega \to [0, \infty]$ is said to be a random time. If in addition there is a filtration $\{\mathcal{F}_t; t \geq 0\}$ such that the event $\{T \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$, then T is said to be a stopping time of $\{\mathcal{F}_t\}$.*

Definition. *Let X be a stochastic process and T be a random time. The random variable X_T is defined on the event $\{T < \infty\}$ by $X_T(\omega) := X_{T(\omega)}(\omega)$.*

Definition. *For each $a \geq 0$ let $\mathcal{S}_a$ denote the stopping times T of $\{\mathcal{F}_t; t \geq 0\}$ such that $T \leq a$ a.s. and let X be a right-continuous stochastic process. Then X is said to be of class DL if the family $(X_T)_{T \in \mathcal{S}_a}$ is uniformly integrable for every $0 < a < \infty$.*

Definition. *An adapted process A is called increasing if*

1. $A_0 = 0$ *a.s.*
2. $t \to A_t$ *is a non-decreasing, right-continuous function a.s.*
3. $E[A_t] < \infty$ *for every* $t \geq 0$.

Theorem (Doob-Meyer Decomposition)**.** *Let $\{\mathcal{F}_t; t \geq 0\}$ satisfy the usual conditions. If X is a right-continuous $\{\mathcal{F}_t\}$-submartingale of class DL, then there exists a right-continuous $\{\mathcal{F}_t\}$-martingale M and an increasing process A adapted to $\{\mathcal{F}_t\}$ such that $X_t = M_t + A_t$ for each $t \geq 0$.*

Definition. *Let X be a right-continuous martingale. M is said to be square-integrable if $E[X_t^2] < \infty$ for every $t \geq 0$.*

Definition. *Let M be a square-integrable martingale with $M_0 = 0$ a.s. The quadratic variation of M is defined to be the process $\langle M \rangle := A$, where A is the increasing process in the Doob-Meyer decomposition of M^2.*

Remark. Let W be a one-dimensional standard Brownian motion. Then W is a square-integrable martingale with $\langle W \rangle_t = t$.

Definition. *Let X and Y be square-integrable martingales with $X_0 = 0$, $Y_0 = 0$ a.s. The cross-variation process of X and Y is defined by*

$$\langle X, Y \rangle_t := \frac{1}{4}(\langle X + Y \rangle_t - \langle X - Y \rangle_t) \ , \qquad t \geq 0.$$

Definition. *Let X be a stochastic process, fix $t > 0$ and let $\Pi = \{t_0, t_1, \ldots, t_n\}$ with $0 = t_0 < t_1 < \ldots < t_n = t$ be a partition of $[0, t]$. The p-th variation for $p > 0$ of X over Π is defined to be*

$$V_t^{(p)}(\Pi) := \sum_{k=1}^{n} |X_{t_k} - X_{t_{k-1}}|^p.$$

Remark. Define the mesh of Π as $\|\Pi\| := \max_{1\leq k\leq n}|t_k - t_{k-1}|$. If $V_t^{(2)}(\Pi)$ converges as $\|\Pi\| \to 0$ (in some sense), then the limit could also be called the quadratic variation of X on $[0,t]$. The following theorem shows that these two definitions of quadratic variation are consistent.

Theorem. *Let M be a square-integrable martingale with $M_0 = 0$ a.s. and let $\Pi_1, \Pi_2, \ldots$ be a sequence of partitions of $[0,t]$ such that $\lim_{n\to\infty} \|\Pi_n\| = 0$. Then $V_t^{(2)}(\Pi_n) \to \langle M\rangle_t$ in probability as $n \to \infty$.*

Definition. *The stochastic process X^T defined for all $t \geq 0$ by $X_t^T := X_{t\wedge T}$ is said to be the process stopped at T.*

Definition. *A stochastic process $\{M_t, \mathcal{F}_t; t \geq 0\}$ is called a local martingale if there exists a non-decreasing sequence of stopping times of $\{\mathcal{F}_t\}$, $\{T_n; n \geq 1\}$, such that the stopped process M^{T_n} is a martingale for each $n \geq 1$ and $\lim_{n\to\infty} T_n = \infty$ a.s.*

Stochastic Integration

A consequence of the non-zero quadratic variation of Brownian motion and the continuity of its sample paths is that it has sample paths of unbounded variation on any interval a.s. As a result, it is not possible to define Riemann-Stieltjes integrals of general continuous stochastic processes with respect to Brownian motion.

In this subsection, the theory of stochastic integration initiated by Itô that avoids the problems inherent in a Riemann-Stieltjes approach is reviewed. The approach is to define stochastic integration with respect to martingale integrators for integrands that are adapted to the same filtration as the martingale. Throughout this subsection, let $\{\mathcal{F}_t; t \geq 0\}$ be a filtration satisfying the usual conditions and let $\{M_t, \mathcal{F}_t; t \geq 0\}$ be a continuous square-integrable martingale.

Definition. *Let X be an $\{\mathcal{F}_t; t \geq 0\}$-adapted process. Denote by*

$$[X]_T^2 := E\left[\int_0^T X_t^2 d\langle M\rangle_t\right]$$

and

$$[X] := \sum_{n=1}^{\infty} 2^{-n}(1 + [X]_n).$$

Definition. *A stochastic process X is said to be progressively measurable with respect to the filtration $\{\mathcal{F}_t; t \geq 0\}$ if for each $t \geq 0$ and $A \in \mathcal{B}(\mathbb{R}^d)$,*

$$\{(s,\omega); s \in [0,t], \omega \in \Omega, X_s(\omega) \in A\} \in \mathcal{B}([0,t]) \otimes \mathcal{F}_t.$$

Remark. The following theorem shows that the condition of progressive measurability is not difficult to attain. Indeed, Brownian motion has a progressively measurable modification.

Theorem. *Let $\{X_t, \mathcal{F}_t; t \geq 0\}$ be an adapted process. If every sample path of X is right-continuous or every sample path of X is left-continuous then X is progressively measurable with respect to $\{\mathcal{F}_t\}$.*

Definition. *Let $\mathcal{L}^*$ denote the set of equivalence classes of progressively measurable processes satisfying $[X]_T < \infty$ for all $T > 0$.*

Definition. *A process X is called simple if there exists a strictly increasing sequence of real numbers $\{t_n\}_{n=0}^{\infty}$ with $t_0 = 0$ and $\lim_{n\to\infty} t_n = \infty$ such that:*

1. *There exists a sequence of random variables $\{\xi_n\}_{n=0}^{\infty}$ and a constant C such that $\sup_{n\geq 0} |\xi_n(\omega)| \leq C$ for every $\omega \in \Omega$.*
2. *ξ_n is $\mathcal{F}_{t_n}$-measurable for every $n \geq 0$.*
3. *X is defined for all $t \geq 0$ and $\omega \in \Omega$ by*

$$X_t(\omega) := \xi_0(\omega)\mathbb{I}_{\{0\}}(t) + \sum_{i=0}^{\infty} \xi_i(\omega)\mathbb{I}_{(t_i,t_{i+1}]}(t).$$

The class of simple processes is denoted by $\mathcal{L}_0$.

Definition. *For $X \in \mathcal{L}_0$, the stochastic integral of X with respect to M, $I_t(X)$, is defined for $t \geq 0$ by*

$$I_t(X) := \sum_{i=0}^{\infty} \xi_i(M_{t\wedge t_{i+1}} - M_{t\wedge t_i}).$$

Remark. The extension of the definition of the stochastic integral from the set $\mathcal{L}_0$ to the set $\mathcal{L}^*$ in a well-defined (i.e. unique) way is based upon the observation that for

$X \in \mathcal{L}_0$,

$$E\left[(I_t(X))^2\right] = E\left[\int_0^t X_s^2 d\langle M\rangle_s\right]$$

and the following result.

Theorem. *$\mathcal{L}_0$ is dense in $\mathcal{L}^*$ with respect to the metric* $\mathrm{d}(X,Y) := [X-Y]$ *for $X, Y \in \mathcal{L}^*$.*

Definition. *For $X \in \mathcal{L}^*$, the stochastic integral of X with respect to M is the unique square-integrable martingale $I(X) = \{I_t(X), \mathcal{F}_t; t \geq 0\}$ which satisfies*

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} \frac{\sqrt{E\left[|I(X^{(n)}) - I(X)|^2\right] \wedge 1}}{2^k} = 0$$

for every sequence $\{X^{(n)}\}_{n=1}^{\infty} \subseteq \mathcal{L}_0$ with $\lim_{n\to\infty}[X^{(n)} - X] = 0$. Denote for $t \geq 0$

$$I_t(X) = \int_0^t X_s dM_s.$$

Theorem. *For $X \in \mathcal{L}^*$,*

$$E\left[(I_t(X))^2\right] = E\left[\int_0^t X_s^2 d\langle M\rangle_s\right]$$

and

$$\langle I(X)\rangle_t = \int_0^t X_s^2 d\langle M\rangle_s.$$

Theorem (Itô's Formula)**.** *Let $\{M_t := (M_t^{(1)}, \ldots, M_t^{(d)}), \mathcal{F}_t; t \geq 0\}$ be a vector of continuous local martingales with $M_0 = 0$ a.s. and $\{A_t := (A_t^{(1)}, \ldots, A_t^{(d)}), \mathcal{F}_t; t \geq 0\}$ a vector of adapted processes of bounded variation with $A_0 = 0$. Set $X_t = X_0 + M_t + A_t$ for $t \geq 0$ where X_0 is an $\mathcal{F}_0$-measurable random vector in $\mathbb{R}^d$ and let $f \in C^{1,2}(\mathbb{R}^+ \times \mathbb{R}^d; \mathbb{R})$. Then a.s. it holds that for all $t \geq 0$,*

$$f(t, X_t) = f(0, X_0) + \int_0^t \frac{\partial}{\partial t} f(s, X_s) ds + \sum_{i=1}^{d} \int_0^t \frac{\partial}{\partial x_i} f(s, X_s) dA_s^{(i)}$$

$$
\begin{aligned}
&+\sum_{i=1}^{d}\int_0^t \frac{\partial}{\partial x_i} f(s,X_s)dM_s^{(i)} \\
&+\sum_{i=1}^{d}\sum_{j=1}^{d}\int_0^t \frac{\partial^2}{\partial x_i \partial x_j} f(s,X_s)d\langle M^{(i)}, M^{(j)}\rangle_s.
\end{aligned}
$$

Remark. We note that Itô's Formula is a fundamental tool in stochastic analysis and makes it easy to consider functions of semi-martingales.

Theorem (Burkholder-Davis-Gundy Inequality)**.** *Let M be a continuous local martingale such that $M_0 = 0$ a.s. Then for every $m > 0$ there exist positive constants k_m and K_m such that for every stopping time T*

$$
k_m E\left[\langle M\rangle_T^m\right] \leq E\left[\sup_{s\in[0,T]} |M_s|^{2m}\right] \leq K_m E\left[\langle M\rangle_T^m\right].
$$

Remark. As we will see, the Burkholder-Davis-Gundy Inequality is a very useful tool and one which we shall make frequent use of. In our uses of the inequality, the continuous local martingale M will be a stochastic integral.

Theorem (Martingale Representation Thoerem)**.** *Let $\{W_t, \mathcal{F}_t; t \geq 0\}$ be a d-dimensional standard Brownian motion where $\{\mathcal{F}_t\}$ is the augmented filtration of W. Then for any square-integrable martingale $\{M_t, \mathcal{F}_t; t \geq 0\}$ with $M_0 = 0$ a.s. and cadlag paths a.s., there exist square-integrable progressively measurable processes $\{Y_t^{(j)}, \mathcal{F}_t; t \geq 0\}$ such that for $t \geq 0$*

$$
M_t = \sum_{j=1}^{d}\int_0^t Y_s^{(j)} dW_s^{(j)}.
$$

Remark. As we will see, the Martingale Representation Thoerem is fundamental to the theory of backward stochastic differential equations.

Now let $\{W_t, \mathcal{F}_t; t \geq 0\}$ be a d-dimensional standard Brownian motion and recall that $\{\mathcal{F}_t\}$ satisfies the usual conditions. Let $\{X_t, \mathcal{F}_t; t \geq 0\}$ be a d-dimensional vector of adapted processes satisfying for $1 \leq i \leq d$ and any $t \geq 0$,

$$
\int_0^t \left(X_s^{(i)}\right)^2 ds < \infty \quad \text{a.s.}
$$

Define

$$Z_t(X) := \exp\left\{\sum_{i=1}^{d}\int_0^t X_s^{(i)} dW_s^{(i)} - \frac{1}{2}\int_0^t |X_s^{(i)}|^2 ds\right\}$$

then

$$Z_t(X) = 1 + \sum_{i=1}^{d}\int_0^t Z_s(X) X_s^{(i)} dW_s^{(i)}.$$

It follows that $Z(X)$ is a continuous local martingale with $Z_0(X) = 1$. If $Z(X)$ is a martingale, define for each $T \geq 0$ the probability measure Q_T on $\mathcal{F}_T$ by $Q_T(A) := E\left[\mathbb{I}_A Z_T(X)\right]$ for each $A \in \mathcal{F}_T$.

Theorem (Novikov's Condition)**.** *If for all $T \geq 0$*

$$E\left[\exp\left\{\frac{1}{2}\int_0^T |X_s|^2 ds\right\}\right] < \infty$$

then $Z(X)$ is a martingale.

Theorem (Bayes' Rule)**.** *Fix $T \geq 0$ and assume that $Z(X)$ is a martingale. If $0 \leq s \leq t \leq T$ and Y is an $\mathcal{F}_t$-measurable random variable with $E_{Q_T}\left[|Y|\right] < \infty$ then*

$$E_{Q_T}\left[Y|\mathcal{F}_s\right] = \frac{1}{Z_s(X)} E\left[Y Z_t(X)|\mathcal{F}_s\right]\ , \qquad P\text{-a.s. and } Q_T\text{-a.s.}$$

Theorem (Girsanov's Theorem)**.** *Assume that $Z(X)$ is a martingale and define the d-dimensional process $\{\tilde{W}_t, \mathcal{F}_t; t \geq 0\}$ by*

$$\tilde{W}_t^{(i)} := W_t^{(i)} - \int_0^t X_s^{(i)} ds; \quad 1 \leq i \leq d\ , \quad t \geq 0.$$

Then for each fixed $T \geq 0$, $\{\tilde{W}_t, \mathcal{F}_t; t \in [0,T]\}$ is a d-dimensional standard Brownian motion on $(\Omega, \mathcal{F}_T, Q_T)$.

Stochastic Differential Equations

Let $b_i(t,x)$, $\sigma_{ij}(t,x)$; $1 \leq i \leq d$, $1 \leq j \leq r$, be Borel-measurable functions from $\mathbb{R}^+ \times \mathbb{R}^d$ into $\mathbb{R}$. Define the vector $b(t,x) := \{b_i(t,x); 1 \leq i \leq d\}$ and the matrix

$\sigma(t,x) := \{\sigma_{ij}(t,x); 1 \le i \le d, 1 \le j \le r\}$. Furthermore, let $W = \{W_t; t \ge 0\}$ be an r-dimensional standard Brownian motion and take $\{\mathcal{F}_t; t \ge 0\}$ to be the augmented filtration of W.

Consider the stochastic differential equation

$$\begin{aligned} dX_s^{t,x} &= b(s, X_s^{t,x})ds + \sigma(s, X_s^{t,x})dW_s \ , \qquad s \in [t,T] \\ X_t^{t,x} &= x. \end{aligned} \tag{2.1}$$

Definition. *A strong solution of the stochastic differential equation (2.1) is an adapted process $\{X_t, \mathcal{F}_t; t \ge 0\}$ with continuous sample paths such that*

1. *For every $1 \le i \le d$, $1 \le j \le r$ and $t \ge 0$*

$$\int_0^t \left\{ |b_i(s, X_s)| + \sigma_{ij}^2(s, X_s) \right\} ds < \infty \quad a.s.$$

2. *a.s. it holds for all $s \ge t$ that*

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

Theorem. *If there exists a constant K such that for every $t \ge 0$ and $x, y \in \mathbb{R}^d$*

$$|b(t,x) - b(t,y)| + \|\sigma(t,x) - \sigma(t,y)\| \le K$$

then the stochastic differential equation (2.1) has a unique up to indistinguishability strong solution.

Theorem (Feynman-Kac)**.** *Suppose that the conditions of the previous theorem hold and define the differential operator*

$$\mathcal{L}u := \sum_{i=1}^d b_i(t,x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^d a_{ij}(t,x) \frac{\partial^2}{\partial x_i \partial x_j} \ , \qquad (a_{ij}(t,x)) := \sigma\sigma^T(t,x).$$

Fix $T > 0$ and assume that there exist constants $L > 0$ and $\lambda \ge 2$ such that the continuous functions $f : \mathbb{R}^d \to \mathbb{R}$, $g : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ and $k : [0,T] \times \mathbb{R}^d \to \mathbb{R}^+$ satisfy

1. *$|f(x)| \le L(1 + |x|^\lambda)$; $\quad x \in \mathbb{R}^d$.*

2. $|g(t,x)| \leq L(1+|x|^\lambda)$; $\quad t \in [0,T]$, $x \in \mathbb{R}^d$.

Suppose further that $u \in C^{1,2}([0,T) \times \mathbb{R}^d; \mathbb{R})$ *satisfies the PDE*

$$-\frac{\partial u(t,x)}{\partial t} + k(t,x)u(t,x) = \mathcal{L}u(t,x) + g(t,x); \quad t \in [0,T), x \in \mathbb{R}^d, \qquad (2.2)$$
$$u(T,x) = f(x); \quad x \in \mathbb{R}^d.$$

If in addition there exist constants $M > 0$ *and* $\mu \geq 0$ *such that*

$$\max_{t\in[0,T]} |u(t,x)| \leq M(1+|x|^{2\mu}); \quad x \in \mathbb{R}^d \ ,$$

then for $t \in [0,T]$ *and* $x \in \mathbb{R}^d$,

$$u(t,x) = E\left[f(X_T^{t,x}) \exp\left\{-\int_t^T k(r, X_r^{t,x})dr\right\} + \int_t^T g(s, X_s^{t,x}) \exp\left\{-\int_t^s k(r, X_r^{t,x})dr\right\} ds\right].$$

Remark. Heuristically, this says that we can solve the PDE (2.2) by letting the stochastic process X explore space until time T and then calculate a functional of its path. The process X explores space in just the right way to generate the differential operator $\mathcal{L}$. The theme of probabilistic representations of the solution of a PDE is fundamental to the topic of this thesis. As we will see in our review of literature in Chapter 3, backward doubly stochastic differential equations (BDSDEs) provide probabilistic representations of a class of stochastic PDEs and so by approximating solutions of certain BDSDEs, we are in fact able to approximate solutions of stochastic PDEs.

Remark. We note that the concept of a probabilistic representation is implicit in the following elementary result on the heat equation: the solution to the initial value problem

$$\frac{\partial u(t,x)}{\partial t} = \frac{1}{2}\frac{\partial^2 u(t,x)}{\partial x^2}, \quad u(0,x) = h(x)$$

is given by

$$u(t,x) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-(x-y)^2/2t} h(y)dy \equiv E\left[h(x+W_t)\right].$$

Backward Stochastic Differential Equations

Backward stochastic differential equations (BSDEs) have a similar form to SDEs with random coefficients in the sense that they contain a Lebesgue integral term and a stochastic integral term. Furthermore, as in the case of the solution of an SDE, the solution of a BSDE is adapted to the filtration of the driving noise.

The backward nomenclature refers to the provision of a terminal condition as part of the specification of a BSDE as opposed to a starting condition. As a result of this, the solution of a BSDE is no longer a single process but in fact a pair of processes. BSDEs have applications to stochastic control and also provide probabilistic representations for a class of semilinear PDEs (see for example [40] or Chapter 3). As we will see in our review of literature in Chapter 3, the advantage of the BSDE representation of PDEs is that it is allows more non-linearity in the coefficients of the PDE.

To introduce BSDEs, let $\{W_t, \mathcal{F}_t; t \geq 0\}$ be a d-dimensional Brownian motion with $\{\mathcal{F}_t\}$ the augmented filtration of W and fix $T > 0$. Let $\xi \in L^2((\Omega, \mathcal{F}_T, P); \mathbb{R})$ and $f : \Omega \times [0,T] \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$. $f(\omega, t, y, z)$ is written as $f(t, y, z)$ (i.e. the dependence on ω is implicit) and it is assumed that

1. For any fixed $y \in \mathbb{R}$ and $z \in \mathbb{R}^d$, the random function $f(., y, z)$ is progressively measurable.
2. $E\left[\int_0^T |f(t,0,0)|^2 dt\right] < \infty$.
3. There exists a constant C such that for all $y_1, y_2 \in \mathbb{R}$, $z_1, z_2 \in \mathbb{R}^d$ almost every $t \in [0,T]$ and a.s.

$$|f(t, y_1, z_1) - f(t, y_2, z_2)| \leq C(|y_1 - y_2| + |z_1 - z_2|).$$

Consider the BSDE

$$dY_t = f(t, Y_t, Z_t)dt - \langle Z_t, dW_t \rangle \ , \quad Y_T = \xi. \tag{2.3}$$

Definition. *A solution to the BSDE (2.3) is a progressively measurable pair* (Y, Z) *satisfying for all* $t \in [0,T]$

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T \langle Z_s, dW_s \rangle$$

such that

$$E\left[\sup_{t\in[0,T]}|Y_t|^2+\int_0^T|Z_t|^2dt\right]<\infty.$$

Theorem. *Given a pair (ξ, f) satisfying the conditions above, there exists a unique solution to the BSDE (2.3).*

Remark. Given that the random terminal condition is specified in the definition of BSDE (2.3), it is not clear how to construct the solution (Y, Z) so that they are both adapted to $\{\mathcal{F}_t\}$. To give the main idea of how this is done, we reproduce part of the proof (taken from [40]) below with several technical points omitted.

Proof. The proof is based upon the fixed point method. Consider the mapping $(U, V) \to (Y, Z)$ defined by

$$Y_t=\xi+\int_t^T f(s,U_s,V_s)ds-\int_t^T \langle Z_s, dW_s\rangle .$$

The pair (Y, Z) is constructed as follows: first define the martingale

$$M_t:=E\left[\xi+\int_0^T f(s,U_s,V_s)ds\middle|\mathcal{F}_t\right].$$

Then, by the Martingale Representation Theorem, there exists an adapted process Z such that

$$M_t=M_0+\int_0^t \langle Z_s, dW_s\rangle .$$

Define the process Y by

$$Y_t=E\left[\xi+\int_t^T f(s,U_s,V_s)ds\middle|\mathcal{F}_t\right].$$

It follows that since $Y_T=\xi$,

$$\begin{aligned}Y_t&=M_t-\int_0^t f(s,U_s,V_s)ds\\&=M_0-\int_0^t f(s,U_s,V_s)ds+\int_0^t \langle Z_s, dW_s\rangle\end{aligned}$$

$$= \xi + \int_t^T f(s, U_s, V_s)ds - \int_t^T \langle Z_s, dW_s \rangle$$

as required. The remainder of the proof shows that this mapping is a contraction and is omitted. □

Backward Martingales

In this section we introduce backward martingales which will be helpful in understanding the theory of backward doubly stochastic differential equations. To this end, let $T > 0$ be some fixed finite time.

Definition. *A family of σ-algebras $\{\mathcal{F}_{t,T}; t \in [0,T]\}$ such that $\mathcal{F}_{t,T} \subseteq \mathcal{F}_{s,T} \subseteq \mathcal{F}$ if $0 \leq s \leq t \leq T$ is said to be a backward filtration of $\mathcal{F}$.*

Remark. For the remainder of this section we will assume a backward filtration, $\{\mathcal{F}_{t,T}; t \in [0,T]\}$, as given.

Definition. *A stochastic process $X = \{X_t; t \in [0,T]\}$ is said to be adapted to $\{\mathcal{F}_{t,T}\}$ if X_t is $\mathcal{F}_{t,T}$-measurable for every $t \in [0,T]$.*

Definition. *A real-valued process $\{M_t; t \in [0,T]\}$ is called a $\{\mathcal{F}_{t,T}\}$ backward martingale if*

1. *$M_t \in L^1((\Omega, \mathcal{F}, P); \mathbb{R})$ for every $t \in [0,T]$.*
2. *M is adapted to $\{\mathcal{F}_{t,T}\}$.*
3. *$E[M_s|\mathcal{F}_{t,T}] = M_t$ a.s. for every $0 \leq s \leq t \leq T$.*

Definition. *A continuous process $\{\overleftarrow{W}_t; t \in [0,T]\}$ taking values in $\mathbb{R}^d$ (with $d \geq 1$) is called a d-dimensional $\{\mathcal{F}_{t,T}\}$ backward Wiener process if*

1. *$\overleftarrow{W}$ is adapted to $\{\mathcal{F}_{t,T}\}$.*
2. *$\overleftarrow{W}_T = 0$ a.s.*
3. *For $0 \leq s \leq t \leq T$, $\overleftarrow{W}_s - \overleftarrow{W}_t$ is independent of $\mathcal{F}_{t,T}$.*
4. *For $0 \leq s \leq t \leq T$, $\overleftarrow{W}_s - \overleftarrow{W}_t$ is a Gaussian random variable with mean zero and covariance matrix $(t-s)I_d$, where I_d denotes the $d \times d$ identity matrix.*

Remark. Let W be a standard one-dimensional Wiener process and define the backward filtration $\mathcal{F}_{t,T} := \sigma\{W_s - W_T; s \in [t,T]\}$. Then the process $\overleftarrow{W}_t := W_t - W_T$ is a one-dimensional $\{\mathcal{F}_{t,T}\}$ backward Wiener process.

Definition. *A process X is called simple with respect to $\{\mathcal{F}_{t,T}\}$ if there exists a strictly increasing sequence of real numbers $\{t_k\}_{k=0}^n$, with $t_0 = 0$ and $t_n = T$ such that:*

1. *There exists a sequence of random variables $\{\xi_k\}_{k=1}^n$ and a constant C such that $\max_{k=1,\dots,n} |\xi_k(\omega)| \leq C$ for every $\omega \in \Omega$.*
2. *ξ_k is $\mathcal{F}_{t_k,T}$-measurable for every $k \geq 1$.*
3. *X is defined for all $t \in [0,T]$ and $\omega \in \Omega$ by*

$$X_t(\omega) := \sum_{k=1}^{n} \xi_k(\omega)\mathbb{I}_{[t_{k-1},t_k)}(t) + \xi_n(\omega)\mathbb{I}_{\{T\}}(t).$$

Definition. *Let $\{M_t; t \in [0,T]\}$ be a continuous square-integrable $\{\mathcal{F}_{t,T}\}$ backward martingale and X be a simple process with respect to $\{\mathcal{F}_{t,T}\}$. Then the backward stochastic integral of X with respect to M, $\overleftarrow{I}_t(X)$, is defined for $t \in [0,T]$ by*

$$\overleftarrow{I}_t(X) := \sum_{k=1}^{n} \xi_k (M_{t\vee t_k} - M_{t\vee t_{k-1}}).$$

Remark. Just as the (forward) stochastic integral is defined as the limit in probability of the stochastic integral of a sequence of simple processes with respect to a (forward) filtration, the backward stochastic integral is defined as the limit in probability of the backward stochastic integral of simple processes with respect to a backward filtration. Indeed, as remarked in [37], if $\overleftarrow{W}$ is an $\{\mathcal{F}_{t,T}\}$ backward Wiener process and X is a continuous process adapted to $\{\mathcal{F}_{t,T}\}$ then the backward stochastic integral of X with respect to $\overleftarrow{W}$ can be defined as

$$\int_t^T X_s \overleftarrow{dW}_s := \lim_{\|\Pi\|\to 0} \sum_{k=1}^{n} X_{t_k} (\overleftarrow{W}_{t\vee t_k} - \overleftarrow{W}_{t\vee t_{k-1}})$$

in probability (where $\|\Pi\| := \max_{1\leq k\leq n}|t_k - t_{k-1}|$ and it is implicit that $n \to \infty$ as $\|\Pi\| \to 0$).

3

Review of Literature

The review of literature in this chapter is separated into three sections. Section 3.1 covers the general theory and applications of backward stochastic differential equations (BSDEs), Section 3.2 covers the approximation of BSDEs and finally Section 3.3 covers backward doubly stochastic differential equations (BDSDEs).

The work in this thesis relies heavily upon three of the references reviewed in this chapter: the paper [49] (Zhang, 2004) on the approximation of BSDEs develops a general strategy based upon deriving a regularity result for Z which we adapt to the BDSDE setting; the papers [36] (Pardoux and Peng, 1994) and [50] (Zhang and Zhao 2007) on the general theory of BDSDEs contain results which we make repeated use of. As a consequence, these references are covered in additional detail in three "Key Reference" subsections.

3.1. BSDEs

Motivated by stochastic control theory, BSDEs were introduced by Pardoux and Peng in 1990 in the paper [34]. They consider BSDEs of the form

$$Y_s = \xi + \int_s^T f(r, Y_r, Z_r)dr + \int_s^T g(r, Y_r, Z_r)dW_r \tag{3.1}$$

for $s \in [0, T]$ where $\{W_t; t \geq 0\}$ is a k-dimensional standard Wiener process on a complete probability space $(\Omega, \mathcal{F}, P)$ with $\{\mathcal{F}_t; t \geq 0\}$ the augmented natural filtration of W; $\xi \in L^2((\Omega, \mathcal{F}_T, P); \mathbb{R}^d)$. $f : \Omega \times [0, T] \times \mathbb{R}^d \times \mathbb{R}^{d\times k} \to \mathbb{R}^d$ and $g : \Omega \times [0, T] \times \mathbb{R}^d \times \mathbb{R}^{d\times k} \to \mathbb{R}^{d\times k}$ are random functions such that

1. f is $\mathcal{P} \otimes \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^{d\times k})/\mathcal{B}(\mathbb{R}^d)$-measurable and g is $\mathcal{P} \otimes \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^{d\times k})/\mathcal{B}(\mathbb{R}^{d\times k})$-measurable where $\mathcal{P}$ denotes the σ-algebra of $\{\mathcal{F}_t\}$-progressively measurable subsets of $\Omega \times [0, T]$.

2. $E\left[\int_0^T \{|f(r,0,0)|^2 + \|g(r,0,0)\|^2\} \, dr\right] < \infty$.

3. For a.e. (t, ω), f and g are globally Lipschitz in y and z.

4. There exists a constant $\alpha > 0$ such that for every y and a.e. (t, ω), $|g(t, y, z_1) - g(t, y, z_2)| \geq \alpha|y_1 - y_2|$.

In this context, they show that there exists a unique $\{\mathcal{F}_t\}$-progressively measurable pair (Y, Z) satisfying $E\left[\int_0^T \{|Y_r|^2 + \|Z_r\|^2\} \, dr\right] < \infty$ which solves (3.1).

In [39] (1991), Peng established the connection between BSDEs and quasilinear PDEs. This was achieved by introducing loosely-coupled BSDEs of the form

$$
\begin{aligned}
X_s^{t,x} &= x + \int_t^s b(r, X_r^{t,x})dr + \int_t^s \sigma(r, X_r^{t,x})dW_r, \\
Y_s^{t,x} &= h(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})dr - \int_s^T Z_r^{t,x} dW_r
\end{aligned}
\tag{3.2}
$$

for $s \in [t, T]$. We note that the terminology "loosely-coupled" refers to the fact that the backward equation for Y and Z depends upon the forward equation for X but the forward equation is independent of the backward equation. Given differentiability conditions on the coefficients and non-degeneracy of σ, Peng showed that if u solves the parabolic PDE

$$
\begin{aligned}
\frac{\partial u(t,x)}{\partial t} &= \mathcal{L}u(t,x) + f(t, x, u(t,x), \nabla u(t,x)\sigma(t,x)), \\
u(T,x) &= h(x)
\end{aligned}
\tag{3.3}
$$

where

$$
\mathcal{L}u := \sum_{i=1}^d b_i(t,x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^d a_{ij}(t,x)\frac{\partial^2}{\partial x_i \partial x_j} \ , \quad (a_{ij}(t,x)) := \sigma\sigma^T(t,x)
$$

then $u(t,x) = Y_t^{t,x}$. Peng also established similar relations for parabolic and elliptic PDEs defined on bounded domains of $\mathbb{R}^d$.

In [35] (1992) Pardoux and Peng made significant inroads in extending the theory of loosely-coupled BSDES. Given differentiability conditions on the coefficients of BSDE (3.2), Pardoux and Peng showed that $u(t,x) := Y_t^{t,x}$ solves the PDE (3.3). An intermediate step to this result was the derivation of the key relationship

$$Z_s^{t,x} = \nabla Y_s^{t,x} (\nabla X_s^{t,x})^{-1} \sigma(X_s^{t,x}).$$

They also showed that if f and h are just Lipschitz continuous in y and z then $u(t,x) := Y_t^{t,x}$ is a viscosity solution of the PDE (3.3). Despite this being a key paper in the theory of BSDEs and their connection to PDEs, we do not provide further details here. This has been done to avoid repetition when in the next section we provide details of their paper [36] on BDSDEs which is very much in the same vein as [35].

Remark. As we will see, in [49] Zhang uses the representation

$$Z_s^{t,x} = \nabla Y_s^{t,x} (\nabla X_s^{t,x})^{-1} \sigma(X_s^{t,x})$$

derived in [35] to construct his key result on the regularity of Z. In [36] Pardoux and Peng derive the same representation of Z for the BDSDE case and we will make use of this representation to prove our result on the regularity of Z.

In [6] (1997) Barles, Buckdahn and Pardoux consider loosely-coupled BSDEs and incorporate a Poisson random measure into the driving noise of both the forward and backward equations. In this setting, the solution of the BSDE is no longer a pair (Y, Z) but now a triple (Y, Z, U). With conditions similar to [35], they show that the BSDE has a unique solution. They then (again in a similar vein to [35]) connect the solution of the BSDE to the viscosity solution of a system of parabolic integral-partial differential equations.

In [25] (2000) Kobylanski considers BSDEs of the form

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s) ds - \int_t^T Z_s dW_s$$

for $t \in [0, T]$ where the terminal condition ξ is bounded and f is continuous and has quadratic growth in Z. She then specializes these conditions to the loosely coupled BSDE setting and connects the solution of the loosely coupled BSDE to the viscosity solution of the corresponding PDE.

In [3] (1993) Antonelli considers adapted solutions to fully coupled forward-backward SDEs (FBSDEs) of the form

$$
\begin{aligned}
U_t &= J_t + \int_0^t f(s, U_s, V_s) dX_s \\
V_t &= E\left[\int_t^T g(s, U_s, V_s) dZ_s + Y \middle| \mathcal{F}_t\right], \quad t \in [0, T], \\
V_T &= Y,
\end{aligned}
$$

where Y is an $\mathcal{F}_T$-measurable random variable, f and g are uniformly Lipschitz in u and v, X and Z are semimartingales and J_t is a cadlag process. We note that here, as both U and V appear in both the forward and backward equations, we refer to the equations as fully coupled. We also note that whilst the equations are in some sense more general (for example the fully coupling and the generalization to semimartingale noise), they are less general in the sense that the driver g of the backward equation only depends upon U and V. We note that this is a significant departure from the loosely coupled case as in the loosely coupled case the backward equation is allowed to depend upon the solution of the forward equation but the forward equation may not depend upon the solution of the backward equation.

In [29] (1994) Ma, Protter and Yong consider fully coupled FBSDEs of the form

$$
\begin{aligned}
X_s^{t,x} &= x + \int_t^s b(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}) dr + \int_t^s \sigma(r, X_r^{t,x}, Y_r^{t,x}) dW_r, \\
Y_s^{t,x} &= h(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}) dr + \int_s^T g(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}) dW_r
\end{aligned}
$$

where W is a d-dimensional Brownian motion and b, σ, f and g are all smooth functions. Under these strong conditions, they show that this very general form of FBSDE has a unique adapted solution triple $(X, Y, Z) : [0, T] \times \Omega \to \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{m \times d}$. They achieve this by following what they call the "Four Step Scheme":

Step 1 Find a smooth mapping $z : [0, T] \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times d}$ satisfying for all $(t, x, y, p) \in [0, T] \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{m \times n}$

$$
p\sigma(t, x, y) + g(t, x, y, z(t, x, y, p)) = 0.
$$

Step 2 Using z, solve the following PDE for $u(t,x)$:

$$
\begin{aligned}
&u_t^k + \frac{1}{2}\mathrm{tr}(u_{xx}\sigma(t,x,u)\sigma(t,x,u)^T)) + \langle b(t,x,u,z(t,x,u,u_x)), u_x^k\rangle \\
&\qquad f^k(t,x,u,z(t,x,u,u_x)) = 0\ , \quad k = 1,\dots,m\ , \quad (t,x) \in [0,T) \times \mathbb{R}^n, \\
&u(T,x) = h(x)\ , \quad x \in \mathbb{R}^n.
\end{aligned}
$$

Step 3 Using u and z, solve the SDE:

$$
X_t = x + \int_0^t \tilde{b}(s,X_s)ds + \int_0^t \tilde{\sigma}(s,X_s)dW_s
$$

where $\tilde{b}(t,x) := b(t,x,u(t,x),z(t,x,u(t,x),u_x(t,x)))$ and $\tilde{\sigma}(t,x) := \sigma(t,x,u(t,x))$.

Step 4 Set

$$
\begin{aligned}
Y_t &:= u(t,X_t), \\
Z_t &:= z(t,X_t,u(t,X_t),u_x(t,X_t)).
\end{aligned}
$$

It is interesting to note the use of the PDE connection in the Four Step Scheme.

In [13] (2002) Delarue considers fully coupled FBSDEs of the form

$$
\begin{aligned}
X_t &= \xi + \int_0^t b(s,X_s,Y_s,Z_s)ds + \int_0^t \sigma(s,X_s,Y_s)dW_s, \\
Y_t &= h(X_T) + \int_t^T f(s,X_s,Y_s,Z_s)ds - \int_t^T Z_s dW_s
\end{aligned}
$$

under conditions weaker than those of [29]. Delarue derives existence and uniqueness of the above FBSDE under fairly standard Lipschitz conditions on the coefficients along with a non-degeneracy condition on σ. We note that as in [29], the PDE connection plays a role in the proof; namely Delarue proves the existence and uniqueness of an adapted triple (X,Y,Z) that solves the FBSDE over a small (enough) time interval and uses the PDE connection to extend the result to an arbitrary (but fixed) time interval. We note that the PDE in this case (ditto [29]) is significantly more general than the case of [35] since the coefficient b is now allowed to depend upon Y and Z and σ is allowed to depend upon Y.

As we have seen, one application of BSDEs is to provide a probabilistic representation for a class of PDEs. The PDE connection is not, however, the only area of application for BSDEs. For example, in [15] (1997) El Karoui et al consider reflected BSDEs with solution triple (Y, Z, K) of the form

$$\begin{aligned} &Y_t = \xi + \int_0^t f(s, Y_s, Z_s) ds + K_T - K_t - \int_t^T Z_s dW_s \ , \quad t \in [0, T], \\ &Y_t \geq S_t \ , \quad t \in [0, T] \end{aligned} \tag{3.4}$$

where S is a given continuous, progressively measurable process known as the obstacle. In addition, the new component to the solution, K, must be continuous and increasing and Y and K must satisfy

$$\int_0^T \{Y_t - S_T\} \, dK_t = 0.$$

They show that if f is Lipschitz in y and z then there exists a unique solution to the reflected BSDE. They then relate the solution component Y to the value function of optimal stopping problems and, by loosely coupling the BSDE (3.4) to the solution of an SDE, the viscosity solution of PDE obstacle problems. Furthermore, in [20] (2007) Hamadene and Jeanblanc apply reflected BSDEs to real option problems such as determining the optimal strategy for electricity production by a power station.

In [11] (2014) Cohen considers BSDEs where the noise is generated by a continuous time Markov chain and the terminal value is prescribed by a stopping time. With this formulation, he finds applications to determining the optimal policy (for each message sending node) for transmitting messages over a finite network and the optimal control (on the speed of individual edge traversals) for traversing a directed graph.

3.2. Approximation of BSDEs

There have been several approaches to solving BSDEs numerically (or at least via discretization scheme upon which a numerical scheme could potentially be based). Most of these approaches fall quite comfortably into one of three camps.

The first camp makes use of the PDE connection to BSDEs and solves the associated PDE numerically. Indeed, one could simply approximate loosley coupled BSDEs of the form

$$X_s^{t,x} = x + \int_t^s b(r, X_r^{t,x})dr + \int_t^s \sigma(r, X_r^{t,x})dW_r \ ,$$
$$Y_s^{t,x} = h(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})dr - \int_t^T Z_r^{t,x}dW_r \tag{3.5}$$

by approximating the related PDE using standard techniques.

A second camp approximates the driving Brownian motion with a simpler stochastic process, such as a random walk and shows that the solution to the simpler equation converges to the solution of the original BSDE in some sense. Whilst this approach is direct and appealing from a mathematical point of view, to date it has tended to require strong assumptions on the coefficients of the BSDE.

A final camp attacks the BSDE problem directly by simply discretizing the BSDE itself. For example, a simple Euler-like discretization scheme of the BSDE given by (3.5) would take the form

$$Y_s^{t,x} \approx Y_u^{t,x} + f(u, X_u^{t,x}, Y_u^{t,x}, Z_u^{t,x})(u-s) - \left\langle Z_s^{t,x}, W_u - W_s \right\rangle$$

for $t \leq s \leq u$. A problem with this approximation, however, is that it fails to ensure that Y_s is $\mathcal{F}_s$-measurable. To fix this, we can take conditional expectations to give the explicit approximation

$$Y_s^{t,x} \approx E\left[Y_u^{t,x} + f(u, X_u^{t,x}, Y_u^{t,x}, Z_u^{t,x})(u-s)|\mathcal{F}_s\right].$$

We note that this is effectively the approach taken by Zhang in [49] and the approach taken in this thesis for the BDSDE case.

We begin our review on the approximation of BSDEs with a paper from the PDE camp. In [14] (1996) Douglas, Ma and Protter consider fully coupled FBSDEs of the form considered in [29]:

$$X_s^{t,x} = x + \int_t^s b(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})dr + \int_t^s \sigma(r, X_r^{t,x}, Y_r^{t,x})dW_r,$$
$$Y_s^{t,x} = h(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})dr + \int_s^T g(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})dW_r$$

where W is a d-dimensional Brownian motion and b, σ, f and g are all smooth functions. To approximate the above equation, they follow the “Four Step Scheme” of [29]. They use standard techniques to approximate the PDE of Step 2 and use the Euler scheme to approximate the SDE of Step 3. We note that in [31] (2006) Milstein and Tretyakov refine this approach to obtain a more efficient scheme by utilising more advanced techniques to approximate the equations in steps 2 and 3 of the “Four Step Scheme” of [29]

The first scheme to directly approximate loosely coupled BSDEs with conditions similar to those required for existence and uniqueness as derived by Pardoux and Peng in [35] was that of Zhang in [49] of which we give a detailed overview in this section. Prior to this, we mention three earlier attempts at direct approximation.

Firstly in [10] (1997) Chevance considers loosely coupled BSDEs of the form

$$\begin{aligned} X_t &= \xi + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s \ , \\ Y_t &= h(X_T) + \int_t^T f(r, X_s, Y_s)ds - \int_t^T Z_s dW_s \end{aligned}$$

where the coefficients b, σ, f and h satisfy strong smoothness conditions. To describe his scheme, let $t = t_0, t_1, \dots, t_n = T$ be a uniform partition of $[0, T]$, $h := \frac{T}{n}$, U^n be a discrete-time approximation of W and $\{\mathcal{F}_j^n\}$ the natural filtration of U^n. Chevance's scheme approximates X and Y with $\widehat{X}$ and $\widehat{Y}$ respectively where $\widehat{X}$ is given by

$$\begin{aligned} &\widehat{X}_0 := \xi, \\ &\widehat{X}_j := \widehat{X}_{j-1} + hb(t_{j-1}, \widehat{X}_{j-1}) + \sqrt{h}\sigma(t_{j-1}, \widehat{X}_{j-1})U_j^n \ , \quad 1 \le j \le n \end{aligned}$$

and $\widehat{Y}$ is given by

$$\begin{aligned} &\widehat{Y}_n := h(\widehat{X}_n), \\ &\widehat{Y}_j := E\left[\widehat{Y}_{j+1} + hf(t_{j+1}, \widehat{X}_{j+1}, \widehat{Y}_{j+1}) \middle| \mathcal{F}_j^n\right] \ , \quad 1 \le j \le n. \end{aligned}$$

In comparison to the results derived in [49], Chevance's scheme requires much stronger conditions on the coefficients and does not allow f to depend upon z.

In [9] (2001) Briand, Delyon and Mémin do construct a scheme that allows f to depend upon z. They consider equations of the form

$$Y_t = \xi + \int_t^T f(Y_s, Z_s)ds - \int_t^T Z_s dW_s$$

where $\xi \in L^2((\Omega, \mathcal{F}_T, P), \mathbb{R})$ and f is Lipschitz in y and z. Their scheme is based upon approximating the standard Brownian motion W with a scaled random walk. More specifically, they uniformly discretize $[0, T]$ into n subintervals, define $h := \frac{T}{n}$ and take $\{\epsilon_k\}_{1 \leq k \leq n}$ to be an i.i.d. Bernoulli symmetric sequence. They define their approximation scheme by

$$\begin{aligned} y_n &= \xi^n, \\ y_k &= y_{k+1} + hf(y_k, z_k) - \sqrt{h} z_k \epsilon_{k+1} \ , \qquad k = n-1, \dots, 0 \\ z_k &= h^{-1/2} E\left[y_{k+1}\epsilon_{k+1} | \mathcal{G}_k\right]. \end{aligned}$$

They then show that if

$$W_t^n := \sqrt{h} \sum_{k=1}^{\lfloor t/h \rfloor} \epsilon_k^n \ , \qquad t \in [0, T] \ ,$$

satisfies $\sup_{t \in [0,T]} |W_t^n - W_t| \to 0$ in probability then

$$\sup_{t \in [0,T]} |Y_t^n - Y_t|^2 + \int_0^T |Z_s^n - Z_s|^2 ds \to 0$$

in probability where $Y_t^n := y_{\lfloor t/h \rfloor}^n$, $Z_t^n := z_{\lfloor t/h \rfloor}^n$.

We note that in [28] (2002) Ma et al also construct an approximation scheme based upon approximating standard Brownian motion W with a scaled random walk. They, however, consider equations of the form

$$Y_t = \xi + \int_t^T f(s, Y_s)ds - \int_t^T Z_s dW_s$$

where $\xi \in L^2((\Omega, \mathcal{F}_T, P), \mathbb{R})$ and f is Lipschitz in s and y. Whilst some of their conditions are weaker than those of [9], their driver f cannot depend upon z - at least not in a nonlinear way.

Key Reference: [49] - Zhang, 2004

To describe the approach of Zhang in [49], let $T > 0$ be a fixed terminal time, $(\Omega, \mathcal{F}, P)$ be a complete probability space on which is defined a standard Brownian motion W and take $\{\mathcal{F}_t; t \geq 0\}$ to be the augmented filtration of W. Denote by $\mathbb{D}$ the space of all real-valued cadlag functions defined on $[0, T]$. Let b,$\sigma : [0, T] \times \mathbb{R} \to \mathbb{R}$ and $f : [0, T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$ be deterministic functions and $\Phi : \mathbb{D}^d \to \mathbb{R}$ a deterministic functional. Zhang then considers the following loosely-coupled BSDE where X, Y and Z are all real-valued stochastic processes:

$$\begin{aligned} X_t &= x + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s \ , \\ Y_t &= \Phi(X) + \int_t^T f(s, X_s, Y_s, Z_s) ds - \int_t^T Z_s dW_s. \end{aligned} \tag{3.6}$$

Zhang assumes that:

1. b, σ and f are uniformly $\frac{1}{2}$ Hölder continuous in t and uniformly Lipschitz continuous in their remaining variables.

2. There exists a constant K such that for all $x_1, x_2 \in \mathbb{D}$,
$$|\Phi(x_1) - \Phi(x_2)| \leq K \sup_{0 \leq t \leq T} |x_1(t) - x_2(t)|.$$

3. There exists a constant K such that
$$\sup_{0 \leq t \leq T} \{|b(t, 0)| + |\sigma(t, 0)| + |f(t, 0, 0, 0)|\} + |\Phi(0)| \leq K.$$

Remark. From the results of [35] and under the above conditions, the SDE and BSDE given by (3.6) have a unique solution.

Zhang defines $\Pi : 0 = t_0 < \ldots < t_n = T$ to be a partition of $[0, T]$ and defines $\Delta t_i := t_i - t_{i-1}$ and $|\Pi| := \max_i \Delta t_i$. Before introducing his discretization scheme, he derives the following result on the regularity of Z.

Remark. The following result is really the key result of the paper and is what makes Zhang's approach work. As previously noted, the result hinges on the representation of Z derived in [35].

Theorem. *Suppose that the above conditions hold, that Z is cadlag and let Π be any partition of $[0,T]$. There exists a constant $C>0$ depending only upon T and K such that*

$$\sum_{i=1}^{n} E\left[\int_{t_{i-1}}^{t_i} \left\{|Z_r - Z_{t_{i-1}}|^2 + |Z_r - Z_{t_i}|^2\right\} dr\right] \leq C(1+|x|^2)|\Pi|. \tag{3.7}$$

Zhang then defines his discretization scheme as follows. Define $\pi(t) := t_{i-1}$ for $t \in [t_{i-1}, t_i)$ and let X^π be the solution to the SDE

$$X_t^\pi = x + \int_0^t b(\pi(s), X_{\pi(s)})ds + \int_0^t \sigma(\pi(s), X_{\pi(s)})dW_s.$$

Now define

$$Y_{t_n}^\pi = \xi^\pi \ , \quad Z_{t_n}^{\pi,1} = 0$$

and for $t \in [t_{i-1}, t_i)$, $i = n, n-1, \ldots, 1$

$$Y_t^\pi = Y_{t_i}^\pi + f(t_i, \Theta_{t_i}^{\pi,1})\Delta t_i - \int_t^{t_i} Z_r^\pi dW_r$$

where

$$\xi^\pi \in L^2((\Omega, \mathcal{F}_T, P); \mathbb{R}) \ , \quad \Theta_{t_i}^{\pi,1} := (X_{t_i}^\pi, Y_{t_i}^\pi, Z_{t_i}^{\pi,1})$$

and

$$Z_{t_i}^{\pi,1} := \frac{1}{\Delta t_{i+1}} E\left[\int_{t_i}^{t_{i+1}} Z_r^\pi dr \middle| \mathcal{F}_{t_i}\right].$$

He then proves the following theorem which provides a bound on the mean square error of the scheme.

Theorem. *Suppose that the above conditions hold, that Z is cadlag and that there is a constant $K_2 > 0$ such that partition Π satisfies $\Delta t_i \geq \frac{|\Pi|}{K_2}$ for $i = 1, \ldots, n$. Then*

$$\max_{0\leq i\leq n} E\left[|Y_{t_i} - Y_{t_i}^\pi|^2\right] + E\left[\int_0^T |Z_r - Z_r^\pi|^2 dr\right] \\ \leq C\left((1+|x|^2)|\Pi| + E\left[|\Phi(X) - \xi^\pi|^2\right]\right)$$

where C depends only upon T, K and K_2.

In [8] (2004) Bouchard and Touzi consider loosely coupled BSDEs of the form

$$X_s^{t,x} = x + \int_t^s b(r, X_r^{t,x})dr + \int_t^s \sigma(r, X_r^{t,x})dW_r \ ,$$
$$Y_s^{t,x} = h(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})dr - \int_t^T Z_r^{t,x} dW_r.$$

They define the following implicit discretization scheme which we note is simpler than the one defined by Zhang in [49].

$$\widehat{Y}_{t_n}^{t,x} := h(\widehat{X}_{t_n}) \ , \quad \widehat{Y}_{t_{i-1}}^{t,x} := E\left[\widehat{Y}_{t_i}^{t,x} | \mathcal{F}_{t_{i-1}}\right] + f(t_i, \widehat{X}_{t_{i-1}}^{t,x}, \widehat{Y}_{t_{i-1}}^{t,x}, \widehat{Z}_{t_{i-1}}^{t,x})\Delta t,$$
$$\widehat{Z}_{t_n}^{t,x} := 0 \ , \quad \widehat{Z}_{t_{i-1}}^{t,x} := \frac{1}{\Delta t} E\left[\widehat{Y}_{t_i}^{t,x} \Delta W_i | \mathcal{F}_{t_{i-1}}\right].$$

From here, they construct a continuous time scheme via the Martingale Representation Theorem. Taking advantage of the key result on the regularity of Z derived in [49], they show a similar form of L^2 convergance as in [49]. They then go on to consider some simulation based implementations of their discretization scheme.

In [17] (2005) Gobet, Lemor and Warin also consider loosely coupled BSDEs of the form

$$X_s^{t,x} = x + \int_t^s b(r, X_r^{t,x})dr + \int_t^s \sigma(r, X_r^{t,x})dW_r \ ,$$
$$Y_s^{t,x} = h(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})dr - \int_t^T Z_r^{t,x} dW_r.$$

They develop a least squares Monte Carlo numerical scheme based upon the implicit scheme of [8]. The basis of the scheme is to generate a number of forward paths (instances of $\widehat{X}$, the numerical approximation of X) and for each path, ω, to calculate $h(X_T(\omega))$. The scheme then proceeds by stepping backwards in time calculating conditional expectations (essentially with respect to $\mathcal{F}_{t_i}$) by regressing upon the values of $\widehat{X}$.

Remark. Least squares Monte Carlo schemes of this kind were made popular by [27] for pricing American options.

In other directions, we mention that in [7] (2008) Bouchard and Elie extend the approach of [49] to loosely coupled FBSDEs with jumps (as considered in [6]) and in [42] (2011), Richou extends the approach of [49] to loosely coupled FBSDEs with bounded terminal condition and driver f with quadratic growth in Z as introduced in [25].

3.3. BDSDEs

Backward doubly stochastic differential equations (BDSDEs) were introduced in 1994 by Pardoux and Peng in their paper [36]. In [36], they extend the setting of BSDEs in [35] by introducing a second noise - hence the term doubly stochastic. They then generalise the connection between BSDEs and PDEs to one between BDSDEs and stochastic PDEs (SPDEs). As this is material is fundamental to the thesis, we now provide a detailed review of [36].

Key Reference: [36] - Pardoux and Peng, 1994

The setting for [36] is as follows. On a probability space $(\Omega, \mathcal{F}, P)$ let $\{W_t; t \geq 0\}$ and $\{B_t; t \geq 0\}$ be mutually independent standard Brownian motions taking values in $\mathbb{R}^d$ and $\mathbb{R}^m$ respectively. Let $\mathcal{N}$ denote the P-null sets of $\mathcal{F}$ and fix $T > 0$. For each $s \in [0, T]$, define $\mathcal{F}_s := \mathcal{F}_s^W \vee \mathcal{F}_{s,T}^B$ where for any process $\{\phi_s\}$, $\mathcal{F}_{r,s}^\phi := \sigma\{\phi_u - \phi_r; u \in [r, s]\} \vee \mathcal{N}$ and $\mathcal{F}_s^\phi := \mathcal{F}_{0,s}^\phi$.

For any $n \in \mathbb{N}$, let $M^2([0,T]; \mathbb{R}^n)$ denote the set of n-dimensional $\{\mathcal{F}_t\}$-adapted processes $\{\phi_t; t \in [0,T]\}$ that satisfy $E\left[\int_0^T |\phi_t|^2 dt\right] < \infty$. Similarly, let $S^2([0,T]; \mathbb{R}^n)$ denote the set of n-dimensional $\{\mathcal{F}_t\}$-adapted processes $\{\phi_t; t \in [0,T]\}$ that satisfy $E\left[\sup_{0 \leq t \leq T} |\phi_t|^2\right] < \infty$.

They define the coefficients of the BDSDE as follows. Let

$$\begin{aligned} f &: \Omega \times [0,T] \times \mathbb{R}^k \times \mathbb{R}^{k \times d} \to \mathbb{R}^k \\ g &: \Omega \times [0,T] \times \mathbb{R}^k \times \mathbb{R}^{k \times d} \to \mathbb{R}^{k \times m} \end{aligned}$$

be jointly measurable and such that for any $(y, z) \in \mathbb{R}^k \times \mathbb{R}^{k \times d}$, $f(., y, z) \in M^2([0,T]; \mathbb{R}^k)$, $g(., y, z) \in M^2([0,T]; \mathbb{R}^{k \times m})$ and:

PP94.1 There exist constants $C > 0$ and $\alpha \in (0, 1)$ such that for any $(\omega, t) \in$

$\Omega \times [0,T]$ and (y_1, z_1), $(y_2, z_2) \in \mathbb{R}^k \times \mathbb{R}^{k\times m}$ the following inequalities hold:

$$|f(t,y_1,z_1) - f(t,y_2,z_2)|^2 \leq C(|y_1 - y_2|^2 + \|z_1 - z_2\|^2)$$
$$|g(t,y_1,z_1) - g(t,y_2,z_2)|^2 \leq C|y_1 - y_2|^2 + \alpha\|z_1 - z_2\|^2.$$

Given $\xi \in L^2((\Omega, \mathcal{F}_T, P); \mathbb{R}^k)$, they consider the BDSDE

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds + \int_t^T g(s, Y_s, Z_s)\overleftarrow{dB}_s - \int_t^T Z_s dW_s \tag{3.8}$$

for $t \in [0,T]$ where $\overleftarrow{dB}_s$ denotes the backward Itô integral with respect to B_s and prove the following existence and uniqueness result.

Theorem. *Suppose that condition PP94.1 holds. Then the BDSDE (3.8) has a unique solution* $(Y,Z) \in S^2([0,T]; \mathbb{R}^k) \times M^2([0,T]; \mathbb{R}^{k\times m})$.

To make the connection to SPDEs, they proceed as follows. Let functions $b \in C_b^3(\mathbb{R}^d, \mathbb{R}^d)$ and $\sigma \in C_b^3(\mathbb{R}^d, \mathbb{R}^{d\times d})$ and for each $s \in [0,T]$ and $x \in \mathbb{R}^d$ denote by $\{X_s^{t,x}; t \in [0,T]\}$ the unique strong solution of the SDE

$$X_t^{t,x} = x,$$
$$dX_s^{t,x} = b(X_s^{t,x})ds + \sigma(X_s^{t,x})dW_s \ , \quad s \in [t,T].$$

The functions f and g now take the less general forms

$$f(s,y,z) := f(s, X_s^{t,x}, y, z),$$
$$g(s,y,z) := g(s, X_s^{t,x}, y, z)$$

where

$$f : [0,T] \times \mathbb{R}^d \times \mathbb{R}^k \times \mathbb{R}^{k\times d} \to \mathbb{R}^k,$$
$$g : [0,T] \times \mathbb{R}^d \times \mathbb{R}^k \times \mathbb{R}^{k\times d} \to \mathbb{R}^{k\times m}$$

and they introduce the function $h : \mathbb{R}^d \to \mathbb{R}^k$ to give the BDSDE

$$Y_t^{t,x} = h(X_T^{t,x}) + \int_t^T f(s, X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})ds$$

$$+\int_t^T g(s, X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})\overleftarrow{dB}_s - \int_t^T Z_s^{t,x} dW_s. \tag{3.9}$$

Furthermore, they assume that for any $s \in [0,T]$, $(x,y,z) \to (f(x,y,z), g(x,y,z))$ is of class C^3, all derivatives are bounded on $[0,T] \times \mathbb{R}^d \times \mathbb{R}^k \times \mathbb{R}^{k\times d}$ and h is of class C^2. They first derive the following representation of Z.

Theorem. *The random field $\{Z_s^{t,x}; 0 \le t \le s \le T, x \in \mathbb{R}^d\}$ has an a.s. continuous version which is given by $Z_s^{t,x} = \nabla Y_s^{t,x}(\nabla X_s^{t,x})^{-1}\sigma(X_s^{t,x})$.*

Remark. Just as the equivalent representation of Z derived in [35] for the BSDE case is used in [49] to derive the regularity result for Z, we use this relation to derive our result on the regularity of Z in the BDSDE case.

Finally, in the following two theorems they relate the BDSDE (3.9) to the following system of quasilinear backward SPDEs:

$$\begin{aligned} u(t,x) = h(x) &+ \int_t^T \left[\mathcal{L}u(s,x) + f(s,x,u(s,x), \nabla u(s,x)\sigma(x))\right] ds \\ &+ \int_t^T g(s,x,u(s,x), \nabla u(s,x)\sigma(x))\overleftarrow{dB}_s \end{aligned} \tag{3.10}$$

where $u : \Omega \times \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^k$, $\mathcal{L}u = (\mathcal{L}u_1, \ldots, \mathcal{L}u_k)^T$ and

$$\mathcal{L}u = \sum_{i=1}^d b_i(x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^d a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j}\ , \quad (a_{ij}(x)) = \sigma\sigma^T(x).$$

Theorem. *Suppose that the above conditions hold and let $\{u(t,x); t \in [0,T], x \in \mathbb{R}^d\}$ be a random field such that $u(t,x)$ is $\mathcal{F}_{t,T}^B$-measurable for each (t,x), $u \in C^{0,2}([0,T] \times \mathbb{R}^d; \mathbb{R}^k)$ a.s, and u satisfies equation (3.10). Then $u(t,x) = Y_t^{t,x}$, where $\{(Y_s^{t,x}, Z_s^{t,x}); 0 \le t \le s \le T, x \in \mathbb{R}^d\}$ is the unique solution of the BDSDE (3.9).*

Theorem. *Suppose that the above conditions hold. Then $\{u(t,x) := Y_t^{t,x}; 0 \le t \le T, x \in \mathbb{R}^d\}$ is the unique classical solution of the system of backward SPDEs (3.10).*

In [5] (2001) Bally and Matoussi translate the loosely coupled BDSDEs of [36] to a weak formulation setting. The benefit of this approach is that it allows significant weakening on the conditions on the coefficients of the BDSDES. To save repetition, however, we do not provide an overview of this paper as the results of [5] are extended with significant overlap by the paper [50] which we cover next. We do note, however, that the key tool that the weak formulation setting enables and which makes the weakening of the conditions on the coefficients possible is an equivalence of norms result for the forward diffusion X (Result A.3 on page 148). It is this equivalence of norms result (or one very similar) that plays a key role in [5], [50] and this thesis. We also note that it was the paper [5] that first considered weak solutions of BDSDEs.

Key Reference: [50] - Zhang and Zhao, 2007

In [50], Zhang and Zhao extend the weak solution formulation of [5] to the infinite dimensional noise and infinite horizon case. Indeed, it is shown in [50] that the solution of an infinite horizon BDSDE at initial time t corresponds to the stationary solution of the corresponding SPDE.

To describe the results of [50] we proceed as follows. On a probability space $(\Omega, \mathcal{F}, P)$ let $\{W_t; t \geq 0\}$ and $\{B_t; t \geq 0\}$ be mutually independent stochastic processes with W a standard Brownian motion valued in $\mathbb{R}^d$ and B a Q-Wiener process valued on a separable Hilbert space U with countable base $\{e_j\}_{j=1}^{\infty}$ with $Qe_j = \lambda_j e_j$ and $\sum_{j=1}^{\infty} \lambda_j < \infty$. B has the expansion (see [12])

$$B_t = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \beta_j(t) e_j$$

where β_j, $j = 1, 2, \ldots$ are mutually independent real-valued Brownian motions on $(\Omega, \mathcal{F}, P)$.

Let $\mathcal{N}$ denote the P-null sets of $\mathcal{F}$ and let us fix $T > 0$. Define

$$\begin{aligned} \mathcal{F}_{t,T} &:= \mathcal{F}_t^W \vee \mathcal{F}_{t,T}^B \ , \quad t \in [0, T]; \\ \mathcal{F}_t &:= \mathcal{F}_t^W \vee \mathcal{F}_{t,\infty}^B \ , \quad t \geq 0. \end{aligned}$$

Here for any process $\{\phi_t; t \geq 0\}$, $\mathcal{F}_{s,t}^{\phi} := \sigma\{\phi_r - \phi_s; r \in [s,t]\} \vee \mathcal{N}$, $\mathcal{F}_t^{\phi} := \mathcal{F}_{0,t}^{\phi}$ and $\mathcal{F}_{t,\infty}^{\phi} := \bigvee_{T \geq 0} \mathcal{F}_{t,T}^{\phi}$. Let the weight function $\rho : \mathbb{R}^d \to \mathbb{R}$ be defined by $\rho(x) := K_\rho e^{\upsilon|x|}$ for constants $\upsilon < 0$ and $K_\rho > 0$ such that $\int_{\mathbb{R}^d} \rho(x) dx = 1$. For $s \geq t$, let $X_s^{t,x}$ be the solution of the SDE

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

where $b \in C_b^2(\mathbb{R}^d;\mathbb{R}^d)$ and $\sigma \in C_b^3(\mathbb{R}^d;\mathbb{R}^d \times \mathbb{R}^d)$.

Zhang and Zhao consider the following BDSDE with infinite-dimensional noise for $s \in [0,T]$:

$$\begin{aligned} Y_s^{t,x} &= h(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})dr \\ &\quad - \int_s^T g(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})\overleftarrow{dB}_r - \int_s^T \left\langle Z_r^{t,x}, dW_r \right\rangle \end{aligned} \tag{3.11}$$

where $\overleftarrow{dB}_r$ denotes the backward stochastic integral with respect to B. Here $h : \Omega \times \mathbb{R}^d \to \mathbb{R}$, $f : [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ and $g : [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to L^2_{U_0}(\mathbb{R})$ where $U_0 = Q^{1/2}(U)$. Setting $g_j := g\sqrt{\lambda_j}e_j : [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ then equation (3.11) is equivalent to

$$\begin{aligned} Y_s^{t,x} &= h(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})dr \\ &\quad - \sum_{j=1}^{\infty} \int_s^T g_j(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})\overleftarrow{dB}_r^j - \int_s^T \left\langle Z_r^{t,x}, dW_r \right\rangle. \end{aligned}$$

They prove weak existence and uniqueness for the following assumptions and definitions:

ZZ07.1 h is $\mathcal{F}^B_{T,\infty} \otimes \mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R})$ measurable and

$$E\left[\int_{\mathbb{R}^d} |h(x)|^2 \rho(x)dx\right] < \infty.$$

ZZ07.2 Functions f and g are Borel-measurable and there exists constants C, C_j, $\alpha_j \geq 0$ with $\sum_{j=1}^{\infty} C_j < \infty$ and $\sum_{j=1}^{\infty} \alpha_j < 1$ such that for any $s \in [0,T]$, and $x \in \mathbb{R}^d$

$$\begin{aligned} &|f(s,x,y_1,z_1) - f(s,x,y_2,z_2)|^2 \leq C(|y_1 - y_2|^2 + |z_1 - z_2|^2) \quad \text{and} \\ &|g_j(s,x,y_1,z_1) - g_j(s,x,y_2,z_2)|^2 \leq C_j|y_1 - y_2|^2 + \alpha_j|z_1 - z_2|^2. \end{aligned}$$

ZZ07.3

$$\int_0^T \int_{\mathbb{R}^d} |f(r,x,0,0)|^2 \rho(x)dxdr < \infty$$

and

$$\int_0^T \int_{\mathbb{R}^d} \|g(r,x,0,0)\|^2_{\mathcal{L}^2_{U_0}(\mathbb{R})} \rho(x) dx dr < \infty.$$

Definition. *Let $\mathbb{S}$ be a Hilbert space with norm $\|\cdot\|_{\mathbb{S}}$ and Borel σ-field $\mathcal{S}$. For $K \in \mathbb{R}^+$, denote by $M^{2,0}([t,T];\mathbb{S})$ the set of $\mathcal{B}([t,T]) \otimes \mathcal{F}/\mathcal{S}$ measurable random processes $\{\phi(s); s \in [t,T]\}$ with values on $\mathbb{S}$ satisfying:*

1. *$\phi(s) : \Omega \to \mathbb{S}$ is $\mathcal{F}_{s,T} \vee \mathcal{F}^B_{T,\infty}$ measurable for $s \in [t,T]$.*
2. *$E\left[\int_t^T \|\phi(s)\|^2_{\mathbb{S}} ds\right] < \infty$.*

Also denote by $S^{2,0}([t,T];\mathbb{S})$ the set of $\mathcal{B}([t,T]) \otimes \mathcal{F}/\mathcal{S}$ measurable random processes $\{\psi(s); s \in [t,T]\}$ with values on $\mathbb{S}$ satisfying:

1. *$\psi(s) : \Omega \to \mathbb{S}$ is $\mathcal{F}_{s,T} \vee \mathcal{F}^B_{T,\infty}$ measurable for $s \in [t,T]$ and $\psi(\cdot,\omega)$ is continuous a.s.*
2. *$E\left[\sup_{s\in[t,T]} \|\psi(s)\|^2_{\mathbb{S}}\right] < \infty$.*

Definition. *A pair of processes*

$$(Y^{t,\cdot}_{\cdot}, Z^{t,\cdot}_{\cdot}) \in S^{2,0}([0,T]; L^2_\rho(\mathbb{R}^d;\mathbb{R})) \times M^{2,0}([0,T]; L^2_\rho(\mathbb{R}^d;\mathbb{R}^d))$$

is called a solution of equation (3.11) if for any $\phi \in C^0_c(\mathbb{R}^d;\mathbb{R})$,

$$\begin{aligned}
\int_{\mathbb{R}_d} Y^{t,x}_s \phi(x) dx = & \int_{\mathbb{R}_d} h(X^{t,x}_T)\phi(x)dx + \int_s^T \int_{\mathbb{R}_d} f(r, X^{t,x}_r, Y^{t,x}_r, Z^{t,x}_r)\phi(x) dx dr \\
& - \sum_{j=1}^{\infty} \int_s^T \int_{\mathbb{R}_d} g_j(r, X^{t,x}_r, Y^{t,x}_r, Z^{t,x}_r)\phi(x) dx \overleftarrow{dB}^j_r \\
& - \int_s^T \left\langle \int_{\mathbb{R}_d} Z^{t,x}_r \phi(x) dx, dW_r \right\rangle \quad a.s.
\end{aligned}$$

Theorem 3.1. *Under conditions (ZZ07.1)-(ZZ07.4), equation (3.11) has a unique solution.*

They then consider the following SPDE and connect its weak solution (defined below) to the previously defined (weak) solution to BDSDE (3.11).

$$
\begin{aligned}
u(t,x) = h(x) &+ \int_t^T \left\{ \mathcal{L}u(s,x) + f(s,x,u(s,x),\sigma^T(x)\nabla u(s,x)) \right\} ds \\
&+ \int_t^T g(s,x,u(s,x),\sigma^T(x)\nabla u(s,x)) \overleftarrow{dB}_s
\end{aligned}
\tag{3.12}
$$

where $u : \Omega \times \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$, $\mathcal{L}u = (\mathcal{L}u_1, \ldots, \mathcal{L}u_k)^T$ with

$$
\mathcal{L}u = \sum_{i=1}^d b_i(x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^d a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j}, \quad (a_{ij}(x)) = \sigma\sigma^T(x).
$$

Definition. *A process u is called a weak solution (solution in $L^2_\rho(\mathbb{R}^d;\mathbb{R})$) of equation (3.12) if $(u, \sigma^T\nabla u) \in M^{2,0}([0,T]; L^2_\rho(\mathbb{R}^d;\mathbb{R})) \times M^{2,0}([0,T]; L^2_\rho(\mathbb{R}^d;\mathbb{R}^d))$ (where $(\sigma^T\nabla u)(s,x)$ is intepreted as $\sigma^T(x)\nabla u(s,x)$) and for an arbitrary $\psi \in C_c^{1,\infty}([0,T] \times \mathbb{R}^d;\mathbb{R})$,*

$$
\begin{aligned}
&\int_t^T \int_{\mathbb{R}^d} u(s,x) \frac{\partial \psi}{\partial s}(s,x) dx ds + \int_{\mathbb{R}^d} u(t,x)\psi(t,x) dx - \int_{\mathbb{R}^d} h(x)\psi(T,x) dx \\
&\quad - \frac{1}{2} \int_t^T \int_{\mathbb{R}^d} (\sigma^T(x)\nabla u(s,x))(\sigma^T(x)\nabla \psi(s,x)) dx ds \\
&\quad - \int_t^T \int_{\mathbb{R}^d} u(s,x) \mathrm{div}((b - \tilde{A})\psi)(s,x) dx ds \\
&= \int_t^T \int_{\mathbb{R}^d} f(s,x,u(s,x),\sigma^T(x)\nabla u(s,x)) \psi(s,x) dx ds \\
&\quad - \sum_{j=1}^\infty \int_t^T \int_{\mathbb{R}^d} g_j(s,x,u(s,x),\sigma^T(x)\nabla u(s,x)) \psi(s,x) dx \overleftarrow{dB}^j_s \quad a.s.
\end{aligned}
$$

$$
\textit{Here} \quad \tilde{A}_j := \frac{1}{2} \sum_{i=1}^d \frac{\partial a_{ij}(x)}{\partial x_i} \quad \textit{and} \quad \tilde{A} = (\tilde{A}_1, \ldots, \tilde{A}_d)^T.
$$

Theorem 3.2. *Assume conditions (ZZ07.1)-(ZZ07.4) hold and define $u(t,x) = Y_t^{t,x}$, where $(Y_s^{t,x}, Z_s^{t,x})$ is the solution of equation (3.11). Then $u(t,x)$ is the unique weak solution of (3.12).*

They then, for some $K > 0$, consider the following BDSDE with infinite-dimensional noise on infinite horizon

$$\begin{aligned} e^{-Ks}Y_s^{t,x} = &\int_s^\infty e^{-Kr}\left\{f(X_r^{t,x},Y_r^{t,x},Z_r^{t,x}) + KY_r^{t,x}\right\}dr \\ &- \int_s^T e^{-Kr} g(X_r^{t,x},Y_r^{t,x},Z_r^{t,x})\overleftarrow{dB}_r^j - \int_s^T e^{-Kr}\left\langle Z_r^{t,x}, dW_r\right\rangle. \end{aligned} \tag{3.13}$$

They assume broadly the same conditions as above along with the monotonicity condition

$$(y_1 - y_2)(f(s,x,y_1,z) - f(s,x,y_2,z)) \le -\mu|y_1-y_2|^2$$

for some $\mu > 0$ sufficiently large. They then show that $u(t,x) := Y_t^{t,x}$, where Y is the solution to (3.13), has a version which is a stationary solution to the SPDE

$$\begin{aligned} u(t,x) = u(T,x) &+ \int_t^T \left\{\mathcal{L}u(s,x) + f(x,u(s,x),\sigma^T(x)\nabla u(s,x))\right\}ds \\ &+ \int_t^T g(x,u(s,x),\sigma^T(x)\nabla u(s,x))\overleftarrow{dB}_s. \end{aligned} \tag{3.14}$$

In [51] (2010) Zhang and Zhao weaken the conditions placed on f in equations (3.13) and (3.14) of [50] from being Lipschitz in y to satisfying the linear growth condition

$$|f(x,y,z)| \le C(1+|y|+|z|)$$

for some constant C. They then again show that $u(t,x) := Y_t^{t,x}$ is a stationary solution to the SPDE (3.14).

In terms of approximations of BDSDEs, [2] (2013) Aman considers BDSDEs of the form

$$\begin{aligned} X_s^{t,x} &= x + \int_t^s b(r,X_r^{t,x})dr + \int_t^s \sigma(r,X_r^{t,x})dW_r \\ Y_s^{t,x} &= h(X_T^{t,x}) + \int_s^T f(r,X_r^{t,x},Y_r^{t,x},Z_r^{t,x})dr \\ &\quad - \int_s^T g(r,X_r^{t,x},Y_r^{t,x})\overleftarrow{dB}_r^j - \int_s^T \left\langle Z_r^{t,x}, dW_r\right\rangle \end{aligned}$$

where the coefficients are Lipschitz. Aman first derives a regularity result for Z of the form derived in [49] for BSDEs although we note that we believe Aman's proof to be incomplete (see the discussion preceding Lemma 6.8). He then defines an approximation scheme similar to the scheme defined in [8] for BSDEs

$$
\begin{aligned}
\widehat{Y}^{t,x}_{t_n} &:= h(\widehat{X}_{t_n}),\\
\widehat{Y}^{t,x}_{t_{i-1}} &:= E\left[\widehat{Y}^{t,x}_{t_i} + g(t_i, \widehat{X}^{t,x}_{t_i}, \widehat{Y}^{t,x}_{t_i})\Delta B_i | \mathcal{F}_{t_{i-1}}\right] + f(t_i, \widehat{X}^{t,x}_{t_{i-1}}, \widehat{Y}^{t,x}_{t_{i-1}}, \widehat{Z}^{t,x}_{t_{i-1}})\Delta t,\\
\widehat{Z}^{t,x}_{t_n} &:= 0,\\
\widehat{Z}^{t,x}_{t_{i-1}} &:= \frac{1}{\Delta t} E\left[\left(\widehat{Y}^{t,x}_{t_i} + g(t_i, \widehat{X}^{t,x}_{t_i}, \widehat{Y}^{t,x}_{t_i})\Delta B_i\right)\Delta W_i | \mathcal{F}_{t_{i-1}}\right]
\end{aligned}
$$

and provides a bound of the L^2 error of the scheme.

Remark. Whilst there is some overlap between the results of this thesis and [2] this is a consequence of both works following the approaches of [49] and [8]. The results of this thesis were derived independently of [2] and we note that our conditions are significantly weaker than those in [2].

4

Notation and Problem Statements

In the first section of this chapter we summarise the main notations used in this thesis. We note that this is not comprehensive - we omit some notations that are used only locally (for example within a single proof). In the second section, we state our assumptions and the questions that we will seek to address in this thesis.

4.1. Notation

- $\{\vec{f}, \vec{g}, \vec{h}\}$: step coefficients, page 47.
- $\{\check{f}, \check{g}, \check{h}\}$: Lipschitz coefficients, page 66.
- $\{\widetilde{f}, \widetilde{g}, \widetilde{h}\}$: smooth coefficients, page 73.
- $\widehat{X}$: Euler approximation of X, page 116.
- $\widehat{Y}$, $\widehat{Z}$: solution of discretization scheme, page 120.
- C_E: partition constant, page 47.
- C_M: truncation constant, page 47.
- C_G: maximum slope of Lipschitz coefficients, page 66.
- $\mathcal{T}$, $\mathcal{X}$: partitions of $[t, T]$ and $[-C_M, C_M]^d$ respectively, page 57.
- $\mathcal{T}^1$: partition of $[t, T]$, page 78.
- $\mathcal{T}^2$: partition of $[t, T]$, page 81.

- $\mathcal{T}^3$: partition of $[t, T]$, page 106.
- $\mathcal{T}^*$: partition of $[t, T]$, page 107.
- $N(\mathcal{X})$: number of elements in the partition $\mathcal{X}$, page 47.
- $\mathcal{E}[\cdot], \mathcal{E}_{\sup}[\cdot], \mathcal{E}_T[\cdot]$: functionals, page 57.
- $\mathfrak{B}^0_{\mathcal{T}}$: the set of boundary points of the partition $\mathcal{T}$, page 65.
- $\mathfrak{B}_{\mathcal{T}}$: the set of points in $[t, T]$ that are within $\frac{C_M}{C_G}$ of a boundary point of the partition $\mathcal{T}$, page 65.
- $\mathfrak{N}_k$: the set of neighbouring partition elements of the partition element $\mathcal{X}_k$, page 65.
- $\mathfrak{B}^0_{\mathcal{X}}$: the set of boundary points of the partition $\mathcal{X}$, page 65.
- $\mathfrak{B}_{\mathcal{X}}$: the set of points in $[-C_M, C_M]^d$ that are within $\frac{C_M}{C_G}$ of a boundary point of the partition $\mathcal{X}$, page 65.
- $\mathfrak{B}_{\mathcal{X},k}$: the set of points in the partition element $\mathcal{X}_k$ that are within $\frac{C_M}{C_G}$ of the element boundary, page 65.
- $\mathfrak{G}_{\mathcal{X}}, \mathfrak{G}_{\mathcal{X},k}$: subsets of $\mathbb{R}^d$, page 106.
- $\mu(A)$: measure function, page 69.
- $\mu_{\mathcal{X}}$: constant, page 79.
- $\Upsilon(K_{\mathfrak{B}})$: constant, page 110.

4.2. Problem Statements

We will now define the problems that we wish to solve. To this end we make the following assumptions based upon those in [50]. On a probability space $(\Omega, \mathcal{F}, P)$ let $\{W_t; t \geq 0\}$ and $\{B_t; t \geq 0\}$ be mutually independent standard Brownian motions taking values in $\mathbb{R}^d$ and $\mathbb{R}^m$ respectively. Let $\mathcal{N}$ denote the P-null sets of $\mathcal{F}$ and fix $T > 0$. For each $s \in [0, T]$, define $\mathcal{F}_s := \mathcal{F}^W_s \vee \mathcal{F}^B_{s,T}$ where for any process $\{\phi_s\}$, $\mathcal{F}^\phi_{r,s} := \sigma\{\phi_u - \phi_r; u \in [r, s]\} \vee \mathcal{N}$ and $\mathcal{F}^\phi_s := \mathcal{F}^\phi_{0,s}$. Let the weight function

$\rho : \mathbb{R}^d \to \mathbb{R}$ be defined by $\rho(x) := K_\rho e^{\upsilon|x|}$ for constants $\upsilon < 0$ and $K_\rho > 0$ such that $\int_{\mathbb{R}^d} \rho(x)dx = 1$.

Let us now define, X, the solution to our forward equation. We note that for the remainder of this thesis, X will always be defined as it is here.

Definition 4.1. *Define X to be the solution to the SDE*

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

where $b \in C_b^3(\mathbb{R}^d, \mathbb{R}^d)$ and $\sigma \in C_b^3(\mathbb{R}^d, \mathbb{R}^{d\times d})$.

Remark. We note that the conditions of Definition 4.1 are sufficient for the equivalence of norms result to hold (as stated in Result A.3 on page 148). This result is a powerful tool and is key for many of the arguments in this thesis.

Definition 4.2. *Define the Lipschitz constant*

$$L_X := \sup_{x_1 \neq x_2 \in \mathbb{R}^d} \left\{ \frac{|b(x_1) - b(x_2)|^2 \vee \|\sigma(x_1) - \sigma(x_2)\|^2}{|x_2 - x_2|^2} \right\}.$$

Remark. It follows from Definition 4.1 that $L_X < \infty$.

4.2.1. Finite Horizon Problem

We will consider the following BDSDE with $0 \leq t \leq s \leq T$ and $\overleftarrow{dB}_r^j$ denoting the backward Itô integral with respect to B_r^j:

$$\begin{aligned} Y_s^{t,x} = h(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})dr - \sum_{j=1}^m \int_s^T g_j(r, X_r^{t,x}, Y_r^{t,x})\overleftarrow{dB}_r^j \\ - \int_s^T \left\langle Z_r^{t,x}, dW_r \right\rangle. \end{aligned} \tag{4.1}$$

Remark. In what follows we will at times refer to the functions f and g_j as $f(r, \theta_r^{t,x})$ and $g_j(r, \theta_r^{t,x})$ with $\theta_r^{t,x} := (X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})$ even though g_j does not depend upon $Z_r^{t,x}$. We allow this abuse of notation, however, for conciseness. We will also use the abbreviation $g := \{g_1, \ldots, g_m\}$.

Definition 4.3. *Any set of coefficients $\{f, g, h\}$ are called* measurable coefficients *if they satisfy following conditions:*

M.1 *$h : \mathbb{R}^d \to \mathbb{R}$ is Borel-measurable and there exist constants $\gamma > 2$ and $C_h < \infty$ such that*

$$\int_{\mathbb{R}^d} |h(x)|^\gamma \rho(x) dx = C_h.$$

M.2 *$f : [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is Borel-measurable and for $j = 1, \dots, m$, $g_j : [0,T] \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ is Borel-measurable. Furthermore, there exist constants $\gamma > 2$ and $C_{f+g} < \infty$ such that*

$$\int_t^T \int_{\mathbb{R}^d} \left\{ |f(r,x,0,0)|^\gamma + \left(\sum_{j=1}^m |g_j(r,x,0)|^2 \right)^{\gamma/2} \right\} \rho(x) dx dr = C_{f+g}.$$

M.3 *There exists a positive constants L and $\epsilon_z > 0$ such that for any $s \in [0,T]$, $x \in \mathbb{R}^d$, y_1, $y_2 \in \mathbb{R}$ and z_1, $z_2 \in [-S_z, S_z]^d$*

$$\begin{aligned} &|f(s,x,y_1,z_1) - f(s,x,y_2,z_2)|^2 \le L(|y_1 - y_2|^2 + |z_1 - z_2|^2) \quad \textit{and} \\ &\sum_{j=1}^m |g_j(s,x,y_1) - g_j(s,x,y_2)|^2 \le L|y_1 - y_2|^2. \end{aligned}$$

M.4 *There exist positive constants L, S_z and ϵ_z such that for any $s \in [0,T]$, $x \in \mathbb{R}^d$, $y \in \mathbb{R}$ and $z_1, z_2 \notin [-S_z, S_z]^d$,*

$$|f(s,x,y,z_1) - f(s,x,y,z_2)|^2 \le L|z_1 - z_2|^{2-\epsilon_z}.$$

Remark. We note that by Theorem 3.1 on page 36 from the review of [50], the conditions of Definition 4.1 and Definition 4.3 are sufficient for the existence of a unique solution to BDSDE (4.1).

We illustrate these conditions with some examples:

Example. *Let $f : \mathbb{R} \to \mathbb{R}$, fix $S_z > 0$ and for $z \in \mathbb{R}$ define*

$$f(z) := \begin{cases} z \sin z + \sqrt{S_z}, & |z| \le S_z \\ \\ S_z \sin S_z + \sqrt{|z|}, & |z| > S_z. \end{cases}$$

Then for $|z_1|, |z_2| \leq S_z$

$$|f(z_1) - f(z_2)|^2 \leq S_z^2 |z_1 - z_2|^2$$

and

$$\begin{aligned} |f(z_1) - f(z_2)|^2 &= (\sqrt{|z_1|} - \sqrt{|z_2|})^2 \\ &\leq |z_1 - z_2| \end{aligned}$$

for $|z_1|, |z_2| \geq S_z$.

Example. *Let* $f : [0,T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, *fix* $S_z > 0$ *and for* $t, x, y, z \in [0,T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$ *define*

$$f(t, x, y, z) := y\mathbb{I}(x > t) + (z^4 \wedge S_z^4)\mathbb{I}(x < t).$$

Example. *Let* $h : \mathbb{R} \to \mathbb{R}$ *be defined by* $h := x^3 \mathbb{I}(|x| > 1)$.

Finite Horizon Problem:

- *Is it possible to approximate the solutions of BDSDE (4.1) with measurable coefficients as defined in Definition 4.3 using a time discretization scheme?*

4.2.2. Infinite Horizon Problem

We will consider the following infinite horizon BDSDE with $0 \leq t \leq s$ and $K > 0$ fixed:

$$\begin{aligned} e^{-Ks} Y_s^{t,x} = &\int_s^\infty e^{-Kr}(f(X_r^{t,x}, Y_r^{t,x}) + KY_r^{t,x})dr - \int_s^\infty e^{-Kr} g(X_r^{t,x}, Y_r^{t,x}) \overleftarrow{dB}_r \\ &- \int_s^\infty e^{-Kr} Z_r^{t,x} dW_r. \end{aligned} \tag{4.2}$$

Definition 4.4. *Given a positive constant* K, *called the* decay factor, *any pair of coefficients* f *and* g *are called* contractive coefficients *if they satisfy the following conditions:*

C.1 $f, g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *are Borel-measurable. Furthermore, there exists a constant* $C_{f+g} < \infty$ *such that* $|f(0,0)| + |g(0,0)| = C_{f+g}$.

C.2 *There exist positive constants L and $L_{g,y}$ such that for any $(x, y) \in \mathbb{R} \times \mathbb{R}$,*

$$\begin{aligned}
&|f(x_1, y_1) - f(x_2, y_2)|^2 \leq L\{|x_1 - x_2|^2 + |y_1 - y_2|^2\}, \\
&|g(x_1, y) - g(x_2, y)|^2 \leq L\{|x_1 - x_2|^2\}, \\
&|g(x, y_1) - g(x, y_2)|^2 \leq L_{g,y}|y_1 - y_2|^2.
\end{aligned}$$

C.3 *There exists a positive constant μ called the* contraction coefficient *satisfying $\mu > L + \frac{1}{2}(K + L_{g,y})$ such that for any $x \in \mathbb{R}^d$,*

$$(y_1 - y_2)(f(x, y_1) - f(x, y_2)) \leq -\mu(y_1 - y_2)^2.$$

Remark. We note that by the results in [50] on infinite horizon BDSDEs, the conditions of Definitions 4.1 and 4.4 are sufficient for the existence of a unique solution to the BDSDE (4.2).

Infinite Horizon Problem:

- *Is it possible to approximate the solutions of BDSDE (4.2) with contractive coefficients as defined in Definition 4.4 using a time discretization scheme?*

Remark. We will tackle the Finite Horizon Problem in Chapters 5 - 7 and the Infinite Horizon Problem in Chapter 8.

5 Continuous-Time Approximations

5.1. Introduction

In this chapter we approximate BDSDEs with measurable coefficients $\{f, g, h\}$ as defined in Definition 4.3 with BDSDEs with successively more regular coefficients.

In Section 5.2, BDSDEs with measurable coefficients are approximated by BDSDEs with step coefficients. The step coefficients are parameterized by two positive constants, C_E and C_M, and defined upon partitions of $[t, T]$ and $[-C_M, C_M]^d$. The constant C_E determines the minimum edge length of each partition element and the constant C_M determines the boundary of the spatial domain of the step coefficients and the maximum absolute value of the step coefficients. As a consequence, by making C_E smaller and C_M larger, one is able to find step coefficients such that the BDSDE approximation is more accurate.

In Section 5.3, BDSDEs with measurable coefficients are approximated by BDSDEs with Lipschitz coefficients. The Lipschitz coefficients are approximations of the step coefficients of Section 5.2 that are Lipschitz in t and x and are parameterized by a positive constant, C_G, which determines the maximum slope of the coefficients. As a consequence, by making C_G larger, one is able to find Lipschitz coefficients such that the BDSDE approximation is more accurate.

In Section 5.4, BDSDEs with measurable coefficients are approximated by BDSDEs with smooth coefficients. The smooth coefficients are simply constructed by mollifying the Lipschitz coefficients of Section 5.3.

Whilst the results of this chapter are potentially interesting in their own right, they are central to the approach taken in this thesis to approximate BDSDEs with measurable coefficients via discretization scheme. Indeed, in Chapter 7 a discretization scheme for BDSDEs with Lipschitz coefficients is constructed and so the approximation of BDSDEs with measurable coefficients by BDSDEs with Lipschitz coefficients is a fundamental step.

We note that the approximation of BDSDEs with measurable coefficients by BDSDEs with step coefficients derived in Section 5.2 is only used within Section 5.3 to obtain the approximation by BDSDEs with Lipschitz coefficients. The approximation of BDSDEs with measurable coefficients by BDSDEs with smooth coefficients derived in Section 5.4 is used in deriving the error estimate for the discretization scheme in Chapter 7 via the regularity result for smooth coefficients of Chapter 6.

5.2. BDSDEs with Step Coefficients

In this section we approximate measurable coefficients with step coefficients and show that the solution of the BDSDE (4.1) with step coefficients approximates the solution of the BDSDE (4.1) with measurable coefficients. To this end, we start with the following definitions.

Definition 5.1. *Given a positive constant C_E and a partition $\mathcal{X}$ of $[-C,C]^d$ for some fixed $C > 0$, we say that $\mathcal{X}$ is a* partition parameterised by C_E *or simply a* parameterised partition *if*

1. *$\mathcal{X}$ consists of a finite number of partition elements, $N(\mathcal{X})$, with each element a left-closed, right-open interval (or a closed interval when the right-most point of the interval is at the partition boundary). We enumerate the partition elements of $\mathcal{X}$ as $\mathcal{X}_1, \dots, \mathcal{X}_{N(\mathcal{X})}$.*

2. *Each element of $\mathcal{X}$ has edges of length at least C_E.*

We call C_E the partition constant*.*

Definition 5.2. *Given positive constants C_E and C_M and parameterised partitions $\mathcal{T}$ and $\mathcal{X}^\phi$ for $\phi = f, g, h$ of $[t,T]$ and $[-C_M, C_M]^d$ respectively, a set of coefficients $\{\vec{f}, \vec{g}, \vec{h}\}$ are* step coefficients parameterised by C_E and C_M *or simply* step coefficients *if they are measurable coefficients as defined in Definition 4.3 on page 42 and additionally satisfy the following conditions. We call C_M the* truncation constant*.*

S.1: *For* $k = 1, \ldots, N(\mathcal{X}^h)$ *there exist constants* $\vec{h}_k$ *such that* $|\vec{h}_k| \leq C_M$ *and*

$$\vec{h}(x) = \sum_{k=1}^{N(\mathcal{X}^h)} \mathbb{I}_{\mathcal{X}_k^h}(x)\vec{h}_k.$$

S.2: *There exist functions* $\vec{f}_{j,k} : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ *and for* $i = 1, \ldots, m$, $\vec{g}^i_{j,k} : \mathbb{R} \to \mathbb{R}$ *such that* $|\vec{f}_{j,k}(y,z)|, |\vec{g}^i_{j,k}(y)| \leq C_M$ *for all* $1 \leq j \leq N(\mathcal{T})$, $1 \leq k \leq N(\mathcal{X})$, $y \in \mathbb{R}$, $z \in \mathbb{R}^d$ *and*

$$\vec{f}(r,x,y,z) = \sum_{j=1}^{N(\mathcal{T})} \sum_{k=1}^{N(\mathcal{X}^f)} \mathbb{I}_{\mathcal{T}_j}(r)\mathbb{I}_{\mathcal{X}_k^f}(x)\vec{f}_{j,k}(y,z),$$
$$\vec{g}_i(r,x,y) = \sum_{j=1}^{N(\mathcal{T})} \sum_{k=1}^{N(\mathcal{X}^{g_i})} \mathbb{I}_{\mathcal{T}_j}(r)\mathbb{I}_{\mathcal{X}_k^{g_i}}(x)\vec{g}^i_{j,k}(y).$$

Definition 5.3. *Given measurable coefficients* $\{f, g, h\}$ *as defined in Definition 4.3 on page 42, positive constants* C_E *and* C_M *and parameterised partitions* $\mathcal{T}$ *and* $\mathcal{X}^\phi$ *for* $\phi = f, g, h$ *of* $[t, T]$ *and* $[-C_M, C_M]^d$ *respectively, we call coefficients* $\{\vec{f}, \vec{g}, \vec{h}\}$ *as constructed below* averaged step coefficients.

Let $\underline{h}$ *be the truncation of* h *defined by*

$$\underline{h}(x) := \begin{cases} (h(x) \vee -C_M) \wedge C_M, & x \in [-C_M, C_M]^d \\ \\ 0, & \text{otherwise.} \end{cases}$$

We define

$$\vec{h}(x) := \sum_{k=1}^{N(\mathcal{X}^h)} \mathbb{I}_{\mathcal{X}_k^h}(x)\vec{h}_k,$$

where for each k, $\vec{h}_k$ *is the* ρ*-weighted mean of* $\underline{h}$ *over* $\mathcal{X}_k$ *given by*

$$\vec{h}_k := \frac{1}{\mu_k} \int_{\mathcal{X}_k^h} \underline{h}(x)\rho(x)dx$$

and

$$\mu_k := \int_{\mathcal{X}_k^h} \rho(x)dx.$$

Similarly, we define for $r \in [t,T]$, $y \in \mathbb{R}$ and $z \in \mathbb{R}^d$,

$$\underline{f}(r,x,y,z) := \begin{cases} (f(r,x,y,z) \vee -C_M) \wedge C_M, & x \in [-C_M, C_M]^d \\ 0, & \textit{otherwise} \end{cases}$$

and

$$\vec{f}(r,x,y,z) := \sum_{j=1}^{N(\mathcal{T})} \sum_{k=1}^{N(\mathcal{X}^f)} \mathbb{I}_{\mathcal{T}_j}(r) \mathbb{I}_{\mathcal{X}_k^f}(x) \vec{f}_{j,k}(y,z)$$

where for each j,k,

$$\vec{f}_{j,k}(y,z) := \frac{1}{\mu_{j,k}} \int_{\mathcal{T}_j} \int_{\mathcal{X}_k^f} \underline{f}(r,x,y,z)\rho(x)dxdr$$

and

$$\mu_{j,k} := (t_{j+1} - t_j) \int_{\mathcal{X}_k^f} \rho(x)dx.$$

The construction for $\vec{g}$ is analogous.

Remark. The coefficients $\{\vec{f}, \vec{g}, \vec{h}\}$ as constructed in Definition 5.3 satisfy the conditions of Definition 5.2. In other words, averaged step coefficients are step coefficients. For example, the Lipschitz condition M.3 is satisfied by $\vec{f}$ since by the Cauchy-Schwarz inequality and for j,k such that $s \in \mathcal{T}_j$, $x' \in \mathcal{X}_k$, $y_1, y_2 \in \mathbb{R}$ and $z_1, z_2 \in [-S_z, S_z]^d$

$$\begin{aligned} &|\vec{f}(s,x',y_1,z_1) - \vec{f}(s,x',y_2,z_2)|^2 \\ &= |\vec{f}_{j,k}(y_1,z_1) - \vec{f}_{j,k}(y_2,z_2)|^2 \\ &= \frac{1}{\mu_{j,k}^2} \left| \int_{\mathcal{T}_j} \int_{\mathcal{X}_k} \underline{f}(r,x,y_1,z_1) - \underline{f}(r,x,y_2,z_2)\rho(x)dxdr \right|^2 \end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{\mu_{j,k}} \int_{\mathcal{T}_j} \int_{\mathcal{X}_k} |\underline{f}(r,x,y_1,z_1) - \underline{f}(r,x,y_2,z_2)|^2 \rho(x) dx dr \\
&\leq \frac{1}{\mu_{j,k}} \int_{\mathcal{T}_j} \int_{\mathcal{X}_k} L \left(|y_1 - y_2|^2 + |z_1 - z_2|^2 \right) \rho(x) dx dr \\
&= L \left(|y_1 - y_2|^2 + |z_1 - z_2|^2 \right).
\end{aligned}
$$

Note that in the above, we have also demonstrated that the Lipschitz condition M.3 is satisfied by the function $\vec{f}_{j,k}$.

In the remainder of this section, we will first prove a few technical lemmas before proving the main result of this section: Theorem 5.9. We begin by noting that it is well known that it is possible to approximate integrable Borel-measurable functions with step functions in the L^1 norm. Lemma 5.4 is a straightforward extension of this result and a sketch proof that is a slight adaptation of that given in [46] on page 131 for the standard L^1 result is provided for completeness.

Given a fixed $S > 0$, let us define $\mathcal{D}_Y(S) := [-S, S]$, $\mathcal{D}_Z(S) := [-S, S]^d$ and $\mathcal{D}(S) := [t, T] \times \mathbb{R}^d \times \mathcal{D}_Y(S) \times \mathcal{D}_Z(S)$. Let us also define the measure λ by its differential $d\lambda := \rho(x) dx \times dy \times dz \times dr$ and note that λ is a finite measure on $\mathcal{D}(S)$ that is absolutely continuous with respect to Lebesgue measure. We define the space $L^p_\lambda(\mathcal{D}(S), \mathbb{R})$ to be the set of Borel-measurable functions $f : \mathcal{D}(S) \to \mathbb{R}$ such that

$$
\int_{\mathcal{D}(S)} |f(r,x,y,z)|^p d\lambda < \infty.
$$

Lemma 5.4. *Let $f \in L^2_\lambda(\mathcal{D}(S), \mathbb{R})$. For each $\epsilon > 0$ we can find a constant $C_M < \infty$, a partition $\mathcal{A} := \{\mathcal{A}_1, \ldots, \mathcal{A}_{N(\mathcal{A})}\}$ of $\mathcal{D}(C_M, S) := [t, T] \times [-C_M, C_M]^d \times \mathcal{D}_Y(S) \times \mathcal{D}_Z(S)$ with a finite number of elements and constants F_j with $|F_j| \leq C_M$ such that the function*

$$
F(r,x,y,z) := \sum_{j=1}^{N(\mathcal{A})} \mathbb{I}_{\mathcal{A}_j}(r,x,y,z) F_j \quad \textit{satisfies}
$$

$$
\int_{\mathcal{D}(S)} |f(r,x,y,z) - F(r,x,y,z)|^2 d\lambda < \epsilon.
$$

Proof. (See [46] for additional detail). Firstly, we may approximate f with a bounded measurable function with compact support f_c, such that for some $C_M < \infty$, $|f_c| \leq C_M$ and $f_c(r,x,y,z) = 0$ for $x \notin [-C_M, C_M]^d$.

We may further approximate f_c with a simple function and since each simple function is the finite sum of multiples of indicator functions, it is sufficient to show that each indicator function can be approximated by the indicator function of a finite number of disjoint boxes (which will constitute the elements of the desired partition $\mathcal{A}$).

For each Borel subset B of $\mathcal{D}(C_M, S)$, we can find an open set O such that $O \supset B$ and $\lambda(O \setminus B) < \frac{\epsilon}{3}$. From the countable union of disjoint open balls making up O, we can select a finite number whose union form O_0 such that $\lambda(O \setminus O_0) < \frac{\epsilon}{3}$. Furthermore, we can find a finite number of boxes whose union form B_0 such that $\lambda(O_0 \Delta B_0) < \frac{\epsilon}{3}$. It then follows that $\lambda(B \Delta B_0) < \epsilon$. □

Remark. Since we only required a finite number of partition elements to construct an approximating step function in Lemma 5.4, we may claim that the elements have a minimum width of C_E for some constant $C_E > 0$.

Lemma 5.5. *Let measurable coefficients $\{f, g, h\}$ as defined in Definition 4.3 on page 42 and positive constants S and ϵ be given. Then there exist positive constants C_E and C_M, constants $\vec{F}_{j,k,l,m}$, $\vec{G}_{j,k,l}$ and $\vec{H}_k$ satisfying $|\vec{F}_{j,k,l,m}|$, $|\vec{G}_{j,k,l}|$, $|\vec{H}_k| \leq C_M$ for each $j = 1, \ldots, N(\mathcal{T})$, $k = 1, \ldots, N(\mathcal{X})$, $l = 1, \ldots, N(\mathcal{Y})$, $m = 1, \ldots, N(\mathcal{Z})$ and parameterised partitions $\mathcal{T}$, $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ of $[t, T]$, $[-C_M, C_M]^d$, $\mathcal{D}_Y(S)$ and $\mathcal{D}_Z(S)$ respectively such that the coefficients $\{\vec{F}, \vec{G}, \vec{H}\}$ defined by*

$$\begin{aligned}
\vec{F}(r, x, y, z) &:= \sum_{j,k,l,m} \mathbb{I}_{\mathcal{T}_j}(r)\mathbb{I}_{\mathcal{X}_k}(x)\mathbb{I}_{\mathcal{Y}_l}(y)\mathbb{I}_{\mathcal{Z}_m}(z)\vec{F}_{j,k,l,m}, \\
\vec{G}(r, x, y) &:= \sum_{j,k,l} \mathbb{I}_{\mathcal{T}_j}(r)\mathbb{I}_{\mathcal{X}_k}(x)\mathbb{I}_{\mathcal{Y}_l}(y)\vec{G}_{j,k,l}, \\
\vec{H}(x) &:= \sum_{k} \mathbb{I}_{\mathcal{X}_k}(x)\vec{H}_k
\end{aligned}$$

satisfy

$$\begin{aligned}
&\int_{\mathcal{D}(S)} |\vec{F}(r, x, y, z) - f(r, x, y, z)|^2 d\lambda < \epsilon, \\
&\int_t^T \int_{\mathcal{D}_Y(S)} \int_{\mathbb{R}^d} |\vec{G}(r, x, y) - g(r, x, y)|^2 \rho(x) dx dy dr < \epsilon, \\
&\int_{\mathbb{R}^d} |\vec{H}(x) - h(x)|^2 \rho(x) dx < \epsilon.
\end{aligned}$$

Proof. We just show the result for f as it is the most complicated case. Firstly, let us note that

$$\begin{aligned}\int_{\mathcal{D}(S)} |f(r,x,y,z)|^2 d\lambda &\le \int_{\mathcal{D}(S)} \left(|f(r,x,0,0)|^2 + L(|y|^2+|z|^2)\right) d\lambda \\ &\le (2S)^{1+d}(C_{f+g} + (T-t)L(1+d)S^2) \\ &< \infty\end{aligned}$$

and so $f \in L^2_\lambda(\mathcal{D}(S),\mathbb{R})$. By Lemma 5.4, for each $\epsilon > 0$ we can find constants $C_M < \infty$ and $C'_E > 0$, a corresponding partition $\mathcal{A} := \{\mathcal{A}_1,\ldots,\mathcal{A}_{N(\mathcal{A})}\}$ of $\mathcal{D}(C_M,S)$ and constants F_j with $|F_j| \le C_M$ such that the function

$$F(r,x,y,z) := \sum_{j=1}^{N(\mathcal{A})} \mathbb{I}_{\mathcal{A}_j}(r,x,y,z)F_j$$

satisfies

$$\int_{\mathcal{D}(S)} |f(r,x,y,z) - F(r,x,y,z)|^2 d\lambda < \epsilon.$$

Let us note that each partition element of $\mathcal{A}$ is a $(2d+2)$-dimensional box. As a consequence, to obtain our partition $\mathcal{T}$ of $[t,T]$ we simply need to project each element of $\mathcal{A}$ onto the line segment $[t,T]$. For example, suppose that

$$\begin{aligned}\mathcal{A}_j := \{&(t_{j,1},t_{j,2}] \times (x_{j,1},x_{j,2}] \times \ldots \times (x_{j,2d-1},x_{j,2d}] \times (y_{j,1},y_{j,2}] \\ &\times (z_{j,1},z_{j,2}] \times \ldots \times (z_{j,2d-1},z_{j,2d}]\}.\end{aligned}$$

Then the projection $P_t : \mathcal{A} \to [t,T]\times[t,T]$ is given by $P_t(\mathcal{A}_j) := \{t_{j,1},t_{j,2}\}$.

Let us define the set $\mathcal{T}' := \{t_0,\ldots,t_{N'}\} := \cup_j P_t(\mathcal{A}_j)$ with any duplicate entries removed. We then take $\mathcal{T} := \{\mathcal{T}_0\ldots,\mathcal{T}_{N(\mathcal{T})}\}$ where $N(\mathcal{T}) := N'-1$, $\mathcal{T}_0 := [t_0,t_1]$ and for $j = 1,\ldots,N(\mathcal{T})$, $\mathcal{T}_j := (t_j,t_{j+1}]$. We obtain $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ analogously.

Since there are a finite number of elements in $\mathcal{A}$ each with a non-zero minimum width, we have that each of $\mathcal{T}$, $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ also consists of a finite number of elements each with a non-zero minimum width. If we take C_E to be the smallest of these then $C_E > 0$ and the proof is complete. □

Lemma 5.6. *Given measurable coefficients $\{f,g,h\}$ as defined in Definition 4.3 on* *page 42 and positive constants S and ϵ, there exist constants C_E and C_M and parameterised partitions $\mathcal{T}$ and $\mathcal{X}^\phi$ for $\phi = f, g, h$ of $[t,T]$ and $[-C_M, C_M]^d$ respectively such that the averaged step coefficients $\{\vec{f}, \vec{g}, \vec{h}\}$ defined in Definition 5.3 satisfy*

$$\int_{\mathcal{D}(S)} |\vec{f}(r,x,y,z) - f(r,x,y,z)|^2 d\lambda < \epsilon,$$
$$\int_t^T \int_{\mathcal{D}_Y(S)} \int_{\mathbb{R}^d} |\vec{g}(r,x,y) - g(r,x,y)|^2 \rho(x) dxdydr < \epsilon \quad \textit{and}$$
$$\int_{\mathbb{R}^d} |\vec{h}(x) - h(x)|^2 \rho(x) dx < \epsilon.$$

Proof. We again just prove the result for f as it is the most complicated case. Let us, by Lemma 5.5, select values of C_E and C_M and corresponding partitions $\mathcal{T}$ and $\mathcal{X}$ of $[t,T]$ and $[-C_M, C_M]^d$ and partitions $\mathcal{Y}$ and $\mathcal{Z}$ of $\mathcal{D}_Y(S)$ and $\mathcal{D}_Z(S)$ such that there exist constants $\vec{F}_{j,k,l,m}$ satisfying $|\vec{F}_{j,k,l,m}| \leq C_M$ and

$$\vec{F}(r,x,y,z) := \sum_{j,k,l,m} \mathbb{I}_{\mathcal{T}_j}(r)\mathbb{I}_{\mathcal{X}_k}(x)\mathbb{I}_{\mathcal{Y}_l}(y)\mathbb{I}_{\mathcal{Z}_m}(z)\vec{F}_{j,k,l,m}$$

satisfies

$$\int_{\mathcal{D}(S)} |\vec{F}(r,x,y,z) - f(r,x,y,z)|^2 d\lambda < \epsilon'.$$

We do not now restrict $\mathcal{Y}$ and $\mathcal{Z}$ to have a minimum element width C_E but assume (since the respective domains $\mathcal{D}(Y)$ and $\mathcal{D}(Z)$ are bounded and so could further refine the partitions $\mathcal{Y}$ and $\mathcal{Z}$ if necessary) that $\mathcal{Y}$ and $\mathcal{Z}$ have a maximum element width of $\sqrt{\epsilon'}$.

Recalling that $\underline{f}$ denotes the C_M-truncation of f, let us now define for each j,k,l,m

$$\dot{f}(r,x,y,z) := \sum_{j,k,l,m} \mathbb{I}_{\mathcal{T}_j}(r)\mathbb{I}_{\mathcal{X}_k}(x)\mathbb{I}_{\mathcal{Y}_l}(y)\mathbb{I}_{\mathcal{Z}_m}(z)\dot{f}_{j,k,l,m}$$

where

$$\dot{f}_{j,k,l,m} := \frac{1}{\mu_{j,k,l,m}} \int_{\mathcal{T}_j} \int_{\mathcal{Z}_m} \int_{\mathcal{Y}_l} \int_{\mathcal{X}_k} \underline{f}(r,x,y,z) d\lambda$$

and

$$\mu_{j,k,l,m} := (t_{j+1} - t_j)(y_{l+1} - y_l) \int_{\mathcal{Z}_m} \int_{\mathcal{X}_k} \rho(x) dx dz.$$

Then we have that since the mean (which in this case is $\dot{f}$) is the best least squares estimator,

$$\begin{aligned} \int_{\mathcal{D}(S)} |\dot{f}(r,x,y,z) - f(r,x,y,z)|^2 d\lambda &\leq \int_{\mathcal{D}(S)} |\vec{F}(r,x,y,z) - f(r,x,y,z)|^2 d\lambda \\ &< \epsilon'. \end{aligned} \tag{5.1}$$

Recalling Definition 5.3, we have that for any j,k,l,m,

$$\dot{f}_{j,k,l,m} = \frac{1}{\mu_{l,m}} \int_{\mathcal{Z}_m} \int_{\mathcal{Y}_l} \vec{f}_{j,k}(y,z) dy dz \quad \text{where} \quad \mu_{l,m} := (y_{l+1} - y_l) \int_{\mathcal{Z}_m} dz.$$

As a consequence, for any j,k,l,m and $r \in \mathcal{T}_j$, $x \in \mathcal{X}_k$, $y \in \mathcal{Y}_l$, $z \in \mathcal{Z}_m$, we have that

$$\begin{aligned} |\vec{f}(r,x,y,z) - \dot{f}(r,x,y,z)|^2 &= |\vec{f}_{j,k}(y,z) - \dot{f}_{j,k,l,m}|^2 \\ &\leq \sup_{y_1,y_2 \in \mathcal{Y}_l, z_1,z_2 \in \mathcal{Z}_m} |\vec{f}_{j,k}(y_1,z_1) - \vec{f}_{j,k}(y_2,z_2)|^2 \\ &\leq L \sup_{y_1,y_2 \in \mathcal{Y}_l, z_1,z_2 \in \mathcal{Z}_m} (|y_1 - y_2|^2 + |z_1 - z_2|^2) \\ &\leq L(1+d)\epsilon' \end{aligned}$$

since $\mathcal{Y}$ and $\mathcal{Z}$ have a maximum element width of $\sqrt{\epsilon'}$. And so,

$$\int_{\mathcal{D}(S)} |\vec{f}(r,x,y,z) - \dot{f}(r,x,y,z)|^2 d\lambda \leq (T-t)(2S)^{1+d} L(1+d)\epsilon'. \tag{5.2}$$

Since ϵ' was arbitrary, the result follows by (5.1) and (5.2). □

Remark. Note that given measurable coefficients $\{f,g,h\}$ and domain $\mathcal{D}(S)$, the step coefficients $\{\vec{f},\vec{g},\vec{h}\}$ are uniquely generated by the constants C_M and C_E and the parameterised partitions $\mathcal{T}$ and $\mathcal{X}^\phi$ for $\phi = f,g,h$ of $[t,T]$ and $[-C_M, C_M]^d$.

Lemma 5.7. *Let measurable coefficients f and g as defined in Definition 4.3 on page 42, domain $\mathcal{D}(S)$ and constant $\epsilon > 0$ all be given. Then there exist positive constants C_E and C_M and parameterised partitions $\mathcal{T}$ and $\mathcal{X}^\phi$ for $\phi = f,g$ of $[t,T]$*

and $[-C_M, C_M]^d$ *such that the averaged step coefficients* $\vec{f}, \vec{g}$ *defined in Definition 5.3 satisfy*

$$\int_t^T \int_{\mathbb{R}^d} \sup_{y \in \mathcal{D}_Y(S), z \in \mathcal{D}_Z(S)} |\vec{f}(r,x,y,z) - f(r,x,y,z)|^2 \rho(x) dx dr < \epsilon \quad \textit{and}$$
$$\int_t^T \int_{\mathbb{R}^d} \sup_{y \in \mathcal{D}_Y(S)} |\vec{g}(r,x,y) - g(r,x,y)|^2 \rho(x) dx dr < \epsilon.$$

Proof. As usual, we just prove the result for f. Let us suppose that the result does not hold; this means that there exists an $\epsilon > 0$ such that for any values of $C_M < \infty$ and $C_E > 0$ and corresponding partitions $\mathcal{T}$ and $\mathcal{X}$ of $[t, T]$ and $[-C_M, C_M]^d$ we have that the uniquely generated averaged step coefficient $\vec{f}$ satisfies

$$\int_t^T \int_{\mathbb{R}^d} \sup_{y \in \mathcal{D}_Y(S), z \in \mathcal{D}_Z(S)} |\vec{f}(r,x,y,z) - f(r,x,y,z)|^2 \rho(x) dx dr \geq \epsilon.$$

Let us define the set

$$\mathcal{C} := \left\{ (r,x) \in [t,T] \times \mathbb{R}^d \,\middle|\, \sup_{y \in \mathcal{D}_Y(S), z \in \mathcal{D}_Z(S)} |\vec{f}(r,x,y,z) - f(r,x,y,z)|^2 \geq \frac{\epsilon}{2(T-t)} \right\}.$$

Then

$$\begin{aligned}\int_{\mathcal{C}^c} \sup_{y \in \mathcal{D}_Y(S), z \in \mathcal{D}_Z(S)} |\vec{f}(r,x,y,z) - f(r,x,y,z)|^2 \rho(x) dx dr &< \int_t^T \int_{\mathbb{R}^d} \frac{\epsilon}{2(T-t)} \rho(x) dx dr \\ &= \frac{\epsilon}{2}.\end{aligned}$$

As a consequence,

$$\int_{\mathcal{C}} \sup_{y \in \mathcal{D}_Y(S), z \in \mathcal{D}_Z(S)} |\vec{f}(r,x,y,z) - f(r,x,y,z)|^2 \rho(x) dx dr > \frac{\epsilon}{2}. \tag{5.3}$$

For fixed r and x, f and $\vec{f}$ are continuous functions of y and z which implies that for each $c := (r_c, x_c) \in \mathcal{C}$, there exist $y_c \in \mathcal{D}_Y(S)$ and $z_c \in \mathcal{D}_Z(S)$ such that

$$|\vec{f}(r_c, x_c, y_c, z_c) - f(r_c, x_c, y_c, z_c)|^2 = \sup_{y \in \mathcal{D}_Y(S), z \in \mathcal{D}_Z(S)} |\vec{f}(r_c, x_c, y, z) - f(r_c, x_c, y, z)|^2.$$

For $\delta > 0$ and $c \in \mathcal{C}$ let $\mathcal{Y}_c(\delta)$ denote any interval of length δ such that $y_c \in \mathcal{Y}_c(\delta)$

and $\mathcal{Z}_c(\delta)$ denote any cuboid of edge length δ such that $z_c \in \mathcal{Z}_c(\delta)$. Then since for fixed r and x, f and $\vec{f}$ are Lipschitz functions of y and z, we have that for $y \in \mathcal{Y}_c(\delta)$ and $z \in \mathcal{Z}_c(\delta)$

$$\begin{aligned}
&|\vec{f}(r_c, x_c, y, z) - f(r_c, x_c, y, z)| \\
&\geq |\vec{f}(r_c, x_c, y_c, z_c) - f(r_c, x_c, y_c, z_c)| - |\vec{f}(r_c, x_c, y_c, z_c) - \vec{f}(r_c, x_c, y, z)| \\
&\quad -|f(r_c, x_c, y_c, z_c) - f(r_c, x_c, y, z)| \\
&\geq |\vec{f}(r_c, x_c, y_c, z_c) - f(r_c, x_c, y_c, z_c)| - 2\delta\sqrt{L(1+d)}. \qquad (5.4)
\end{aligned}$$

If we choose

$$\delta \leq \frac{\sqrt{\epsilon}}{4\sqrt{2(T-t)L(1+d)}}$$

then we have by the definition of $\mathcal{C}$ that

$$\begin{aligned}
|\vec{f}(r_c, x_c, y_c, z_c) - f(r_c, x_c, y_c, z_c)| &= \sup_{y \in \mathcal{D}_Y(S), z \in \mathcal{D}_Z(S)} |\vec{f}(r_c, x_c, y, z) - f(r_c, x_c, y, z)| \\
&\geq \sqrt{\frac{\epsilon}{2(T-t)}} \\
&\geq 4\sqrt{L(1+d)}\delta.
\end{aligned}$$

As a consequence, it follows by (5.4) that for $y \in \mathcal{Y}_c(\delta)$ and $z \in \mathcal{Z}_c(\delta)$,

$$|\vec{f}(r_c, x_c, y, z) - f(r_c, x_c, y, z)|^2 \geq \frac{1}{2} \sup_{y \in \mathcal{D}_Y(S), z \in \mathcal{D}_Z(S)} |\vec{f}(r_c, x_c, y, z) - f(r_c, x_c, y, z)|^2.$$

We then have by (5.3) that

$$\begin{aligned}
&\int_{\mathcal{D}(S)} |\vec{f}(r, x, y, z) - f(r, x, y, z)|^2 d\lambda \\
&\geq \int_{\mathcal{C} \times \mathcal{D}_Y(S) \times \mathcal{D}_Z(S)} \mathbb{I}_{\mathcal{Y}_c(\delta)}(y) \mathbb{I}_{\mathcal{Z}_c(\delta)}(z) |\vec{f}(r, x, y, z) - f(r, x, y, z)|^2 d\lambda \\
&\geq \frac{1}{2} \int_{\mathcal{C} \times \mathcal{D}_Y(S) \times \mathcal{D}_Z(S)} \mathbb{I}_{\mathcal{Y}_c(\delta)}(y) \mathbb{I}_{\mathcal{Z}_c(\delta)}(z) \\
&\quad \sup_{y' \in \mathcal{D}_Y(S), z' \in \mathcal{D}_Z(S)} |\vec{f}(r, x, y', z') - f(r, x, y', z')|^2 d\lambda
\end{aligned}$$

$$= \frac{\delta^{1+d}}{2} \int_{\mathcal{C}} \sup_{y \in \mathcal{D}_Y(S), z \in \mathcal{D}_Z(S)} |\vec{f}(r,x,y,z) - f(r,x,y,z)|^2 \rho(x) dx dr$$

$$> \frac{\epsilon \delta^{1+d}}{4}.$$

As a consequence, there is no averaged step coefficient $\vec{f}$ that can approximate f in $L^2_\lambda(\mathcal{D}(S), \mathbb{R})$ which is a contradiction by Lemma 5.6. □

Definition 5.8. *For a random function $\psi : \Omega \times \mathbb{R}^d \to \mathbb{R}$, define*

$$\mathcal{E}\left[\psi(\omega, x)\right] := E\left[\int_{\mathbb{R}^d} \psi(\omega, x) \rho(x) dx\right].$$

For a random function $\psi : \Omega \times [t,T] \times \mathbb{R}^d \to \mathbb{R}$, define

$$\mathcal{E}_{\sup}\left[\psi(\omega, r, x)\right] := E\left[\sup_{t \le r \le T} \int_{\mathbb{R}^d} \psi(\omega, r, x) \rho(x) dx\right] \quad \textit{and}$$

$$\mathcal{E}_T\left[\psi(\omega, r, x)\right] := E\left[\int_t^T \int_{\mathbb{R}^d} \psi(\omega, r, x) \rho(x) dx dr\right].$$

Theorem 5.9. *Let (Y, Z) denote the solution to the BDSDE (4.1) with measurable coefficients $\{f, g, h\}$ as defined in Definition 4.3 on page 42. Then for each $\epsilon > 0$ there exist positive constants C_E and C_M and parameterised partitions $\mathcal{T}$ and $\mathcal{X}^\phi$ (for $\phi = f, g, h$) of $[t,T]$ and $[-C_M, C_M]^d$ such that $(\vec{Y}, \vec{Z})$, the solution to the BDSDE with averaged step coefficients $\{\vec{f}, \vec{g}, \vec{h}\}$ defined in Definition 5.3, satisfies*

$$\mathcal{E}_{\sup}\left[|\vec{Y}^{t,x}_r - Y^{t,x}_r|^2\right] + \mathcal{E}_T\left[|\vec{Z}^{t,x}_r - Z^{t,x}_r|^2\right] < \epsilon.$$

Proof. In the following proof, C will denote a generic constant that can vary from line to line but will not depend upon C_E or C_M. Let us denote by $\vec{\theta} := (X, \vec{Y}, \vec{Z})$. By Itô's formula (Result A.4) we have that for any $s \in [t,T]$ and a.e. $x \in \mathbb{R}^d$,

$$\begin{aligned} &|\vec{Y}^{t,x}_s - Y^{t,x}_s|^2 + \int_s^T |\vec{Z}^{t,x}_r - Z^{t,x}_r|^2 dr \\ &= |\vec{h}(X^{t,x}_T) - h(X^{t,x}_T)|^2 + 2\int_s^T (\vec{Y}^{t,x}_r - Y^{t,x}_r)(\vec{f}(r, \vec{\theta}^{t,x}_r) - f(r, \theta^{t,x}_r)) dr \\ &\quad + \sum_{j=1}^m \int_s^T |\vec{g}_j(r, \vec{\theta}^{t,x}_r) - g_j(r, \theta^{t,x}_r)|^2 dr - M^{t,x}_s \end{aligned}$$

where

$$
\begin{aligned}
M_s^{t,x} := & \ 2 \sum_{j=1}^{m} \int_s^T (\vec{Y}_r^{t,x} - Y_r^{t,x}) \left(\vec{g}_j(r, \vec{\theta}_r^{t,x}) - g_j(r, \theta_r^{t,x}) \right) \overleftarrow{dB}_r^j \\
& + 2 \int_s^T \left\langle (\vec{Y}_r^{t,x} - Y_r^{t,x})(\vec{Z}_r^{t,x} - Z_r^{t,x}), dW_r \right\rangle .
\end{aligned}
$$

Our strategy is to first apply Gronwall's inequality to transform the equation above into an amenable form. We will then prove the inequality without the sup using the equivalence of norms result (Result A.3). Finally, we will apply the Burkholder-–Davis–Gundy inequality to obtain the result with the sup.

Step 1: By Young's inequality, for any $\delta > 0$

$$
\begin{aligned}
& 2(\vec{Y}_r^{t,x} - Y_r^{t,x})(\vec{f}(r, \vec{\theta}_r^{t,x}) - f(r, \theta_r^{t,x})) \\
& \leq \frac{1}{\delta} |\vec{Y}_r^{t,x} - Y_r^{t,x}|^2 + \delta |\vec{f}(r, \vec{\theta}_r^{t,x}) - f(r, \theta_r^{t,x})|^2 \\
& \leq \frac{1}{\delta} |\vec{Y}_r^{t,x} - Y_r^{t,x}|^2 + 2\delta \left(|\vec{f}(r, \vec{\theta}_r^{t,x}) - \vec{f}(r, \theta_r^{t,x})|^2 + |\vec{f}(r, \theta_r^{t,x}) - f(r, \theta_r^{t,x})|^2 \right) \\
& \leq \frac{1}{\delta} |\vec{Y}_r^{t,x} - Y_r^{t,x}|^2 + 2\delta L \left(|\vec{Y}_r^{t,x} - Y_r^{t,x}|^2 + |\vec{Z}_r^{t,x} - Z_r^{t,x}|^2 \right) \\
& \quad + 2\delta |\vec{f}(r, \theta_r^{t,x}) - f(r, \theta_r^{t,x})|^2 .
\end{aligned}
$$

Similarly, we have that

$$
\sum_{j=1}^{m} |\vec{g}_j(r, \vec{\theta}_r^{t,x}) - g_j(r, \theta_r^{t,x})|^2 \leq 2L |\vec{Y}_r^{t,x} - Y_r^{t,x}|^2 + 2 \sum_{j=1}^{m} |\vec{g}_j(r, \theta_r^{t,x}) - g_j(r, \theta_r^{t,x})|^2 .
$$

It follows that

$$
\begin{aligned}
& |\vec{Y}_s^{t,x} - Y_s^{t,x}|^2 + \int_s^T |\vec{Z}_r^{t,x} - Z_r^{t,x}|^2 dr \\
& \leq |\vec{h}(X_T^{t,x}) - h(X_T^{t,x})|^2 + \int_s^T \left(\frac{1}{\delta} + 2(1+\delta)L \right) |\vec{Y}_r^{t,x} - Y_r^{t,x}|^2 \\
& \quad + 2\delta L |\vec{Z}_r^{t,x} - Z_r^{t,x}|^2 + 2\delta |\vec{f}(r, \theta_r^{t,x}) - f(r, \theta_r^{t,x})|^2 \\
& \quad + 2 \sum_{j=1}^{m} |\vec{g}_j(r, \theta_r^{t,x}) - g_j(r, \theta_r^{t,x})|^2 dr - M_s^{t,x} .
\end{aligned}
$$

Consequently, choosing $\delta > 0$ so that $2\delta L = 1 - \delta'$ for some $0 < \delta' < 1$, we have that

$$\begin{aligned}
&|\vec{Y}_s^{t,x} - Y_s^{t,x}|^2 + \delta' \int_s^T |\vec{Z}_r^{t,x} - Z_r^{t,x}|^2 dr \\
&\leq |\vec{h}(X_T^{t,x}) - h(X_T^{t,x})|^2 + C \int_s^T \Big(|\vec{Y}_r^{t,x} - Y_r^{t,x}|^2 + |\vec{f}(r, \theta_r^{t,x}) - f(r, \theta_r^{t,x})|^2 \\
&\quad + \sum_{j=1}^m |\vec{g}_j(r, \theta_r^{t,x}) - g_j(r, \theta_r^{t,x})|^2 \Big) \, dr - M_s^{t,x}.
\end{aligned} \tag{5.5}$$

Taking expectations, applying spatial integrals and Fubini's theorem and then multiplying throughout by $\frac{1}{\delta'}$ gives us (for a new constant C) that

$$\begin{aligned}
&\mathcal{E}\left[|\vec{Y}_s^{t,x} - Y_s^{t,x}|^2\right] + E\left[\int_s^T \int_{\mathbb{R}^d} |\vec{Z}_r^{t,x} - Z_r^{t,x}|^2 \rho(x) dx dr\right] \\
&\leq C \Bigg(\mathcal{E}\left[|\vec{h}(X_T^{t,x}) - h(X_T^{t,x})|^2\right] \\
&\quad + \mathcal{E}_T \left[|\vec{f}(r, \theta_r^{t,x}) - f(r, \theta_r^{t,x})|^2 + \sum_{j=1}^m |\vec{g}_j(r, \theta_r^{t,x}) - g_j(r, \theta_r^{t,x})|^2 \right] \\
&\quad + \int_s^T \mathcal{E}\left[|\vec{Y}_r^{t,x} - Y_r^{t,x}|^2\right] dr \Bigg).
\end{aligned} \tag{5.6}$$

Applying Gronwall's inequality to the function of s, $\mathcal{E}\left[|\vec{Y}_s^{t,x} - Y_s^{t,x}|^2\right]$, we have that for any $s \in [t, T]$ (and again for a new constant C),

$$\begin{aligned}
&\mathcal{E}\left[|\vec{Y}_s^{t,x} - Y_s^{t,x}|^2\right] \\
&\leq C \Bigg(\mathcal{E}\left[|\vec{h}(X_T^{t,x}) - h(X_T^{t,x})|^2\right] \\
&\quad + \mathcal{E}_T \left[|\vec{f}(r, \theta_r^{t,x}) - f(r, \theta_r^{t,x})|^2 + \sum_{j=1}^m |\vec{g}_j(r, \theta_r^{t,x}) - g_j(r, \theta_r^{t,x})|^2 \right] \Bigg).
\end{aligned} \tag{5.7}$$

Now,

$$\int_s^T \mathcal{E}\left[|\vec{Y}_r^{t,x} - Y_r^{t,x}|^2\right] dr \leq (T - t) \sup_{t \leq s \leq T} \mathcal{E}\left[|\vec{Y}_s^{t,x} - Y_s^{t,x}|^2\right]$$

and so applying equation (5.7) to equation (5.6) gives us that

$$\mathcal{E}\left[|\vec{Y}_s^{t,x} - Y_s^{t,x}|^2\right] + \mathcal{E}_T\left[|\vec{Z}_r^{t,x} - Z_r^{t,x}|^2\right]$$

$$\leq C \left(\mathcal{E}\left[|\vec{h}(X_T^{t,x}) - h(X_T^{t,x})|^2 \right] + \mathcal{E}_T\left[|\vec{f}(r,\theta_r^{t,x}) - f(r,\theta_r^{t,x})|^2 + \sum_{j=1}^{m} |\vec{g}_j(r,\theta_r^{t,x}) - g_j(r,\theta_r^{t,x})|^2 \right] \right). \tag{5.8}$$

Step 2: By the equivalence of norms (Result A.3), there exists a constant $C > 0$ such that

$$\mathcal{E}\left[|\vec{h}(X_T^{t,x}) - h(X_T^{t,x})|^2 \right] \leq C \int_{\mathbb{R}^d} |\vec{h}(x) - h(x)|^2 \rho(x) dx.$$

Since, by Lemma 5.6, $\vec{h}$ approximates h in $L^2_\rho(\mathbb{R}^d, \mathbb{R})$, for any $\epsilon' > 0$ we can find a truncation constant $C_M(h) \in (0,\infty)$, a partition constant $C_E(h) > 0$ and a corresponding partition $\mathcal{X}^h$ of $[-C_M(h), C_M(h)]^d$ so that

$$\int_{\mathbb{R}^d} |\vec{h}(x) - h(x)|^2 \rho(x) dx < \epsilon'. \tag{5.9}$$

For f and g we note that for any $S \in [S_z, \infty)$, $\mathcal{D}_Y(S) := [-S,S]$ and $\mathcal{D}_Z(S) := [-S,S]^d$,

$$\begin{aligned}
&|\vec{f}(r,\theta_r^{t,x}) - f(r,\theta_r^{t,x})|^2 \\
&= |\vec{f}(r,\theta_r^{t,x}) - f(r,\theta_r^{t,x})|^2 \left(\mathbb{I}_{\mathcal{D}_Y(S)}(Y_r^{t,x}) + \mathbb{I}_{\mathcal{D}_Y(S)^c}(Y_r^{t,x}) \right) \left(\mathbb{I}_{\mathcal{D}_Z(S)}(Z_r^{t,x}) + \mathbb{I}_{\mathcal{D}_Z(S)^c}(Z_r^{t,x}) \right).
\end{aligned}$$

We will proceed by just examining the cases $\{\mathcal{D}_Y(S) \times \mathcal{D}_Z(S)\}$ and $\{\mathcal{D}_Y(S)^c \times \mathcal{D}_Z(S)^c\}$ as the cases $\{\mathcal{D}_Y(S) \times \mathcal{D}_Z(S)^c\}$ and $\{\mathcal{D}_Y(S)^c \times \mathcal{D}_Z(S)\}$ follow similarly.

Case 1 - $\mathcal{D}_Y(S) \times \mathcal{D}_Z(S)$: Let us define

$$\begin{aligned}
F(r,x) &:= \sup_{y \in \mathcal{D}_Y(S), z \in \mathcal{D}_Z(S)} |\vec{f}(r,x,y,z) - f(r,x,y,z)|^2 \\
&\leq 2 \sup_{y \in \mathcal{D}_Y(S), z \in \mathcal{D}_Z(S)} |\vec{f}(r,x,y,z)|^2 + 2 \sup_{y \in \mathcal{D}_Y(S), z \in \mathcal{D}_Z(S)} |f(r,x,y,z)|^2 \\
&\leq 2C_M^2 + 4|f(r,x,0,0)|^2 + 4(d+1)LS^2
\end{aligned}$$

and so

$$\int_t^T \int_{\mathbb{R}^d} |F(r,x)| \rho(x) dx dr < \infty$$

and Result A.3 (equivalence of norms) can be applied. It then follows by the equivalence of norms and Lemma 5.7 that for any $\epsilon' > 0$,

$$\begin{aligned}
&\mathcal{E}_T\left[|\vec{f}(r,\theta_r^{t,x}) - f(r,\theta_r^{t,x})|^2 \mathbb{I}_{\mathcal{D}_Y(S)}(Y_r^{t,x})\mathbb{I}_{\mathcal{D}_Z(S)}(Z_r^{t,x})\right] \\
&\leq \mathcal{E}_T\left[F(r,X_r^{t,x})\right] \\
&\leq C\int_t^T\int_{\mathbb{R}^d} F(r,x)\rho(x)dxdr \\
&< \epsilon', \qquad (5.10)
\end{aligned}$$

if we choose (dependent upon S) a large enough value of truncation constant $C_M(f)$, a small enough value of partition constant $C_E(f) \in (0, C_E(h)]$ and corresponding partitions $\mathcal{T}^f$ and $\mathcal{X}^f$ of $[t,T]$ and $[-C_M(f), C_M(f)]^d$. Note that we will choose $\epsilon' > 0$ to be a fraction of the ϵ from the statement of our result.

Case 2 - $\mathcal{D}_Y(S)^c \times \mathcal{D}_Z(S)^c$: Let us first note that by Fubini's theorem and Chebyshev's inequality,

$$\begin{aligned}
\mathcal{E}_T\left[\mathbb{I}_{\mathcal{D}_Y(S)^c}(Y_r^{t,x})\right] &= \int_t^T\int_{\mathbb{R}^d} E\left[\mathbb{I}_{\mathcal{D}_Y(S)^c}(Y_r^{t,x})\right]\rho(x)dxdr \\
&= \int_t^T\int_{\mathbb{R}^d} P\left(|Y_r^{t,x}| > S\right)\rho(x)dxdr \\
&\leq \frac{1}{S^2}\int_t^T\int_{\mathbb{R}^d} E\left[|Y_r^{t,x}|^2\right]\rho(x)dxdr \\
&= \frac{1}{S^2}\mathcal{E}_T\left[|Y_r^{t,x}|^2\right].
\end{aligned}$$

Similarly, $\mathcal{E}_T\left[\mathbb{I}_{\mathcal{D}_Z^c}(Z_r^{t,x})\right] \leq \frac{1}{S^2}\mathcal{E}_T\left[|Z_r^{t,x}|^2\right]$. Now,

$$\begin{aligned}
&|\vec{f}(r,\theta_r^{t,x}) - f(r,\theta_r^{t,x})|^2\mathbb{I}_{\mathcal{D}_Y^c}(Y_r^{t,x})\mathbb{I}_{\mathcal{D}_Z^c}(Z_r^{t,x}) \\
&\leq 3\Big(|\vec{f}(r,\theta_r^{t,x}) - \vec{f}(r,X_r^{t,x},0,Z_r^{t,x})|^2 + |\vec{f}(r,X_r^{t,x},0,Z_r^{t,x}) - f(r,X_r^{t,x},0,Z_r^{t,x})|^2 \\
&\quad + |f(r,X_r^{t,x},0,Z_r^{t,x}) - f(r,\theta_r^{t,x})|^2\Big)\mathbb{I}_{\mathcal{D}_Y(S)^c}(Y_r^{t,x})\mathbb{I}_{\mathcal{D}_Z(S)^c}(Z_r^{t,x}) \\
&\leq 6L|Y_r^{t,x}|^2\mathbb{I}_{\mathcal{D}_Y(S)^c}(Y_r^{t,x}) + 3|\vec{f}(r,X_r^{t,x},0,Z_r^{t,x}) - f(r,X_r^{t,x},0,Z_r^{t,x})|^2\mathbb{I}_{\mathcal{D}_Z(S)^c}(Z_r^{t,x}).
\end{aligned}$$

By Results A.1 and A.5 (on the moments of X and Y) and the equivalence of norms, we know that $\mathcal{E}_T\left[|Y_r^{t,x}|^{2p}\right] < \infty$ for $2p \in [2,\gamma]$. Consequently, by Hölder's inequality we have that for $p \in (1, \frac{\gamma}{2}]$ and q conjugate to p

$$\begin{aligned}\mathcal{E}_T\left[|Y_r^{t,x}|^2 \mathbb{I}_{\mathcal{D}_Y(S)^c}(Y_r^{t,x})\right] &\leq \left(\mathcal{E}_T\left[|Y_r^{t,x}|^{2p}\right]\right)^{1/p}\left(\mathcal{E}_T\left[\mathbb{I}_{\mathcal{D}_Y(S)^c}(Y_r^{t,x})\right]\right)^{1/q} \\ &\leq \frac{C}{S^{2/q}}.\end{aligned}$$

Similarly, we have that since $S \geq S_z$

$$\begin{aligned}&|\vec{f}(r, X_r^{t,x}, 0, Z_r^{t,x}) - f(r, X_r^{t,x}, 0, Z_r^{t,x})|^2 \mathbb{I}_{\mathcal{D}_Z(S)^c}(Z_r^{t,x}) \\ &\leq 3\Big(|\vec{f}(r, X_r^{t,x}, 0, Z_r^{t,x}) - \vec{f}(r, X_r^{t,x}, 0, z(r,x)^*)|^2 \\ &\quad + |\vec{f}(r, X_r^{t,x}, 0, z(r,x)^*) - f(r, X_r^{t,x}, 0, z(r,x)^*)|^2 \\ &\quad + |f(r, X_r^{t,x}, 0, z(r,x)^*) - f(r, X_r^{t,x}, 0, Z_r^{t,x})|^2\Big)\mathbb{I}_{\mathcal{D}_Z(S)^c}(Z_r^{t,x}) \\ &\leq 6L|Z_r^{t,x}|^{2-\epsilon_z}\mathbb{I}_{\mathcal{D}_Z(S)^c}(Z_r^{t,x}) + 3|\vec{f}(r, X_r^{t,x}, 0, z(r,x)^*) - f(r, X_r^{t,x}, 0, z(r,x)^*)|^2\end{aligned}$$

where $z(r,x)^*$ denotes the unique closest point in $\mathcal{D}_Z(S)$ to $Z_r^{t,x}$ and so

$$|Z_r^{t,x} - z(r,x)^*| \leq |Z_r^{t,x}|.$$

Similarly to Case 1, by equivalence of norms, Lemma 5.7 and since $z(r,x)^* \in \mathcal{D}_Z(S)$,

$$\begin{aligned}\mathcal{E}_T\left[|\vec{f}(r, X_r^{t,x}, 0, z(r,x)^*) - f(r, X_r^{t,x}, 0, z(r,x)^*)|^2\right] &\leq \mathcal{E}_T\left[F(r, X_r^{t,x})\right] \\ &< \epsilon',\end{aligned}$$

for any ϵ' given the choices of $C_E(f)$, $C_M(f)$, $\mathcal{T}^f$ and $\mathcal{X}^f$ in Case 1.

Furthermore, by Hölder's inequality, selecting $p = \frac{2}{2-\epsilon_z}$ and q conjugate to p, we have that

$$\begin{aligned}\mathcal{E}_T\left[|Z_r^{t,x}|^{2-\epsilon_z}\mathbb{I}_{\mathcal{D}_Z^c}(Z_r^{t,x})\right] &\leq \left(\mathcal{E}_T\left[|Z_r^{t,x}|^2\right]\right)^{1/p}\left(\mathcal{E}_T\left[\mathbb{I}_{\mathcal{D}_Z(S)^c}(Z_r^{t,x})\right]\right)^{1/q} \\ &\leq \frac{C}{S^{2/q}}.\end{aligned}$$

Arguing similarly for the cases $\{\mathcal{D}_Y(S) \times \mathcal{D}_Z(S)^c\}$ and $\{\mathcal{D}_Y(S)^c \times \mathcal{D}_Z(S)\}$ we conclude that we can find $C_E(f)$, $C_M(f)$, $\mathcal{T}^f$ and $\mathcal{X}^f$ so that

$$\mathcal{E}_T\left[|\vec{f}(r, \theta_r^{t,x}) - f(r, \theta_r^{t,x})|^2\right] \leq \frac{C}{S^{2/q}} + \epsilon'. \tag{5.11}$$

The coefficients $\{g_j\}$ can be treated using the same argument.

As we progress through $\phi = f, g, h$, let us select non-increasing values of partition constant C_E, non-decreasing values of truncation constant C_M and successively refined partitions $\mathcal{T}$. Since the number of functions is finite, we are able to find $C_M < \infty$, $C_E > 0$ and corresponding partitions $\mathcal{T}$ and $\mathcal{X}^\phi$ of $[t,T]$ and $[-C_M, C_M]^d$ so that all of the above inequalities are satisfied for $\vec{\phi}$ each defined on $\mathcal{T}$ and $\mathcal{X}^\phi$. Consequently, by (5.8), (5.9), (5.10), (5.11) and choosing S large enough, the result without the sup now follows:

$$\mathcal{E}\left[|\vec{Y}_s^{t,x} - Y_s^{t,x}|^2\right] + \mathcal{E}_T\left[|\vec{Z}_r^{t,x} - Z_r^{t,x}|^2\right] \leq \epsilon. \tag{5.12}$$

Step 3: Applying spatial integrals and Fubini's theorem to equation (5.5), taking the sup over $s \in [t,T]$ and then taking expectations gives

$$\begin{aligned}
&\mathcal{E}_{\sup}\left[|\vec{Y}_s^{t,x} - Y_s^{t,x}|^2\right] \\
&\leq \mathcal{E}\left[|\vec{h}(X_T^{t,x}) - h(X_T^{t,x})|^2\right] \qquad (5.13)\\
&\quad + C\mathcal{E}_T\left[|\vec{Y}_r^{t,x} - Y_r^{t,x}|^2 + |\vec{f}(r,\theta_r^{t,x}) - f(r,\theta_r^{t,x})|^2 + \sum_{j=1}^m |\vec{g}_j(r,\theta_r^{t,x}) - g_j(r,\theta_r^{t,x})|^2\right] \\
&\quad + \mathcal{E}_{\sup}\left[-M_s^{t,x}\right].
\end{aligned}$$

By Step 2, it is sufficient to show that $\mathcal{E}_{\sup}[-M_s^{t,x}] < \epsilon$. Now, by the stochastic Fubini theorem we have that

$$\begin{aligned}
\int_{\mathbb{R}^d} M_s^{t,x}\rho(x)dx = {} & 2\sum_{j=1}^m \int_s^T \int_{\mathbb{R}^d} (\vec{Y}_r^{t,x} - Y_r^{t,x})(\vec{g}_j(r,\vec{\theta}_r^{t,x}) - g_j(r,\theta_r^{t,x}))\rho(x)dx\overleftarrow{dB}_r^j \\
&+2\int_s^T \left\langle \int_{\mathbb{R}^d} (\vec{Y}_r^{t,x} - Y_r^{t,x})(\vec{Z}_r^{t,x} - Z_r^{t,x})\rho(x)dx, dW_r \right\rangle.
\end{aligned}$$

Furthermore, by the Burkholder–Davis–Gundy inequality, the Cauchy-Schwarz inequality and Young's inequality, we have that for each $j = 1, \ldots, m$ and any $\delta > 0$

$$\begin{aligned}
&E\left[\sup_{t\leq s\leq T}\left|\int_s^T \int_{\mathbb{R}^d} (\vec{Y}_r^{t,x} - Y_r^{t,x})(\vec{g}_j(r,\vec{\theta}_r^{t,x}) - g_j(r,\theta_r^{t,x}))\rho(x)dx\overleftarrow{dB}_r^j\right|\right] \\
&\leq CE\left[\left(\int_t^T \left|\int_{\mathbb{R}^d} (\vec{Y}_r^{t,x} - Y_r^{t,x})(\vec{g}_j(r,\vec{\theta}_r^{t,x}) - g_j(r,\theta_r^{t,x}))\rho(x)dx\right|^2 dr\right)^{1/2}\right]
\end{aligned}$$

$$\begin{aligned}
&\leq CE\left[\left(\int_t^T\left(\int_{\mathbb{R}^d}|\vec{Y}_r^{t,x}-Y_r^{t,x}|^2\rho(x)dx\right)\right.\right.\\
&\qquad\left.\left.\left(\int_{\mathbb{R}^d}|\vec{g}_j(r,\vec{\theta}_r^{t,x})-g_j(r,\theta_r^{t,x})|^2\rho(x)dx\right)dr\right)^{1/2}\right]\\
&\leq CE\left[\left(\sup_{t\leq s\leq T}\int_{\mathbb{R}^d}|\vec{Y}_s^{t,x}-Y_s^{t,x}|^2\rho(x)dx\right)^{1/2}\right.\\
&\qquad\left.\left(\int_t^T\int_{\mathbb{R}^d}|\vec{g}_j(r,\vec{\theta}_r^{t,x})-g_j(r,\theta_r^{t,x})|^2\rho(x)dxdr\right)^{1/2}\right]\\
&\leq C\delta\mathcal{E}_{\sup}\left[|\vec{Y}_s^{t,x}-Y_s^{t,x}|^2\right]+\frac{C}{\delta}\mathcal{E}_T\left[|\vec{g}_j(r,\vec{\theta}_r^{t,x})-g_j(r,\theta_r^{t,x})|^2\right]\\
&\leq C\delta\mathcal{E}_{\sup}\left[|\vec{Y}_s^{t,x}-Y_s^{t,x}|^2\right]+\frac{2C}{\delta}\mathcal{E}_T\left[L|\vec{Y}_r^{t,x}-Y_r^{t,x}|^2+|\vec{g}_j(r,\theta_r^{t,x})-g_j(r,\theta_r^{t,x})|^2\right].
\end{aligned}$$

The $\mathcal{E}_T\left[|\vec{Y}_r^{t,x}-Y_r^{t,x}|^2\right]$ term can be handled by Fubini's theorem and (5.12). Additionally, we have already argued in Step 2 that

$$\mathcal{E}_T\left[|\vec{g}_j(r,\theta_r^{t,x})-g_j(r,\theta_r^{t,x})|^2\right]\leq\frac{C}{S^{2/q}}+\epsilon'.$$

We can similarly show that

$$\begin{aligned}
&E\left[\sup_{t\leq s\leq T}\left|\int_s^T\left\langle\int_{\mathbb{R}^d}(\vec{Y}_r^{t,x}-Y_r^{t,x})(\vec{Z}_r^{t,x}-Z_r^{t,x})\rho(x)dx,dW_r\right\rangle\right|\right]\\
&\leq C\delta\mathcal{E}_{\sup}\left[|\vec{Y}_s^{t,x}-Y_s^{t,x}|^2\right]+\frac{C}{\delta}\mathcal{E}_T\left[|\vec{Z}_r^{t,x}-Z_r^{t,x}|^2\right]\\
&\leq C\delta\mathcal{E}_{\sup}\left[|\vec{Y}_s^{t,x}-Y_s^{t,x}|^2\right]+\epsilon
\end{aligned}$$

by equation (5.12). The result now follows by choosing a small enough δ so that the $\mathcal{E}_{\sup}\left[|\vec{Y}_r^{t,x}-Y_r^{t,x}|^2\right]$ terms can be moved to the left hand side of equation (5.13). □

Definition 5.10. *Given measurable coefficients $\{f,g,h\}$ as defined in Definition 4.3 on page 42 and averaged step coefficients $\{\vec{f},\vec{g},\vec{h}\}$ as defined in Definition 5.3 with partition constant C_E, truncation constant C_M and parameterised partitions $\mathcal{T}$ and $\mathcal{X}^\phi$ (for $\phi=f,g,h$), define the function $\beta:\mathbb{R}^+\times\mathbb{R}^+\to\mathbb{R}^+$ to represent the error term of Theorem 5.9:*

$$\beta(C_E,C_M):=\mathcal{E}_{\sup}\left[|\vec{Y}_r^{t,x}-Y_r^{t,x}|^2\right]+\mathcal{E}_T\left[|\vec{Z}_r^{t,x}-Z_r^{t,x}|^2\right].$$

Remark. Given a required accuracy of ϵ and measurable coefficients $\{f, g, h\}$ as defined in Definition 4.3 on page 42 as inputs, we have constants C_M and C_E, corresponding partitions $\mathcal{T}$ and $\mathcal{X}^\phi$ for $\phi = f, g, h$ of $[t, T]$ and $[-C_M, C_M]^d$ respectively and step coefficients $\{\vec{f}, \vec{g}, \vec{h}\}$ as defined in Definition 5.2 as outputs. In subsequent sections, we will assume that these outputs have now been fixed and will make use only of Theorem 5.9.

Remark. Whilst we have shown that we can make ϵ arbitrarily small by making C_E small and C_M large, the relation between the constants C_E and C_M and the error function β is not straighforward. Indeed, the relation depends on the measurable coefficients $\{f, g, h\}$ which can be highly non-linear.

5.3. BDSDEs with Lipschitz Coefficients

In this section we approximate step coefficients with Lipschitz coefficients and show that the solution of the BDSDE with Lipschitz coefficients approximates the solution of the BDSDE with step coefficients. Combining this with the main result of Section 5.2 we conclude that the solution of the BDSDE with Lipschitz coefficients approximates the solution of the BDSDE with measurable coefficients. To this end, we start with the following definitions.

Definition 5.11. *Let a truncation constant C_M and partition constant C_E be given and let $\mathcal{T}$ and $\mathcal{X}$ be parameterised partitions as defined in Definition 5.1 of $[t, T]$ and $[-C_M, C_M]^d$ respectively. Given a constant $C_G \geq \frac{2C_M}{C_E}$ we define the following sets (where* dist *denotes the standard Euclidean distance between sets):*

$$
\begin{aligned}
\mathfrak{B}^0_{\mathcal{T}} &:= \{t_0, t_1, \ldots, t_{N(\mathcal{T})}\}, \\
\mathfrak{B}_{\mathcal{T}} &:= \bigcup_k \left\{ s \in [t, T] \,\middle|\, |s - t_k| \leq \frac{C_M}{C_G} \right\}, \\
\mathfrak{N}_k &:= \{\mathcal{X}_j \in \mathcal{X} \,|\mathrm{dist}(\mathcal{X}_j, \mathcal{X}_k) = 0\}, \\
\mathfrak{B}^0_{\mathcal{X}} &:= \bigcup_k \{x \in \overline{\mathcal{X}}_k \,|\mathrm{dist}(x, \mathfrak{N}_k) = 0\}, \\
\mathfrak{B}_{\mathcal{X}} &:= \left\{ x \in [-C_M, C_M]^d \,\middle|\, \mathrm{dist}(x, \mathfrak{B}^0_{\mathcal{X}}) \leq \frac{C_M}{C_G} \right\} \\
\mathfrak{B}_{\mathcal{X},k} &:= \mathfrak{B}_{\mathcal{X}} \cap \mathcal{X}_k.
\end{aligned}
$$

In words these are as follows:
$\mathfrak{B}^0_{\mathcal{T}}$ *is the set of boundary points of the partition* $\mathcal{T}$.
$\mathfrak{B}_{\mathcal{T}}$ *is the set of points in* $[t,T]$ *that are within* $\frac{C_M}{C_G}$ *of a boundary point of the partition* $\mathcal{T}$.
$\mathfrak{N}_k$ *is the set of neighbouring partition elements of the partition element* $\mathcal{X}_k$.
$\mathfrak{B}^0_{\mathcal{X}}$ *is the set of boundary points of the partition* $\mathcal{X}$.
$\mathfrak{B}_{\mathcal{X}}$ *is the set of points in* $[-C_M,C_M]^d$ *that are within* $\frac{C_M}{C_G}$ *of a boundary point of the partition* $\mathcal{X}$.
$\mathfrak{B}_{\mathcal{X},k}$ *is the set of points in the partition element* $\mathcal{X}_k$ *that are within* $\frac{C_M}{C_G}$ *of the element boundary.*

Remark. We note that $l(\mathfrak{B}_{\mathcal{T}})$ is $\sim \frac{C_M}{C_G}$ and $\mu(\mathfrak{B}_{\mathcal{X}})$ is $\sim \left(\frac{C_M}{C_G}\right)^d$. For motivation on why we define these sets, see the remark after the following definition.

Definition 5.12. *Let* $\{\vec{f},\vec{g},\vec{h}\}$ *denote step coefficients as defined in Definition 5.2 with truncation constant* C_M*, partition constant* C_E *and parameterised partitions* $\mathcal{T}$ *and* $\mathcal{X}^\phi$ *for* $\phi = f,g,h$ *of* $[t,T]$ *and* $[-C_M,C_M]^d$ *respectively. Let* C_G *be a constant satisfying* $C_G \geq \frac{2C_M}{C_E}$. *Then* $\{\check{f},\check{g},\check{h}\}$ *are called* Lipschitz approximations with maximum slope C_G of $\{\vec{f},\vec{g},\vec{h}\}$ *or simply* Lipschitz coefficients *if they are measurable coefficients as defined in Definition 4.3 on page 42 and additionally satisfy the following conditions:*

1. *For any* $x_1, x_1 \in \mathbb{R}^d$
$$|\check{h}(x_1) - \check{h}(x_2)| \leq C_G|x_1 - x_2|.$$

2. *For any* $t_1,t_2 \in [0,T]$, $x,z \in \mathbb{R}^d$, $y \in \mathbb{R}$ *and* $j = 1,\dots,m$
$$|\check{f}(t_1,x,y,z) - \check{f}(t_2,x,y,z)| \leq C_G|t_1 - t_2|$$
and
$$|\check{g}_j(t_1,x,y,z) - \check{g}_j(t_2,x,y,z)| \leq C_G|t_1 - t_2|.$$

3. *For any* $t \in [0,T]$, $x_1,x_2,z \in \mathbb{R}^d$, $y \in \mathbb{R}$ *and* $j = 1,\dots,m$
$$|\check{f}(t,x_1,y,z) - \check{f}(t,x_2,y,z)| \leq C_G|x_1 - x_2|$$

and

$$|\check{g}_j(t, x_1, y, z) - \check{g}_j(t, x_2, y, z)| \leq C_G |t_1 - t_2|.$$

4. *For all* $x \notin \mathfrak{B}_{\mathcal{X}^h}$, $\check{h}(x) = \vec{h}(x)$. *For all* $t \notin \mathfrak{B}_{\mathcal{T}}$, $x \notin \mathfrak{B}_{\mathcal{X}^f}$, $y \in \mathbb{R}$ *and* $z \in \mathbb{R}^d$, $\check{f}(t, x, y, z) = \vec{f}(t, x, y, z)$. *For all* $t \notin \mathfrak{B}_{\mathcal{T}}$, $x \notin \mathfrak{B}_{\mathcal{X}^g}$ *and* $y \in \mathbb{R}$, $\check{g}(t, x, y) = \vec{g}(t, x, y)$.

Remark. The final condition of the above definition requires that the sets $\mathfrak{B}_{\mathcal{T}}$ and $\mathfrak{B}_{\mathcal{X}^{f,g,h}}$ contain the parts of the domain $[t, T] \times \mathbb{R}^d$ upon which the step coefficients $\{\vec{f}, \vec{g}, \vec{h}\}$ and the Lipschitz coefficients $\{\check{f}, \check{g}, \check{h}\}$ are not equal. As a consequence, the regions where the Lipschitz coefficient have a slope of C_G are restricted to $\mathfrak{B}_{\mathcal{T}}$ and $\mathfrak{B}_{\mathcal{X}^{f,g,h}}$. As we will see, to make the Lipschitz approximations more accurate, we must make C_G larger. Since the term C_G is present in error terms in subsequent chapters, it is desirable to be able to quantify the effect of increasing C_G which is precisely what the sets $\mathfrak{B}_{\mathcal{T}}$ and $\mathfrak{B}_{\mathcal{X}^{f,g,h}}$ allow.

Remark. We note that given a set of Lipschitz coefficients $\{\check{f}, \check{g}, \check{h}\}$ as defined above, there is implicitly a fixed partition constant C_E, truncation constant C_M, maximum slope C_G and parameterised partitions $\mathcal{T}$ and $\mathcal{X}^\phi$ for $\phi = f, g, h$ of $[t, T]$ and $[-C_M, C_M]^d$ respectively. Furthermore, the sets defined in Definition 5.11 are also fixed.

Example. *As an example, we give an explicit definition of a possible choice for* $\check{h}$ *when* $d = 1$ *(see also Figure 5.1).*

$$\check{h}(x) := \begin{cases} \vec{h}(x_{k-1}) + \left(\vec{h}(x_k) - \vec{h}(x_{k-1})\right) \frac{C_G(x - x_k) + C_M}{2C_M}, & x \in \left[x_k - \frac{C_M}{C_G}, x_k + \frac{C_M}{C_G}\right) \\ \vec{h}(x_k), & x \in \left[x_k + \frac{C_M}{C_G}, x_{k+1} - \frac{C_M}{C_G}\right) \\ \vec{h}(x_k) + \left(\vec{h}(x_{k+1}) - \vec{h}(x_k)\right) \frac{C_G(x - x_{k+1}) + C_M}{2C_M}, & x \in \left[x_{k+1} - \frac{C_M}{C_G}, x_{k+1} + \frac{C_M}{C_G}\right). \end{cases}$$

Remark. We require that $C_G \geq \frac{2C_M}{C_E}$ otherwise it may not be possible to perform the interpolation.

Lemma 5.13. *Let step coefficients* $\{\vec{f}, \vec{g}, \vec{h}\}$ *as defined in Definition 5.2 with truncation constant* $C_M < \infty$*, partition constant* $C_E > 0$ *and corresponding partitions* $\mathcal{T}$

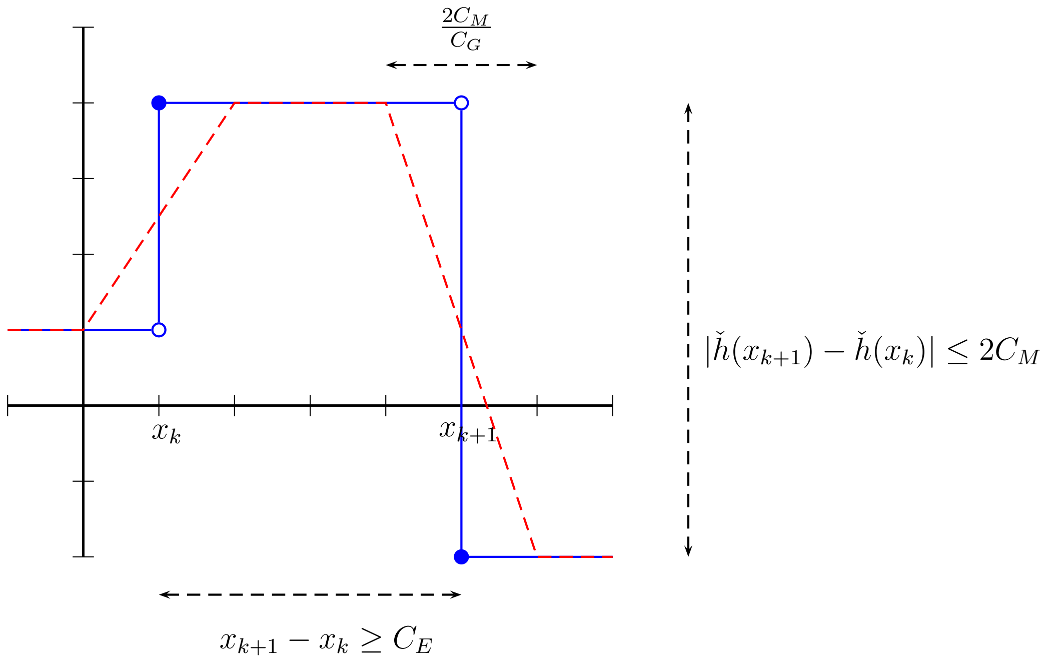

Figure 5.1.: h (solid line) and $\widehat{h}$ (broken line) for $d = 1$. We note that the interpolation is split evenly between adjacent partition elements and that we can choose to interpolate with different slopes between elements as long as they do not exceed C_G.

and $\mathcal{X}^\phi$ for $\phi = f, g, h$ of $[t, T]$ and $[-C_M, C_M]^d$ respectively be given. For each constant $C_G \geq \frac{2C_M}{C_E}$ it is possible to find Lipschitz approximations $\{\check{f}, \check{g}, \check{h}\}$ of $\{\vec{f}, \vec{g}, \vec{h}\}$ with maximum slope C_G.

Proof. Let us first consider the case $\check{h}$. For all $x \in \mathfrak{B}^c_{\mathcal{X}^h}$, define $\check{h}(x) := \vec{h}(x)$. This defines $\check{h}(x)$ on unconnected cuboids of $\mathbb{R}^d$. It remains to define $\check{h}$ for $x \in \mathfrak{B}_{\mathcal{X}^h}$ in such a way that $\check{h}$ is a Lipschitz function of x with Lipschitz constant C_G.

Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be adjacent partition elements of $\mathcal{X}^h$. By the definition of $\mathfrak{B}_{\mathcal{X}^h}$, there is a distance of at least $\frac{2C_M}{C_G}$ between the points of $\mathcal{X}_1 \cap \mathfrak{B}^c_{\mathcal{X}^h}$ and the points of $\mathcal{X}_2 \cap \mathfrak{B}^c_{\mathcal{X}^h}$.

Suppose that $x_1 \in \mathcal{X}_1$ is on the border of $\mathfrak{B}_{\mathcal{X}^h}$ that is facing $\mathcal{X}_2$ and similarly $x_2 \in \mathcal{X}_2$ is on the border of $\mathfrak{B}_{\mathcal{X}^h}$ that is facing $\mathcal{X}_1$. Then since $|\vec{h}(x_1) - \vec{h}(x_2)| \leq 2C_M$ we can choose $\check{h}$ such that it linearly interpolates between x_1 and x_2 in such a manner that $\check{h}$ is a Lipschitz function with Lipschitz constant C_G. For simplicity, let us always interpolate evenly between adjacent partitions (see Figure 5.1 for an example of this). As a consequence, since we can always construct such a function $\check{h}$ and since $\check{h}$ satisfies the conditions of Definition 5.12, the proof for $\check{h}$ is complete.

Let us now consider $\check{f}$ ($\check{g}$ is the same). We note that since $(t, x) \in [0, T] \times \mathbb{R}^d \subset \mathbb{R}^{d+1}$ and d was arbitrary, there is no difficulty in adding a dependence upon t. For $\check{f}$, instead of linearly interpolating between two constants, we now just interpolate between two functions - namely the functions $\vec{f}_{j,k}(y, z)$ from Definition 5.2.

Since a linear combination of two measurable coefficients (as defined in Definition 4.3 on page 42) is itself a measurable coefficient, it follows that $\check{f}$ is a measurable coefficient and the proof is complete.

□

Definition 5.14. *For a set $A \subset \mathbb{R}^d$, we define*

$$\mu(A) := \int_{\mathbb{R}^d} \mathbb{I}_A(x)\rho(x)dx.$$

Lemma 5.15. *Let $(\vec{Y}, \vec{Z})$ denote the solution to the BDSDE with step coefficients $\{\vec{f}, \vec{g}, \vec{h}\}$ as defined in Definition 5.2, and $(\check{Y}, \check{Z})$ the solution to the BDSDE with Lipschitz coefficients $\{\check{f}, \check{g}, \check{h}\}$ as defined in Definition 5.12. Then there exists a*

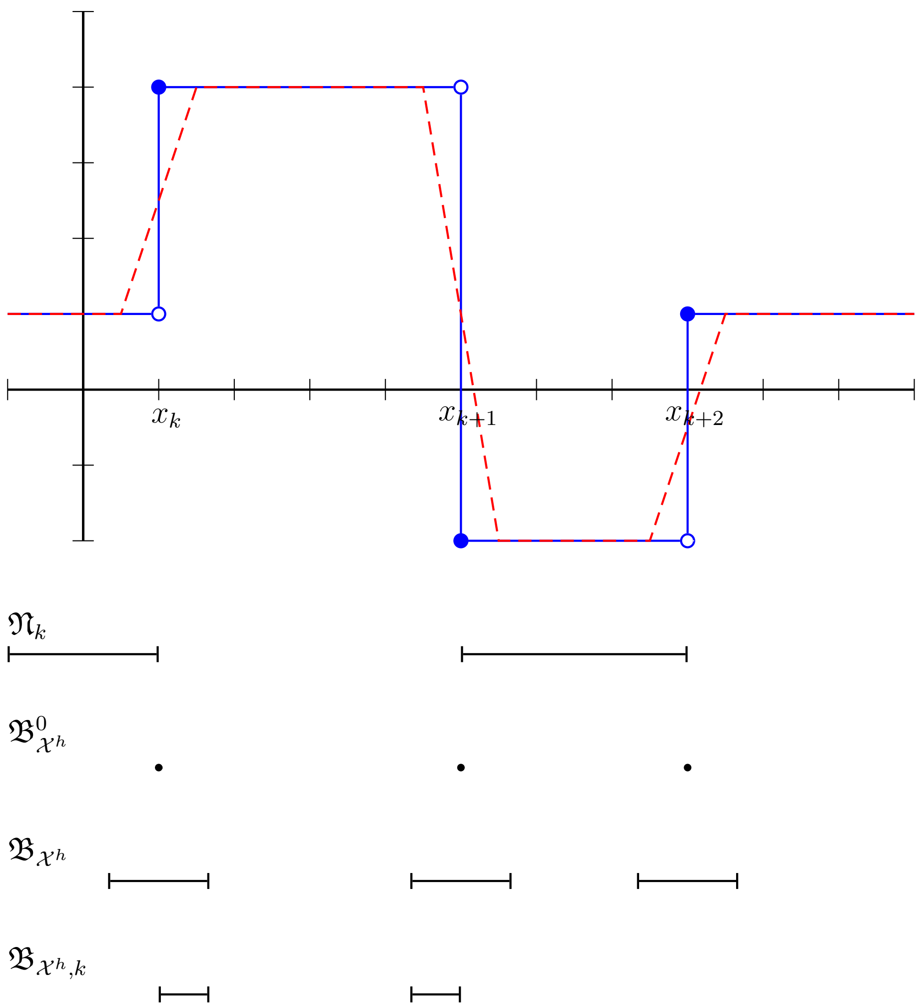

Figure 5.2.: A visualization of the sets from Definition 5.11 related to $\mathcal{X}_k := [x_k, x_{k-1})$.

constant $C > 0$ independent of C_E, C_M and C_G such that

$$\mathcal{E}_{\sup}\left[|\check{Y}_r^{t,x} - \vec{Y}_r^{t,x}|^2\right] + \mathcal{E}_T\left[|\check{Z}_r^{t,x} - \vec{Z}_r^{t,x}|^2\right] \leq CC_M^2\left(l(\mathfrak{B}_{\mathcal{T}}) + \mu(\mathfrak{B}_{\mathcal{X}})\right)$$

where l represents Lebesgue measure and μ is defined in Definition 5.14.

Proof. By Itô's formula (Result A.4) we have that for any $s \in [t, T]$ and a.e. $x \in \mathbb{R}^d$,

$$\begin{aligned}
&|\check{Y}_s^{t,x} - \vec{Y}_s^{t,x}|^2 + \int_s^T |\check{Z}_r^{t,x} - \vec{Z}_r^{t,x}|^2 dr \\
&= |\check{h}(X_T^{t,x}) - \vec{h}(X_T^{t,x})|^2 + 2\int_s^T (\check{Y}_r^{t,x} - \vec{Y}_r^{t,x})(\check{f}(r, \check{\theta}_r^{t,x}) - \vec{f}(r, \vec{\theta}_r^{t,x}))dr \\
&\quad + \sum_{j=1}^m \int_s^T |\check{g}_j(r, \check{\theta}_r^{t,x}) - \vec{g}_j(r, \vec{\theta}_r^{t,x})|^2 dr - M_s^{t,x}
\end{aligned}$$

where

$$\begin{aligned}
M_s^{t,x} := 2\sum_{j=1}^m \int_s^T (\check{Y}_r^{t,x} - \vec{Y}_r^{t,x})\left(\check{g}_j(r, \check{\theta}_r^{t,x}) - \vec{g}_j(r, \vec{\theta}_r^{t,x})\right)\overleftarrow{dB}_r^j \\
+ 2\int_s^T \left\langle (\check{Y}_r^{t,x} - \vec{Y}_r^{t,x})(\check{Z}_r^{t,x} - \vec{Z}_r^{t,x}), dW_r \right\rangle.
\end{aligned}$$

Following the same argument based upon Gronwall's inequality as in the proof of Theorem 5.9 we have that

$$\begin{aligned}
&\mathcal{E}\left[|\check{Y}_s^{t,x} - \vec{Y}_s^{t,x}|^2\right] + \mathcal{E}_T\left[|\check{Z}_r^{t,x} - \vec{Z}_r^{t,x}|^2\right] \\
&\leq C\left(\mathcal{E}\left[|\check{h}(X_T^{t,x}) - \vec{h}(X_T^{t,x})|^2\right]\right. \\
&\quad \left. + \mathcal{E}_T\left[|\check{f}(r, \vec{\theta}_r^{t,x}) - \vec{f}(r, \vec{\theta}_r^{t,x})|^2 + \sum_{j=1}^m |\check{g}_j(r, \vec{\theta}_r^{t,x}) - \vec{g}_j(r, \vec{\theta}_r^{t,x})|^2\right]\right).
\end{aligned}$$

By equivalence of norms and the definition of Lipschitz coefficients it follows that

$$\begin{aligned}
\mathcal{E}\left[|\check{h}(X_T^{t,x}) - \vec{h}(X_T^{t,x})|^2\right] &\leq 4C_M^2 \mathcal{E}\left[\mathbb{I}_{\mathfrak{B}_{\mathcal{X}}}(X_T^{t,x})\right] \\
&\leq CC_M^2 \int_{\mathbb{R}^d} \mathbb{I}_{\mathfrak{B}_{\mathcal{X}}}(x)\rho(x)dx \\
&= CC_M^2 \mu(\mathfrak{B}_{\mathcal{X}}).
\end{aligned}$$

Similarly, for $\phi = f, g$

$$\begin{aligned}\mathcal{E}_T\left[|\check{\phi}(r,\vec{\theta}_r^{t,x}) - \vec{\phi}(r,\vec{\theta}_r^{t,x})|^2\right] &\leq 4C_M^2\mathcal{E}_T\left[\mathbb{I}_{\mathfrak{B}_\mathcal{T}}(r) + \mathbb{I}_{\mathfrak{B}_\mathcal{X}}(X_r^{t,x})\right]\\ &\leq 4C_M^2\left(\int_t^T \mathbb{I}_{\mathfrak{B}_\mathcal{T}}(r)dr + C(T-t)\int_{\mathbb{R}^d}\mathbb{I}_{\mathfrak{B}_\mathcal{X}}(x)\rho(x)dxdr\right)\\ &= 4C_M^2\left(l(\mathfrak{B}_\mathcal{T}) + C(T-t)\mu(\mathfrak{B}_\mathcal{X})\right)\end{aligned}$$

and the result without the sup follows.

The result with the sup follows, as in the proof of Theorem 5.9, by applying the Burkholder–Davis–Gundy inequality to $M_s^{t,x}$. □

An easy corollary of Theorem 5.9 and Lemma 5.15 is the following theorem which is the main result of this section.

Theorem 5.16. *Let (Y, Z) denote the solution to the BDSDE with measurable coefficients $\{f, g, h\}$ as defined in Definition 4.3 on page 42 and $(\check{Y}, \check{Z})$ the solution to the BDSDE with Lipschitz coefficients $\{\check{f}, \check{g}, \check{h}\}$ as defined in Definition 5.12. Then there exists a constant $C > 0$ independent of C_E, C_M and C_G such that*

$$\mathcal{E}_{\sup}\left[|\check{Y}_r^{t,x} - Y_r^{t,x}|^2\right] + \mathcal{E}_T\left[|\check{Z}_r^{t,x} - Z_r^{t,x}|^2\right] < \beta(C_E, C_M) + CC_M^2\left(l(\mathfrak{B}_\mathcal{T}) + \mu(\mathfrak{B}_\mathcal{X})\right)$$

where l represents Lebesgue measure, μ is defined in Definition 5.14 and β is defined in Definition 5.10.

Remark. As previously noted, we can make the term $\beta(C_E, C_M)$ arbitrarily small by making C_E small and C_M large and selecting the partitions $\mathcal{T}$ and $\mathcal{X}$ accordingly. For a fixed value of C_G this will, however, likely make the terms $l(\mathfrak{B}_\mathcal{T})$ and $\mu(\mathfrak{B}_\mathcal{X})$ larger. The actual effect on altering C_E and C_M on the terms $\beta(C_E, C_M)$, $l(\mathfrak{B}_\mathcal{T})$ and $\mu(\mathfrak{B}_\mathcal{X})$ is highly dependent on the coefficients and the choices $\mathcal{T}$ and $\mathcal{X}$. It is of course possible to place upper bounds on the effects but these will be gross overestimates in most cases and quite unhelpful. For example, doubling the value of C_M could obviously double the value of $l(\mathfrak{B}_\mathcal{T})$ but this would certainly not be typical.

Despite this, it is easy to see that we can always find values of C_E, C_M and C_G to make the error bound in Theorem 5.16 arbitrarily small. To see this, suppose that $\epsilon > 0$ is given and we require that

$$\mathcal{E}_{\sup}\left[|\check{Y}_r^{t,x} - Y_r^{t,x}|^2\right] + \mathcal{E}_T\left[|\check{Z}_r^{t,x} - Z_r^{t,x}|^2\right] < \epsilon.$$

By Theorem 5.9, we can find values of C_E and C_M so that $\beta(C_E, C_M) < \frac{\epsilon}{2}$. Now let us fix, C_E and C_M to these values. For fixed C_E and C_M, we can by Lemma 5.15 find a value of C_G such that $CC_M^2\left(l(\mathfrak{B}_\mathcal{T}) + \mu(\mathfrak{B}_\mathcal{X})\right) < \frac{\epsilon}{2}$ and our desired inequality is satisfied.

5.4. BDSDEs with Smooth Coefficients

In this section we approximate Lipschitz coefficients with smooth coefficients and show that the solution of the BDSDE with Lipschitz coefficients approximates the solution of the BDSDE with smooth coefficients. The motiviation for this is to be able to apply the regularity result for BDSDEs with smooth coefficients derived in Chapter 6 when deriving an upper bound for the error term in our discretization scheme in Chapter 7. An easy corollary of approximating the solution of BDSDEs with Lipschitz coefficients with the solution of BDSDEs with smooth coefficients is that we can approximate the solution of BDSDE with measurable coefficients with the solution of BDSDEs with smooth coefficients.

Lemma 5.17. *Suppose that $f : \mathbb{R}^d \to \mathbb{R}$ is a Hölder-continuous function with exponent α and let us define for each $\delta > 0$, the smooth function*

$$f_\delta(x) := \delta^{-d} \int_{\mathbb{R}^d} \lambda\left(\frac{x - x'}{\delta}\right) f(x')dx'$$

where

$$\lambda(x) := \begin{cases} C \exp\left\{\frac{1}{|x|^2-1}\right\}, & |x| < 1 \\ 0, & |x| \geq 1 \end{cases}$$

and C is chosen so that $\int_{\mathbb{R}^d} \lambda(x) = 1$. Then f_δ is also a Hölder-continuous function with exponent α and $f_\delta \to f$ uniformly on $\mathbb{R}^d$ as $\delta \to 0$.

Proof. The proof is practically the same as the standard result for continuous functions with compact support (see for example [43], p.19). The difference is that we no longer require f to have compact support since the Hölder-continuity of f is sufficient to imply uniform continuity on the entire domain. □

Definition 5.18. *Let $\{\check{f}, \check{g}, \check{h}\}$ be Lipschitz coefficients as defined in Definition 5.12. We call $\{\widetilde{f}, \widetilde{g}, \widetilde{h}\}$ their* smooth approximations parameterised by δ *or simply* smooth coefficients *if they are defined for some $\delta > 0$ by*

$$\begin{aligned}
\widetilde{f}(r,x,y,z) &:= \delta^{-(d+1)} \int_{t-\delta}^{T+\delta} \int_{\mathbb{R}^d} \lambda\left(\frac{r-r'}{\delta}\right) \lambda\left(\frac{x-x'}{\delta}\right) \check{f}(r',x',y,z) dx' dr', \\
\widetilde{g}(r,x,y) &:= \delta^{-(d+1)} \int_{t-\delta}^{T+\delta} \int_{\mathbb{R}^d} \lambda\left(\frac{r-r'}{\delta}\right) \lambda\left(\frac{x-x'}{\delta}\right) \check{g}(r',x',y) dx' dr' \quad \textit{and} \\
\widetilde{h}(x) &:= \delta^{-d} \int_{\mathbb{R}^d} \lambda\left(\frac{x-x'}{\delta}\right) \check{h}(x') dx'
\end{aligned}$$

Remark. Since we are required in Definition 5.18 to integrate over the enlarged interval $[t-\delta, T+\delta]$, we extend $\check{f}$ and $\check{g}$ as follows:

$$\begin{aligned}
&\check{f}(r,x,y,z) := \check{f}(t,x,y,z) \text{ and } \check{g}(r,x,y) := \check{g}(t,x,y) \text{ for } r < t, \\
&\check{f}(r,x,y,z) := \check{f}(T,x,y,z) \text{ and } \check{g}(r,x,y) := \check{g}(T,x,y) \text{ for } r > T.
\end{aligned}$$

Theorem 5.19. *Let $(\check{Y}, \check{Z})$ denote the solution to the BDSDE with Lipschitz coefficients $\{\check{f}, \check{g}, \check{h}\}$ as defined in Definition 5.12. Then for each $\epsilon > 0$, there exists a $\delta > 0$ such that the solution $(\widetilde{Y}, \widetilde{Z})$ to the BDSDE with smooth coefficients $\{\widetilde{f}, \widetilde{g}, \widetilde{h}\}$ parameterised by δ as defined in Definition 5.18 satisfies*

$$\mathcal{E}_{\sup}\left[|\widetilde{Y}_r^{t,x} - \check{Y}_r^{t,x}|^2\right] + \mathcal{E}_T\left[|\widetilde{Z}_r^{t,x} - \check{Z}_r^{t,x}|^2\right] < \epsilon.$$

Proof. By Itô's formula (Result A.4) we have that for any $s \in [t,T]$ and a.e. $x \in \mathbb{R}^d$,

$$\begin{aligned}
&|\widetilde{Y}_s^{t,x} - \check{Y}_s^{t,x}|^2 + \int_s^T |\widetilde{Z}_r^{t,x} - \check{Z}_r^{t,x}|^2 dr \\
&= |\widetilde{h}(X_T^{t,x}) - \check{h}(X_T^{t,x})|^2 + 2\int_s^T (\widetilde{Y}_r^{t,x} - \check{Y}_r^{t,x})(\widetilde{f}(r, \widetilde{\theta}_r^{t,x}) - \check{f}(r, \check{\theta}_r^{t,x}))dr \\
&\quad + \sum_{j=1}^m \int_s^T |\widetilde{g}_j(r, \widetilde{\theta}_r^{t,x}) - \check{g}_j(r, \check{\theta}_r^{t,x})|^2 dr - M_s^{t,x}
\end{aligned}$$

where

$$M_s^{t,x} := 2\sum_{j=1}^m \int_s^T (\widetilde{Y}_r^{t,x} - \check{Y}_r^{t,x}) \left(\widetilde{g}_j(r, \widetilde{\theta}_r^{t,x}) - \check{g}_j(r, \check{\theta}_r^{t,x})\right) \overleftarrow{dB}_r^j$$

$$+ 2 \int_s^T \left\langle (\widetilde{Y}_r^{t,x} - \check{Y}_r^{t,x})(\widetilde{Z}_r^{t,x} - \check{Z}_r^{t,x}), dW_r \right\rangle .$$

Following the same argument based upon Gronwall's inequality as in the proof of Theorem 5.9 we have that

$$\begin{aligned}
&\mathcal{E}\left[|\widetilde{Y}_s^{t,x} - \check{Y}_s^{t,x}|^2\right] + \mathcal{E}_T\left[|\widetilde{Z}_r^{t,x} - \check{Z}_r^{t,x}|^2\right] \\
&\le C \Bigg(\mathcal{E}\left[|\widetilde{h}(X_T^{t,x}) - \check{h}(X_T^{t,x})|^2\right] \\
&\quad + \mathcal{E}_T \left[|\widetilde{f}(r, \check{\theta}_r^{t,x}) - \check{f}(r, \check{\theta}_r^{t,x})|^2 + \sum_{j=1}^m |\widetilde{g}_j(r, \check{\theta}_r^{t,x}) - \check{g}_j(r, \check{\theta}_r^{t,x})|^2 \right] \Bigg) .
\end{aligned}$$

By Lemma 5.17, $|\widetilde{h}(x) - \check{h}(x)|$ can be made arbitrarily small uniformly in x. As a consequence, we have that

$$\mathcal{E}\left[|\widetilde{h}(X_T^{t,x}) - \check{h}(X_T^{t,x})|^2\right] \le \epsilon.$$

Similarly, $|\widetilde{f}(r,x,y,z) - \check{f}(r,x,y,z)|$ and $|\widetilde{g}_j(r,x,y) - \check{g}_j(r,x,y)|$ can also be made arbitrarily small uniformly in r, x, y and z. It is easy to see that the result without the sup now follows.

The result with the sup follows, as in the proof of Theorem 5.9, by applying the Burkholder–Davis–Gundy inequality to $M_s^{t,x}$. □

The following corollary follows directly from Theorem 5.16 and Theorem 5.19.

Corollary 5.20. *Let (Y, Z) denote the solution to the BDSDE with measurable coefficients $\{f, g, h\}$ as defined in Definition 4.3. There exist smooth coefficients $\{\widetilde{f}, \widetilde{g}, \widetilde{h}\}$ as defined in Definition 5.18 parameterised by a $\delta > 0$ and a constant $C > 0$ independent of C_E, C_M, C_G and δ such that the solution $(\widetilde{Y}, \widetilde{Z})$ to the BDSDE with smooth coefficients satisfies*

$$\mathcal{E}_{\sup}\left[|\widetilde{Y}_r^{t,x} - Y_r^{t,x}|^2\right] + \mathcal{E}_T\left[|\widetilde{Z}_r^{t,x} - Z_r^{t,x}|^2\right] \le \beta(C_E, C_M) + C C_M^2 \left(l(\mathfrak{B}_{\mathcal{T}}) + \mu(\mathfrak{B}_{\mathcal{X}}) \right)$$

where l represents Lebesgue measure, μ is defined in Definition 5.14 and ϵ is defined in Definition 5.10.

Remark. We note that in the corollary above, the inclusion of C_E, C_M and C_G is due to the implicit dependence of the smooth coefficients $\{\widetilde{f}, \widetilde{g}, \widetilde{h}\}$ upon them.

5.5. Conclusion

The main results of this chapter are:

- Theorem 5.16: the approximation of BDSDEs with measurable coefficients by BDSDEs with Lipschitz coefficients.
- Theorem 5.19: the approximation of BDSDEs with Lipschitz coefficients by BDSDEs with smooth coefficients.

By Theorem 5.16 we are able to shift our discretization problem to the Lipschitz coefficient setting and by Theorem 5.19 we are able to shift to a smooth coefficient setting when this is convenient.

Remark. Let us recall Theorem 3.2 on page 37 from the review of [50] on the connection between BDSDEs and SPDEs. In light of Theorem 3.2, the results of this chapter can be recast as results on the approximation of SPDEs with measurable coefficients by SPDEs with Lipschitz / smooth coefficients.

6 Regularity of Y and Z

6.1. Introduction

In this chapter we derive a regularity result for Y given Lipschitz coefficients and for Z given smooth coefficients as they will be central to deriving the error bound of the discretization scheme in Chapter 7. To achieve this, we will need to construct a suitable refinement of $\mathcal{T}$ for which these regularity results hold and which we can subsequently base our discretization scheme upon.

Remark. These regularity results are an extension of those found in [49]. Regularity results for conditions stronger than ours are derived in [2]. As we will note, however, there is a missing step in the proof of the regularity of Z in [2] which we address.

Remark. The majority of the definitions and results in this chapter are parameterised by a value Δt. We stress that the definitions and results hold for any $\Delta t \in (0, C_E]$. In particuar, the value of Δt *is not fixed*. Instead, the value of C_E is now fixed but we are free to choose any $\Delta t \in (0, C_E]$ to plug into the definitions and results of this chapter.

6.2. Regularity of Y for BDSDEs with Lipschitz Coefficients

In this section our main result is a regularity result for Y given Lipschitz coefficients. Prior to proving this theorem, we will introduce the concept of a partition without cluster and prove a couple of technical lemmas.

From here on we will assume that our step coefficients as defined in Definition 5.2 on page 47 and their Lipschitz approximations as defined in Definition 5.12 on page 66 are fixed. As a consequence, the constants C_E, C_M and C_G are now fixed as are the partitions $\mathcal{T}$ and $\mathcal{X}^\phi$ for $\phi = f, g, h$ of $[t,T]$ and $[-C_M, C_M]^d$ respectively.

Definition 6.1. *Suppose that $\mathcal{U}$ is a set of partitions of $[t,T]$ each of which is a refinement of $\mathcal{T}$. $\mathcal{U}$ is called a* family of partitions without cluster of order κ *if there exists a constant $\kappa \geq 1$ independent of the partition constant C_E and truncation constant C_M such that for any $\Delta t > 0$ there exists a partition $\mathcal{U}_j := \{u_0, u_1, \ldots, u_{N(\mathcal{U}_j)}\} \in \mathcal{U}$ satisfying $\max_i(u_i - u_{i-1}) \leq \Delta t$ and $N(\mathcal{U}_j) \leq \kappa \left\lceil \frac{T-t}{\Delta t} \right\rceil$. $\mathcal{U}_j$ is in turn called a* partition without cluster of order κ and granularity Δt.

Example. *If $\mathcal{T}$ is the trivial partition ($\mathcal{T} := \{t, T\}$) then the set of all uniform partitions of $[t,T]$ is a family of partitions without cluster of order one.*

Remark. The reason for introducing the concept of partitions without cluster will become apparent as we proceed but let us have a brief preview. The functions f and g are allowed to be irregular in both t and x. Our strategy to cope with this irregularity with respect to x will be to utilise the spatial integral in the formulation of our norm and appeal to the equivalence of norms result. Effectively, this will cause the irregularities to be "averaged out". With the time irregularity, however, we will not be able to use a similar argument. Instead, whenever we encounter an irregular region, we will drop the step size in this region to compensate. The trick will be to demonstrate that we do not have to add too many extra discretization points for this to work - this is where the concept of partitions without cluster becomes helpful.

We will construct our partition of $[t,T]$ via successive refinements of $\mathcal{T}$. We recall that by definition, each interval of $\mathcal{T}$ is of length at least C_E. We now define our first refinement of $\mathcal{T}$

Definition 6.2. *We denote by $\mathcal{T}^1$ any refinement of $\mathcal{T}$ where each interval of $\mathcal{T}$ of length l satisfying $(n+1)C_E > l \geq nC_E$ for some integer $n \geq 2$, is divided into $n-1$ subintervals of length C_E and a single subinterval of length $< 2C_E$.*

Remark. A consequence of Definition 6.2 is that $\mathcal{T}^1$ consists of intervals all of length $l \in [C_E, 2C_E)$. This means that in the terminology of Definition 6.1, $\mathcal{T}^1$ is a partition without cluster of order 2 and granularity $2C_E$.

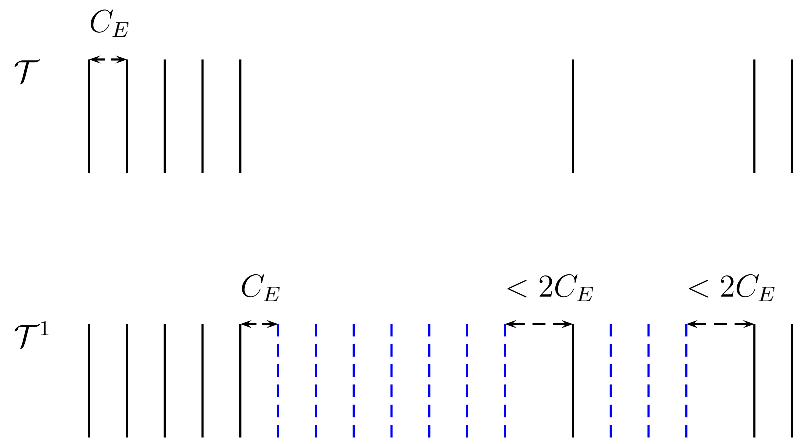

Figure 6.1.: Construction of $\mathcal{T}^1$ from $\mathcal{T}$: each interval of $\mathcal{T}^1$ has width $\in [C_E, 2C_E)$.

Definition 6.3. *We denote*

$$\mu_{\mathcal{X}} := \sqrt{\mu(\mathfrak{B}_{\mathcal{X}^h})} + \sqrt{\mu(\mathfrak{B}_{\mathcal{X}^f})} + \sum_{j=1}^{m} \sqrt{\mu(\mathfrak{B}_{\mathcal{X}^{g_j}})}.$$

Lemma 6.4. *Let $(\widetilde{Y}, \widetilde{Z})$ denote the solution to the BDSDE with smooth coefficients $\{\widetilde{f}, \widetilde{g}, \widetilde{h}\}$ as defined in Definition 5.18. Then for each $p \geq 2$, there exists a constant $C > 0$ independent of C_E, C_M and C_G such that for any $s \in [t, T]$*

$$\mathcal{E}\left[|\nabla \widetilde{Y}_s^{t,x}|^p\right] \leq C C_G^p (m^{p/2-1} + 1)\mu_{\mathcal{X}}.$$

Proof. In this proof we adapt the approach of [49]. We have (see [36], p. 223) that

$$\nabla \widetilde{Y}_s^{t,x} = (\widetilde{h}_x(X_T^{t,x}))^T \nabla X_T^{t,x} + \int_s^T \widetilde{f}'(r, \widetilde{\theta}_r^{t,x}) \nabla \widetilde{\theta}_r^{t,x} dr - \sum_{j=1}^{m} \int_s^T \widetilde{g}_j'(r, \widetilde{\theta}_r^{t,x}) \nabla \widetilde{\theta}_r^{t,x} \overleftarrow{dB}_r^j$$
$$- \int_s^T \left(\nabla \widetilde{Z}_r^{t,x}\right)^T dW_r$$

where we have used the shorthand for $\widetilde{\phi} = \widetilde{f}, \widetilde{g}$

$$\widetilde{\phi}'(r, \widetilde{\theta}_r^{t,x}) \nabla \widetilde{\theta}_r^{t,x} := (\widetilde{\phi}_x(r, \widetilde{\theta}_r^{t,x}))^T \nabla X_r^{t,x} + \widetilde{\phi}_y(r, \widetilde{\theta}_r^{t,x}) \nabla \widetilde{Y}_r^{t,x} + (\widetilde{\phi}_z(r, \widetilde{\theta}_r^{t,x}))^T \nabla \widetilde{Z}_r^{t,x}.$$

We note that the first and third terms are the product of a vector and a matrix and the second term is the product of a scalar and a vector.

By Result A.5 on the moments of $\widetilde{Y}$ and since each of the derivatives of $\widetilde{f}$, $\widetilde{g}$ and $\widetilde{h}$ are bounded, we have that

$$\mathcal{E}\left[|\nabla\widetilde{Y}_r^{t,x}|^p\right] \leq C\mathcal{E}\left[|(h'(X_T^{t,x}))^T\nabla X_T^{t,x}|^p\right] \\ +C\mathcal{E}_T\left[|(\widetilde{f}_x(r,\widetilde{\theta}_r^{t,x}))^T\nabla X_r^{t,x}|^p + \left(\sum_{j=1}^m |(\widetilde{g}_{j,x}(r,\widetilde{\theta}_r^{t,x}))^T\nabla X_r^{t,x}|^2\right)^{p/2}\right].$$

By the Cauchy-Schwarz inequality, the equivalence of norms, and Result A.1 on the moments of ∇X it follows that

$$\begin{aligned}
&\mathcal{E}\left[|(\widetilde{h}'(X_T^{t,x}))^T\nabla X_T^{t,x}|^p\right] \\
&\leq 2\mathcal{E}\left[C_G^p\mathbb{I}_{\mathfrak{B}_{\mathcal{X}^h}}(X_T^{t,x})|\nabla X_T^{t,x}|^p\right] \\
&\leq CC_G^p\left(\int_{\mathbb{R}^d}\mathbb{I}_{\mathfrak{B}_{\mathcal{X}^h}}(x)\rho(x)dx\,\mathcal{E}\left[|\nabla X_T^{t,x}|^{2p}\right]\right)^{1/2} \\
&\leq CC_G^p\sqrt{\mu(\mathfrak{B}_{\mathcal{X}^h})}.
\end{aligned}$$

Note that as a result of the mollification process, $|h'(X_T^{t,x})| > 0$ for regions outside of $\mathfrak{B}_{\mathcal{X}^h}$ (which we can make small by making δ in Definition 5.18 small). This is the reason for introducing the factor of two in the inequality above. By Jensen's inequality,

$$\left(\sum_{j=1}^m |(\widetilde{g}_{j,x}(r,\widetilde{\theta}_r^{t,x}))^T\nabla X_r^{t,x}|^2\right)^{p/2} \leq m^{p/2-1}\sum_{j=1}^m |(\widetilde{g}_{j,x}(r,\widetilde{\theta}_r^{t,x}))^T\nabla X_r^{t,x}|^p$$

and so

$$\begin{aligned}
&\mathcal{E}_T\left[\left(\sum_{j=1}^m |(\widetilde{g}_{j,x}(r,\widetilde{\theta}_r^{t,x}))^T\nabla X_r^{t,x}|^2\right)^{p/2}\right] \\
&\leq Cm^{p/2-1}C_G^p\sum_{j=1}^m\mathcal{E}_T\left[\mathbb{I}_{\mathfrak{B}_{\mathcal{X}^{g_j}}}(X_r^{t,x})|\nabla X_r^{t,x}|^p\right] \\
&\leq Cm^{p/2-1}C_G^p\sum_{j=1}^m\left(\int_t^T\int_{\mathbb{R}^d}\mathbb{I}_{\mathfrak{B}_{\mathcal{X}^{g_j}}}(x)\rho(x)dxdr\mathcal{E}_T\left[|\nabla X_r^{t,x}|^{2p}\right]\right)^{1/2} \\
&\leq Cm^{p/2-1}C_G^p\sum_{j=1}^m\sqrt{\mu(\mathfrak{B}_{\mathcal{X}^{g_j}})}.
\end{aligned}$$

The result follows via a similar argument for f. $\square$

Definition 6.5. *Let $\breve{g}$ denote a Lipschitz coefficient as defined in Definition 5.12 on page 66 and let $\mathcal{T}^1$ be as defined in Definition 6.2. Suppose that $\mathcal{T}^1$ is represented by the points r_i where $t = r_0 < r_1 < \ldots < r_n = T$.*

Define

$$U := \int_t^T \int_{\mathbb{R}^d} \sum_{j=1}^m |\breve{g}_j(r, x, 0)|^2 \rho(x) dx dr \leq 2C_{f+g}$$

and for $i = 1, \ldots, n$

$$u_i := \int_{r_{i-1}}^{r_i} \int_{\mathbb{R}^d} \sum_{j=1}^m |\breve{g}_j(r, x, 0)|^2 \rho(x) dx dr$$

so that $U = \sum_{i=1}^n u_i$.

We define the refinement $\mathcal{T}^{1.1}$ of $\mathcal{T}^1$ by partitioning each interval $[r_{i-1}, r_i)$ where $u_i > \frac{U}{n}$ into $\left\lceil \frac{nu_i}{U} \right\rceil$ equally spaced subintervals.

Given a $\Delta t \in (0, C_E]$, we define the refinement $\mathcal{T}^{1.2}$ of $\mathcal{T}^{1.1}$ by dividing each of the $2n$ intervals of $\mathcal{T}^{1.1}$ into subintervals of maximum length Δt.

Define $N := \left\lceil \frac{T-t}{\Delta t} \right\rceil$. Then we define the refinement $\mathcal{T}^2$ of $\mathcal{T}^{1.2}$ by uniformly dividing each interval of $\mathcal{T}^{1.2}$ into exactly $\left\lceil \frac{2N}{n} \right\rceil$ subintervals.

Lemma 6.6. *Let $\breve{g}$ denote a Lipschitz coefficient as defined in Definition 5.12 on page 66 with partition constant C_E, truncation constant C_M and maximum slope C_G and let $\mathcal{T}^2$ be as defined in Definition 6.5. Then there exists a constant $C > 0$ independent of C_E, C_M and C_G such that for any $\Delta t \in (0, C_E]$, $\mathcal{T}^2$ is a partition without cluster of order 6 and granularity Δt such that for $i = 1, \ldots, N(\mathcal{T}^2)$*

$$\int_{t_{i-1}}^{t_i} \int_{\mathbb{R}^d} \sum_{j=1}^m |\breve{g}_j(r, x, 0)|^2 \rho(x) dx dr \leq C \Delta t.$$

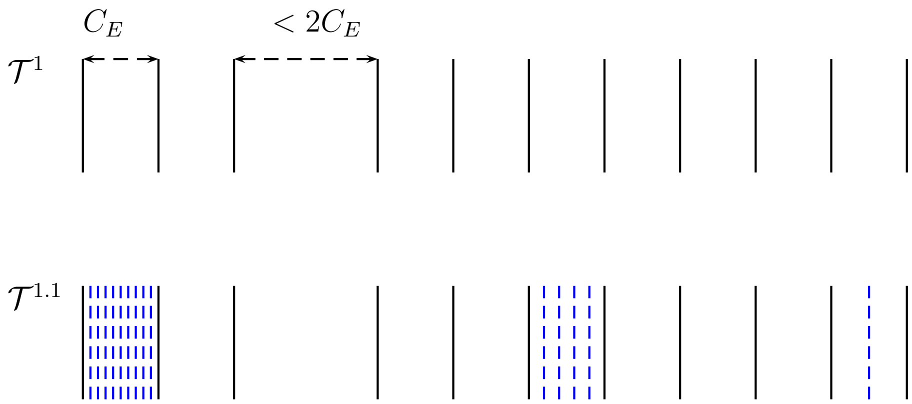

Figure 6.2.: Construction of $\mathcal{T}^{1.1}$ from $\mathcal{T}^1$: each interval of $\mathcal{T}^1$ that has a high value of u_i (as defined in Definition 6.5) is split into subintervals.

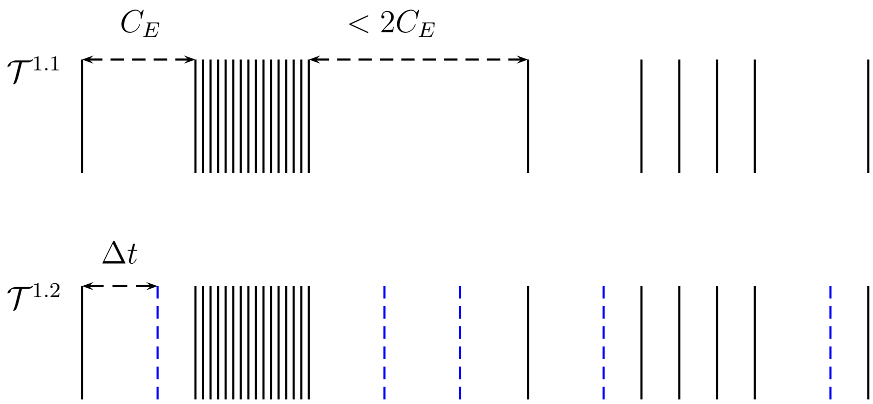

Figure 6.3.: Construction of $\mathcal{T}^{1.2}$ from $\mathcal{T}^{1.1}$: each interval of $\mathcal{T}^{1.2}$ has maximum width Δt.

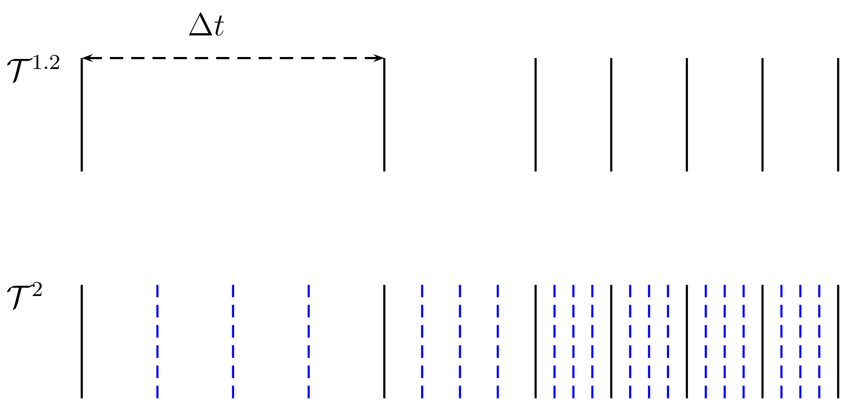

Figure 6.4.: Construction of $\mathcal{T}^2$ from $\mathcal{T}^{1.2}$: each interval of $\mathcal{T}^{1.2}$ is split into an equal number of subintervals.

Proof. Let us use the notation of Definition 6.5. By definition, for any j, k, l, any $r \in \mathcal{T}_k$ and any $x \in \mathcal{X}_l$, $\breve{g}_j(r, x, 0)$ is constant. We will first show that we can add an additional n points to the partition $\mathcal{T}^1$ to construct the refinement $\mathcal{T}^{1.1} := \{s_0, s_1, \ldots, s_{2n}\}$ such that for $i = 1, \ldots, 2n$,

$$v_i := \int_{s_{i-1}}^{s_i} \int_{\mathbb{R}^d} \sum_{j=1}^m |\breve{g}_j(r, x, 0)|^2 \rho(x) dx dr \leq \frac{U}{n}.$$

To see this, note that for every u_i such that $u_i > \frac{U}{n}$, it is sufficient to partition each corresponding interval $(r_{i-1}, r_i]$ into $\lceil \frac{nu_i}{U} \rceil$ equally spaced subintervals (since $(r_{i-1}, r_i] = \mathcal{T}^1_{i-1} \subset \mathcal{T}_j$ for some j and $\breve{g}$ is constant on $\mathcal{T}_j$ for fixed x). To achieve this we require $\lceil \frac{nu_i}{U} \rceil - 1 \leq \lfloor \frac{nu_i}{U} \rfloor$ additional points. This means that the total number of extra points we require is bounded by

$$\sum_{i=1}^n \left\lfloor \frac{nu_i}{U} \right\rfloor \leq \frac{n}{U} \sum_{i=1}^n u_i \leq n$$

as required.

We now divide each of the $2n$ intervals of $\mathcal{T}^{1.1}$ into subintervals of maximum length Δt to obtain $\mathcal{T}^{1.2}$. Since $\mathcal{T}^1$ has n steps of length at least C_E, we have that $C_E \leq \frac{T-t}{n}$. To achieve an interval length $l \leq \Delta t$, since each interval of $\mathcal{T}^{1.1}$ has length $l < 2C_E$ it is sufficient to divide each interval into at most $\lceil \frac{2C_E}{\Delta t} \rceil$ subintervals. If we define $N := \lceil \frac{T-t}{\Delta t} \rceil$ then $\frac{1}{\Delta t} \leq \frac{N}{T-t}$ and

$$\frac{2C_E}{\Delta t} \leq \frac{2(T-t)}{n} \frac{N}{T-t} = \frac{2N}{n}.$$

It follows that $\lceil \frac{2C_E}{\Delta t} \rceil \leq \lceil \frac{2N}{n} \rceil$ and it is sufficient to divide each of the $2n$ intervals of $\mathcal{T}^{1.1}$ into $\lceil \frac{2N}{n} \rceil$ subintervals which requires at most $\frac{2N}{n} 2n = 4N$ additional points.

We now construct the refinement $\mathcal{T}^2$ of $\mathcal{T}^{1.2}$ by uniformly dividing each interval of $\mathcal{T}^{1.2}$ into exactly $\lceil \frac{2N}{n} \rceil$ subintervals. This means that $\mathcal{T}^2 := \{t_0, t_1, \ldots, t_{N(\mathcal{T}^2)}\}$ is a partition without cluster of order 6 and granularity Δt such that for $i = 1, \ldots, N(\mathcal{T}^2)$, we have that $t_i - t_{i-1} \leq \Delta t$ and

$$\int_{t_{i-1}}^{t_i} \int_{\mathbb{R}^d} \sum_{j=1}^m |\breve{g}_j(r, x, 0)|^2 \rho(x) dx dr \leq \frac{U}{n} \frac{n}{2N} \leq \frac{U \Delta t}{2(T-t)}.$$

□

Theorem 6.7. *Let $(\check{Y}, \check{Z})$ denote the solution to the BDSDE with Lipschitz coefficients $\{\check{f}, \check{g}, \check{h}\}$ as defined in Definition 5.12 on page 66. Then there exists a constant $C > 0$ independent of C_E, C_M, C_G such that for any $\Delta t \in (0, C_E]$ and for each $i = 1, \ldots, n$ with $\mathcal{T}^2 := \{t_0, t_1, \ldots, t_n\}$ as defined in Definition 6.5*

$$E\left[\sup_{t_{i-1}\le s\le t_i}\int_{\mathbb{R}^d}|\check{Y}_s^{t,x} - \check{Y}_{t_i}^{t,x}|^2\rho(x)dx\right] \le C\Delta t C_G^2(m+1)\sqrt{\mu_{\mathcal{X}}}$$

where $\mu_{\mathcal{X}}$ as defined in Definition 6.3.

Proof. In this proof we adapt the approach of [49]. For $s \in [t_{i-1}, t_i]$ and a.e. $x \in \mathbb{R}^d$, we have by Jensen's inequality and the Cauchy-Schwarz inequality that

$$\begin{aligned}\frac{1}{3}|\check{Y}_s^{t,x} - \check{Y}_{t_i}^{t,x}|^2 \le \Delta t\int_{t_{i-1}}^{t_i}\left|\check{f}(r, \check{\theta}_r^{t,x})\right|^2 dr + m\sum_{j=1}^m\left|\int_s^{t_i}\check{g}_j(r, \check{\theta}_r^{t,x})\overleftarrow{dB}_r^j\right|^2 \\ +\left|\int_s^{t_i}\left\langle \check{Z}_r^{t,x}, dW_r\right\rangle\right|^2.\end{aligned}$$

Applying spatial integrals, Fubini's theorem, the Burkholder–Davis–Gundy inequality and the equivalence of norms gives us that

$$\begin{aligned}
&E\left[\sup_{t_{i-1}\le s\le t_i}\int_{\mathbb{R}^d}|\check{Y}_s^{t,x} - \check{Y}_{t_i}^{t,x}|^2\rho(x)dx\right] \\
&\le CE\left[\int_{t_{i-1}}^{t_i}(\Delta t|\check{f}(r, \check{\theta}_r^{t,x})|^2 + m\sum_{j=1}^m|\check{g}_j(r, \check{\theta}_r^{t,x})|^2 + |\check{Z}_r^{t,x}|^2)dr\right] \\
&\le CE\left[\int_{t_{i-1}}^{t_i}(2\Delta t|\check{f}(r, X_r^{t,x}, 0, 0)|^2 + 2m\sum_{j=1}^m|\check{g}_j(r, X_r^{t,x}, 0)|^2\right. \\
&\quad \left. +2(m+\Delta t)L|\check{Y}_r^{t,x}|^2 + (1 + 2\Delta t L)|\check{Z}_r^{t,x}|^2)dr\right] \\
&\le C\int_{t_{i-1}}^{t_i}\int_{\mathbb{R}^d}(\Delta t|\check{f}(r, x, 0, 0)|^2 + m\sum_{j=1}^m|\check{g}_j(r, x, 0)|^2)\rho(x)dxdr \\
&\quad + CE\left[\int_{t_{i-1}}^{t_i}\int_{\mathbb{R}^d}\left(m|\check{Y}_r^{t,x}|^2 + |\check{Z}_r^{t,x}|^2\right)\rho(x)dxdr\right] \\
&\le Cm\Delta t + CE\left[\int_{t_{i-1}}^{t_i}\int_{\mathbb{R}^d}|\check{Z}_r^{t,x}|^2\rho(x)dxdr\right]
\end{aligned}$$

by Lemma 6.6 and Result A.5.

Now let $(\widetilde{Y}, \widetilde{Z})$ denote the solution to the BDSDE with smooth coefficients $\{\widetilde{f}, \widetilde{g}, \widetilde{h}\}$ as defined in Definition 5.18 on page 73. Then (see [36], p. 223) we have that

$$\widetilde{Z}_s^{t,x} = \nabla \widetilde{Y}_s^{t,x} (\nabla X_s^{t,x})^{-1} \sigma(X_s^{t,x}).$$

By Fubini's theorem, the Cauchy-Schwarz inequality and Young's inequality it follows that

$$\begin{aligned}
&E\left[\int_{t_{i-1}}^{t_i} \int_{\mathbb{R}^d} |\widetilde{Z}_r^{t,x}|^2 \rho(x) dx dr\right] \\
&\leq \Delta t \sup_{s \in [t_{i-1}, t_i]} \mathcal{E}\left[|\widetilde{Z}_r^{t,x}|^2\right] \\
&\leq \Delta t \sup_{s \in [t_{i-1}, t_i]} \sqrt{\mathcal{E}\left[|\nabla \widetilde{Y}_s^{t,x}|^4\right]} \sqrt{\mathcal{E}\left[|(\nabla X_s^{t,x})^{-1}|^8 + |\sigma(X_s^{t,x})|^8\right]}.
\end{aligned}$$

The result now follows by Theorem 5.19, Lemma 6.4, Result A.1 and the Lipschitz property of σ. □

6.3. Regularity of Z for BDSDEs with Smooth Coefficients

In this section our main result is a regularity result for Z given smooth coefficients. Prior to proving this theorem, we will discuss the missing step in the corresponding proof in [2] and prove a technical lemma that fills this gap.

Let $(\widetilde{Y}, \widetilde{Z})$ denote the solution to the BDSDE with smooth coefficients $\{\widetilde{f}, \widetilde{g}, \widetilde{h}\}$ as defined in Definition 5.18 on page 73. To derive our regularity result for $\widetilde{Z}$ we will make use of the relation (see [36], p. 223)

$$\widetilde{Z}_s^{t,x} = (\nabla \widetilde{Y}_s^{t,x})^T (\nabla X_s^{t,x})^{-1} \sigma(X_s^{t,x}).$$

Let us define the notation (which is similar to that used in [49])

$$\begin{aligned}
\widetilde{\gamma}_s^{t,x} &:= \nabla \widetilde{Y}_s^{t,x}, \\
\widetilde{\zeta}_s^{t,x} &:= \nabla \widetilde{Z}_s^{t,x}, \\
\widetilde{F}_x(r) &:= \widetilde{f}_x(r, \widetilde{\theta}_r^{t,x}),
\end{aligned}$$

$$\widetilde{G}_{j,x}(r) := \widetilde{g}_{j,x}(r, \widetilde{\theta}_r^{t,x}),$$
$$\widetilde{H}_x := \widetilde{h}_x(X_T^{t,x})$$

and similarly for $\widetilde{F}_y(r)$, $\widetilde{F}_z(r)$ and $\widetilde{G}_y(r)$. For the sake of conciseness, we can rewrite the linear BDSDE (again see [36], p. 223)

$$\begin{aligned}\nabla\widetilde{Y}_s^{t,x} &= (\widetilde{h}_x(X_T^{t,x}))^T\nabla X_T^{t,x} \\ &\quad + \int_s^T ((\widetilde{f}_x(r,\widetilde{\theta}_r^{t,x}))^T\nabla X_r^{t,x} + \widetilde{f}_y(r,\widetilde{\theta}_r^{t,x})\nabla\widetilde{Y}_r^{t,x} + (\widetilde{f}_z(r,\widetilde{\theta}_r^{t,x}))^T\nabla\widetilde{Z}_r^{t,x})dr \\ &\quad - \sum_{j=1}^m \int_s^T ((\widetilde{g}_{j,x}(r,\widetilde{\theta}_r^{t,x}))^T\nabla X_r^{t,x} + \widetilde{g}_{j,y}(r,\widetilde{\theta}_r^{t,x})\nabla\widetilde{Y}_r^{t,x})\overleftarrow{dB}_r^j \\ &\quad - \int_s^T \left(\nabla\widetilde{Z}_r^{t,x}\right)^T dW_r\end{aligned}$$

in this notation as

$$\begin{aligned}\widetilde{\gamma}_s^{t,x} &= \widetilde{H}_x + \int_s^T (\widetilde{F}_x(r) + \widetilde{F}_y(r)\widetilde{\gamma}_r^{t,x} + \widetilde{F}_z(r)\widetilde{\zeta}_r^{t,x})dr \\ &\quad - \sum_{j=1}^m \int_s^T (\widetilde{G}_{j,x}(r) + \widetilde{G}_{j,y}(r)\widetilde{\gamma}_r^{t,x})\overleftarrow{dB}_r^j - \int_s^T \left(\widetilde{\zeta}_r^{t,x}\right)^T dW_r. \end{aligned} \tag{6.1}$$

If we now define

$$\Lambda_s := \exp\left\{\int_t^s (\widetilde{F}_y(r) - \frac{1}{2}\sum_{j=1}^m |\widetilde{G}_{j,y}(r)|^2)dr - \sum_{j=1}^m \int_t^s \widetilde{G}_{j,y}(r)\overleftarrow{dB}_r^j\right\}$$

and could apply integration by parts directly, we would have that

$$\begin{aligned}\Lambda_s\widetilde{\gamma}_s^{t,x} &= \Lambda_T\widetilde{H}_x + \int_s^T \Lambda_r\left(\widetilde{K}(r) + \widetilde{F}_z(r)\widetilde{\zeta}_r^{t,x}\right)dr - \sum_{j=1}^m \int_s^T \Lambda_r\widetilde{G}_{j,x}(r)\overleftarrow{dB}_r^j \\ &\quad - \int_s^T \Lambda_r\left(\widetilde{\zeta}_r^{t,x}\right)^T dW_r,\end{aligned}$$

where $\widetilde{K}(r) := \widetilde{F}_x(r) - \widetilde{G}_x(r)\widetilde{G}_y(r)$. We note that it appears that this approach has been followed in [2]. The problem with this approach is that Λ_s is not even

$\mathcal{F}^W_{t,T} \vee \mathcal{F}^B_{s,T}$-measurable which means that the term

$$\sum_{j=1}^{m} \int_s^T \Lambda_r \widetilde{G}_{j,x}(r) \overleftarrow{dB}^j_r$$

is not well defined. It is possible to obtain a slightly modified version of this equation as the following lemma shows. To achieve this let us define

$$\overleftarrow{\Lambda}_s := \exp\left\{ -\int_s^T (\widetilde{F}_y(r) - \frac{1}{2}\sum_{j=1}^{m} |\widetilde{G}_{j,y}(r)|^2) dr + \sum_{j=1}^{m} \int_s^T \widetilde{G}_{j,y}(r) \overleftarrow{dB}^j_r \right\}$$

and observe that Λ_s is $\mathcal{F}^W_{t,s} \vee \mathcal{F}^B_{t,T}$-measurable, $\overleftarrow{\Lambda}_s$ is $\mathcal{F}^W_{t,T} \vee \mathcal{F}^B_{s,T}$-measurable and $\Lambda_s = \Lambda_T \overleftarrow{\Lambda}_s$.

Lemma 6.8. *Let $(\widetilde{Y}, \widetilde{Z})$ denote the solution to the BDSDE with smooth coefficients $\{\widetilde{f}, \widetilde{g}, \widetilde{h}\}$ as defined in Definition 5.18 on page 73. Then*

$$\begin{aligned} \Lambda_s \widetilde{\gamma}^{t,x}_s = \Lambda_T \widetilde{H}_x &+ \int_s^T \Lambda_r \left(\widetilde{K}(r) + \widetilde{F}_z(r) \widetilde{\zeta}^{t,x}_r \right) dr - \Lambda_T \sum_{j=1}^{m} \int_s^T \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r) \overleftarrow{dB}^j_r \\ &- \int_s^T \Lambda_r \left(\widetilde{\zeta}^{t,x}_r \right)^T dW_r. \end{aligned}$$

Proof. We note that the proof of this result closely follows the proof of Itô's formula given in [23], p. 149.

We will show that

$$\begin{aligned} \overleftarrow{\Lambda}_s \widetilde{\gamma}^{t,x}_s = \widetilde{H}_x &+ \int_s^T \overleftarrow{\Lambda}_r \left(\widetilde{K}(r) + \widetilde{F}_z(r) \widetilde{\zeta}^{t,x}_r \right) dr - \sum_{j=1}^{m} \int_s^T \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r) \overleftarrow{dB}^j_r \\ &- \overleftarrow{\Lambda}_T \int_s^T \Lambda_r \left(\widetilde{\zeta}^{t,x}_r \right)^T dW_r \end{aligned}$$

from which the desired result readily follows upon multiplying both sides by Λ_T.

Let us note that since $\widetilde{f}_y$ and $\widetilde{g}_y$ are bounded, we can show that for each $p \geq 1$

there exists a constant $C > 0$ such that

$$E\left[\sup_{t\le s\le T}\left\{|\Lambda_s|^p + |\overleftarrow{\Lambda}_s|^p\right\}\right] \le C.$$

Now, let us define for each $n \ge 1$ the stopping time

$$T_n := \sup\{s \in [t,T]; |\overleftarrow{\Lambda}_s| + |\widetilde{\gamma}_s^{t,x}| \ge n\}. \tag{6.2}$$

Since $\lim_{n\to\infty} T_n = s$ almost surely, it is sufficient to prove the result for the stopped processes $\overleftarrow{\Lambda}_{s\vee T_n}$ and $\widetilde{\gamma}^{t,x}_{s\vee T_n}$. As a consequence, we may assume that $\overleftarrow{\Lambda}$ and $\widetilde{\gamma}^{t,x}$ are bounded. Now,

$$\overleftarrow{\Lambda}_s = 1 - \int_s^T \left(\widetilde{F}_y(r) - \sum_{j=1}^m |\widetilde{G}_{j,y}(r)|^2\right)\overleftarrow{\Lambda}_r dr + \sum_{j=1}^m \int_s^T \widetilde{G}_{j,y}(r)\overleftarrow{\Lambda}_r \overleftarrow{dB}^j_r. \tag{6.3}$$

Since the coefficients in equations (6.1) and (6.3) are almost surely continuous, we can find a sequence of partitions $\Pi^{(n)}$ of $[s,T]$ with $N(n)$ steps (which we abbreviate as N) and random variables $R^{(n)}$ and $S^{(n)}$ such that

$$\begin{aligned}
\overleftarrow{\Lambda}_s &= 1 - \sum_{k=1}^N \left(\widetilde{F}_y(t_k) - \sum_{j=1}^m |\widetilde{G}_{j,y}(t_k)|^2\right)\overleftarrow{\Lambda}_{t_k}\Delta t_k \\
&\quad + \sum_{j=1}^m\sum_{k=1}^N \widetilde{G}_{j,y}(t_{k+1})\overleftarrow{\Lambda}_{t_{k+1}}\Delta\overleftarrow{B}^j_k + R^{(n)}, \\
\widetilde{\gamma}_s^{t,x} &= \widetilde{H}_x + \sum_{k=1}^N\left(\widetilde{F}_x(t_k) + \widetilde{F}_y(t_k)\widetilde{\gamma}^{t,x}_{t_k} + \widetilde{F}_z(t_k)\widetilde{\zeta}^{t,x}_{t_k}\right)\Delta t_k \\
&\quad - \sum_{j=1}^m\sum_{k=1}^N\left(\widetilde{G}_{j,x}(t_{k+1}) + \widetilde{G}_{j,y}(t_{k+1})\widetilde{\gamma}^{t,x}_{t_{k+1}}\right)\Delta\overleftarrow{B}^j_k - \sum_{k=1}^N\left(\widetilde{\zeta}^{t,x}_{t_k}\right)^T\Delta W_k + S^{(n)}
\end{aligned}$$

where $\Delta t_k := t_{k+1} - t_k$, $\Delta\overleftarrow{B}^j_k := \overleftarrow{B}^j_{t_{k+1}} - \overleftarrow{B}^j_{t_k}$, $\Delta W_k := W_{t_{k+1}} - W_{t_k}$ and $R^{(n)}$, $S^{(n)} \to 0$ in mean square as $n \to \infty$.

To see this, note that if we define $\widetilde{G}^{(n)}_{j,y}(r) := \widetilde{G}_{j,y}(t_{k+1})$ and $\overleftarrow{\Lambda}^{(n)}_r := \overleftarrow{\Lambda}_{t_{k+1}}$ for

$r \in [t_k, t_{k+1})$ then

$$E\left[\left|\sum_{j=1}^{m}\sum_{k=1}^{N}\widetilde{G}_{j,y}(t_{k+1})\overleftarrow{\Lambda}_{t_{k+1}}\Delta\overleftarrow{B}_k^j - \sum_{j=1}^{m}\int_s^T \widetilde{G}_{j,y}(r)\overleftarrow{\Lambda}_r d\overleftarrow{B}_r^j\right|^2\right]$$
$$\leq C\sum_{j=1}^{m} E\left[\int_s^T \left|\widetilde{G}_{j,y}^{(n)}(r)\overleftarrow{\Lambda}_r^{(n)} - \widetilde{G}_{j,y}(r)\overleftarrow{\Lambda}_r\right|^2 dr\right]$$
$$\to 0$$

by the dominated convergence theorem and the almost sure continuity of $\widetilde{G}_{j,y}(r)$ and $\overleftarrow{\Lambda}_r$.

Similarly, for $k = 1, \ldots, N$, there exist random variables $r_k^{(n)}$ and $s_k^{(n)}$ such that

$$\overleftarrow{\Lambda}_{t_k} = \overleftarrow{\Lambda}_{t_{k+1}} - \left(\widetilde{F}_y(t_k) - \sum_{j=1}^{m}|\widetilde{G}_{j,y}(t_k)|^2\right)\overleftarrow{\Lambda}_{t_k}\Delta t_k + \sum_{j=1}^{m}\widetilde{G}_{j,y}(t_{k+1})\overleftarrow{\Lambda}_{t_{k+1}}\Delta\overleftarrow{B}_k^j + r_k^{(n)},$$
$$\widetilde{\gamma}_{t_k}^{t,x} = \widetilde{\gamma}_{t_{k+1}}^{t,x} + \left(\widetilde{F}_x(t_k) + \widetilde{F}_y(t_k)\widetilde{\gamma}_{t_k}^{t,x} + \widetilde{F}_z(t_k)\widetilde{\zeta}_{t_k}^{t,x}\right)\Delta t_k$$
$$- \sum_{j=1}^{m}\left(\widetilde{G}_{j,x}(t_{k+1}) + \widetilde{G}_{j,y}(t_{k+1})\widetilde{\gamma}_{t_{k+1}}^{t,x}\right)\Delta\overleftarrow{B}_k^j - \left(\widetilde{\zeta}_{t_k}^{t,x}\right)^T \Delta W_k + s_k^{(n)}$$

and $r_k^{(n)}, s_k^{(n)} \to 0$ in mean square as $n \to \infty$.

Rearranging gives us that

$$\overleftarrow{\Lambda}_{t_{k+1}} + \sum_{j=1}^{m}\widetilde{G}_{j,y}(t_{k+1})\overleftarrow{\Lambda}_{t_{k+1}}\Delta\overleftarrow{B}_k^j = \overleftarrow{\Lambda}_{t_k} + \left(\widetilde{F}_y(t_k) - \sum_{j=1}^{m}|\widetilde{G}_{j,y}(t_k)|^2\right)\overleftarrow{\Lambda}_{t_k}\Delta t_k - r_k^{(n)}$$

and

$$\widetilde{\gamma}_{t_{k+1}}^{t,x} - \sum_{j=1}^{m}\left(\widetilde{G}_{j,x}(t_{k+1}) + \widetilde{G}_{j,y}(t_{k+1})\widetilde{\gamma}_{t_{k+1}}^{t,x}\right)\Delta\overleftarrow{B}_k^j$$
$$= \widetilde{\gamma}_{t_k}^{t,x} - \left(\widetilde{F}_x(t_k) + \widetilde{F}_y(t_k)\widetilde{\gamma}_{t_k}^{t,x} + \widetilde{F}_z(t_k)\widetilde{\zeta}_{t_k}^{t,x}\right)\Delta t_k + \left(\widetilde{\zeta}_{t_k}^{t,x}\right)^T \Delta W_k - s_k^{(n)}.$$

And so, discarding the $(\Delta t_k)^2$, $\Delta t_k \Delta W_k$, $\Delta t_k s_k^{(n)}$ and $\Delta t_k r_k^{(n)}$ terms, we have that

$$\overleftarrow{\Lambda}_{t_{k+1}}\widetilde{\gamma}_{t_{k+1}}^{t,x} - \sum_{j=1}^{m}\left(\widetilde{G}_{j,x}(t_{k+1}) + \widetilde{G}_{j,y}(t_{k+1})\widetilde{\gamma}_{t_{k+1}}^{t,x}\right)\overleftarrow{\Lambda}_{t_{k+1}}\Delta\overleftarrow{B}_k^j$$

$$
\begin{aligned}
&\quad + \sum_{j=1}^{m} \widetilde{G}_{j,y}(t_{k+1}) \overleftarrow{\Lambda}_{t_{k+1}} \widetilde{\gamma}^{t,x}_{t_{k+1}} \Delta \overleftarrow{B}^{j}_{k} \\
&\quad - \left(\sum_{j=1}^{m} \widetilde{G}_{j,y}(t_{k+1}) \overleftarrow{\Lambda}_{t_{k+1}} \Delta \overleftarrow{B}^{j}_{k} \right) \left(\sum_{j=1}^{m} \left(\widetilde{G}_{j,x}(t_{k+1}) + \widetilde{G}_{j,y}(t_{k+1}) \widetilde{\gamma}^{t,x}_{t_{k+1}} \right) \Delta \overleftarrow{B}^{j}_{k} \right) \\
&= \overleftarrow{\Lambda}_{t_k} \widetilde{\gamma}^{t,x}_{t_k} - \left(\widetilde{F}_x(t_k) + \widetilde{F}_y(t_k) \widetilde{\gamma}^{t,x}_{t_k} + \widetilde{F}_z(t_k) \widetilde{\zeta}^{t,x}_{t_k} \right) \overleftarrow{\Lambda}_{t_k} \Delta t_k \\
&\quad + \overleftarrow{\Lambda}_{t_k} \left(\widetilde{\zeta}^{t,x}_{t_k} \right)^T \Delta W_k + \left(\widetilde{F}_y(t_k) - \sum_{j=1}^{m} |\widetilde{G}_{j,y}(t_k)|^2 \right) \overleftarrow{\Lambda}_{t_k} \widetilde{\gamma}^{t,x}_{t_k} \Delta t_k \\
&\quad - s^{(n)}_k \overleftarrow{\Lambda}_{t_k} - r^{(n)}_k \widetilde{\gamma}^{t,x}_{t_k} - r^{(n)}_k \left(\widetilde{\zeta}^{t,x}_{t_k} \right)^T \Delta W_k + r^{(n)}_k s^{(n)}_k .
\end{aligned}
$$

Upon rearranging again and recalling that $\overleftarrow{\Lambda}_{t_k} = \overleftarrow{\Lambda}_T \Lambda_{t_k}$, it follows that

$$
\begin{aligned}
&\overleftarrow{\Lambda}_s \widetilde{\gamma}^{t,x}_s - \widetilde{H}_x \\
&= \sum_{k=1}^{N} \left(\overleftarrow{\Lambda}_{t_k} \widetilde{\gamma}^{t,x}_{t_k} - \overleftarrow{\Lambda}_{t_{k+1}} \widetilde{\gamma}^{t,x}_{t_{k+1}} \right) \\
&= \sum_{k=1}^{N} \left(\widetilde{F}_x(t_k) + \sum_{j=1}^{m} |\widetilde{G}_{j,y}(t_k)|^2 \widetilde{\gamma}^{t,x}_{t_k} + \widetilde{F}_z(t_k) \widetilde{\zeta}^{t,x}_{t_k} \right) \overleftarrow{\Lambda}_{t_k} \Delta t_k \\
&\quad - \sum_{k=1}^{N} \left(\sum_{j=1}^{m} \widetilde{G}_{j,y}(t_{k+1}) \overleftarrow{\Lambda}_{t_{k+1}} \Delta \overleftarrow{B}^{j}_{k} \right) \left(\sum_{j=1}^{m} \left(\widetilde{G}_{j,x}(t_{k+1}) + \widetilde{G}_{j,y}(t_{k+1}) \widetilde{\gamma}^{t,x}_{t_{k+1}} \right) \Delta \overleftarrow{B}^{j}_{k} \right) \\
&\quad - \sum_{j=1}^{m} \sum_{k=1}^{N} \widetilde{G}_{j,x}(t_{k+1}) \overleftarrow{\Lambda}_{t_{k+1}} \Delta \overleftarrow{B}^{j}_{k} - \overleftarrow{\Lambda}_T \sum_{k=1}^{N} \Lambda_{t_k} \left(\widetilde{\zeta}^{t,x}_{t_k} \right)^T \Delta W_k \\
&\quad + \sum_{k=1}^{N} \left(s^{(n)}_k \overleftarrow{\Lambda}_{t_k} + r^{(n)}_k \widetilde{\gamma}^{t,x}_{t_k} + r^{(n)}_k \left(\widetilde{\zeta}^{t,x}_{t_k} \right)^T \Delta W_k - r^{(n)}_k s^{(n)}_k \right) \\
&=: A_1 - A_2 - A_3 - A_4 + A_5.
\end{aligned}
$$

We will show the convergence in $L^1(\Omega, \mathbb{R})$ for each A_i.

$$
\begin{aligned}
A_1 &:= \sum_{k=1}^{N} \left(\widetilde{F}_x(t_k) + \sum_{j=1}^{m} |\widetilde{G}_{j,y}(t_k)|^2 \widetilde{\gamma}^{t,x}_{t_k} + \widetilde{F}_z(t_k) \widetilde{\zeta}^{t,x}_{t_k} \right) \overleftarrow{\Lambda}_{t_k} \Delta t_k \\
&\to \int_s^T \left(\widetilde{F}_x(r) + \sum_{j=1}^{m} |\widetilde{G}_{j,y}(r)|^2 \widetilde{\gamma}^{t,x}_r + \widetilde{F}_z(r) \widetilde{\zeta}^{t,x}_r \right) \overleftarrow{\Lambda}_r dr
\end{aligned}
$$

in $L^2(\Omega, \mathbb{R})$ as $n \to \infty$ (and so $N := N(n) \to \infty$). This is easy to see - we can just follow the argument above that showed that $S^{(n)} \to 0$ in $L^2(\Omega, \mathbb{R})$ as $n \to \infty$.

For A_2, first

$$\begin{aligned}
&\sum_{j=1}^{m} \sum_{k=1}^{N} \widetilde{G}_{j,y}(t_{k+1}) \left(\widetilde{G}_{j,x}(t_{k+1}) + \widetilde{G}_{j,y}(t_{k+1}) \widetilde{\gamma}^{t,x}_{t_{k+1}} \right) \overleftarrow{\Lambda}_{t_{k+1}} \left(\Delta \overleftarrow{B}^{j}_{k} \right)^2 \\
&\to \sum_{j=1}^{m} \int_s^T \widetilde{G}_{j,y}(r) \left(\widetilde{G}_{j,x}(r) + \widetilde{G}_{j,y}(r) \widetilde{\gamma}^{t,x}_{r} \right) \overleftarrow{\Lambda}_{r} dr
\end{aligned}$$

in $L^2(\Omega, \mathbb{R})$ as $n \to \infty$. This is standard (see [23]) using the boundedness of $\overleftarrow{\Lambda}$ and $\widetilde{\gamma}^{t,x}$ afforded by introducing T_n in (6.2). As a consequence, it is sufficient to show that in mean square,

$$\sum_{k=1}^{N} \sum_{i \neq j} \widetilde{G}_{j,y}(t_{k+1}) \left(\widetilde{G}_{i,x}(t_{k+1}) + \widetilde{G}_{i,y}(t_{k+1}) \widetilde{\gamma}^{t,x}_{t_{k+1}} \right) \overleftarrow{\Lambda}_{t_{k+1}} \Delta \overleftarrow{B}^{i}_{k} \Delta \overleftarrow{B}^{j}_{k} \to 0.$$

Denoting $G(t_k)_{i,j} := \widetilde{G}_{j,y}(t_k) \left(\widetilde{G}_{i,x}(t_k) + \widetilde{G}_{i,y}(t_k) \widetilde{\gamma}^{t,x}_{t_k} \right) \overleftarrow{\Lambda}_{t_k}$, we have that

$$\begin{aligned}
&E\left[\left(\sum_{k=1}^{N} \sum_{i \neq j} G(t_{k+1})_{i,j} \Delta \overleftarrow{B}^{i}_{k} \Delta \overleftarrow{B}^{j}_{k} \right)^2 \right] \\
&= E\left[\sum_{k=1}^{N} \left(\sum_{i \neq j} G(t_{k+1})_{i,j} \Delta \overleftarrow{B}^{i}_{k} \Delta \overleftarrow{B}^{j}_{k} \right)^2 \right] \\
&\quad + 2E\left[\sum_{k<l} \left(\sum_{i \neq j} G(t_{k+1})_{i,j} \Delta \overleftarrow{B}^{i}_{k} \Delta \overleftarrow{B}^{j}_{k} \right) \left(\sum_{i \neq j} G(t_{l+1})_{i,j} \Delta \overleftarrow{B}^{i}_{l} \Delta \overleftarrow{B}^{j}_{l} \right) \right] \\
&= E\left[\sum_{k=1}^{N} \sum_{i \neq j} \left(G(t_{k+1})_{i,j} + G(t_{k+1})_{j,i} \right)^2 \left(\Delta \overleftarrow{B}^{i}_{k} \right)^2 \left(\Delta \overleftarrow{B}^{j}_{k} \right)^2 \right] \\
&\le C \max_k \{\Delta t_k\} \to 0 \text{ as } n \to \infty.
\end{aligned}$$

To see this, first note that

$$\begin{aligned}
&\left(\sum_{i \neq j} G(t_{k+1})_{i,j} \Delta \overleftarrow{B}^{i}_{k} \Delta \overleftarrow{B}^{j}_{k} \right)^2 \\
&= \sum_{i \neq j} \left(G(t_{k+1})_{i,j} + G(t_{k+1})_{j,i} \right)^2 \left(\Delta \overleftarrow{B}^{i}_{k} \right)^2 \left(\Delta \overleftarrow{B}^{j}_{k} \right)^2
\end{aligned}$$

$$+\sum_{i_1\neq j_1, i_2\neq j_2} G(t_{k+1})_{i_1,j_1} G(t_{k+1})_{i_2,j_2} \Delta\overleftarrow{B}_k^{i_1} \Delta\overleftarrow{B}_k^{j_1} \Delta\overleftarrow{B}_k^{i_2} \Delta\overleftarrow{B}_k^{j_2}$$

where additionally $i_1 \neq j_2$ and/or $i_2 \neq j_1$. This latter sum has zero expectation since, taking an example term with $l := i_2 = j_1$, we have that

$$\begin{aligned}
&E\left[G(t_{k+1})_{i_1,l} G(t_{k+1})_{l,j_2} \left(\Delta\overleftarrow{B}_k^{l}\right)^2 \Delta\overleftarrow{B}_k^{i_1} \Delta\overleftarrow{B}_k^{j_2}\right] \\
&= E\left[G(t_{k+1})_{i_1,l} G(t_{k+1})_{l,j_2} E\left[\left(\Delta\overleftarrow{B}_k^{l}\right)^2 \Delta\overleftarrow{B}_k^{i_1} \Delta\overleftarrow{B}_k^{j_2} \middle| \mathcal{F}_{t,T}^W \vee \mathcal{F}_{t_{k+1},T}^B\right]\right] \\
&= 0.
\end{aligned}$$

Secondly, we have that since $k < l$, $i_1 \neq j_1$ and $i_2 \neq j_2$

$$\begin{aligned}
&E\left[G(t_{k+1})_{i_1,j_1} \Delta\overleftarrow{B}_k^{i_1} \Delta\overleftarrow{B}_k^{j_1} G(t_{l+1})_{i_2,j_2} \Delta\overleftarrow{B}_l^{i_2} \Delta\overleftarrow{B}_l^{j_2}\right] \\
&= E\left[G(t_{k+1})_{i_1,j_1} G(t_{l+1})_{i_2,j_2} \Delta\overleftarrow{B}_k^{i_1} \Delta\overleftarrow{B}_k^{j_1} E\left[\Delta\overleftarrow{B}_l^{i_2} \Delta\overleftarrow{B}_l^{j_2} \middle| \mathcal{F}_{t,T}^W \vee \mathcal{F}_{t_{l+1},T}^B\right]\right] \\
&= 0.
\end{aligned}$$

Finally,

$$\begin{aligned}
&\sum_{k=1}^{N} \sum_{i\neq j} E\left[\left(G(t_{k+1})_{i,j} + G(t_{k+1})_{j,i}\right)^2 \left(\Delta\overleftarrow{B}_k^{i}\right)^2 \left(\Delta\overleftarrow{B}_k^{j}\right)^2\right] \\
&= \sum_{k=1}^{N} \sum_{i\neq j} E\left[\left(G(t_{k+1})_{i,j} + G(t_{k+1})_{j,i}\right)^2 E\left[\left(\Delta\overleftarrow{B}_k^{i}\right)^2 \left(\Delta\overleftarrow{B}_k^{j}\right)^2 \middle| \mathcal{F}_{t,T}^W \vee \mathcal{F}_{t_{k+1},T}^B\right]\right] \\
&\leq \max_k\{\Delta t_k\} \sum_{k=1}^{N} \sum_{i\neq j} E\left[\left(G(t_{k+1})_{i,j} + G(t_{k+1})_{j,i}\right)^2 \Delta t_k\right] \\
&\leq C \max_k\{\Delta t_k\}
\end{aligned}$$

since

$$E\left[\sum_{k=1}^{N} \left(G(t_{k+1})_{i,j} + G(t_{k+1})_{j,i}\right)^2 \Delta t_k\right] \to E\left[\int_s^T \left(G(r)_{i,j} + G(r)_{j,i}\right)^2 dr\right] < \infty.$$

$$A_3 := \sum_{j=1}^{m} \sum_{k=1}^{N} \widetilde{G}_{j,x}(t_{k+1}) \overleftarrow{\Lambda}_{t_{k+1}} \Delta\overleftarrow{B}_k^{j} \to \sum_{j=1}^{m} \int_s^T \widetilde{G}_{j,x}(r) \overleftarrow{\Lambda}_r \overleftarrow{dB}_r^{j}$$

in $L^2(\Omega,\mathbb{R})$ as $n\to\infty$ is standard.

$$A_4 := \overleftarrow{\Lambda}_T \sum_{k=1}^{N} \Lambda_{t_k} \left(\widetilde{\zeta}_{t_k}^{t,x}\right)^T \Delta W_k \to \overleftarrow{\Lambda}_T \int_s^T \Lambda_r \left(\widetilde{\zeta}_r^{t,x}\right)^T dW_r$$

in $L^1(\Omega,\mathbb{R})$ as $n\to\infty$. To see this, note that

$$\begin{aligned} &E\left[\left|\overleftarrow{\Lambda}_T \sum_{k=1}^{N} \Lambda_{t_k} \left(\widetilde{\zeta}_{t_k}^{t,x}\right)^T \Delta W_k - \overleftarrow{\Lambda}_T \int_s^T \Lambda_r \left(\widetilde{\zeta}_r^{t,x}\right)^T dW_r\right|\right] \\ &\le E\left[\left|\overleftarrow{\Lambda}_T\right|^2\right]^{1/2} E\left[\left|\sum_{k=1}^{N} \Lambda_{t_k} \left(\widetilde{\zeta}_{t_k}^{t,x}\right)^T \Delta W_k - \int_s^T \Lambda_r \left(\widetilde{\zeta}_r^{t,x}\right)^T dW_r\right|^2\right]^{1/2} \end{aligned}$$

and since $(\widetilde{\gamma},\widetilde{\zeta})$ is the solution of a BDSDE with smooth coefficients, we can (see [36]) take $\widetilde{\zeta}$ to be continuous.

We will now show that

$$A_5 := \sum_{k=1}^{N} s_k^{(n)} \overleftarrow{\Lambda}_{t_k} + r_k^{(n)} \widetilde{\gamma}_{t_k}^{t,x} + r_k^{(n)} \left(\widetilde{\zeta}_{t_k}^{t,x}\right)^T \Delta W_k - r_k^{(n)} s_k^{(n)} \to 0$$

in $L^1(\Omega,\mathbb{R})$ as $n\to\infty$. It is sufficient to show that

$$\sum_{k=1}^{N} E\left[|r_k^{(n)}|^2 + |s_k^{(n)}|^2\right] \to 0$$

as $n\to\infty$. To see that this is sufficient, note for example that

$$\begin{aligned} E\left[\left|\sum_{k=1}^{N} r_k^{(n)} \left(\widetilde{\zeta}_{t_k}^{t,x}\right)^T \Delta W_k\right|\right] &\le E\left[\left(\sum_{k=1}^{N} \left|r_k^{(n)}\right|^2\right)^{1/2} \left(\sum_{k=1}^{N} \left|\left(\widetilde{\zeta}_{t_k}^{t,x}\right)^T \Delta W_k\right|^2\right)^{1/2}\right] \\ &\le \left(E\left[\sum_{k=1}^{N} \left|r_k^{(n)}\right|^2\right]\right)^{1/2} \left(E\left[\sum_{k=1}^{N} \left|\left(\widetilde{\zeta}_{t_k}^{t,x}\right)^T \Delta W_k\right|^2\right]\right)^{1/2} \end{aligned}$$

and

$$E\left[\sum_{k=1}^{N} \left|\left(\widetilde{\zeta}_{t_k}^{t,x}\right)^T \Delta W_k\right|^2\right] = E\left[\sum_{k=1}^{N} \left|\widetilde{\zeta}_{t_k}^{t,x}\right|^2 \Delta t_k\right] \to E\left[\int_s^T \left|\widetilde{\zeta}_r^{t,x}\right|^2 dr\right] < \infty$$

as $n \to \infty$. We now follow the same arguments as we did to show that

$$E\left[|R^{(n)}|^2 + |S^{(n)}|^2\right] \to 0$$

as $n \to \infty$. For a step function ϕ defined on $\Pi^{(n)}$ we define $\phi^{(n)}(r) := \phi(t_k)$ and $\phi^{(n+)}(r) := \phi(t_{k+1})$ for $r \in [t_k, t_{k+1}) \subset \Pi^{(n)}$. Let us consider $r_k^{(n)}$ (the argument for $s_k^{(n)}$ is similar):

$$\begin{aligned}
&E\left[\left|r_k^{(n)}\right|^2\right] \\
&= E\left[\left|\overleftarrow{\Lambda}_{t_k} - \overleftarrow{\Lambda}_{t_{k+1}} + \left(\widetilde{F}_y(t_k) - \sum_{j=1}^m \left|\widetilde{G}_{j,y}(t_k)\right|^2\right)\overleftarrow{\Lambda}_{t_k}\Delta t_k\right.\right. \\
&\quad \left.\left. - \sum_{j=1}^m \widetilde{G}_{j,y}(t_{k+1})\overleftarrow{\Lambda}_{t_{k+1}}\Delta \overleftarrow{B}_k^j\right|^2\right] \\
&= E\left[\left|\int_{t_k}^{t_{k+1}} \left(\left(\widetilde{F}_y^{(n)}(r) - \sum_{j=1}^m \left|\widetilde{G}_{j,y}^{(n)}(r)\right|^2\right)\overleftarrow{\Lambda}_r^{(n)} - \left(\widetilde{F}_y(r) - \sum_{j=1}^m \left|\widetilde{G}_{j,y}(r)\right|^2\right)\overleftarrow{\Lambda}_r\right) dr\right.\right. \\
&\quad \left.\left. - \sum_{j=1}^m \int_{t_k}^{t_{k+1}} \left(\widetilde{G}_{j,y}^{(n+)}(r)\overleftarrow{\Lambda}_r^{(n)} - \widetilde{G}_{j,y}(r)\overleftarrow{\Lambda}_r\right) \overleftarrow{dB}_r^j\right|^2\right].
\end{aligned}$$

Now,

$$\begin{aligned}
&E\left[\left|\int_{t_k}^{t_{k+1}} \left\{\widetilde{F}_y^{(n)}(r)\overleftarrow{\Lambda}_r^{(n)} - \widetilde{F}_y(r)\overleftarrow{\Lambda}_r\right\} dr\right|^2\right] \\
&\le CE\left[\int_{t_k}^{t_{k+1}} \left|\widetilde{F}_y^{(n)}(r)\overleftarrow{\Lambda}_r^{(n)} - \widetilde{F}_y(r)\overleftarrow{\Lambda}_r\right|^2 dr\right] \\
&\le CE\left[\int_{t_k}^{t_{k+1}} \left\{\left|\widetilde{F}_y^{(n)}(r)\left(\overleftarrow{\Lambda}_r^{(n)} - \overleftarrow{\Lambda}_r\right)\right|^2 + \left|\overleftarrow{\Lambda}_r\left(\widetilde{F}_y^{(n)}(r) - \widetilde{F}_y(r)\right)\right|^2\right\} dr\right]
\end{aligned}$$

and similarly for $\widetilde{G}$. Furthermore,

$$\begin{aligned}
&E\left[\left|\sum_{j=1}^m \int_{t_k}^{t_{k+1}} \widetilde{G}_{j,y}^{(n)}(r)\overleftarrow{\Lambda}_r^{(n)} - \widetilde{G}_{j,y}(r)\overleftarrow{\Lambda}_r \overleftarrow{dB}_r^j\right|^2\right] \\
&\le C\sum_{j=1}^m E\left[\int_{t_k}^{t_{k+1}} \left|\widetilde{G}_{j,y}^{(n)}(r)\overleftarrow{\Lambda}_r^{(n)} - \widetilde{G}_{j,y}(r)\overleftarrow{\Lambda}_r\right|^2 dr\right]
\end{aligned}$$

$$= C\Delta t_k \sum_{j=1}^{m} E\left[\sup_{r\in[s,T]} \left|\widetilde{G}_{j,y}(t_k)\overleftarrow{\Lambda}_{t_k} - \widetilde{G}_{j,y}(r)\overleftarrow{\Lambda}_r\right|^2\right].$$

It follows that

$$\begin{aligned}
&\sum_{k=1}^{N} E\left[|r_k^{(n)}|^2\right] \\
&\leq C\Bigg\{E\Bigg[\int_s^T \Bigg\{\left|\widetilde{F}_y^{(n)}(r)\left(\overleftarrow{\Lambda}_r^{(n)} - \overleftarrow{\Lambda}_r\right)\right|^2 + \left|\overleftarrow{\Lambda}_r\left(\widetilde{F}_y^{(n)}(r) - \widetilde{F}_y(r)\right)\right|^2 \\
&\quad + \left|\sum_{j=1}^{m}|\widetilde{G}_{j,y}^{(n)}(r)|^2\left(\overleftarrow{\Lambda}_r^{(n)} - \overleftarrow{\Lambda}_r\right)\right|^2 + \left|\sum_{j=1}^{m}\overleftarrow{\Lambda}_r\left(|\widetilde{G}_{j,y}^{(n)}(r)|^2 - |\widetilde{G}_{j,y}(r)|^2\right)\right|^2\Bigg\}dr\Bigg] \\
&\quad + E\left[\sum_{j=1}^{m}\sup_{r\in[s,T]}\left|\widetilde{G}_{j,y}(t_k)\overleftarrow{\Lambda}_{t_k} - \widetilde{G}_{j,y}(r)\overleftarrow{\Lambda}_r\right|^4\right]^{1/2}\Bigg\}
\end{aligned}$$

$\to 0$ as $n \to \infty$ by the dominated convergence theorem and the almost sure continuity of f_y, g_y and $\overleftarrow{\Lambda}$. Arguing almost verbatim from [23], we now note that by taking some subsequence of partitions the result holds almost surely for each $s \in [t, T]$ and so the processes on either side of our equality are modifications of each other and since they are both almost surely continuous, they are indistinguishable. □

We now prove our result on the regularity of Z with the help of Lemma 6.8.

Theorem 6.9. *Let $(\widetilde{Y}, \widetilde{Z})$ denote the solution to the BDSDE with smooth coefficients $\{\widetilde{f}, \widetilde{g}, \widetilde{h}\}$ as defined in Definition 5.18 and $\{t_0, t_1, \ldots, t_n\}$ be a partition without cluster with granularity Δt as defined in Definition 6.1. Then there exists a constant $C > 0$ independent of C_E, C_M, C_G and Δt such that*

$$\sum_{i=1}^{n} E\left[\int_{t_{i-1}}^{t_i}\int_{\mathbb{R}^d}\left(|\widetilde{Z}_r^{t,x} - \widetilde{Z}_{t_{i-1}}^{t,x}|^2 + |\widetilde{Z}_r^{t,x} - \widetilde{Z}_{t_i}^{t,x}|^2\right)\rho(x)dxdr\right] \leq Cm\Delta t C_G^2\sqrt{\mu_{\mathcal{X}}}$$

where $\mu_{\mathcal{X}}$ is as defined in Definition 6.3.

Proof. In this proof we adapt the approach of [49] and so our approach is similar to that found in [2]. We separate the proof into two steps - in the first, we derive an alternative representation for Z and in the second we utilise this representation to obtain the desired inequality.

<u>Step 1</u>: By Lemma 6.8 we have that

$$\Lambda_s \widetilde{\gamma}_s^{t,x} = \Lambda_T \widetilde{H}_x + \int_s^T \Lambda_r \left(\widetilde{K}(r) + \widetilde{F}_z(r) \widetilde{\zeta}_r^{t,x} \right) dr - \Lambda_T \sum_{j=1}^m \int_s^T \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r) \overleftarrow{dB}_r^j$$
$$- \int_s^T \Lambda_r \left(\widetilde{\zeta}_r^{t,x} \right)^T dW_r.$$

Let us now define

$$M_s := \exp \left\{ \int_0^s \left\langle \widetilde{F}_z(r), dW_r \right\rangle - \frac{1}{2} \int_0^s |\widetilde{F}_z(r)|^2 dr \right\}$$

with $\widetilde{F}_z(r) \equiv 0$ for $r < t$ and note that M is a martingale with respect to the filtration $\mathcal{G}_s := \mathcal{F}_s^W \vee \mathcal{F}_T^B$. By Girsanov's theorem, we have that $\{\widetilde{W}_s, \mathcal{G}_s; 0 \leq s \leq T\}$ defined for $i = 1, \ldots, d$ by

$$\widetilde{W}_s^i := W_s^i - \int_0^s \widetilde{F}_{z_i}(r) dr$$

is a d-dimensional BM on $(\Omega, \mathcal{G}_T, Q)$, where Q is defined for $A \in \mathcal{G}_T$ by $Q(A) := E\left[\mathbb{I}_A M_T\right]$. It follows that

$$\Lambda_s \widetilde{\gamma}_s^{t,x} = \Lambda_T \widetilde{H}_x + \int_s^T \Lambda_r \widetilde{K}(r) dr - \Lambda_T \sum_{j=1}^m \int_s^T \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r) \overleftarrow{dB}_r^j$$
$$- \int_s^T \Lambda_r \left(\bar{Z}_r^{t,x} \right)^T d\widetilde{W}_r.$$

Putting

$$\begin{aligned}
\gamma_s^{t,x} &:= \Lambda_s \widetilde{\gamma}_s^{t,x} + \int_t^s \Lambda_r \widetilde{K}(r) dr - \Lambda_T \sum_{j=1}^m \int_t^s \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r) \overleftarrow{dB}_r^j, \\
\xi^{t,x} &:= \gamma_T^{t,x}, \\
\zeta_s^{t,x} &:= \Lambda_s \left(\bar{Z}_s^{t,x} \right)^T,
\end{aligned} \tag{6.4}$$

we have that

$$\gamma_s^{t,x} = \xi^{t,x} - \int_s^T \zeta_r^{t,x} d\widetilde{W}_r.$$

Now,

$$\int_s^T \zeta_r^{t,x} d\widetilde{W}_r = \Lambda_T \widetilde{H}_x - \Lambda_s \widetilde{\gamma}_s^{t,x} + \int_s^T \Lambda_r \widetilde{K}(r)dr - \Lambda_T \sum_{j=1}^m \int_s^T \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r) \overleftarrow{dB}_r^j$$

and since the right hand side is square-integrable it follows that the left hand side is a martingale and not just a local martingale with respect to the filtration $\mathcal{G}$. As a consequence, we may write

$$\gamma_s^{t,x} = E_Q\left[\xi^{t,x}|\mathcal{G}_s\right]$$

and so

$$\widetilde{\gamma}_s^{t,x} = \Lambda_s^{-1} E_Q\left[\xi^{t,x}|\mathcal{G}_s\right] - \Lambda_s^{-1} \int_t^s \Lambda_r \widetilde{K}(r)dr + \Lambda_s^{-1}\Lambda_T \sum_{j=1}^m \int_t^s \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r) \overleftarrow{dB}_r^j.$$

By Bayes' rule, we have that

$$E_Q\left[\xi^{t,x}|\mathcal{G}_s\right] = M_s^{-1} E\left[M_T \xi^{t,x}|\mathcal{G}_s\right].$$

It then follows by the Martingale Representation Theorem that

$$\widetilde{\gamma}_s^{t,x} = \Lambda_s^{-1} M_s^{-1} \xi_s^{t,x} - \Lambda_s^{-1} \int_t^s \Lambda_r \widetilde{K}(r)dr + \Lambda_s^{-1}\Lambda_T \sum_{j=1}^m \int_t^s \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r) \overleftarrow{dB}_r^j.$$

where

$$\xi_s^{t,x} := E\left[M_T \xi^{t,x}|\mathcal{G}_s\right] = E\left[M_T \xi^{t,x}|\mathcal{G}_t\right] + \int_t^s \eta_r dW_r \tag{6.5}$$

for some η satisfying

$$E\left[\int_t^T |\eta_r|^2 dr\right] < \infty.$$

It follows that

$$\widetilde{Z}_s^{t,x} = (\nabla \widetilde{Y}_s^{t,x})^T (\nabla X_s^{t,x})^{-1} \sigma(X_s^{t,x})$$

$$= \left(M_s^{-1}\xi_s^{t,x} - \int_t^s \Lambda_r \widetilde{K}(r)dr + \Lambda_T \sum_{j=1}^m \int_t^s \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r) \overleftarrow{dB}_r^j \right)^T V_s$$

where $V_s := \Lambda_s^{-1}(\nabla X_s^{t,x})^{-1}\sigma(X_s^{t,x})$.

Step 2: Let us suppose that $t \leq s \leq u \leq T$ such that $u - s \leq \Delta t$. Then

$$|Z_u^{t,x} - Z_s^{t,x}| \leq I_1(s,u) + I_2(s,u) + I_3(s,u) + I_4(s,u)$$

where

$$\begin{aligned}
I_1(s,u) &:= \left| \left(\int_t^u \Lambda_r \widetilde{K}(r)dr \right)^T V_u - \left(\int_t^s \Lambda_r \widetilde{K}(r)dr \right)^T V_s \right|, \\
I_2(s,u) &:= \left| \left(\Lambda_T \sum_{j=1}^m \int_t^u \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r) \overleftarrow{dB}_r^j \right)^T V_u - \left(\Lambda_T \sum_{j=1}^m \int_t^s \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r) \overleftarrow{dB}_r^j \right)^T V_s \right|, \\
I_3(s,u) &:= |\xi_u^{t,x}||M_u^{-1}V_u - M_s^{-1}V_s|, \\
I_4(s,u) &:= |\xi_u^{t,x} - \xi_s^{t,x}||M_s^{-1}V_s|.
\end{aligned}$$

Let us define for $r \leq s$, $\Lambda_r^s := \Lambda_s \Lambda_r^{-1}$. Then since each of f_y, f_z and g_y is bounded, we can show that for each $p \geq 1$ there exists a constant $C > 0$ such that

$$\begin{aligned}
&E\left[\sup_{t \leq r \leq s \leq T} \left\{ |\Lambda_r^s|^p + |\Lambda_s^{-1}|^p + |\overleftarrow{\Lambda}_s|^p + |M_s|^p + |M_s^{-1}|^p \right\} \right] \leq C, \\
&E\left[|\Lambda_u^T - \Lambda_s^T|^p + |\Lambda_u^{-1} - \Lambda_s^{-1}|^p + |M_u - M_s|^p + |M_u^{-1} - M_s^{-1}|^p \right] \leq C(\Delta t)^{p/2}.
\end{aligned} \tag{6.6}$$

We omit the proof for I_1 and proceed to I_2 which is similar.

Case 1 (I_2):

$$I_2(s,u) \leq \left| \sum_{j=1}^m \int_s^u \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r) \overleftarrow{dB}_r^j \right| |\Lambda_T V_u| + \left| \sum_{j=1}^m \int_t^s \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r) \overleftarrow{dB}_r^j \right| |\Lambda_T V_u - \Lambda_T V_s| .$$

Now,

$$\begin{aligned}
&\mathcal{E}\left[\left|\sum_{j=1}^{m}\int_s^u \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r)\overleftarrow{dB}_r^j\right|^2 |\Lambda_T V_u|^2\right] \\
&\leq \left(\mathcal{E}\left[\left|\sum_{j=1}^{m}\int_s^u \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r)\overleftarrow{dB}_r^j\right|^4\right]\right)^{1/2} \left(\mathcal{E}\left[|\Lambda_T V_u|^4\right]\right)^{1/2}.
\end{aligned}$$

Furthermore, by (6.6) and the Cauchy-Schwarz and Jensen's inequalities,

$$\begin{aligned}
&\mathcal{E}\left[\left|\sum_{j=1}^{m}\int_s^u \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r)\overleftarrow{dB}_r^j\right|^4\right] \\
&= \mathcal{E}\left[\sum_{j=1}^{m}\left|\int_s^u \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r)\overleftarrow{dB}_r^j\right|^4 + 6\sum_{j=2}^{m}\sum_{i=1}^{j-1}\left|\int_s^u \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r)\overleftarrow{dB}_r^j\right|^2 \left|\int_s^u \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r)\overleftarrow{dB}_r^i\right|^2\right] \\
&\leq (1+3(m-1))\mathcal{E}\left[\sum_{j=1}^{m}\left|\int_s^u \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r)\overleftarrow{dB}_r^j\right|^4\right] \\
&\leq Cm\mathcal{E}\left[\left(\sum_{j=1}^{m}\int_s^u |\overleftarrow{\Lambda}_r|^2 |\widetilde{G}_{j,x}(r)|^2 dr\right)^2\right] \\
&\leq Cm^2\sum_{j=1}^{m}\mathcal{E}\left[\left(\int_s^u |\overleftarrow{\Lambda}_r|^2 |\widetilde{G}_{j,x}(r)|^2 dr\right)^2\right] \\
&\leq Cm^2\sum_{j=1}^{m}\mathcal{E}\left[\int_s^u |\overleftarrow{\Lambda}_r|^4 dr\int_s^u |\widetilde{G}_{j,x}(r)|^4 dr\right] \\
&\leq Cm^2\sum_{j=1}^{m}\left(\mathcal{E}\left[\left(\int_s^u |\overleftarrow{\Lambda}_r|^4 dr\right)^2\right]\mathcal{E}\left[\left(\int_s^u |\widetilde{G}_{j,x}(r)|^4 dr\right)^2\right]\right)^{1/2} \\
&\leq Cm^2\Delta t\sum_{j=1}^{m}\left(\mathcal{E}\left[\int_s^u |\overleftarrow{\Lambda}_r|^8 dr\right]\mathcal{E}\left[\int_s^u |\widetilde{G}_{j,x}(r)|^8 dr\right]\right)^{1/2} \\
&\leq Cm^2(\Delta t)^2\sum_{j=1}^{m}\left(\sup_{r\in[s,u]}\mathcal{E}\left[|\overleftarrow{\Lambda}_r|^8\right]\sup_{r\in[s,u]}\mathcal{E}\left[|\widetilde{G}_{j,x}(r)|^8\right]\right)^{1/2} \\
&\leq Cm^2(\Delta t)^2 C_G^4\sum_{j=1}^{m}\sqrt{\mu(\mathfrak{B}_{\mathcal{X}^{g_j}})} \\
&\leq Cm^2(\Delta t)^2 C_G^4 \mu_{\mathcal{X}}
\end{aligned}$$

since

$$\mathcal{E}\left[|\widetilde{G}_{j,x}(r)|^8\right] \leq C_G^8 \mathcal{E}\left[\mathbb{I}_{\mathfrak{B}_{\mathcal{X}^{g_j}}}(X_r^{t,x})\right] \leq C C_G^8 \mu(\mathfrak{B}_{\mathcal{X}^{g_j}}).$$

It is easy to show that $\mathcal{E}\left[|\Lambda_T V_{t_i}|^4\right] < \infty$ and so

$$\mathcal{E}\left[\left|\sum_{j=1}^m \int_s^u \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r) \overleftarrow{dB}_r^j\right|^2 |\Lambda_T V_{t_i}|^2\right] \leq C m \Delta t C_G^2 \sqrt{\mu_{\mathcal{X}}}.$$

Similarly,

$$\begin{aligned}
&\mathcal{E}\left[\left|\sum_{j=1}^m \int_t^s \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r) \overleftarrow{dB}_r^j\right|^2 |\Lambda_T V_u - \Lambda_T V_s|^2\right] \\
&\leq \left(\mathcal{E}\left[\left|\sum_{j=1}^m \int_t^s \overleftarrow{\Lambda}_r \widetilde{G}_{j,x}(r) \overleftarrow{dB}_r^j\right|^4\right]\right)^{1/2} \left(\mathcal{E}\left[|\Lambda_T V_u - \Lambda_T V_s|^4\right]\right)^{1/2} \\
&\leq C m C_G^2 \sqrt{\mu_{\mathcal{X}}} \sqrt{\mathcal{E}\left[|\Lambda_T V_u - \Lambda_T V_s|^4\right]}.
\end{aligned}$$

Now

$$\begin{aligned}
&\mathcal{E}\left[|\Lambda_T V_u - \Lambda_T V_s|^4\right] \\
&= \mathcal{E}\left[\left|\Lambda_u^T (\nabla X_u^{t,x})^{-1} \sigma(X_u^{t,x}) - \Lambda_s^T (\nabla X_s^{t,x})^{-1} \sigma(X_s^{t,x})\right|^4\right] \\
&\leq C\left\{\mathcal{E}\left[\left|\Lambda_u^T - \Lambda_s^T\right|^4 \left|(\nabla X_u^{t,x})^{-1} \sigma(X_u^{t,x})\right|^4\right]\right. \\
&\quad \left. + \mathcal{E}\left[\left|\Lambda_s^T\right|^4 \left|(\nabla X_u^{t,x})^{-1} \sigma(X_u^{t,x}) - (\nabla X_s^{t,x})^{-1} \sigma(X_s^{t,x})\right|^4\right]\right\} \\
&\leq C\left\{\left(\mathcal{E}\left[\left|\Lambda_u^T - \Lambda_s^T\right|^8\right] \mathcal{E}\left[\left|(\nabla X_u^{t,x})^{-1} \sigma(X_u^{t,x})\right|^8\right]\right)^{1/2}\right. \\
&\quad \left. + \left(\mathcal{E}\left[\left|\Lambda_s^T\right|^8\right] \mathcal{E}\left[\left|(\nabla X_u^{t,x})^{-1} \sigma(X_u^{t,x}) - (\nabla X_s^{t,x})^{-1} \sigma(X_s^{t,x})\right|^8\right]\right)^{1/2}\right\}.
\end{aligned}$$

By Results A.1 and A.2,

$$\begin{aligned}
&\mathcal{E}\left[\left|(\nabla X_u^{t,x})^{-1} \sigma(X_u^{t,x}) - (\nabla X_s^{t,x})^{-1} \sigma(X_s^{t,x})\right|^8\right] \\
&\leq C\left\{\mathcal{E}\left[\left|(\nabla X_u^{t,x})^{-1} - (\nabla X_s^{t,x})^{-1}\right|^8 \left|\sigma(X_s^{t,x})\right|^8\right]\right. \\
&\quad \left. + \mathcal{E}\left[\left|(\nabla X_u^{t,x})^{-1}\right|^8 \left|\sigma(X_u^{t,x}) - \sigma(X_s^{t,x})\right|^8\right]\right\}
\end{aligned}$$

$$
\begin{aligned}
&\le C\Bigg\{\left(\mathcal{E}\left[\left|(\nabla X_u^{t,x})^{-1}-(\nabla X_s^{t,x})^{-1}\right|^{16}\right]\mathcal{E}\left[\left|\sigma(X_s^{t,x})\right|^{16}\right]\right)^{1/2}\\
&\quad+\left(\mathcal{E}\left[\left|(\nabla X_u^{t,x})^{-1}\right|^{16}\right]\mathcal{E}\left[\left|\sigma(X_u^{t,x})-\sigma(X_s^{t,x})\right|^{16}\right]\right)^{1/2}\Bigg\}\\
&\le C(\Delta t)^4.
\end{aligned}
$$

Recalling Result A.1 and equation (6.6) we have that $\mathcal{E}\left[|\Lambda_T V_u-\Lambda_T V_s|^4\right]\le C(\Delta t)^2$ and so

$$
\mathcal{E}\left[|I_2(s,u)|^2\right]\le Cm\Delta t C_G^2\sqrt{\mu_{\mathcal{X}}}.
$$

Since $\mathcal{U}$ is a partition without cluster it follows that

$$
\sum_{i=1}^n \sup_{t_{i-1}\le u\le t_i}\mathcal{E}\left[|I_2(t_{i-1},u)|^2\right]\le CmC_G^2\sqrt{\mu_{\mathcal{X}}}.
$$

<u>Case 2 (I_3)</u>: Since $s\le u$ and V and M are adapted to the filtration $\mathcal{G}$, we have by equations (6.4) and (6.5) that

$$
\begin{aligned}
&\mathcal{E}\left[|I_3(s,u)|^2\right]\\
&=\mathcal{E}\left[|\xi_u^{t,x}|^2|M_u^{-1}V_u-M_s^{-1}V_s|^2\right]\\
&\le\mathcal{E}\left[E\left[|M_T\xi^{t,x}|^2|\mathcal{G}_u\right]|M_u^{-1}V_u-M_s^{-1}V_s|^2\right]\\
&=\mathcal{E}\left[|M_T\xi^{t,x}|^2|M_u^{-1}V_u-M_s^{-1}V_s|^2\right]\\
&\le\left(\mathcal{E}\left[|\xi^{t,x}|^4\right]\right)^{1/2}\left(\mathcal{E}\left[|M_T|^4|M_u^{-1}V_u-M_s^{-1}V_s|^4\right]\right)^{1/2}\\
&\le C\left(\mathcal{E}\left[|\Lambda_T\widetilde{H}_x|^4+\left|\int_t^T\Lambda_r\widetilde{K}(r)dr\right|^4+\left|\Lambda_T\sum_{j=1}^m\int_t^T\overleftarrow{\Lambda}_r\widetilde{G}_{j,x}(r)\overleftarrow{dB}_r^j\right|^4\right]\right)^{1/2}\\
&\quad\left(\mathcal{E}\left[|M_T|^4|M_u^{-1}V_u-M_s^{-1}V_s|^4\right]\right)^{1/2}\\
&\le CmC_G^2\sqrt{\mu_{\mathcal{X}}}\left(\mathcal{E}\left[|M_T|^4|M_u^{-1}V_u-M_s^{-1}V_s|^4\right]\right)^{1/2}
\end{aligned}
$$

by arguments similar to those in Case 1. Furthermore,

$$
\begin{aligned}
\mathcal{E}\left[|M_T|^4|M_u^{-1}V_u-M_s^{-1}V_s|^4\right]&\le\left(\mathcal{E}\left[|M_T|^8\right]\mathcal{E}\left[|M_u^{-1}V_u-M_s^{-1}V_s|^8\right]\right)^{1/2}\\
&\le C(\Delta t)^2
\end{aligned}
$$

again by arguments similar to those on Case 1. As a consequence,

$$\mathcal{E}\left[|I_3(s,u)|^2\right] \leq Cm\Delta t C_G^2 \sqrt{\mu_{\mathcal{X}}}.$$

As in Case 1, since $\mathcal{U}$ is a partition without cluster it follows that

$$\sum_{i=1}^{n} \sup_{t_{i-1}\leq u\leq t_i} \mathcal{E}\left[|I_3(t_{i-1},u)|^2\right] \leq CmC_G^2\sqrt{\mu_{\mathcal{X}}}.$$

<u>Case 3 (I_4)</u>: Let us denote by

$$\begin{aligned}
\Gamma_s &:= \sup_{t\leq r\leq s}\left\{1+\left|\nabla X_r^{t,x}\right|+\left|(\nabla X_r^{t,x})^{-1}\right|+\left|\Lambda_r^{-1}\right|+\left|M_r^{-1}\right|\right\},\\
\overline{\Gamma}_s &:= \sup_{t\leq r\leq s}\left\{1+\left|X_r^{t,x}\right|\right\}.
\end{aligned}$$

Recalling that $\xi_u^{t,x}-\xi_s^{t,x}=\int_s^u \eta_r dW_r$,

$$\begin{aligned}
\sup_{t_{i-1}\leq u\leq t_i} E\left[|I_4(t_{i-1},u)|^2\right] &= \sup_{t_{i-1}\leq u\leq t_i} E\left[|M_{t_{i-1}}^{-1}V_{t_{i-1}}|^2|\xi_u^{t,x}-\xi_{t_{i-1}}^{t,x}|^2\right]\\
&\leq C \sup_{t_{i-1}\leq u\leq t_i} E\left[\Gamma_{t_{i-1}}^6\overline{\Gamma}_{t_{i-1}}^2 E\left[|\xi_u^{t,x}|^2-|\xi_{t_{i-1}}^{t,x}|^2\middle|\,\mathcal{G}_{t_{i-1}}\right]\right]\\
&= C \sup_{t_{i-1}\leq u\leq t_i} E\left[\Gamma_{t_{i-1}}^6\overline{\Gamma}_{t_{i-1}}^2 E\left[\int_{t_{i-1}}^{u}|\eta_r|^2dr\middle|\,\mathcal{G}_{t_{i-1}}\right]\right]\\
&\leq CE\left[\Gamma_{t_{i-1}}^6\overline{\Gamma}_{t_{i-1}}^2 E\left[\int_{t_{i-1}}^{t_i}|\eta_r|^2dr\middle|\,\mathcal{G}_{t_{i-1}}\right]\right]\\
&= CE\left[\Gamma_{t_{i-1}}^6\overline{\Gamma}_{t_{i-1}}^2\left(|\xi_{t_i}^{t,x}|^2-|\xi_{t_{i-1}}^{t,x}|^2\right)\right].
\end{aligned}$$

Since Γ and $\overline{\Gamma}$ are positive and non-decreasing, it follows that

$$\begin{aligned}
&\sum_{i=1}^{n} E\left[\Gamma_{t_{i-1}}^6\overline{\Gamma}_{t_{i-1}}^2\left(|\xi_{t_i}^{t,x}|^2-|\xi_{t_{i-1}}^{t,x}|^2\right)\right]\\
&= E\left[\Gamma_{t_{n-1}}^6\overline{\Gamma}_{t_{n-1}}^2|\xi_{t_n}^{t,x}|^2-\Gamma_{t_0}^6\overline{\Gamma}_{t_0}^2|\xi_{t_0}^{t,x}|^2-\sum_{i=1}^{n-1}|\xi_{t_i}^{t,x}|^2\left(\Gamma_{t_i}^6\overline{\Gamma}_{t_i}^2-\Gamma_{t_{i-1}}^6\overline{\Gamma}_{t_{i-1}}^2\right)\right]\\
&\leq E\left[\Gamma_T^6\overline{\Gamma}_T^2|\xi_T^{t,x}|^2\right].
\end{aligned}$$

As a consequence,

$$\begin{aligned}\sum_{i=1}^{n} \sup_{t_{i-1}\le u\le t_i} \mathcal{E}\left[|I_4(t_{i-1},u)|^2\right] &\le C\mathcal{E}\left[\Gamma_T^6\overline{\Gamma}_T^2\left|\xi_T^{t,x}\right|^2\right] \\ &\le C\sqrt{\mathcal{E}\left[\Gamma_T^{12}\overline{\Gamma}_T^4\right]\mathcal{E}\left[|\xi_T^{t,x}|^4\right]}.\end{aligned}$$

By equation (6.6) and Result A.1 we know that the first expectation is bounded. For the second, we have that

$$\mathcal{E}\left[|\xi_T^{t,x}|^4\right] \le \mathcal{E}\left[|M_T\xi^{t,x}|^4\right] \le \sqrt{\mathcal{E}\left[|M_T|^8\right]\mathcal{E}\left[|\xi^{t,x}|^8\right]}.$$

By equation (6.6) and the arguments found in Case 1 it follows that

$$\sum_{i=1}^{n} \sup_{t_{i-1}\le u\le t_i} \mathcal{E}\left[|I_4(t_{i-1},u)|^2\right] \le CmC_G^2\sqrt{\mu_{\mathcal{X}}}.$$

For $r \in [t_{i-1}, t_i]$,

$$|\widetilde{Z}_r^{t,x} - \widetilde{Z}_{t_i}^{t,x}|^2 \le 2|\widetilde{Z}_r^{t,x} - \widetilde{Z}_{t_{i-1}}^{t,x}|^2 + 2|\widetilde{Z}_{t_i}^{t,x} - \widetilde{Z}_{t_{i-1}}^{t,x}|^2$$

and so

$$\begin{aligned}&\sum_{i=1}^{n} E\left[\int_{t_{i-1}}^{t_i}\int_{\mathbb{R}^d}\left(|\widetilde{Z}_r^{t,x} - \widetilde{Z}_{t_{i-1}}^{t,x}|^2 + |\widetilde{Z}_r^{t,x} - \widetilde{Z}_{t_i}^{t,x}|^2\right)\rho(x)dxdr\right] \\ &\le \sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} \mathcal{E}\left[\left(3|\widetilde{Z}_r^{t,x} - \widetilde{Z}_{t_{i-1}}^{t,x}|^2 + 2|\widetilde{Z}_{t_i}^{t,x} - \widetilde{Z}_{t_{i-1}}^{t,x}|^2\right)\right] dr \\ &\le 5\Delta t \sum_{i=1}^{n} \sup_{t_{i-1}\le u\le t_i} \mathcal{E}\left[|\widetilde{Z}_u^{t,x} - \widetilde{Z}_{t_{i-1}}^{t,x}|^2\right] \\ &\le C\Delta t \sum_{i=1}^{n}\sum_{j=1}^{4} \sup_{t_{i-1}\le u\le t_i} \mathcal{E}\left[|I_j(t_{i-1},u)|^2\right] \\ &\le Cm\Delta t C_G^2\sqrt{\mu_{\mathcal{X}}}\end{aligned}$$

as required. □

6.4. Conclusion

The main results of this chapter are:

- Theorem 6.7: the regularity of $\check{Y}$ for BDSDEs with Lipschitz coefficients.
- Theorem 6.9: the regularity of $\widetilde{Z}$ for BDSDEs with smooth coefficients.

These results are key to deriving the mean square error bound of our approximation scheme in the next chapter.

7 Discretization Scheme

7.1. Introduction

In this chapter we define our discretization scheme for BDSDEs with Lipschitz coefficients (as defined in Definition 5.12 on page 66) in Section 7.3 and derive a bound for the mean square error of the scheme in Section 7.4. To equip us for this, we first finalise our partition of $[t, T]$ and derive some technical lemmas in Section 7.2.

In this chapter we will assume as fixed Lipschitz coefficients. As a consequence, the values of partition constant C_E (as defined in Definition 5.1 on page 47), truncation constant C_M (as defined in Definition 5.2 on page 47) and maximum slope C_G (as defined in Definition 5.12 on page 66) are all fixed. The majority of results in this chapter are parameterised by a value Δt. As we will see, the only restriction with any connection to previous chapters is that $\Delta t \leq \frac{C_M^2}{C_G^2} \wedge C_E$. We are free to choose any value of Δt that satisfies this inequality.

The main results of this chapter make use of the results of Chapter 6. The results of Chapter 6 are also parameterised by by a value Δt. The only restriction upon Δt in Chapter 6 is that $\Delta t \in (0, C_E]$.

As a consequence, within this chapter, we are free to select any value of Δt such that $\Delta t \leq \frac{C_M^2}{C_G^2} \wedge C_E$ and plug this value into the results of Chapter 6. We stress that there is no restriction to our choice of Δt from any previous chapters other than the aforementioned inequality.

7.2. Preliminary Results

Let us denote throughout this chapter by $(\check{Y}, \check{Z})$ the solution of the BDSDE (4.1) with Lipschitz coefficients $\{\check{f}, \check{g}, \check{h}\}$ as defined in Definition 5.12 on page 66 and recall the filtration $\mathcal{G}_s := \mathcal{F}_s^W \vee \mathcal{F}_T^B$. We will make use of the following shorthand: $E_i[.] := E[.|\mathcal{G}_{t_i}]$.

The argument that follows will depend upon measuring the irregularity of the coefficients $\{\check{f}, \check{g}, \check{h}\}$ in some sense. To this end, let us recall Definition 5.11 on page 65 and further define:

Definition 7.1. *Let $\mathcal{X} \in \{\mathcal{X}^{\check{f}}, \mathcal{X}^{\check{g}}, \mathcal{X}^{\check{h}}\}$. Then we define*

$$\mathfrak{G}_{\mathcal{X}} := \left\{ x \in \mathbb{R}^d \middle| \operatorname{dist}(x, \mathfrak{B}_{\mathcal{X}}) > \frac{C_M}{C_G} \right\}$$
$$\mathfrak{G}_{\mathcal{X},k} := \left\{ x \in \mathfrak{G}_{\mathcal{X}} \middle| \operatorname{dist}(x, \mathfrak{B}_{\mathcal{X}}) \in \left(\frac{kC_M}{C_G}, \frac{(k+1)C_M}{C_G} \right] \right\}$$

for $k \geq 1$.

We will now make a final refinement of our partition $\mathcal{T}^2$ as defined in Definition 6.5 on page 81. As already discussed, the constants C_E, C_M and C_G are now fixed. Without loss of generality, let us assume that $C_E \leq 1$. We are now free to select in what follows any value of Δt such that $\Delta t \leq \frac{C_M^2}{C_G^2} \wedge C_E$. We tacitly assume from now that this is the case.

Definition 7.2. *We denote by $\mathcal{T}^3$ any refinement of $\mathcal{T}^2$ that is a partition without cluster of granularity Δt such that for each i, either $\mathcal{T}_i^3 \subset \mathfrak{B}_{\mathcal{T}}$ or $\mathcal{T}_i^3 \cap \mathfrak{B}_{\mathcal{T}} = \emptyset$.*

Remark. In words, Definition 7.2 means that each element of $\mathcal{T}^3$ must either lie entirely within $\frac{C_M}{C_G}$ of a boundary point of the partition $\mathcal{T}$ or lie entirely at least $\frac{C_M}{C_G}$ from every boundary point of the partition $\mathcal{T}$.

To refine $\mathcal{T}^2$ so that for each i either $\mathcal{T}_i^3 \subset \mathfrak{B}_{\mathcal{T}}$ or $\mathcal{T}_i^3 \cap \mathfrak{B}_{\mathcal{T}} = \emptyset$ it is sufficient to add twice as many partition elements as there were in the partition $\mathcal{T}$. To see this note that for each k, each element $\mathcal{T}_k \in \mathcal{T}$ has just two boundary points; as a consequence, there can be at most two sub-elements of $\mathcal{T}_k$, $\mathcal{T}_{k_1}^2$, $\mathcal{T}_{k_2}^2 \in \mathcal{T}^2$, such that $\mathcal{T}_{k_i}^2 \not\subset \mathfrak{B}_{\mathcal{T}}$ and $\mathcal{T}_{k_i}^2 \cap \mathfrak{B}_{\mathcal{T}} \neq \emptyset$. It follows that, $\mathcal{T}^3$ is also a partition without cluster of granularity Δt.

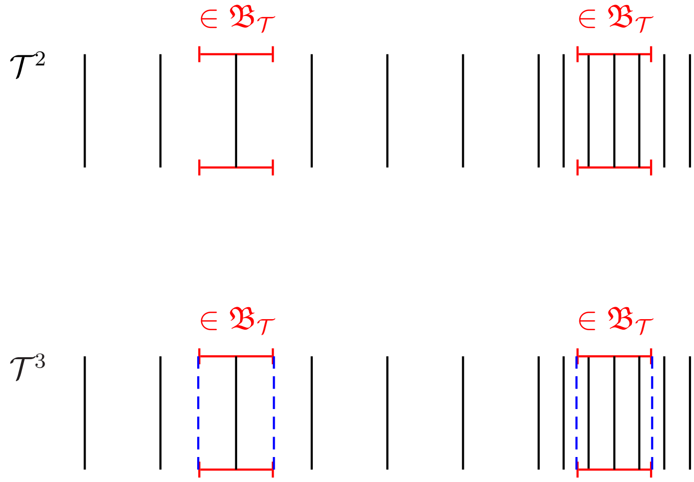

Figure 7.1.: Construction of $\mathcal{T}^3$ from $\mathcal{T}^2$: each interval of $\mathcal{T}^3$ lies either entirely within $\mathfrak{B}_{\mathcal{T}}$ entirely without.

Definition 7.3. *Let the partition $\mathcal{T}^3$ be as defined in Definition 7.2 and denote by $\mathcal{J}$ those i such that $\mathcal{T}_i^3 \subset \mathfrak{B}_{\mathcal{T}}$. Then we define $\mathcal{T}^*$ to be the refinement of $\mathcal{T}^3$ constructed by dividing for each $i \in \mathcal{J}$ the interval $\mathcal{T}_i^3$ into*

$$\left\lceil \frac{\Delta t_i}{\sqrt{\Delta t l(\mathfrak{B}_{\mathcal{T}})}} \right\rceil$$

equal subintervals.

Lemma 7.4. *The partition $\mathcal{T}^*$ as defined in Definition 7.3 is a partition without cluster of granularity Δt such that for any partition element $\mathcal{T}_i^* := [t_i, t_{i+1})$ of $\mathcal{T}^*$ and for any $s \in \mathcal{T}_i^*$ we have that for $\phi = \check{f}, \check{g}$*

$$I_i := \mathcal{E}\left[|\phi(s, \check{\theta}_s^{t,x}) - \phi(t_i, \check{\theta}_s^{t,x})|^2\right] \leq \Delta t C_G^2 l(\mathfrak{B}_{\mathcal{T}}). \tag{7.1}$$

Proof. When $\mathcal{T}_i^3 \cap \mathfrak{B}_{\mathcal{T}} = \emptyset$ then $I_i = 0$ and when $\mathcal{T}_i^3 \subset \mathfrak{B}_{\mathcal{T}}$, $I_i \leq (C_G \Delta t_i)^2$ where $\Delta t_i := t_i - t_{i-1}$. Let us again denote by $\mathcal{J}$ those i such that $\mathcal{T}_i^3 \subset \mathfrak{B}_{\mathcal{T}}$. As a consequence, it is sufficient to ensure that $\Delta t_i \leq \sqrt{\Delta t l(\mathfrak{B}_{\mathcal{T}})}$ whenever $i \in \mathcal{J}$. It is thus sufficient to divide each such $\mathcal{T}_i^3 \subset \mathfrak{B}_{\mathcal{T}}$ into

$$\left\lceil \frac{\Delta t_i}{\sqrt{\Delta t l(\mathfrak{B}_{\mathcal{T}})}} \right\rceil$$

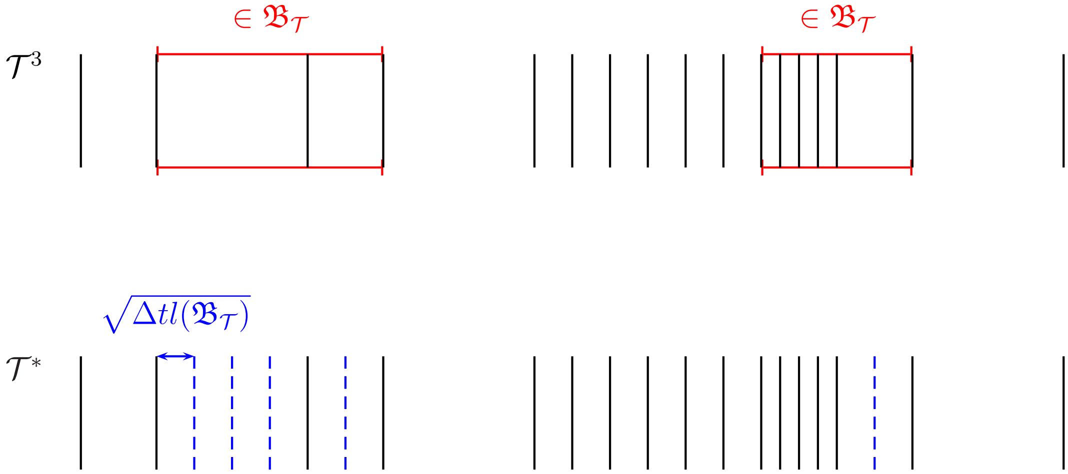

Figure 7.2.: Construction of $\mathcal{T}^*$ from $\mathcal{T}^3$.

subintervals. This means that to ensure that $I_i \leq \Delta t C_G^2 l(\mathfrak{B}_{\mathcal{T}})$ upon $\mathcal{T}_i^*$, it is sufficient to add

$$\left\lfloor \frac{\Delta t_i}{\sqrt{\Delta t l(\mathfrak{B}_{\mathcal{T}})}} \right\rfloor$$

subintervals to the partition.

Now

$$\sum_{i \in \mathcal{J}} \Delta t_i = l(\mathfrak{B}_{\mathcal{T}})$$

and so the total number of additional intervals required, N^*, satisfies

$$N^* \leq \sum_{i \in \mathcal{J}} \frac{\Delta t_i}{\sqrt{\Delta t l(\mathfrak{B}_{\mathcal{T}})}} = \sqrt{\frac{l(\mathfrak{B}_{\mathcal{T}})}{\Delta t}} \leq \frac{\sqrt{T-t}}{\Delta t}.$$

As a consequence, since $\mathcal{T}^3$ is a partition without cluster of granularity Δt, so is $\mathcal{T}^*$. $\square$

Lemma 7.5. *Let $\mathcal{X} \in \{\mathcal{X}^f, \mathcal{X}^g, \mathcal{X}^h\}$. Then for each $j \geq 1$*

$$\mu_x(\mathfrak{G}_{\mathcal{X},j}) \leq (1+e^\nu)\mu_x(\mathfrak{B}_{\mathcal{X}}) \quad \textit{and} \quad \mu_x(\mathfrak{G}_{\mathcal{X}}^c) \leq (2+e^\nu)\mu_x(\mathfrak{B}_{\mathcal{X}}).$$

Proof. For any subset $A \subset [-C_M, C_M]^d$,

$$\mu_x(A) := \int_{\mathbb{R}^d} \mathbb{I}_A(x)\rho(x)dx = \sum_{k\geq 0}\int_{\mathcal{X}_k} \mathbb{I}_A(x)\rho(x)dx.$$

Without loss of generality, let us assume that $0 \in \mathcal{X}_0$ and that $\mathcal{X}_0$ has maximum edge length $\leq 2C_E$. If this is not the case, it is easy to see that we can refine $\mathcal{X}$ so that this is the case without violating the restriction that each edge of $\mathcal{X}$ must have length at least C_E.

We first show that for any j,

$$\sum_{k\geq 1}\int_{\mathcal{X}_k} \mathbb{I}_{\mathfrak{G}_{\mathcal{X},j}}(x)\rho(x)dx \leq \mu_x(\mathfrak{B}_{\mathcal{X}}).$$

Let $k \geq 1$, and suppose that the partition element $\mathcal{X}_k$ is the d-dimensional box with dimensions $\delta_1 \times \delta_2 \times \ldots \times \delta_d$.

The boundary of $\mathcal{X}_k$ consists of $2d$, $(d-1)$-dimensional surfaces each of which comes as a symmetric pair with each member of a pair having the same $(d-1)$-dimensional surface area. For example, when $d = 1$ the boundary consists of two points, when $d = 2$ the boundary consists of the outline of a rectangle and when $d = 3$ the boundary consists of the surface of a rectangular cuboid.

Since $k \neq 0$ we know that $0 \notin \mathcal{X}_k$ and so for each pair we can identify the unique member that is closest to 0. Let ψ_i, $i = 1, \ldots d$, be the member of each of the surface pairs that is closest to the origin and let us denote by

$$B_{2i} := \left\{ x \in \mathcal{X}_k \,\middle|\, \operatorname{dist}(x, \psi_i) \leq \frac{C_M}{C_G} \right\} \quad \text{and} \quad B_{2i-1} := \left\{ x \in \mathcal{X}_{k_i} \,\middle|\, \operatorname{dist}(x, \psi_i) \leq \frac{C_M}{C_G} \right\}$$

where $\mathcal{X}_{k_i}$ denotes the unique partition element that shares ψ_i as a boundary with $\mathcal{X}_k$.

We now denote by $G_{k,j} := \mathfrak{G}_{\mathcal{X},j} \cap \mathcal{X}_k$, the points of $\mathcal{X}_k$ that are between $\frac{jC_M}{C_G}$ and $\frac{(j+1)C_M}{C_G}$ from the nearest point of the boundary of $\mathcal{X}_k$. $G_{k,j}$ can be separated into d pairs of boxes:

$$(G_1, G_2), (G_3, G_4), \ldots, (G_{2d-1}, G_{2d}).$$

where each of these pairs runs parallel to a distinct boundary surface, ψ_i. Without loss of generality, let G_{2i-1} and G_{2i} run parallel to ψ_i for each i. If we now associate G_j with B_j for $j = 1, \ldots, 2d$ it is clear that $l(B_j) > l(G_j)$ where l denotes Lebesgue measure, and $\text{dist}(0, B_j) < \text{dist}(0, G_j)$ and so $\mu(B_j) > \mu(G_j)$ (see Figure 7.3 for an illustration of this association when $d = 2$). It now follows that since the B_j are distinct, disjoint and each is a subset of $\mathfrak{B}_{\mathcal{X}}$,

$$\sum_{k\geq 1} \int_{\mathcal{X}_k} \mathbb{I}_{\mathfrak{G}_{\mathcal{X},j}}(x)\rho(x)dx \leq \sum_{k\geq 0} \int_{\mathcal{X}_k} \mathbb{I}_{\mathfrak{B}_{\mathcal{X}}}(x)\rho(x)dx = \mu_x(\mathfrak{B}_{\mathcal{X}}).$$

For $k = 0$,

$$\begin{aligned}\int_{\mathcal{X}_0} \mathbb{I}_{\mathfrak{G}_{\mathcal{X},j}}(x)\rho(x)dx &= \int_{\mathcal{X}_0} \mathbb{I}_{\mathfrak{G}_{\mathcal{X},j}}(x)e^{-\nu|x|}dx \\ &\leq \int_{\mathcal{X}_0} \mathbb{I}_{\mathfrak{B}_{\mathcal{X}}}(x)e^{-\nu(|x|-C_E)}dx \\ &\leq e^{\nu C_E}\mu_x(\mathfrak{B}_{\mathcal{X}})\end{aligned}$$

and the first result follows.

The second result follows by noting that $\mathfrak{G}_{\mathcal{X}}^c = \mathfrak{B}_{\mathcal{X}} \cup (\mathfrak{G}_{\mathcal{X}}^c \setminus \mathfrak{B}_{\mathcal{X}})$. Following the same argument as for $\mathfrak{G}_{\mathcal{X},j}$ we have that $\mu_x(\mathfrak{G}_{\mathcal{X}}^c \setminus \mathfrak{B}_{\mathcal{X}}) \leq (1 + e^{\nu})\mu_x(\mathfrak{B}_{\mathcal{X}})$. $\square$

For the sake of brevity we now define a parameterised constant that will be used repeatedly.

Definition 7.6. *Given a constant $K_{\mathfrak{B}} > 0$, we define for $\mathcal{X} = \mathcal{X}^f \cup \mathcal{X}^g \cup \mathcal{X}^h$*

$$\Upsilon(K_{\mathfrak{B}}) := \left(l(\mathfrak{B}_{\mathcal{T}}) + \mu_x(\mathfrak{B}_{\mathcal{X}})^{\frac{1}{2+K_{\mathfrak{B}}}}\right) \vee \sqrt{\mu_{\mathcal{X}}}.$$

Lemma 7.7. *For each $K_{\mathfrak{B}} > 0$, there exists a constant $C > 0$ independent of C_E, C_M, C_G and Δt such that*

$$\mathcal{E}\left[|\check{h}(X_T^{t,x}) - \check{h}(\widehat{X}_T^{t,x})|^2\right] \leq C\Delta t\Upsilon(K_{\mathfrak{B}})C_G^2$$

and for $\phi = \check{f}, \check{g}$, $i = 1, \ldots n$ and $s \in [t_{i-1}, t_i]$

$$\mathcal{E}\left[|\phi(s, \check{\theta}_s^{t,x}) - \phi(t_i, \widehat{X}_{t_i}^{t,x}, \check{Y}_s^{t,x}, \check{Z}_s^{t,x})|^2\right] \leq C\Delta t\Upsilon(K_{\mathfrak{B}})C_G^2.$$

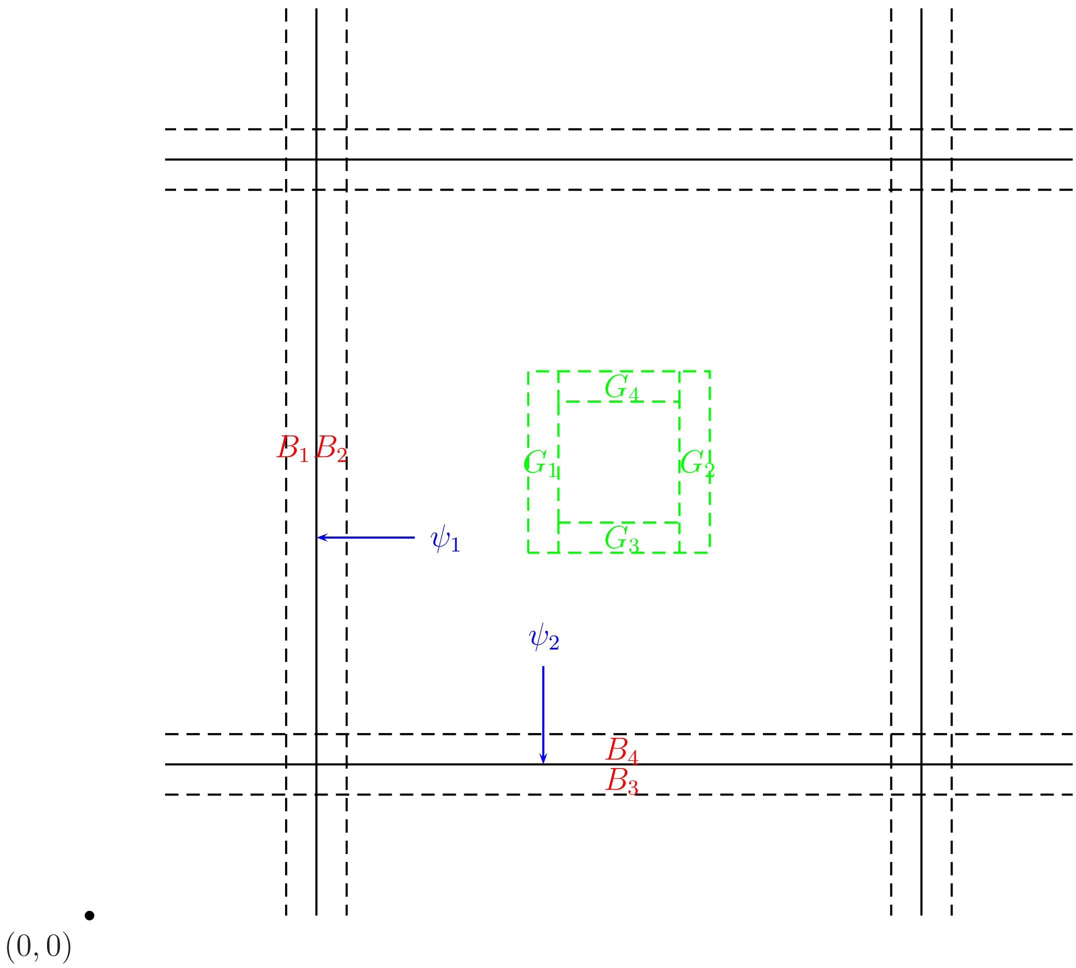

Figure 7.3.: Finding an upper bound for $\mu_x(G_{\mathcal{X},j})$: ψ_i denotes the member of each of the surface pairs that is closest to the origin; B_i denotes a member of $\mathfrak{B}_{\mathcal{X}}$; G_i denotes an element of $\mathfrak{G}_{\mathcal{X}}$ associated with B_i.

Proof. We prove the result for ϕ as $\check{h}$ is simpler. To this end, we note that

$$\begin{aligned}|\phi(s,\theta_s^{t,x}) - \phi(t_i,\widehat{X}_{t_i}^{t,x},\check{Y}_s^{t,x},\check{Z}_s^{t,x})|^2 \leq\ & 3|\phi(s,\check{\theta}_s^{t,x}) - \phi(t_i,\check{\theta}_s^{t,x})|^2 \\ &+3|\phi(t_i,\check{\theta}_s^{t,x}) - \phi(t_i,X_{t_i}^{t,x},\check{Y}_s^{t,x},\check{Z}_s^{t,x})|^2 \\ &+3|\phi(t_i,X_{t_i}^{t,x},\check{Y}_s^{t,x},\check{Z}_s^{t,x}) - \phi(t_i,\widehat{X}_{t_i}^{t,x},\check{Y}_s^{t,x},\check{Z}_s^{t,x})|^2.\end{aligned}$$

As a consequence, we have the following cases to consider:

1. $|\phi(s,\check{\theta}_s^{t,x}) - \phi(t_i,\check{\theta}_s^{t,x})|^2$.
2. $|\phi(t_i,\check{\theta}_s^{t,x}) - \phi(t_i,X_{t_i}^{t,x},\check{Y}_s^{t,x},\check{Z}_s^{t,x})|^2$.
3. $|\phi(t_i,X_{t_i}^{t,x},\check{Y}_s^{t,x},\check{Z}_s^{t,x}) - \phi(t_i,\widehat{X}_{t_i}^{t,x},\check{Y}_s^{t,x},\check{Z}_s^{t,x})|^2$.

We note that Case 3 is analagous to the result for h so we will not prove the result for h separately.

<u>Case 1</u>: This follows by Lemma 7.4.

<u>Case 3</u>: For $\mathcal{X} := \mathcal{X}^{\phi}$,

$$\begin{aligned}&\mathcal{E}\left[|\phi(t_i,X_{t_i}^{t,x},\check{Y}_s^{t,x},\check{Z}_s^{t,x}) - \phi(t_i,\widehat{X}_{t_i}^{t,x},\check{Y}_s^{t,x},\check{Z}_s^{t,x})|^2\right] \\ &\leq C_G^2\mathcal{E}\left[|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^2\left(\mathbb{I}_{\mathfrak{G}_{\mathcal{X}}^c}(X_{t_i}^{t,x}) + \sum_k \mathbb{I}_{\mathfrak{G}_{\mathcal{X},k}}(X_{t_i}^{t,x})\mathbb{I}_{\left(\frac{kC_M}{C_G},\infty\right)}(|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|)\right)\right]\end{aligned}$$

where we have argued that we have an upper bound if for all instances where either $X_{t_i}^{t,x}$ does not lie in $\mathfrak{G}_{\mathcal{X}}$ or $X_{t_i}^{t,x}$ does lie in $\mathfrak{G}_{\mathcal{X}}$ but $\widehat{X}_{t_i}^{t,x}$ is far enough away from $X_{t_i}^{t,x}$ to have possibly either entered into or crossed over $\mathfrak{B}_{\mathcal{X}}$ we assign a Lipschitz constant of C_G.

By Hölder's inequality, equivalence of norms and Lemmas 7.10 and 7.5 we have that for conjugate indices p and q

$$\begin{aligned}&\mathcal{E}\left[|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^2\mathbb{I}_{\mathfrak{G}_{\mathcal{X}}^c}(X_{t_i}^{t,x})\right] \\ &\leq \left(\mathcal{E}\left[|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^{2p}\right]\right)^{1/p}\left(\mathcal{E}\left[\mathbb{I}_{\mathfrak{G}_{\mathcal{X}}^c}(X_{t_i}^{t,x})\right]\right)^{1/q} \\ &\leq C\left((\Delta t_i)^p\int_{\mathbb{R}^d}(1+|x|^{2p})\rho(x)dx\right)^{1/p}\left(\int_{\mathbb{R}^d}\mathbb{I}_{\mathfrak{G}_{\mathcal{X}}^c}(x)\rho(x)dx\right)^{1/q}\end{aligned}$$

$$
\begin{aligned}
&\leq C\Delta t \mu_x(\mathfrak{G}_{\mathcal{X}}^c)^{1/q} \\
&= C\Delta t \mu_x(\mathfrak{G}_{\mathcal{X}}^c)^{1/(2+K_{\mathfrak{B}})} \\
&\leq C\Delta t \mu_x(\mathfrak{B}_{\mathcal{X}})^{1/(2+K_{\mathfrak{B}})}
\end{aligned}
$$

choosing $q = 2 + K_{\mathfrak{B}}$.

For the second term, we similarly have that for conjugate indices a and b

$$
\begin{aligned}
&\mathcal{E}\left[|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^2 \mathbb{I}_{\mathfrak{G}_{\mathcal{X},k}}(X_{t_i}^{t,x}) \mathbb{I}_{\left(\frac{kC_M}{C_G},\infty\right)}(|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|)\right] \\
&\leq \left(\mathcal{E}\left[|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^{2a}\right]\right)^{1/a} \left(\mathcal{E}\left[\mathbb{I}_{\mathfrak{G}_{\mathcal{X},k}}(X_{t_i}^{t,x}) \mathbb{I}_{\left(\frac{kC_M}{C_G},\infty\right)}(|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|)\right]\right)^{1/b} \\
&\leq C\Delta t \left(\mathcal{E}\left[\mathbb{I}_{\mathfrak{G}_{\mathcal{X},k}}(X_{t_i}^{t,x}) \mathbb{I}_{\left(\frac{kC_M}{C_G},\infty\right)}(|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|)\right]\right)^{1/b}.
\end{aligned}
$$

By Fubini's theorem, equivalence of norms, Chebyshev's inequality and Lemma 7.5 we have that for conjugate indices p and q,

$$
\begin{aligned}
&\mathcal{E}\left[\mathbb{I}_{\mathfrak{G}_{\mathcal{X},k}}(X_{t_i}^{t,x}) \mathbb{I}_{\left(\frac{kC_M}{C_G},\infty\right)}(|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|)\right] \\
&\leq \left(\mathcal{E}\left[\mathbb{I}_{\mathfrak{G}_{\mathcal{X},k}}(X_{t_i}^{t,x})\right]\right)^{1/p} \left(\mathcal{E}\left[\mathbb{I}_{\left(\frac{kC_M}{C_G},\infty\right)}(|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|)\right]\right)^{1/q} \\
&\leq C\left(\int_{\mathbb{R}^d} \mathbb{I}_{\mathfrak{G}_{\mathcal{X},k}}(x)\rho(x)dx\right)^{1/p} \left(\int_{\mathbb{R}^d} P\left(|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}| > \frac{kC_M}{C_G}\right)\rho(x)dx\right)^{1/q} \\
&\leq C(\mu_x(\mathfrak{G}_{\mathcal{X},k}))^{1/p} \left(\frac{C_G^2}{k^2 C_M^2}\mathcal{E}\left[|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^2\right]\right)^{1/q} \\
&\leq C(\mu_x(\mathfrak{B}_{\mathcal{X}}))^{1/p} \left(\frac{C_G^2 \Delta t}{k^2 C_M^2}\right)^{1/q} \\
&\leq C(\mu_x(\mathfrak{B}_{\mathcal{X}}))^{1/p} \left(\frac{1}{k^2}\right)^{1/q}
\end{aligned}
$$

since we have assumed $\Delta t \leq \frac{C_M^2}{C_G^2}$. Consequently,

$$
\begin{aligned}
&\sum_{k\geq 1} \mathcal{E}\left[|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^2 \mathbb{I}_{\mathfrak{G}_{\mathcal{X},k}}(X_{t_i}^{t,x}) \mathbb{I}_{\left(\frac{kC_M}{C_G},\infty\right)}(|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|)\right] \\
&\leq C\Delta t (\mu_x(\mathfrak{B}_{\mathcal{X}}))^{\frac{1}{bp}} \sum_{k\geq 1} k^{\frac{-2}{bq}}
\end{aligned}
$$

$$\leq C\Delta t(\mu_x(\mathfrak{B}_{\mathcal{X}}))^{\frac{1}{(2+K_{\mathfrak{B}})}}$$

choosing b and p so that $bp = 2 + K_{\mathfrak{B}}$ and $bq < 2$.

Case 2: Although we could get a stronger result for this case using the Markovian property of X and a slightly more complex argument, the desired inequality follows using essentially the same argument as in Case 3. □

We now derive a specialized form of the discrete Gronwall inequality. For the standard discrete Gronwall inequality see, for example, [49], p. 479.

Lemma 7.8. *Suppose that for $j = 0, \ldots, n$ $a_j, b_j, c_j \geq 0$ and there exist positive constants C and r such that either*

$$a_j + b_j \leq (1 + C\Delta t)(a_{j+1} + b_{j+1}) + Cc_{j+1} + C(\Delta t)^{1+r} \tag{7.2}$$

or

$$a_j + b_j \leq (1 + C\Delta t)(a_{j+1} + b_{j+1}) + \frac{1}{2}\sum_{l=j+2}^{j+k} b_l + C\sum_{l=j+1}^{j+k} c_l + C(\Delta t)^{1+r}, \tag{7.3}$$

In the latter case, we assume also that for $i = j+1, \ldots, j+k-1$,

$$a_i + b_i \leq (1 + C\Delta t)\left(a_{i+1} + \frac{1}{2}b_{i+1}\right) + Cc_{i+1} + C(\Delta t)^{1+r}.$$

Then there exists a constant C' independent of Δt such that

$$\max_{0 \leq i \leq n}(a_i + b_i) \leq C'\left(a_n + b_n + \sum_{l=1}^{n} c_l + (\Delta t)^r\right).$$

Proof. If (7.3) holds, we have that

$$\begin{aligned} a_j + b_j &\leq (1 + C\Delta t)^2 (a_{j+2} + b_{j+2}) + \frac{1}{2}\sum_{l=j+3}^{j+k} b_l + C\sum_{l=j+1}^{j+k} c_l + (1 + C\Delta t)Cc_{j+2} \\ &\quad + C(\Delta t)^{1+r}(1 + (1 + C\Delta t)) \\ &\leq (1 + C\Delta t)^k (a_{j+k} + b_{j+k}) + C\sum_{l=j+1}^{j+k} c_l(1 + (1 + C\Delta t)^{l-(j+1)}) \end{aligned}$$

$$
\begin{aligned}
&+C\,(\Delta t)^{1+r}\sum_{i=0}^{k-1}(1+C\Delta t)^{i}\\
&\leq (1+C\Delta t)^{k}\left(a_{j+k}+b_{j+k}+C\sum_{l=j+1}^{j+k}c_l+kC\,(\Delta t)^{1+r}\right).
\end{aligned}
$$

Now, if for positive integers k_1 and k_2,

$$
a_j+b_j\leq (1+C\Delta t)^{k_1}\left(a_{j+k_1}+b_{j+k_1}+C\sum_{l=j+1}^{j+k_1}c_l+k_1C\,(\Delta t)^{1+r}\right)
$$

and

$$
a_{j-k_2}+b_{j-k_2}\leq (1+C\Delta t)^{k_2}\left(a_j+b_j+C\sum_{l=j-k_2+1}^{j}c_l+k_2C\,(\Delta t)^{1+r}\right)
$$

then

$$
\begin{aligned}
&a_{j-k_2}+b_{j-k_2}\\
&\leq (1+C\Delta t)^{k_1+k_2}\left(a_{j+k_1}+b_{j+k_1}+C\sum_{l=j+1}^{j+k_1}c_l+k_1C\,(\Delta t)^{1+r}\right)\\
&\quad+(1+C\Delta t)^{k_2}\,C\left(\sum_{l=j-k_2+1}^{j}c_l+k_2\,(\Delta t)^{1+r}\right)\\
&\leq (1+C\Delta t)^{k_1+k_2}\left(a_{j+k_1}+b_{j+k_1}+C\sum_{l=j-k_2+1}^{j+k_1}c_l+(k_1+k_2)C\,(\Delta t)^{1+r}\right).
\end{aligned}
$$

As a consequence, if we decompose $n=k_1+k_2+\ldots+k_m$ where $k_i=1$ indicates that (7.2) holds, then recalling that κ denotes the order of the partition, we have that for any $j=n-k_1-k_2-\ldots-k_l$,

$$
\begin{aligned}
a_j+b_j&\leq (1+C\Delta t)^{n-j}\left(a_n+b_n+C\sum_{l=j+1}^{n}c_l+(n-j)C\,(\Delta t)^{1+r}\right)\\
&\leq (1+C\Delta t)^{n}\left(a_n+b_n+C\sum_{l=1}^{n}c_l+nC\,(\Delta t)^{1+r}\right)
\end{aligned}
$$

$$\le \left(1 + \frac{C\kappa(T-t)}{n}\right)^n \left(a_n + b_n + C\sum_{l=1}^{n} c_l + C\kappa(T-t)\,(\Delta t)^r\right)$$
$$\le e^{C\kappa(T-t)} \left(a_n + b_n + C\sum_{l=1}^{n} c_l + C\kappa(T-t)\,(\Delta t)^r\right).$$

□

Finally, to define our discretization scheme for $(\check{Y}, \check{Z})$ we will require a discretization scheme X. For the sake of simplicity, we select the Euler approximation:

Definition 7.9. *Let X be defined as in Definition 4.1 on page 42. We denote by $\widehat{X}$ the Euler approximation of X (see for example [24]) given by*

$$\widehat{X}^{t,x}_{t_0} := x,$$
$$\widehat{X}^{t,x}_{t_{i+1}} := \widehat{X}^{t,x}_{t_i} + b(\widehat{X}^{t,x}_{t_i})\Delta t_i + \sigma(\widehat{X}^{t,x}_{t_i})\Delta W_{i+1}$$

where $\Delta W_{i+1} := W_{t_{i+1}} - W_{t_i}$.

The following result provides an upper bound on the higher order moments of the error of the Euler scheme. A bound on the mean of the absolute value of the error is derived in [24]. We note that the following result is likely to be already known (although an existing proof was not located) and that the proof is similar to that of the aforementioned result in [24]. We do provide a proof, however, for completeness.

Lemma 7.10. *Let X be defined as in Definition 4.1 on page 42 and $t =: t_0, t_1, \dots t_n := T$ be a partition without cluster of order κ and granularity Δt. For every $p \ge 2$, there is a constant C independent of Δt such that for every $i = 1, \dots, n$*

$$E\left[|X^{t,x}_{t_i} - \widehat{X}^{t,x}_{t_i}|^p\right] \le C(1 + |x|^p)(\Delta t)^{p/2}.$$

Proof. For ease of presentation, we prove the result for dimension $d = 1$. If we define $\widehat{X}$ by $\widehat{X}^{t,x}_t := x$ and for $s \in (t_i, t_{i+1}]$, $i = 0, \dots, n-1$

$$\widehat{X}^{t,x}_s := \widehat{X}^{t,x}_{t_i} + b(\widehat{X}^{t,x}_{t_i})(s - t_i) + \sigma(\widehat{X}^{t,x}_{t_i})(W_s - W_{t_i})$$

then our notation is consistent with the Euler scheme. By Itô's Formula, for $s \in$

$(t_i, t_{i+1}]$

$$
\begin{aligned}
|X_s^{t,x} - \widehat{X}_s^{t,x}|^p = {} & |X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^p \\
& + \int_{t_i}^{s} p|X_r^{t,x} - \widehat{X}_r^{t,x}|^{p-2}(X_r^{t,x} - \widehat{X}_r^{t,x})(b(X_r^{t,x}) - b(\widehat{X}_{t_i}^{t,x}))dr \\
& + \int_{t_i}^{s} p|X_r^{t,x} - \widehat{X}_r^{t,x}|^{p-2}(X_r^{t,x} - \widehat{X}_r^{t,x})(\sigma(X_r^{t,x}) - \sigma(\widehat{X}_{t_i}^{t,x}))dW_r \\
& + \int_{t_i}^{s} \frac{p}{2}(p-2)|X_r^{t,x} - \widehat{X}_r^{t,x}|^{p-2}|\sigma(X_r^{t,x}) - \sigma(\widehat{X}_{t_i}^{t,x})|^2 dr
\end{aligned}
$$

It follows that

$$
\begin{aligned}
E\left[|X_{t_{i+1}}^{t,x} - \widehat{X}_{t_{i+1}}^{t,x}|^p\right] \leq {} & E\left[|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^p\right] \\
& + E\left[\int_{t_i}^{t_{i+1}} p|X_r^{t,x} - \widehat{X}_r^{t,x}|^{p-1}|b(X_r^{t,x}) - b(\widehat{X}_{t_i}^{t,x})|dr\right] \\
& + E\left[\int_{t_i}^{t_{i+1}} \frac{p}{2}(p-2)|X_r^{t,x} - \widehat{X}_r^{t,x}|^{p-2}|\sigma(X_r^{t,x}) - \sigma(\widehat{X}_{t_i}^{t,x})|^2)dr\right]
\end{aligned}
$$

Recalling the definition of L_X from Definition 4.2 on page 42,

$$
\begin{aligned}
& |X_r^{t,x} - \widehat{X}_r^{t,x}|^{p-1}|b(X_r^{t,x}) - b(\widehat{X}_{t_i}^{t,x})| \\
& \quad \leq \sqrt{L_X}|X_r^{t,x} - \widehat{X}_r^{t,x}|^{p-1}\left(|X_r^{t,x} - X_{t_i}^{t,x}| + |X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|\right) \\
& \quad \leq \sqrt{L_X}\left(2|X_r^{t,x} - \widehat{X}_r^{t,x}|^p + |X_r^{t,x} - X_{t_i}^{t,x}|^p + |X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^p\right).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& |X_r^{t,x} - \widehat{X}_r^{t,x}|^{p-2}|\sigma(X_r^{t,x}) - \sigma(\widehat{X}_{t_i}^{t,x})|^2 \\
& \quad \leq 2L_X\left(2|X_r^{t,x} - \widehat{X}_r^{t,x}|^p + |X_r^{t,x} - X_{t_i}^{t,x}|^p + |X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^p\right).
\end{aligned}
$$

As a consequence, for a positive constant C independent of Δt,

$$
\begin{aligned}
& E\left[|X_{t_{i+1}}^{t,x} - \widehat{X}_{t_{i+1}}^{t,x}|^p\right] \\
& \quad \leq E\left[|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^p\right] \\
& \qquad + CE\left[\int_{t_i}^{t_{i+1}} \left\{|X_r^{t,x} - \widehat{X}_r^{t,x}|^p + |X_r^{t,x} - X_{t_i}^{t,x}|^p + |X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^p\right\} dr\right]
\end{aligned}
$$

$$\leq (1 + C\Delta t)E\left[|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^p\right] + C(1 + |x|^p)(\Delta t)^{p/2+1}$$
$$+CE\left[\int_{t_i}^{t_{i+1}} |X_r^{t,x} - \widehat{X}_r^{t,x}|^p dr\right]$$

by Result A.2. By Gronwall's inequality it follows that, again with C a positive constant independent of Δt,

$$E\left[|X_{t_{i+1}}^{t,x} - \widehat{X}_{t_{i+1}}^{t,x}|^p\right] \leq (1 + C\Delta t)E\left[|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^p\right] + C(1 + |x|^p)(\Delta t)^{p/2+1}.$$

If we write this as

$$I_{i+1} \leq (1 + C\Delta t)I_i + \alpha\Delta t$$

it follows that

$$\begin{aligned} I_{i+1} &\leq (1 + C\Delta t)((1 + C\Delta t)I_{i-1} + \alpha\Delta t) + K\Delta t \\ &\leq \alpha\Delta t \sum_{j=0}^{n-1}(1 + C\Delta t)^j \end{aligned}$$

since $I_0 = 0$. Since the partition is a partition without cluster of granularity κ, we have that $\Delta t \leq \frac{\kappa(T-t)}{n}$. As a consequence,

$$\begin{aligned} \sum_{j=0}^{n-1}(1 + C\Delta t)^j &\leq \sum_{j=0}^{n-1}\left(1 + \frac{C\kappa(T-t)}{n}\right)^j \\ &\leq n\left(1 + \frac{C\kappa(T-t)}{n}\right)^n \\ &\leq ne^{C\kappa(T-t)} \\ &\leq \frac{\kappa(T-t)}{\Delta t}e^{C\kappa(T-t)}. \end{aligned}$$

It follows that for $i = 1, \ldots, n$ there is a positive constant C independent of Δt such that

$$I_i \leq C\alpha = C(1 + |x|^p)(\Delta t)^{p/2}$$

as required. □

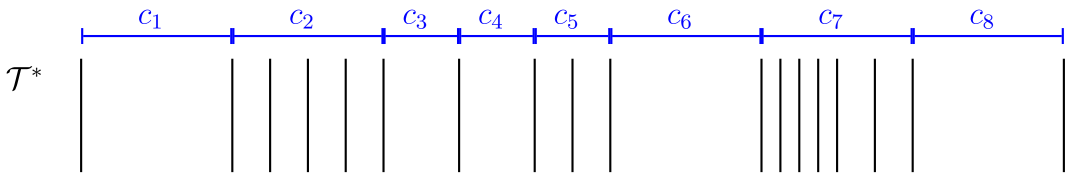

Figure 7.4.: Construction of chains $c_1, c_2, \ldots, c_8$ on $\mathcal{T}^*$.

7.3. Definition of the Scheme

Chains

Before we define our discretization scheme, we need to split the partition $\mathcal{T}^*$ (defined in Definition 7.3 on page 107) into a sequence of chains as described below. To understand why this is neccessary, we note that in [49], the derivation of the error bound relies upon the partition of time being K-uniform. In our notation, a partition $\mathcal{T}$ of $[t, T]$ is K-uniform if there is a $K > 0$ such that $\Delta t_i \geq \frac{\Delta t}{K}$ for all $i = 1, \ldots, n$ (where n denotes the number of intervals in $\mathcal{T}^*$). To cope with the irregular coefficients in this thesis, however, we allow individual elements of $\mathcal{T}^*$ to be arbitrarily small and as a consequence, it is not possible for $\mathcal{T}^*$ to be K-uniform. By using the concept of a chain we can circumvent the requirement of having a K-uniform partition.

Definition 7.11. *To split the partition $\mathcal{T}^* : t = t_0 < t_1 < \ldots < t_n = T$ (defined in Definition 7.3) into a sequence of chains we start by first placing the first partition element $\mathcal{T}_1^* = [t_0, t_1)$ into its own chain and closing the chain. We then start the following procedure with $i = 2$:*

1. *Open a new chain and add the partition element $\mathcal{T}_i^* := [t_{i-1}, t_i)$.*
2. *If $t_{i-1} - t_{i-2} \leq t_i - t_{i-1}$ then close the chain and restart the process with $i = i+1$.*
3. *Otherwise if $t_{i-1} - t_{i-2} > t_i - t_{i-1}$ then continue to add partition elements $\mathcal{T}_{i+1}^*, \mathcal{T}_{i+2}^*, \ldots, \mathcal{T}_{i+k}^*$ until $t_{i-1} - t_{i-2} \leq t_{i+k} - t_{i-1}$ and close the chain. Now restart the process for $i = i + k + 1$.*

Discretization Scheme

Definition 7.12. *The discretization scheme for $(\check{Y}, \check{Z})$ is initialised with*

$$\widehat{Y}^{t,x}_{t_n} := \check{h}(\widehat{X}_{t_n}) \ , \quad \widehat{Z}^{t,x}_{t_n} := 0.$$

If we define

$$\widehat{U}^{t,x}_{t_i} := \widehat{Y}^{t,x}_{t_i} + \check{f}(t_i, \widehat{\theta}^{t,x}_{t_i})\Delta t_i - \sum_{j=1}^{m} \check{g}_j(t_i, \widehat{\theta}^{t,x}_{t_i})\Delta B^j_i$$

where $\widehat{\theta}^{t,x}_{t_i} := (\widehat{X}^{t,x}_{t_i}, \widehat{Y}^{t,x}_{t_i}, \widehat{Z}^{t,x}_{t_i})$ and $\Delta B^j_i := B^j_{t_i} - B^j_{t_{i-1}}$ then

$$\widehat{Y}^{t,x}_{t_{i-1}} := E_{i-1}\left[\widehat{U}^{t,x}_{t_i}\right].$$

If t_{i-1} is not the start point of a chain of length greater than one, then we define

$$\widehat{Z}^{t,x}_{t_{i-1}} := \frac{1}{\Delta t_i} E_{i-1}\left[\widehat{U}^{t,x}_{t_i}\Delta W_i\right].$$

When t_{i-1} is the start point of a chain of length $k > 1$, we define

$$\widehat{V}^{t,x}_{t_{i+k-1}} := \widehat{Y}^{t,x}_{t_{i+k-1}} + \sum_{l=0}^{k-1} \check{f}(t_{i+l}, \widehat{\theta}^{t,x}_{t_{i+l}})\Delta t_{i+l} - \sum_{l=0}^{k-1}\sum_{j=1}^{m} \check{g}_j(t_{i+l}, \widehat{\theta}^{t,x}_{t_{i+l}})\Delta B^j_{i+l}$$

and

$$\widehat{Z}^{t,x}_{t_{i-1}} := \frac{1}{\sum_{l=0}^{k-1}\Delta t_{i+l}} E_{i-1}\left[\widehat{V}^{t,x}_{t_{i+k-1}} \sum_{l=0}^{k-1}\Delta W_{i+l}\right].$$

Remark. The scheme above follows a mixture of the approaches followed by [49] and [8] and consequently, in the case where each chain is of length one, is similar to the scheme described in [2].

Continuous-time Representation

We now construct a continuous-time representation of our discretization scheme. We note that this approach is similar to that taken in [8] for their implicit discretization scheme for BSDEs. This continuous-time representation will be used in the derivation of the error bound of our scheme.

To this end, if we now define for $s \in [t_{i-1}, t_i]$

$$M_s := E\left[\widehat{U}^{t,x}_{t_i}\middle|\,\mathcal{G}_s\right]$$

then

$$M_{t_{i-1}} = \widehat{Y}^{t,x}_{t_{i-1}} \quad \text{and} \quad M_{t_i} = \widehat{U}^{t,x}_{t_i}.$$

By an extension of the Martingale Representation Theorem (see [36], p. 212), there exists a process $\bar{Z}_s$ adapted to $\mathcal{G}_s$ such that

$$M_s = M_{t_{i-1}} + \int_{t_{i-1}}^{s} \left\langle \bar{Z}^{t,x}_r, dW_r \right\rangle. \tag{7.4}$$

Definition 7.13. *We denote by $\bar{Z}$ the stochastic process defined by equation (7.4).*

It follows that

$$M_{t_{i-1}} = M_{t_i} - \int_{t_{i-1}}^{t_i} \left\langle \bar{Z}^{t,x}_r, dW_r \right\rangle$$

and so

$$\begin{aligned} \widehat{Y}^{t,x}_{t_{i-1}} &= \widehat{U}^{t,x}_{t_i} - \int_{t_{i-1}}^{t_i} \left\langle \bar{Z}^{t,x}_r, dW_r \right\rangle \\ &= \widehat{Y}^{t,x}_{t_i} + \check{f}(t_i, \widehat{\theta}^{t,x}_{t_i})\Delta t_i - \sum_{j=1}^{m} \check{g}_j(t_i, \widehat{\theta}^{t,x}_{t_i})\Delta B^j_i - \int_{t_{i-1}}^{t_i} \left\langle \bar{Z}^{t,x}_r, dW_r \right\rangle. \end{aligned} \tag{7.5}$$

As a consequence, if we define for $s \in [t_{i-1}, t_i)$

$$\widehat{Y}^{t,x}_s := \widehat{Y}^{t,x}_{t_i} + \check{f}(t_i, \widehat{\theta}^{t,x}_{t_i})(t_i - s) - \sum_{j=1}^{m} \check{g}_j(t_i, \widehat{\theta}^{t,x}_{t_i})(B^j_{t_i} - B^j_s) - \int_s^{t_i} \left\langle \bar{Z}^{t,x}_r, dW_r \right\rangle$$

then $\widehat{Y}$ is a continuous time process that takes the correct values at each discretization point and consequently our notation is consistent.

If t_{i-1} is not the start point of a chain of length greater than one we have by (7.5) the relation

$$\begin{aligned}
\widehat{Z}^{t,x}_{t_{i-1}} &:= \frac{1}{\Delta t_i} E_{i-1}\left[\widehat{U}^{t,x}_{t_i} \Delta W_i\right] \\
&= \frac{1}{\Delta t_i} E_{i-1}\left[\left(\widehat{Y}^{t,x}_{t_{i-1}} + \int_{t_{i-1}}^{t_i} \left\langle \bar{Z}^{t,x}_r, dW_r \right\rangle\right) \Delta W_i\right] \\
&= \frac{1}{\Delta t_i} E_{i-1}\left[\int_{t_{i-1}}^{t_i} \bar{Z}^{t,x}_r dr\right]
\end{aligned}$$

and so $\widehat{Z}^{t,x}_{t_{i-1}}$ minimizes the mean-square error among all $\mathcal{G}_{t_{i-1}}$-measurable random variables in estimating $\bar{Z}^{t,x}_r$ over $[t_{i-1}, t_i]$.

When t_{i-1} is the start point of a chain of length $k > 1$ we have by (7.5) that since $E_{i-1}\left[\widehat{Y}^{t,x}_{t_{i-1}} \sum_{l=0}^{k-1} \Delta W_{i+l}\right] = 0$,

$$\begin{aligned}
\widehat{Z}^{t,x}_{t_{i-1}} &= \frac{1}{\sum_{l=0}^{k-1} \Delta t_{i+l}} E_{i-1}\left[\left(\widehat{V}^{t,x}_{t_{i+k-1}} - \widehat{Y}^{t,x}_{t_{i-1}}\right) \sum_{l=0}^{k-1} \Delta W_{i+l}\right] \\
&= \frac{1}{\sum_{l=0}^{k-1} \Delta t_{i+l}} E_{i-1}\left[\left(\left(\widehat{U}^{t,x}_{t_i} - \widehat{Y}^{t,x}_{t_{i-1}}\right) + \left(\widehat{U}^{t,x}_{t_{i+1}} - \widehat{Y}^{t,x}_{t_i}\right) + \dots\right.\right. \\
&\qquad \left.\left. + \left(\widehat{U}^{t,x}_{t_{i+k-1}} - \widehat{Y}^{t,x}_{t_{i+k-2}}\right)\right) \sum_{l=0}^{k-1} \Delta W_{i+l}\right] \\
&= \frac{1}{\sum_{l=0}^{k-1} \Delta t_{i+l}} E_{i-1}\left[\int_{t_{i-1}}^{t_{i+k-1}} \left\langle \bar{Z}^{t,x}_r, dW_r \right\rangle \sum_{l=0}^{k-1} \Delta W_{i+l}\right] \\
&= \frac{1}{\sum_{l=0}^{k-1} \Delta t_{i+l}} E_{i-1}\left[\int_{t_{i-1}}^{t_{i+k-1}} \bar{Z}^{t,x}_r dr\right]
\end{aligned}$$

and so $\widehat{Z}^{t,x}_{t_{i-1}}$ minimizes the mean-square error among all $\mathcal{G}_{t_{i-1}}$-measurable random variables in estimating $\bar{Z}^{t,x}_r$ over $[t_{i-1}, t_{i+k-1}]$.

7.4. Error Estimate

Lemma 7.14. *Let $(\check{Y}, \check{Z})$ denote the solution to the BDSDE with Lipschitz coefficients $\{\check{f}, \check{g}, \check{h}\}$ as defined in Definition 5.12 on page 66, $\widehat{Y}$ the approximation to $\check{Y}$ defined in Definition 7.12 and $\bar{Z}$ the approximation to $\check{Z}$ defined in Definition 7.13. For each $K_{\mathfrak{B}} > 0$ there exist positive constants C and Δt_0 independent of C_E, C_M*

and C_G *such that for any* $\Delta t \in \left(0, \Delta t_0 \wedge \frac{C_M^2}{C_G^2} \wedge C_E\right]$

$$\max_{0\le i\le n} \mathcal{E}\left[|\check{Y}_{t_i}^{t,x} - \widehat{Y}_{t_i}^{t,x}|^2\right] + \mathcal{E}_T\left[|\check{Z}_r^{t,x} - \bar{Z}_r^{t,x}|^2\right] \le C\Delta t(C_G^2\Upsilon(K_{\mathfrak{B}}) + 1)$$

where $\Upsilon(K_{\mathfrak{B}})$ *is as defined in Definition 7.6.*

Proof. This proof follows a mixture of the approaches followed by [49] and [8] and consequently is similar to the corresponding proof in [2] (although, of course, the schemes are slightly different and there is no concept of a chain in [2]).

For $i = 1, \dots, n$, $s \in [t_{i-1}, t_i)$, we have that

$$\begin{aligned}\check{Y}_s^{t,x} - \widehat{Y}_s^{t,x} = \check{Y}_{t_i}^{t,x} - \widehat{Y}_{t_i}^{t,x} &+ \int_s^{t_i} \left\{\check{f}(r, \check{\theta}_r^{t,x}) - \check{f}(t_i, \widehat{\theta}_{t_i}^{t,x})\right\} dr \\ &- \sum_{j=1}^m \int_s^{t_i} \left\{\check{g}_j(r, \check{\theta}_r^{t,x}) - \check{g}_j(t_i, \widehat{\theta}_{t_i}^{t,x})\right\} \overleftarrow{dB}_r^j - \int_s^{t_i} \left\langle \check{Z}_r^{t,x} - \bar{Z}_r^{t,x}, dW_r\right\rangle.\end{aligned}$$

By our standard argument we have that for any $\delta > 0$,

$$\begin{aligned}J_s &:= \mathcal{E}\left[|\check{Y}_s^{t,x} - \widehat{Y}_s^{t,x}|^2\right] + E\left[\int_s^{t_i}\int_{\mathbb{R}^d} |\check{Z}_r^{t,x} - \bar{Z}_r^{t,x}|^2\rho(x)dxdr\right] \\ &\le \mathcal{E}\left[|\check{Y}_{t_i}^{t,x} - \widehat{Y}_{t_i}^{t,x}|^2\right] + E\left[\int_s^{t_i}\int_{\mathbb{R}^d}\left(\frac{1}{\delta}|\check{Y}_r^{t,x} - \widehat{Y}_r^{t,x}|^2 + \delta|\check{f}(r, \check{\theta}_r^{t,x}) - \check{f}(t_i, \widehat{\theta}_{t_i}^{t,x})|^2\right.\right. \\ &\quad \left.\left. + \sum_{j=1}^m |\check{g}_j(r, \check{\theta}_r^{t,x}) - \check{g}_j(t_i, \widehat{\theta}_{t_i}^{t,x})|^2\right)\rho(x)dxdr\right].\end{aligned}$$

Now,

$$\begin{aligned}|\check{f}(r, \check{\theta}_r^{t,x}) - \check{f}(t_i, \widehat{\theta}_{t_i}^{t,x})|^2 &\le \left(1 + \frac{1}{\delta}\right)|\check{f}(r, \check{\theta}_r^{t,x}) - \check{f}(t_i, \widehat{X}_{t_i}^{t,x}, \check{Y}_r^{t,x}, \check{Z}_r^{t,x})|^2 \\ &\quad + (1+\delta)|\check{f}(t_i, \widehat{X}_{t_i}^{t,x}, \check{Y}_r^{t,x}, \check{Z}_r^{t,x}) - \check{f}(t_i, \widehat{\theta}_{t_i}^{t,x})|^2 \\ &\le \left(1 + \frac{1}{\delta}\right)|\check{f}(r, \check{\theta}_r^{t,x}) - \check{f}(t_i, \widehat{X}_{t_i}^{t,x}, \check{Y}_r^{t,x}, \check{Z}_r^{t,x})|^2 \\ &\quad + (1+\delta)L\left(|\check{Y}_r^{t,x} - \widehat{Y}_{t_i}^{t,x}|^2 + |\check{Z}_r^{t,x} - \widehat{Z}_{t_i}^{t,x}|^2\right),\end{aligned}$$

$$\begin{aligned}|\check{g}_j(r, \check{\theta}_r^{t,x}) - \check{g}_j(t_i, \widehat{\theta}_{t_i}^{t,x})|^2 &\le \left(1 + \frac{1}{\delta}\right)|\check{g}_j(r, \check{\theta}_r^{t,x}) - \check{g}_j(t_i, \widehat{X}_{t_i}^{t,x}, \check{Y}_r^{t,x}, \check{Z}_r^{t,x})|^2 \\ &\quad + (1+\delta)L\left(|\check{Y}_r^{t,x} - \widehat{Y}_{t_i}^{t,x}|^2\right),\end{aligned}$$

$$|\check{Y}_r^{t,x} - \widehat{Y}_{t_i}^{t,x}|^2 \leq \left(1 + \frac{1}{\delta}\right)|\check{Y}_r^{t,x} - \check{Y}_{t_i}^{t,x}|^2 + (1+\delta)|\check{Y}_{t_i}^{t,x} - \widehat{Y}_{t_i}^{t,x}|^2.$$

Consequently, we have by Theorem 6.7 and Lemma 7.7 that

$$\begin{aligned}
J_s \leq\ & \mathcal{E}\left[|\check{Y}_{t_i}^{t,x} - \widehat{Y}_{t_i}^{t,x}|^2\right] \\
& + E\left[\int_s^{t_i}\int_{\mathbb{R}^d}\left(\frac{1}{\delta}|\check{Y}_r^{t,x} - \widehat{Y}_r^{t,x}|^2 + \delta\left(1+\frac{1}{\delta}\right)|\check{f}(r,\check{\theta}_r^{t,x}) - \check{f}(t_i,\widehat{X}_{t_i}^{t,x},\check{Y}_r^{t,x},\check{Z}_r^{t,x})|^2\right.\right. \\
& + L\delta(1+\delta)\left(|\check{Y}_r^{t,x} - \widehat{Y}_{t_i}^{t,x}|^2 + |\check{Z}_r^{t,x} - \widehat{Z}_{t_i}^{t,x}|^2\right) \\
& + \left(1+\frac{1}{\delta}\right)\sum_{j=1}^m |\check{g}_j(r,\check{\theta}_r^{t,x}) - \check{g}_j(t_i,\widehat{X}_{t_i}^{t,x},\check{Y}_r^{t,x},\check{Z}_r^{t,x})|^2 \\
& \left.\left. + (1+\delta)\left(mL|\check{Y}_r^{t,x} - \widehat{Y}_{t_i}^{t,x}|^2\right)\rho(x)dxdr\right] \\
\leq\ & \mathcal{E}\left[|\check{Y}_{t_i}^{t,x} - \widehat{Y}_{t_i}^{t,x}|^2\right] \\
& + E\left[\int_s^{t_i}\int_{\mathbb{R}^d}\left(\frac{1}{\delta}|\check{Y}_r^{t,x} - \widehat{Y}_r^{t,x}|^2 + (\delta+m)\left(1+\frac{1}{\delta}\right)C\Delta t C_G^2\Upsilon(K_{\mathfrak{B}})\right.\right. \\
& + L(m+\delta)(1+\delta)\left(\left(1+\frac{1}{\delta}\right)|\check{Y}_r^{t,x} - \check{Y}_{t_i}^{t,x}|^2 + (1+\delta)|\check{Y}_{t_i}^{t,x} - \widehat{Y}_{t_i}^{t,x}|^2\right) \\
& \left.+ (1+\delta)L\delta|\check{Z}_r^{t,x} - \widehat{Z}_{t_i}^{t,x}|^2\rho(x)dxdr\right] \\
\leq\ & C'\int_s^{t_i} J_r dr + (1+C\Delta t)\mathcal{E}\left[|\check{Y}_{t_i}^{t,x} - \widehat{Y}_{t_i}^{t,x}|^2\right] \\
& + (1+\delta)L\delta E\left[\int_s^{t_i}\int_{\mathbb{R}^d}|\check{Z}_r^{t,x} - \widehat{Z}_{t_i}^{t,x}|^2\rho(x)dxdr\right] + C(\Delta t)^2 C_G^2\Upsilon(K_{\mathfrak{B}}).
\end{aligned}$$

It then follows by Gronwall's inequality that

$$\begin{aligned}
J_{t_{i-1}} \leq\ & (1+C'\Delta t)\left(\mathcal{E}\left[|\check{Y}_{t_i}^{t,x} - \widehat{Y}_{t_i}^{t,x}|^2\right]\right. \\
& \left. + (1+\delta)L\delta E\left[\int_{t_{i-1}}^{t_i}\int_{\mathbb{R}^d}|\check{Z}_r^{t,x} - \widehat{Z}_{t_i}^{t,x}|^2\rho(x)dxdr\right]\right) \\
& + C(\Delta t)^2(C_G^2\Upsilon(K_{\mathfrak{B}}) + 1).
\end{aligned} \tag{7.6}$$

To see this note that as we are interested in convergence as $\Delta t \to 0$ and since C' does not depend upon Δt we may assume that $C'\Delta t \leq 1$. It then follows that for

$\Delta t \le \Delta t_0$ with $\Delta t_0 = \frac{1}{C'}$

$$\begin{aligned}
e^{C'\Delta t} &= 1 + C'\Delta t + \frac{(C'\Delta t)^2}{2!} + \dots \\
&= 1 + C'\Delta t \left(1 + \frac{C'\Delta t}{2!} + \frac{(C'\Delta t)^2}{3!} + \dots\right) \\
&\le 1 + e^{C'\Delta t} C'\Delta t \\
&\le 1 + 3C'\Delta t.
\end{aligned}$$

Now, for the solution (Y^*, Z^*) of any associated BDSDE with mollified coefficients we have that for any $\delta > 0$

$$\begin{aligned}
|\check{Z}_r^{t,x} - \widehat{Z}_{t_i}^{t,x}|^2 &\le \left(1 + \frac{1}{\delta}\right) |\check{Z}_r^{t,x} - \check{Z}_{t_i}^{t,x,*}|^2 + (1+\delta)|\check{Z}_{t_i}^{t,x,*} - \widehat{Z}_{t_i}^{t,x}|^2 \\
&\le \left(2 + \frac{2}{\delta}\right) \left(|\check{Z}_r^{t,x} - \check{Z}_r^{t,x,*}|^2 + |\check{Z}_r^{t,x,*} - \check{Z}_{t_i}^{t,x,*}|^2\right) + (1+\delta)|\check{Z}_{t_i}^{t,x,*} - \widehat{Z}_{t_i}^{t,x}|^2.
\end{aligned}$$

Now, if t_i is not the start point of a chain of length greater than one,

$$\begin{aligned}
\mathcal{E}\left[|\check{Z}_{t_i}^{t,x,*} - \widehat{Z}_{t_i}^{t,x}|^2\right] &= \mathcal{E}\left[\left|\check{Z}_{t_i}^{t,x,*} - \frac{1}{\Delta t_{i+1}} E_i\left[\int_{t_i}^{t_{i+1}} \bar{Z}_r^{t,x} dr\right]\right|^2\right] \\
&= \mathcal{E}\left[\left|\frac{1}{\Delta t_{i+1}} E_i\left[\int_{t_i}^{t_{i+1}} \check{Z}_{t_i}^{t,x,*} - \bar{Z}_r^{t,x} dr\right]\right|^2\right] \\
&\le \frac{1}{(\Delta t_{i+1})^2} \mathcal{E}\left[E_i\left[\left|\int_{t_i}^{t_{i+1}} \check{Z}_{t_i}^{t,x,*} - \bar{Z}_r^{t,x} dr\right|^2\right]\right] \\
&= \frac{1}{(\Delta t_{i+1})^2} \mathcal{E}\left[\left|\int_{t_i}^{t_{i+1}} \check{Z}_{t_i}^{t,x,*} - \bar{Z}_r^{t,x} dr\right|^2\right] \\
&\le \frac{1}{\Delta t_{i+1}} E\left[\int_{t_i}^{t_{i+1}} \int_{\mathbb{R}^d} |\check{Z}_{t_i}^{t,x,*} - \bar{Z}_r^{t,x}|^2 \rho(x) dx dr\right] \\
&\le \frac{1}{\Delta t_{i+1}} E\left[\int_{t_i}^{t_{i+1}} \int_{\mathbb{R}^d} \left(\left(2 + \frac{2}{\delta}\right) \left(|\check{Z}_{t_i}^{t,x,*} - \check{Z}_r^{t,x,*}|^2 + |\check{Z}_r^{t,x,*} - \check{Z}_r^{t,x}|^2\right)\right.\right. \\
&\quad \left.\left. + \ (1+\delta)|\check{Z}_r^{t,x} - \bar{Z}_r^{t,x}|^2\right) \rho(x) dx dr\right] \\
&\le \frac{1}{\Delta t_{i+1}} E\left[\int_{t_i}^{t_{i+1}} \int_{\mathbb{R}^d} \left(C\left(|\check{Z}_{t_i}^{t,x,*} - \check{Z}_r^{t,x,*}|^2 + |\check{Z}_r^{t,x,*} - \check{Z}_r^{t,x}|^2\right)\right.\right. \\
&\quad \left.\left. + \ (1+\delta)|\check{Z}_r^{t,x} - \bar{Z}_r^{t,x}|^2\right) \rho(x) dx dr\right]
\end{aligned}$$

where we have used Fubini's theorem and the Cauchy-Schwarz inequality. Since

$\Delta t_i \leq \Delta t_{i+1}$ (which we know because t_i is not the start point of a chain of length greater than one), we have that

$$\begin{aligned}
&E\left[\int_{t_{i-1}}^{t_i}\int_{\mathbb{R}^d}|\check{Z}_r^{t,x}-\widehat{Z}_{t_i}^{t,x}|^2\rho(x)dxdr\right]\\
&\leq CE\left[\int_{t_{i-1}}^{t_{i+1}}\int_{\mathbb{R}^d}\left(|\check{Z}_{t_i}^{t,x,*}-\check{Z}_r^{t,x,*}|^2+|\check{Z}_r^{t,x,*}-\check{Z}_r^{t,x}|^2\right)\rho(x)dxdr\right]\\
&\quad+(1+\delta)^2E\left[\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}^d}|\check{Z}_r^{t,x}-\bar{Z}_r^{t,x}|^2\rho(x)dxdr\right].
\end{aligned}$$

Consequently, recalling (7.6) and choosing δ so that

$$(1+C'\Delta t)(1+\delta)^3L\delta=\frac{1}{4}$$

we have that

$$A_{i-1}+B_{i-1}+\frac{1}{4}B_i\leq(1+C\Delta t)\,A_i+\frac{1}{2}B_i+CC_i+C(\Delta t)^2(C_G^2\Upsilon(K_{\mathfrak{B}})+1) \quad (7.7)$$

where we have defined

$$A_i:=\mathcal{E}\left[|\check{Y}_{t_i}^{t,x}-\widehat{Y}_{t_i}^{t,x}|^2\right],$$

$$B_i:=E\left[\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}^d}|\check{Z}_r^{t,x}-\bar{Z}_r^{t,x}|^2\rho(x)dxdr\right]$$

and

$$C_i:=E\left[\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}^d}\left(|\check{Z}_{t_i}^{t,x,*}-\check{Z}_r^{t,x,*}|^2+|\check{Z}_r^{t,x,*}-\check{Z}_r^{t,x}|^2\right)\rho(x)dxdr\right]$$

Similarly, when t_i is the start point of a chain of length $k>1$,

$$\begin{aligned}
\mathcal{E}\left[|\check{Z}_{t_i}^{t,x,*}-\widehat{Z}_{t_i}^{t,x}|^2\right]&=\mathcal{E}\left[\left|\check{Z}_{t_i}^{t,x,*}-\frac{1}{\sum_{l=1}^k\Delta t_{i+l}}E_i\left[\int_{t_i}^{t_{i+k}}\bar{Z}_r^{t,x}dr\right]\right|^2\right]\\
&\leq\frac{1}{\sum_{l=1}^k\Delta t_{i+l}}E\left[\int_{t_i}^{t_{i+k}}\int_{\mathbb{R}^d}\left(C\left(|\check{Z}_{t_i}^{t,x,*}-\check{Z}_r^{t,x,*}|^2+|\check{Z}_r^{t,x,*}-\check{Z}_r^{t,x}|^2\right)\right.\right.\\
&\quad\left.\left.+\,(1+\delta)|\check{Z}_r^{t,x}-\bar{Z}_r^{t,x}|^2\right)\rho(x)dxdr\right].
\end{aligned}$$

Since $\Delta t_i \le \sum_{l=1}^{k} \Delta t_{i+l}$, we have that

$$\begin{aligned}
&E\left[\int_{t_{i-1}}^{t_i}\int_{\mathbb{R}^d}|\check{Z}_r^{t,x}-\widehat{Z}_{t_i}^{t,x}|^2\rho(x)dxdr\right]\\
&\le CE\left[\int_{t_{i-1}}^{t_{i+k}}\int_{\mathbb{R}^d}\left(|\check{Z}_{t_i}^{t,x,*}-\check{Z}_r^{t,x,*}|^2+|\check{Z}_r^{t,x,*}-\check{Z}_r^{t,x}|^2\right)\rho(x)dxdr\right]\\
&\quad +(1+\delta)^2E\left[\int_{t_i}^{t_{i+k}}\int_{\mathbb{R}^d}|\check{Z}_r^{t,x}-\bar{Z}_r^{t,x}|^2\rho(x)dxdr\right]
\end{aligned}$$

and

$$\begin{aligned}
&A_{i-1}+B_{i-1}+\frac{1}{4}B_i\\
&\le (1+C\Delta t)\,A_i+\frac{1}{2}B_i+\frac{1}{2}\sum_{j=i+1}^{i+k}(B_j+CC_j)+C(\Delta t)^2(C_G^2\Upsilon(K_{\mathfrak{B}})+1)
\end{aligned} \tag{7.8}$$

Defining $I_i := A_i + B_i$, we have by Lemma 7.8 that

$$\max_{0\le i\le n} I_i \le C(I_n+\sum_{i=1}^{n}C_i+\Delta t(C_G^2\Upsilon(K_{\mathfrak{B}})+1)).$$

By Theorems 5.19 and 6.9 (and choosing ϵ small enough), $\sum_{i=1}^{n} C_i < C\Delta t C_G^2\Upsilon(K_{\mathfrak{B}})$. By Lemma 7.7 and since $B_n = 0$,

$$I_n = E\left[\int_{\mathbb{R}^d}|\check{h}(X_T^{t,x})-\check{h}(\widehat{X}_T^{t,x})|^2\rho(x)dx\right]\le C\Delta t C_G^2\Upsilon(K_{\mathfrak{B}}).$$

It follows that

$$\max_{0\le i\le n} I_i \le C\Delta t(C_G^2\Upsilon(K_{\mathfrak{B}})+1)$$

which gives the result for Y. Summing both sides of (7.7) and (7.8) gives us that

$$\sum_{i=0}^{n-2} I_i+\frac{1}{4}E\left[\int_t^T\int_{\mathbb{R}^d}|\check{Z}_r^{t,x}-\bar{Z}_r^{t,x}|^2\rho(x)dxdr\right]\le\sum_{i=0}^{n-1}(1+C\Delta t)I_i+C\Delta t(C_G^2\Upsilon(K_{\mathfrak{B}})+1).$$

Consequently, we have that

$$\begin{aligned}
&E\left[\int_t^T \int_{\mathbb{R}^d} |\check{Z}_r^{t,x} - \bar{Z}_r^{t,x}|^2 \rho(x)dxdr\right] \\
&\leq C(1 + C\Delta t)I_{n-1} + I_0 + C\Delta t \sum_{i=0}^{n-1} I_i + C\Delta t(C_G^2 \Upsilon(K_{\mathfrak{B}}) + 1) \\
&\leq C\Delta t(C_G^2 \Upsilon(K_{\mathfrak{B}}) + 1)
\end{aligned}$$

as required. □

Let us now define for $s \in [t_{i-1}, t_i)$, the step processes $\widehat{X}_s^{t,x} := \widehat{X}_{t_{i-1}}^{t,x}$, $\widehat{Y}_s^{t,x} := \widehat{Y}_{t_{i-1}}^{t,x}$, $\widehat{Z}_s^{t,x} := \widehat{Z}_{t_{i-1}}^{t,x}$ where $\widehat{X}_{t_{i-1}}^{t,x}$ is the Euler approximation to X and $(\widehat{Y}, \widehat{Z})$ are the approximation to $(\check{Y}, \check{Z})$ defined in Definition 7.12. Furthermore, we define $\widehat{\theta}_s^{t,x} := (\widehat{X}_s^{t,x}, \widehat{Y}_s^{t,x}, \widehat{Z}_s^{t,x})$.

Theorem 7.15. *Let $(\check{Y}, \check{Z})$ denote the solution to the BDSDE with Lipschitz coefficients $\{\check{f}, \check{g}, \check{h}\}$ as defined in Definition 5.12 on page 66 and let $\widehat{X}, \widehat{Y}, \widehat{Z}$ be defined as above. For each $K_{\mathfrak{B}} > 0$ there exist positive constants C and Δt_0 independent of C_E, C_M and C_G such that for any $\Delta t \in \left(0, \Delta t_0 \wedge \frac{C_M^2}{C_G^2} \wedge C_E\right]$*

$$\max_{1 \leq i \leq n} E\left[\sup_{t_{i-1} \leq s \leq t_i} \int_{\mathbb{R}^d} |\check{Y}_s^{t,x} - \widehat{Y}_s^{t,x}|^2 \rho(x)dx\right] + \mathcal{E}_T\left[|\check{Z}_r^{t,x} - \widehat{Z}_r^{t,x}|^2\right] \leq C\Delta t(C_G^2 \Upsilon(K_{\mathfrak{B}}) + 1)$$

where $\Upsilon(K_{\mathfrak{B}})$ is as defined in Definition 7.6.

Proof. This proof is very similar to corresponding proof in [49]. For $1 \leq i \leq n$,

$$\begin{aligned}
&E\left[\sup_{t_{i-1} \leq s \leq t_i} \int_{\mathbb{R}^d} |\check{Y}_s^{t,x} - \widehat{Y}_s^{t,x}|^2 \rho(x)dx\right] \\
&= E\left[\sup_{t_{i-1} \leq s \leq t_i} \int_{\mathbb{R}^d} |\check{Y}_s^{t,x} - \widehat{Y}_{t_{i-1}}^{t,x}|^2 \rho(x)dx\right] \\
&\leq 2E\left[\sup_{t_{i-1} \leq s \leq t_i} \int_{\mathbb{R}^d} |\check{Y}_s^{t,x} - \check{Y}_{t_{i-1}}^{t,x}|^2 \rho(x)dx\right] + 2\mathcal{E}\left[|\check{Y}_{t_{i-1}}^{t,x} - \widehat{Y}_{t_{i-1}}^{t,x}|^2\right].
\end{aligned}$$

and the result for Y follows by Theorem 6.7 and Lemma 7.14.

For the solution $(\widetilde{Y}, \widetilde{Z})$ of any associated BDSDE with mollified coefficients we have that

$$E\left[\int_{t_{i-1}}^{t_i}\int_{\mathbb{R}^d}|\bar{Z}_r^{t,x}-\widehat{Z}_{t_{i-1}}^{t,x}|^2\rho(x)dxdr\right]\leq E\left[\int_{t_{i-1}}^{t_i}\int_{\mathbb{R}^d}|\bar{Z}_r^{t,x}-\widetilde{Z}_{t_{i-1}}^{t,x}|^2\rho(x)dxdr\right]$$

since, as previously noted, $\widehat{Z}_{t_{i-1}}$ minimizes the mean square error over all $\mathcal{G}_{t_{i-1}}$-measurable random variables in estimating $\bar{Z}$ over $[t_{i-1},t_i]$. Consequently, it follows that

$$\begin{aligned}
&\mathcal{E}_T\left[|\check{Z}_r^{t,x}-\widehat{Z}_r^{t,x}|^2\right]\\
&\leq 2\mathcal{E}_T\left[|\check{Z}_r^{t,x}-\bar{Z}_r^{t,x}|^2\right]+2\mathcal{E}_T\left[|\bar{Z}_r^{t,x}-\widehat{Z}_r^{t,x}|^2\right]\\
&=2\mathcal{E}_T\left[|\check{Z}_r^{t,x}-\bar{Z}_r^{t,x}|^2\right]+2\sum_{i=1}^n E\left[\int_{t_{i-1}}^{t_i}\int_{\mathbb{R}^d}|\bar{Z}_r^{t,x}-\widehat{Z}_{t_{i-1}}^{t,x}|^2\rho(x)dxdr\right]\\
&\leq 2\mathcal{E}_T\left[|\check{Z}_r^{t,x}-\bar{Z}_r^{t,x}|^2\right]+2\sum_{i=1}^n E\left[\int_{t_{i-1}}^{t_i}\int_{\mathbb{R}^d}|\bar{Z}_r^{t,x}-\widetilde{Z}_{t_{i-1}}^{t,x}|^2\rho(x)dxdr\right]\\
&\leq 2\mathcal{E}_T\left[|\check{Z}_r^{t,x}-\bar{Z}_r^{t,x}|^2\right]\\
&\quad+6\sum_{i=1}^n E\left[\int_{t_i}^{t_{i-1}}\int_{\mathbb{R}^d}\left(|\bar{Z}_r^{t,x}-\check{Z}_r^{t,x}|^2+|\check{Z}_r^{t,x}-\widetilde{Z}_r^{t,x}|^2+|\widetilde{Z}_r^{t,x}-\widetilde{Z}_{t_{i-1}}^{t,x}|^2\right)\rho(x)dxdr\right]\\
&=8\mathcal{E}_T\left[|\check{Z}_r^{t,x}-\bar{Z}_r^{t,x}|^2\right]\\
&\quad+6\sum_{i=1}^n E\left[\int_{t_i}^{t_{i-1}}\int_{\mathbb{R}^d}|\check{Z}_r^{t,x}-\widetilde{Z}_r^{t,x}|^2+|\widetilde{Z}_r^{t,x}-\widetilde{Z}_{t_{i-1}}^{t,x}|^2\rho(x)dxdr\right]
\end{aligned}$$

and the result for $\check{Z}$ follows by Theorems 5.19 and 6.9 and Lemma 7.14. □

Remark. In the above result, the approximation improves as $\Delta t\to 0$ but worsens as the maximum slope $C_G\to\infty$ (as defined in Definition 5.12 on page 66). In Theorem 5.16, however, the approximation improves as $C_G\to\infty$. As a consequence, to have overall convergence of the scheme, we require that $\Delta t C_G^2\to 0$ as $\Delta t\to 0$ and $C_G\to\infty$. We could achieve this, for example, by having $\Delta t\sim\frac{1}{C_G^3}$.

It follows that if to construct an approximation we really do need to have $C_E\to 0$ or $C_M\to\infty$ (and so $C_G\to\infty$ at a relatively fast rate) then the upper bound given by the previous result on the rate of convergence could be *very* poor. Of course, this is not really a surprise since we are providing an upper bound for cases which include very irregular coefficients.

Combining Theorems 7.15 and 5.16 (on page 72) gives us the main result of the thesis:

Theorem 7.16. *Let (Y, Z) denote the solution to the BDSDE with measurable coefficients $\{f, g, h\}$ as defined in Definition 4.3 on page 42 and let $\widehat{X}, \widehat{Y}, \widehat{Z}$ be defined as above. For each $K_{\mathfrak{B}} > 0$ there exist positive constants C and Δt_0 independent of C_E, C_M and C_G such that for any $\Delta t \in \left(0, \Delta t_0 \wedge \frac{C_M^2}{C_G^2} \wedge C_E\right]$*

$$\max_{1 \le i \le n} E\left[\sup_{t_{i-1} \le s \le t_i} \int_{\mathbb{R}^d} |Y_s^{t,x} - \widehat{Y}_s^{t,x}|^2 \rho(x) dx\right] + \mathcal{E}_T\left[|Z_r^{t,x} - \widehat{Z}_r^{t,x}|^2\right]$$

$$\le C\Delta t(C_G^2 \Upsilon(K_{\mathfrak{B}}) + 1) + 2\beta(C_E, C_M) + CC_M^2\left(l(\mathfrak{B}_{\mathcal{T}}) + \mu(\mathfrak{B}_{\mathcal{X}})\right)$$

where $\Upsilon(K_{\mathfrak{B}})$ is as defined in Definition 7.6, l represents Lebesgue measure, μ is defined in Definition 5.14 on page 69 and β is defined in Definition 5.10 on page 64.

7.5. Conclusion

The main results of this chapter are Theorem 7.15 which gives a bound of the mean square error of our discretization scheme for BDSDEs with Lipschitz coefficients and Theorem 7.16 which combines Theorems 7.15 and 5.16 to give a bound of the mean square error of our discretization scheme for BDSDEs with measurable coefficients.

Remark. Let us recall Theorem 3.2 on page 37 restated in the review of [50] on the connection between BDSDEs and SPDEs. In light of Theorem 3.2, the main result of this chapter can be recast as a result on the approximation of SPDEs with measurable coefficients via a time discretization scheme.

In practice, the application of Theorem 7.16 separates into three stages:

1. In the first stage, we approximate our measurable coefficients with a set of Lipschitz coefficients and in doing construct a partition of $[t, T]$, $\mathcal{T}$, which defines certain properties of the Lipschitz coefficients.

2. In the second stage, based upon certain properties of the Lipschitz coefficients, we refine the partition $\mathcal{T}$ to obtain $\mathcal{T}^*$ so that when discretized using $\mathcal{T}^*$, the Lipschitz coefficients have certain nice properties.

3. In the final stage, the intervals of $\mathcal{T}^*$ are grouped into chains and our discretization scheme is applied.

The first stage is highly dependent upon the measurable coefficients and so should be considered in the context of a specific problem. As a consequence, we now provide a brief summary of the final two stages where we assume as given a partition $\mathcal{T}$ of $[t, T]$ parameterised by partition constant C_E as defined in Definition 5.1 on page 47 and Lipschitz coefficients $\{\check{f}, \check{g}, \check{h}\}$ as defined in Definition 5.12 on page 66.

1. Construct $\mathcal{T}^1$ from $\mathcal{T}$ as specified in Definition 6.2 on page 78 which is straightforward.

2. Construct $\mathcal{T}^2$ from $\mathcal{T}^1$ as specified in Definition 6.5 on page 81. The intermediate step of constructing $\mathcal{T}^{1.1}$ requires calculating for each interval $[r_{i-1}, r_i) \in \mathcal{T}^1$ the value

$$u_i := \int_{r_{i-1}}^{r_i} \int_{\mathbb{R}^d} \sum_{j=1}^{m} |\check{g}_j(r, x, 0)|^2 \rho(x) dx dr.$$

   We note that this may well require the use of numerical integration.

3. Construct $\mathcal{T}^3$ from $\mathcal{T}^2$ as specified in Definition 7.2 on page 106 which is straightforward.

4. Construct $\mathcal{T}^*$ from $\mathcal{T}^3$ as specified in Definition 7.3 on page 107. We note that this requires the computation of the value

$$\frac{\Delta t_i}{\sqrt{\Delta t l(\mathfrak{B}_{\mathcal{T}})}}$$

   for each $\mathcal{T}_i^3 \subset \mathfrak{B}_{\mathcal{T}}$. We note, however, that this would not be hard to do programmatically since the values Δt_i and $l(\mathfrak{B}_{\mathcal{T}})$ are readily calculated.

5. Construct a chain on $\mathcal{T}^*$ as defined in Definition 7.11 on page 119 which is straightforward.

6. Compute the Euler approximation of X given by given by

$$\begin{aligned} \widehat{X}_{t_0}^{t,x} &:= x, \\ \widehat{X}_{t_{i+1}}^{t,x} &:= \widehat{X}_{t_i}^{t,x} + b(\widehat{X}_{t_i}^{t,x})\Delta t_i + \sigma(\widehat{X}_{t_i}^{t,x})\Delta W_{i+1} \end{aligned}$$

where $\Delta W_{i+1} := W_{t_{i+1}} - W_{t_i}$. To implement this, one would typically simulate a large number of Brownian paths with values at the points in time defined by $\mathcal{T}^*$ and calculate $\widehat{X}$ for each of these paths at the same time points.

7. Initialise the backward scheme with $\widehat{Y}^{t,x}_{t_n} := \check{h}(\widehat{X}_{t_n})$ and $\widehat{Z}^{t,x}_{t_n} := 0$.

8. For $i = n, \dots, 1$, calculate $\widehat{Y}^{t,x}_{t_{i-1}}$ using the formula

$$\widehat{Y}^{t,x}_{t_{i-1}} := E_{i-1}\left[\widehat{Y}^{t,x}_{t_i} + \check{f}(t_i, \widehat{\theta}^{t,x}_{t_i})\Delta t_i - \sum_{j=1}^{m} \check{g}_j(t_i, \widehat{\theta}^{t,x}_{t_i})\Delta B^j_i\right]$$

where and $\Delta B^j_i := B^j_{t_i} - B^j_{t_{i-1}}$.

9. For $i = n, \dots, 1$, calculate $\widehat{Z}^{t,x}_{t_{i-1}}$ using the formula:

   a) (if t_{i-1} is not the start point of a chain of length greater than one)

$$\widehat{Z}^{t,x}_{t_{i-1}} := \frac{1}{\Delta t_i} E_{i-1}\left[\widehat{Y}^{t,x}_{t_i} + \check{f}(t_i, \widehat{\theta}^{t,x}_{t_i})\Delta t_i - \sum_{j=1}^{m} \check{g}_j(t_i, \widehat{\theta}^{t,x}_{t_i})\Delta B^j_i \Delta W_i\right].$$

   b) (if t_{i-1} is the start point of a chain of length $k > 1$)

$$\widehat{Z}^{t,x}_{t_{i-1}} := \frac{1}{\sum_{l=0}^{k-1} \Delta t_{i+l}} E_{i-1}\left[\widehat{V}^{t,x}_{t_{i+k-1}} \sum_{l=0}^{k-1} \Delta W_{i+l}\right]$$

   where

$$\widehat{V}^{t,x}_{t_{i+k-1}} := \widehat{Y}^{t,x}_{t_{i+k-1}} + \sum_{l=0}^{k-1} \check{f}(t_{i+l}, \widehat{\theta}^{t,x}_{t_{i+l}})\Delta t_{i+l} - \sum_{l=0}^{k-1}\sum_{j=1}^{m} \check{g}_j(t_{i+l}, \widehat{\theta}^{t,x}_{t_{i+l}})\Delta B^j_{i+l}.$$

Remark. We note that one possible approach to calculate the conditional expectations above could be to perform a least squares regression as explored in [17] for BSDEs.

Remark. We note that it is not strictly necessary to use the Euler approximation to construct $\widehat{X}$. Indeed, any approximation of X that satisfies the inequality of Lemma 7.10 would be sufficient.

8 Infinite Horizon Case

8.1. Introduction

In this chapter we take a first look at the approximation of infinite horizon BDSDEs (as introduced in [50]) via time discretization. For the sake of simplicity, we restrict our presentation to the case where W and B are one-dimensional standard Brownian motions and the coefficients f and g are Lipschitz in t and x. We do not, however, restrict our attention to the BSDE case as we have seen in the previous chapters that the transition from BSDEs to BDSDEs is not without casualties (for example the dependence of g upon Z).

Indeed, it seems that the strategy used to derive the regularity of Z is somewhat problematic in the infinite horizon case. As a consequence, in this first look, we will also remove the dependence of f upon Z. The exact problem statement for this chapter is given by the Infinite Horizon Problem of Section 4.2.2. Specifically, we will approximate the BDSDE (4.2) with contractive coefficients as defined in Definition 4.4 on page 44. We note that the contraction coefficient μ of Definition 4.4 is similar in spirit to the contraction defined for the SDE case in [48].

In [50], it was shown (under more general contractive assumptions on the coefficients) that the BDSDE (4.2) has a unique solution (Y, Z) and $u(t, .) := Y_t^{t,\cdot}$ is a stationary solution of the corresponding SPDE. As a consequence, if we are able to approximate the BDSDE (4.2), we are able to approximate the eventual state of the corresponding SPDE.

From [50] we know that we can approximate the BDSDE (4.2) with the following finite horizon BDSDE

$$e^{-Ks}Y_s^{t,x} = \int_s^T e^{-Kr}(f(X_r^{t,x}, Y_r^{t,x}) + KY_r^{t,x})dr - \int_s^T e^{-Kr}g(X_r^{t,x}, Y_r^{t,x})\overleftarrow{dB}_r \\ - \int_s^T e^{-Kr}Z_r^{t,x}dW_r. \tag{8.1}$$

The difference between the BDSDE (8.1) and the case we have considered in previous chapters is that our terminal time T is no longer fixed and can be arbitrarily large so we must take care that any bounds we subsequently derive do not depend upon T.

8.2. Preliminary Results

We note that nothing in this section is *really* new. Indeed, similar arguments are present in our previous chapters, [50], [49], [16], [23] and [24]. We include the following proofs (which are just minor adaptations of other proofs) for the sake of clarity.

Lemma 8.1. *Let X be defined as in Definition 4.1 on page 42. There exist positive constants C and K such that for any $s \geq t$*

$$E\left[e^{-Ks}|X_s^{t,x}|^2\right] \leq C(1+|x|^2)$$

Proof. Let us recall L_X from Definition 4.2 on page 42. By Itô's formula and Young's inequality we have that

$$\begin{aligned} E\left[|X_s^{t,x}|^2\right] &\leq |x|^2 + E\left[\int_t^s |X_r^{t,x}|^2 + |b(X_r^{t,x})|^2 + |\sigma(X_r^{t,x})|^2 dr\right] \\ &\leq |x|^2 + 2s(|b(0)|^2 + |\sigma(0)|^2) + E\left[\int_t^s (1+4L_X)|X_r^{t,x}|^2 dr\right] \\ &\leq |x|^2 + 2Cs + (1+4L_X)\int_t^s E\left[|X_r^{t,x}|^2\right] dr. \end{aligned}$$

By Gronwall's inequality, it follows that

$$E\left[|X_s^{t,x}|^2\right] \leq (|x|^2 + 2Cs)e^{(1+4L_X)s}.$$

The result follows upon choosing $K \geq 1 + 4L_X$. □

Remark. For $\phi = f, g$,

$$|\phi(X_r^{t,x}, 0)|^2 \leq C(1 + |X_r^{t,x}|^2).$$

It follows that if we choose $K > 1 + 4L_X$, then there exists a constant $C < \infty$ such that

$$E\left[\int_t^\infty e^{-Kr}\left(|f(X_r^{t,x}, 0)|^2 + |g(X_r^{t,x}, 0)|^2\right) dr\right] \leq C(1 + |x|^2) \tag{8.2}$$

- a condition that is more in line with that found in [50].

Lemma 8.2. *Let X be defined as in Definition 4.1 on page 42. Suppose that positive constants T and Δt are given and define $n := \frac{T-t}{\Delta t}$ and for $k = 0, \ldots, n$, $t_k := t + k\Delta t$. There exist positive constants C and K independent of T and Δt such that*

$$\sum_{i=1}^n E\left[\int_{t_{i-1}}^{t_i} e^{-Kt_i}|X_r^{t,x} - X_{t_i}^{t,x}|^2 dr\right] \leq C\Delta t(1 + |x|^2).$$

Proof. Let $s \in [t_{i-1}, t_i]$, then

$$|X_s^{t,x} - X_{t_i}^{t,x}|^2 \leq 2\left|\int_s^{t_i} b(X_r^{t,x})dr\right|^2 + 2\left|\int_s^{t_i} \sigma(X_r^{t,x})dW_r\right|^2$$

and so

$$\begin{aligned}
E\left[|X_s^{t,x} - X_{t_i}^{t,x}|^2\right] &\leq 2\Delta t E\left[\int_s^{t_i} |b(X_r^{t,x})|^2 dr\right] + 2E\left[\int_s^{t_i} |\sigma(X_r^{t,x})|^2 dr\right] \\
&\leq CE\left[\int_s^{t_i} (1 + |X_r^{t,x}|^2) dr\right] \\
&\leq C\Delta t + C\int_s^{t_i} E\left[|X_r^{t,x}|^2\right] dr.
\end{aligned}$$

Now, by Lemma 8.1, there exists a constant $K_1 > 0$ such that

$$\begin{aligned}
E\left[e^{-K_1 t_i}|X_s^{t,x} - X_{t_i}^{t,x}|^2\right] &\leq e^{-K_1 t_i} C\Delta t + C\int_s^{t_i} e^{-K_1 t_i} E\left[|X_r^{t,x}|^2\right] dr \\
&\leq C\Delta t(1 + |x|^2).
\end{aligned}$$

If we now choose $K = K_1 + \alpha$ for some $\alpha > 0$, it follows that

$$\sum_{i=1}^{n} E\left[\int_{t_{i-1}}^{t_i} e^{-Kt_i}|X_r^{t,x} - X_{t_i}^{t,x}|^2 dr\right] < C\Delta t(1+|x|^2)\int_t^\infty e^{-\alpha r}dr$$
$$= \frac{C}{\alpha}\Delta t(1+|x|^2)$$

as required □

Lemma 8.3. *Let X be defined as in Definition 4.1 on page 42, let $\widehat{X}$ be the Euler approximation to X as defined in Definition 7.9 on page 116 and suppose that a positive constant T is given. There exist positive constants C, K and Δt_0 independent of T such that for any $\Delta t \in (0, \Delta t_0]$*

$$\sum_{i=1}^{n} E\left[\int_{t_{i-1}}^{t_i} e^{-Kt_i}|X_r^{t,x} - \widehat{X}_{t_i}^{t,x}|^2 dr\right] \le C\Delta t(1+|x|^2)$$

where $n := \frac{T-t}{\Delta t}$ and for $k = 0, \ldots, n$, $t_k := t + k\Delta t$.

Proof. Firstly, let us note that

$$|X_r^{t,x} - \widehat{X}_{t_i}^{t,x}|^2 \le 2|X_r^{t,x} - X_{t_i}^{t,x}|^2 + 2|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^2.$$

It follows by Lemma 8.2 that it is sufficient to show that

$$\sum_{i=1}^{n} E\left[e^{-Kt_i}|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^2\right] \le C(1+|x|^2).$$

It is easy to see by Young's inequality, Hölder's inequality and Itô's isometry that

$$\begin{aligned}
&E\left[|X_{t_{i+1}}^{t,x} - \widehat{X}_{t_{i+1}}^{t,x}|^2\right] \\
&\le (1+\Delta t)E\left[|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^2\right] + \left(2 + \frac{1}{\Delta t}\right)E\left[\left|\int_{t_i}^{t_{i+1}} (b(X_r^{t,x}) - \widehat{b}(X_{t_i}^{t,x}))dr\right|^2\right] \\
&\quad +2E\left[\left|\int_{t_i}^{t_{i+1}} (\sigma(X_r^{t,x}) - \widehat{\sigma}(X_{t_i}^{t,x}))dW_r\right|^2\right] \\
&\le (1+\Delta t)E\left[|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^2\right] + L_X(3+2\Delta t)E\left[\int_{t_i}^{t_{i+1}} |X_r^{t,x} - \widehat{X}_{t_i}^{t,x}|^2 dr\right] \\
&\le (1+\Delta t\{1+(1+\epsilon)L_X(3+2\Delta t)\})E\left[|X_{t_i}^{t,x} - \widehat{X}_{t_i}^{t,x}|^2\right]
\end{aligned}$$

$$+\left(1+\frac{1}{\epsilon}\right) L_X(3+2\Delta t)E\left[\int_{t_i}^{t_{i+1}} |X_r^{t,x} - X_{t_i}^{t,x}|^2 dr\right].$$

It follows that

$$\begin{aligned}
J_{i+1} &:= e^{-K_1 t_{i+1}} E\left[|X_{t_{i+1}}^{t,x} - \widehat{X}_{t_{i+1}}^{t,x}|^2\right] \\
&\le e^{-K_1\Delta t}\left(1+\Delta t\left\{1+(1+\epsilon)L_X(3+2\Delta t)\right\}\right) J_i + CE\left[\int_{t_i}^{t_{i+1}} e^{-K_1 t_{i+1}} |X_r^{t,x} - X_{t_i}^{t,x}|^2 dr\right].
\end{aligned}$$

Now, for $x \ge 0$, $e^x \ge 1+x$ and so

$$e^{-K_1\Delta t} \le \frac{1}{1+K_1\Delta t}.$$

It follows that if we choose ϵ and Δt small enough so that

$$1+4L_X \ge 1+(1+\epsilon)L_X(3+2\Delta t)$$

and

$$K_1 > 1+4L_X$$

then

$$\begin{aligned}
J_{i+1} &\le J_i + CE\left[\int_{t_i}^{t_{i+1}} |X_r^{t,x} - X_{t_i}^{t,x}|^2 dr\right] \\
&\le J_0 + C\sum_{i=1}^{n} E\left[\int_{t_{i-1}}^{t_i} e^{-K_1 t_i} |X_r^{t,x} - X_{t_i}^{t,x}|^2 dr\right] \\
&\le C\Delta t(1+|x|^2)
\end{aligned}$$

by Lemma 8.2.

If we now choose $K = K_1 + \alpha$ for some $\alpha > 0$, then

$$\begin{aligned}
\sum_{j=1}^{n} E\left[e^{-Kt_j} |X_{t_j}^{t,x} - \widehat{X}_{t_j}^{t,x}|^2\right] &\le \sum_{j=1}^{n} e^{-\alpha j\Delta t} J_j \\
&\le C\Delta t(1+|x|^2)\sum_{j=1}^{n} e^j - \alpha\Delta t
\end{aligned}$$

$$
\begin{aligned}
&\leq C\Delta t(1+|x|^2)\sum_{j=1}^{n}(1+\alpha\Delta t)^{-j} \\
&\leq C\Delta t\frac{1}{\alpha\Delta t} \\
&\leq C
\end{aligned}
$$

as required. □

Lemma 8.4. *Let X be defined as in Definition 4.1 on page 42, (Y, Z) be the solution of the BDSDE (8.1) with contractive coefficients f, g parameterised by decay factor K as defined in Definition 4.4 on page 44 and suppose that a positive constant T is given. There exists a positive constant C and a positive value for the decay factor K both independent of T such that*

$$
E\left[\int_t^T e^{-Kr}(|f(X_r^{t,x}, Y_r^{t,x})|^2 + |g(X_r^{t,x}, Y_r^{t,x})|^2)dr\right] \leq C(1+|x|^2).
$$

Proof. We consider f only as g is identical. Now,

$$
|f(X_r^{t,x}, Y_r^{t,x})|^2 \leq C(|f(X_r^{t,x}, 0)|^2 + |Y_r^{t,x}|^2).
$$

From [50], we know that

$$
E\left[\int_t^T e^{-Kr}|Y_r^{t,x}|^2 dr\right] \leq CE\left[\int_t^T e^{-Kr}(|f(X_r^{t,x}, 0)|^2 + |g(X_r^{t,x}, 0)|^2)dr\right].
$$

The result now follows by (8.2). □

Definition 8.5. *The f and g be contractive coefficients as defined in Definition 4.4 on page 44 and (Y, Z) the solution to BDSDE (8.1). We define*

$$
\begin{aligned}
\widetilde{Y}_s^{t,x} &:= e^{-Ks}Y_s^{t,x}, \\
\widetilde{Z}_s^{t,x} &:= e^{-Ks}Z_s^{t,x}, \\
\widetilde{f}(s,x,y) &:= e^{-Ks}f(x, e^{Ks}y) + Ky, \\
\widetilde{g}(s,x,y) &:= e^{-Ks}g(x, e^{Ks}y).
\end{aligned}
$$

We still refer to K as the decay factor and μ (from Definition 4.4) as the contraction coefficient.

Using the notation of Definition 8.5, let us now rewrite (8.1) as

$$\widetilde{Y}_s^{t,x} = \int_s^T \widetilde{f}(r, X_r^{t,x}, \widetilde{Y}_r^{t,x})dr - \int_s^T \widetilde{g}(r, X_r^{t,x}, \widetilde{Y}_r^{t,x})\overleftarrow{dB}_r - \int_s^T \widetilde{Z}_r^{t,x} dW_r. \tag{8.3}$$

Remark. We note the following relations that follow directly from Definition 8.5 (where $\phi = f, g$ and without loss of generality, $s_2 > s_1$):

$$\begin{aligned}
|\widetilde{\phi}(s_1, x, y) - \widetilde{\phi}(s_2, x, y)|^2 &\leq 2e^{-2Ks_1}\left(1 - e^{-K(s_2 - s_1)}\right)^2 |\widetilde{\phi}(s_1, x, y)|^2 \\
|\widetilde{\phi}(s, x_1, y) - \widetilde{\phi}(s, x_2, y)|^2 &\leq e^{-2Ks} L|x_1 - x_2|^2, \\
|\widetilde{f}(s, x, y_1) - \widetilde{f}(s, x, y_2)|^2 &\leq 2(L + K^2)|y_1 - y_2|^2, \\
(y_1 - y_2)(\widetilde{f}(s, x, y_1) - \widetilde{f}(s, x, y_2)) &\leq (K - \mu)|y_1 - y_2|^2, \\
|\widetilde{g}(s, x, y_1) - \widetilde{g}(s, x, y_2)|^2 &\leq L_{g,y}|y_1 - y_2|^2.
\end{aligned}$$

Since, for any $x \geq 0$

$$\begin{aligned}
1 - e^{-x} &= e^{-x}(e^x - 1) \\
&= e^{-x}x\left(1 + \frac{x}{2!} + \frac{x}{3!} + \dots\right) \\
&\leq e^{-x}xe^x \\
&= x
\end{aligned}$$

it follows that

$$|\widetilde{\phi}(s_1, x, y) - \widetilde{\phi}(s_2, x, y))|^2 \leq 2e^{-2Ks_1}K^2(s_2 - s_1)^2|\widetilde{\phi}(s_1, x, y)|^2. \tag{8.4}$$

Lemma 8.6. *Let $\widetilde{Y}, \widetilde{Z}, \widetilde{f}$ and $\widetilde{g}$ be defined as in Definition 8.5 parameterised by decay factor K. Suppose that positive constants T and Δt are given and define $n := \frac{T-t}{\Delta t}$ and for $k = 0, \dots, n$, $t_k := t + k\Delta t$. There exists a positive constant C and a positive value for the decay factor K both independent of T and Δt such that*

$$\sum_{i=1}^{n} E\left[\int_{t_{i-1}}^{t_i} |\widetilde{Y}_r^{t,x} - \widetilde{Y}_{t_i}^{t,x}|^2 dr\right] \leq C\Delta t(1 + |x|^2).$$

Proof. For $s \in [t_{i-1}, t_i]$,

$$\frac{1}{3}|\widetilde{Y}_s^{t,x} - \widetilde{Y}_{t_i}^{t,x}|^2 \leq \Delta t \int_{t_{i-1}}^{t_i} |\widetilde{f}(r, X_r^{t,x}, \widetilde{Y}_r^{t,x})|^2 dr + \left| \int_s^{t_i} \widetilde{g}(r, X_r^{t,x}, \widetilde{Y}_r^{t,x}) \overleftarrow{dB}_r \right|^2 + \left| \int_s^{t_i} \widetilde{Z}_r^{t,x} dW_r \right|^2$$

and so

$$E\left[|\widetilde{Y}_s^{t,x} - \widetilde{Y}_{t_i}^{t,x}|^2\right] \leq CE\left[\int_{t_{i-1}}^{t_i} |\widetilde{f}(r, X_r^{t,x}, \widetilde{Y}_r^{t,x})|^2 + |\widetilde{g}(r, X_r^{t,x}, \widetilde{Y}_r^{t,x})|^2 + |\widetilde{Z}_r^{t,x}|^2 dr\right].$$

It follows that

$$\begin{aligned} E\left[\int_{t_{i-1}}^{t_i} |\widetilde{Y}_r^{t,x} - \widetilde{Y}_{t_i}^{t,x}|^2 dr\right] &= \int_{t_{i-1}}^{t_i} E\left[|\widetilde{Y}_r^{t,x} - \widetilde{Y}_{t_i}^{t,x}|^2\right] dr \\ &\leq \Delta t \sup_{s \in [t_{i-1}, t_i]} E\left[|\widetilde{Y}_s^{t,x} - \widetilde{Y}_{t_i}^{t,x}|^2\right] \\ &\leq C\Delta t E\left[\int_{t_{i-1}}^{t_i} |\widetilde{f}(r, X_r^{t,x}, \widetilde{Y}_r^{t,x})|^2 + |\widetilde{g}(r, X_r^{t,x}, \widetilde{Y}_r^{t,x})|^2 + |\widetilde{Z}_r^{t,x}|^2 dr\right]. \end{aligned}$$

As a consequence, it is sufficient to show that there exists a C independent of T such that

$$E\left[\int_t^T |\widetilde{f}(r, X_r^{t,x}, \widetilde{Y}_r^{t,x})|^2 + |\widetilde{g}(r, X_r^{t,x}, \widetilde{Y}_r^{t,x})|^2 + |\widetilde{Z}_r^{t,x}|^2 dr\right] \leq C.$$

The desired bound for $\widetilde{f}$ and $\widetilde{g}$ can be deduced from Lemma 8.4 and the bound for $\widetilde{Z}$ is derived in [50]. □

8.3. Discretization Scheme

Definition 8.7. *Let $\widetilde{Y}, \widetilde{Z}, \widetilde{f}$ and $\widetilde{g}$ be defined as in Definition 8.5 and $\widehat{X}$ be defined as in Definition 7.9 on page 116. We define our discretization scheme for equation (8.3) as:*

$$\widehat{Y}_T^{t,x} := 0 \ , \quad \widehat{Y}_{t_{i-1}}^{t,x} := E_{i-1}\left[\widehat{Y}_{t_i}^{t,x} + \widetilde{f}(t_i, \widehat{X}_{t_i}^{t,x}, \widehat{Y}_{t_i}^{t,x})\Delta t_i - \widetilde{g}(t_i, \widehat{X}_{t_i}^{t,x}, \widehat{Y}_{t_i}^{t,x})\Delta B_i\right]. \quad (8.5)$$

Theorem 8.8. *Let $\widetilde{Y}, \widetilde{Z}, \widetilde{f}$ and $\widetilde{g}$ be defined as in Definition 8.5 parameterised by decay factor K and contraction coefficient μ and let $\widehat{Y}$ be as defined in Definition 8.7. Suppose that $K > \frac{1}{2} + 2L_X$, $\mu > K + L_{g,y} + L$ and that a positive T is given. Then there exist positive constants C and Δt_0 independent of T such that for any $\Delta t \in (0, \Delta t_0]$*

$$\max_{0\le i\le n} E\left[|\widetilde{Y}^{t,x}_{t_i} - \widehat{Y}^{t,x}_{t_i}|^2\right] \le C\Delta t(1+|x|^2)$$

where $n := \frac{T-t}{\Delta t}$ and for $k = 0, \ldots, n$, $t_k := t + k\Delta t$.

Proof. For $i = 1, \ldots, n$, $s \in [t_{i-1}, t_i)$, we have that

$$\begin{aligned}\widetilde{Y}^{t,x}_{t_{i-1}} - \widehat{Y}^{t,x}_{t_{i-1}} = \widetilde{Y}^{t,x}_{t_i} - \widehat{Y}^{t,x}_{t_i} &+ \int_{t_{i-1}}^{t_i} \left\{\widetilde{f}(r, X^{t,x}_r, \widetilde{Y}^{t,x}_r) - \widetilde{f}(t_i, \widehat{X}^{t,x}_{t_i}, \widehat{Y}^{t,x}_{t_i})\right\} dr \\ &- \int_{t_{i-1}}^{t_i} \left\{\widetilde{g}(r, X^{t,x}_r, \widetilde{Y}^{t,x}_r) - \widetilde{g}(t_i, \widehat{X}^{t,x}_{t_i}, \widehat{Y}^{t,x}_{t_i})\right\} \overleftarrow{dB}_r - \int_{t_{i-1}}^{t_i} \widetilde{Z}^{t,x}_r dW_r.\end{aligned}$$

Instead of now applying Itô's formula as was done in Lemma 7.14 (and in [8] and [2]), we now instead follow the strategy of [49] by moving the final stochastic integral to the left-hand-side, squaring and taking expectations to give:

$$\begin{aligned}J_{i-1} &:= E\left[|\widetilde{Y}^{t,x}_{t_{i-1}} - \widehat{Y}^{t,x}_{t_{i-1}}|^2\right] \\ &\le E\left[|\widetilde{Y}^{t,x}_{t_{i-1}} - \widehat{Y}^{t,x}_{t_{i-1}}|^2\right] + E\left[\int_{t_{i-1}}^{t_i} |\widetilde{Z}^{t,x}_r|^2 dr\right] \\ &= E\left[\left|(\widetilde{Y}^{t,x}_{t_i} - \widehat{Y}^{t,x}_{t_i}) + \int_{t_{i-1}}^{t_i} \left\{\widetilde{f}(r, X^{t,x}_r, \widetilde{Y}^{t,x}_r) - \widetilde{f}(t_i, \widehat{X}^{t,x}_{t_i}, \widehat{Y}^{t,x}_{t_i})\right\} dr\right.\right. \\ &\quad \left.\left. - \int_{t_{i-1}}^{t_i} \left\{\widetilde{g}(r, X^{t,x}_r, \widetilde{Y}^{t,x}_r) - \widetilde{g}(t_i, \widehat{X}^{t,x}_{t_i}, \widehat{Y}^{t,x}_{t_i})\right\} \overleftarrow{dB}_r\right|^2\right].\end{aligned}$$

It follows by the Cauchy-Schwarz inequality, Itô's isometry, Young's inequality and since $(\widetilde{Y}^{t,x}_{t_i} - \widehat{Y}^{t,x}_{t_i})$ is $\mathcal{F}_{t_i}$-measurable that

$$\begin{aligned}J_{i-1} \le\ & E\left[|\widetilde{Y}^{t,x}_{t_i} - \widehat{Y}^{t,x}_{t_i}|^2\right] \\ &+\Delta t\left(1 + \frac{1}{\epsilon_1}\right) E\left[\int_{t_{i-1}}^{t_i} |\widetilde{f}(r, X^{t,x}_r, \widetilde{Y}^{t,x}_r) - \widetilde{f}(t_i, \widehat{X}^{t,x}_{t_i}, \widehat{Y}^{t,x}_{t_i})|^2 dr\right] \\ &+(1+\epsilon_1)E\left[\int_{t_{i-1}}^{t_i} |\widetilde{g}(r, X^{t,x}_r, \widetilde{Y}^{t,x}_r) - \widetilde{g}(t_i, \widehat{X}^{t,x}_{t_i}, \widehat{Y}^{t,x}_{t_i})|^2 dr\right]\end{aligned} \tag{8.6}$$

$$+2E\left[\int_{t_{i-1}}^{t_i}(\widetilde{Y}_{t_i}^{t,x}-\widehat{Y}_{t_i}^{t,x})(\widetilde{f}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})-\widetilde{f}(t_i,\widehat{X}_{t_i}^{t,x},\widehat{Y}_{t_i}^{t,x}))dr\right].$$

We now examine each of the final three terms in (8.6).

$\underline{|\widetilde{f}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})-\widetilde{f}(t_i,\widehat{X}_{t_i}^{t,x},\widehat{Y}_{t_i}^{t,x})|^2}$:

$$\begin{aligned}
|\widetilde{f}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})&-\widetilde{f}(t_i,\widehat{X}_{t_i}^{t,x},\widehat{Y}_{t_i}^{t,x})|\\
&\leq|\widetilde{f}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})-\widetilde{f}(t_i,X_r^{t,x},\widetilde{Y}_r^{t,x})|\\
&\quad+|\widetilde{f}(t_i,X_r^{t,x},\widetilde{Y}_r^{t,x})-\widetilde{f}(t_i,\widehat{X}_{t_i},\widetilde{Y}_r^{t,x})|\\
&\quad+|\widetilde{f}(t_i,\widehat{X}_{t_i},\widetilde{Y}_r^{t,x})-\widetilde{f}(t_i,\widehat{X}_{t_i},\widetilde{Y}_{t_i}^{t,x})|\\
&\quad+|\widetilde{f}(t_i,\widehat{X}_{t_i},\widetilde{Y}_{t_i}^{t,x})-\widetilde{f}(t_i,\widehat{X}_{t_i},\widehat{Y}_{t_i}^{t,x})|.
\end{aligned}$$

It follows from (8.4) that

$$\begin{aligned}
&|\widetilde{f}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})-\widetilde{f}(t_i,\widehat{X}_{t_i}^{t,x},\widehat{Y}_{t_i}^{t,x})|^2\\
&\qquad\leq C\Big\{e^{-2Kr}K^2(\Delta t)^2|\widetilde{f}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})|^2+e^{-2Kt_i}L|X_r^{t,x}-\widehat{X}_{t_i}^{t,x}|^2\\
&\qquad\quad+2(L+K^2)|\widetilde{Y}_r^{t,x}-\widetilde{Y}_{t_i}^{t,x}|^2+2(L+K^2)|\widetilde{Y}_{t_i}^{t,x}-\widehat{Y}_{t_i}^{t,x}|^2\Big\}.
\end{aligned}$$

$\underline{|\widetilde{g}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})-\widetilde{g}(t_i,\widehat{X}_{t_i}^{t,x},\widehat{Y}_{t_i}^{t,x})|^2}$:
Similarly,

$$\begin{aligned}
&|\widetilde{g}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})-\widetilde{g}(t_i,\widehat{X}_{t_i}^{t,x},\widehat{Y}_{t_i}^{t,x})|^2\\
&\qquad\leq\left(3+\frac{1}{\delta}\right)\Big\{e^{-2Kr}K^2(\Delta t)^2|\widetilde{g}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})|^2+e^{-Kt_i}L|X_r^{t,x}-\widehat{X}_{t_i}^{t,x}|^2\\
&\qquad\quad+L_{g,y}|\widetilde{Y}_r^{t,x}-\widetilde{Y}_{t_i}^{t,x}|^2\Big\}+(1+3\delta)L_{g,y}|\widetilde{Y}_{t_i}^{t,x}-\widehat{Y}_{t_i}^{t,x}|^2.
\end{aligned}$$

$\underline{2(\widetilde{Y}_{t_i}^{t,x}-\widehat{Y}_{t_i}^{t,x})(\widetilde{f}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})-\widetilde{f}(t_i,\widehat{X}_{t_i}^{t,x},\widehat{Y}_{t_i}^{t,x}))}$:

$$\begin{aligned}
&2(\widetilde{Y}_{t_i}^{t,x}-\widehat{Y}_{t_i}^{t,x})(\widetilde{f}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})-\widetilde{f}(t_i,\widehat{X}_{t_i}^{t,x},\widehat{Y}_{t_i}^{t,x}))\\
&=2(\widetilde{Y}_{t_i}^{t,x}-\widehat{Y}_{t_i}^{t,x})\Big\{(\widetilde{f}(r,X_r^{t,x},\widetilde{Y}_{t_i}^{t,x})-\widetilde{f}(r,X_r^{t,x},\widehat{Y}_{t_i}^{t,x}))\\
&\quad+(\widetilde{f}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})-\widetilde{f}(r,X_r^{t,x},\widetilde{Y}_{t_i}^{t,x}))+(\widetilde{f}(r,X_r^{t,x},\widehat{Y}_{t_i}^{t,x})-\widetilde{f}(t_i,\widehat{X}_{t_i}^{t,x},\widehat{Y}_{t_i}^{t,x}))\Big\}.
\end{aligned}$$

Now,

$$2(\widetilde{Y}_{t_i}^{t,x}-\widehat{Y}_{t_i}^{t,x})(\widetilde{f}(r,X_r^{t,x},\widetilde{Y}_{t_i}^{t,x})-\widetilde{f}(r,X_r^{t,x},\widehat{Y}_{t_i}^{t,x}))\leq 2(K-\mu)|\widetilde{Y}_{t_{i-1}}^{t,x}-\widehat{Y}_{t_{i-1}}^{t,x}|^2.$$

Furthermore, by Young's inequality,

$$\begin{aligned}2(\widetilde{Y}_{t_i}^{t,x}-\widehat{Y}_{t_i}^{t,x})(\widetilde{f}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})-\widetilde{f}(r,X_r^{t,x},\widetilde{Y}_{t_i}^{t,x}))\leq{}& \epsilon_2(\widetilde{Y}_{t_i}^{t,x}-\widehat{Y}_{t_i}^{t,x})^2\\ &+\frac{1}{\epsilon_2}2(L+K^2)(\widetilde{Y}_r^{t,x}-\widetilde{Y}_{t_i}^{t,x})^2.\end{aligned}$$

And finally,

$$\begin{aligned}&2(\widetilde{Y}_{t_i}^{t,x}-\widehat{Y}_{t_i}^{t,x})(\widetilde{f}(r,X_r^{t,x},\widehat{Y}_{t_i}^{t,x})-\widetilde{f}(t_i,\widehat{X}_{t_i}^{t,x},\widehat{Y}_{t_i}^{t,x}))\\ &\leq \epsilon_3(\widetilde{Y}_{t_i}^{t,x}-\widehat{Y}_{t_i}^{t,x})^2+\frac{1}{\epsilon_3}(\widetilde{f}(r,X_r^{t,x},\widehat{Y}_{t_i}^{t,x})-\widetilde{f}(t_i,\widehat{X}_{t_i}^{t,x},\widehat{Y}_{t_i}^{t,x}))^2\\ &\leq \epsilon_3(\widetilde{Y}_{t_i}^{t,x}-\widehat{Y}_{t_i}^{t,x})^2+\frac{3}{\epsilon_3}\Big[\Big\{e^{-2Kr}K^2(\Delta t)^2|\widetilde{f}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})|^2\\ &\quad+e^{-2Kt_i}L|X_r^{t,x}-\widehat{X}_{t_i}^{t,x}|^2\Big\}\Big]\end{aligned}$$

Substituting all of this back into (8.6), we have that

$$\begin{aligned}J_{i-1}\leq{}& E\left[|\widetilde{Y}_{t_i}^{t,x}-\widehat{Y}_{t_i}^{t,x}|^2\right]\{1+\Delta t(C\Delta t+(1+\epsilon_1)(1+3\delta)L_{g,y}+2K-2\mu+\epsilon_2+\epsilon_3))\}\\ &+C(\Delta t)^2E\left[\int_{t_{i-1}}^{t_i}e^{-2Kr}(|\widetilde{f}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})|^2+|\widetilde{g}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})|^2)dr\right]\\ &+Ce^{-2Kt_i}(\Delta t)^2\\ &+Ce^{-2Kt_i}E\left[\int_{t_{i-1}}^{t_i}|X_r^{t,x}-\widehat{X}_{t_i}^{t,x}|^2dr\right]\\ &+CE\left[\int_{t_{i-1}}^{t_i}|\widetilde{Y}_r^{t,x}-\widetilde{Y}_{t_i}^{t,x}|^2dr\right]\end{aligned}$$

Let us choose Δt, ϵ_1, ϵ_2, ϵ_3 and δ small enough so that

$$2\mu=C\Delta t+(1+\epsilon_1)(1+3\delta)L_{g,y}+2K+\epsilon_2+\epsilon_3+\alpha$$

for some $\alpha>0$. This gives us that

$$J_{i-1}\leq(1-\alpha\Delta t)J_i+e^{-2Kt_i}C(\Delta t)^2+A_{t_i}$$

where

$$\begin{aligned}
A_{t_i} &:= CE\left[\int_{t_{i-1}}^{t_i}(e^{-2Kt_i}|X_r^{t,x}-\widehat{X}_{t_i}^{t,x}|^2+|\widetilde{Y}_r^{t,x}-\widetilde{Y}_{t_i}^{t,x}|^2)dr\right] \\
&\quad +C(\Delta t)^2E\left[\int_{t_{i-1}}^{t_i}e^{-2Kr}(|\widetilde{f}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})|^2+|\widetilde{g}(r,X_r^{t,x},\widetilde{Y}_r^{t,x})|^2)dr\right] \\
&\leq CE\left[\int_{t_{i-1}}^{t_i}(e^{-2Kt_i}|X_r^{t,x}-\widehat{X}_{t_i}^{t,x}|^2+|\widetilde{Y}_r^{t,x}-\widetilde{Y}_{t_i}^{t,x}|^2)dr\right] \\
&\quad +C(\Delta t)^2E\left[\int_{t_{i-1}}^{t_i}e^{-4Kr}(|f(X_r^{t,x},Y_r^{t,x})|^2+|Y_r^{t,x}|^2+|g(X_r^{t,x},Y_r^{t,x})|^2)dr\right].
\end{aligned}$$

It follows that

$$\begin{aligned}
J_{i-1} &\leq J_n+Ce^{-2Kt_i}(\Delta t)^2\sum_{k=0}^{\infty}(1-\alpha\Delta t)^k+\sum_{k=0}^{n}A_{t_i} \\
&\leq C\Delta t(1+|x|^2)
\end{aligned}$$

since

$$J_n=0,$$

$$\sum_{k=0}^{\infty}(1-\alpha\Delta t)^k=\frac{1}{\alpha\Delta t},$$

and

$$\sum_{k=0}^{n}A_{t_i}\leq C\Delta t(1+|x|^2) \tag{8.7}$$

by Lemmas 8.3, 8.6 and 8.4. Since i was arbitrary in the above, it follows that

$$\max_{0\leq i\leq n}J_i\leq C\Delta t(1+|x|^2)$$

as required. □

An easy corollary of Theorem 8.8 is the following theorem which is the main result of this chapter.

Theorem 8.9. (Y,Z) *be the solution of the BDSDE (8.1) with contractive coefficients parameterised by decay factor K and contraction coefficient μ as defined in Definition*

4.4 on page 44 and let $\widehat{Y}$ *be as defined in Definition 8.7. Suppose that* $K > \frac{1}{2} + 2L_X$, $\mu > K + L_{g,y} + L$ *and that a positive* T *is given. Then there exist positive constants* C *and* Δt_0 *independent of* T *such that for any* $\Delta t \in (0, \Delta t_0]$

$$E\left[|Y_t^{t,x} - \widehat{Y}_t^{t,x}|^2\right] \leq C\Delta t(1 + |x|^2)$$

where $n := \frac{T-t}{\Delta t}$ *and for* $k = 0, \ldots, n$, $t_k := t + k\Delta t$.

8.4. Conclusion

The main result of this chapter is Theorem 8.9 which shows that we are able to approximate infinite horizon BDSDEs with contractive coefficients via a time discretization scheme. We do note, however, that this chapter is just a first look at the problem and as a consequence, the assumptions made are quite restrictive.

Remark. In light of the connection made between infinite horizon BDSDEs with contractive coeficients and SPDEs in [50], we note that the main result of this chapter can be recast as a result on the approximation of the stationary solution of SPDEs with contractive coefficients via a time discretization scheme.

9 Discussion

With Theorem 7.16 we see that we have been able to answer in the affirmative the question posed in Section 4.2.1 on the approximation of finite horizon BDSDEs via time discretization scheme. Furthermore, with Theorem 8.9 we have been able to answer in the affirmative the question posed in Section 4.2.2 on the approximation of infinite horizon BDSDEs via time discretization scheme although the conditions assumed are quite restrictive.

In the finite horizon case, if we compare our assumptions (as defined in Definition 4.3) with those of [5] we note that in [5]: $\gamma = 2$ in conditions M.1 and M.2, g is permitted to depend upon Z and condition M.4 is not required. Each of these differences has been a result of our method of proof and there is no compelling reason to believe why a modified method could not yield the weaker assumptions of [5]. As a consequence, a reasonable subsequent question would be:

- *Is it possible to approximate the solutions of BDSDEs with coefficients as defined in [5] using a time discretization scheme?*

We also note that in [50], the setting in [5] is extended to include infinite-dimensional noise and infinite horizon. As a consequence, reasonable questions would be:

- *Is it possible to approximate the solutions of BDSDEs with coefficients as defined in [50] with infinite-dimensional noise using a time discretization scheme?*
- *Is it possible to approximate the solutions of infinite horizon BDSDEs with coefficients as defined in [50] using a time discretization scheme?*

The study of BSDEs has also been extended to consider PDEs with boundary conditions as opposed to the free space formulation of this thesis (see for example [38]). A suitable question would be:

- *Is it possible to approximate the solutions of SPDEs with boundary conditions using a BDSDE approach?*

Whilst this thesis proposes a discretization scheme for the solution of BDSDEs, the actual calculation of the resulting conditional expectation is not explored. As a consequence, a reasonable question would be:

- *Is it possible to construct a numerical scheme for the solution of SPDEs based upon the discretization scheme of this thesis?*

A

Useful Results

In this section we collate some useful results. We note that we state the results in a form that is sufficient for our use - the results may hold under more general conditions in the references.

Result A.1 (**Moments of** X, [16])**.** *Let X be defined as in Definition 4.1 on page 42. For each $x \in \mathbb{R}^d$,*

$$\sup_{s\in[t,T]} \{|X_s^{t,x}| + \|\nabla X_s^{t,x}\| + \|(\nabla X_s^{t,x})^{-1}\|\} \in \bigcap_{p\geq 1} L^p(\Omega).$$

Result A.2 (**Regularity of** X, [16])**.** *Let X be defined as in Definition 4.1 on page 42. For each $p \geq 2$, there exists a constant $C > 0$ such that for any $t \leq r < s \leq T$,*

$$E\left[|X_s^{t,x} - X_r^{t,x}|^p\right] \leq C(1+|x|^p)(s-r)^{p/2} \quad and$$
$$E\left[\|(\nabla X_s^{t,x})^{-1} - (\nabla X_r^{t,x})^{-1}\|^p\right] \leq C(s-r)^{p/2}.$$

Result A.3 (**Equivalence of norms**, [5])**.** *Let X be defined as in Definition 4.1 on page 42 and let $s \in [t,T]$. If $\phi : \Omega \times \mathbb{R}^d \to \mathbb{R}$ is independent of $\mathcal{F}_{t,s}^W$ and $\phi\rho \in L^1((\Omega, \mathcal{F}, P) \times \mathbb{R}^d; \mathbb{R})$, then there exist constants $C_1, C_2 > 0$ such that*

$$C_1 E\left[\int_{\mathbb{R}^d} |\phi(x)|\rho(x)dx\right] \leq E\left[\int_{\mathbb{R}^d} |\phi(X_s^{t,x})|\rho(x)dx\right] \leq C_2 E\left[\int_{\mathbb{R}^d} |\phi(x)|\rho(x)dx\right].$$

Moreover, if $\psi : \Omega \times [t,T] \times \mathbb{R}^d \to \mathbb{R}$, $\psi(s,.)$ is independent of $\mathcal{F}_{t,s}^W$ and $\psi\rho \in$

$L^1(\Omega \times [t,T] \times \mathbb{R}^d)$, *then there exist constants* $C_1, C_2 > 0$ *such that*

$$C_1 E\left[\int_t^T \int_{\mathbb{R}^d} |\psi(r,x)|\rho(x)dxdr\right] \le E\left[\int_t^T \int_{\mathbb{R}^d} |\psi(r,X_r^{t,x})|\rho(x)dxdr\right]$$
$$\le C_2 E\left[\int_t^T \int_{\mathbb{R}^d} |\psi(r,x)|\rho(x)dxdr\right].$$

Result A.4 (**Extension of Itô's formula**, [36])**.** *For any* $n \in \mathbb{N}$, *let* $M^2([0,T];\mathbb{R}^n)$ *denote the set of n-dimensional* $\{\mathcal{F}_t\}$*-adapted processes* $\{\phi_t; t \in [0,T]\}$ *that satisfy* $E\left[\int_0^T |\phi_t|^2 dt\right] < \infty$. *Similarly, let* $S^2([0,T];\mathbb{R}^n)$ *denote the set of n-dimensional* $\{\mathcal{F}_t\}$*-adapted processes* $\{\phi_t; t \in [0,T]\}$ *that satisfy* $E\left[\sup_{0\le t\le T} |\phi_t|^2\right] < \infty$. *Let* $\alpha \in S^2([0,T];\mathbb{R}^k)$, $\beta \in M^2([0,T];\mathbb{R}^k)$, $\gamma \in M^2([0,T];\mathbb{R}^{k\times m})$, $\delta \in M^2([0,T];\mathbb{R}^{k\times d})$ *be such that for* $t \in [0,T]$

$$\alpha_t = \alpha_0 + \int_0^t \beta_s ds + \int_0^t \gamma_s \overleftarrow{dB}_s + \int_0^t \delta_s dW_s$$

where $\overleftarrow{dB}_s$ *denotes the backward Itô integral and* dW_s *the standard forward Itô integral. Then for* $\phi \in C^2(\mathbb{R}^k, \mathbb{R})$,

$$\phi(\alpha_t) = \phi(\alpha_0) + \int_0^t \langle \phi'(\alpha_s), \beta_s\rangle ds + \int_0^t \left\langle \phi'(\alpha_s), \gamma_s \overleftarrow{dB}_s \right\rangle + \int_0^t \langle \phi'(\alpha_s), \delta_s dW_s\rangle$$
$$-\frac{1}{2}\int_0^t Tr(\phi''(\alpha_s)\gamma_s\gamma_s^T)ds + \frac{1}{2}\int_0^t Tr(\phi''(\alpha_s)\delta_s\delta_s^T)ds.$$

Result A.5 (**Moments of** Y **and** Z, [36])**.** *Let* (Y,Z) *denote the solution to the BDSDE (4.1) with measurable coefficients* $\{f,g,h\}$ *as defined in Definition 4.3 on page 42. Then for each* $p \in [2,\gamma]$, *(with* γ *as in Definition 4.3) there exists a constant* $C > 0$ *such that for a.e.* $x \in \mathbb{R}^d$,

$$E\left[\sup_{t\le s\le T} |Y_s^{t,x}|^p + \left(\int_t^T |Z_r^{t,x}|^2 dr\right)^{p/2}\right]$$
$$\le CE\left[|h(X_T^{t,x})|^p + \int_t^T |f(r,X_r^{t,x},0,0)|^p + \left(\sum_{j=1}^m |g_j(r,X_r^{t,x},0)|^2\right)^{p/2} dr\right].$$

Result A.6 (**Representation of** Z, [36])**.** *Let* (Y,Z) *denote the solution to the BDSDE (4.1) with smooth coefficients as defined in Definition 5.18 on page 73. Then*

Z has an a.s. continuous version which is given by

$$Z_s^{t,x} = \nabla Y_s^{t,x} (\nabla X_s^{t,x})^{-1} \sigma(X_s^{t,x}).$$

Bibliography

[1] Allen, E. J., Novosel, S. J., and Zhang, Z. Finite element and difference approximation of some linear stochastic partial differential equations. *Stochastics and Stochastic Reports 64* (1998), 117–142.

[2] Aman, A. A numerical scheme for backward doubly stochastic differential equations. *Bernoulli 19*, 1 (2013), 93–114.

[3] Antonelli, F. Backward-forward stochastic differential equations. *The Annals of Applied Probability 3*, 3 (1993), 777–793.

[4] Ash, R. B., and Doléans-Dade, J. *Probability and Measure Theory.* Academic Press, 2000.

[5] Bally, V., and Matoussi, A. Weak solutions for spdes and backward doubly stochastic differential equations. *Journal of Theoretical Probability 14* (2001), 125–164.

[6] Barles, G., Buckdahn, R., and Pardoux, E. Backward stochastic differential equations and integral-partial differential equations. *Stochastics and Stochastic Reports 60* (1997), 57–83.

[7] Bouchard, B., and Elie, R. Discrete-time approximation of decoupled forward–backward sde with jumps. *Stochastic Processes and their Applications 118* (2008), 53–75.

[8] Bouchard, B., and Touzi, N. Discrete-time approximation and monte-carlo simulation of backward stochastic differential equations. *Stochastic Processes and their Applications 111* (2004), 175–206.

[9] Briand, P., Delyon, B., and Mémin, J. Donsker-type theorem for bsdes. *Electronic Communications in Probability 6* (2001), 1–14.

[10] Chevance, D. Numerical methods for backward stochastic differential equations. In *Numerical methods in finance*. Cambridge University Press, 1997, pp. 232–244.

[11] Cohen, S. N. Undiscounted markov chain bsdes to stopping times. *Journal of Applied Probability 51*, 1 (2014), 262–281.

[12] Da Prato, G., and Zabczyk, J. *Stochastic Equations in Infinite Dimensions.* Cambridge University Press, 2014.

[13] Delarue, F. On the existence and uniqueness of solutions to fbsdes in a non-degenerate case. *Stochastic Processes and their Applications 99* (2002), 209–286.

[14] Douglas, J., Ma, J., and Protter, P. Numerical methods for forward-backward stochastic differential equations. *The Annals of Applied Probability 6*, 3 (1996), 940–968.

[15] El Karoui, N., Kapoudjian, E., Pardoux, E., Peng, S., and Quenez, M. Reflected solutions of backward sdes and related obstacle problems for pdes. *The Annals of Applied Probability 25*, 2 (1997), 702–737.

[16] Gihman, I. I., and Skorohod, A. V. *Stochastic Differential Equations.* Springer, 1972.

[17] Gobet, E., Lemor, J., and Warin, X. A regression-based monte carlo method to solve backward stochastic differential equations. *The Annals of Applied Probability 15*, 3 (2005), 2172–2202.

[18] Gyöngy, I. Lattice approximations for stochastic quasi-linear parabolic partial differential equations driven by space-time white noise i. *Potential Analysis 9* (1998), 1–25.

[19] Gyöngy, I. Lattice approximations for stochastic quasi-linear parabolic partial differential equations driven by space-time white noise ii. *Potential Analysis 11* (1999), 1–37.

[20] Hamadene, S., and Jeanblanc, M. On the starting and stopping problem: application in reversible investments. *Mathematics of Operations Research 32*, 1 (2007), 182–192.

[21] Jentzen, A., and Kloeden, P. Taylor expansions of solutions of stochastic partial differential equations with additive noise. *The Annals of Probability 38*, 2 (2010), 532–569.

[22] Kallianpur, G., and Sundar, P. *Stochastic Analysis and Diffusion Processes*. Oxford University Press, 2014.

[23] Karatzas, I., and Shreve, S. *Brownian Motion and Stochastic Calculus*. Springer, 1998.

[24] Kloeden, P. E., and Platen, E. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992.

[25] Kobylanski, M. Backward stochastic differential equations and partial differential equations with quadratic growth. *The Annals of Probability 28*, 2 (2000), 558–602.

[26] Kunita, H. *Stochastic Flows and Stochastic Differential Equations*. Cambridge University Press, 1997.

[27] Longstaff, F., and Schwartz, E. S. Valuing american options by simulation: a simple least squares approach. *The Review of Financial Studies 14* (2001), 113–147.

[28] Ma, J., Protter, P., Martin, J. S., and Torres, S. Numerical method for backward stochastic differential equations. *The Annals of Applied Probability 12*, 1 (2002), 302–316.

[29] Ma, J., Protter, P., and Yong, J. Solving forward-backward stochastic differential equations explicitly - a four step scheme. *Probability Theory and Related Fields 98* (1994), 339–359.

[30] Millet, A., and Morien, P. On implicit and explicit discretization schemes for parabolic spdes in any dimension. *Stochastic Processes and their Applications 115* (2005), 1073–1106.

[31] Milstein, G. N., and Tretyakov, M. V. Numerical algorithms for forward-backward stochastic differential equations. *SIAM Journal on Scientific Computing 28*, 2 (2006), 561–582.

[32] Milstein, G. N., and Tretyakov, M. V. Solving parabolic stochastic partial differential equations via averaging over characteristics. *Mathematics of Computation 78*, 268 (2009), 2075–2106.

[33] Pardoux, E. Backward stochastic differential equations and viscosity solutions of systems of semilinear parabolic and elliptic pdes of second order. *Progress in Probability 42* (1998), 79–127.

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

[35] Pardoux, E., and Peng, S. Backward stochastic differential equations and quasilinear parabolic partial differential equations. *Lecture Notes in Control and Information Sciences 176* (1992), 200–217.

[36] Pardoux, E., and Peng, S. Backward doubly stochastic differential equations and systems of quasilinear spdes. *Probability Theory and Related Fields 98* (1994), 209–227.

[37] Pardoux, E., and Protter, P. A two-sided stochastic integral and its calculus. *Probability Theory and Related Fields 76* (1987), 15–49.

[38] Pardoux, E., and Zhang, S. Generalized bsdes and nonlinear neumann boundary value problems. *Probability Theory and Related Fields 110* (1998), 535–558.

[39] Peng, S. Probabilistic interpretation for systems of quasilinear parabolic partial differential equations. *Stochastics 37* (1991), 61–74.

[40] Pham, H. *Continuous-time Stochastic Control and Optimization with Financial Applications*. Springer, 2009.

[41] PROTTER, P. E. *Stochastic Integration and Differential Equations*. Springer, 2005.

[42] RICHOU, A. Numerical simulation of bsdes with drivers of quadratic growth. *The Annals of Applied Probability 21*, 5 (2011), 1933–1964.

[43] ROBINSON, J. *Infinite-Dimensional Dynamical Systems*. Cambridge University Press, 2001.

[44] SHARDLOW, T. Weak convergence of a numerical method for a stochastic heat equation. *BIT Numerical Mathematics 43* (2003), 179–193.

[45] STRAUSS, W. A. *Partial Differential Equations*. Wiley, 2008.

[46] TAYLOR, S. J. *Introduction to Measure and Integration*. Cambridge University Press, 1973.

[47] WALSH, J. B. Finite element methods for parabolic stochastic pde's. *Potential Analysis 23* (2005), 1–43.

[48] YEVIK, A., AND ZHAO, H. Numerical approximations to the stationary solutions of stochastic differential equations. *SIAM Journal on Numerical Analysis 49*, 4 (2011), 1397–1416.

[49] ZHANG, J. A numerical scheme for bsdes. *The Annals of Applied Probability 14*, 1 (2004), 459–488.

[50] ZHANG, Q., AND ZHAO, H. Stationary solutions of spdes and infinite horizon bdsdes. *Journal of Functional Analysis 252* (2007), 171–219.

[51] ZHANG, Q., AND ZHAO, H. Stationary solutions of spdes and infinite horizon bdsdes with non-lipschitz coefficients. *Journal of Differential Equations 248* (2010), 953–991.