People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

University of Saida - Dr Moulay Tahar.
Faculty: Mathematics, Computer Science and Telecommunications
Department: Mathematics
No. Assigned by the library

Dissertation presented with a view to obtaining the diploma of

Academic Master
Specialty: Stochastic Analysis, Statistics of Process and Applications

Presented by
Hafsa **L**akhache[1]

Under the supervision of

Dr. Bouanani Hafida and **Dr. Kebiri Omar**

Forward-Backward Stochastic Differential Equations in Optimal Control

---

Defended on 18/06/2026 in front of the committee

Pr. Kandouci Abdeldjebbar	University of Saida Dr. Moulay Tahar	President
Dr. Bouanani Hafida	University of Saida Dr. Moulay Tahar	Supervisor
Dr. Kebiri Omar	Brandenburg University of Technology	Co-supervisor
Dr. Mekkaoui Imene	University of Saida Dr. Moulay Tahar	Examiner

2025/2026

[1]e-mail: hafsalakhache@gmail.com

Acknowledgments

(In the Name of Allah, the Most Gracious, the Most Merciful)

I would like to express my deepest and most sincere gratitude to my supervisors:

Dr. Hafida Bouanani, Professor at the University of Saida, for her invaluable guidance, unwavering support, and precious scientific and professional advice throughout the completion of this work. Her patience, encouragement, and trust have been a constant source of inspiration during this academic journey.

Dr. Omar Kebiri, Professor at Brandenburg University of Technology Cottbus–Senftenberg, for his generous support and for the knowledge he kindly shared with me. His guidance greatly enriched both my academic experience and personal growth.

Your helpful advice and suggestions were of great assistance throughout the completion of this thesis, for which I remain deeply grateful.

I would also like to thank the distinguished members of the jury for dedicating their valuable time to reading and evaluating this work.

My sincere thanks go to **Professor Abdeljebbar Kandouci**, Professor at the University of Saida and Director of the Laboratory of Stochastic Models, Statistics and Applications, for the honor of presiding over this thesis jury and for his continuous encouragement and academic support throughout my studies.

I am equally grateful to **Dr. Imene Mekkaoui**, Professor at the University of Saida, for accepting to examine this work and for her valuable remarks, thoughtful observations, and appreciated encouragement.

I would also like to thank all members of the Laboratory of Stochastic Models, Statistics and Applications (LMSSA).

Finally, there is a quotation that particularly resonated with me in the context of this work:

"Life must be lived forward, but can only be understood backward."

— Søren Kierkegaard

Dedication

Before this achievement carried my name, it carried your sacrifices, your prayers, and your endless love.

To my beloved parents, thank you for being my strength, my support, and the reason I never gave up. No words will ever truly express the gratitude and love I hold for you.

To my brother Djeloul, for guiding me toward this path and encouraging me to believe in myself and in my abilities.

And to the memory of my beloved brother Sofiane, who supported me with endless love throughout every stage of my life. Words can never truly express the place you hold in my heart or the pain your absence has left behind. I deeply wish you were here beside me to witness this moment and share this achievement with me. Though you are no longer here, your love, your kindness, and your memory will forever remain part of who I am.

To my beloved sisters, K, N, and Kh, and to my nieces and nephews, for their love, kindness, and constant support throughout this journey.

To all those who believed in me, this achievement belongs to you as much as it belongs to me.

Abstract

This thesis presents a rigorous and comprehensive investigation of Forward-Backward Stochastic Differential Equations (FBSDEs) and their application to stochastic optimal control theory, effectively bridging abstract stochastic analysis, deterministic partial differential equations (PDEs), and numerical computing. We first construct the foundational probabilistic framework by developing necessary elements of stochastic calculus, forward Stochastic Differential Equations (SDEs), and Backward Stochastic Differential Equations (BSDEs). Building upon these components, we analyze the structural properties of coupled and decoupled FBSDE systems, establishing the existence and uniqueness of their solutions. We further explore the deep interplay between probabilistic trajectories and analytical mechanics by mapping these stochastic processes to linear and non-linear PDEs through the classical and non-linear Feynman-Kac formulas.

Leveraging this analytical foundation, we formulate the stochastic optimal control problem through two core paradigms: the local approach via the Stochastic Maximum Principle (SMP) and the global approach via Dynamic Programming and the Hamilton-Jacobi-Bellman (HJB) equation. An important theoretical aspect of this work is to illustrate the connection established in the original work between the Stochastic Maximum Principle (SMP) and the Hamilton-Jacobi-Bellman (HJB) framework, where the adjoint process corresponds to the spatial gradient of the value function. Finally, recognizing that analytical solutions to fully coupled FBSDEs are exceptionally rare, we transition from theory to computation. We implement advanced time-discretization schemes and backward numerical methods to resolve terminal-value constraints. The accuracy, convergence, and practical utility of these numerical algorithms are successfully validated through two distinct case studies: a non-linear, coupled controlled SIR epidemiological model, and a benchmark Linear-Quadratic (LQ) control problem.

Keywords: Forward-Backward Stochastic Differential Equations, Optimal Control, Stochastic Maximum Principle, Hamilton-Jacobi-Bellman Equation, Partial Differential Equations, Feynman-Kac formula.

Contents

List of Figures

List of Algorithms

Introduction

Background and Motivation

In most real-world systems, uncertainty is inherent. It introduces significant challenges in the search for optimal solutions, as phenomena arising in fields such as finance, physics, and biology require mathematical frameworks capable of handling randomness. Forward-backward stochastic differential equations (FBSDEs) provide such a framework.

While theoretical models often assume idealized conditions, real-world systems are inherently uncertain and dynamic. Whether on a high-frequency trading floor or a quadrotor drone navigating a gusty urban environment, uncertainty remains the only constant. In robotics, the challenge of optimal control goes beyond simply reaching a destination; it requires doing so efficiently while the environment continuously perturbs the system. Classical methods often struggle to balance immediate decisions with long-term objectives under such unpredictable conditions.
This is where forward-backward SDEs (FBSDEs) provide a powerful framework. They decompose the control problem into two interconnected processes: a forward SDE describing the system's noisy evolution in time, and a backward stochastic differential equation that determines the optimal control strategy by propagating information from the terminal objective back to the present. By coupling these two directions, FBSDEs establish a robust mathematical structure that anticipates uncertainty rather than merely reacting to it.
Although their applications to modern systems such as autonomous flight are recent, the mathematical foundation of FBSDEs is rooted in a century-long development of stochastic calculus, beginning with the study of FSDEs and evolving toward increasingly sophisticated tools for modeling and controlling systems under uncertainty.

The study of stochastic differential equations (SDEs) has been a cornerstone of modern probability theory since the foundational work of *K. Itô* in the 1940s [34]. Itô's introduction of stochastic calculus provided mathematicians and scientists with a rigorous framework for modeling systems driven by random noise. More broadly, the theory of stochastic integration was developed through the contributions of several seminal contributors, including *N. Wiener* [63] (1923), *A. N. Kolmogorov* [40] (1931), *W. Feller* [23] (1936), *P.Lévy* [44] (1948), and *K. Itô* [34, 35, 36] (1942, 1944, 1951). Although the list of contributions is extensive, Itô's theory remains the most influential and is widely regarded as the foundation of modern stochastic calculus. Forward stochastic differential equations describe the evolution of a stochastic process from an initial condition, like ordinary differential equations in the deterministic setting. However, many practical problems in optimal control, mathematical finance, and game theory require a different formulation in which the terminal value of the process is specified instead of the initial condition. This fundamental limitation of purely forward SDEs motivated the development of backward stochastic differential equations (BSDEs) and subsequently, forward-backward stochastic differential equations (FBSDEs).

The Backward Stochastic Differential Equations (BSDEs) were introduced in the linear case by *Bismut* [10] as the adjoint equation associated with the stochastic version of Pontryagin's maximum principle in control theory. In addition, the well-posedness result for general linear BSDEs was established by *Bensoussan* in 1983 [5], who used the martingale representation theorem to prove the existence and uniqueness of the solutions. The general nonlinear case of existence and uniqueness of the solution (under Lipschitz condition) was resolved in the seminal paper by *Pardoux and Peng* [52]. Since then, the interest for BSDEs has increased regularly, due to the connections of this subject with mathematical finance, stochastic control, and partial differential equations (PDEs), we refer the reader to *El Karoui, Peng, Quenez*[20], *El Karoui, Quenez* [21], and for the connection with viscosity solution to PDEs see *Pardoux*[51] and the references therein. Another important result *Peng* discovered that the adapted solution of BSDE admits a probabilistic interpretation of solutions to semilinear and quasilinear parabolic PDEs [56, 54]. After this, an extensive study of BSDEs was initiated, and potential for its application was found in applied and theoretical areas such as stochastic control, mathematical finance, and differential geometry.

Furthermore, the development of FBSDEs is naturally motivated by stochastic optimal control theory, particularly from the adjoint equations that appear in Pontryagin 's type maximum principle. The earliest version of such a system was introduced by *Bismut* in 1973 [9, 10], consisting of a FSDE and a linear BSDE. Subsequently, the first result for a coupled FBSDE where the coefficients of each equation depend on the solution of the other was obtained by *Antonelli* [1] in his Ph.D. thesis in 1993.

Since then, many authors have studied this system using various methods, such as the four-step scheme [45] and the method of continuation [31], as well as many other approaches [17, 46, 57]. Furthermore, in the FBSDEs theory framework, particularly, in a Markovian framework, the solution of a BSDE describes the viscosity solution of the associated semi-linear partial differential equation (PDE in short). For more details, we refer to the paper by *El Karoui, Peng and Quenez* [20] and the references therein. *Peng* [55] obtained the Hamilton-Jacobi-Bellman equation and proved that the value function is its viscosity solution. Many authors have studied the theory of BSDEs and its applications to stochastic control (see Peng [53]), finance (see El Karoui, Peng, and Quenez [20]), and for partial differential equations theory (see Peng [54], etc.). Since then, BSDEs have been widely used in stochastic control, and especially in mathematical finance.

The foundations of optimal control were laid as early as 1697 *Bernoulli's* work on variational problems, eventually evolving into the modern stochastic control framework popularized in the mid-20th century [61]. Over time, this field of study has continued to develop and has attracted the attention of many researchers, for a comprehensive historical account, we refer the reader to the end of chapter 2 in [65]. However, it is true, except that its real rise occurred in the 1950s and 1960s, driven by the work of Pontryagin and his group on the maximum principle. Since then, this theory has drawn significant interest in both pure and applied mathematics. In particular, the development of stochastic modeling in fields such as economics, finance, and engineering has highlighted the need to extend classical optimal control methods to systems driven by randomness. In such cases, the dynamics of a controlled system are typically described by SDE.

Consequently, the study of stochastic optimal control combines probabilistic techniques with methods of PDEs. Two fundamental approaches have been developed to analyze such problems, the first is the Stochastic Maximum Principle (SMP), which provides the necessary conditions for optimality and extends the classical deterministic maximum principle introduced by *Pontryagin* to include stochastic systems. There are many works on this topic [4, 11, 29, 41], including the contribution by *I. EKELAND* in 1972 and 1979, who studied this principle in cases where the diffusion σ does not contain the control variable [19]. A difficulty in treating is the case where the diffusion σ depends on the control variable u, *Bensoussan* studied such a case [4, 11], his method strongly depends on the control being convex, next *S. Peng* gives the principle in a more general framework where σ depends on u and the control set is not necessarily convex [53]. In this framework, the adjoint processes associated with the optimal control problem are described by BSDE.

The second approach is the method of dynamic programming, the basic idea of which was initiated by *R.Bellman* in the early 1950s [2]. This method characterizes the value function of the control problem as the solution of the Hamilton-Jacobi-Bellman equation, which is generally a nonlinear PDE. In many cases, the value function may fail to be smooth; therefore, the concept of viscosity solutions, developed by *Crandall and Lions* [15], is employed.

Despite the rich theoretical background, numerical methods for FBSDEs remain an active area of research because analytical solutions are rarely available. For forward SDE, the state process is typically approximated using a time discretization scheme such as the Euler-Maruyama method or the Milstein scheme [39, 60, 30]. For backward SDE, numerical approximation is usually based on a backward time discretization, such as backward Euler discretization [14]. The computation of the conditional expectations arising in the backward iteration are typically approximated using a regression technique, for example, the Least Squares Monte Carlo method [33]. Combining these numerical procedures leads to a fully discrete numerical approximation of the FBSDE system.
In the decoupled case, numerous numerical methods have been developed, including the Malliavin calculus method [16], the quantization method [50], and regression-based approaches [25]; see also [33] and the references therein for a comprehensive overview. The coupled case is more challenging due to the intertwined nature of the forward and backward equations. Notable methods include the four-step scheme-based approach [47], the Markovian iteration scheme [3], and Fourier expansion

techniques [48]. More recently, neural network-based algorithms, such as the deep BSDE method, have gained popularity due to their high accuracy and effectiveness in handling high-dimensional problems [27, 26, 28].

Dissertation Objectives

The primary objective of this dissertation is to provide a comprehensive and rigorous treatment of FBSDE in the context of optimal control theory.
The dissertation systematically develops the theoretical foundations of FBSDEs before exploring their connections to optimal control problems and numerical methods.
The contributions of this dissertation include:

1. A detailed exposition of existence and uniqueness theorems for both decoupled and coupled FBSDEs, with careful attention to the conditions required for solvability and the techniques used in the proofs.

2. A thorough exploration of the connection between FBSDEs and stochastic optimal control through both the maximum principle and dynamic programming approaches.

3. An analysis of numerical methods for FBSDEs, including discretization schemes.

Structure of the dissertation

This dissertation is structured across three chapters. It begins with an introduction that provides the theoretical framework and the project objectives. Following this,

- ***Chapter 1*** Explores the fundamental theoretical concepts required for the remainder of the thesis. It begins with a review of stochastic calculus, including the basic tools such as probability space and stochastic process, such as Brownian motion and Itô calculus. The chapter then discusses stochastic differential equations (SDEs). This is followed by an introduction to Backward stochastic differential equations (BSDEs), which play a central role in stochastic control theory. Building on these concepts, the chapter presents forward-backward stochastic differential equations (FBSDEs), including their definitions and the existence and uniqueness of the solution. Furthermore, the chapter discusses the relation between FBSDE and PDE (linear and nonlinear cases).

- ***Chapter 2*** Develops the optimal control framework using FBSDEs. The chapter begins with the formulation of the optimal control, where the stochastic control problem is introduced. It then presents the Stochastic Maximum Principle (SMP) as a fundamental tool for deriving necessary conditions for optimality. Next, it explores the Dynamic programming approach and derives the Hamilton-Jacobi-Bellman (HJB) equation. Then, it concludes with a connection between the two approaches.

- ***Chapter 3*** Focuses on numerical methods and simulations for solving FBSDEs. It starts by introducing discretization schemes for the forward equations, including Euler-Maruyama and Milstein schemes. Then, it describes the backward numerical methods for BSDEs and the approximation of conditional expectation. Next, it develops numerical algorithms for both coupled and decoupled FBSDE systems. Finally, illustrates the proposed methods through a linear-Quadratic (LQ) example and a controlled SIR epidemiological model, and numerical simulations.

Chapter 1

Forward-Backward Stochastic Differential Equations

The aim of this chapter is to present a rigorous foundation of Forward Backward Stochastic Differential Equations (FBSDEs for short). Beginning with the essential theoretical background in stochastic calculus and Stochastic differential equations (SDEs for short), we extend our analysis to Backward stochastic differential equations (BSDEs), which are uniquely defined by their terminal conditions. Finally, we study FBSDE, a powerful mathematical structure used to address complex problems in optimal control.

1.1 Stochastic Calculus

The definitions of this section are borrowed from [8], [7] and [49].

Probability space

Definition 1.1.1. ***($\sigma-$algebra) or ($\sigma-$field)***

Let Ω be a nonempty sample space, a collection of subsets $\mathcal{F}$ of Ω is called a σ-algebra if it satisfies the following properties:

1. $\Omega \in \mathcal{F}$,
2. $A \in \mathcal{F} \implies A^c \in \mathcal{F}$,
3. $A_n \in \mathcal{F},\ n \geq 1 \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

The pair $(\Omega, \mathcal{F})$ is called a measurable space, and the sets in $\mathcal{F}$ are referred to as $\mathcal{F}$-measurable sets. Furthermore, if $\mathcal{G}$ is a σ-algebra on Ω such that $\mathcal{G} \subseteq \mathcal{F}$, then $\mathcal{G}$ is called a sub-σ-algebra of $\mathcal{F}$.

Definition 1.1.2. *The Borel σ-algebra on a topological space Ω (typically $\mathbb{R}$), denoted by $\mathcal{B}(\Omega)$, is the smallest σ-algebra containing all open subsets of Ω. Any set $B \in \mathcal{B}(\Omega)$ is referred to as a Borel set. $\mathcal{B}(\mathbb{R})$ consists of all open sets, closed countable unions, and intersection sets.*

Definition 1.1.3. *A* probability measure *is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ such that:*

- $\mathbb{P}(\varnothing) = 0,\ \mathbb{P}(\Omega) = 1.$
- *If $\{A_i\}_{i\geq 1} \subset \mathcal{F}$ is a sequence of disjoint sets ($A_i \cap A_j = \varnothing \quad$ for all $i \neq j$) in $\mathcal{F}$, then:*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \quad \text{(countable additivity).}$$

The triplet $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space.

Definition 1.1.4. ***(Independence of Events)*** *A family of events $\{A_i : i \in I\}$ is independent if for every finite subset $J \subset I$,*

$$\mathbb{P}\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J} \mathbb{P}(A_i).$$

Definition 1.1.5. ***(Null Sets and Completeness)*** *If we have an event A and $\mathbb{P}(A) = 0$, then A is a $\mathbb{P}$-null event (or null set).*
The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be complete if for every set $A \in \mathcal{F}$ with $\mathbb{P}(A) = 0$, then for any $B \subset A$, it follows that $B \subset \mathcal{F}$ (consequently $\mathbb{P}(B) = 0$).

Definition 1.1.6. *Given $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, we define the collection of all subsets of $\mathbb{P}-$null sets as:*

$$\mathcal{N} := \{B \subset \Omega : \exists A \in \mathcal{F},\ \mathbb{P}(A) = 0,\ B \subset A\}.$$

The smallest σ-algebra containing both $\mathcal{F}$ and $\mathcal{N}$ is:

$$\mathcal{F}^* := \mathcal{F} \vee N.$$

For any $A^* \in \mathcal{F}^*$, there exists $A, B \in \mathcal{F}$, such that: $A^* \subseteq B$ and $\mathbb{P}(B \setminus A) = 0$. We then extend $\mathbb{P}$ to $\mathcal{F}^*$ by setting $\mathbb{P}(A^*) = \mathbb{P}(A)$. The resulting triple $(\Omega, \mathcal{F}^*, \mathbb{P})$ is a complete probability space.

Definition 1.1.7. *A* random variable *is a measurable function $X : \Omega \to \mathbb{R}$, and*

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F}.$$

A probability measure μ_X on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined by:

$$\mu_X(B) = \mathbb{P}(X^{-1}(B)), \quad \forall B \in \mathcal{B}(\mathbb{R}),$$

μ_X is called the distribution *of X.*
In the case of a random vector $X = (X_1, ..., X_n)$, the distribution μ_X is uniquely determined by the cumulative distribution function $F : \mathbb{R}^n \to [0, 1]$, defined as:

$$F(x_1, \ldots, x_n) := \mathbb{P}(X_i \leq x_i, 1 \leq i \leq\ n).$$

This function is non-decreasing, right continuous, and satisfies:

$$\lim_{x\to-\infty} F(x) = 0, \qquad \lim_{x\to\infty} F(x) = 1.$$

Definition 1.1.8. ***(Expectation)*** *Let X be a random variable, if $\int_\Omega |X(\omega)|\, dP(\omega) < \infty$, then the number*

$$\mathbb{E}[X] := \int_\Omega X(\omega)\, dP(\omega),$$

is called the expectation *of X.*
More generally, if $g : \mathbb{R}^m \to \mathbb{R}$ is a Borel measurable function and

$$\int_\Omega |g(X(\omega))|\, dP(\omega) < \infty,$$

then the expectation of $g(X)$ is defined by:

$$\mathbb{E}[g(X)] := \int_\Omega g(X(\omega))\, dP(\omega),$$

using the distribution μ_X, then the expectation of $g(X)$ is defined by:

$$\mathbb{E}[g(X)] := \int_\Omega g(X(\omega))\, dP(\omega) = \int_{\mathbb{R}^m} g(X(\omega))\, d\mu_X(x).$$

If two random variables $X,\ Y : \Omega \to \mathbb{R}$ are independent, then:

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y],$$

provided that $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[|Y|] < \infty$.

In this part, we define four essential modes of convergence for a sequence of *random variables* (X_n).

Definition 1.1.9.

- ***Almost sure convergence:***

$X_n \to X$ *a.s.* *if* $\mathbb{P}(\omega : X_n(\omega) \to X(\omega)) = 1$.

- ***Convergence in probability:***

$$X_n \xrightarrow{\mathbb{P}} X \quad \Longleftrightarrow \quad \forall \varepsilon > 0,\ \lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$

Note that:(Almost sure convergence $\Longrightarrow$ convergence in probability)

- ***Convergence in $\mathcal{L}^p$:***

$$X_n \xrightarrow{\mathcal{L}^p} X, \Longleftrightarrow \quad \textit{if}: \lim_{n\to\infty} \mathbb{E}[||X_n - X||^p] = 0, \quad p \geq 1.$$

- ***Weak convergence:***

$X_n \to X, \Longleftrightarrow$ *if*: $\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$ *as* $n \to \infty$ *(or converge in distribution)*

Theorem 1.1.1. *The various modes of convergence are connected in a specific order, and this relationship can be illustrated as:*

$$X_n \xrightarrow{a.s.} X \quad \Longrightarrow \quad X_n \xrightarrow{\mathbb{P}} X \quad \Longrightarrow \quad X_n \xrightarrow{d} X.$$

Definition 1.1.10. ***(Conditional expectation)*** *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space,* $\mathcal{G}$ *be a sub-*σ*-algebra and a random variable* $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$*,*[1] *we say that a random variable* Y *is the* conditional expectation *of* X *with respect to* $\mathcal{G}$*, denoted by* $\mathbb{E}[X \mid \mathcal{G}]$*, if the following conditions are satisfied:*

1. *Y is $\mathcal{G}$-measurable.*
2. *for all $G \in \mathcal{G}$:*

$$\int_G Y d\mathbb{P} = \int_G X\, d\mathbb{P}.$$

Note that the expectation of X, denoted by $\mathbb{E}[X]$ is a number, while the conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$ is a random variable.
The existence and uniqueness of $\mathbb{E}[X \mid \mathcal{G}]$ are guaranteed by the Radon-Nikodym Theorem [8].
We now summarize the essential properties of condition expectations.

Proposition 1.1.1. *[18] Let* X, Y *be two random variables on the probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *taking their values in* $\mathbb{R}$*. Let* $\mathcal{G} \subset \mathcal{F}$ *be a sub-*σ*-algebra. Then:*

1. ***Measurability property.*** *If X is $\mathcal{G}$-measurable, then*

$$\mathbb{E}[X \mid \mathcal{G}] = X.$$

2. ***Independence property.*** *If X and $\mathcal{G}$ are independent, then*

$$\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X].$$

3. *If Y is $\mathcal{G}$-measurable and $\mathbb{E}[|XY|] < \infty$, then for any random variable such that XY is integrable, we have:*

$$\mathbb{E}[XY \mid \mathcal{G}] = Y\, \mathbb{E}[X \mid \mathcal{G}].$$

4. ***Tower property.*** *If $\mathcal{H} \subset \mathcal{G}$ is another sub-σ-algebra, then*

$$\mathbb{E}[X \mid \mathcal{H}] = \mathbb{E}[\, \mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}\,].$$

5. ***Monotonicity.*** *If $X \leq Y$, then*

$$\mathbb{E}[X \mid \mathcal{G}] \leq \mathbb{E}[Y \mid \mathcal{G}].$$

6. ***The low of total expectation***

$$\mathbb{E}\big[\mathbb{E}[X \mid \mathcal{G}]\big] = \mathbb{E}[X].$$

[1] The set of all integrable random variables, which are $\mathcal{F}$ measurable, is defined as:

$$L^1(\Omega, \mathcal{F}, \mathbb{P}) = \big\{X : \Omega \to \mathbb{R} \mid \mathbb{E}[|X|] = \int_\Omega X(\omega)\, dP(\omega) < \infty\big\}.$$

7. ***Linearity:*** *For all* $a, b \in \mathbb{R}$,

$$\mathbb{E}[aX + bY \mid \mathcal{G}] = a\,\mathbb{E}[X \mid \mathcal{G}] + b\,\mathbb{E}[Y \mid \mathcal{G}].$$

Proposition 1.1.2. *[18]* ***(Jensen's Inequality)*** *Let* $\phi : \mathbb{R} \to \mathbb{R}$ *be a convex function such that:* $\mathbb{E}[\phi(X)] < \infty$, *and* $\mathbb{E}[X] < \infty$, *then:*

$$\phi(\mathbb{E}[X \mid \mathcal{G}]) \leq \mathbb{E}[\phi(X) \mid \mathcal{G}] \quad a.s.$$

Corollary 1.1.1. *[18] Taking* $\phi(.) = |.|^p$, *for any* $p \geq 1$, *provided that* $\mathbb{E}[|X|^p]$ *exists, we have:*

$$|\mathbb{E}[X \mid \mathcal{G}]|^p \leq \mathbb{E}[|X|^p \mid \mathcal{G}].$$

Stochastic Process

Definition 1.1.11. *A* filtration *on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *is an increasing family of sub-*σ*-algebras* $\{\mathcal{F}_t\}_{t\geq 0}$ *such that:*

$$\mathcal{F}_s \subseteq \mathcal{F}_t, \qquad s \leq t,$$

representing information available up to time t.
The probability space equipped with a filtration is written as follows $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ *is called a* filtered probability space.
• *The* canonical filtration *of* X *is the smallest* σ*-field under which* X_s *is measurable for all* $0 \leq s \leq t$ *defined by:*

$$\mathcal{F}_t^X = \sigma(X_s, 0 \leq s \leq t), \quad \forall t \in [0, T],$$

is called the natural filtration.
Furthermore, a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ *is said to satisfy the usual condition if:*

- $(\Omega, \mathcal{F}, \mathbb{P})$ *is a complete probability space and* $\mathcal{F}_0$ *contains all* $\mathbb{P}$*-null sets.*
- *We say that the* filtration $(\mathcal{F}_t)_{t\geq 0}$ *is right continuous if* $\mathcal{F}_t = \mathcal{F}_{t^+} := \bigcap_{s>t} \mathcal{F}_s$ *for any* $t \in [0, T)$.

Definition 1.1.12. *A* stochastic process *is a family of random variables*

$$X = \{X_t : t \in T\},$$

defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *and assuming values in* $\mathbb{R}^n$.
The parameter space T *is usually* $\mathbb{R}^+$. *Note that for each fixed* $t \in T$, *we have a random variable* $\omega \to X_t(\omega);\quad \omega \in \Omega$. *On the other hand, fixing* $\omega \in \Omega$ *we can consider the function*

$$t \to X_t(\omega); \quad t \in T,$$

which is called a sample path *(realization) of the process* X_t. *Sometimes it is convenient to write* $X(t, \omega)$ *instead of* $X_t(\omega)$. *Thus, we may also regard the process as a function of two variables* $(t, \omega) \to X(t, \omega)$ *from* $T \times \Omega$ *to* $\mathbb{R}^n$.

The Conditional Dominated Convergence Theorem is a cornerstone for the analysis of FBSDEs. It allows for the interchanging of the limit and the conditional expectation operator $E[\cdot|\mathcal{F}_t]$, provided the sequence of processes is bounded by an integrable process $Y \in L^1(\Omega)$.

Theorem 1.1.2. *[18] **(Conditional Dominated Convergence Theorem)** Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables such that $X_n \to X$ almost surely. If there exists an integrable random variable Y such that $|X_n| \leq Y$ a.s for all n, then:*

$$\mathbb{E}[X_n \mid \mathcal{F}] \xrightarrow{n\to+\infty} \mathbb{E}[X \mid \mathcal{F}].$$

Definition 1.1.13. *Let $X = \{X_t\}_{t\in[0,T]}$ and $Y = \{Y_t\}_{t\in[0,T]}$ be two stochastic processes defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We distinguish between the following concepts:*

- ***Equivalent:*** *If they have the same finite-dimensional distributions.*
- ***Modification:*** *They are modifications if for every t,*

$$\mathbb{P}(X_t = Y_t) = 1.$$

- ***Indistinguishable:*** *They are indistinguishable if*

$$\mathbb{P}\big(X_t = Y_t,\ \forall t\big) = 1.$$

Definitions: Let $(X_t)_{t\geq 0}$ be a stochastic process taking values in the n-dimensional real space R^n :

- **Measurable process:** X_t is said to be measurable if the joint map from the time sample space to the value space is measurable with respect to the product σ-algebra; implies that the map $(t,\omega) \to X_t(\omega)$ from $[0,T]\times\Omega$ to $\mathbb{R}^n$ must be $(\mathcal{B}([0,T])\otimes\mathcal{F})/\mathcal{B}(\mathbb{R}^n)$-measurable.
- **Adapted:** We say that the process X_t is adapted with respect to filtration $(\mathcal{F})_{t\geq 0}$ if $\forall t \in [0,T]$ the map (random variable) $\omega \to X_t(\omega)$ is $\mathcal{F}_t/\mathcal{B}(\mathbb{R}^n)$-measurable.
- **Progressively measurable:** For all $t \in [0,T]$, The map $(s,\omega) \to X_s(\omega)$ is measurable on $\mathcal{B}([0,t])\otimes\mathcal{F}_t/\mathcal{B}(\mathbb{R}^n)$.
  Note that : a *progressively measurable* process is both adapted and measurable

$$\textbf{Progressively measurable} \implies \begin{cases} \textbf{Adapted}, \\ \textbf{measurable}. \end{cases}$$

Now let us define three distinct forms of continuity for a *stochastic process* X_t:

- X_t is said to be *a.s continuous* if the *path* $t \to X_t(\omega)$ is a continuous function.
- X_t is said to be *a.s* right continuous with left Limits (RCLL) if (almost) all its trajectories are RCLL.
- X_t is said to be *stochastic continuous* if for all time $t \geq 0$ and for all small number $\varepsilon > 0$,

$$\lim_{s\to t} \mathbb{P}[|X_t - X_s| > \varepsilon] = 0.$$

Definition 1.1.14. *A* random variable $\tau : \Omega \to [0, \infty]$ *is a* stopping time *if :*

$$\{\tau \leq t\} \in \mathcal{F}_t, \qquad \forall t \geq 0.$$

A stopped process $X_\tau(\omega)$ *is defined by:*

$$X_\tau(\omega) = X_{\tau \wedge t}(\omega) = \begin{cases} X_\tau & \text{if} \quad \{\tau \leq t\}, \\ X_t & \text{if} \quad \{\tau > t\}. \end{cases}$$

Definition 1.1.15. ***(Brownian motion)*** *A stochastic process* $B = (B_t, t \geq 0)$ *is a* Brownian motion *if it satisfies the following properties:*

(i) Start at the origin: $B_0 = 0 \quad a.s.$

(ii) Independent increments: For any times $0 \leq t_1 \leq t_2 < ... < t_n$ *the increments* $B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2}, ..., B_{t_n} - B_{t_{n-1}}$ *are independent random variables,*

(iii) For all $t \geq s \geq 0, \quad B_t - B_s \sim \mathcal{N}(0, t-s)$, $\mathbb{E}[B_t] = 0$, *and* $Var(B_t) = t$.

(iv) The function $t \to B_t$ *is continuous almost surely (continuous trajectories).*

Remark 1.1.1. *The Brownian motion can also be defined as a centered Gaussian process for which*

$$\mathbb{E}[B_s B_t] = min(s, t) = s \wedge t, \qquad s, t \geq 0.$$

Remark 1.1.2. *Throughout this Master's project, we typically work with natural filtration generated by the Brownian motion, denoted by* $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ *defined as:*

$$\mathcal{F}_t = \sigma(B_s : 0 \leq s \leq t) \vee \mathcal{N}.$$

Where $\mathcal{N}$ *is the set of all* $\mathbb{P}$*-null sets.*
The n*-dimensional Brownian motion is defined by:* $(B_t)_{t \geq 0} = (B_t^1, ..., B_t^n)$.

Proposition 1.1.3. *[49] If* $f : \mathbb{R} \to \mathbb{R}$

$$\mathbb{E}[f(B_t)] = \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} f(x) e^{-\frac{x^2}{2t}} \, dx.$$

For example: choosing $f(u) = u^2$, *then:* $\mathbb{E}[B_t^2] = t$.

Definition 1.1.16. *Let* X *be a continuous stochastic process. Consider a sequence of subdivisions* $\Delta_n = (0 = t_0 < t_1 < ... < t_n = t)$ *of the interval* $[0, t]$. *We define the infinitesimal variation of order* p $(p > 0)$ *as:*

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

If $V_t^p(X_t)$ *admits a limit when* $\Delta_n \to 0 \quad (as \quad n \to \infty)$ *and this limit is independent of the choice of subdivision, it is called the variation of order* p *on* $[0, t]$, *denoted by* $\langle X, X \rangle_t^p$.
In particular, the most important instances of the p^{th} *order variation are:*

1. $p = 1$ *is called the* total variation *of* X_t.

2. $p = 2$ *is called the* quadratic variation *and we denote it by* $\langle X, X \rangle_t$.

Example: For Brownian motion $B_t \in \mathbb{R}$ the *quadratic variation* is:

$$\langle B, B \rangle_t = t.$$

Definition 1.1.17. *Let* $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ *be a filtered probability space. A process* $X = (X_t)_{t \geq 0}$ *is said to be:*

- *a* martingale *with respect to the filtration* $(\mathcal{F}_t)_{t\geq 0}$ *if:*

  1. X_t *is integrable for each* $t \geq 0$*;*
  2. $(X_t)_{t\geq 0}$ *is adapted to the filtration* $(\mathcal{F}_t)_{t\geq 0}$*;*
  3. $X_s = \mathbb{E}[\, X_t \mid \mathcal{F}_s \,]$ *for all* $s \leq t$.

- *a* submartingale *if* $\mathbb{E}[X_t \mid \mathcal{F}_s] \geq X_s$ *for* $s \leq t$*,*

Definition 1.1.18. *Let* $M = (M_t)_{t\geq 0}$ *be a continuous adapted process, we call it a* continuous locale martingale *if there exists an increasing sequence of stopping time* $(\tau_n)_{n\geq 0}$ *such that* $\tau_n \to \infty \quad (a.s.)$ *and, for every* n*, the stopped process* X^{τ_n} *(defined as* $X_{\tau_n}\mathbf{1}_{\{\tau_n > 0\}}$ *) is a* martingale*. We say that the sequence of stopping times* τ_n *reduces* M*.*

Definition 1.1.19. *A stochastic process* $X = \{X_t, t \geq 0\}$ *is called* Semimartingale*. If it can be written in the following form:*

$$X_t = X_0 + M_t + A_t,$$

where M_t *is a* continuous locale martingale *and* A_t *is a process of* finite variation *with* $M_0 = A_0 = 0$*, and* X_0 *is the initial condition.*
This is the most important class of process in Itô's calculus.

Proposition 1.1.4. *[58] Let* $(B_t)_{t\geq 0}$ *be a Brownian motion on the probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, then:*

1. $(B_t)_{t\geq 0}$ *is* martingale*.*

2. $(B_t^2 - t)_{t\geq 0}$ *is* martingale*.*

The following theorem is fundamental to the theory of BSDEs.

Theorem 1.1. *[37] Let* $B = \{B_t = (B_t^1, .., B_t^n,\ \mathcal{F}_t;\ 0 \leq t \leq \infty)\}$ *be a* n*-dimensional Brownian motion on* $(\Omega, \mathcal{F}, \mathbb{P})$*, and let* $\{\mathcal{F}_t\}$ *be the augmentation under* $\mathbb{P}$ *of the filtration* $\{\mathcal{F}_t^B\}$ *generated by* B*, then for any square-integrable martingale* $M = \{M_t, \mathcal{F}_t;\ 0 \leq t \leq \infty\}$ *with* $M_0 = 0$ *and RCLL paths a.s., there exists a progressively measurable process* $Y^j = \{Y^j, \mathcal{F}_t; 0 \leq t \leq \infty\}$*, such that:*

$$\mathbb{E}\int_0^T (Y_t^j)^2 \ dt < \infty; \quad 1 \leq j \leq n,$$

for every $0 < T < \infty$ *and:*

$$M_t = \sum_{j=1}^{n} \int_0^t Y_s^j \ dB_s^j; \quad 0 \leq t \leq \infty.$$

Remark 1.1.3. *An important example: we are given a random variable ξ on $(\Omega, \mathcal{F})$ that is integrable $(\mathbb{E}[\xi] < \infty)$; then the process defined by:*

$$X_t = \mathbb{E}[\xi \mid \mathcal{F}_t], \quad t \in T,$$

is a martingale*.*

Theorem 1.2. *[43] **(Doob's inequality)** If X_t is a right continuous martingale, then for all $p > 1,\ t \geq 0$:*

$$(\mathbb{E}[|\sup_{0\leq s\leq t} X_s|^p]) \leq \left(\frac{p}{p-1}\right)^p \sup_{0\leq s\leq t} (\mathbb{E}[|X_s|^p]).$$

Furthermore, let M_t be a continuous, square-integrable martingale with respect to a filtration $(\mathcal{F}_t)_{t\geq 0}$ with $M_0 = 0$ a.s. Then:

- $\mathbb{E}[\sup_{0\leq s\leq t} |M_s| \geq \lambda] \leq \frac{1}{\lambda}\mathbb{E}[|M_t|], \quad \forall t > 0, \quad \lambda > 0.$
- $\mathbb{E}(\sup_{0\leq s\leq t} |M_s|^2) \leq 4\mathbb{E}(|M_t|^2).$

Theorem 1.3. *[42]**Burkhoder-Davis-Gundy (BDG) inequality** For every $p > 0$, there exist positive constants $c_p,\ C_p > 0$, for any continuous martingale M with $M_0 = 0$:*

$$c_p\mathbb{E}[\langle M, M\rangle_\infty^{p/2}] \leq \mathbb{E}[\sup_{t\geq 0} |M_t|^p] \leq C_p\mathbb{E}[\langle M, M\rangle_\infty^{p/2}].$$

Lemma 1.1. *[42] **Gronwall** Let $T > 0$, and let g be a measurable nonnegative, bounded function on $[0, T]$, for all $t \in [0, T]$. Suppose there exist constants $a \geq 0$ and $b \geq 0$, such that:*

$$g(t) \leq a + b\int_0^t g(s)ds.$$

Then,

$$g(t) \leq a \quad exp(bt), \quad forall \quad t \in [0, T].$$

Itô Formula: The integration with respect to the time increment dt is defined through the classical Riemann or Lebesgue integral. In stochastic calculus, one must integrate an adapted stochastic process with respect to the random increments of a Brownian motion dB_t. This operation is known as the Itô integral.

In this subsection, we are interested in the stochastic integral with respect to Brownian motion.

Definition 1.1.20. *Let $B = (B_t)_{t\geq 0}$ be a Brownian motion, and let $X = (X_t)_{t\geq 0}$ be an adapted process on the probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ satisfying the L^2 integrability condition.*
The Itô's integral of X with respect to B is the process $I = (I_t)_{t\geq 0}$:

$$I = \int_0^t X_s dB_s, \quad t \in [0, T].$$

Lemma 1.2. *[6] Let $u(t,x) \in \mathcal{C}^{1,2}([0,\infty)\times\mathbb{R};\mathbb{R})$ be a function that is once continuously differentiable in t and twice continuously differentiable in x. Let X_t be a stochastic process defined by:*

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dB_s,$$

where μ_s and σ_s are adapted processes. Then the process $Y = (u(t,X_t))_{t\geq 0}$ is a semimartingale satisfying:

$$Y_t = Y_0 + \int_0^t \frac{\partial u}{\partial s}(s,X_s)ds + \int_0^t \frac{\partial u}{\partial x}(s,X_s)\mu_s ds + \int_0^t \frac{\partial u}{\partial x}(s,X_s)\sigma_s dB_s + \frac{1}{2}\int_0^t \frac{\partial^2 u}{\partial x^2}(s,X_s)\sigma_s^2 ds. \quad (1.1)$$

Examples:

1. For example, taking $X_t = B_t$ and $f \in \mathcal{C}^2(\mathbb{R},\mathbb{R})$ we have:

$$f(B_t) = f(0) + \int_0^t f'(B_s)\, dB_s + \tfrac{1}{2}\int_0^t f''(B_s)\, ds. \quad (1.2)$$

2. Let $(X_t)_{t\geq 0} = (X_t^1, ..., X_t^p)$ be a vector of continuous semimartingales and $F : \mathbb{R}^p \to \mathbb{R}$ be of class $\mathcal{C}^2$, then for every $t \geq 0$

$$F(X_t) = F(X_0) + \sum_{i=1}^p \int_0^t \frac{\partial F}{\partial x_i}(X_s)\, dX_s^i + \tfrac{1}{2}\sum_{i,j=1}^p \int_0^t \frac{\partial^2 F}{\partial x_i \partial x_j}(X_s)\, d\langle X^i, X^j\rangle_s. \quad (1.3)$$

1.2 Stochastic Differential Equations

In this section, we recall some elements of the theory of stochastic differential equations with random coefficients driven by Brownian motion.

Definition 1.2.1. *Let consider $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$ be a filtered probability space satisfying the usual condition, and let:*

- *Drift coefficient $\mu(t,x) = (\mu^i(s,X_s))_{1\leq i\leq d}$ is a measurable vector in $\mathbb{R}^d$.*
- *Diffusion coefficient $\sigma(t,x) = (\sigma^{ij}(s,X_s))_{1\leq i\leq d 1\leq j\leq n}$ is a measurable $d\times n$ matrix.*
- *$B = (B^1, ..., B^n)$ is a standard n-dimensional Brownian motion.*

Consider the following SDE:

$$dX_t = \mu(t,X_t)\, dt + \sigma(t,X_t)\, dB_t. \quad (1.4)$$

The notation 1.4 is symbolic because dB_t does not truly have a meaning (Brownian motion is not differentiable). It should be written in the form:

$$X_t = X_0 + \int_0^t \mu(s,X_s)\, ds + \int_0^t \sigma(s,X_s)\, dB_s. \quad (1.5)$$

X_0 is the initial condition, (1.5) is well-defined.
Which the SDE (1.4) can be written component-wise as:

$$dX_t^i = \mu_i(t, X_t)\, dt + \sum_{j=1}^{n} \sigma_{ij}(t, X_t)\, dB_t^j, \qquad 1 \le i \le d, \tag{1.6}$$

equivalently:

$$X_t^i = X_0^i + \int_0^t \mu^i(s, X_s)\, ds + \sum_{j=1}^{n} \int_0^t \sigma^{ij}(s, X_s)\, dB_s^j, \quad 1 \le i \le d, \tag{1.7}$$

The existence and uniqueness of the solution

Definition 1.2.2. *[37]**(Strong solution)** A process $X = (X_t)_{t\geq 0}$ is called a* strong solution *of* (1.4) *if for a fixed probability space and a fixed Brownian motion, there exists a process X satisfying:*

1. *X_t is $\mathcal{F}_t$-adapted and continuous.*
2. *$\mathbb{P}\big(\int_0^t |\mu(s, X_s)| + |\sigma(s, X_s)|^2\, ds < \infty \big) = 1$.*
3. *$X_t = x_0 + \int_0^t \mu(s, X_s)\ ds + \int_0^t \sigma(s, X_s) dB_s, \quad \mathbb{P} - a.s.$*

Definition 1.2.3. ***(Weak solution)*** *A SDE is said to admit a weak solution if there exists a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, a Brownian motion B, and a continuous process X that satisfy* (1.5) *almost surely. Therefore, a weak solution is a collection of objects $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, X)$.*

The existence (and uniqueness) of a strong solution to the (1.4) is ensured by the following hypothesis:
Hypothesis: For a constant $K > 0$, and for all $t \geq 0$, $x, y \in \mathbb{R}^d$

(H1)

$$||\mu(t, x) - \mu(t, y)|| + ||\sigma(t, x) - \sigma(t, y)|| \le K||x - y||, \tag{Lipschitz condition}$$

(H2)

$$||\mu(t, x)||^2 + ||\sigma(t, x)||^2 \le K^2(1 + ||x||^2). \tag{Linear growth condition}$$

Let ξ be a random vector valued in $\mathbb{R}^d$ with finite second moment:

$$\mathbb{E}||\xi||^2 < \infty.$$

Theorem 1.2.1. *[37] Under assumptions **(H1)-(H2)**, the SDE* (1.4) *admits a unique strong solution with initial condition ξ. Moreover, this process is square-integrable, such that:*

$$\mathbb{E}||X_t||^2 \le C(1 + \mathbb{E}||\xi||^2)e^{ct}; \quad 0 \le t \le T. \tag{1.8}$$

for every $T > 0$, there exists a constant C, depending only on K and T.

Proof. The SDE to be solved is:

$$X_t = X_0 + \int_0^t \mu(s, X_s) ds + \int_0^t \sigma(s, X_s) dB_s.$$

(i) **Uniqueness proof:** Suppose $X = (X_t)_{t\in[0,T]}$ and $Y = (Y_t)_{t\in[0,T]}$ are two solution, with $X_0 = Y_0 = \xi$ by applying the inequality $(a+b)^2 \leq 2a^2 + 2b^2$ to the integral form, we have:

$$||X_t - Y_t||^2 \leq 2 \, || \int_0^t [\mu(s, X_s) - \mu(s, Y_s)] \; ds \; ||^2 + 2 \, || \int_0^t [\sigma(s, X_s) - \sigma(s, Y_s)] \; dB_s \; ||^2,$$

by taking the expectation and applying Cauchy-Schwarz (for the drift) and Itô Isometry / BDG inequality (for the diffusion), we obtain:

$$\mathbb{E}[||X_t - Y_t||^2] \leq 2T \; \mathbb{E} \int_0^t ||\mu(s, X_s) - \mu(s, Y_s)||^2 ds + 2 \; \mathbb{E} \int_0^t ||\sigma(s, X_s) - \sigma(s, Y_s)||^2 ds.$$

Using the (1.2):

$$\mathbb{E}[||X_t - Y_t||^2] \leq C \int_0^t \mathbb{E}[||X_t - Y_t||^2] \; ds,$$

where $C = \max(2TK^2, 2K^2)$. By Gronwall's Lemma (1.1), we have: $\mathbb{E}[||X_t - Y_t||^2] = 0$ for all t.
Since the paths are continuous, this implies

$$\mathbb{P}\{ \sup_{t\in[0,T]} |X_t - Y_t| = 0\} = 1$$

(ii) **Existence proof (Picard iterations):** Let us define a sequence of approximations (x_t^n) as follows:

$$X_t^n = \xi + \int_0^t \mu(s, X_s^{n-1}) \; ds + \int_0^t \sigma(s, X_s^{n-1}) \; dB_s.$$

Using the same technique as in the uniqueness proof, we can show by induction that:

$$\mathbb{E}[||X_t^{n+1} - X_t^n||^2] \leq C \int_0^t \mathbb{E}[||X_s^n - X_s^{n-1}||^2] \; ds \implies \mathbb{E}[|X_t^{n+1} - X_t^n|^2] \leq \frac{(MT)^{n+1}}{(n+1)!}.$$

This shows that X^n is a Cauchy sequence in $L^2(\Omega)$. It converges to a limit process X_t, which satisfies the integral SDE (1.5) by passing to the limit as $n \to \infty$.

(iii) The bound $\mathbb{E}||X_t||^2 \leq C(1 + \mathbb{E}||\xi||^2)e^{ct}$ is derived by applying Itô formula to $||X_t||^2$ and applying Gronwall's Lemma once more.

□

1.3 Backward Stochastic Differential Equations

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space and a random variable ξ is $\mathcal{F}_T$-measurable ($\xi \in \mathcal{F}_T$), we consider the following differential equation:

$$-\frac{dY_t}{dt} = f(Y_t), \quad t \in [0, T], \qquad Y_T = \xi. \tag{1.9}$$

Consider the simplest case where ($f = 0$), it is impossible to find an $\mathcal{F}_t-$ adapted solution Y, then the only solution is $Y_t = \xi$, which is not adapted (because ($\xi \in \mathcal{F}_T$)). Then, the natural way to make the solution be $\mathcal{F}_t-$adapted is the martingale :

$$Y_t = \mathbb{E}(\xi \mid \mathcal{F}_t).$$

The Martingale representation Theorem (1.1) allows us to construct a unique, square integrable and adapted process $Z = (Z_t)_{t\in[0,T]}$ such that:

$$Y_t = \mathbb{E}(\xi \mid \mathcal{F}_t) = \mathbb{E}[\xi] + \int_0^t Z_r dB_r,$$

by setting $t = T$, we have $Y_T = \mathbb{E}(\xi \mid \mathcal{F}_T) = \xi$, then:

$$Y_t = \xi - \int_t^T Z_r dB_r,$$

which implies

$$-dY_t = -Z_t dB_t, \quad \text{with} \qquad Y_T = \xi. \tag{1.10}$$

Consequently, the full BSDE is formulated by allowing the function f to depend on Y_t and Z_t:

$$\begin{cases} -dY_t = f(t, Y_t, Z_t)dt - Z_t dB_t, \\ Y_T = \xi, \end{cases} \tag{1.11}$$

which is equivalent to:

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s dB_s. \tag{1.12}$$

We borrow the following notation from [52]:
$\mathcal{S}^2(\mathbb{R}^k)$ is the space of progressively measurable processes Y such that:

$$||Y||^2_{\mathcal{S}^2} = \mathbb{E}\left[\sup_{0\le t\le T} |Y_t|^2\right] < \infty.$$

$$\mathcal{S}^2_c(\mathbb{R}^k) : \text{subspace of continuous process.}$$

$\mathcal{M}^2(\mathbb{R}^{k\times d})$ is the space of progressively measurable Z such that:

$$\|Z\|_{\mathcal{M}^2} = \mathbb{E}\left[\int_0^T \|Z_t\|^2 \, dt\right] < \infty,$$

where $\|z\|^2 = \text{trace}(zz^*)$ for $z \in \mathbb{R}^{k\times d}$, and $M^2(\mathbb{R}^{k\times d})$ denotes the $\mathcal{M}^2(\mathbb{R}^{k\times d})$ equivalent classes.

Remark 1.3.1. *For a matrix $z \in \mathbb{R}^{k\times d}$, the notation $\|z\|^2 = \text{trace}(zz^\top)$ defines the norm. If z^* is the transpose of z, then:*

$$\|z\|^2 = \sum_{i=1}^{k}\sum_{j=1}^{d} |z_{i,j}|^2.$$

Definition 1.3.1. *[13] A pair process $(Y_t, Z_t)_{0\le t\le T}$ is called a solution of the BSDE (1.12) if:*

1. *Y and Z are progressively measurable with values in $\mathbb{R}^k$ and $\mathbb{R}^{k\times d}$ respectively.*
2. $$\int_0^T (|f(r, Y_r, Z_r)| + \|Z_t\|^2)dr < \infty \quad \mathbb{P}.a.s.$$
3. *The pair (Y, Z) satisfies the integral equation (1.12) for all $t \in [0, T]$ $\mathbb{P}.a.s.$*

Hypothesis: We assume that the generator $f : [0, T] \times \mathbb{R}^k \times \mathbb{R}^{k\times d} \to \mathbb{R}^k$

(A1) *Lipschitz condition:* There exists $L > 0$ such that for all $y_1, y_2 \in \mathbb{R}^k$ and $z_1, z_2 \in \mathbb{R}^{k\times d}$

$$|f(t, y_1, z_1) - f(t, y_2, z_2)| \le L(|y_1 - y_2| + \|z_1 - z_2\|), \qquad L > 0,$$

(A2) *Integrability condition:*

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

Existence and uniqueness theorem:

Theorem 1.4. *[52] Under hypothesis **(A1)-(A2)** and for the terminal condition $\xi \in L^2(\mathcal{F}_T)$ the BSDE (1.12) has a unique solution $(Y, Z) \in \mathcal{S}^2 \times M^2$.*
Before proving this theorem, let us establish the following lemma:

Proposition 1.3.1. *[13] Suppose that there exists a positive process $\{f_t\}_{t\ge 0} \in M^2(\mathbb{R})$ and a constants $\lambda > 0$ such that:*

$$|f(t, y, z)| \le f_t + \lambda(|y| + \|z\|), \quad \forall (t, y, z) \in [0, T] \times \mathbb{R}^k \times \mathbb{R}^{k\times d}.$$

If $(Y_t, Z_t)_{0\le t\le T}$ is a solution of BSDE, such that $Z \in M^2$, then $Y \in \mathcal{S}_c^2$.

Lemma 1.3.1. *[13] For $Y \in \mathcal{S}^2(\mathbb{R})$ and $Z \in M^2(\mathbb{R}^{k\times d})$ then $\{\int_0^t Y_s Z_s dB_s, t \in [0, T]\}$ is a uniformaly integrable martingale.*

Lemma 1.3.2. *[13] Simple case where f not depend on y and z, Let $\{F_t\}_{0\le t\le T} \in M^2(\mathbb{R}^k)$ and the BSDE:*

$$Y_t = \xi + \int_t^T F_r\, dr - \int_t^T Z_r\, dB_r, \quad 0 \le t \le T. \tag{1.13}$$

Let $\xi \in L^2(\mathcal{F}_T)$ and $\{F_t\}_{0\le t\le T} \in M^2(\mathbb{R}^k)$. The BSDE (1.13) has a unique solution (Y, Z) such that $Z \in M^2$.

The method of proof is similar to the other existence uniqueness theorems in SDEs. We find a suitable complete space and a suitable mapping from that space into itself, and then invoke the contraction mapping theorem. We will only sketch the proof using Banach's fixed-point theorem.

Proof. Notation:
The Banach space $\mathcal{B}^2 = \mathcal{S}_c^2 \times M^2$,
with the norm:

$$\|(U,V)\|_\alpha = \mathbb{E}\big[\sup_{0\le t\le T} e^{\alpha t}|U_t|^2 + \int_0^T e^{\alpha t}\|V_r\|^2\,dr\big]^{\frac{1}{2}},$$

BSDE to solve:

$$Y_t = \xi + \int_t^T f(r,Y_r,Z_r)\,dr - \int_t^T Z_r\,dB_r.$$

1. Define a contraction mapping Ψ:
   For $(U,V)\in\mathcal{B}^2$, define $(Y,Z)=\Psi(U,V)$ as the solution of:

   $$Y_t = \xi + \int_t^T f(r,U_r,V_r)\,dr - \int_t^T Z_r\,dB_r.$$

2. Ψ is well-defined from $\mathcal{B}^2$ to $\mathcal{B}^2$.
   Verification: Set $F_r = f(r,U_r,V_r)$. Since f is Lipshitz

   $$|F_r| \le |f(r,0,0)| + L|U_r| + L\|V_r\|.$$

   By Lemma 1.3.2 and Proposition 1.3.1, there exists solutions $(Y,Z)\in\mathcal{B}^2$.

3. Let (U,V), $(U',V')\in\mathcal{B}^2$ and define: $(Y,Z)=\Psi(U,V)$, $(Y',Z')=\Psi(U',V')$, $\hat{y}=Y-Y'$, $\hat{z}=Z-Z'$, $\hat{y}_T=\xi-\xi=0$.

   $$d\hat{y}_t = -\{f(t,U_t,V_t)-f(t,U'_t,V'_t)\}\,dt + \hat{z}_t\,dB_t.$$

   Applying Itô formula to $e^{\alpha t}|\hat{y}_t|^2$:

   $$d(e^{\alpha t}|\hat{y}_t|^2) = \alpha e^{\alpha t}|\hat{y}_t|^2\,dt + \|\hat{z}_t\|^2\,dt + 2e^{\alpha t}\,\hat{y}_t\,d\hat{y}_t.$$

   After simplification and integration from t to T:

   $$e^{\alpha t}|\hat{y}_t|^2+\int_t^T e^{\alpha r}\|\hat{z}_r\|^2\,dr = \int_t^T e^{\alpha r}(-\alpha|\hat{y}_r|^2+2\hat{y}_r.\{f(t,U_t,V_t)-f(t,U'_t,V'_t)\})\,dt-\int_t^T 2e^{\alpha r}\,\hat{y}_r.\hat{z}_r\,dB_r.$$

4. Using Lipschitz condition: Since f is Lipschitz, note $\hat{u}=U-U'$, $\hat{v}=V-V'$, with constant L:

   $$e^{\alpha t}|\hat{y}_t|^2+\int_t^T e^{\alpha r}\|\hat{z}_r\|^2\,dr \le \int_t^T e^{\alpha r}(-\alpha|\hat{y}_r|^2+2L|\hat{y}_r||\hat{u}_r|+2L|\hat{y}_r|\|\hat{v}_r\|)\,dr-\int_t^T 2e^{\alpha r}\,\hat{y}_r.\hat{z}_r\,dB_r.$$

   Applying the inequality: for all $\varepsilon>0$, we have $2ab\le\frac{a^2}{\varepsilon}+\varepsilon b^2$:

   $$e^{\alpha t}|\hat{y}_t|^2+\int_t^T e^{\alpha r}\|\hat{z}_r\|^2\,dr \le \int_t^T e^{\alpha r}(-\alpha+2\frac{L^2}{\varepsilon})|\hat{y}_r|^2\,dr-\int_t^T 2e^{\alpha r}\,\hat{y}_r.\hat{z}_r\,dB_r+\underbrace{\varepsilon\int_t^T e^{\alpha r}(|\hat{u}_r|^2+\|v_r\|^2)\,dr}_{R_\varepsilon},$$

choose $\alpha = 2\frac{L^2}{\varepsilon}$ to cancel the $|\hat{y}_t|^2$ terms, and from Lemma 1.3.1 $M_t = \int_0^t 2e^{\alpha r}\ \hat{y}_r.\hat{z}_r\ dB_r$ is local martingal.
Taking expectation at $t = 0$:

$$\mathbb{E}[\int_0^T e^{\alpha r}\|\hat{z}_r\|^2\ dr] \leq \mathbb{E}[R_\varepsilon].$$

5. Contraction conclusion:
Applying BDG inequality:

$$\mathbb{E}[\sup_{0\leq t\leq T} e^{\alpha t}|\hat{y}_t|^2] \leq \mathbb{E}[R_\varepsilon] + C\mathbb{E}[(\int_0^T e^{2\alpha r}|\hat{y}_r|^2\|\hat{z}_r\|^2\ dr)^{1/2}].$$

Using $ab \leq \frac{a^2}{2} + \frac{b^2}{2}$ and absorbing terms:

$$\mathbb{E}[\sup_{0\leq t\leq T} e^{\alpha t}|\hat{y}_t|^2 + \int_0^T e^{\alpha r}\|\hat{z}_r\|^2\ dr] \leq (3 + C^2)\mathbb{E}[R_\varepsilon].$$

return to the definition of R_ε, then:

$$\|(\hat{y}, \hat{z})\|_\alpha^2 \leq \varepsilon(3 + C^2)(1 \vee T)\|(\hat{u}, \hat{v})\|_\alpha^2,$$

choose ε such that $\varepsilon(3 + C^2)(1 \vee T) = \frac{1}{2}$. Then Ψ is a strict contraction.

6. Banach Fixed-Point Theorem:
Since $\mathcal{B}^2$ is a Banach space with norm $\|.\|_\alpha$ and $\Psi : \mathcal{B}^2 \to \mathcal{B}^2$ is a contraction. By Banach's fixed-point theorem, there exists a unique solution $(Y, Z) \in \mathcal{B}^2$.

□

Remark 1.3.2. *We shall see that the role of Z, more specifically that of the term $\int_t^T Z_r dB_r$, is to ensure that the process Y is $\mathcal{F}_t-$ adapted.*

1.4 Forward–Backward Stochastic Differential Equations

Definition 1.1. *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a complete filtered probability space, with filtration $\mathbb{F}$ that satisfies the usual conditions. We consider the following system:*

$$\begin{cases} dX_t = b(t, X_t, Y_t, Z_t)\,dt + \sigma(t, X_t, Y_t, Z_t)\,dB_t, \\ -dY_t = f(t, X_t, Y_t, Z_t)\,dt - Z_t\,dB_t, \\ X_s = x,\ \ Y_T = \Phi(X_T). \end{cases} \tag{1.14}$$

where:

- $X_t^{s,x}$ *is the forward (state) process that satisfies:*

$$X_t^{s,x} = x + \int_s^t b(r, X_r^{s,x}, Y_r^{s,x}, Z_r^{s,x})\,dr + \int_s^t \sigma(r, X_r^{s,x}, Y_r^{s,x}, Z_r^{s,x})\,dB_r, \quad s \leq t \leq T. \tag{1.15}$$

- $Y_t^{s,x}$ *is the backward (costate) process that satisfies:*

$$Y_t^{s,x} = \Phi(X_T^{s,x}) + \int_s^t f(r, X_r^{s,x}, Y_r^{s,x}, Z_r^{s,x})\, dr - \int_s^t Z_r^{s,x} dB_r. \tag{1.16}$$

where $X_t^{s,x}$, $Y_t^{s,x}$ *in* $\mathcal{S}^2(s,T;\mathbb{R}^d), \mathcal{S}^2(s,T;\mathbb{R}^k)$ *respectively.*

Remark 1.4.1. *The spaces notation can be found in Notations*

Classification of Forward–Backward Stochastic Differential Equations

There are two types of FBSDEs: decoupled and coupled FBSDEs.

1. **Decoupled FBSDEs:** An FBSDE is said to be decoupled if the forward component X is independent of the backward component (Y, Z).
   The system is defined by the following equation:

$$\begin{cases} dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t, \\ dY_t = -f(t, X_t, Y_t, Z_t)\, dt + Z_t\, dB_t, \\ Y_T = \Phi(X_T). \end{cases} \tag{1.17}$$

   In this case, the forward SDE is solved first using classical theory (Theorem (1.2.1)). Once the process X is obtained, it is substituted into the BSDE, which can be solved using Theorem (1.4).

2. **Coupled FBSDEs:**
   Fully coupled: An FBSDE is said to be *fully coupled* when both the drift and diffusion coefficients of the forward equation depend on all variables (X, Y, Z). The general system is given as follows:

$$\begin{cases} dX_t = b(t, X_t, Y_t, Z_t)\, dt + \sigma(t, X_t, Y_t, Z_t)\, dB_t, \\ dY_t = -f(t, X_t, Y_t, Z_t)\, dt + Z_t\, dB_t, \\ Y_T = \Phi(X_T). \end{cases} \tag{1.18}$$

   or

$$\begin{cases} X_r^{t,x} = x + \int_t^r b(X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})\, ds + \int_t^r \sigma(X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})\, dB_s, \\ Y_t^{t,x} = \Phi(X_T^{t,x}) - \int_t^T f(X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})\, ds + \int_t^T Z_s\, dB_s, \end{cases} \tag{1.19}$$

   where, the deterministic functions b, σ, f and Φ are defined by:

$$\begin{aligned} &b : \mathbb{R}^d \times \mathbb{R}^p \times \mathbb{R}^{p\times m} \longrightarrow \mathbb{R}^d, \\ &\sigma : \mathbb{R}^d \times \mathbb{R}^p \times \mathbb{R}^{p\times m} \longrightarrow \mathbb{R}^{d\times m}, \\ &f : \mathbb{R}^d \times \mathbb{R}^p \times \mathbb{R}^{p\times m} \longrightarrow \mathbb{R}^p, \\ &\Phi : \mathbb{R}^d \longrightarrow \mathbb{R}^p. \end{aligned}$$

   The processes $X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}$ are $\mathcal{F}_t$-adapted square integrable processes[12].

- $\mathcal{S}^2(t,T;\mathbb{R}^m)$ denote the set of $\mathbb{R}^m$-valued, $\mathbb{F}$-adapted, continuous processes $(X_s, s \in [t,T])$ which satisfy $\mathbb{E}[\sup_{t\leq s\leq T} |X_s|^2] < \infty$.
- $\mathcal{H}^2(t,T;\mathbb{R}^m)$ is the set of $\mathbb{R}^m$-valued, $\mathbb{F}$-predictable processes $(Z_s, s \in [t,T])$ which satisfy $\mathbb{E}[\int_t^T |Z_s|^2 ds] < \infty$.
- $\mathcal{M}^2(t,T;\mathbb{R}^m)$ denotes the set of all $\mathbb{R}^m$-valued, square integrable RCLL martingales $M = (M_s)_{s\in[t,T]}$ with respect to $\mathbb{F}$, with $M_t = 0$.
- $\mathcal{K}_t^{d,k,p\times m} = \mathcal{S}^2(t,T;\mathbb{R}^d) \times \mathcal{S}^2(t,T;\mathbb{R}^k) \times \mathcal{H}^2(t,T;\mathbb{R}^{p\times m})$.

Definition 1.4.1. *A solution of FBSDE* (1.19) *is a process* $(X^{t,x}, Y^{t,x}, Z^{t,x}) \in \mathcal{K}_t^{m,p,p\times m}$*, which satisfies* (1.19).

Two main approaches have been developed for solving coupled FBSDEs. The first is the continuation method, introduced by *Hu and Peng* [31] and further developed by *Peng and Wu* [57]. This method requires a suitable *monotonicity condition* on the coefficients to ensure the existence and uniqueness of the solution. A comprehensive treatment of this method can be found in [31], [57].
The second approach is the *Four-step scheme* developed by Ma, Protter, and Yong [45]. This scheme transforms the FBSDE into a partial differential equation (PDE) problem.

Existence and uniqueness of the solution of the FBSDEs:
We introduce a direct method for solving FBSDEs. It is called *Four Step Scheme*, because it contains four major steps.
Four Step Scheme: [45] Let us consider the FBSDE in which the forward equation is *non-degenerate*. *"Jin Ma, Philip Protter and Jiongmin Yong"* prove that in this case the adapted solution can always be sought in an "ordinary" sense over an arbitrary prescribed time duration via a direct "Four step scheme". Using this scheme, they further prove that the backward components of the adapted solution are determined explicitly by the forward component via the solution of a certain quasi-linear parabolic PDE system.
Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space satisfying the usual conditions, let B be a standard $d-$dimensional Brownian motion. Let us consider the following forward-backward SDE:

$$\begin{cases} X_t = x + \int_0^t b(s, X_s, Y_s, Z_s)\, ds + \int_0^t \sigma(s, X_s, Y_s, Z_s)\, dB_s, \\ Y_t = \Phi(X_T) - \int_t^T f(s, X_s, Y_s, Z_s)\, ds + \int_t^T g(s, X_s, Y_s, Z_s)\, dB_s, \end{cases} \tag{1.20}$$

equivalently

$$\begin{cases} dX_t = b(t, X_t, Y_t, Z_t)\, dt + \sigma(t, X_t, Y_t, Z_t)\, dB_t, \\ dY_t = -f(t, X_t, Y_t, Z_t)\, dt + g(t, X_t, Y_t, Z_t)\, dB_t, \quad t \in [0,T], \\ X_0 = x, \quad Y_T = \Phi(X_T) \end{cases} \tag{1.21}$$

Here, the processes X, Y and Z take values in $\mathbb{R}^n, \mathbb{R}^m$ and $\mathbb{R}^{m\times d}$, respectively; and the functions b, σ, f, g and Φ take valus in $\mathbb{R}^n, \mathbb{R}^m, \mathbb{R}^{n\times d}, \mathbb{R}^{m\times d}$ and $\mathbb{R}^m$. In what follows, we use the usual Euclidean norms in $\mathbb{R}^m$ and $\mathbb{R}^n$ and for $z \in \mathbb{R}^{m\times d}$ (resp $\mathbb{R}^{n\times d}$) we define $|z| = \{tr(zz^T)\}^{1/2}$. Then $\mathbb{R}^{m\times d}$ (resp $\mathbb{R}^{n\times d}$) is a Hilbert.

Definition 1.4.2. *A triple of processes* $(X, Y, Z) : [0,T] \times \Omega \to \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{m\times d}$ *is called an ordinary adapted solution of the FBSDE* (1.20)*, if it is* $\{\mathcal{F}_t\}$*-adapted and square integrable, such that it satisfies* (1.20) $\mathbb{P}$*-almost surely.*

Suppose that (X_t, Y_t, Z_t) is an adapted solution of (1.20) or equivalently (1.21). We assume that Y and X are related by:

$$Y_t = \theta(t, X_t) \quad \forall t \in [0,T] \quad \mathbb{P}\text{ -a.s.} \tag{1.22}$$

where θ is some function to be determined. Suppose that all the functions involved are smooth, say at least $\mathcal{C}^2$; then by applying Itô formula. We have for $1 \le k \le m$.

$$\begin{aligned} dY_t^k &= d\theta^k(t, X_t) \\ &= \Big\{\theta_t^k(t, X_t) + \langle \theta_x^k(t, X_t), b(t, X_t, \theta(t, X_t), Z_t)\rangle \\ &\quad + \frac{1}{2}\, tr\big[\theta_{xx}^k(t, X_t)\, \sigma(t, X_t, \theta(t, X_t))^T\big]\Big\}\, dt \\ &\quad + \langle \theta_x^k(t, X_t), \sigma(t, X_t, \theta(t, X_t))\, dB_t\rangle. \end{aligned} \tag{1.23}$$

comparing (1.23) and (1.21), we see that if θ is the right choice, then it necessary that, for $k = 1, \dots, m$

$$\begin{cases} -f^k(t, X_t, \theta(t, X_t)) = \theta_t^k(t, X_t) + \langle \theta_x^k(t, X_t), b(t, X_t, \theta(t, X_t), Z_t)\rangle \\ +\frac{1}{2} tr[\theta_{xx}^k(t, X_t)\sigma(t, X_t, \theta(t, X_t))\sigma(t, X_t, \theta(t, X_t))^T]; \\ \theta(T, X_T) = \Phi(X_T). \end{cases} \tag{1.24}$$

and

$$-g(t, X_t, Y_t, Z_t) = \theta_x(t, X_t)\sigma(t, X_t, \theta(t, X_t)). \tag{1.25}$$

The above arguments suggest that we design the following "Four Step Scheme" to solve the FBSDE (1.20).

Four step scheme

Step 1: Find a smoth mapping $z \in [0,T] \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times d}$ satisfying

$$p\sigma(t, x, y) + g(t, x, y, z(t, x, y, p)) = 0; \quad \forall (t, x, y, p) \in [0,T] \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{m\times n}. \tag{1.26}$$

Step 2: Using the function z above, solve the following parabolic system for $\theta(t, x)$:

$$\begin{cases} \theta_t^k + \frac{1}{2} tr(\theta_{xx}^k \sigma(t, x, \theta)\sigma(t, x, \theta)^T) + \langle b(t, x, \theta, z(t, x, \theta, \theta_x)), \theta_x\rangle \\ +g^k(t, x, \theta, z(t, x, \theta, \theta_x)) = 0, \quad k = 1, \dots, m, \quad (t, x) \in (0,T) \times \mathbb{R}^n, \\ \theta(T, x) = \Phi(x), \quad x \in \mathbb{R}^n \end{cases} \tag{1.27}$$

Step 3: Using θ and z, solve the following forward SDE:

$$X_t = x + \int_0^t \tilde{b}(s, X_s)\, ds + \int_0^t \tilde{\sigma}(s, X_s)\, dB_s, \tag{1.28}$$

where $\tilde{b}(t, x) = b(t, x, \theta(t, x), z(t, x, \theta(t, x), \theta_x(t, x)))$ and $\tilde{\sigma}(t, x) = \sigma(t, x, \theta(t, x))$.

Step 4: Set

$$\begin{cases} Y_t = \theta(t, X_t), \\ Z_t = z(t, X_t, \theta(t, X_t), \theta_x(t, X_t)). \end{cases} \tag{1.29}$$

Then if this scheme is realizable, (X_t, Y_t, Z_t) would give an adapted solution of (1.20).

In the study of coupled FBSDEs, the classification of the system as *degenerate* or *non-degenerate* is determined by the properties of the diffusion coefficient matrix σ.

Remark 1.4.2. *The non-degenerate means that the diffusion matrix σ is non-degenerate, which satisfies the non-degeneracy condition 1.4.*

The following presentation is adopted from [12].

***Non-degenerate case*:**Let us consider the following results of Delarue's [17]. Delarue has proved the existence and uniqueness of a solution to FBSDE, focusing on the non-degeneracy case, and the matrix σ is independent of Z.

The system studied is a fully coupled FBSDE on a time duration $[0, T]$ defined as follow:

$$\begin{cases} X_t = \xi + \int_0^t b(s, X_s, Y_s, Z_s)\, ds + \int_0^t \sigma(s, X_s, Y_s)\, dB_s, \quad \forall t \in [0, T] \\ Y_t = \Phi(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)\, ds - \int_t^T Z_s\, dB_s, \quad \forall t \in [0, T] \\ \mathbb{E} \int_t^T (|X_t|^2 + |Y_t|^2 + |Z_t|^2)\, dt < \infty. \end{cases} \tag{1.30}$$

Main Assumptions

The proof relies on two sets of assumptions, denoted (A1) and (A2). The assumptions (A1) consists of local solvability conditions, which are the Lipschitz and monotonicity conditions (we refer the reader to [17] for the precise statement).

*Assumption **(A2)** (Global Extension):*

These assumptions contain the non-degeneracy condition to ensure global existence.

There exist two constants K and $\lambda > 0$, such that the functions b, σ, f and Φ satisfy the following assumptions:

(A.2.1) The functions b, σ, f and Φ are bounded and satisfy for any (x, y, z) and $(x', y', z') \in \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d$:

$$\begin{aligned} &|b(x, y, z) - b(x', y', z')| \leq K(|x - x'| + |y - y'| + |z - z'|), \\ &|\sigma(x, y) - \sigma(x', y')| \leq K^2(|x - x'|^2 + |y - y'|^2), \\ &|f(x, y, z) - f(x', y', z')| \leq K(|x - x'| + |y - y'| + |z - z'|), \\ &|\Phi(x) - \Phi(x')| \leq K|x - x'|. \end{aligned}$$

(A.2.2) *Non-degeneracy condition:* For every $(t, x, y) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}; \forall \zeta \in \mathbb{R}^d$

$$\langle \zeta, \sigma\sigma^T(t, x, y)\zeta \rangle \geq \lambda |\zeta|^2.$$

Theorem 1.5. *[17] Under Assumption **(A2)**, for any terminal time $T > 0$ and any initial condition ξ with finite second moment, The FBSDE (1.30) admits a unique adapted solution $(X_t, Y_t, Z_t)_{0 \leq t \leq T}$.*

Sketch of Proof:

The proof is conducted in two distinct steps: establishing local existence via the fixed-point theorem, and extending to global existence using PDE.

Step1: Existence and Uniqueness in Small Time Duration:

The first step proves that for a sufficiently small time horizon $T \leq C_k$, the system (1.30) admits a unique solution (Theorem 1.1 in [17]).

- The Mapping Ψ:
  We define a mapping Ψ on the space $\mathcal{K} = \mathcal{S}^2_T(\mathbb{R}^d) \times \mathcal{S}^2_T(\mathbb{R}^p) \times \mathcal{H}^2_T(\mathbb{R}^{d\times p})$. For a given triplet $(X, Y, Z) \in \mathcal{K}$, we define $(\bar{X}, \bar{Y}, \bar{Z}) = \Psi(X, Y, Z)$ as follows:
  • $\bar{X}$ is the solution to the SDE: $d\bar{X}_t = b(t, \bar{X}, Y_t, Z_t)\, dt + \sigma(t, \bar{X}_t, Y_t)\, dB_t$ with initial condition ξ. Under (A1) this has a unique solution.
  • $(\bar{Y}, \bar{Z})$ is the solution to the BSDE:
  $\bar{Y}_t = \Phi(\bar{X}_T) + \int_t^T f(s, \bar{X}_s, \bar{Y}_s, \bar{Z}_s)\, ds - \int_t^T \bar{Z}_s\, dB_s$ under (A1) and given $\bar{X}$, this admits a unique solution.

- Contraction Argument:
  Using Itô's formula and BDG inequalities, we establish the difference of two iterations $(\bar{X}, \bar{Y}, \bar{Z})$ and $(\bar{U}, \bar{V}, \bar{W})$. Then the mapping Ψ is contractive from $\mathcal{K}$ to itself. By Banach's fixed point theorem, for every $T \leq C_k$ there exists a unique solution to the problem (1.30).

We establish a priori estimates to ensure that the norms of X_t and Y_t remain bounded within the space $\mathcal{K} = \mathcal{S}^2_T(\mathbb{R}^d) \times \mathcal{S}^2_T(\mathbb{R}^p) \times \mathcal{H}^2_T(\mathbb{R}^{d\times p})$. Specifically, from Delarue's local analysis (Corollary 1.4), for any initial condition $(t, x) \in [0, T] \times \mathbb{R}^d$, the local solution satisfies the following a priori estimates:

1. There exists a constant c_K only depending on K:

$$\mathbb{E}[\sup_{0\leq s\leq T} |X_s^{t,x}|^2] + \mathbb{E}[\sup_{0\leq s\leq T} |Y_s^{t,x}|^2] + \mathbb{E}[\int_0^T |Z_s^{t,x}|^2 ds] \leq c_K(1 + |x|^2). \tag{1.31}$$

2. There exists a constant $c_k^{(2)}$ only depending on K such that $\forall (t, t') \in [0, T]^2, \quad \forall (x, x') \in (\mathbb{R}^d)^2$,

$$\mathbb{E}[\sup_{0\leq s\leq T} |X_s^{t',x'} - X_s^{t,x}|^2] + \mathbb{E}[\sup_{0\leq s\leq T} |Y_s^{t',x'} - Y_s^{t,x}|^2] + \mathbb{E}[\int_0^T |Z_s^{t',x'} - Z_s^{t,x}|^2 ds] \leq c_k|x - x'|^2 + c_k^{(2)}(1 + |x|^2)|t' - t|.$$

Step2: Global Extension via PDE connection:

After establishing local existence and uniqueness on sufficiently small time intervals, the objective is to extend this result to an arbitrary time horizon $[0, T]$

- The Associated PDE system:

$$(E') \quad \begin{cases} \forall (t, x) \in [0, T] \times \mathbb{R}^d, \forall \ell \in \{1, \dots, q\} : \\ \frac{\partial \theta^\ell}{\partial t}(t, x) + \frac{1}{2}\sum_{i,j=1}^d a_{i,j}(t, x, \theta(t, x))\frac{\partial^2 \theta^\ell}{\partial x_i \partial x_j}(t, x) \\ + \sum_{i=1}^d b_i(t, x, \theta(t, x), \nabla_x \theta(t, x)\sigma(t, x, \theta(t, x)))\frac{\partial \theta^\ell}{\partial x_i}(t, x) \\ + f^\ell(t, x, \theta(t, x), \nabla_x \theta(t, x)\sigma(t, x, \theta(t, x))) = 0 \\ \theta(T, x) = \Phi(x) \end{cases} \tag{1.32}$$

$\forall(t,x,y) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^p, \quad a(t,x,y) = \sigma\sigma^\top(t,x,y)$ admits a unique solution $\theta \in C^{1,2}$. To formally justify the link between the FBSDE and the associated PDE system (1.32), we rely on the continuity and regularity of the decoupling field θ. Because $Y_s^{t,x}$ is $\mathcal{F}_t$ measurable and $\mathcal{F}_0$ is trivial, $Y_0^{0,x}$ is deterministic, allowing us to define $\theta(t,x) = Y_s^{t,x}$. The a priori estimates above guarantee that (t,x) is continuous and locally Lipschitz in x. By applying Itô's formula to $\theta(t,x)$ and taking conditional expectations. This Itô justification formally proves that any local FBSDE solution must satisfy the representation $Y_s = \theta(s, X_s)$, thereby linking the stochastic problem directly to the deterministic PDE.

- Decoupling Field: The solution (X,Y,Z) is linked to the PDE solution by the representation formula:
$$Y_t = \theta(t,X_t), \text{ and } \quad Z_t = \nabla_x\theta(t,X_t)\sigma(t,X_t,Y_t).$$
- Gradient estimates: Delarue constructs a sequence of regularized coefficients $(b_n, f_n, \sigma_n, \Phi_n)$ satisfying (A2) and converges to (b, f, σ, Φ) almost everywhere. Using classical result (e.g. Ladyzenskaja et al, 1968), the solutions θ_n to the regularized PDEs satisfy uniform gradient bounds: $\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |\nabla_x\theta_n(t,x)| \leq M$.
By passing to the limit as $n \to \infty$, the true decoupling field θ inherits this exact uniform Lipschitz bound. This uniform bound is the crucial mechanism that prevents the gradient from blowing up during a backward induction. Because the Lipschitz constant of θ is strictly capped by M, the effective Lipschitz constant of the system on any subsequent local time step is bounded by $\max(K, M)$. Consequently, the maximum allowable time $\delta = C_K$ for the Picard contraction in step 1 remains strictly positive and does not shrink to zero. This allows the interval $[0,T]$ to be partitioned into a finite number of sub-intervals of length δ. The local solutions can therefore be successfully stitched together via backward induction, yielding a unique global solution over entire horizon $[0,T]$.

□

While Delarue's approach provides a robust link between FBSDEs and quasilinear PDEs via the decoupling field $\theta(t,x)$, it is restricted by the requirement of a non-degenerate diffusion matrix. To address cases where σ may be degenerate, Peng and Wu [57] introduced a purely probabilistic method for fully coupled FBSDEs over an arbitrarily large time duration. Their approach is based on a monotonicity condition and direct energy estimates to prove global existence.

Generalized framework for FBSDE system:
The following framework is based on the so-called *continuation method*, introduced by *Peng and Wu* [57].
For a given $1 \times d$ matrix G (with G^T be the transpose of G) and $\gamma := (x,y,z)$, we put

$$A(t,\gamma) = \begin{pmatrix} -G^T f \\ Gb \\ G\sigma \end{pmatrix} (t,\gamma).$$

We assume that there exists a $1 \times d$ full rank matrix G such that the following assumptions are satisfied.

H1
1. $A(t,\gamma)$ is uniformly Lipschitz in γ uniformly on t, and for any γ,
$$A(.,\gamma) \in \mathcal{H}^2(0,T;\mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d)$$
2. $\Phi(x)$ is uniformly Lipschitz and for any $x \in \mathbb{R}^d, \Phi(x) \in L^2(\Omega, \mathcal{F}_T, \mathbb{P}; \mathbb{R})$. We denote by K the Lipschitz constant of A and Φ.

H2
1. $\langle A(t,\gamma) - A(t,\hat{\gamma}), \gamma - \hat{\gamma} \leq -\beta_1|G\bar{x}|^2 - \beta_2(|G^T\bar{y}|^2 + |G^T\bar{z}|^2)$.
2. $\langle \Phi(x) - \Phi(\hat{x}), G(x-\hat{x})\rangle \geq \mu_1|G\bar{x}|^2, \bar{x} = x - \hat{x}, \bar{y} = y - \hat{y}, \bar{z} = z - \hat{z}$, where β_1, β_2, μ_1 are strictly positive constants.

Theorem 1.4.1. *[57] We suppose that the assumptions $(H1-H2)$ hold. Then, there exists a unique adapted solution (X,Y,Z) to the FBSDE (1.19).*

1.5 Connection with linear and Non-linear Partial Differential Equations

Definition 1.5.1. *[62]* ***Generator*** *Let $\{X_s^{t,x}, s \geq t\}$ be the strong solution of (1.4). For a function $f : \mathbb{R}^n \to \mathbb{R}$, we define the function $\mathcal{L}f$ by:*

$$\mathcal{L}f(x) = \lim_{h\to 0} \frac{\mathbb{E}[f(X_{t+h}^{t,x})] - f(x)}{h}$$

if the limit exists. Clearly, $\mathcal{L}f$ is well defined for all bounded $\mathcal{C}^2$-function with bounded derivatives and:

$$\mathcal{L}f(x) = b(t,x).Df(x) + \frac{1}{2}Tr[\sigma\sigma^T(t,x)D^2f(x)] \tag{1.33}$$

where Df and D^2f denote the gradient and Hessian of f, respectively, the linear differential operator $\mathcal{L}$ is called the generator of X.

The operator $\mathcal{L}$ plays a central role in connecting stochastic processes with linear PDEs.

Definition 1.5.2. *[62] Let $v(t,x)$ be a function of classe $\mathcal{C}^{1,2}([0,T] \times \mathbb{R}^n)$ defined by: $v(t,x) := \mathbb{E}[\Phi(X_T^{t,x})]$. Then v solves the partial differential equation:*

$$\begin{cases} \frac{\partial v(t,x)}{\partial t} + \mathcal{L}v(t,x) = 0, & (t,x) \in [0,T) \times \mathbb{R}^n, \\ v(T,x) = \Phi(x), & x \in \mathbb{R}^n. \end{cases}$$

This establishes a fundamental connection between the generator and linear PDEs.
The Feynman-Kac Representation: Let us consider the following linear partial differential equation:

$$\frac{\partial v(t,x)}{\partial t} + \mathcal{L}v(t,x) - k(t,x)v(t,x) + f(t,x) = 0, \quad v(T,x) = \Phi(x), \quad (t,x) \in [0,T] \times \mathbb{R}^n \tag{1.34}$$

$\mathcal{L}$ is the generator (1.33), Φ is given function ($\Phi : \mathbb{R}^n \to \mathbb{R}$), function $k, f : [0,T] \times \mathbb{R}^d \to \mathbb{R}$, function $b, \sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\mathcal{M}_{\mathbb{R}}(d,d)$, respectively.

Theorem 1.6. *Let the coefficients b, σ be continuous and satisfy the Lipschitz condition in x uniformly in t, $\int_0^T (|b(t,0)|^2 + |\sigma(t,0)|^2)\, dt < \infty$. Assume that the function k is uniformly bounded and f has quadratic growth in x uniformly in t. Let v be a $\mathcal{C}^{1,2}([0,T] \times \mathbb{R}^n)$ solution of (1.34) with quadratic growth in x uniformly in t, then:*

$$v(t,x) = \mathbb{E}[\int_t^T \beta_s^{t,x} f(s, X_s^{t,x})\, ds + \beta_T^{t,x} \Phi(X_T^{t,x})], \quad t \le T \quad x \in \mathbb{R}^d, \tag{1.35}$$

where $\{X_s^{t,x}, s \ge t\}$ is the solution of the SDE with the initial data $X_t^{t,x} = x$ and $\beta_s^{t,x} := e^{\int_t^s k(u, X_u^{t,x})\, du}$ for $t \le s \le T$.

The above Feynman-Kac representation formula has an important numerical implication. Indeed, it opens the door to the use of *Monte Carlo* methods in order to obtain a numerical approximation of the solution of the partial differential equation (1.34) [62].

A Feynman-Kac Type Representation through FBSDEs:

There is an innate relation between stochastic differential equations and second-order partial differential equations (PDEs) of parabolic or elliptic type. Specifically, solutions to a certain class of non-linear PDEs can be represented by solutions to forward-backward stochastic differential equations (FBSDEs), in the same spirit as demonstrated by the well-known Feynman-Kac formulas (Karatzas and Shreve, 1991) for linear PDEs [22]. Then, we proceed to link the solutions of forward and backward processes with the solution of the PDEs, within the framework of a *nonlinear Feynman-Kac formula.*

Consider a complete filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. While FSDEs have a fairly straightforward definition, in the sense that both the SDE and the filtration evolve forward in time, this is not the case for BSDEs. Indeed, since solutions to BSDEs need to satisfy a terminal condition, integration needs to be performed backwards in time in some sense, yet the filtration still evolves forward in time. It turns out that a terminal value problem involving BSDEs admits an adapted solution if we back-propagate the conditional expectation of the process, that is, if we set

$$Y_s = \mathbb{E}[Y_s \mid \mathcal{F}_s].$$

In a sense, systems of FBSDEs (1.14) describe two-point boundary value problems[22]. Moreover, if the functions b, σ, f and Φ are deterministic, in the sense that they do not depend explicitly on $\omega \in \Omega$, then the solution (Y, Z) exhibits the Markovian property, this implies the solution can be expressed as deterministic functions of time and the state process (El Karoui, Peng, and Quenez, 1997).

Lemma 1.3. *[22] (The Markovian Property) There exists a deterministic decoupling function $\theta : [0,T] \times \mathbb{R}^n \to \mathbb{R}$ and $\theta_x : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ such that the solution $(Y^{t,x}, Z^{t,x})$ of the BSDE (1.16) is*

$$Y_s^{t,x} = \theta(s, X_s^{t,x}), \quad Z_s^{t,x} = \sigma^\top(s, X_s^{t,x}) \theta_x(s, X_s^{t,x}),$$

for all $s \in [t,T]$.

The connection between the FBSDE system and nonlinear parabolic PDEs is formalized through the nonlinear Feynman-Kac formula. Indeed, the following lemma can be proven by an application of Itô formula (see El Karoui et al., 1997; Ma and Yong, 1999; Yong and Zhou, 1999):

Lemma 1.4. *[22](Nonlinear Feynman-Kac) Consider the Cauchy problem:*

$$\begin{cases} \theta_t + \frac{1}{2}tr(\theta_{xx}\sigma(t,x)\sigma^\top(t,x)) + \theta_x^\top b(t,x) + f(t,x,\theta,\sigma^\top(t,x)\theta_x) = 0, & (t,x) \in [0,T) \times \mathbb{R}^n, \\ \theta(T,x) = \Phi(x), \quad x \in \mathbb{R}^n, \end{cases} \tag{1.36}$$

wherein the functions b, σ, f and Φ satisfy mild regularity conditions[2]*. Then (1.36) admits a unique (viscosity*[3]*) solution $\theta : [0,T] \times \mathbb{R}^n \to \mathbb{R}$, which has the following probabilistic representation:*

$$\theta(t,x) = Y_t^{t,x}, \quad \forall (t,x) \in [0,T] \times \mathbb{R}^n \tag{1.37}$$

where (X,Y,Z) is the unique adapted solution of the FBSDE system. Furthermore,

$$(Y_s^{t,x}, Z_s^{t,x}) = \big(\theta(s, X_s^{t,x}), \sigma^\top(s, X_s^{t,x})\theta_x(s, X_s^{t,x})\big), \tag{1.38}$$

for all $s \in [t,T]$, and if (1.36) admits a classical solution, then (1.37) provides that classical solution.

In this chapter, we developed the mathematical framework needed to study stochastic systems. We began with stochastic calculus, then introduced stochastic differential equations (SDEs), which describe the forward evolution of a state process X_t. Next, we presented backward stochastic differential equations (BSDEs), where the solution (Y_t, Z_t) is determined by a terminal condition at time T. Finally, we combined these ideas to define Forward Backward stochastic differential equations (FBSDEs), which link the forward and backward components in a coupled system.

These equations provide the main tools for studying the optimal control problem. In the next chapter, the forward equation will represent the controlled state process, while the backward equation will describe the adjoint process.

[2]The regularity conditions requires the functions σ, b, f and Φ to be continuous, σ and b to be uniformly Lipschitz in x, and f to be Lipschitz in (y,z), uniformly w.r.t (t,x)(see remark in [22]).

[3]See Remark 2 in [22] and for the definition of the viscosity solution in the next Chapter 2.3.1

Chapter 2

Optimal Control Framework Using FBSDEs

This chapter introduces the framework of stochastic optimal control and its characterization using Forward-Backward SDEs. We review the basics of optimal control and present the Stochastic Maximum Principle (SMP) as a tool for finding necessary conditions for optimality. The dynamic programming approach, including the Hamilton-Jacobi-Bellman (HJB) equation, is then discussed, highlighting its connection to SMP.

2.1 Optimal Control

The stochastic optimal control problem concerns the optimization of a performance criterion over a class of admissible controls while the state evolves according to a stochastic differential equation. This framework is formulated and illustrated through several examples drawn from practical applications.

In this section, we assume that the filtration $\mathbb{F}$ is the canonical filtration of the Brownian motion B. The following definitions are taken from [62]

Definition 2.1.1. ***Control process*** *Given a subset U of $\mathbb{R}^k$, we denote by $\mathcal{U}$ the set of all progressively measurable processes $u = \{u_t, t < T\}$ valued in U. The elements of $\mathcal{U}$ are called control processes.*

The set $\mathcal{U}$ defined by:

$$\mathcal{U}[0,T] = \{u : [0,T] \times \Omega \to U \mid u(.) \text{ is progressively measurable}\}.$$

Presentation of the problem:

Definition 2.1.2. ***the State equation (Controlled process)*** *Let*

$$b : (t,x,v) \in [0,T] \times \mathbb{R}^d \times U \to b(t,x,v) \in \mathbb{R}^d,$$

and

$$\sigma : (t, x, v) \in [0, T] \times \mathbb{R}^d \times U \to \sigma(t, x, v) \in \mathbb{R}^{d\times m},$$

be two continuous functions satisfying the following conditions:

$$|b(t, x, v) - b(t, y, v)| + |\sigma(t, x, v) - \sigma(t, y, v)| \le K\ |x - y|; \tag{2.1}$$

$$|b(t, x, v)| + |\sigma(t, x, v)| \le K\ (1 + |x| + |v|), \tag{2.2}$$

for some constant K independent of (t, x, v).
For each control process $u \in \mathcal{U}$, we consider the controlled stochastic differential equation:

$$dX_t = b(t, X_t, u_t)\ dt + \sigma(t, X_t, u_t)\ dB_t, \qquad X_s = x. \tag{2.3}$$

The unique solution to SDE (2.3) is denoted by $\{X_r^{s,x;u}, r \in [s, T]\}$.
Next, we introduce **the cost functional**: let l be the *running cost* defined by:

$$l : [0, T] \times \mathbb{R}^d \times U \to \mathbb{R}, \quad \Phi : \mathbb{R}^d \to \mathbb{R},$$

where Φ called *the terminal cost*, and the cost functional is defined as follows:

$$J(s, x; u) = \mathbb{E}[\int_s^T l(r, X_r^{s,x;u}, u_r)\ dr + \Phi(X_T^{s,x;u})].$$

Definition 2.1.3. *[65] A control u is called an admissible control, and the couple $(X^{s,x;u}, u)$ is called an admissible pair if :*

(i) $u \in \mathcal{U}[s, T]$.

(ii) $X^{s,x;u}$ *is the unique solution of* (2.3).

(iii) some prescribed state constraints are satisfied.

(iv) $l(s, X_s^{s,x;u}, u_s) \in L^1_{\mathcal{F}}(0, T; \mathbb{R})$ *and* $\Phi(X_T) \in L^1_{\mathcal{F}_T}(\Omega; \mathbb{R})$.

The subset of $\mathcal{U}$ for which the SDE has a unique solution called the set of admissible controls and is denoted by $\mathcal{U}_{ad}[0, T]$.
The Stochastic optimal control problem is formulated as the minimization of the cost functional:

$$J(s, x; \bar{u}) = \inf_{u\in\mathcal{U}_{ad}} J(s, x; u). \tag{2.4}$$

Any control $\bar{u}$ is called an *optimal control*, and the corresponding state process $X^{s,x;\bar{u}}$ is called *the optimal state process.*

Remark 2.1.1. *A notable observation is that one can switch between maximization and minimization problems through the relation:*

$$\min\{J\} = -\max\{-J\}. \tag{2.5}$$

Two main approaches exist for solving stochastic optimal control problems: the *stochastic maximum principle* (SMP for short) and *dynamic programming* (DP for short).
The maximum principle, analogous to the Pontryagin maximum principle in deterministic control, derives necessary conditions for optimality by introducing *adjoint processes*. This approach naturally leads to a system of FBSDEs. The dynamic programming approach, on the other hand, characterizes the value function as the solution of a *Hamilton-Jacobi-Bellman* PDE. These two approaches are deeply connected, and the relationship between them illuminates the structure of FBSDEs and their role in optimal control theory. We develop both approaches in detail, emphasizing their interconnections and relative advantages for different problem classes.

2.2 Stochastic Maximum Principle

One of the fundamental approaches to solving optimization problems is to derive necessary conditions that any optimal solution must satisfy. The maximum principle, formulated and derived by Pontryagin and his group in the 1950s, states that any optimal control and its corresponding state trajectory must satisfy an extended Hamiltonian system, which is a two-point boundary value problem, plus a maximum condition of a function called the Hamiltonian.

Statement of the Stochastic maximum principle:
Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a given filtered probability space satisfying the usual conditions, on which an m-dimensional standard Brownian motion is defined. Consider the stochastic controlled system:

$$dX_t = b(t, X_t, u_t)\ dt + \sigma(t, X_t, u_t)\ dB_t. \tag{2.6}$$

where

$$b(t, x, u) : [0, T] \times \mathbb{R}^d \times U \to \mathbb{R}^d$$
$$\sigma(t, x, u) : [0, T] \times \mathbb{R}^d \times U \to \mathbb{R}^{d\times m}, \quad \sigma = (\sigma^1, \sigma^2, .., \sigma^d).$$

We adopt the following statement from [32].
The set of admissible controls is defined as

$$\mathcal{U}[0, T] = \{u : [0, T] \times \Omega \to U \mid u \text{ is } \mathcal{F}_t\text{-adapted and } \mathbb{E}\left[\int_0^T \|u_t\|^2\ dt\right] < \infty\}.$$

Assume that $U \in \mathbb{R}^d$ is a convex set. The stochastic optimal control problem is to minimize the cost functional over $\mathcal{U}[0, T]$, which is mathematically formulated as follows:

$$\begin{cases} J(u) = \mathbb{E}[\int_0^T l(s, X_s, u_s)\ ds + \Phi(X_T)] \\ \inf\{J(u); u \in \mathcal{U}\} \end{cases} \tag{2.7}$$

To derive the SMP, we introduce the Hamiltonian:

$$\begin{aligned} H(t, x, u, p, q) = p^\top b(t, x, u) + Tr(q^\top \sigma(t, x, u)) - l(t, x, u), \\ \forall (t, x, u, p, q) \in [0, T] \times \mathbb{R}^d \times U \times \mathbb{R}^d \times \mathbb{R}^{d\times m}. \end{aligned} \tag{2.8}$$

Associated with the stochastic control problem (2.7), the adjoint equation are defined as follows:

$$\begin{cases} dp_t &= -H_x(t, X_t, u_t, p_t, q_t)\, dt + q_t\, dB_t, \qquad t \in [0, T], \\ p_T &= -\Phi_x(X_T), \end{cases} \tag{2.9}$$

where H_x is the derivative of Hamiltonian H. It is notable that equation (2.9) is a BSDE whose solution is formed by a pair of processes, $(p, q) \in L^2_{\mathcal{F}}(0, T; \mathbb{R}^n) \times (L^2_{\mathcal{F}}(0, T; \mathbb{R}^n))^d$. Concerning the well-posedness of BSDE, we have the following assumption.

Assumption:

Let $\varphi = b, \sigma, l$ and Φ. The map φ is $\mathcal{C}^2$ in x, and $\varphi(t, 0, u)$ is bounded for any $(t, u) \in [0, T] \times U$. Moreover, φ, φ_x and φ_{xx} are uniformly Lipschitz in x and u.

Theorem 2.2.1. *[32, 65] **(Stochastic Maximum Principle)***

Let Assumption 2.2 hold, and $(\bar{X}_t, \bar{u}_t, \bar{p}_t, \bar{q}_t)$ be an admissible 4-tuple. Suppose that Φ is convex, H is concave for all $t \in [0, T]$ $\mathbb{P}$ almost surely, and the maximum condition

$$H(t, \bar{X}_t, \bar{u}_t, \bar{p}_t, \bar{q}_t) = \max_{u \in U} H(t, \bar{X}_t, u, \bar{p}_t, \bar{q}_t) \quad \forall t \in [0, T], \quad \mathbb{P}\text{−}a.s. \tag{2.10}$$

holds. Then, $(\bar{X}_t, \bar{u}_t)$ is an optimal pair of the problem 2.7.

The proof can be found in [65].

Under sufficient smoothness and concavity assumptions of H, the optimization (2.10) is uniquely solved by the following first-order conditions

$$H_u(t, \bar{X}_t, \bar{u}_t, \bar{p}_t, \bar{q}_t) = 0.$$

Assuming this equation can be solved explicitly, the optimal control admits the following representation through a mapping $\vartheta : (t, \bar{X}_t, \bar{p}_t, \bar{q}_t) \to \bar{u}_t$, such that

$$\bar{u}_t = \vartheta(t, \bar{X}_t, \bar{p}_t, \bar{q}_t), \quad t \in [0, T].$$

We refer to ϑ as the feedback map. For a broad class of practically relevant control problems, this mapping can be obtained in closed form. Substituting the feedback control into the state and adjoint equations yields the fully coupled FBSDE:

$$\begin{cases} dX_t = b(t, X_t, p_t, q_t)\, dt + \sigma(t, X_t, p_t, q_t)\, dB_t, \\ dp_t = -F(t, X_t, p_t, q_t)\, dt + q_t dB_t, \quad t \in [0, T] \\ X_0 = x_0, \quad p_T = -\Phi_x(X_T). \end{cases} \tag{2.11}$$

where

$$F(t, x, p, q) = H_x(t, x, \vartheta(t, x, p, q), p, q).$$

After obtaining the adapted solution (X_t, p_t, q_t) of the FBSDE, the optimal control is reconstructed via the feedback map. Consequently, the pair $(\bar{X}_t, \bar{u}_t)$ solves the original stochastic control problem.

While the Stochastic Maximum Principle (SMP) characterizes optimality through the local behavior of adjoint processes along a specific trajectory, an alternative and more global perspective is provided by the Dynamic Programming Principle (DPP), which leads us to the derivation of the Hamilton-Jacobi-Bellman (HJB) equation.

2.3 Dynamic Programming and the HJB Equation

2.3.1 Dynamic Programming Principle

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space satisfying the usual condition, on which is defined an m-dimensional standard Brownian motion. Let $T > 0$ and a metric space U. For any $(s, y) \in [0, T) \times \mathbb{R}^d$, consider the following stochastic controlled system:

$$\begin{cases} dX_t = b(t, X_t, u_t)\ dt + \sigma(t, X_t, u_t)\ dB_t, & t \in [s, T] \\ X_s = y, \end{cases} \tag{2.12}$$

along with the cost functional:

$$J(s, y; u) = \mathbb{E}[\int_s^T l(t, X_t, u_t)\ dt + \Phi(X_T)]. \tag{2.13}$$

Fixing $s \in [0, T)$, we denote the weak control set by $\mathcal{U}_w[s, T]$. In the weak formulation, the admissible controls are described by 5-tuples $(\Omega, \mathcal{F}, \mathbb{P}, B, u)$, where the probability space, the Brownian motion, and the control process are considered as part of the optimization variables. The optimal control problem is:
(S_{sy}) For given $(s, y) \in [0, T) \times \mathbb{R}^d$, find a $\bar{u} \in \mathcal{U}_w[s, T]$ such that

$$J(s, y; \bar{u}) = \inf_{u \in \mathcal{U}_w[s,T]} J(s, y; u). \tag{2.14}$$

Assumptions:[65]
Consider the following assumptions:

(H.a) (U, d) is a polish space and $T > 0$.

(H.b) The maps $b : [0, T] \times \mathbb{R}^d \times U \to \mathbb{R}^d$, $\sigma : [0, T] \times \mathbb{R}^d \times U \to \mathbb{R}^{d \times m}$, $l : [0, T] \times \mathbb{R}^d \times U \to \mathbb{R}$ and $\Phi : \mathbb{R}^d \to \mathbb{R}$ are uniformly continuous, and there exists a constant $L > 0$ such that for $\varphi(t, x, u) = b(t, x, u), \sigma(t, x, u), l(t, x, u), \Phi(x)$, we have

$$\begin{cases} |\varphi(t, x, u) - \varphi(t, \hat{x}, \hat{u})| \le L|x - \hat{x}|, & \forall t \in [0, T], \quad x, \hat{x} \in \mathbb{R}^d, \quad u \in U, \\ |\varphi(t, 0, u)| \le L, \qquad \forall (t, u) \in [0, T] \times U. \end{cases} \tag{2.15}$$

Under **(H.a)-(H.b)**, for any $(s, y) \in [0, T) \times \mathbb{R}^d$ and $u(\cdot) \in \mathcal{U}_w[s, T]$, (2.12) admits a unique solution $X(\cdot) = X(\cdot; s, y, u(\cdot))$. The value function is defined by:

$$\begin{cases} V(s,y) = \inf_{u \in \mathcal{U}_w[s,T]} J(s,y;u), \quad \forall (s,y) \in [0,T) \times \mathbb{R}^d, \\ V(T,y) = \Phi(y), \quad \forall y \in \mathbb{R}^d. \end{cases} \tag{2.16}$$

The Dynamic Programming Principle (DPP) constitutes the fundamental tool for analyzing stochastic optimal control problems. It formalizes the concept of time-consistency of optimal strategies. The following theorem establishes the stochastic version of Bellman's principle of optimality.

Theorem 2.3.1. *[65]**(Principle of optimality)** Under assumptions (H.a)-(H.b), for any* $(s, y) \in [0, T) \times \mathbb{R}^d$*, the value function satisfies:*

$$V(s,y) = \inf_{u \in \mathcal{U}_w[s,T]} \mathbb{E}\Big[\int_s^{\hat{s}} l(t, X_t^{s,y,u}, u_t)\, dt + V(\hat{s}, X_{\hat{s}}^{s,y,u}) \Big], \qquad \forall 0 \le s \le \hat{s} \le T. \tag{2.17}$$

While the above theorem holds for any admissible control, the following, along with the optimal trajectory (optimal pair), the value function admits a martingale representation.

Theorem 2.3.2. *[65] Let (H.a)-(H.b) hold. if* $(\bar{X}, \bar{u})$ *is optimal for problem* (S_{sy})*, then*

$$V(t, \bar{X}_t) = \mathbb{E}\Big[\int_t^T l(r, \bar{X}_r, \bar{u}_r)\, dr + \Phi(\bar{X}_T) \mid \mathcal{F}_t^s \Big], \qquad \mathbb{P} - a.s, \quad \forall t \in [s, T]. \tag{2.18}$$

2.3.2 The Hamilton-Jacobi-Bellman Equation

The Dynamic Programming Principle provides the analytical foundation for deriving the Hamilton-Jacobi-Bellman (HJB) equation, which is a nonlinear partial differential equation whose solution coincides with the value function of a stochastic control problem. In what follows, we adopt the formulation introduced by Yong and Zhou [65] and derive the HJB equation in its Hamiltonian representation.

In what follows, we omit the dependence of the solution of the PDE on its variables for simplicity.

Proposition 2.3.1. *[65] Assume (H.a)-(H.b) hold and that* $V \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R}^n)$*. Then the value function satisfies the terminal value problem of a second-order partial differential equation:*

$$\begin{cases} -V_t + \sup_{u \in U} G(t, x, u, -V_x, -V_{xx}) = 0, \quad (t,x) \in [0,T) \times \mathbb{R}^d, \\ V(T,x) = \Phi(x), \end{cases} \tag{2.19}$$

where

$$G(t,x,u,p,P) = \frac{1}{2} Tr(P\sigma(t,x,u)\sigma(t,x,u)^\top) + \langle p, b(t,x,u) \rangle - l(t,x,u). \tag{2.20}$$

Note that $p = -V_x, \ P = -V_{xx}$.

Equivalently, defining the Hamiltonian in minimization form:

$$H(t,x,u,V_x,V_{xx}) = \frac{1}{2} Tr(\sigma(t,x,u)\sigma(t,x,u)^\top V_{xx}) + \langle b(t,x,u), V_x \rangle + l(t,x,u).$$

Then our HJB equation becomes:

$$\begin{cases} V_t + \inf_{u \in U} H(t, x, u, V_x, V_{xx}) = 0, \\ V(T, x) = \Phi(x), \end{cases} \tag{2.21}$$

where V_x and V_{xx} are the gradient and the Hessian of V, respectively.
Derivation of the HJB equation (Sketch of the proof):
The proof proceeds in two parts, establishing respectively the inequalities ≥ 0 and ≤ 0 in the HJB equation, which together yield the result.

1. **The inequality** $-V_t + \sup_{u \in U} G \geq 0$.
   Fix $(t, x) \in [0, T) \times \mathbb{R}^d$ and let $\delta > 0$ be sufficiently small, and let $u \in U$ and let $X(\cdot)$ be the state trajectory corresponding to the control $u_t \equiv u \in \mathcal{U}_w$. Applying the DPP (Theorem 2.3.1) with $s = t, \hat{s} = t + \delta$:
   $$V(t, y) \leq \mathbb{E}\Big[\int_t^{t+\delta} l(s, X_s, u_s) \; ds + V(t + \delta, X_{t+\delta}) \Big]. \tag{2.22}$$
   Since $V \in \mathcal{C}^{1,2}$, Itô's formula applied to the function $V(s, X_s^{t,x,u})$ on $[t, t + \delta]$ gives:
   $$V(t + \delta, X(t + \delta)) = V(t, y) + \int_t^{t+\delta} V_t + \langle b, V_x \rangle + \frac{1}{2} \mathrm{Tr}(\sigma\sigma^\top V_{xx}) \, ds + M_\delta, \tag{2.23}$$
   where $M_\delta = \int_t^{t+\delta} V_x(s, X_s)\sigma(s, X_s, u) \; dB_s$ is a martingale with $\mathbb{E}[M_\delta] = 0$. Substituting (2.23) into (2.22) and canceling $V(t, y)$:
   $$0 \leq \mathbb{E}\Big[\int_t^{t+\delta} (l(s, X_s, u) + V_t(s, X_s) + \langle b, V_x \rangle + \frac{1}{2} Tr(\sigma\sigma V_{xx}))(s, X_s, u) \Big].$$
   Dividing by $\delta > 0$ and rearranging:
   $$0 \leq \frac{1}{\delta} \mathbb{E}\Big[\int_t^{t+\delta} \{-V_t + G(s, X_s, u, -V_x, -V_{xx})\}(s, X_s, u) \Big].$$
   Letting $\delta \to 0$:
   $$0 \leq -V_t(s, y) + G(t, y, u, -V_x(t, y), -V_{xx}(t, y)).$$
   Taking the supremum over U:
   $$0 \leq -V_t(s, y) + \sup_{u \in U} G(t, y, u, -V_x(t, y), -V_{xx}(t, y)).$$

2. **The inequality** $-V_t + \sup_{u \in U} \leq 0$.
   Let $\epsilon > 0$. By the definition of the infimum in the DPP, for $\delta > 0$ small enough, there exists an optimal control u^ϵ such that:
   $$V(t, y) + \epsilon\delta \geq \mathbb{E}\Big[\int_t^{t+\delta} l(s, X_s^\epsilon, u_s^\epsilon) \; ds + V(t + \delta, X_{t+\delta}^\epsilon) \Big]. \tag{2.24}$$

Where X^ϵ is the state trajectory corresponding to u^ϵ. Applying Itô's formula to $V(s, X_s^\epsilon)$, taking expectations and rearranging:

$$-\epsilon \leq \frac{1}{\delta}\mathbb{E}\bigg[\int_t^{t+\delta} \{-V_t + G(s, X_s^\epsilon, u^\epsilon, -V_x, -V_{xx})\}\, ds\bigg].$$

Letting $\delta \to 0$, the uniform continuity of b, σ, l in (H.b) ensures:

$$\lim_{\delta\to 0} \sup_{y\in\mathbb{R}^d}, u \in U\ |\varphi(s, y, u) - \varphi(t, y, u)| = 0, \quad \varphi \in \{b, \sigma, l\}, \tag{2.25}$$

the integrand converges uniformly. Passing to the limit:

$$-\epsilon \leq -V_t(t, y) + \sup_{u\in U} G(t, y, u, -V_x(t, y), -V_{xx}(t, y)).$$

Since $\epsilon > 0$, letting $\epsilon \to 0^+$:

$$0 \geq -V_t(t, y) + \sup_{u\in U} G(t, y, u, -V_x(t, y), -V_{xx}(t, y)). \tag{2.26}$$

3. **Conclusion:** Combining (1) and (2.26):

$$-V_t(t, y) + \sup_{u\in U} G(t, y, u, -V_x(t, y), -V_{xx}(t, y)) = 0, \qquad \forall (t, y) \in [0, T) \times \mathbb{R}^d. \tag{2.27}$$

The terminal condition $V(T, x) = \Phi(x)$ follows directly from the definition of the value function. □

2.3.3 Connection between SMP and DP equations

In this section, we establish the deep theoretical link between the Stochastic Maximum Principle (SMP) and Dynamic Programming. This connection is often referred to as the stochastic version of the method of characteristics.

Assume that the value function $V(t, x)$, derived from the HJB equation (2.19), is sufficiently smooth ($\mathcal{C}^{1,2}$), (Further, if $V \in \mathcal{C}^{1,3}([0, T] \times \mathbb{R}^d)$ and V_{tx} is also continuous), we can identify the adjoint processes (p_t, q_t) of the SMP as follows:

- p_t corresponds to the spatial gradient of the value function evaluated along the optimal state trajectory $\bar{X}_t$:

$$-p_t = V_x(t, \bar{X}_t), \quad \forall t \in [s, T], \quad \mathbb{P}\text{−a.s.}$$

- q_t captures the diffusion term and is related to the Hessian of the value function:

$$-q_t = V_{xx}(t, \bar{X}_t)\sigma(t, \bar{X}_t, \bar{u}_t), \quad \text{a.e.} t \in [s, T], \quad \mathbb{P}\text{−a.s.}$$

where V_x, V_{xx} denotes the gradient of V with respect to x and Hessian matrix, respectively.

This identification transforms the adjoint BSDE into an identity that follows from the HJB equation. For the more comprehensive, we refer the reader to Chapter 5 [65]

Verification Theorem

The Verification Theorem establishes that a smooth solution of the HJB equation is indeed the optimal value function of a stochastic control problem. While the HJB equation provides necessary conditions that the value function must satisfy, the verification theorem proves that any function satisfying the HJB equation is sufficient to represent the true optimal value.
Then, the classical stochastic verification theorem, following Yong and Zhou [65], can be stated as follows.

Theorem 2.3.3. *[65] Suppose the assumptions hold (H.a-H.b). Let $v \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R}^n)$ be a solution of the HJB equation. Then for any admissible control $u(\cdot) \in \mathcal{U}_w[s,T]$ and any initial condition $(s,y) \in [0,T] \times \mathbb{R}^n$,*

$$v(s,y) \leq J(s,y;u(\cdot)).$$

Moreover, an admissible pair $(\bar{x}(\cdot), \bar{u}(\cdot))$ is optimal for the stochastic control problem if and only if

$$\begin{aligned} v_t(t,\bar{x}(t)) &= \max_{u \in U} G(t,\bar{x}(t),u,-v_x(t,\bar{x}(t)),-v_{xx}(t,\bar{x}(t))) \\ &= G(t,\bar{x}(t),\bar{u}(t),-v_x(t,\bar{x}(t)),-v_{xx}(t,\bar{x}(t))), \\ & a.e. t \in [s,T], \ \mathbb{P}-a.s. \end{aligned}$$

Sketch of the proof
Let $u(\cdot)$ be any admissible control and $x(t)$ the corresponding state trajectory. Applying Itô's formula to the process $v(t,x(t))$ from s to T, we obtain

$$\begin{aligned} v(s,y) = \mathbb{E}\big[\Phi(x(T))\big] - \mathbb{E}\int_s^T \Big\{ & v_t(t,x(t)) + \langle v_x(t,x(t)),\, b(t,x(t),u(t))\rangle \\ & + \frac{1}{2}\operatorname{Tr}\Big(\sigma(t,x(t),u(t))^\top v_{xx}(t,x(t))\,\sigma(t,x(t),u(t))\Big)\Big\}\,dt \end{aligned} \tag{2.28}$$

$$= J(s,y;u(\cdot)) + \mathbb{E}\int_s^T \Big\{ -v_t(t,x(t)) + G\big(t,x(t),u(t),-v_x(t,x(t)),-v_{xx}(t,x(t))\big)\Big\}\,dt$$

$$\leq J(s,y;u(\cdot)) + \mathbb{E}\int_s^T \Big\{ -v_t(t,x(t)) + \sup_{u \in U} G\big(t,x(t),u,-v_x(t,x(t)),-v_{xx}(t,x(t))\big)\Big\}\,dt = J(s,y;u(\cdot)).$$

Therefore, $v(s,y) \leq J(s,y;u(\cdot))$, and if there exists a control u^* that achieves the supremum in the Hamiltonian G, then the inequality becomes an equality. Consequently,

$$v(s,y) = J(s,y;u^*(\cdot)) = V(s,y),$$

which shows that v coincides with the value function and that $u^*(\cdot)$ is an optimal control. □
The Verification theorem plays a fundamental role in stochastic optimal control theory. It proves that a solution to the nonlinear HJB equation is actually the optimal value function V. Furthermore, it identifies the optimal control as the one that maximizes the generalized Hamiltonian G. In this way, the theorem completes the dynamic programming framework by establishing the global optimality of the candidate solution. While the above theorem requires $v \in \mathcal{C}^{1,2}$, in particular, value functions are not smooth. Thus, the theory of Viscosity solution is necessary, which we explore in the following.

Viscosity Solutions

In stochastic optimal control problems, the value function V is generally not smooth enough to satisfy the HJB equation in the classical sense. Although V satisfies the properties (continuity and growth)[1], it often fails to be of class $\mathcal{C}^{1,2}$. Consequently, the first and second-order derivatives appearing in the HJB equation (2.19) do not exist in the classical sense. To overcome this difficulty, the notion of viscosity solutions is introduced. This framework allows us to satisfy the HJB in a weak sense, and the value function is characterized as the unique solution of the corresponding HJB equation. For more information on viscosity solution, see Fleming and Soner [24] and the references therein.

Definition 2.3.1. *[65] A function $v \in \mathcal{C}([0,T] \times \mathbb{R}^n)$ is called a viscosity subsolution of (2.19) if*

$$v(T,x) \leq \phi(x), \quad \forall x \in \mathbb{R}^n, \tag{2.29}$$

for any smooth test function $\varphi \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R}^n)$, such that $v - \varphi$ has a local maximum at $(t,x) \in [0,T) \times \mathbb{R}^n$, we have

$$-\varphi_t + \sup_{u \in U} G(t,x,u,-\varphi_x(t,x),-\varphi_{xx}(t,x)) \leq 0. \tag{2.30}$$

The function v is called a viscosity supersolution if the inequalities (2.29)-(2.30) the " $\leq$ " are changed to " $\geq$ " and "local maximum" changed to "local minimum". Further, if v is both a viscosity subsolution and a viscosity supersolution of (2.19), then it is called a viscosity solution.

Then we have the following theorem.

Theorem 2.3.4. *[65] Suppose (H.a)-(H.b) holds. Then the value function V is a viscosity solution of the HJB equation.*

For the proof, see [65] and for a comprehensive treatement see [24]

> The analysis of the stochastic control problem throughout this chapter has established a robust theoretical triangle between Dynamic Programming, the Stochastic Maximum Principle, and Forward-Backward Stochastic Differential Equations.
> The Stochastic Maximum Principle provided necessary optimality conditions through Hamiltonian maximization. On the Dynamic Programming side, the HJB equation was derived from the principle of optimality as a high-dimensional semilinear parabolic PDE encoding the value function. The power of the bridge between these two approaches was made explicit through the identification of the adjoint process derived from the SMP, which represents the spatial gradient of the value function obtained from the HJB equation. Having established that solving a stochastic optimal control problem is equivalent to solving a system of FBSDEs, the next phase of this work focuses on the numerical approximation of these FBSDEs.

[1] see Proposition 3.1 in [65] for the properties of the value function.

Chapter 3

Numerical Methods and Simulations

In general, stochastic differential equations and stochastic control problems do not admit explicit analytical solutions. For this reason, numerical methods are required to approximate their solutions. In this Chapter, we first present numerical schemes for SDEs, namely the Euler-Maruyama and Milstein methods. We then introduce a numerical approach for BSDEs based on the Least Squares Monte Carlo method. Finally, extend the numerical study to a fully coupled FBSDE system.

3.1 Discretization schemes for FBSDEs

Numerical methods for SDEs

Euler-Maruyama Scheme

Consider the following SDE:

$$d\hat{X}_t = b(t, \hat{X}_t)\ dt + \sigma(t, \hat{X}_t)\ dB_t,$$

with initial condition $\hat{X}_0 = x_0$. The Euler-Maruyama method is a numerical approximation technique used for solving FSDEs, developed by the mathematician *G. Maruyama* as an extension of the Euler method, as a numerical integration methodology for estimating solutions for a system of SDEs from a given initial value $X_0 = x_0$. To apply the numerical method over the interval $[0, T]$, we first discretize the integral. Let $0 = t_0 < t_1 < \cdots < t_n = T$ be a partition with constant step size $\Delta t = t_{i+1} - t_i = T/n, \quad i = 0, 1, \ldots, n-1.$ which implies that $t_{i+1} = t_i + \Delta t = i\Delta t$. The corresponding Brownian increments are defined as $\Delta B_i = B(t_i + \Delta t) - B(t_i) = B_{t_{i+1}} - B_{t_i}$. Due to the fundamental properties of Brownian motion[1], the sequence $\{\Delta B_i\}_{i=0}^{n-1}$ forms a collection of independent and identically distributed random variables. Moreover, since each increment corresponds to a time interval of length Δt, it follows that

$$\Delta B_i \sim\ \mathcal{N}(0, \Delta t).$$

[1]Specifically, the independence and stationarity of its increments.

For each trajectory, the approximation $X_{t_{i+1}}$ is computed from the previous value X_{t_i} alone via the recurrence

$$X_{t_{i+1}} = X_{t_i} + b(t_i,\, X_{t_i})\,\Delta t + \sigma(t_i,\, X_{t_i})\,\Delta B_i, \qquad i = 0, 1, \ldots, n-1.$$

For clarity, the Euler-Maruyama algorithm is summarized below [64].

Convergence analysis

Algorithm 1 Euler–Maruyama Method

Define drift $b(t, X(t))$ and diffusion $\sigma(t, X(t))$
Input: Time horizon T, number of steps N
$\Delta t = T/N$
$t_0 = 0$, $X_{t_0} = x_0$
for $i = 0$ to $N-1$ **do**
  $\xi_i \sim \mathcal{N}(0, 1)$
  $X_{t_{i+1}} = X_{t_i} + b(t_i, X_{t_i})\,\Delta t + \sigma(t_i, X_{t_i})\sqrt{\Delta t}\,\xi_i$
  $t_{i+1} = t_i + \Delta t$
  Print t_{i+1} and $X_{t_{i+1}}$
end for

Two distinct notions of convergence are standard in the numerical analysis of SDEs, reflecting different uses of the approximation [30, 39].

Strong convergence: The scheme $\{X_{t_i}\}$ is said to converge *strongly* with order $\gamma > 0$ at time T if there exists a positive constant C, independent of Δt, such that

$$\mathbb{E}[|\hat{X}_T - X_{t_N}|] \leq C\Delta t^{\gamma}.$$

Weak convergence: The scheme converges *weakly* to X at time T as Δ with order $\beta > 0$, if for every sufficiently smooth test function $g \in \mathcal{C}^{(\beta+1)}(\mathbb{R}, \mathbb{R})$, there exists $C > 0$, which does not depend on Δt such that

$$|\mathbb{E}[g(\hat{X}_T)] - \mathbb{E}[g(X_{t_N})]| \leq C\Delta t^{\beta}.$$

Although the Euler-Maruyama method has a weak convergence of order 1, the strong convergence order is $1/2$. This fact was proved in Gikhman and Skorokhod (1972) under appropriate conditions on the functions b and σ. To build a strong order 1 method for SDEs, another term in the "stochastic Taylor series"[2] must be added to the method [60].

While the Euler approximation is the simplest useful time-discrete approximation, it is generally not particularly efficient numerically. We shall thus present another time-discrete approximation.

[2] We refer the reader to Chapter 5 [39] for a good comprehensive treatment.

Milstein scheme

The Milstein scheme proposed by Milstein, which turns out to converge strongly of order 1, the approximating trajectory has the following form

$$X_{t_{i+1}} = X_{t_i} + b(t_i,\, X_{t_i})\,\Delta t_i + \sigma(t_i,\, X_{t_i})\,\Delta B_i + \frac{1}{2}\sigma(t_i,\, X_{t_i})\partial_x\sigma(t_i,\, X_{t_i})(\Delta B_i^2 - \Delta t_i) \qquad i = 0, 1, \ldots, n-1. \tag{3.1}$$

Note that the Milstein method is identical to the Euler-Maruyama method if the diffusion term does not depend on X.

Convergence analysis and comparison on linear SDE

To illustrate various methods, we examine a simple example in some detail. We shall consider the following linear SDE:

$$dX(t) = aX(t)\;dt + bX(t)dB(t), \tag{3.2}$$

with $X(0) = x_0$ for $t \in [0, T]$ with the initial value $x_0 \in \mathbb{R}$. We know that this SDE has the explicit solution

$$X(t) = x_0 \exp\left((a - \frac{1}{2}b^2)t + bB(t)\right). \tag{3.3}$$

Knowing the solution explicitly gives us the possibility of comparing the Euler-Maruyama and Milstein approximations with the exact solution and calculating the error. To simulate a trajectory of the Euler approximation for a given time discretization, we simply start from the initial value $Y(0) = X(0)$, and proceed recursively to generate the next value

$$Y_{i+1} = Y_i + aY_i\Delta_i + bY_i\Delta B_i, \quad i = 0, 1, 2, \ldots \tag{3.4}$$

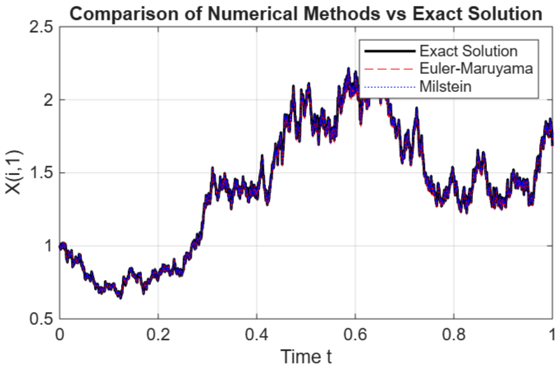
The Milstein method becomes

$$Y_{i+1} = Y_i + aY_i\Delta_i + bY_i\Delta B_i + \frac{1}{2}b^2 Y_i(\Delta B_i^2 - \Delta t_i). \tag{3.5}$$

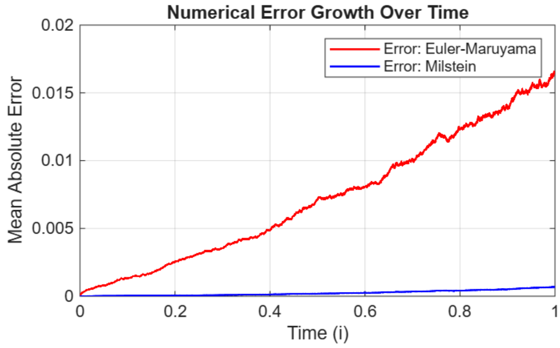
Table 3.1: Average error at $T = 1$ of approximate solutions of (3.2). The error scales as $\Delta t^{1/2}$ for Euler–Maruyama and Δt for Milstein.

Δt	Euler–Maruyama	Milstein
10^{-1}	0.566461	0.430923
10^{-2}	0.237800	0.107478
10^{-3}	0.053800	0.006502
10^{-4}	0.016347	0.000750

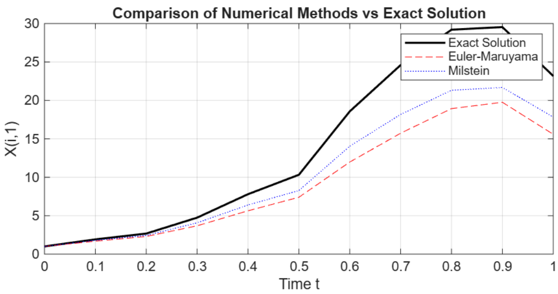
Note: Table 3.1 shows the average error at final time $T = 1$ for both numerical schemes. We observe that the Euler–Maruyama method exhibits a convergence rate consistent with $\mathcal{O}(\Delta t^{1/2})$, while the Milstein method achieves a higher convergence rate of $\mathcal{O}(\Delta t)$, confirming its improved accuracy.



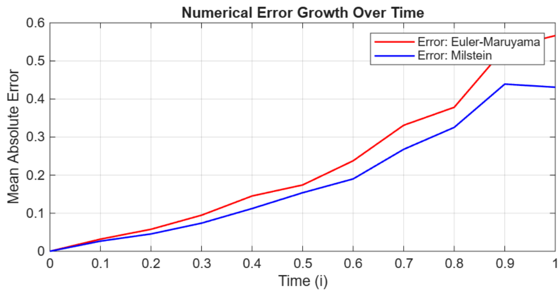
(a) Comparison ($N = 10000$)



(b) Numerical Error ($N = 10000$)



(c) Comparison ($N = 10$)



(d) Numerical Error ($N = 10$)

Figure 3.1: Comparison of Euler-Maruyama and Milstein schemes with the exact solution, and corresponding numerical error evolution for different discretization levels.

N here represents the number of subdivisions of the time interval, the initial condition $X_0 = 1$, and the coefficients $a = 1.5,\ b = 0.8$.

Note: Illustrate the performance of the Euler-Maruyama and Milstein schemes in approximating the exact solution for two discretization levels, $N = 10000$ and $N = 10$. For the finer discretization, both methods closely follow the exact trajectory, with the Milstein scheme exhibiting a smaller numerical error over time. In contrast, for the coarse discretization, the approximation errors become more pronounced, particularly for the Euler-Maruyama method. These results confirm that reducing the time step significantly improves the approximation accuracy for both schemes. Furthermore, the Milstein method consistently produces a lower mean absolute error due to the inclusion of higher-order stochastic correction terms, highlighting its strong convergence behavior.

3.2 Backward numerical methods

Due to the complexity of BSDEs, it is generally not possible to obtain an analytical solution. Consider the BSDE in the integral form:

$$\hat{Y}(t) = \Phi(\hat{X}(T)) - \int_t^T f(s, \hat{X}_s, \hat{Y}_s, \hat{Z}_s)\, ds + \int_t^T \hat{Z}_s\, dB_s,$$

Therefore, one must resort to numerical methods to approximate processes (Y, Z), such as the backward Euler method, which works backward in time. In the present context, there are certainly two general categories of explicit and implicit discretization schemes for the FBSDE [14], which can be summarized as follows

Algorithm 2 Backward Scheme for BSDE

Initialization: Approximate the terminal condition $Y^n_{t_m} = \Phi(X^m_{t_n})$ with the Euler-Maruyama scheme X^m.

for $k = (n-1)$ **to** 0 **do**

$$Z^m_{t_k} = \frac{1}{t_{k+1} - t_k} \mathbb{E}\left[Y^m_{t_{k+1}} (B_{t_{k+1}} - B_{t_k}) \,\middle|\, \mathcal{F}_{t_k} \right], \tag{3.6}$$

$$Y^m_{t_k} = \begin{cases} \mathbb{E}\left[Y^m_{t_{k+1}} + f(t_k, X^m_{t_k}, Y^m_{t_{k+1}}, Z^m_{t_k})(t_{k+1} - t_k) \,\middle|\, \mathcal{F}_{t_k} \right], & \text{(explicit)}, \\ \mathbb{E}\left[Y^m_{t_{k+1}} \,\middle|\, \mathcal{F}_{t_k} \right] + f(t_k, X^m_{t_k}, Y^m_{t_k}, Z^m_{t_k})(t_{k+1} - t_k), & \text{(implicit)}. \end{cases} \tag{3.7}$$

end

The implicit scheme often provides better properties and performance relative to the explicit scheme, with these benefits coming in exchange for additional computing effort for solving the defining for $Y^n_{t_k}$.
To discretize the BSDE, we introduce the notation, $Y_i := Y_{t_i}, \quad Z_i := Z_{t_i}$, for a given discrete time grid, recalling that the adapted BSDE solutions satisfy, $Y_s = \mathbb{E}[Y_s \mid \mathcal{F}_s], \quad Z_s = \mathbb{E}[Z_s \mid \mathcal{F}_s]$, the backward equation can be approximated by

$$Y_i = \mathbb{E}[Y_{i+1} + f(t_{i+1}, X_{i+1}, Y_{i+1}, Z_{i+1})\Delta t_i \mid \mathcal{F}_{t_i}], \quad i = N-1, \cdot, 0. \tag{3.8}$$

Moreover, the stochastic integral term disappears under conditional expectation since the Brownian increment ΔB_i has zero mean. By virtue of the nonlinear Feynman-Kac representation 1.35, the process Z corresponds to the term $\sigma^\top(s, X^{t,x}_s) v_x(s, X^{t,x}_s)$. Therefore, the process $Z_i = \mathbb{E}[Z_i \mid \mathcal{F}_{t_i}] = \mathbb{E}[\sigma^\top(t_i, X_i) v_x(t_i, X_i) \mid \mathcal{F}_{t_i}] = \sigma^\top(t_i, X_i) v_x(t_i, X_i)$. The backward iteration is initialized by the terminal condition $Y_T = \Phi(X_T)$ and $Z_T = \sigma^\top \Phi_x(X_T)$ for a $\Phi(\cdot)$ is differentiable almost everywhere. To approximate the condition expectation (3.8), we employ the *Least-Square Monte Carlo* (LSMC) method adopted from [22], which we briefly review in what follows.

Least Square Monte Carlo

The LSMC method addresses the numerical approximation of conditional expectations of the form $\mathbb{E}[Y \mid X]$, for square integrable random variables X and Y. The conditional expectation can be characterized as the best approximation in L^2, that is, $\mathbb{E}[Y \mid X] = \phi^*(X)$, where ϕ^* solves the infinite-dimensional minimization problem

$$\phi^* = \arg\min_{\phi} \mathbb{E}[|\phi(X) - Y|^2],$$

and ϕ ranges over all measurable functions with $\mathbb{E}[|\phi(X)|^2] < \infty$. A finite-dimensional approximation of this problem is obtained by representing the function $\phi(\cdot)$ as a linear combination of predetermined basis functions, $\phi(x) = \sum_{k=1}^{K} \alpha_k \varphi_k(x) = \varphi(x)\alpha$, and $\alpha = (\alpha_1, \dots, \alpha_k)$ is the vector of unknown coefficients, k being the dimension of the basis. Monte Carlo simulation is then used to generate M independent trajectories of the forward process, $\{X_i^{(m)}\}_{i=0}^{N}, m = 1, \dots, M$. At each time step t_i, the conditional expectation is approximated through linear regression using the simulated samples, specifically, the coefficients are computed by solving the least-squares problem

$$\alpha_i^* = \arg\min_{\alpha_i} \frac{1}{M} \sum_{m=1}^{M} |\varphi_i(X^{(m)})\alpha_i - Y_i^{(m)}|^2.$$

Equivalently, introducing the matrix

$$A_i = \Big(\varphi_k(X_i^{(m)})\Big)_{m=1,\dots,M;\ k=1,\dots,K},$$

and the vector

$$b_i = (Y_{i+1}^{(m)} + \Delta t_i f(t_{i+1}, X_i^{(m)}, Y_{i+1}^{(m)}, Z_{i+1}^{(m)}))_{m=1,\dots,M},$$

the regression problem can be written in matrix form as

$$\alpha_i^* = \arg\min_{\alpha \in \mathbb{R}^K} \|A_i \alpha - b_i\|^2.$$

Assuming that the coefficient matrix $A_i \in \mathbb{R}^{M \times K}$, $K \leq M$ has maximum rank K, then the solution is given by

$$\alpha_i^* = (A_i^\top A_i)^{-1} A_i^\top b_i.$$

Linear BSDEs: Now we present the special case of linear BSDEs [13], which means that the generator f depends linearly on Y and Z. We consider the case $k = 1$, and Y is real-valued and Z is a matrix of size $1 \times d$.

Proposition 3.2.1. *[13] Let $\{(a_t, b_t)\}_{t\in[0,T]}$ be a progressively measurable and bounded process taking values in $\mathbb{R} \times \mathbb{R}^d$. Let $\{C_t\}_{t\in[0,T]}$ belong to $M^2(\mathbb{R})$, and let ξ be an $\mathcal{F}_T-$measurable square-integrable real random variable. The linear BSDE*

$$Y_t = \xi + \int_t^T \{a_r Y_r + Z_r b_r + c_r\}\, dr - \int_t^T Z_r dB_r, \tag{3.9}$$

admits a unique solution satisfying

$$\forall t \in [0,T], \quad Y_t = \Gamma_t^{-1}\mathbb{E}\left[\xi\Gamma_T + \int_t^T c_r\Gamma_r\ dr \middle| \mathcal{F}_t\right], \tag{3.10}$$

where, for every $t \in [0,T]$,

$$\Gamma_t = \exp\left(\int_0^t b_r.dB_r - \frac{1}{2}\int_0^t |b_r|^2\ dr + \int_0^t a_r\ dr\right). \tag{3.11}$$

For the proof, we refer the reader to [13].

3.3 Numerical methods for FBSDE (coupled and decoupled)

1. **Decoupled FBSDE:** Consider the decoupled FBSDE:

$$\begin{cases} \hat{X}_t = x + \int_0^t b(s,\hat{X}_s)\,ds + \int_0^t \sigma(s,\hat{X}_s)\,dB_s, \\ \hat{Y}_t = \Phi(\hat{X}_T) - \int_t^T f(s,\hat{X}_s,\hat{Y}_s,\hat{Z}_s)\,ds + \int_t^T \hat{Z}_s\,dB_s. \end{cases} \tag{3.12}$$

We now discuss the numerical discretisation of the decoupled FBSDE, which allows it to be discretised using an explicit time-stepping scheme. The least-square Monte Carlo scheme is based on the Euler discretisation.

$$\begin{aligned} X_{i+1} &= X_i + \Delta t\ b(t_i,X_i) + \sqrt{\Delta t}\ \sigma(t_i,X_i)\xi_{i+1} \\ Y_{i+1} &= Y_i - \Delta t\ f(X_i,Y_i,Z_i) + \sqrt{\Delta t}Z_i \cdot \xi_{i+1}, \end{aligned} \tag{3.13}$$

where (X_i,Y_i,Z_i) denotes the numerical discretisation of the joint process (X_s,Y_s,Z_s), adopting the LSMC discussed above, we have the following algorithm based on [38]:

Algorithm 3 Least-squares Monte Carlo

1: Define K, M, N and $\Delta t = T/M$.
2: Set initial condition $x \in \mathbb{R}^d$.
3: Choose radial basis functions

$$\{\phi_k \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}) : k = 1, \dots, K\}.$$

4: Generate M independent realisations $X^{(1)}, \dots, X^{(M)}$ of length N from

$$X_{i+1} = X_i + \Delta t\, b(t_i, X_i) + \sqrt{\Delta t}\, \sigma(t_i, X_i)\xi_{i+1}, \qquad X_0 = x.$$

5: Initialise BSDE by

$$Y_N^{(m)} = \Phi(X_N^{(m)}), \qquad Z_N^{(m)} = \sigma(X_N^{(m)})^T \nabla\Phi(X_N^{(m)}).$$

6: **for** $n = N-1, \dots, 1$ **do**
7: Assemble linear system $A_n \hat{\alpha}(t_n) = b_n$ according to
8: Evaluate $Y_n^{(m)}$ and $Z_n^{(m)}$ according to

$$Y_n^{(m)} = \sum_{k=1}^{K} \alpha_k(t_n)\, \phi_k(X_n^{(m)}),$$

$$Z_n^{(m)} = \sigma(X_n^{(m)})^\top \sum_{k=1}^{K} \alpha_k(t_n)\, \nabla\phi_k(X_n^{(k)}).$$

9: If necessary, adapt basis functions ϕ_k.
10: **end for**

Example of decoupled FBSDE:
To illustrate the numerical method developed above, we present the classical Black-Scholes option pricing model, following [13]. In finance, an important question is determining the price of an option (a financial product). Let us take the simplest case, that of the Black-Scholes model and a "European call".
The price $(S(t))_{t\geq 0}$ is governed by a geometric Brownian motion, satisfying the following SDE:

$$dS(t) = S(t)(\mu\ dt + \sigma dB(t)), \quad S(0) = x_0,$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ denotes the constant market volatility. Applying Itô's formula, the explicit solution for the stock price at any time $t \geq 0$ is given by:

$$S(t) = x_0 \exp\left((\mu - \frac{\sigma^2}{2})t + \sigma B(t)\right).$$

Alongside this investment, there is a risk-free investment (for example, a savings account) with a constant rate of return equal to r, the price of this asset is given by:

$$dE(t) = rE(t)dt, \quad E(0) = y,$$

which admits the explicit solution

$$E(t) = ye^{rt}.$$

A strategy is defined as a pair of processes $(p(t), q(t))_{t\geq 0}$, adapted to the filtration generated by the Brownian motion B, where $q(t)$ represents the number of risk-free asset units and $p(t)$ represents the number of risky asset units, that is, the number of shares in the portfolio at time t. Therefore, the value of the portfolio at time t is:

$$V(t) = q(t)E(t) + p(t)S(t).$$

We consider self-financing strategies, which means that the evolution of the portfolio value satisfies:

$$dV(t) = q(t)dE(t) + p(t)dS(t) = rq(t)E(t)\ dt + p(t)S(t)(\mu\ dt + \sigma dB(t)).$$

Since $q(t)E(t) = V(t) - p(t)S(t)$, we can write, by denoting $\pi(t) = p(t)S(t)$, which represents the amount of money invested in stocks, we obtain

$$dV(t) = rV(t)dt + \pi(t)(\mu - r)dt + \pi(t)\sigma dB(t).$$

Setting, $Z(t) = \pi(t)\sigma, \quad \theta = \frac{\mu - r}{\sigma}$, where θ is called the risk premium, we get

$$dV(t) = rV(t)\ dt + \theta Z(t)dt + Z(t)dB(t), \tag{3.14}$$

a common problem in finance is the pricing of options. A European call option with maturity T and stock price K is a contract giving its holder the right, but not the obligation, to buy one share of the stock at price K at time T. The seller of the option must therefore pay the holder the amount $\xi := (S(T) - K)^+$, which represents the profit from exercising the option. More generally, one may consider a contingent claim whose payoff is a positive random variable ξ depending on $(S(t))_{0\leq t\leq T}$. The fundamental question is: at what initial price v should the option be sold ? The answer rests on the duplication principle. The seller receives v at time $t = 0$ and invests it in the market by following the strategy $(Z(t))_{0\leq t\leq T}$ yet to be determined, so that the resulting portfolio value satisfies the dynamics in equation (3.14) and matches the contingent claim at maturity $V(T) = \xi$. This is precisely the problem of finding an adapted pair $(V(t), Z(t))_{0\leq t\leq T}$ satisfying

$$\begin{cases} dV(t) = (rV(t) + \theta Z(t))\ dt + Z(t)dB(t), \\ V(T) = \xi. \end{cases} \tag{3.15}$$

Equation (3.15) is a BSDE, more precisely a linear BSDE, because the driver f is linear on $(V(t), Z(t))$. By coupling the forward asset price dynamics with the backward equation, the pricing problem is naturally formulated as a decoupled FBSDE system of the form:

$$\begin{cases} dS(t) = \mu S(t)dt + \sigma S(t)dB(t), \\ dV(t) = (rV(t) + \theta Z(t))\ dt + Z(t)dB(t), \\ S(0) = x_0, \quad V(T) = \xi. \end{cases} \tag{3.16}$$

Writing the integral form of equation (3.15):

$$V(t) = \xi - \int_t^T (rV(s) + \theta Z(s))\, ds - \int_t^T Z(s)dB(s).$$

By the proposition 3.2.1, the BSDE admits a unique solution, with the parameters $a_t = -r,\ b_t = -\theta,\ c_t = 0$. The proposition gives us:

$$\Gamma(t) = \exp\Big(-\theta B(t) - \frac{1}{2}\theta^2 t - rt \Big).$$

And:

$$V(t) = \Gamma(t)^{-1}\mathbb{E}[\xi\Gamma(T) \mid \mathcal{F}_t].$$

Then, with simplifying:

$$V(t) = e^{-r(T-t)} e^{\theta B(t)+\frac{1}{2}\theta^2 t}\mathbb{E}\Big[\xi e^{-\theta B(T)-\frac{1}{2}\theta^2 T} \mid \mathcal{F}_t\Big] \tag{3.17}$$

which is the exact solution of the linear BSDE of the Black-Scholes model. We now illustrate the numerical behavior of the solution through simulations. We present an approximation of one simulated trajectory of the stock price process $S(t)$, the corresponding option price process $V(t)$, and the amount of money invested in the risky asset, represented by $\pi(t)$.

Note: Figure 3.2 tracks the price fluctuations of the underlying stock over time. Because it

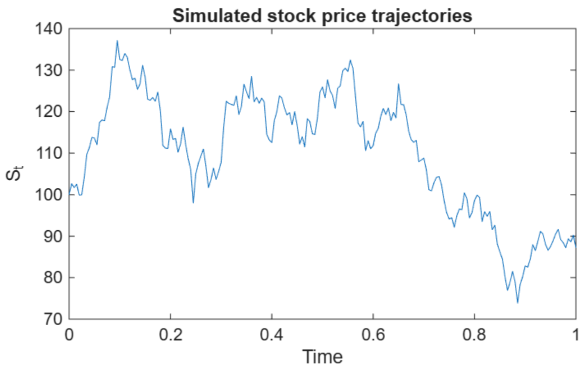


Figure 3.2: The Risky Asset (Stock Price $S(t)$)

is a risky market asset, its price moves unpredictably based on market forces.

Note: Figure 3.3 represents the price of the option contract that the buyer pays. The value initially increases as market conditions change, but ultimately decreases to zero. It hits zero because once the contract crosses its final time horizon, all uncertainty is removed-the buyer already knows the final price of the stock. If the option is out of the money at that deadline, the contract holds no value and expires worthless.

Note: Figure 3.4 represents the total monetary amount invested in the risky asset at any given time (Quantity of Shares × Stock Price). This graph explicitly outlines the dynamic trading strategy the investor must follow. It serves as a financial roadmap, showing exactly

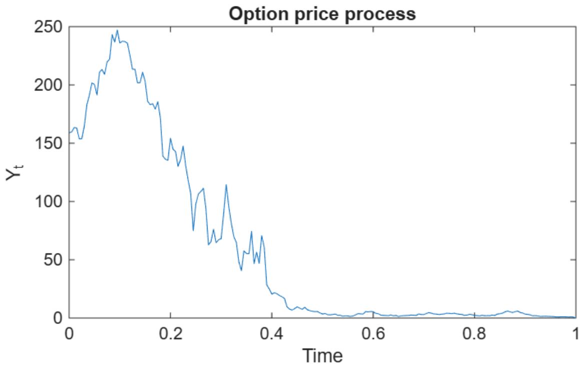


Figure 3.3: The Option Price (The Premium)

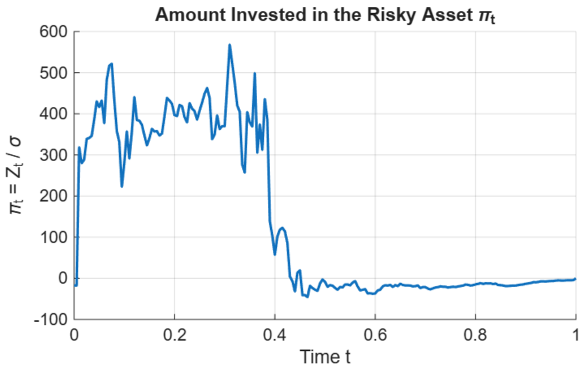


Figure 3.4: The Trading Strategy (Asset Value Held)

how much of the risky asset must be bought or sold at each moment to perfectly manage and balance the risk of the option contract.

2. **Coupled FBSDE:** The primary numerical challenge in solving the coupled FBSDE system lies in the mutual dependence between the forward and backward components. Specifically, the drift b and diffusion σ coefficients of the forward equation depend explicitly on the backward solution (Y_t, Z_t), while the driver f and the terminal condition Φ depend on the forward process X_t. The fundamental idea to solve such a system is to replace the simultaneous coupled system with a sequence of decoupled solvable problems, using the trajectories computed in the previous iteration. Rather than trying to solve for (X, Y, Z) all at once, we begin by choosing an initial guess for the backward components, denoted as $(Y_t^0, Z_t^0)_{t\in[0,T]}$. At any given iteration p, the forward SDE can be solved independently because the backward components inside its coefficients are "frozen" as known input (Y^{p-1}, Z^{p-1}) from the previous iteration. This decouples the FSDE, allowing it to be simulated using standard discretization schemes such as the Euler-Maruyama 3.1 or Milstein 3.1 schemes. Once the forward trajectory X^p is obtained, the backward equation becomes a standard BSDE for (Y_t^p, Z_t^p), which can then

be solved backward from the known terminal condition $Y_T^p = \Phi(X_T^p)$ using standard conditional expectation approximations like the LSMC method 3.2. The process repeats until the sequence of solutions converges, for a small time horizon or weakly coupled system, even a crude initialization produces rapid convergence. As summarized in the following algorithm.

Algorithm 4 Forward-backward method

Require: Time horizon T, iteration number N, initial condition x_0

1: Choose an initial guess (Y^0, Z^0)

2: **for** $p = 1, \dots, N$ **do**

3: **Forward step:** Solve

$$dX_t^p = b(t, X_t^p, Y_t^{p-1}, Z_t^{p-1})\,dt + \sigma(t, X_t^p, Y_t^{p-1}, Z_t^{p-1})\,dW_t, \quad X_0^p = x_0.$$

4: **Backward step:** Solve

$$\begin{cases} dY_t^p = -g(t, X_t^p, Y_t^p, Z_t^p)\,dt + Z_t^p\,dW_t, \\ Y_T^p = \Phi(X_T^p). \end{cases}$$

5: Update

$$(X, Y, Z) \leftarrow (X^p, Y^p, Z^p)$$

6: **end for**

7: **return** (X^N, Y^N, Z^N)

The numerical framework described above provides a robust mechanism for solving general coupled FBSDEs, but its utility is best illustrated through application to concrete dynamical systems.

Controlled SIR model

A prime example is the optimal control of epidemic dynamics, specifically the stochastic **Susceptible-Infected-Recovered** (SIR) model. Applying the SMP to this control problem naturally yields a coupled system of FBSDEs.

Mathematical models are simplified representations of how an infection spreads across a population over time, and generally come in two forms: stochastic and deterministic models [59]. In a closed population of fixed size is partitioned at each time t into three compartments: the susceptible individuals S_t, who are at risk of infection, the infected individuals I_t, who carry and transmit the disease; and the recovered individuals R_t, who have acquired immunity. The dynamics are governed by the contact rate of susceptibles with infectives (transmission) β and the recovery rate of infectives γ, and a mortality rate μ and birth rate Λ. The control $u(t)$ represents an intervention policy (e.g., vaccination or quarantine) that reduces the transmission

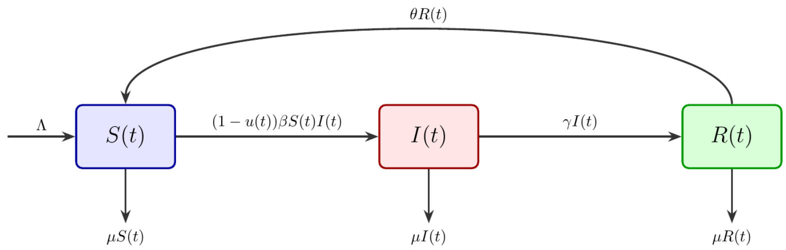


Figure 3.5: The flowchart of SIR model

rate β. The following set of SDEs describes the stochastic controlled SIR model on $t \in [s, T]$:

$$\begin{cases} dS(t) = [\Lambda - (1-u(t))\beta S(t)I(t) + \theta R(t) - \mu S(t)]\ dt + \delta_1 S(t)I(t)\ dB(t), \\ dI(t) = [(1-u(t))\beta S(t)I(t) - \gamma I(t) - \mu I(t)]\ dt + \delta_2 I(t)dB(t) \\ dR(t) = [\gamma I(t) - \theta R(t) - \mu R(t)]\ dt + \delta_3 R(t)dB(t). \end{cases} \tag{3.18}$$

where $S(s) > 0,\ I(s) > 0,\ R(s) \geq 0$. For a 1−dimensional Brownian motion. We define the vector:

$$X(t) = [S(t), I(t), R(t)]^\top.$$

The controlled system can be written as:

$$dX(t) = b(X(t), u(t))\ dt + \sigma(X(t))\ dB(t), \tag{3.19}$$

with initial condition $X(s) = [S(s), I(s), R(s)] = x_0$, where b and σ are vectors with components:

$$b_1(X(t), u(t)) = \Lambda - (1-u(t))\beta S(t)I(t) + \theta R(t) - \mu S(t),$$

$$b_2(X(t), u(t)) = (1-u(t))\beta S(t)I(t) - \gamma I(t) - \mu I(t),$$

$$b_3(X(t), u(t)) = \gamma I(t) - \theta R(t) - \mu R(t),$$

$$\sigma_1(X(t)) = \delta_1 S(t)I(t),$$

$$\sigma_2(X(t)) = \delta_2 I(t),$$

$$\sigma_3(X(t)) = \delta_3 R(t).$$

The cost functional is given as follows:

$$J(s, x_0; u) = \mathbb{E}\Big[\int_s^T (\frac{1}{2}u(t)^2 + I(t))\ dt + I(T)\Big].$$

Our goal is to find an optimal control $\bar{u} \in U$ such that:

$$J(s, x; \bar{u}) = \inf_{u \in U} J(s, x; u),$$

where U is an admissible control set defined by $U = [0, 1]$.
To solve such a problem. First, we use the stochastic maximum principle. We define the Hamiltonian H as follows:

$$H(x, u, p, q) = p^\top b + q^\top \sigma + l(x, u),$$

where the running cost $l(x, u) = \frac{1}{2}u(t)^2 + I(t)$ and terminal cost $\Phi(X(T)) = I(T)$, and:

$$p(t) = \begin{pmatrix} p_1(t) \\ p_2(t) \\ p_3(t) \end{pmatrix}, \quad q(t) = \begin{pmatrix} q_1(t) \\ q_2(t) \\ q_3(t) \end{pmatrix}.$$

Explicitly,

$$\begin{aligned} H = {} & \frac{1}{2}u(t)^2 + I(t) \\ & + p_1(t)\Big[\Lambda - (1 - u(t))\beta S(t)I(t) + \theta R(t) - \mu S(t)\Big] \\ & + p_2(t)\Big[(1 - u(t))\beta S(t)I(t) - \gamma I(t) - \mu I(t)\Big] \\ & + p_3(t)\Big[\gamma I(t) - \theta R(t) - \mu R(t)\Big] \\ & + q_1(t)\,\delta_1 S(t)I(t) \\ & + q_2(t)\,\delta_2 I(t) \\ & + q_3(t)\,\delta_3 R(t) \end{aligned}$$

Following the SMP, the adjoint equations are:

$$dp_i(t) = -\frac{\partial H}{\partial X_i}dt + q_i(t)dB(t), \quad i = 1, 2, 3.$$

The terminal cost is $\Phi(X(T)) = I(T)$, then:

$$p(T) = \frac{\partial \Phi(X(T))}{\partial X} = [0, 1, 0]^\top.$$

$$\begin{cases} dp_1(t) = -\{p_1(t)[-(1 - u(t))\beta I(t) - \mu] + p_2(t)[(1 - u(t))\beta I(t)] + q_1(t)\delta_1 I(t)\}\ dt + q_1(t)dB(t), \\ dp_2(t) = -\{1 - p_1(t)(1 - u(t))\beta S(t) + p_2(t)[(1 - u(t))\beta S(t) - \gamma - \mu] + p_3(t)\gamma + q_1(t)\delta_1 S(t) + q_2(t)\delta_2\}\ dt \\ +q_2(t)dB(t), \\ dp_3(t) = -\{p_1(t)\theta - p_3(t)\theta - p_3(t)\mu + q_3(t)\delta_3\}\ dt + q_3(t)dB(t). \end{cases}$$

Differentiating the Hamiltonian with respect to u, we get the optimal control

$$\bar{u} = \max\Bigg(0, \Bigg(\min(1, -\beta S(t)I(t)[p_1(t) - p_2(t)])\Bigg)\Bigg).$$

After substituting the optimal control into the state equation, we get a coupled FBSDE system:

$$\begin{cases} dS(t) = [\Lambda - (1 - \bar{u}(t))\beta S(t)I(t) + \theta R(t) - \mu S(t)]\ dt + \delta S(t)I(t)\ dB(t), \\ dI(t) = [(1 - \bar{u}(t))\beta S(t)I(t) - \gamma I(t) - \mu I(t)]\ dt + \delta_2 I(t)dB(t), \\ dR(t) = [\gamma I(t) - \theta R(t) - \mu R(t)]\ dt + \delta_3 R(t)dB(t). \end{cases}$$

with initial condition x_0, and the backward system:

$$\begin{cases} dp_1(t) = -\{p_1(t)[-(1-\bar{u}(t))\beta I(t) - \mu] + p_2(t)[(1-\bar{u}(t))\beta I(t)] + q_1(t)\delta_1 I(t)\}\ dt + q_1(t)dB(t), \\ dp_2(t) = -\{1 - p_1(t)(1-\bar{u}(t))\beta S(t) + p_2(t)[(1-\bar{u}(t))\beta S(t) - \gamma - \mu] + p_3(t)\gamma + q_1(t)\delta_1 S(t) + q_2(t)\delta_2\}\ dt \\ +q_2(t)dB(t), \\ dp_3(t) = -\{p_1(t)\theta - p_3(t)\theta - p_3(t)\mu + q_3(t)\delta_3\}\ dt + q_3(t)dB(t). \end{cases}$$

with terminal condition

$$p(T) = \begin{pmatrix} p_1(T) \\ p_2(T) \\ p_3(T) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Then the system can be written as:

$$\begin{cases} dX(t) = b(X(t), \bar{u}(t))\ dt + \sigma(X(t))dB(t), \\ dp(t) = -\nabla_x H(X(t), \bar{u}(t), p(t), q(t))\ dt + q(t)dB(t), \\ \bar{u}(t) = \arg\min_{u\in[0,1]} H(X(t), u, p(t), q(t)). \end{cases} \tag{3.20}$$

Solving the problem with HJB: We define the value function:

$$V(s, x_0) = \inf_{u\in U} \mathbb{E}\Big[\int_s^T (\frac{1}{2}u(t)^2 + I(t))\ dt + I(T) \mid X(s) = x_0\Big].$$

with terminal condition:

$$V(T, X(T)) = I(T).$$

We define the Hamiltonian:

$$H(t, x, u, \nabla_x V, \nabla_{xx} V) = l(x, u) + \langle b(x, u), \nabla_x V\rangle + \frac{1}{2}Tr(\sigma\sigma^\top \nabla_{xx} V),$$

where $\nabla_x V$, $\nabla_{xx} V$ represents the gradient and the hessian of V, respectively.
By the Dynamic Programming Principle 2.3.1 and Itô formula, the value function satisfies the following HJB equation:

$$V_t + \inf_{u\in U}\{H(t, x, u, \nabla_x V, \nabla_{xx} V)\} = 0.$$

Explicitly,

$$V_t + \inf_{u\in U}\{b_1 V_S + b_2 V_I + b_3 V_R + \frac{1}{2}[\delta_1^2 S(t)^2 I(t)^2 V_{SS} + \delta_2^2 I(t)^2 V_{II} + \delta_3^2 R(t)^2 V_{RR}] + \frac{1}{2}u(t)^2 + I(t)\} = 0.$$

Minimizing with respect to the control, we get the optimal control:

$$\bar{u}(t) = -\beta S(t) I(t)(V_S - V_I).$$

Substituting the result back into the HJB:

$$\begin{aligned}
&V_t + V_S\Big[\Lambda - \big(1 + \beta S(t)I(t)(V_S - V_I)\big)\beta S(t)I(t) - \theta R(t) - \mu S(t)\Big] \\
&\quad + V_I\Big[1 + \beta S(t)I(t)(V_S - V_I)\beta S(t)I(t) - \gamma I(t) - \mu I(t)\Big] \\
&\quad + V_R\Big[\gamma I(t) - \theta R(t) - \mu R(t)\Big] \\
&\quad + \frac{1}{2}\delta_1^2 S(t)^2 I(t)^2 V_{SS} + \delta_2^2 I(t)^2 V_{II} + \delta_3^2 R(t)^2 V_{RR} \\
&\quad + \frac{1}{2}\beta^2 S(t)^2 I(t)^2 \big(V_S^2 + V_I^2 - 2V_S V_I\big) + I(t) = 0.
\end{aligned}$$

Then, the FSDE becomes:

$$dX(t) = \begin{pmatrix} \Lambda + \beta S(t)I(t) + \theta R(t) - \mu S(t) \\ \beta S(t)I(t) - \gamma I(t) - \mu I(t) \\ \gamma I(t) - \theta R(t) - \mu R(t) \end{pmatrix} dt + \begin{pmatrix} \delta_1 S(t)I(t) \\ \delta_2 I(t) \\ \delta_3 R(t) \end{pmatrix} dB(t), \tag{3.21}$$

By nonlinear Feynman-Kac formula 1.5, we have:

$$Y(t) = V(t, X(t)), \;\; Z(t) = \sigma^\top(X(t))\nabla_x V,$$

where $Y(t) \in \mathbb{R}$ and $Z(t) \in \mathbb{R}^3$ with: $Z_1(t) = \sigma_1 V_S,\; Z_2(t) = \sigma_2 V_I,\; Z_3(t) = \sigma_3 V_R$. The backward SDE is given by:

$$\begin{aligned}
dY(t) = \Bigg\{ & I(t) - \beta^2 S(t)^2 I(t)^2 \left(\frac{Z_1(t)}{\sigma_1} - \frac{Z_2(t)}{\sigma_2}\right)\frac{Z_1(t)}{\sigma_1} \\
& + \beta^2 S(t)^2 I(t)^2 \left(\frac{Z_1(t)}{\sigma_1} - \frac{Z_2(t)}{\sigma_2}\right)\frac{Z_2(t)}{\sigma_2} \\
& + \frac{1}{2}\beta^2 S(t)^2 I(t)^2 \left(\frac{Z_1(t)^2}{\sigma_1^2} - \frac{Z_2(t)^2}{\sigma_2^2} - 2\frac{Z_1(t)Z_2(t)}{\sigma_1\sigma_2}\right)\Bigg\} dt + Z(t)\, d\tilde{B}(t).
\end{aligned} \tag{3.22}$$

Where $\tilde{B} = \begin{pmatrix} B \\ B \\ B \end{pmatrix}$. Since the resulting FBSDE system (3.21) with (3.22) is a fully coupled system through the optimal control state $\bar{u}(t)$, an analytical solution is generally unavailable. To validate the theoretical frameworks applied to this model, we perform numerical solutions.

Interpretation: Figure 3.6 represents the solution of the SDE that appears in the nonlinear Feynman-Kac formula over a horizon ($T = 1$). *The red lines (top)* represent the $S(t)$. It starts at $S(0) = 0.8$ (80%) and fluctuates with noise. *The green lines (middle)* represent the infected population $I(t)$. It starts at $I(0) = 0.13$ (13%). *The yellow lines* represent the recovered population $R(t)$. It starts at $R(0) = 0.07$ (07%).

Interpretation: Figure 3.7 represents the optimal solution and the solution without control ($u = 0$) of our SIR model. We get the optimal solution by replacing the control with the optimal one, and this is done by using the value of the components $Z(t)$ computed from our coupled FBSDE system,

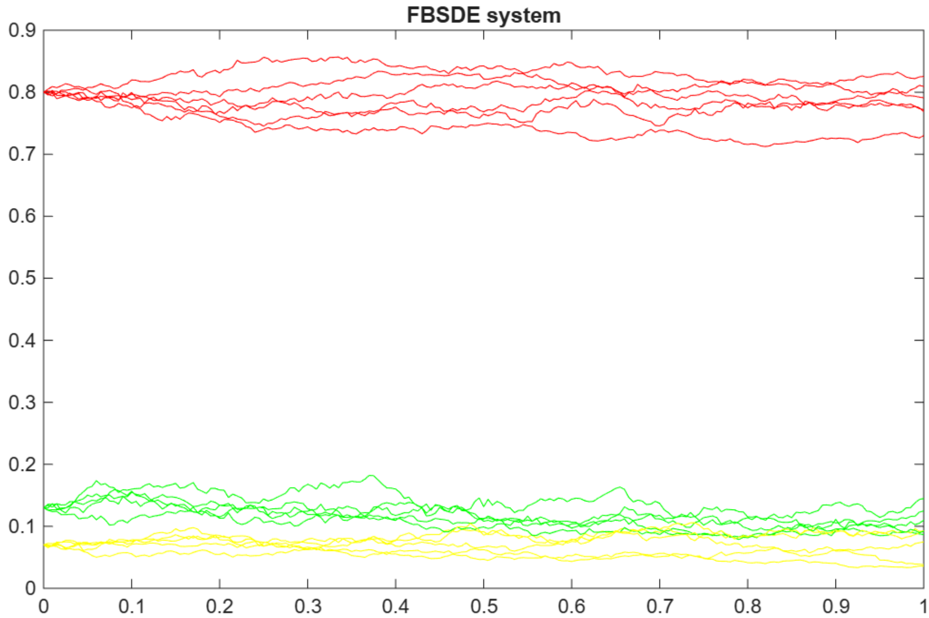


Figure 3.6: Simulation of 5 trajectories of $X(t)$

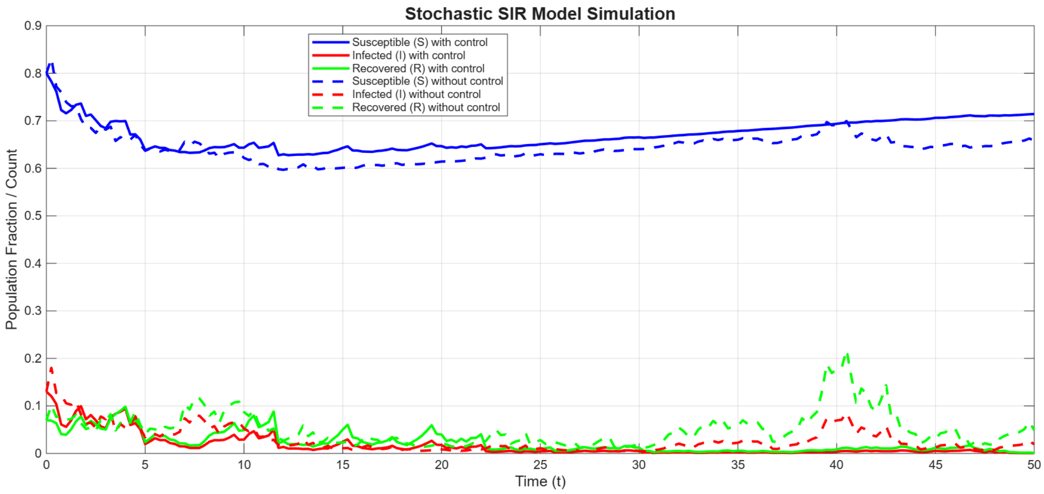


Figure 3.7: Simulation of controlled SIR model

with a horizon of $T = 50$ days. It is clear that the trajectory of the infected without control is above the trajectory of the infected with control. This shows clearly the effect of the optimal control. In addition, the graph shows that the trajectory of the optimal infection goes to zero when t grows.
Interpretation: Figure 3.8 shows that the trajectories of Y with $Y(0) \approx 0.27$ which is the value of our optimal cost. Because our problem aims to minimize the presence of the disease at the end of the horizon, the driver forces all stochastic paths of Y_t to converge precisely to the terminal cost $\Phi(X(T)) = I(T)$ (e.g the final infected). The downward trend across the trajectories indicates a

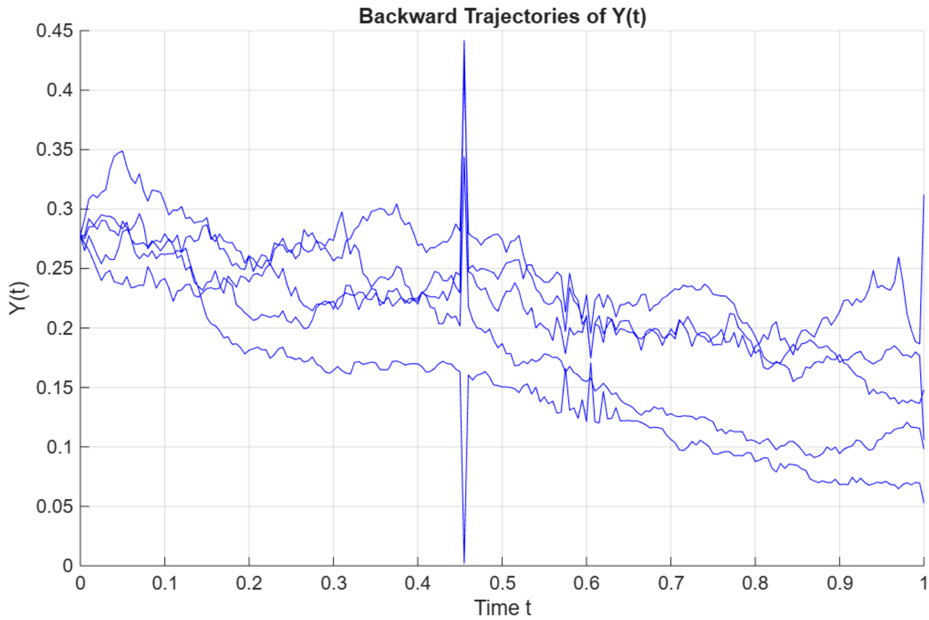


Figure 3.8: Simulations of 5 trajectories of Y_t.

decrease in the expected final risk over time under optimal control.

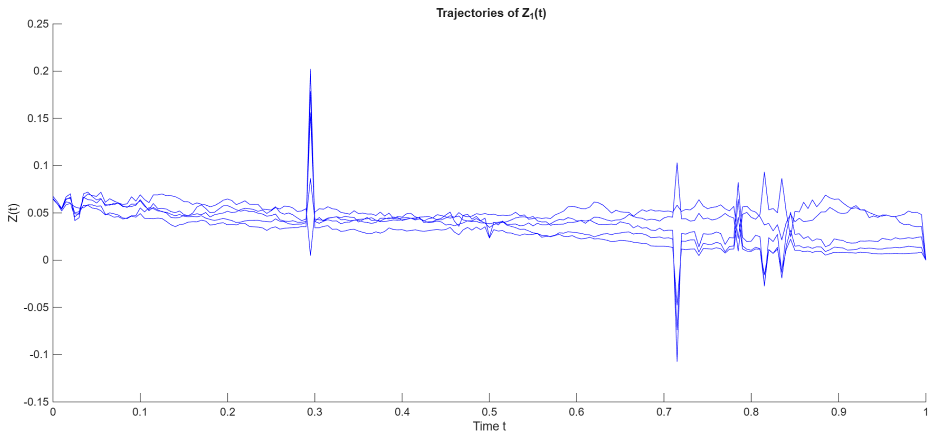


Figure 3.9: Simulations of 5 trajectories of the first component of Z_t

Interpretation: Figure 3.9 represents the trajectories of the first components of Z. This shows the effect of the optimal control because of the relation of the optimal control with Z. The larger Z_1 is, the smaller the optimal control becomes.

Interpretation: Figure 3.10 represents the trajectories of the second components of Z. This shows the effect of the optimal control because of the relation of the optimal control with Z. The larger Z_2 is, the larger the optimal control becomes.

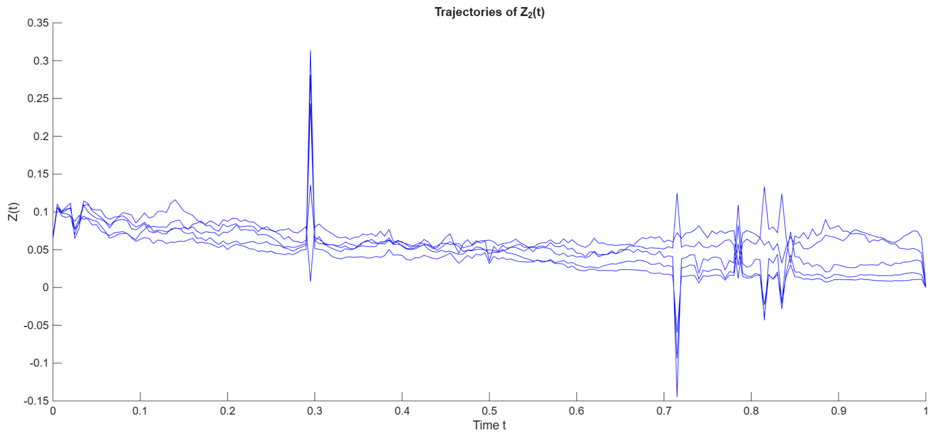


Figure 3.10: Simulations of 5 trajectories of the second component of Z_t

3.4 Linear-Quadratic (LQ) Example

Stochastic Optimal control

The following example is based on [46].

Consider the following controlled stochastic differential equation:

$$\begin{cases} dX(s) = [aX(s) + bu(s)]ds + dB(s), \quad s \in [t,T], \\ X(t) = x, \end{cases} \tag{3.23}$$

where $X(\cdot)$ is the state process, $u(\cdot)$ is the control process. Both are assumed to be $\{\mathcal{F}_t\}_{t\geq 0}$ adapted and square integrable. For simplicity, we will focus on the one-dimensional case (i.e, X, u, and B are one-dimensional, and a and b are given constants). The optimal control problem is then to minimize the quadratic cost functional subject to the state dynamics above, which is defined as follows:

$$J(t,x;u) = \frac{1}{2}\mathbb{E}\left[\int_t^T [|X(s)|^2 + |u(s)|^2]ds + |X(T)|^2\right]. \tag{3.24}$$

This problem is called *stochastic linear-quadratic (LQ) problem*, because the state is linear and the cost function is quadratic. There exists a unique solution to this optimal control problem (the mapping $u \to J(t,x;u)$ is convex). The goal is to determine this optimal control.

1. *Solving the problem via SMP:* Defining the Hamiltonian:

$$H(x,u,y,z) = \frac{1}{2}(x^2+u^2) + y(ax+bu) + z,$$

where $y(\cdot)$ and $z(\cdot)$ are the adjoint processes. By SMP, the adjoint process $Y(\cdot)$ satisfies the BSDE:

$$\begin{aligned} dY(s) &= -H_x(X(s),u(s),Y(s),Z(s))\,ds + Z(s)dB(s), \\ &= -(aY(s)+X(s))\,ds + Z(s)dB(s), \quad s \in [t,T]. \end{aligned} \tag{3.25}$$

Since the terminal cost $\Phi(X(T)) = \frac{1}{2}X(T)^2$, then the terminal condition:

$$Y(T) = \Phi_x(X(T)) = X(T).$$

The optimal control is obtained by the necessary condition:

$$\frac{\partial H}{\partial u} = 0,$$

since

$$\frac{\partial H}{\partial u} = u + bY,$$

it follows that the optimal control satisfies:

$$\bar{u}(t) = -bY(t), \quad s \in [t, T]. \tag{3.26}$$

Substituting (3.26) back into the state equation (3.23), we obtain the following *optimality system*, which is a **coupled forward-backward SDE system**:

$$\begin{cases} dX(s) = [aX(s) - b^2Y(s)]\, ds + dB(s), \\ dY(s) = -[aY(s) + X(s)]\, ds + Z(s)dB(s), \qquad s \in [t, T], \\ X(t) = x, \quad Y(T) = X(T). \end{cases} \tag{3.27}$$

It is clear that if we can prove that (3.27) admits an adapted solution $(X(\cdot), Y(\cdot), Z(\cdot))$, then (3.26) gives an optimal control, solving the original stochastic optimal control problem.

2. *Solving the problem via HJB:* Consider the state equation (3.23) with the cost functional (3.24). Define the value function:

$$V(t, x) = \inf_{u \in \mathcal{U}} \mathbb{E}\left[\frac{1}{2}\int_t^T (X(s)^2 + u(s)^2)\, ds + \frac{1}{2}X(T)^2 \mid X(t) = x\right], \tag{3.28}$$

with terminal condition:

$$V(T, x) = \frac{1}{2}x^2.$$

$\mathcal{U}$ is the set of admissible controls;

By the Dynamic Programming Principle 2.3.1 and Itô formula, the value function satisfies the following HJB equation:

$$V_t + \inf_u [H(t, x, u, V_x, V_{xx})] = 0,$$

since the drift $b(X(t), u(t)) = aX(t) + bu(t)$ and the diffusion $\sigma(X(t), u(t)) = 1$, and the running cost $l(X(t), u(t)) = \frac{1}{2}(X(t)^2 + u(t)^2)$ then the Hamiltonian is defined as:

$$H(t, x, u, V_x, V_{xx}) = \frac{1}{2}V_{xx} + (ax + bu)V_x + \frac{1}{2}(x^2 + u^2).$$

The HJB equation is:

$$V_t + \inf_u \left[(ax + bu)V_x + \frac{1}{2}V_{xx} + \frac{1}{2}(x^2 + u^2)\right] = 0.$$

Minimize with respect to the control:

$$\frac{\partial}{\partial u}(buV_x + \frac{1}{2}u^2) = 0.$$

This gives:

$$\bar{u} = -bV_x.$$

Substituting the minimizer back:

$$b\bar{u}V_x + \frac{1}{2}\bar{u}^2 = -b^2V_x^2 + \frac{1}{2}b^2V_x^2 = -\frac{1}{2}b^2V_x^2.$$

The HJB becomes:

$$V_t + axV_x + \frac{1}{2}V_{xx} + \frac{1}{2}x^2 - \frac{1}{2}b^2V_x^2 = 0. \tag{3.29}$$

3. *Connection with FBSDE:* For moving from the PDE to the FBSDE, we use the nonlinear Feynman-Kac formula 1.35, which is:

$$Y(s) = V(s, X(s)), \qquad Z(s) = \sigma^\top(s, X(s))V_x(s, X(s)).,$$

First, substituting the optimal control into the state, we get the FSDE:

$$dX(s) = [aX(s) - b^2V_x]\ ds + dB(s),$$

In our state, diffusion $\sigma = 1$, then $Z(t) = V_x(t, X(t))$. By applying Itô's formula to $Y(t)$, we get:

$$dY(s) = V_t ds + V_x dX(s) + \frac{1}{2}V_{xx}ds.$$

Substituting $dX(s)$, we get:

$$dY(s) = [V_t + aX(s)V_x + \frac{1}{2}V_{xx} - b^2V_x^2]ds + V_x dB(s). \tag{3.30}$$

From the HJB equation (3.29):

$$V_t + aX(t)V_x + \frac{1}{2}V_{xx} = -\frac{1}{2}X(t)^2 + \frac{1}{2}b^2V_x^2,$$

then the equation (3.30) becomes:

$$dY(t) = [-\frac{1}{2}X(t)^2 + \frac{1}{2}b^2V_x^2 - b^2V_x^2]\ dt + Z(t)dB(t).$$

Simplify and use the fact that $V_x = Z(t)$:

$$dY(t) = -\frac{1}{2}(X(t)^2 + b^2Z(t)^2)\ dt + Z(t)dB(t).$$

The terminal condition is obtained by:

$$Y(T) = V(T, X(T)) = \frac{1}{2}X(T)^2.$$

The optimal control problem is equivalent to the coupled system:

$$\begin{cases} dX(s) = [aX(s) - b^2 Z(s)]dt + dB(s), \\ dY(s) = -\frac{1}{2}[X(s)^2 + b^2 Z(s)^2]\ dt + Z(t)dB(s), \qquad s \in [t, T], \\ X(t) = x, \qquad Y(T) = \frac{1}{2}X(T)^2. \end{cases} \tag{3.31}$$

with optimal control:

$$\bar{u}(t) = -bZ(t). \tag{3.32}$$

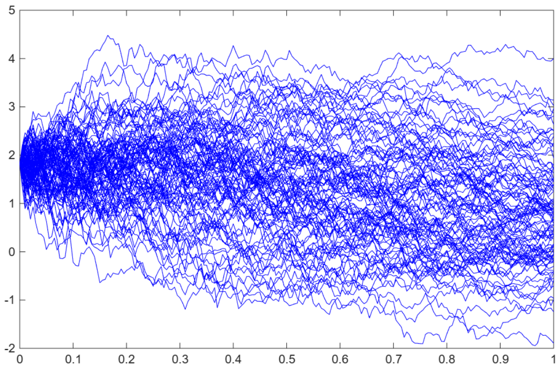


Figure 3.11: Simulations of 100 trajectories of the process $Y(t)$

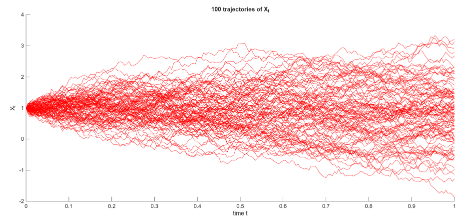


Figure 3.12: Simulations of 100 trajectories of the processes $X(t)$

Interpretation: Figure 3.11 represents the trajectory of Y, where they are arises at $t = 0$ to one point which represents the optimal cost. Figure 3.12, on the other hand, illustrates the trajectories of the optimal solutions.
After obtained the adapted solution $(X(\cdot), Y(\cdot), Z(\cdot))$ of the coupled FBSDE system (3.27), we now consider the corresponding optimal control (3.26), which is computed directly from the solution of the backward equation. The following figure displays the trajectory of $\bar{u}(\cdot)$.

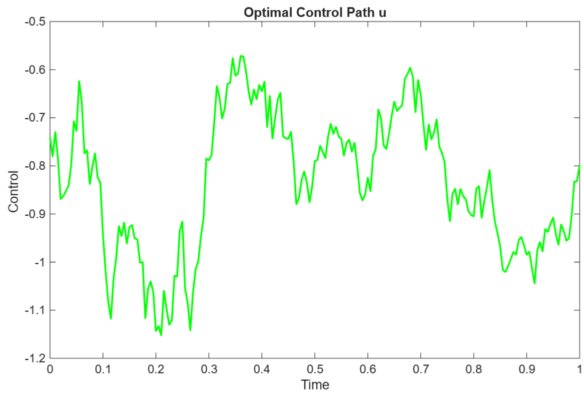


Figure 3.13: Trajectory of the optimal control $\bar{u}(t) = -bY(t)$ obtained from the solution of the coupled FBSDE system.

Notations

Appendix A

$(\Omega, \mathcal{F}, \mathbb{P})$ Probability space.

$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ Filtered probability space.

$(B_t)_{t\geq 0}$ Brownian motion.

$\mathbb{E}[X]$ Expectation of a random variable X.

$\mathbb{E}[X \mid \mathcal{F}_t]$ Conditional expectation given $\mathcal{F}_t$.

$(\mathcal{F}_t)_{t\geq 0}$ Filtration.

$\mathbb{F}$ Natural filtration generated by Brownian motion.

H Hamiltonian function associated with the Stochastic Maximum Principle.

G Generalized Hamiltonian function associated with HJB.

$J(u.)$ Cost functional corresponding to the control $u.$.

$\mathcal{N}$ Collection of all $\mathbb{P}$-null sets of $\mathcal{F}$.

$\mathbb{R}$ The set of real numbers.

$\mathbb{R}^n$ n-dimensional Euclidean space.

$\mathbb{R}^{n\times d}$ Set of all real $n \times d$ matrices.

$\mathcal{S}^n$ Set of all real $n \times n$ symmetric matrices.

$\mathcal{S}^2(t, T; \mathbb{R}^m)$ denote the set of $\mathbb{R}^m$-valeud, $\mathbb{F}$-adapted, continuous processes $(X_s, s \in [t, T])$ which satisfy $\mathbb{E}[\sup_{t\leq s\leq T} |X_s|^2] < \infty$.

U Non-empty subset of $\mathbb{R}^m$ representing the control set.

$\mathcal{U}$ Set of admissible controls.

$L^1(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ Set of random variables X such that

$$\mathbb{E}[|X|] = \int_\Omega |X(\omega)|\, dP(\omega) < \infty.$$

$L^2(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ Set of $\mathbb{R}^n$-valued random variables X such that

$$\mathbb{E}[|X|^2] = \int_\Omega |X(\omega)|^2\, dP(\omega) < \infty.$$

$\mathcal{C}([0,T] \times \mathbb{R}^n)$ Space of continuous functions from $[0,T] \times \mathbb{R}^n$ to $\mathbb{R}$.

$\mathcal{C}^{1,2}([0,T] \times \mathbb{R}^n)$ Functions continuously differentiable in the first variable and twice continuously differentiable in the second variable.

$F : \mathbb{R}^p \to \mathbb{R}$ is of $\mathcal{C}^2$ if all its first and second-order partial derivatives exist and are continuous on $\mathbb{R}^p$.

$\varphi_x,\ \varphi_{xx}$ **or** $\nabla_x \varphi,\ \nabla_x^2 \varphi$ Gradient and Hessian of $\varphi(t,x)$, respectively.

Appendix B

Abreviations:

SDEs Stochastic Differential Equations.

BSDEs Backward Stochastic Differential Equations.

FBSDEs Forward–Backward Stochastic Differential Equations.

SMP Stochastic Maximum Principle.

DPP Dynamic Programming Principle.

HJB Hamilton–Jacobi–Bellman equation.

LSMC Least Square Monte Carlo.

a.e. Almost everywhere.

a.s. Almost surely.

r.v. Random variable.

$\mathrm{Tr}(A)$ Trace of a matrix A.

$x^\top$ Transpose of the vector (or matrix) x.

$\langle \cdot, \cdot \rangle$ Inner product in Euclidean space.

$\|\varphi\|^2 = Tr(\varphi\varphi^\top) = \sum_{i,j} \varphi_{i,j}^2.$

Conclusion

This thesis has presented a comprehensive study of Forward-Backward Stochastic Differential Equations (FBSDEs) and their pivotal role within the framework of optimal control theory. By systematically bridging the gap between foundational stochastic analysis, deterministic partial differential equations, and numerical simulations, this work underscores the profound theoretical elegance and practical utility of FBSDE systems. The structural progression of this research reflects the inherent lifecycle of an applied mathematics problem, moving methodically from rigorous foundational proofs and structural optimization theory to targeted algorithmic design and practical validation.

The mathematical foundation established in the initial phase of this research provided the necessary analytical machinery to explore these complex systems. Through a deep dive into stochastic calculus, we consolidated the essential tools of Brownian motion, stochastic integration, and Itô's formula, which allowed for a natural progression into standard forward stochastic differential equations. Recognizing that terminal-value problems require an inverted information flow, the inquiry was extended to backward stochastic differential equations, focusing specifically on the adapted nature of their solution pairs. By integrating these forward and backward trajectories, we thoroughly addressed the existence and uniqueness of solutions for both decoupled and coupled FBSDE structures. This theoretical journey culminated in an examination of the deep interplay between probabilistic methods and deterministic analysis, demonstrating how these joint stochastic processes map directly to linear and non-linear partial differential equations through the classical and non-linear Feynman-Kac formulas.

With these analytical foundations secure, this work introduced the core objective of the thesis: optimizing stochastic dynamical systems under specified performance criteria. We approached this optimal control problem from two historically distinct yet deeply connected paradigms, beginning with the Stochastic Maximum Principle to derive necessary optimality conditions where the resulting adjoint equations naturally take the form of BSDEs. Parallel to this local approach, we utilized the Dynamic Programming Principle to derive the governing Hamilton-Jacobi-Bellman equation, which characterizes optimal behavior globally via a non-linear partial differential equation. A central achievement of this work was establishing the explicit mathematical bridge between these two frameworks. We successfully demonstrated that the adjoint process from the Stochastic Maximum Principle is intrinsically linked to the spatial gradient of the value function in the Hamilton-Jacobi-Bellman framework, thereby proving that pathwise optimization duals are identical to value-function sensitivities.

Because analytical solutions to fully coupled FBSDEs are exceptionally rare, the final phase of this thesis translated abstract mathematical theory into robust computational algorithms. We evaluated time-discretization schemes to approximate the forward and backward paths over discrete horizons, and explored the algorithmic implementation of conditional expectations required to resolve the terminal-value constraints. These techniques culminated in comprehensive numerical methods capable of solving both coupled and decoupled systems. To validate the practical utility of our numerical solvers, we implemented two distinct case studies. First, a controlled SIR model applied FBSDE dynamics to an epidemiological framework, demonstrating how non-linear, coupled systems can be actively controlled to optimize societal outcomes under uncertainty. Second, a classic Linear-Quadratic example allowed us to rigorously verify the accuracy, convergence rates, and stability of our numerical algorithms against known analytical baselines, with the simulations closely mirroring our theoretical predictions and reaffirming that FBSDEs remain an indispensable frontier in the study of stochastic systems and optimization.

Bibliography

[1] F. Antonelli. *Backward forward stochastic differential equations.* Purdue University, 1993.

[2] R. Bellman. Dynamic programming, princeton univ. *Press Princeton, New Jersey*, 39, 1957.

[3] C. Bender and J. Zhang. Time discretization and markovian iteration for coupled fbsdes. *The Annals of Applied Probability*, 18(1), Feb. 2008.

[4] A. Bensoussan. Lectures on stochastic control in nonlinear filtering and stochastic control. lectures notes in mathematics, 1982.

[5] A. Bensoussan. Stochastic maximum principle for distributed parameter systems. *Journal of the Franklin Institute*, 315(5-6):387–406, 1983.

[6] N. Berglund. Martingales et calcul stochastique. *arXiv preprint arXiv:1312.7799*, 2013.

[7] N. Berrouis. *The study of optimal controls for forward backward doubly stochastic differential equations.* Doctorate of science thesis, Department of Mathematics, Mohamed Khider University, Biskra, Algeria, 2022.

[8] P. Billingsley. *Probability and measure.* John Wiley & Sons, 2017.

[9] J.-M. Bismut. Conjugate convex functions in optimal stochastic control. *Journal of mathematical analysis and applications*, 44(2):384–404, 1973.

[10] J.-M. Bismut. *Théorie probabiliste du contrôle des diffusions*, volume 167. American Mathematical Soc., 1976.

[11] J.-M. Bismut. An introductory approach to duality in optimal stochastic control. *SIAM review*, 20(1):62–78, 1978.

[12] H. Bouanani. *Optimal control for forward backward stochastic differential equation driven by G-Brownian motion.* PhD thesis, University of Saida – Dr. Moulay Tahar, Department of Mathematics, 2021.

[13] P. Briand. Equations différentielles stochastiques rétrogrades. *Mars*, 2001.

[14] J. Chessari, R. Kawai, Y. Shinozaki, and T. Yamada. Numerical methods for backward stochastic differential equations: A survey. *Probability Surveys*, 20:486–567, 2023.

[15] M. G. Crandall and P.-L. Lions. Viscosity solutions of hamilton-jacobi equations. *Transactions of the American mathematical society*, 277(1):1–42, 1983.

[16] D. Crisan, K. Manolarakis, and N. Touzi. On the monte carlo simulation of bsdes: An improvement on the malliavin weights. *Stochastic Processes and their Applications*, 120(7):1133–1158, 2010.

[17] F. Delarue. On the existence and uniqueness of solutions to fbsdes in a non-degenerate case. *Stochastic processes and their applications*, 99(2):209–286, 2002.

[18] R. Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[19] I. Ekeland. Nonconvex minimization problems. 1979.

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

[21] N. El Karoui and M.-C. Quenez. Non-linear pricing theory and backward stochastic differential equations. In *Financial Mathematics: Lectures given at the 3rd Session of the Centro Internazionale Matematico Estivo (CIME) held in Bressanone, Italy, July 8–13, 1996*, pages 191–246. Springer, 2006.

[22] I. Exarchos and E. A. Theodorou. Stochastic optimal control via forward and backward stochastic differential equations and importance sampling. *Automatica*, 87:159–165, 2018.

[23] W. Feller. Zur theorie der stochastischen prozesse: Existenz-und eindeutigkeitssätze. *Mathematische Annalen*, 113(1):113–160, 1937.

[24] W. H. Fleming and H. M. Soner. *Controlled Markov processes and viscosity solutions*. Springer, 2006.

[25] E. Gobet, J.-P. Lemor, and X. Warin. A regression-based monte carlo method to solve backward stochastic differential equations. *The Annals of Applied Probability*, 15(3), Aug. 2005.

[26] J. Han, A. Jentzen, and W. E. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34):8505–8510, 2018.

[27] J. Han, A. Jentzen, et al. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Communications in mathematics and statistics*, 5(4):349–380, 2017.

[28] J. Han and J. Long. Convergence of the deep bsde method for coupled fbsdes. *Probability, Uncertainty and Quantitative Risk*, 5(1):5, 2020.

[29] U. G. Haussmann. General necessary conditions for optimal control of stochastic systems. In *Stochastic systems: modeling, identification and optimization, II*, pages 30–48. Springer, 2009.

[30] D. J. Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM review*, 43(3):525–546, 2001.

[31] Y. Hu and S. Peng. Solution of forward-backward stochastic differential equations. *Probability Theory and Related Fields*, 103(2):273–283, 1995.

[32] Z. Huang, B. Negyesi, and C. W. Oosterlee. Convergence of the deep bsde method for stochastic control problems formulated through the stochastic maximum principle. *Mathematics and Computers in Simulation*, 227:553–568, 2025.

[33] Z. Huang and C. W. Oosterlee. Convergence of the markovian iteration for coupled fbsdes via a differentiation approach, 2025.

[34] K. Itô. Differential equations determining markov processes. 244(1077):1352–1400, 1942.

[35] K. Itô. 109. stochastic integral. *Proceedings of the Imperial Academy*, 20(8):519–524, 1944.

[36] K. Itô. *On stochastic differential equations*, volume 4. American Mathematical Society New York, 1951.

[37] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, New York, 2 edition, 1991.

[38] O. Kebiri, L. Neureither, and C. Hartmann. Adaptive importance sampling with forward-backward stochastic differential equations, 2019.

[39] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 1992.

[40] A. Kolmogoroff. Über die analytischen methoden in der wahrscheinlichkeitsrechnung. *Mathematische Annalen*, 104(1):415–458, 1931.

[41] H. Kushner. Necessary conditions for continuous parameter stochastic optimization problems. *SIAM Journal on Control*, 10(3):550–565, 1972.

[42] J.-F. Le Gall. *Mouvement brownien, martingales et calcul stochastique*, volume 71 of *Mathématiques et Applications*. Springer, 2013.

[43] O. Lévêque. Cours de probabilités et calcul stochastique. 2005.

[44] P. Lévy. *Processus Stochastiques et Mouvement Brownien*. Paris, Gauthier-Villars, 1948.

[45] J. Ma, P. Protter, and J. Yong. Solving forward backward stochastic differential equations explicitly a four step scheme. *Probability theory and related fields*, 98(3):339–359, 1994.

[46] J. Ma and J. Yong. *Forward–Backward Stochastic Differential Equations and Their Applications*. Springer, Berlin, 1999.

[47] G. Milstein and M. V. Tretyakov. Numerical algorithms for forward-backward stochastic differential equations. *SIAM Journal on Scientific Computing*, 28(2):561–582, 2006.

[48] B. Negyesi and C. W. Oosterlee. A numerical fourier cosine expansion method with higher order taylor schemes for fully coupled fbsdes. *arXiv preprint arXiv:2501.10988*, 2025.

[49] B. Øksendal. Stochastic differential equations. In *Stochastic differential equations: an introduction with applications*, pages 38–50. Springer, 2003.

[50] G. Pagès and A. Sagna. Improved error bounds for quantization based numerical schemes for bsde and nonlinear filtering. *Stochastic Processes and their Applications*, 128(3):847–883, 2018.

[51] É. Pardoux. Backward stochastic differential equations and viscosity solutions of systems of semilinear parabolic and elliptic pdes of second order. In *Stochastic Analysis and Related Topics VI: Proceedings of the Sixth Oslo Silivri Workshop Geilo 1996*, pages 79–127. Springer, 1998.

[52] E. Pardoux and S. Peng. Adapted solution of a backward stochastic differential equation. *Systems & control letters*, 14(1):55–61, 1990.

[53] S. Peng. A general stochastic maximum principle for optimal control problems. *SIAM Journal on control and optimization*, 28(4):966–979, 1990.

[54] S. Peng. Probabilistic interpretation for systems of quasilinear parabolic partial differential equations. *Stochastics and stochastics reports (Print)*, 37(1-2):61–74, 1991.

[55] S. Peng. A generalized dynamic programming principle and hamilton-jacobi-bellman equation. *Stochastics: An International Journal of Probability and Stochastic Processes*, 38(2):119–134, 1992.

[56] S. Peng et al. A nonlinear feynman-kac formula and applications. In *Proceedings of Symposium of System Sciences and Control Theory*, pages 173–184. World Scientific, 1992.

[57] S. Peng and Z. Wu. Fully coupled forward-backward stochastic differential equations and applications to optimal control. *SIAM Journal on Control and Optimization*, 37(3):825–843, 1999.

[58] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.

[59] H. S. Rodrigues. Optimal control and numerical optimization applied to epidemiological models, 2014.

[60] T. Sauer. *Numerical Solution of Stochastic Differential Equations in Finance*, pages 529–550. 07 2012.

[61] H. J. Sussmann and J. C. Willems. 300 years of optimal control: from the brachystochrone to the maximum principle. *IEEE Control Systems Magazine*, 17(3):32–44, 2002.

[62] N. Touzi. *Optimal stochastic control, stochastic target problems, and backward SDE*, volume 29. Springer Science & Business Media, 2012.

[63] N. Wiener. Differential-space. *Journal of Mathematics and Physics*, 2(1-4):131–174, 1923.

[64] J. T. Wu, K. Leung, and G. M. Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *The lancet*, 395(10225):689–697, 2020.

[65] J. Yong and X. Y. Zhou. *Stochastic controls: Hamiltonian systems and HJB equations*, volume 43. Springer Science & Business Media, 1999.