

People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

University of Saida - Dr. Moulay Tahar.
Faculty of MIT.
Department of Mathematics.



No. Assigned by the library.

--	--	--	--	--	--	--	--	--	--



Master's dissertation Submitted for the Degree of
Academic Master

Discipline: MATHEMATICS
Specialty: Stochastic Analysis, Process Statistics, and Applications (SASPA)

by

Mohamed Amine MOKHTARI ¹

Under the supervision of

Dr. Khalida BACHIR CHERIF

Theme:

**Stochastic Optimal Control and
Machine Learning**

Defended on September 28th, 2026 before the jury composed of

Dr. Kouider DJERFI	University of Saida Dr. Moulay Tahar	President
Dr. Khalida BACHIR CHERIF	University of Saida Dr. Moulay Tahar	Supervisor
Dr. Imen MEKKAOU	University of Saida Dr. Moulay Tahar	Examiner

Year, 2025/2026.

Abstract

This thesis presents a modern machine learning framework for solving stochastic optimal control problems. The system dynamics follow stochastic differential equations (SDEs), and the objective is to determine control strategies that optimize a specified performance measure. We focus specifically on high-dimensional problems, approaching them through the stochastic maximum principle while leveraging deep learning to enable efficient computation.

Keywords: Stochastic differential equations, Stochastic maximum principle, Stochastic control, deep neural networks

Acknowledgements

The last duty of a person who has received kindness is to express gratitude.

First and foremost, I would like to express my deepest gratitude and infinite respect to Almighty God, who has granted me the strength, courage, and patience to complete this work successfully.

I would like to express my sincere gratitude to my supervisor, **Dr. Khalida BACHIR CHERIF**, for her availability, valuable guidance, patience, and scientific rigor. She has kindly guided me throughout this project. I sincerely appreciate her support and supervision.

I would also like to thank all the teachers who have contributed to my education, from my earliest years of study until now. Their courses, passion, and dedication have inspired me and prepared me for the challenges of research.

My sincere thanks also go to the members of the examination committee **Pr. Kouider DJERFI, Dr. Imen MEKKAOU** who agreed to read and evaluate this work. Their comments and suggestions are greatly appreciated.

A heartfelt thank you to my family, especially my dear father, my beloved mother, and my brothers and sisters. Their moral support, constant encouragement, and unconditional love have been my driving force throughout this journey. No words can fully express my gratitude.

Dedication

I dedicate this work to:

- † To my dear parents, a source of strength, tenderness, and dignity.
- † To my family, for their unconditional love, constant support, and for always being my source of strength and encouragement throughout my journey.
- † To all my teachers, especially those who have guided, trained, and mentored me throughout my academic journey..
- † To my friends and classmates, for their sincere friendship, support, and the unforgettable moments we shared together.
- † To all those who have contributed, directly or indirectly, to my success and achievement.

Thank you from the bottom of my heart.

List of Figures

3.1	Machine Learning and Deep Learning	35
3.2	A simple neural network with two hidden layers (in red), and a one-dimensional output layer	39
4.1	The whole network architecture for Algorithm 1	47
4.2	Relative error for $n = 100$	57
4.3	Mean and scope of the cost functional ($n = 100$)	57
4.4	Mean and scope of the cost functional for 10 independent runs	59
4.5	Mean and scope of relative errors of the cost functional	60
4.6	mean and scope of the cost functional among 10 independent runs (variances comparison)	62

List of Tables

4.1	Comparison of our different algorithms for $n = 5$	55
4.2	Comparison between ODE45 and the algorithms for different dimensions	58

Contents

List of Figures	4
List of Tables	5
Introduction	8
1 Stochastic Control Problem	11
1.1 Mathematical formulation of SOCP	11
1.1.1 Strong formulation	11
1.1.2 Weak Formulation	13
1.2 Existence of Optimal Controls	15
1.2.1 Existence under Strong Formulation	15
1.2.2 Existence under Weak Formulation	16
1.3 Stochastic Control Models	17
1.3.1 Random duration	17
1.3.2 Risk-Sensitive Controls	18
2 Maximum Principle and Stochastic Hamiltonian Systems	21
2.1 Statement of the Stochastic Maximum Principle	21
2.2 Adjoint equations	23
2.3 The maximum principle and stochastic Hamiltonian systems .	25
2.3.1 A worked-out example	30
3 Machine Learning and Deep Learning	34
3.1 What Are Machine Learning and Deep Learning ?	34
3.2 Foundations of Machine Learning	35
3.2.1 Types of Machine Learning	36

3.2.2	Generalization	37
3.2.3	Choosing the Right Model	37
3.3	Deep Learning	38
3.3.1	Activation Functions	39
3.3.2	Loss	40
3.3.3	Optimization	40
4	Deep Learning Methods for Stochastic Optimal Control	42
4.1	Motivation and Challenges	42
4.2	Background and problem formulation	43
4.3	Numerical algorithms	45
4.3.1	Algorithm 1: Numerical algorithm with 1-NNet	45
4.3.2	Algorithm 2: Numerical algorithm with 2-NNets	48
4.3.3	Algorithm 3: Numerical Algorithm with explicit ex- pression of $\tilde{H}$	51
4.4	Numerical results	53
4.4.1	A low-dimensional example	54
4.4.2	Some high-dimensional stochastic optimal control prob- lems	56
4.5	Conclusion	63

Introduction

Stochastic optimal control is a fundamental area of applied mathematics that deals with decision-making problems under uncertainty. In many real-world systems, randomness arises naturally due to noise, incomplete information, or external disturbances. As a result, the evolution of such systems cannot be accurately described by deterministic models alone, which motivates the use of stochastic differential equations as a natural modeling framework [24, 16].

Stochastic optimal control theory has been mainly developed through two fundamental approaches: Pontryagin’s Maximum Principle (PMP) and Bellman’s Dynamic Programming Principle (DPP). These two frameworks constitute the theoretical foundation of modern stochastic control and have been successfully applied in various fields such as finance, engineering, robotics, and economics.

Several numerical methods have been proposed for solving stochastic optimal control problems, including Markov chain approximation methods, finite-difference schemes, and probabilistic numerical methods based on dynamic programming. Although these approaches provide accurate approximations in low-dimensional settings, they become computationally inefficient when the dimension of the state space increases.

In particular, classical dynamic programming methods require the numerical solution of Hamilton–Jacobi–Bellman (HJB) equations on high-dimensional grids. Consequently, the computational complexity grows exponentially with respect to the dimension, leading to the well-known curse of dimensionality.

The main objective of stochastic optimal control theory is to determine a control strategy that optimizes a given performance criterion, typically expressed as the expectation of a cost functional. The controlled dynamics are described by stochastic differential equations, and the optimization is

performed over a set of admissible controls. This leads to highly non-trivial mathematical problems that require tools from probability theory, stochastic calculus, and functional analysis [32].

This work is structured into four main chapters. The first chapter introduces the stochastic control problem, including the formulation of controlled stochastic differential equations, the definition of admissible control processes, and the associated cost functional. This provides the basic mathematical framework for the subsequent analysis.

The second chapter focuses on the essential tool in optimal control which is Stochastic Maximum Principle (SMP) and the associated stochastic Hamiltonian system. The SMP provides necessary conditions for optimality and plays a central role in characterizing optimal controls. It transforms the original control problem into a system of forward-backward stochastic differential equations (FBSDEs), as developed in the works of Bismut and Peng, and further developed in [32].

The third chapter examines the fundamental concepts of machine learning (ML) and deep learning (DL), providing a comprehensive understanding for both researchers and beginners aiming to build a strong foundation in these fields. In particular, it aims to help readers understand when and how to effectively apply different models.

The chapter also explores key techniques for improving model performance, including regularization and hyperparameter optimization. Furthermore, it highlights the essential components of deep learning, such as activation functions, loss functions, and optimization algorithms, while emphasizing foundational methods that continue to be relevant and effective today.

The final chapter is devoted to modern computational approaches, particularly deep learning methods for stochastic optimal control. Classical analytical and numerical techniques often suffer from the curse of dimensionality, making them inefficient in high-dimensional settings. Recent advances in machine learning, especially deep neural networks, have shown great potential in overcoming these limitations. In particular, methods developed by Han, Jentzen, and E demonstrate how deep learning can be used to approximate solutions of high-dimensional stochastic control problems through their connection with backward stochastic differential equations (BSDEs) [18].

In conclusion, this study bridges classical stochastic control theory with modern deep learning techniques, providing both a theoretical foundation and a computational perspective for solving complex stochastic optimization problems.

Chapter 1

Stochastic Control Problem

1.1 Mathematical formulation of SOCP

1.1.1 Strong formulation

Given a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ satisfying the usual conditions, on which an m -dimensional standard Brownian motion $W(\cdot)$ is defined, consider the following controlled stochastic differential equation:

$$\begin{cases} dx(t) = b(t, x(t), u(t)) dt + \sigma(t, x(t), u(t)) dW(t), & t \in [0, T], \\ x(0) = x_0 \in \mathbb{R}^n. \end{cases} \quad (1.1)$$

where,

$$b : [0, T] \times \mathbb{R}^n \times U \rightarrow \mathbb{R}^n, \quad \sigma : [0, T] \times \mathbb{R}^n \times U \rightarrow \mathbb{R}^{n \times m},$$

with U is a given separable metric space and $T \in (0, \infty)$ being fixed. The function $u(\cdot)$ is called the control representing the action, decision, or policy of the decision-makers (controllers). At any time instant the controller is knowledgeable about some information (as specified by the information field $\{\mathcal{F}_t\}_{t \geq 0}$) of what has happened up to that moment, but not able to foretell what is going to happen afterwards due to the uncertainty of the system.

This nonanticipative restriction in mathematical terms can be represented as “ $u(\cdot)$ is $\{\mathcal{F}_t\}_{t \geq 0}$ -adapted” . Namely, the control $u(\cdot)$ is taken from the set

$$\mathcal{U}[0, T] \triangleq \{u : [0, T] \times \Omega \rightarrow U \mid u(\cdot) \text{ is } \{\mathcal{F}_t\}_{t \geq 0}\text{-adapted}\}.$$

Any $u(\cdot) \in \mathcal{U}[0, T]$ is called a feasible control.

In addition, we may have some state constraints. For any $u(\cdot) \in \mathcal{U}[0, T]$, equation (1.1) is one with random coefficients. Let $S(t) : [0, T] \rightarrow 2^{\mathbb{R}^n}$ be a given multifunction. The state constraint may be given by

$$x(t) \in S(t), \quad \forall t \in [0, T], \quad \text{P-a.s.} \quad (1.2)$$

Note that some other types of state constraints are also possible. Next, we introduce the cost functional as follows:

$$J(u(\cdot)) = \mathbb{E} \left\{ \int_0^T f(t, x(t), u(t)) dt + h(x(T)) \right\}. \quad (1.3)$$

Definition 1.1.1. [31] Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ be given satisfying the usual conditions and let $W(t)$ be a given m -dimensional standard $\{\mathcal{F}_t\}_{t \geq 0}$ -Brownian motion. A control $u(\cdot)$ is called an s -admissible control, and $(x(\cdot), u(\cdot))$ an s -admissible pair, if

- (i) $u(\cdot) \in \mathcal{U}[0, T]$;
- (ii) $x(\cdot)$ is the unique solution of equation (1.1)
- (iii) some prescribed state constraints (for example, 1.2) are satisfied;
- (iv) $f(\cdot, x(\cdot), u(\cdot)) \in L^1_{\mathcal{F}}(0, T; \mathbb{R})$ and $h(x(T)) \in L^1_{\mathcal{F}_T}(\Omega; \mathbb{R})$.

The set of all s -admissible controls is denoted by $\mathcal{U}_{\text{ad}}^s[0, T]$.

Our stochastic optimal control problem under strong formulation can be stated as follows:

Problem (SS). Minimize (1.3) over $\mathcal{U}_{\text{ad}}^s[0, T]$.

The goal is to find $\bar{u}(\cdot) \in \mathcal{U}_{\text{ad}}^s[0, T]$ (if it ever exists), such that

$$J(\bar{u}(\cdot)) = \inf_{u(\cdot) \in \mathcal{U}_{\text{ad}}^s[0, T]} J(u(\cdot)). \quad (1.4)$$

Problem (SS) is said to be s -finite if the right-hand side of (1.4) is finite, and it is said to be (uniquely) s -solvable if there exists a (unique) $\bar{u}(\cdot) \in \mathcal{U}_{\text{ad}}^s[0, T]$ such that (1.4) holds.

Any $\bar{u}(\cdot) \in \mathcal{U}_{\text{ad}}^s[0, T]$ satisfying (1.4) is called an s -optimal control. The corresponding state process $\bar{x}(\cdot)$ and the state-control pair $(\bar{x}(\cdot), \bar{u}(\cdot))$ are called an s -optimal state process and an s -optimal pair, respectively.

1.1.2 Weak Formulation

We note that in the strong formulation, the underlying filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ along with the Brownian motion $W(\cdot)$ are all fixed. However, in certain situations it will be convenient or necessary to vary $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ as well as $W(\cdot)$ and to consider them as parts of the control. This is the case.

Definition 1.1.2. [31] *A 6-tuple*

$$\pi = (\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P}, W(\cdot), u(\cdot))$$

is called a w -admissible control, and $(x(\cdot), u(\cdot))$ a w -admissible pair, if

(i) $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ *is a filtered probability space satisfying the usual conditions;*

(ii) $W(\cdot)$ *is an m -dimensional standard Brownian motion defined on*

$$(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P});$$

(iii) $u(\cdot)$ *is an $\{\mathcal{F}_t\}_{t \geq 0}$ -adapted process on $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in U ;*

(iv) $x(\cdot)$ *is the unique solution of equation (1.1) on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ under $u(\cdot)$;*

(v) *some prescribed state constraints (for example, (1.2)) are satisfied;*

(vi) $f(\cdot, x(\cdot), u(\cdot)) \in L^1_{\mathcal{F}}(0, T; \mathbb{R})$ *and* $h(x(T)) \in L^1(\Omega; \mathbb{R})$.

Here, the spaces $L^1_{\mathcal{F}}(0, T; \mathbb{R})$ and $L^1(\Omega; \mathbb{R})$ are defined on the given filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ associated with the 6-tuple π .

The set of all w -admissible controls is denoted by $\mathcal{U}_{\text{ad}}^w[0, T]$.

Sometimes, we might write $u(\cdot) \in \mathcal{U}_{\text{ad}}^w[0, T]$ instead of $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P}, W(\cdot), u(\cdot)) \in \mathcal{U}_{\text{ad}}^w[0, T]$, if it is clear from the context that weak formulation is under consideration.

Our stochastic optimal control problem under the weak formulation can be stated as follows.

Problem (WS). Minimize (1.3) over $\mathcal{U}_{\text{ad}}^w[0, T]$.

Namely, one seeks $\bar{\pi} \in \mathcal{U}_{\text{ad}}^w[0, T]$ (if it exists) such that

$$J(\bar{\pi}) = \inf_{\pi \in \mathcal{U}_{\text{ad}}^w[0, T]} J(\pi). \quad (1.5)$$

As in the strong formulation, Problem (WS) is said to be w -finite if the right-hand side of (1.5) is finite. One may also define the w -solvability, the w -optimal control, the w -optimal state process, and the w -optimal pair accordingly.

We emphasize here that the strong formulation is the one that stems from the practical world, whereas the weak formulation sometimes serves as an auxiliary but effective mathematical model aiming at ultimately solving problems with the strong formulation. A main reason why this might work is that the objective of a stochastic control problem is to minimize the mathematical expectation of a certain random variable that depends only on the distribution of the processes involved. Therefore, if the solutions of (1.1) in different probability spaces have the same probability distribution, then one has more freedom in choosing a convenient probability space to work with. A good example is when one tries to employ the dynamic programming principle to solve the problem. It should also be noted, however, that the weak formulation fails if any of the given coefficients b , σ , f , and h are also random (namely, they depend on ω explicitly), because in this case the probability space has to be specified and fixed a priori.

The following type of the state equation is of particular interest:

$$\begin{cases} dx(t) = [A(t)x(t) + B(t)u(t)]dt + \sum_{j=1}^m [C_j(t)x(t) + D_j(t)u(t)]dW^j(t), \\ x(0) = x_0, \end{cases} \quad (1.6)$$

where $A(\cdot)$, $B(\cdot)$, $C_j(\cdot)$, and $D_j(\cdot)$ are matrix-valued functions of suitable Sizes. The associated cost functional is of a quadratic form :

$$J(u(\cdot)) = \mathbb{E} \int_0^T \frac{1}{2} \left(\left[\langle Q(t)x(t), x(t) \rangle + \langle R(t)u(t), u(t) \rangle \right] \right) dt + \frac{1}{2} \langle Gx(T), x(T) \rangle. \quad (1.7)$$

Problem (SS) or (WS) with state equation (1.6) and cost functional (1.7) is called a stochastic linear quadratic optimal control problem (stochastic LQ

problem, for short). We note that in the above formulation, $T > 0$ is a fixed constant. Thus, Problem (SS) or (WS) is also referred to as a fixed duration problem. Sometimes, we may face problems with nonfixed durations.

1.2 Existence of Optimal Controls

In this section we are going to discuss the existence of optimal controls. The essence of the theory is the following: A lower semicontinuous function defined on some compact metric space attains its minimum.

1.2.1 Existence under Strong Formulation

We now consider the stochastic case. Let us first look at the existence under strong formulation.

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ be given as before and let $W(\cdot)$ be a given one-dimensional Brownian motion. Consider the following stochastic linear controlled system:

$$\begin{cases} dx(t) = [Ax(t) + Bu(t)] dt + [Cx(t) + Du(t)] dW(t), & t \in [0, T], \\ x(0) = x_0, \end{cases} \quad (1.8)$$

where A, B, C, D are matrices of suitable sizes. The state $x(\cdot)$ takes values in $\mathbb{R}^n$, and the control $u(\cdot)$ is in

$$\mathcal{U}^L[0, T] \triangleq \left\{ u(\cdot) \in L^2_{\mathcal{F}}(0, T; \mathbb{R}^k) \mid u(t) \in U, \text{ a.e. } t \in [0, T], \mathbb{P}\text{-a.s.} \right\}, \quad (1.9)$$

with $U \subset \mathbb{R}^k$. Note that we have an additional constraint that a control must be square-integrable, just to ensure the existence of solutions of (1.9) under $u(\cdot)$. If U is bounded, then this restriction is satisfied automatically.

The cost functional is

$$J(u(\cdot)) = \mathbb{E} \left\{ \int_0^T f(x(t), u(t)) dt + h(x(T)) \right\}. \quad (1.10)$$

with $f : \mathbb{R}^n \times U \rightarrow \mathbb{R}$ and $h : \mathbb{R}^n \rightarrow \mathbb{R}$.

The optimal control problem can be stated as follows.

Problem (SL). Minimize $J(\bar{u})$ subject to (1.9) over $\mathcal{U}^L[0, T]$.

Now we introduce the following assumptions:

(H1) The set $U \subseteq \mathbb{R}^k$ is convex and closed, and the functions f and h are convex. and for some $\delta, K > 0$,

$$f(x, u) \geq \delta |u|^2 - K, \quad h(x) \geq -K, \quad \forall (x, u) \in \mathbb{R}^n \times U. \quad (1.11)$$

(H2) The set $U \subseteq \mathbb{R}^k$ is convex and compact, and the functions f and h are convex.

When either (H1) or (H2) is assumed, Problem **(SL)** is called a *stochastic linear-convex optimal control problem*, since the controlled system is linear and the cost functional is convex. A special case is when both f and h are convex quadratic functions. Problem **(SL)** is then reduced to a *stochastic linear-quadratic problem*.

Theorem 1.1. [31] *Under either (H1) or (H2), if Problem **(SL)** is finite, then it admits an optimal control.*

1.2.2 Existence under Weak Formulation

In this subsection we examine the existence of optimal controls under the weak formulation. Let us make the following assumptions:

(SE1) (U, d) is a compact metric space and $T > 0$.

(SE2) The maps b, σ, f , and h are all continuous, and there exists a constant $L > 0$ such that, for $\varphi(t, x, u) = (b(t, x, u), \sigma(t, x, u), f(t, x, u), h(x))$

$$\left\{ \begin{array}{l} |\varphi(t, x, u) - \varphi(t, \hat{x}, \hat{u})| \leq L |x - \hat{x}|, \\ \forall t \in [0, T], \quad x, \hat{x} \in \mathbb{R}^n, \quad u, \hat{u} \in U, \\ |\varphi(t, 0, u)| \leq L, \quad \forall (t, u) \in [0, T] \times U. \end{array} \right. \quad (1.12)$$

For every $(t, x) \in [0, T] \times \mathbb{R}^n$, the set

$$(b, \sigma \sigma^T, f)(t, x, U) \triangleq \left\{ (b(t, x, u), (\sigma \sigma^T)^{ij}(t, x, u), f(t, x, u)) \mid \right.$$

$$\left. u \in U, \quad i = 1, \dots, n, \quad j = 1, \dots, m \right\}$$

is convex in $\mathbb{R}^{n+nm+1}$.

(SE4) $S(t) \equiv \mathbb{R}^n$.

Remark In proving the existence of optimal controls, one typically seeks a certain compactness structure. The weak formulation enables us to find the compactness of the image measure of some stochastic processes involved on a certain functional space. However, since the control $u(\cdot)$ is measurable only in t and there is no convenient compactness property on the space of merely measurable functions, we need to embed it in a larger space with proper compactness.

Theorem 1.2. [31] *Under assumptions (SE1)–(SE4), if Problem (WS) is finite, then it admits an optimal control.*

1.3 Stochastic Control Models

1.3.1 Random duration

In Problem (SS) formulated in Section 1, the time duration under consideration (called a control horizon) is a fixed, deterministic finite interval $[0, T]$. In many real applications the control horizon may be a random duration $[0, \tau]$ where the terminal time τ is a random variable at which the state of the controlled system changes critically and the control beyond τ may no longer be meaningful or necessary. For example, in the insurance company model (Section 3.3 [31]) the control after the time when the company is bankrupt becomes unnecessary. Therefore, it makes better sense to consider the problem in $[0, \tau]$ where τ is the bankruptcy time defined as

$$\tau = \inf\{t \geq 0 : x(t) = 0\}.$$

To formulate a general model that incorporates a random control horizon, consider the following stochastic differential equation:

$$\begin{cases} dx(t) = b(t, x(t), u(t)) dt + \sigma(t, x(t), u(t)) dW(t), \\ x(0) = x_0 \in \mathcal{O}. \end{cases} \quad (1.13)$$

on a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ on which an m -dimensional Brownian motion $W(\cdot)$ is defined. Here $\mathcal{O} \subseteq \mathbb{R}^n$ is a given open set. An admissible control $u(\cdot)$ is defined similarly as in Section 1.1.1.

Define the first exit time

$$\tau \triangleq \inf\{t > 0 : x(t) \notin \mathcal{O}\}.$$

Note that τ implicitly depends on the control taken.

Given the cost functional

$$J(u(\cdot)) = \mathbb{E} \left[\int_0^\tau f(t, x(t), u(t)) dt + h(x(\tau)) \right], \quad (1.14)$$

the problem is to minimize (1.14) over the set of admissible controls.

An interesting case is that in which $f \equiv 1$ and $h \equiv 0$, which corresponds to a stochastic time optimal control problem with the target set $M = \mathcal{O}^c$.

Another special model has $\mathcal{O} = \mathbb{R}^n$, in which case $\tau = \infty$. It is called an infinite horizon problem if the integral in (1.14) remains finite (as is the case when, for example, the function f is of the form $e^{-\gamma t}g(t, x, u)$ with g bounded and $\gamma > 0$).

1.3.2 Risk-Sensitive Controls

In the risk-sensitive control model, both the system dynamics and an admissible control are defined in the same way as those of Problem (SS) in Section 1.1.1. The only difference lies in the way of defining the performance criterion. Instead of minimizing the direct cost as in (1.3), the controller aims to minimize some disutility of the cost in a risk-sensitive control problem.

To be more precise, given a monotonically increasing function

$$\varphi : \mathbb{R} \rightarrow \mathbb{R},$$

the objective is to find an admissible control $u(\cdot)$ so as to minimize

$$J(u(\cdot)) = \mathbb{E} \left[\varphi \left(\int_0^T f(t, x(t), u(t)) dt + h(x(T)) \right) \right], \quad (1.15)$$

subject to the system (1.1).

In economic theory the disutility function φ is usually taken to be either convex or concave, corresponding respectively to the risk-averse or risk-seeking attitude of the decision-maker. That is why the model is called a risk-sensitive one.

Let us elaborate this point by an intuitive argument. Define

$$X \equiv \int_0^T f(t, x(t), u(t)) dt + h(x(T)),$$

and suppose φ is differentiable at $\mathbb{E}[X]$. Then we have the Taylor expansion

$$\varphi(X) \approx \varphi(\mathbb{E}X) + \varphi'(\mathbb{E}X)(X - \mathbb{E}X) + \frac{1}{2}\varphi''(\mathbb{E}X)(X - \mathbb{E}X)^2 + \dots$$

If φ is strictly convex near $\mathbb{E}X$, then $\varphi''(\mathbb{E}X) > 0$, which introduces a penalty to the variance term $(X - \mathbb{E}X)^2$ in the overall cost. This means that the controller tries to avoid large deviations of X from its mean $\mathbb{E}X$, that is, the controller is risk-averse. Conversely, if φ is strictly concave, the controller becomes risk-seeking. Near $\mathbb{E}X$, then $\varphi''(\mathbb{E}X) < 0$, which reduces the overall cost for large $|X - \mathbb{E}X|$. This is the so-called *risk-seeking* case.

Finally, if $\varphi''(\mathbb{E}X)$ is close to or equal to 0, then

$$\mathbb{E}[\varphi(X)] \approx \varphi(\mathbb{E}X),$$

in which case the risk-sensitive model reduces to the standard one in Section 1 (recall that φ is monotonically increasing). This corresponds to the *risk-neutral* situation.

A commonly used and extensively studied disutility function is the following exponential function:

$$\varphi^\theta(x) = \theta \exp(\theta x), \quad (1.16)$$

where $\theta \in \mathbb{R}$ is a parameter representing the risk-sensitivity degree of the criterion. In particular, φ_θ is convex (respectively, concave) when $\theta > 0$ (respectively, $\theta < 0$). When θ is sufficiently small, the problem can be well approximated by the standard stochastic control problem introduced in Section 1.

Another frequently used disutility function is the so-called *NARA (Non-constant Absolute Risk Aversion) utility*:

$$\varphi^\gamma(x) = \begin{cases} \frac{1}{\gamma}x^\gamma, & \gamma \neq 0, \\ \ln x, & \gamma = 0, \end{cases} \quad x > 0, \quad (1.17)$$

where γ is the risk parameter. The cases

$$\gamma < 1, \quad \gamma > 1, \quad \gamma = 1$$

correspond respectively to risk-seeking, risk-averse, and risk-neutral situations.

Chapter 2

Maximum Principle and Stochastic Hamiltonian Systems

2.1 Statement of the Stochastic Maximum Principle

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$ be a given filtered probability space satisfying the usual conditions, on which an m -dimensional standard Brownian motion $W(t)$ (with $W(0) = 0$) is given. We consider the following stochastic controlled system:

$$\begin{cases} dx(t) = b(t, x(t), u(t)) dt + \sigma(t, x(t), u(t)) dW(t), \\ x(0) = x_0, \end{cases} \quad (2.1)$$

with the cost functional

$$J(u(\cdot)) = \mathbb{E} \left[\int_0^T f(t, x(t), u(t)) dt + h(x(T)) \right], \quad t \in [0, T]. \quad (2.2)$$

In the above, $b : [0, T] \times \mathbb{R}^n \times U \rightarrow \mathbb{R}^n$, $\sigma : [0, T] \times \mathbb{R}^n \times U \rightarrow \mathbb{R}^{n \times m}$, $f : [0, T] \times \mathbb{R}^n \times U \rightarrow \mathbb{R}$, and, $h : \mathbb{R}^n \rightarrow \mathbb{R}$. We define

$$\begin{cases} b(t, x, u) = \begin{pmatrix} b^1(t, x, u) \\ \vdots \\ b^n(t, x, u) \end{pmatrix}, \\ \sigma(t, x, u) = (\sigma^1(t, x, u), \dots, \sigma^m(t, x, u)), \\ \sigma^j(t, x, u) = \begin{pmatrix} \sigma^{1j}(t, x, u) \\ \vdots \\ \sigma^{nj}(t, x, u) \end{pmatrix}, \quad 1 \leq j \leq m. \end{cases} \quad (2.3)$$

Let us make the following assumptions

(S0) $\{\mathcal{F}_t\}_{t \geq 0}$ is the natural filtration generated by $W(t)$, augmented by all the P -null sets in $\mathcal{F}$.

(S1) (U, d) is a separable metric space and $T > 0$.

(S2) The maps b, σ, f , and h are measurable, and there exist a constant $L > 0$ and a modulus of continuity $\bar{\omega} : [0, \infty) \rightarrow [0, \infty)$ such that for $\varphi(t, x, u) = b(t, x, u), \sigma(t, x, u), f(t, x, u), h(x)$, we have

$$\begin{cases} |\varphi(t, x, u) - \varphi(t, \hat{x}, \hat{u})| \leq L|x - \hat{x}| + \bar{\omega}(d(u, \hat{u})), & \forall t \in [0, T], x, \hat{x} \in \mathbb{R}^n, u, \hat{u} \in U, \\ |\varphi(t, 0, u)| \leq L, & \forall (t, u) \in [0, T] \times U \end{cases} \quad (2.4)$$

(S3) The maps b, σ, f , and h are C^2 in x . Moreover, there exist a constant $L > 0$ and a modulus of continuity $\bar{\omega} : [0, \infty) \rightarrow [0, \infty)$ such that for $\psi = b, \sigma, f, h$, we have

$$\begin{cases} |\psi_x(t, x, u) - \psi_x(t, \hat{x}, \hat{u})| \leq L|x - \hat{x}| + \bar{\omega}(d(u, \hat{u})) \\ |\psi_{xx}(t, x, u) - \psi_{xx}(t, \hat{x}, \hat{u})| \leq \bar{\omega}|x - \hat{x}| + d(u, \hat{u}), \end{cases} \quad \forall t \in [0, T], x, \hat{x} \in \mathbb{R}^n, u, \hat{u} \in U. \quad (2.5)$$

Assumption (S0) signifies that the system noise is the only source of uncertainty in the problem, and the past information about the noise is available to the controller. This assumption will be crucial below.

Now we define

$$\mathcal{U}[0, T] \triangleq \{u : [0, T] \times \Omega \rightarrow U \mid u \text{ is } \{\mathcal{F}_t\}_{t \geq 0} \text{-adapted}\}. \quad (2.6)$$

Given $u(\cdot) \in \mathcal{U}[0, T]$, equation (2.1) is an SDE with random coefficients. we see that under (S1)-(S2), for any $u(\cdot) \in \mathcal{U}[0, T]$, the state equation (2.1) admits a unique solution $x(\cdot) \equiv x(\cdot; u(\cdot))$ and the cost functional (2.2) is well-defined. In the case that $x(\cdot)$ is the solution of (2.1) corresponding to $u(\cdot) \in \mathcal{U}[0, T]$, we call $(x(\cdot), u(\cdot))$ an *admissible pair*, and $x(\cdot)$ an *admissible state* process (trajectory). Our optimal control problem can be stated as follows.

Problem (S). Minimize (2.2) over $\mathcal{U}[0, T]$.

Any $\bar{u}(\cdot) \in \mathcal{U}[0, T]$ satisfying

$$J(\bar{u}(\cdot)) = \inf_{u(\cdot) \in \mathcal{U}[0, T]} J(u(\cdot)) \quad (2.7)$$

is called an optimal control. The corresponding $\bar{x}(\cdot) \equiv x(\cdot; \bar{u}(\cdot))$ and $(\bar{x}(\cdot), \bar{u}(\cdot))$ are called an optimal state process/trajectory and optimal pair, Our next goal is to derive a set of necessary conditions for stochastic optimal controls, we need some preparations.

2.2 Adjoint equations

In this section we will introduce adjoint equations involved in a stochastic maximum principle and the associated stochastic Hamiltonian system. Recall that $S^n = \{A \in \mathbb{R}^{n \times n} \mid A^T = A\}$ and let $(x(\cdot), u(\cdot))$ be a given optimal pair.

We have seen that in the deterministic case the adjoint variable $p(\cdot)$ plays a central role in the maximum principle. The adjoint equation that $p(\cdot)$ satisfies is a backward ordinary differential equation (meaning that the terminal value is specified). It is nevertheless equivalent to a forward equation if we reverse the time. In the stochastic case, however, one cannot simply reverse the time, as it may destroy the nonanticipativeness of the solutions. Instead, we introduce the following terminal value problem for a stochastic differential

equation:

$$\begin{cases} dp(t) = -\left\{ b_x(t, \bar{x}(t), \bar{u}(t))^\top p(t) + \sum_{j=1}^m \sigma_x^j(t, \bar{x}(t), \bar{u}(t))^\top q^j(t) \right. \\ \quad \left. - f_x(t, \bar{x}(t), \bar{u}(t)) \right\} dt + q(t) dW(t), \quad t \in [0, T], \\ p(T) = -h_x(Y(T)). \end{cases} \quad (2.8)$$

Here the unknown is a pair of $\{F_t\}_{t \geq 0}$ -adapted processes $(p(\cdot), q(\cdot))$. We call the above a backward stochastic differential equation (BSDE, for short). The key issue here is that the equation is to be solved backwards (since the terminal value is given); however, the solution $(p(\cdot), q(\cdot))$ is required to be $\{F_t\}_{t \geq 0}$ -adapted. Any pair of processes $(p(\cdot), q(\cdot)) \in L^2_{\mathcal{F}}(0, T; \mathbb{R}^n) \times (L^2_{\mathcal{F}}(0, T; \mathbb{R}^n))^m$ satisfying (2.8) is called an adapted solution of (2.8). Under (S0)-(S3), for any $(\bar{x}(\cdot), \bar{u}(\cdot)) \in L^2_{\mathcal{F}}(0, T; \mathbb{R}^n) \times \mathcal{U}[0, T]$, (2.8) admits a unique adapted solution $(p(\cdot), q(\cdot))$. Note that (S0) cannot be omitted. The adjoint variable $p(\cdot)$ in the deterministic case corresponds to the so-called shadow price or the marginal value of the resource represented by the state variable in economic theory. The maximum principle is nothing but the so-called duality principle: Minimizing the total cost amounts to maximizing the total contribution of the marginal value [31]. Nevertheless, in the stochastic situation, the controller has to balance carefully the scale of control and the degree of uncertainty if a control made is going to affect the volatility of the system (i.e., if the diffusion coefficient depends on the control variable). Therefore, the marginal value alone may not be able to fully characterize the trade-off between the cost and control gain in an uncertain environment. One has to introduce another variable to reflect the uncertainty or the risk factor in the system. This is done by introducing an additional adjoint equation as follows:

$$\begin{cases} dP(t) = -\left\{ b_x(t, \bar{x}(t), \bar{u}(t))^\top P(t) + P(t) b_x(t, \bar{x}(t), \bar{u}(t)) \right. \\ \quad + \sum_{j=1}^m \sigma_x^j(t, \bar{x}(t), \bar{u}(t))^\top P(t) \sigma_x^j(t, \bar{x}(t), \bar{u}(t)) \\ \quad + \sum_{j=1}^m [\sigma_x^j(t, \bar{x}(t), \bar{u}(t))^\top Q^j(t) + Q^j(t) \sigma_x^j(t, \bar{x}(t), \bar{u}(t))] \\ \quad \left. + H_{xx}(t, \bar{x}(t), \bar{u}(t), p(t), q(t)) \right\} dt + \sum_{j=1}^m Q^j(t) dW^j(t), \\ P(T) = -h_{xx}(Y(T)), \end{cases} \quad (2.9)$$

2.3 The maximum principle and stochastic Hamiltonian system 25

where the Hamiltonian H is defined by

$$H(t, x, u, p, q) = \langle p, b(t, x, u) \rangle + \text{tr}[q^\top \sigma(t, x, u)] - f(t, x, u), \quad (2.10)$$

$$(t, x, u, p, q) \in [0, T] \times \mathbb{R}^n \times U \times \mathbb{R}^n \times \mathbb{R}^{n \times m},$$

and $(p(\cdot), q(\cdot))$ is the solution to (2.8). In the above (2.9), the unknown is again a pair of processes $(P(\cdot), Q(\cdot)) \in L^2_{\mathcal{F}}(0, T; S^n) \times (L^2_{\mathcal{F}}(0, T; S^n))^m$.

Incidentally, the Hamiltonian $H(t, x, u, p, q)$ defined by (2.10) coincides with $H(t, x, u, p)$ when $\sigma = 0$.

Note that equation (2.9) is also a BSDE with matrix-valued unknowns. As with (2.8), under assumptions (S0) – (S3), there exists a unique adapted solution $(P(\cdot), Q(\cdot))$ to (2.9). We refer to (2.8) (resp. (2.9)) as the first-order (resp. second-order) adjoint equations, and to $p(\cdot)$ (resp. $P(\cdot)$) as the first-order (resp. second-order) adjoint process. In what follows, if $(\bar{x}(\cdot), \bar{u}(\cdot))$ is an optimal (resp. admissible) pair, and $(p(\cdot), q(\cdot))$ and $(P(\cdot), Q(\cdot))$ are adapted solutions of (2.8) and (2.9), respectively, then $(\bar{x}(\cdot), \bar{u}(\cdot), p(\cdot), q(\cdot), P(\cdot), Q(\cdot))$ is called an optimal 6-tuple (resp. admissible 6-tuple).

2.3 The maximum principle and stochastic Hamiltonian systems

Now we are going to state the Pontryagin-type maximum principle for optimal stochastic controls. At first glance, it might be quite natural for one to expect that a stochastic maximum principle should maximize the Hamiltonian H defined by (2.10). Unfortunately, this is not true if the diffusion coefficient σ depends on the control. Here is an example.

Example 1. Consider the following control system ($n = m = 1$):

$$\begin{cases} dx(t) = u(t) dW(t), & t \in [0, 1], \\ x(0) = 0, \end{cases} \quad (2.11)$$

with the control domain being $U = [-1, 1]$ and the cost functional being

$$J(u(\cdot)) = \mathbb{E} \left[\int_0^1 \left(x(t)^2 - \frac{1}{2} u(t)^2 \right) dt + x(1)^2 \right]. \quad (2.12)$$

2.3 The maximum principle and stochastic Hamiltonian system 26

Substituting

$$x(t) = \int_0^t u(s) dW(s)$$

into the cost functional, we obtain via a simple calculation

$$J = \mathbb{E} \left[\int_0^1 \left(\frac{3}{2} - t \right) u(t)^2 dt \right]. \quad (2.13)$$

Hence, the optimal control is $\bar{u}(t) = 0$ with the optimal state trajectory $\bar{x}(t) \equiv 0$. However, the corresponding Hamiltonian is

$$H(t, x, u, p(t), q(t)) = -\frac{1}{2}u^2 + q(t)u. \quad (2.14)$$

This is a convex function in u , which does not attain a maximum at $\bar{u}(t) = 0$ for any time t .

To obtain the correct form of the stochastic maximum principle, one should add the risk adjustment, which is related to the diffusion coefficient, to the Hamiltonian to reflect the controller's risk-seeking or risk-averse attitude. To do this, let us introduce the generalized Hamiltonian

$$\begin{aligned} G(t, x, u, p, P) &\triangleq \frac{1}{2} \text{tr}(\sigma(t, x, u)^\top P \sigma(t, x, u)) + \langle p, b(t, x, u) \rangle - f(t, x, u), \\ (t, x, u, p, P) &\in [0, T] \times \mathbb{R}^n \times U \times \mathbb{R}^n \times S^n. \end{aligned} \quad (2.15)$$

Note that $G(t, x, u, p, P)$ depends on P , but not on q . Comparing (2.10), (2.15), we see that

$$\begin{cases} G(t, x, u, p, P) = H(t, x, u) + \text{tr}[\sigma(t, x, u)^\top P \sigma(t, x, u)], \\ H(t, x, u, p, q) = H(t, x, u) + \text{tr}[q^\top \sigma(t, x, u)]. \end{cases}$$

We need to pay attention to the difference between $G(t, x, u, p, P)$ and $H(t, x, u, p, q)$. The function $G(t, x, u, p, P)$ will appear in the HJB equation of stochastic optimal control problems studied in the next chapter.

Next, associated with an optimal 6-tuple $(\bar{x}(\cdot), \bar{u}(\cdot), p(\cdot), q(\cdot), P(\cdot), Q(\cdot))$, we define an $\mathcal{H}$ -function

2.3 The maximum principle and stochastic Hamiltonian system 27

$$\begin{aligned}
\mathcal{H}(t, x, u) &\triangleq H(t, x, u, p(t), q(t)) - \frac{1}{2} \text{tr} [\sigma(t, \bar{x}(t), \bar{u}(t))^T P(t) \sigma(t, \bar{x}(t), \bar{u}(t))] \\
&\quad + \frac{1}{2} \text{tr} [\sigma(t, x, u) - \sigma(t, \bar{x}(t), \bar{u}(t))]^T P(t) [\sigma(t, x, u) - \sigma(t, \bar{x}(t), \bar{u}(t))] \\
&\equiv -\frac{1}{2} \text{tr} [\sigma(t, x, u)^T P(t) \sigma(t, x, u)] + \langle p(t), b(t, x, u) \rangle - f(t, x, u) \\
&\quad + \text{tr} [q(t)^T \sigma(t, x, u)] - \text{tr} [\sigma(t, x, u)^T P(t) \sigma(t, \bar{x}(t), \bar{u}(t))] \\
&\equiv G(t, x, u, p(t), P(t)) + \sigma(t, x, u)^T [q(t) - P(t) \sigma(t, \bar{x}(t), \bar{u}(t))].
\end{aligned}
\tag{2.16}$$

Notice that an $\mathcal{H}$ -function may be defined similarly associated with any admissible 6-tuple $(x(\cdot), u(\cdot), p(\cdot), q(\cdot), P(\cdot), Q(\cdot))$.

Theorem 2.1. (Stochastic Maximum Principle) [31]

Let (S0) – (S3) hold. Let $(\bar{x}(\cdot), \bar{u}(\cdot))$ be an optimal pair of Problem (S).

Then there are pairs of processes

$$\begin{cases} (p(\cdot), q(\cdot)) \in L^2_{\mathcal{F}}(0, T; \mathbb{R}^n) \times (L^2_{\mathcal{F}}(0, T; \mathbb{R}^n))^m, \\ (P(\cdot), Q(\cdot)) \in L^2_{\mathcal{F}}(0, T; S^n) \times (L^2_{\mathcal{F}}(0, T; S^n))^m, \end{cases}
\tag{2.17}$$

where

$$\begin{cases} q(\cdot) = (q^1(\cdot), \dots, q^m(\cdot)), & Q(\cdot) = (Q^1(\cdot), \dots, Q^m(\cdot)), \\ q^j(\cdot) \in L^2_{\mathcal{F}}(0, T; \mathbb{R}^n), & Q^j(\cdot) \in L^2_{\mathcal{F}}(0, T; S^n), \quad 1 \leq j \leq m, \end{cases}
\tag{2.18}$$

satisfying the first-order and second-order adjoint equations (2.8) and (2.9), respectively, such that

$$\begin{aligned}
&H(t, \bar{x}(t), \bar{u}(t), p(t), q(t)) - H(t, \bar{x}(t), u, p(t), q(t)) \\
&\quad - \text{tr} [\sigma(t, Y(t), U(t)) - \sigma(t, \bar{x}(t), u)]^T P(t) [\sigma(t, \bar{x}(t), \bar{u}(t)) - \sigma(t, \bar{x}(t), u)] \geq 0,
\end{aligned}
\tag{2.19}$$

$$\forall u \in U, \text{ a.e. } t \in [0, T], \mathbb{P}\text{-a.s.}$$

or, equivalently,

$$\mathcal{H}(t, \bar{x}(t), \bar{u}(t)) = \max_{u \in U} \mathcal{H}(t, \bar{x}(t), u), \text{ a.e. } t \in [0, T], \mathbb{P}\text{-a.s.} \tag{2.20}$$

2.3 The maximum principle and stochastic Hamiltonian system 28

The inequality (2.19) is called the variational inequality, and (2.20) is called the maximum condition. Note that the third term on the left-hand side of (2.19) reflects the risk adjustment, which must be present when σ depends on u .

Now let us look at Example (1) again. The first-order and second-order adjoint equations associated with the optimal pair $(\bar{x}(t), \bar{u}(t)) \equiv (0, 0)$ are

$$\begin{cases} p(1) = 0, \\ dp(t) = q(t)dW(t), \end{cases} \quad (2.21)$$

and

$$\begin{cases} dP(t) = 2dt + Q(t)dW(t), \\ P(1) = -2. \end{cases} \quad (2.22)$$

Thus, $(p(t), q(t)) = (0, 0)$ and $(P(t), Q(t)) = (2t - 4, 0)$ are the (unique) adapted solutions. The left-hand side of (2.19) now reads

$$-\frac{1}{2}\bar{u}(t)^2 + q(t)\bar{u}(t) - \frac{1}{2}u^2 - q(t)u - \frac{1}{2}[\bar{u}(t) - u]^2 P(t) = \left(\frac{3}{2} - t\right)u^2 \geq 0, \quad \forall t \in [0, 1].$$

We may also calculate

$$\mathcal{H}(t, \bar{x}(t), u) = \frac{1}{2}(P(t) + 1)u^2 + q(t)u = \frac{1}{2}(2t - 3)u^2. \quad (2.23)$$

The above function $\mathcal{H}$ is concave in u for any $t \in [0, 1]$, and $\bar{u}(t) \equiv 0$ does maximize $\mathcal{H}$. Note that for the present case (see (2.14)),

$$H(t, Y(t), u, p(t), q(t)) = \frac{1}{2}u^2.$$

Thus, one sees clearly how the second-order adjoint process (which represents the risk factor) plays a role in turning the convex function $u \mapsto H(t, \bar{x}(t), u, p(t), q(t))$ into the concave one $u \mapsto \mathcal{H}(t, \bar{x}(t), u)$ shown above.

Let us single out two important special cases.

Case 1 : The diffusion does not contain the control variable, i.e.,

$$\sigma(t, x, u) \equiv \sigma(t, x), \quad \forall (t, x, u) \in [0, T] \times \mathbb{R}^n \times U. \quad (2.24)$$

2.3 The maximum principle and stochastic Hamiltonian system 29

In this case, the maximum condition (2.20) reduces to

$$H(t, \bar{x}(t), \bar{u}(t), p(t), q(t)) = \max_{u \in U} H(t, \bar{x}(t), u, p(t), q(t)), \quad \text{a.e. } t \in [0, T], \text{ } P\text{-a.s.} \quad (2.25)$$

which is parallel to the deterministic case (no risk adjustment is required). We note that in this case, equation (2.9) for $(P(\cdot), Q(\cdot))$ is not needed. Thus, the twice differentiability of the functions b, σ, f , and h in x is not necessary here.

Case 2: The control domain $U \subseteq \mathbb{R}^k$ is convex and all the coefficients are C^1 in u . Then (2.19) implies

$$\langle H_u(t, \bar{x}(t), \bar{u}(t), p(t), q(t)), u - \bar{u}(t) \rangle \leq 0, \forall u \in U, \quad \text{a.e. } t \in [0, T], \text{ } P\text{-a.s.} \quad (2.26)$$

This is called a local form (in contrast to the global form (2.19) or (2.20)) of the maximum principle. Note that the local form does not involve the second-order adjoint process $P(\cdot)$ either.

Analogous to the deterministic case, the system (2.1) along with its first-order adjoint system can be written as follows:

$$\begin{cases} dx(t) = H_p(t, x(t), u(t), p(t), q(t))dt + H_q(t, x(t), u(t), p(t), q(t))dW(t), \\ dp(t) = -H_x(t, x(t), u(t), p(t), q(t))dt + q(t)dW(t), \quad t \in [0, T], \\ x(0) = x_0, \quad p(T) = -h_x(x(T)). \end{cases} \quad (2.27)$$

The combination of (2.27), (2.9), and (2.19) (or (2.20)) is called an (extended) stochastic Hamiltonian system, with its solution being a 6-tuple

$$(x(\cdot), u(\cdot), p(\cdot), q(\cdot), P(\cdot), Q(\cdot)).$$

Therefore, we can rephrase Theorem 2.1 as the following.

Theorem 2.2. *Let (S0)-(S3) hold. Let Problem (S) admit an optimal pair $(\bar{x}(\cdot), \bar{u}(\cdot))$. Then the optimal 6-tuple $(\bar{x}(\cdot), \bar{u}(\cdot), p(\cdot), q(\cdot), P(\cdot), Q(\cdot))$ of Problem (S) solves the stochastic Hamiltonian system (2.27), (2.9), and (2.19) (or (2.20)).*

2.3.1 A worked-out example

Let us now give a more substantial example to demonstrate how Theorem 2.1 can help in finding an optimal control.

Example 2. Consider the following control system ($n = m = 1$):

$$\begin{cases} dx(t) = u(t)dt + u(t)dW(t), & t \in [0, 1], \\ x(0) = 1, \end{cases} \quad (2.28)$$

with the control domain being $U = \mathbb{R}$ and the cost functional being

$$J(u(\cdot)) = \mathbb{E} \left[-\frac{1}{2} \int_0^1 ru(t)^2 dt + \frac{1}{2} x(1)^2 \right]. \quad (2.29)$$

where $r \in (0, 1)$ satisfies

$$\ln r + 2 - r < 0 \quad (\text{or } 0 < r < 0.1586 \text{ approximately}). \quad (2.30)$$

Suppose $(\bar{x}(\cdot), \bar{u}(\cdot))$ is an optimal pair (which we are going to identify). Then the corresponding adjoint equations are

$$\begin{cases} dp(t) = q(t)dW(t), & t \in [0, 1], \\ p(1) = -\bar{x}(1), \end{cases} \quad (2.31)$$

and

$$\begin{cases} dP(t) = Q(t)dW(t), & t \in [0, 1], \\ P(1) = -1. \end{cases} \quad (2.32)$$

Clearly, $(P(t), Q(t)) = (-1, 0)$ is the only adapted solution to (3.32). The corresponding $\mathcal{H}$ -function is

$$\mathcal{H}(t, x, u) = -\frac{1}{2}(1-r)u^2 + (p(t) + q(t) + r\bar{u}(t))u. \quad (2.33)$$

Note that the above is a concave function of u due to the fact that $r < 1$. So by Theorem 3.2, a necessary condition for $\bar{u}(\cdot)$ to be optimal is

$$-(1-r)\bar{u}(t) + p(t) + q(t) + r\bar{u}(t) = 0,$$

$$\bar{u}(t) = -\frac{p(t) + q(t)}{r}. \quad (2.34)$$

By plugging (2.34) into (2.28) and putting it together with (2.32), we obtain the corresponding stochastic Hamiltonian system:

$$\begin{cases} d\bar{x}(t) = \frac{p(t) + q(t)}{r} dt + \frac{p(t) + q(t)}{r} dW(t), \\ dp(t) = q(t)dW(t), \\ \bar{x}(0) = 1, \quad p(1) = -\bar{x}(1). \end{cases} \quad (2.35)$$

Once we find an adapted solution $(\bar{x}(\cdot), p(\cdot), q(\cdot))$ of the above, a candidate for the optimal control can then be given by (2.34). We now try to find an adapted solution of (2.35). Suppose $(\bar{x}(\cdot), p(\cdot), q(\cdot))$ is an adapted solution such that the following relation holds:

$$p(t) = \theta(t)\bar{x}(t), \quad \forall t \in [0, 1], \quad P\text{-a.s.} \quad (2.36)$$

for some deterministic function $\theta(\cdot)$. We would like to determine the equation that $\theta(\cdot)$ should satisfy. To this end, we differentiate (2.36) using Itô's formula:

$$dp(t) = \left[\dot{\theta}(t)\bar{x}(t) - \theta(t)\frac{p(t)+q(t)}{r} \right] dt - \theta(t)\frac{p(t)+q(t)}{r} dW(t). \quad (2.37)$$

Comparing this with (2.31), we obtain

$$\begin{cases} \dot{\theta}(t)\bar{x}(t) = \frac{\theta(t)(p(t)+q(t))}{r}, \\ q(t) = -\frac{\theta(t)(p(t)+q(t))}{r} \end{cases} \quad (2.38)$$

Consequently

$$\begin{cases} q(t) = -\frac{p(t)\theta(t)}{r+\theta(t)} = \frac{\theta^2(t)\bar{x}(t)}{r+\theta(t)} \\ \frac{p(t)+q(t)}{r} = \frac{\theta(t)\bar{x}(t)}{r+\theta(t)} \end{cases} \quad (2.39)$$

Then, combining the first relation in (2.37) and the second in (2.38), we see that if $\theta(\cdot)$ is the right choice, it should satisfy the following:

$$\begin{cases} \dot{\theta}(t) = \frac{\theta(t)^2}{r+\theta(t)}, \quad t \in [0, 1] \\ \theta(t) = -1. \end{cases} \quad (2.40)$$

Now, suppose for the time being that (2.39) admits a solution $\theta(\cdot)$ on $[0, 1]$. Then it is necessary that (noting that $r + \theta(1) < 0$)

$$r + \theta(t) < 0, \quad \forall t \in [0, 1]. \quad (2.41)$$

In this case, we have a candidate for the optimal control:

$$\bar{u}(t) = \frac{\theta(t)}{r + \theta(t)} \bar{x}(t), \quad t \in [0, 1], \quad P\text{-a.s.} \quad (2.42)$$

which is in state feedback form. The corresponding state process is exactly $\bar{x}(t)$, which is the solution to the first equation in (2.35), or to the following:

$$\begin{cases} dY(t) = -\frac{\theta(t)}{r + \theta(t)} \bar{x}(t) dt - \frac{\theta(t)}{r + \theta(t)} \bar{x}, dW(t) \\ \bar{x}(0) = 1. \end{cases} \quad (2.43)$$

Let us now show that $(\bar{x}(\cdot), \bar{u}(\cdot))$ obtained is indeed an optimal pair. In fact, a simple calculation via It's formula yields that for any admissible pair $(x(\cdot), u(\cdot))$, comparing this with (2.31), we obtain

$$\left\{ d(\theta(t)x(t)^2) = \left\{ \theta(t)u(t)^2 + 2\theta(t)u(t)x(t) + \frac{\theta(t)^2 + x(t)^2}{r + \theta(t)} \bar{x}(t) dW(t) \right\} dt + \{ \dots \} dW(t) \right.$$

Integrating from 0 to 1, taking expectations, and noting (2.39) and (2.35), it follows that

$$\mathbb{E}x(1)^2 = -\theta(0) - \mathbb{E} \int_0^1 \left\{ \theta(t)u(t)^2 + 2\theta(t)u(t)x(t) + \frac{\theta(t)^2 + x(t)^2}{r + \theta(t)} \right\} dt. \quad (2.44)$$

Substituting this into the right-hand side of (2.29) yields

$$J(u(\cdot)) = -\frac{1}{2} \mathbb{E} \int_0^1 (r + \theta(t)) \left(u(t) + \frac{\theta(t)}{r + \theta(t)} x(t) \right)^2 dt - \frac{1}{2} \theta(0). \quad (2.45)$$

Therefore, by (2.40), the minimum value of the cost functional will be achieved at $\bar{u}(t) = -\frac{\theta(t)\bar{x}(t)}{r + \theta(t)}$, which is exactly given by (2.41).

We now show that equation (2.39) does admit a solution $\theta(\cdot)$ with the property (2.40). To this end, first note that (2.39) is equivalent to

$$\ln(-\theta(t)) - \frac{r}{\theta(t)} = t - 1 + r. \quad (2.46)$$

Define

$$f^t(\lambda) = \ln \lambda + \frac{r}{\lambda} - t + 1 - r, \quad \lambda \in (0, +\infty).$$

Noting (2.30), for all $t \in [0, 1]$, it follows that

$$f^t(r) = \ln r + 2 - t - r < 0,$$

$$f^t(1) = 1 - t > 0,$$

we conclude that there is $-\theta(t) \in (r, 1)$ (i.e., $r + \theta(t) < 0$) such that

$$f^t(-\theta(t)) = 0, \quad t \in [0, 1].$$

Moreover,

$$\frac{df^t(\lambda)}{d\lambda} = \frac{\lambda - r}{\lambda^2} \begin{cases} < 0, & \text{if } \lambda < r, \\ > 0, & \text{if } \lambda > r, \\ = 0, & \text{if } \lambda = r. \end{cases} \quad (2.47)$$

Thus the $\theta(t)$ satisfying $f^t(-\theta(t)) = 0$ and $r + \theta(t) < 0$ is unique. Finally,

$$f^1(1) = 0 \quad \text{implies} \quad \theta(1) = -1.$$

This establishes the existence of solutions to (2.39) subject to (2.40).

To conclude, an optimal feedback control is given by (2.41), and the optimal value is $-\frac{1}{2}\theta(0)$ (see (2.43) with $\theta(\cdot)$ solved via (2.39)).

Chapter 3

Machine Learning and Deep Learning

This chapter explores the foundational concepts of machine learning (ML) and deep learning (DL), providing both a detailed understanding for both researchers and beginners aiming to develop a solid grasp of when and how to use different models effectively. Our aim is to provide a comprehensive guide for choosing, tuning, and improving ML and DL models across various applications.

3.1 What Are Machine Learning and Deep Learning ?

Machine learning (ML) and deep learning (DL) have rapidly transformed from academic papers to real applications driving innovation across industries. From powering recommendation systems and personal assistants to enabling breakthroughs in healthcare, ML and DL have proven their ability to solve complex problems in ways traditional methods cannot. Despite the vast advancements in both fields, the foundational concepts are valuable for building robust, efficient, and interpretable models.

The importance of understanding these foundations cannot be overstated. Regardless of the reader's level of expertise, a solid grasp of the underlying

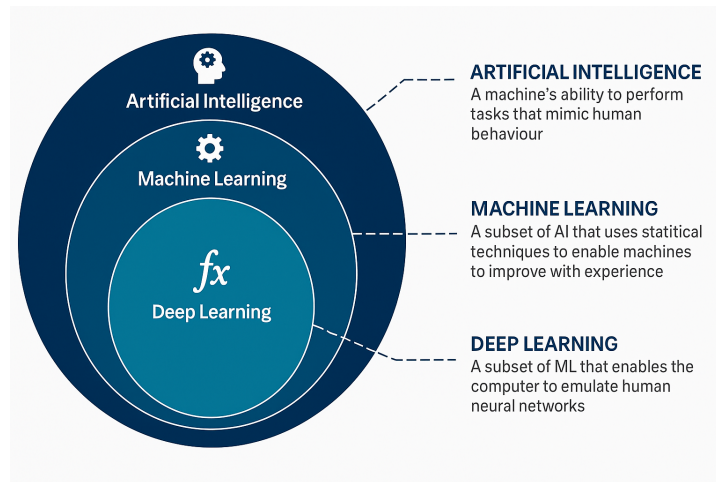


Figure 3.1: Machine Learning and Deep Learning

mechanisms of ML and DL models is essential for selecting the appropriate model for a given task, optimizing its performance, and ensuring its interpretability. In an era where complex models are often perceived as black boxes, this knowledge provides the necessary insight to build trust, evaluate, and refine AI systems effectively.

3.2 Foundations of Machine Learning

Machine learning overcomes such limitations. Generally speaking, ML means that a computer program's performance improves with experience with respect to some class of tasks and performance measures [30]. As such, it aims at automating the task of analytical model building to perform cognitive tasks like object detection or natural language translation. This is achieved by applying algorithms that iteratively learn from problem-specific training data, which allows computers to find hidden insights and complex patterns without explicitly being programmed [9]. Especially in tasks related to high-dimensional data such as classification, regression, and clustering, ML shows good applicability. By learning from previous computations and extracting regularities from massive databases, it can help to produce reliable and re-

peatable decisions. For this reason, ML algorithms have been successfully applied in many areas, such as fraud detection, credit scoring, next-best offer analysis, speech and image recognition, or natural language processing (NLP).

3.2.1 Types of Machine Learning

There are three main types of ML: supervised, unsupervised, and reinforcement learning.

- **Supervised learning:** Supervised learning requires a training data set that covers examples for the input as well as labeled answers or target values for the output. An example could be the prediction of active users subscribed to a market platform in a month's time as output (considered as the target variable or y variable) based on different input characteristics, such as the number of sold products or positive user reviews (often referred to as input features or x variables). The pairs of input and output data in the training set are then used to calibrate the open parameters of the ML model. Once the model has been successfully trained, it can be used to predict the target variable y given new or unseen data points of the input features x . Regarding the type of supervised learning, we can further distinguish between regression problems, where a numeric value is predicted (e.g., number of users), and classification problems, where the prediction result is a categorical class affiliation such as "lookers" or "buyers".
- **Unsupervised learning:** Unsupervised learning takes place when the learning system is supposed to detect patterns without any pre-existing labels or specifications. Thus, training data only consists of variables x with the goal of finding structural information of interest, such as groups of elements that share common properties (known as clustering) or data representations that are projected from a high-dimensional space into a lower one (known as dimensionality reduction) [9]. A prominent example of unsupervised learning in electronic markets is applying clustering techniques to group customers or markets into segments for the purpose of a more target-group specific communication.

- **Reinforcement learning**; In a reinforcement learning system, instead of providing input and output pairs, we describe the current state of the system, specify a goal, provide a list of allowable actions and their environmental constraints for their outcomes, and let the ML model experience the process of achieving the goal by itself using the principle of trial and error to maximize a reward.

3.2.2 Generalization

One of the big challenges in machine learning is to improve generalization, which is the ability to put general rules from seen data to deal with new, unseen data. Basically, we split our data into three subsets: the training, validation, and test sets. Knowing the difference between the training error and the validation error can help us identify two common phenomena: underfitting and overfitting. Underfitting occurs when a model cannot learn from the training data. When the model is able to perfectly predict the training data, it might begin to overfit the data, *i.e.* the model knows more than enough so it fails to generalize.

Another idea that is related to overfitting and underfitting is **bias-variance tradeoff**. Firstly, **bias error** happens when the learning algorithm makes incorrect assumptions, secondly, **variance** says how much our results will change if we train the model on a slightly different dataset.

3.2.3 Choosing the Right Model

Choosing the right machine learning model depends on the type of data, the problem, the required accuracy, interpretability, and available resources. The most advanced model is not always the best choice; the model must fit the specific context.

First, consider the data. For small datasets, simpler models such as linear regression or decision trees are often better because they are less likely to overfit. For large datasets, more complex models such as random forests or deep learning can discover complex patterns. The type and number of features also influence the choice. Regularization and dimensionality reduction

can help with high-dimensional data.

The type of problem is also important. For classification, models such as logistic regression and decision trees are common, while (**CNNs**) are useful for image recognition. For regression, linear models work well for simple relationships, while random forests or support vector regression are suitable for non-linear problems. For clustering, methods such as **k-means** and **DBSCAN** can be used.

Finally, there is a trade-off between performance and interpretability. Simple models are easier to understand, while complex models can provide better accuracy but are harder to interpret. Computational resources must also be considered, since complex models require more time and computing power.

3.3 Deep Learning

Deep learning drives many applications and services, including digital assistants, voice-enabled TV remotes, credit card fraud detection, self-driving cars, and generative AI. Unlike traditional machine learning, deep learning can handle unstructured data, such as text and images, without extensive preprocessing, also deep learning algorithms automate feature extraction, which reduces dependency on human expertise. As shown in Figure 3.2, the core of deep learning is a neural network that is made up of layers (input, hidden, and output) that contain nodes. Each node computes its output based on a set of weights (or parameters) applied to the output of the previous layer.

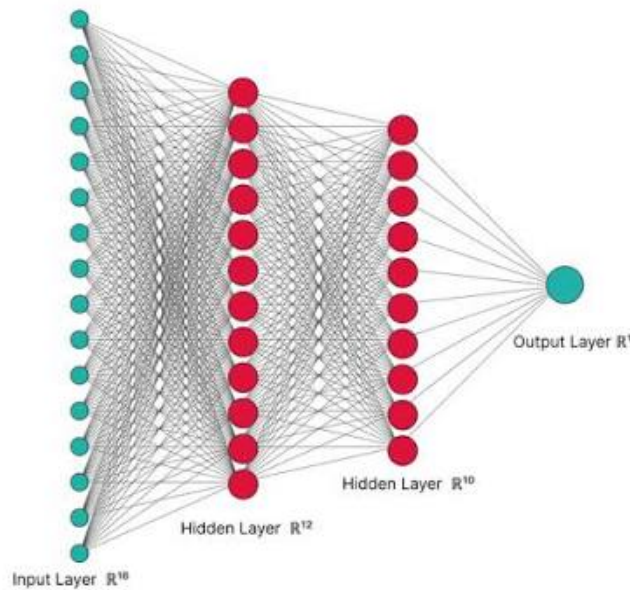


Figure 3.2: A simple neural network with two hidden layers (in red), and a one-dimensional output layer

3.3.1 Activation Functions

Activation function is a fancy way of saying that we are making the output of each neuron non linear because we want to learn non-linear relationships between the input and the output. Without a nonlinear activation function, even the deepest network is as good as a single one.

There are 3 types of activation functions: binary step function, linear function, and the most important one â non-linear functions.

binary step function depends on a threshold value that decides whether a neuron should be activated or not. Basically, we compare the inputs with a certain threshold; if the input is greater than it, then the neuron is activated, else it is deactivated, meaning that its output is not passed on to the next hidden layer.

linear function is another type of activation functions where the activation is proportional to the input. The function does not do anything to the weighted sum of the input, it simply outputs the value it was given.

3.3.2 Loss

Let's go back to supervised learning to understand some important concepts, such as loss and optimization. As a reminder, in supervised learning, we want to predict the label y given a data point x . Therefore, we make a prediction $\hat{y} = f(x)$. But how can we learn useful information from our prediction? First, we need to define our model and go to use simple linear regression as an example.

$$\hat{y} = w^T x,$$

with w is a vector containing the weights/parameters. To measure how good or bad our predictions are, we use a loss function. Common loss functions include **Mean Squared Error (MSE)** which squares the difference between predicted and actual values. Also, we have **Cross-Entropy Loss** which is mostly used for classification problems and measures the probability distribution difference between predictions and true labels. With our linear regression model, the mean squared error loss function is :

$$\hat{L}(X, y) = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2.$$

We minimize this loss by finding the point at which the gradient of the loss with respect to the weights w is zero (i.e., $\nabla_w \hat{L} = 0$)

3.3.3 Optimization

In any machine learning and deep learning model, we want to minimize the loss function by updating the parameters iteratively during training. In traditional machine learning, we can easily optimize the loss and reach the optimal point â the point with the least value of loss â but due to the number of parameters in neural networks, finding the global minimum becomes difficult.

So, to find a good optimizer, we need a more complicated technique like **gradient-based optimization** which finds the minimum by dividing the loss landscape into steps with size α or learning rate. A most basic gradient algorithm is **stochastic gradient descent (SGD)** where at each step, SGD randomly samples a batch of training data (a smaller group of the data

set) and computes the gradient loss with respect to the model parameters $\nabla_w \hat{L}(X_B; w)$.

Chapter 4

Deep Learning Methods for Stochastic Optimal Control

4.1 Motivation and Challenges

We mainly consider the following stochastic optimal control problem which was introduced in chapter 2.

$$\inf_{u(\cdot) \in \mathcal{U}_{ad}[0, T]} \mathbb{E} \left\{ \int_0^T f(t, x_t, u_t) dt + h(x_T) \right\}, \quad (4.1)$$

$$s.t. \quad x_t = x_0 + \int_0^t b(s, x_s, u_s) ds + \int_0^t \sigma(s, x_s, u_s) dW_s \quad (4.2)$$

The (extended) Hamiltonian system is characterized by the following FBSDE with a maximum condition [20]

$$\left\{ \begin{array}{l} dx_t^* = b(t, x_t^*, u_t^*) dt + \sigma(t, x_t^*, u_t^*) dW_t, \\ dp_t^* = -H_x(t, x_t^*, u_t^*, p_t^*, q_t^*) dt + q_t^* dW_t, \\ x_0^* = x, \quad p_T^* = -h_x(x_T^*), \\ H(t, x_t^*, u_t^*, p_t^*, q_t^*) = \max_{u \in U} H(t, x_t, u, p_t, q_t). \end{array} \right. \quad (4.3)$$

We first reformulate this later as a new optimal control problem,

$$\inf_{\tilde{p}_0, \{\tilde{q}_t\}_{0 \leq t \leq T}} \mathbb{E} \left[| -h_x(\tilde{x}(T)) - \tilde{p}_T |^2 \right] \quad (4.4)$$

$$\begin{aligned}
s.t. \quad \tilde{x}_t &= x_0 + \int_0^t b(s, \tilde{x}_s, \tilde{u}_s) ds + \int_0^t \sigma(s, \tilde{x}_s, \tilde{u}_s) dW_s, \\
\tilde{p}_t &= \tilde{p}_0 - \int_0^t H_x(s, \tilde{x}_s, \tilde{u}_s, \tilde{p}_s, \tilde{q}_s) + \int_0^t \tilde{q}_s dW_s, \\
\tilde{u}_s &= \arg \max_{u \in U} H(t, \tilde{x}_t, u, \tilde{p}_t, \tilde{q}_t),
\end{aligned}$$

where the process $\{\tilde{q}_t\}_{0 \leq t \leq T}$ and initial state $\tilde{p}_0$ are regarded as controls.

4.2 Background and problem formulation

Considering controlled stochastic differential equation (1.1), and the cost functional (1.3).

Our stochastic optimal control problem can be stated as minimizing (1.3) over $\mathcal{U}_{ad}[0, T]$. The goal is to find $u^*(\cdot) \in \mathcal{U}_{ad}[0, T]$ (if it exists) such that

$$J(u^*(\cdot)) = \inf_{u(\cdot) \in \mathcal{U}_{ad}[0, T]} \mathbb{E} \left\{ \int_0^T f(t, x_t, u_t) dt + h(x_T) \right\}. \quad (4.5)$$

Any $u^*(\cdot) \in \mathcal{U}_{ad}[0, T]$ satisfying (4.5) is called an optimal control.

Now, we reformulate the control problem (4.5) as a new problem based on the SMP and its corresponding stochastic Hamiltonian system. Considering the extended stochastic Hamiltonian system (4.3), which is essentially a coupled FBSDE with a maximum condition. Suppose that there exists a solution $(x^*(\cdot); u^*(\cdot); p^*(\cdot); q^*(\cdot))$ for FBSDE (4.3).

As is known, the FBSDE can be regarded as a stochastic optimal control problem [25]. Based on this idea, we have the following state equation with a maximum condition which is equivalent to (4.3),

$$\begin{cases} d\tilde{x}_t = b(t; \tilde{x}_t; \tilde{u}_t) dt + \sigma(t; \tilde{x}_t; \tilde{u}_t) dW_t, \\ d\tilde{p}_t = -H_x(t; \tilde{x}_t; \tilde{u}_t; \tilde{p}_t; \tilde{q}_t) dt + \tilde{q}_t dW_t, \\ \tilde{x}_0 = x_0; \quad \tilde{p}_0 = p_0, \\ H(t; \tilde{x}_t; \tilde{u}_t; \tilde{p}_t; \tilde{q}_t) = \max_{u \in U} H(t; \tilde{x}_t; u; \tilde{p}_t; \tilde{q}_t). \end{cases} \quad (4.6)$$

where $(\tilde{p}_0; \tilde{q})$ is the pair of control valued in $\mathbb{R}^n \times \mathbb{R}^{n \times d}$. As $\tilde{u}_t$ can be represented as

$$\tilde{u}_t = \arg \max_{u \in U} H(t; \tilde{x}_t; u; \tilde{p}_t; \tilde{q}_t), \quad (4.7)$$

then we get a new variational problem which is a reformulation of the control problem (4.5):

$$\begin{aligned}
& \inf_{\tilde{p}_0; \{\tilde{q}_t\}_{0 \leq t \leq T}} E[| -h_x(\tilde{x}_T) - \tilde{p}_T |^2] \quad (4.8) \\
& s.t. \tilde{x}_t = x_0 + \int_0^t b(t; \tilde{x}_s; \tilde{u}_s) ds + \int_0^t \sigma(t; \tilde{x}_s; \tilde{u}_s) dW_s, \\
& \tilde{p}_t = \tilde{p}_0 - \int_0^t H_x(s; \tilde{x}_s; \tilde{u}_s; \tilde{p}_s; \tilde{q}_s) ds + \int_0^t \tilde{q}_s dW_s, \\
& \tilde{u}_t = \arg \max_{u \in U} H(t; \tilde{x}_t; u; \tilde{p}_t; \tilde{q}_t).
\end{aligned}$$

According to the following Theorem 4.1 and Corollary 4.2.1, we can prove that the optimal control $\tilde{u}$ of (4.8) can be obtained when $E[| -h_x(\tilde{x}_T) - \tilde{p}_T |^2] = 0$. Supposing that $\bar{H}$ in (2.11) is differentiable in $x; p; q$, we have

$$\begin{cases} \bar{H}_p(t; x; p; q) = H_p(t; x; u^*; p; q) = b(t; x; u^*), \\ \bar{H}_q(t; x; p; q) = H_q(t; x; u^*; p; q) = \sigma(t; x; u^*), \\ \bar{H}_x(t; x; p; q) = H_x(t; x; u^*; p; q), \\ u^* = \arg \max_{u \in U} H(t; x; u; p; q), \end{cases} \quad (4.9)$$

for any $(t; x; u; p; q) \in [0; T] \times \mathbb{R}^n \times U \times \mathbb{R}^n \times \mathbb{R}^{n \times d}$. Then (2.27) can be rewritten as

$$\begin{cases} dx_t^* = \bar{H}_p(t; x_t^*; p_t^*; q_t^*) dt + \bar{H}_q(t; x_t^*; p_t^*; q_t^*) dW_t, \\ dp_t^* = -\bar{H}_x(t; x_t^*; p_t^*; q_t^*) dt + q_t^* dW_t, \\ x_0^* = x_0; \quad p_T^* = -h_x(x_T^*), \end{cases} \quad (4.10)$$

which is a FBSDE without constraint. And the problem (4.8) is equivalent to

$$\begin{aligned}
& \inf_{\tilde{p}_0; \{\tilde{q}_t\}_{0 \leq t \leq T}} E[| -h_x(\tilde{x}_T) - \tilde{p}_T |^2] \quad (4.11) \\
& s.t. \tilde{x}_t = x_0 + \int_0^t \bar{H}_p(s; \tilde{x}_s; \tilde{p}_s; \tilde{q}_s) ds + \int_0^t \bar{H}_q(s; \tilde{x}_s; \tilde{p}_s; \tilde{q}_s) dW_s, \\
& \tilde{p}_t = \tilde{p}_0 - \int_0^t \bar{H}_x(s; \tilde{x}_s; \tilde{p}_s; \tilde{q}_s) ds + \int_0^t \tilde{q}_s dW_s.
\end{aligned}$$

Define the cost functional of the stochastic optimal control problem (4.2) as

$$V(0; x_0) = \inf_{\tilde{p}_0; \{\tilde{q}_t\}_{0 \leq t \leq T}} E \left[|-h_x(\tilde{x}_T) - \tilde{p}_T|^2 \right],$$

Theorem 4.1. [20] Assume $h_x(x)$ is continuous in x , the map $\bar{H} = \max_{u \in \mathbf{U}} H(t, x, u, p, q)$ is differentiable in $x; p; q$, and its derivatives are continuous in $x; p; q$, such that for $\phi = \bar{H}_x; \bar{H}_p; \bar{H}_q$ and some constant $C > 0$,

$$|\phi(t; x; p; q) - \phi(t; x'; p'; q')| \leq C(|x - x'| + |p - p'| + |q - q'|),$$

$$|\phi(t; 0; 0; 0)|; \quad |\bar{H}_q(t; x; p; 0)| \leq C,$$

$$\forall x; x'; p; p' \in \mathbb{R}^n; \quad q; q' \in \mathbb{R}^{n \times d},$$

then the FBSDE (4.10) is solvable over $[0; T]$ if and only if $V(0; x_0) = 0$.

Corollary 4.2.1. [20] Suppose the assumptions in Theorem 4.1 hold, if there exists a solution $(\tilde{x}_t; \tilde{u}_t; \tilde{p}_t; \tilde{q}_t)_{0 \leq t \leq T}$ of (4.6) satisfying

$$E \left[|-h_x(\tilde{x}_T) - \tilde{p}_T|^2 \right] = 0, \quad (4.12)$$

then $(\tilde{x}_t; \tilde{u}_t; \tilde{p}_t; \tilde{q}_t)$ is a solution of (2.27), and the cost functional (1.4) can be obtained by

$$J(u^*(\cdot)) = J(\tilde{u}(\cdot)) = E \left\{ \int_0^T f(t; \tilde{x}_t; \tilde{u}_t) dt + h(\tilde{x}_T) \right\}. \quad (4.13)$$

4.3 Numerical algorithms

In this section, we propose three algorithms for solving the new variational problem through deep learning and show the neural network structure of Algorithm 1. The network structures of the other two algorithms are similar with that of Algorithm 1.

4.3.1 Algorithm 1: Numerical algorithm with 1-Net

Let π be a partition of the time interval, $0 = t_0 < t_1 < t_2 < \dots < t_{N-1} < t_N = T$ of $[0, T]$. Define $\Delta t_i = t_{i+1} - t_i$ and $\Delta W_{t_i} = W_{t_{i+1}} - W_{t_i}$, where $W_{t_i} \sim N(0, t_i)$, $i = 0, 1, 2, \dots, N-1$.

We also denote $\delta = \sup_{0 \leq i \leq N-1} \Delta t_i$.

Then the Euler scheme of the forward SDE (4.6) and the maximum condition (4.7) can be written as

$$\begin{cases} \tilde{x}_{t_{i+1}}^\pi = \tilde{x}_{t_i}^\pi + b(t_i, \tilde{x}_{t_i}^\pi, \tilde{u}_{t_i}^\pi) \Delta t_i + \sigma(t_i, \tilde{x}_{t_i}^\pi, \tilde{u}_{t_i}^\pi) \Delta W_{t_i}, \\ \tilde{p}_{t_{i+1}}^\pi = \tilde{p}_{t_i}^\pi - H_x(t_i, \tilde{x}_{t_i}^\pi, \tilde{u}_{t_i}^\pi, \tilde{p}_{t_i}^\pi, \tilde{q}_{t_i}^\pi) \Delta t_i + \tilde{q}_{t_i}^\pi \Delta W_{t_i}, \\ \tilde{x}_0^\pi = x_0, \quad \tilde{p}_0^\pi = \tilde{p}_0, \\ \tilde{u}_{t_i}^\pi = \arg \max_{u \in U} H(t_i, \tilde{x}_{t_i}^\pi, u, \tilde{p}_{t_i}^\pi, \tilde{q}_{t_i}^\pi). \end{cases} \quad (4.14)$$

We regard $\{\tilde{q}_{t_i}^\pi\}_{0 \leq i < N}$ as a control and assume it satisfies

$$\tilde{q}_{t_i}^\pi = \phi_1(t_i, \tilde{x}_{t_i}^\pi, \tilde{u}_{t_i}^\pi, \tilde{p}_{t_i}^\pi, \theta_{t_i}^1). \quad (4.15)$$

where $\tilde{q}_{t_i}^\pi$ is a feedback control of the states $\tilde{x}_{t_i}^\pi, \tilde{p}_{t_i}^\pi$ and the control $\tilde{u}_{t_i}^\pi$.

Here we give some explanations to recognize the difference between the controls $\tilde{u}_{t_i}^\pi$ and $\tilde{q}_{t_i}^\pi$.

$\tilde{u}_{t_i}^\pi$ is the control of the original control problem (1.1) where it should be solved by the maximum condition (4.7) in (4.6). And $\tilde{q}_{t_i}^\pi$ is the control of the new control problem (4.8).

Note that the processes $\tilde{q}_{t_i}^\pi$ and $\tilde{u}_{t_i}^\pi$ are interdependent according to (4.7) and (4.15). Plugging (4.7) in (4.15) and according to the implicit function theorem, we can get

$$\begin{aligned} \tilde{q}_{t_i}^\pi &= \phi(t_i, \tilde{x}_{t_i}^\pi, \tilde{p}_{t_i}^\pi, \theta_{t_i}^1), \\ &= \phi^1(t_i, \tilde{x}_{t_i}^\pi, \arg \max_{u \in U} H(t_i, \tilde{x}_{t_i}^\pi, u, \tilde{p}_{t_i}^\pi, \tilde{q}_{t_i}^\pi), \tilde{p}_{t_i}^\pi, \theta_{t_i}^1), \end{aligned} \quad (4.16)$$

where ϕ is a new unknown function.

We develop a neural network (1-NNet) for simulating the feedback control $\tilde{q}^\pi$.

Different from our previous work in [21], a single network is constructed for all the time-points and the time t_i is regarded as an input of the neural network.

The network consists of five layers including one $(1 + n + n)$ -dimensional input layer, three $(10 + n + n)$ -dimensional hidden layers and a $(n \times d)$ -dimensional output layer.

All parameters of the network are represented as θ .

Algorithm 1 Numerical algorithm with 1-NNet

```

1: Input: The Brownian motion  $\Delta W_{t_i}$ , initial parameters  $(\theta^0, \tilde{p}_0^{0,\pi})$ , learning rate  $\eta$ ;
2: Output: The 4-tuple processes  $(\tilde{x}_{t_i}^{l,\pi}, \tilde{u}_{t_i}^{l,\pi}, \tilde{p}_{t_i}^{l,\pi}, \tilde{q}_{t_i}^{l,\pi})$ .
3: for  $l = 0$  to  $maxstep$  do
4:    $\tilde{x}_{t_0}^{l,\pi} = x_0, \tilde{p}_{t_0}^{l,\pi} = \tilde{p}_0^{l,\pi};$ 
5:   for  $i = 0$  to  $N - 1$  do
6:      $\tilde{q}_{t_i}^{l,\pi} = \phi(t_i, \tilde{x}_{t_i}^{l,\pi}, \tilde{p}_{t_i}^{l,\pi}; \theta^l);$ 
7:      $\tilde{u}_{t_i}^{l,\pi} = \arg \max_{u \in U} H(t_i, \tilde{x}_{t_i}^{l,\pi}, u, \tilde{p}_{t_i}^{l,\pi}, \tilde{q}_{t_i}^{l,\pi});$ 
8:      $\tilde{x}_{t_{i+1}}^{l,\pi} = \tilde{x}_{t_i}^{l,\pi} + b(t_i, \tilde{x}_{t_i}^{l,\pi}, \tilde{u}_{t_i}^{l,\pi})\Delta t_i + \sigma(t_i, \tilde{x}_{t_i}^{l,\pi}, \tilde{u}_{t_i}^{l,\pi})\Delta W_{t_i};$ 
9:      $\tilde{p}_{t_{i+1}}^{l,\pi} = \tilde{p}_{t_i}^{l,\pi} - H_x(t_i, \tilde{x}_{t_i}^{l,\pi}, \tilde{u}_{t_i}^{l,\pi}, \tilde{p}_{t_i}^{l,\pi}, \tilde{q}_{t_i}^{l,\pi})\Delta t_i + \tilde{q}_{t_i}^{l,\pi}\Delta W_{t_i};$ 
10:   end for
11:    $J(\tilde{u}^{l,\pi}(\cdot)) = \frac{1}{M} \sum_{j=1}^M \left[ \frac{T}{N} \sum_{i=0}^{N-1} f(t_i, \tilde{x}_{t_i}^{l,\pi}, \tilde{u}_{t_i}^{l,\pi}) + h(\tilde{x}_T^{l,\pi}) \right];$ 
12:    $loss = \frac{1}{M} \sum_{j=1}^M \left[ | -h_x(\tilde{x}_T^{l,\pi}) - \tilde{p}_T^{l,\pi} |^2 \right];$ 
13:    $(\theta^{l+1}, \tilde{p}_0^{l+1,\pi}) = (\theta^l, \tilde{p}_0^{l,\pi}) - \eta \nabla loss.$ 
14: end for

```

Remark There is one difficulty in this algorithm. As shown in line 5 of Algorithm 1, the solution of the maximum condition is an extremum problem of multivariate functions and in most cases it has no analytical solution, which means that we can not get the explicit value of $\tilde{u}_{t_i}^{l,\pi}$ to calculate the forward process. When the explicit solution is not available, there are some ways to get the approximated solution, such as the BFGS and its extended methods [23], the gradient descent methods, the Sequential Least Squares Programming (SLSQP) and so on. On the other hand, when the explicit solution of $\tilde{u}$ is available, the Hamiltonian system (4.6) is equivalent to a FBSDE, and this situation will be discussed in subsection 4.3.3.

4.3.2 Algorithm 2: Numerical algorithm with 2-NNets

As mentioned in Algorithms 1, when the explicit solution of the maximum condition is not available, some approximation methods should be used. However, it is very time consuming to calculate the approximate maximum condition for high-dimensional cases. In order to solve this problem, we

develop a numerical algorithm with two neural networks (2-NNets) in this subsection. Here we consider a kind of stochastic optimal control problem, where the convex control domain $U = \mathbb{R}^k$, all the coefficients are C^1 in u and the optimal control $\tilde{u}$ falls inside the boundary of the control domain, then $H(t, x_t^*, u_t^*, p_t^*, q_t^*) = 0$ implies

$$H_u(t; x_t^*; u_t^*; p_t^*; q_t^*) = 0; \quad \forall u \in U, \quad a.e. \ t \in [0; T], \quad P\text{-a.s.} \quad (4.18)$$

Thus the corresponding stochastic Hamiltonian system (2.27) can be represented as

$$\begin{cases} dx_t^* = b(t; x_t^*; u_t^*)dt + \sigma(t; x_t^*; u_t^*)dW_t, \\ dp_t^* = -H_x(t; x_t^*; u_t^*; p_t^*; q_t^*)dt + q_t^*dW_t, \\ x_0^* = x_0, \quad p_T^* = -h_x(x_T^*), \\ H_u(t; x_t^*; u_t^*; p_t^*; q_t^*) = 0, \end{cases} \quad \begin{matrix} t \in [0; T], \\ \\ \\ \forall u \in U. \end{matrix} \quad (4.19)$$

The Hamiltonian system (4.19) with constraint $H_u(t; x_t^*; u_t^*; p_t^*; q_t^*) = 0$ was initially considered by [31, 3], see also [4], and correspond to a wide range of stochastic optimal control problems. For solving the Hamiltonian system (4.19), we reformulate it by the following new control problem

$$\inf_{\tilde{p}_0; \{\tilde{q}_t\}_{0 \leq t \leq T}; \{\tilde{u}_t\}_{0 \leq t \leq T}} E \left[|-h_x(\tilde{x}_T) - \tilde{p}_T|^2 + \lambda \int_0^T H_u(t; \tilde{x}_t; \tilde{u}_t; \tilde{p}_t; \tilde{q}_t)^2 dt \right]; \quad (4.20)$$

$$\begin{aligned} s.t. \tilde{x}_t &= x_0 + \int_0^t b(s; \tilde{x}_s; \tilde{u}_s)ds + \int_0^t \sigma(s; \tilde{x}_s; \tilde{u}_s)dW_s, \\ \tilde{p}_t &= \tilde{p}_0 - \int_0^t H_x(s; \tilde{x}_s; \tilde{u}_s; \tilde{p}_s; \tilde{q}_s)ds + \int_0^t \tilde{q}_s dW_s, \end{aligned}$$

where $\tilde{p}_0; \{\tilde{q}_t\}_{0 \leq t \leq T}; \{\tilde{u}_t\}_{0 \leq t \leq T}$ are the controls and λ is the hyper-parameter. As long as the cost functional (4.20) converges to 0, the 4-tuple $(\tilde{x}_t; \tilde{u}_t; \tilde{p}_t; \tilde{q}_t)$ converges to $(x_t^*; u_t^*; p_t^*; q_t^*)$.

The discrete Euler scheme is given as

$$\begin{cases} \tilde{x}_{t_i+1}^\pi = \tilde{x}_{t_i}^\pi + b(t_i; \tilde{x}_{t_i}^\pi; \tilde{u}_{t_i}^\pi)\Delta t_i + \sigma(t_i; \tilde{x}_{t_i}^\pi; \tilde{u}_{t_i}^\pi)\Delta W_{t_i}, \\ \tilde{p}_{t_i+1}^\pi = \tilde{p}_{t_i}^\pi - H_x(t_i; \tilde{x}_{t_i}^\pi; \tilde{u}_{t_i}^\pi; \tilde{p}_{t_i}^\pi; \tilde{q}_{t_i}^\pi)\Delta t_i + \tilde{q}_{t_i}^\pi \Delta W_{t_i}, \\ \tilde{x}_0^\pi = x_0, \quad \tilde{p}_0^\pi = \tilde{p}_0, \end{cases} \quad (4.21)$$

and the loss function is

$$loss = \frac{1}{M} \sum_{j=1}^M \left[| -h_x(\tilde{x}_T^\pi) - \tilde{p}_T^\pi |^2 + \lambda \sum_{i=0}^{N-1} H_u(t_i, \tilde{x}_{t_i}^\pi, \tilde{u}_{t_i}^\pi, \tilde{p}_{t_i}^\pi, \tilde{q}_{t_i}^\pi)^2; \right] \quad (4.22)$$

where the time-divided coefficient $\frac{T}{N}$ is merged into the coefficient λ .

Different from Algorithms 1, we regard the two processes $\{\tilde{q}_{t_i}^\pi; \tilde{u}_{t_i}^\pi\}_{0 \leq i \leq N-1}$ as controls, which means that two neural networks (2-NNets) should be constructed to simulate $\tilde{q}_{t_i}^\pi; \tilde{u}_{t_i}^\pi$, respectively.

And similar with Algorithms 1, we regard $\tilde{q}_{t_i}^\pi; \tilde{u}_{t_i}^\pi$ as feedback controls of the state $\tilde{x}_{t_i}^\pi$ and the time t_i , then we construct a common neural network for all time steps, respectively,

$$\begin{aligned} \tilde{q}_{t_i}^\pi &= \phi_1(t_i; \tilde{x}_{t_i}^\pi; \theta_q), \\ \tilde{u}_{t_i}^\pi &= \phi_2(t_i; \tilde{x}_{t_i}^\pi; \theta_u). \end{aligned} \quad (4.23)$$

The 2-NNets contain both one $(n+1)$ -dim input layers and three $(n+10)$ -dim hidden layers, the output layers are $(n \times d)$ -dim for $\tilde{q}_{t_i}^\pi$ and k -dim for $\tilde{u}_{t_i}^\pi$ respectively. The loss function is given as (4.22).

The pseudo-code is given as Algorithm 2.

Algorithm 2 Numerical algorithm with 2-NNets

- 1: **Input:** The Brownian motion ΔW_{t_i} , initial parameters $(\theta^{q,0}, \theta^{u,0}, \tilde{p}_0^{0,\pi})$, learning rate η and hyperparameter λ ;
 - 2: **Output:** The processes $\tilde{x}_{t_i}^{l,\pi}$ and $\tilde{p}_T^{l,\pi}$.
 - 3: **for** $l = 0$ **to** $maxstep$ **do**
 - 4: $\tilde{x}_{t_0}^{l,\pi} = x_0$, $\tilde{p}_{t_0}^{l,\pi} = \tilde{p}_0^{l,\pi}$ and $H = 0$;
 - 5: **for** $i = 0$ **to** $N - 1$ **do**
 - 6: $\tilde{q}_{t_i}^{l,\pi} = \phi_1(t_i, \tilde{x}_{t_i}^{l,\pi}; \theta^{q,l})$;
 - 7: $\tilde{u}_{t_i}^{l,\pi} = \phi_2(t_i, \tilde{x}_{t_i}^{l,\pi}; \theta^{u,l})$;
 - 8: $\tilde{x}_{t_{i+1}}^{l,\pi} = \tilde{x}_{t_i}^{l,\pi} + b(t_i, \tilde{x}_{t_i}^{l,\pi}, \tilde{u}_{t_i}^{l,\pi})\Delta t_i + \sigma(t_i, \tilde{x}_{t_i}^{l,\pi}, \tilde{u}_{t_i}^{l,\pi})\Delta W_{t_i}$;
 - 9: $\tilde{p}_{t_{i+1}}^{l,\pi} = \tilde{p}_{t_i}^{l,\pi} - H_x(t_i, \tilde{x}_{t_i}^{l,\pi}, \tilde{u}_{t_i}^{l,\pi}, \tilde{p}_{t_i}^{l,\pi}, \tilde{q}_{t_i}^{l,\pi})\Delta t_i + \tilde{q}_{t_i}^{l,\pi}\Delta W_{t_i}$;
 - 10: $H = H + |H_u(t_i, \tilde{x}_{t_i}^{l,\pi}, \tilde{u}_{t_i}^{l,\pi}, \tilde{p}_{t_i}^{l,\pi}, \tilde{q}_{t_i}^{l,\pi})|^2$;
 - 11: **end for**
 - 12: $J(\tilde{u}^{l,\pi}(\cdot)) = \frac{1}{M} \sum_{j=1}^M \left[\frac{T}{N} \sum_{i=0}^{N-1} f(t_i, \tilde{x}_{t_i}^{l,\pi}, \tilde{u}_{t_i}^{l,\pi}) + h(\tilde{x}_T^{l,\pi}) \right]$;
 - 13: $loss = \frac{1}{M} \sum_{j=1}^M \left[| - h_x(\tilde{x}_T^{l,\pi}) - \tilde{p}_T^{l,\pi} |^2 + \lambda H \right]$;
 - 14: $(\theta^{q,l+1}, \theta^{u,l+1}, \tilde{p}_0^{l+1,\pi}) = (\theta^{q,l}, \theta^{u,l}, \tilde{p}_0^{l,\pi}) - \eta \nabla loss$.
 - 15: **end for**
-

In Algorithm 2, instead of solving the maximum condition explicitly or approximately, we only need to consider the control condition $H_u(t; x; u; p; q) = 0$. Thus Algorithm 2 can deal with a wide range of high-dimensional problems more effectively even if the optimal control $\tilde{u}$ can not be solved explicitly.

4.3.3 Algorithm 3: Numerical Algorithm with explicit expression of $\bar{H}$

The above mentioned Algorithm 2 provides a method for solving a general kind of high-dimensional stochastic optimal control problems when the optimal control $\tilde{u}$ has not an explicit solution. In this subsection, we introduce another algorithm for solving high-dimensional cases. We show that as long as the function $\bar{H}$ is known, we can solve a class of high-dimensional stochastic optimal control problems through the deep-learning method established in our previous work [22].

In more details, we consider the stochastic optimal control problem (1.4)

and assume that $\bar{H}$ is given. Then the corresponding Hamiltonian system with $\bar{H}$ is defined as follows

$$\begin{cases} dx_t^* = \bar{H}_p(t; x_t^*; p_t^*; q_t^*)dt + \bar{H}_q(t; x_t^*; p_t^*; q_t^*)dW_t, \\ dp_t^* = -\bar{H}_x(t; x_t^*; p_t^*; q_t^*)dt + q_t^*dW_t, & t \in [0; T], \\ x_0^* = x_0, & p_T^* = -h_x(x_T^*), \end{cases} \quad (4.24)$$

which is essentially an FBSDE. If (4.24) satisfies the monotonic conditions [29, 26], we can solve it in high dimensions through the deep-learning method proposed in [22] for solving the FBSDEs. But different with that in [22], a single DNN is constructed for all the time points to improve the computing efficiency, as is in Algorithm 1 and 2. Then the optimal state processes $(x^*(\cdot); p^*(\cdot); q^*(\cdot))$ can be obtained.

Finally we get the optimal control $u^*(\cdot)$ through the maximum condition,

$$u_t^* = \arg \max_{u \in U} H(t; x_t^*; u; p_t^*; q_t^*), \quad (4.25)$$

where H is given in (2.10).

The pseudo-code is shown in Algorithm 3. To solve the extremum problem described at line 11 in Algorithm 3, we can use the similar methods as that in Algorithm 1, such as BFGS, SLSQP. However, an important difference is that we need to calculate the extremum problem only once in the whole algorithm, as the optimal state processes $(x^*(\cdot); p^*(\cdot); q^*(\cdot))$ have already been obtained by solving the Hamiltonian system (4.24) with deep learning.

Algorithm 3 Numerical algorithm with explicit expression of H

Input: The Brownian motion ΔW_{t_i} , initial parameters $(\theta^0, \tilde{p}_0^{0,\pi})$, learning rate η ;

Output: The triple processes $(\tilde{x}_{t_i}^{l,\pi}, \tilde{p}_{t_i}^{l,\pi}, \tilde{q}_{t_i}^{l,\pi})$.

for $l = 0$ **to** $maxstep$ **do**

$$\tilde{x}_{t_0}^{l,\pi} = x_0, \tilde{p}_{t_0}^{l,\pi} = \tilde{p}_0^{l,\pi};$$

for $i = 0$ **to** $N - 1$ **do**

$$\tilde{q}_{t_i}^{l,\pi} = \phi(t_i, \tilde{x}_{t_i}^{l,\pi}, \tilde{p}_{t_i}^{l,\pi}; \theta^l);$$

$$\tilde{x}_{t_{i+1}}^{l,\pi} = \tilde{x}_{t_i}^{l,\pi} + H_p(t_i, \tilde{x}_{t_i}^{l,\pi}, \tilde{p}_{t_i}^{l,\pi}, \tilde{q}_{t_i}^{l,\pi})\Delta t_i + H_q(t_i, \tilde{x}_{t_i}^{l,\pi}, \tilde{p}_{t_i}^{l,\pi}, \tilde{q}_{t_i}^{l,\pi})\Delta W_{t_i};$$

$$\tilde{p}_{t_{i+1}}^{l,\pi} = \tilde{p}_{t_i}^{l,\pi} - H_x(t_i, \tilde{x}_{t_i}^{l,\pi}, \tilde{p}_{t_i}^{l,\pi}, \tilde{q}_{t_i}^{l,\pi})\Delta t_i + \tilde{q}_{t_i}^{l,\pi}\Delta W_{t_i};$$

end for

$$\text{loss} = \frac{1}{M} \sum_{j=1}^M \left[| -h_x(\tilde{x}_T^{l,\pi}) - \tilde{p}_T^{l,\pi} |^2 \right];$$

$$(\theta^{l+1}, \tilde{p}_0^{l+1,\pi}) = (\theta^l, \tilde{p}_0^{l,\pi}) - \eta \nabla \text{loss};$$

end for

$$\tilde{u}_{t_i}^{l,\pi} = \arg \max_{u \in U} H(t_i, \tilde{x}_{t_i}^{l,\pi}, u, \tilde{p}_{t_i}^{l,\pi}, \tilde{q}_{t_i}^{l,\pi});$$

$$J(\tilde{u}^{l,\pi}(\cdot)) = \frac{1}{M} \sum_{j=1}^M \left[\frac{T}{N} \sum_{i=0}^{N-1} f(t_i, \tilde{x}_{t_i}^{l,\pi}, \tilde{u}_{t_i}^{l,\pi}) + h(\tilde{x}_T^{l,\pi}) \right];$$

The choice of algorithm can be determined according to the nature of the function H in the maximum condition and the character of the stochastic optimal control problem. When $\tilde{u}$ can be solved explicitly, it's better to choose Algorithm 3, then the optimal stochastic control problem is degenerated into the problem of solving FBSDEs. When $\tilde{u}$ does not have an explicit representation, all of the three proposed algorithms can be chosen.

4.4 Numerical results

In this section, we apply our proposed algorithms in solving some stochastic optimal control problems and show the numerical results. Firstly, we give a low-dimensional example to compare the performance of the three algorithms. Then we show some high dimensional cases and give the corresponding results. If not specially mentioned, we set the dimensions $n = k$. And all the examples in this section are calculated with the number of time-points $N = 25$ and the batch size of 64. The learning rates and the sample

number in the training set varies to adjust to different examples. The sample number in the test set is set to be 512. The ReLU activation function and Adam optimizer are adopted in the network architectures and the data is normalized before each layer. In order to get more general results, we calculate the means of numerical results from 15 independent runs of each algorithm.

4.4.1 A low-dimensional example

Consider the following LQ control system:

$$\begin{cases} dx_t = \left(-\frac{1}{4}x_t + u_t\right) dt + \left(\frac{1}{5}x_t + u_t\right) dW_t, \\ x(0) = x_0, \end{cases} \quad (4.26)$$

and the cost functional is defined as

$$J(0; x_0; u(\cdot)) = E\left(\frac{1}{2} \int_0^T \left[\left\langle \frac{1}{2}x_t; x_t \right\rangle + \langle 2u_t; u_t \rangle\right] dt + \frac{1}{2} \langle Qx_T; x_T \rangle\right); \quad (4.27)$$

where we set the dimensions $n = k$; $d = 1$ and Q is deterministic matrix taking value in $\mathbb{R}^{n \times n}$.

The control domain is $U = \mathbb{R}^n$, and the Hamiltonian H is

$$H(t; x; u; p; q) = \left\langle p; -\frac{1}{4}x + u \right\rangle + \left\langle q; \frac{1}{5}x + u \right\rangle - \frac{1}{4} \langle x; x \rangle - \langle u; u \rangle.$$

The explicit representation of the optimal control u^* is

$$u^* = \frac{1}{2}(p + q).$$

Therefore the function $\bar{H}$ has an explicit form

$$\bar{H}(x; p; q) = -\frac{1}{2} \left\langle \frac{1}{2}x; x \right\rangle + \left\langle p; -\frac{1}{4}x \right\rangle + \left\langle q; \frac{1}{5}x \right\rangle + \frac{1}{4} \langle p + q; p + q \rangle.$$

The corresponding Hamiltonian system is

$$\begin{cases} dx_t^* = \left(-\frac{1}{4}x_t^* + u_t^*\right) dt + \left(\frac{1}{5}x_t^* + u_t^*\right) dW_t, \\ -dp_t^* = \left(-\frac{1}{2}x_t^* - \frac{1}{4}p_t^* + \frac{1}{5}q_t^*\right) dt - q_t^* dW_t, \\ x_0^* = x_0, \quad p_T^* = -Qx_T^*, \\ u_t^* = \frac{1}{2}(p_t^* + q_t^*). \end{cases} \quad (4.28)$$

We set $x_0 = 1$ in this example. It can be verified that equation (4.28) satisfies the monotonic condition and has a unique solution $(x^*(\cdot); p^*(\cdot); q^*(\cdot))$. The reader can check it in **Appendix B** [20].

Supposing the solution of FBSDE (4.28) is in the following form:

$$p_t^* = -K_t x_t^*; \quad q_t^* = -M_t x_t^*.$$

Combing it with (4.28), we obtain a Riccati equation

$$\begin{cases} \dot{K}_t - \frac{1}{2}K_t^2 - \frac{1}{2}K_t + (\frac{1}{5}E_n - \frac{1}{2}K_t)M_t + \frac{1}{2}E_n = 0, \\ \frac{1}{2}K_t^2 - \frac{1}{5}K_t + \frac{1}{2}K_t M_t + M_t = 0, \\ K_T = Q, \end{cases} \quad (4.29)$$

where $\dot{K}_t$ is the derivative of K_t with respect to t , E_n is an n -order unit matrix. Equation (4.29) is a deterministic one-order ordinary differential equation and we can get its numerical solution with the four-order Runge-Kutta methods by using the ODE45 method in Matlab (ODE45 in brief for easy expression). Therefore the numerical solutions of ODE45 is used as a benchmark to be compared with that of our algorithms in the LQ control problem.

Firstly, we set $T = 0.1$, $Q = E_n$. The numerical solution of equation (4.29) with ODE45 is $K_0 = 0.9586E_n$ for each dimension, and the value of p_0^* with ODE45 is $p_0^* = -K(0)x_0 = -0.9586$; i.e. p_0^* is a n -dim vector with all its elements equal to -0.9586 .

We give the numerical results comparison of two proposed algorithms in Table 4.1 with 3000 iteration steps when $n = 5$. And the numerical solution $p_0^* = -0.9586$ with ODE45 is used as the benchmark for calculating the relative errors. The numerical experiments are performed in PYTHON on a **HP EliteBook 840 G8** computer with a 2.40 Gigahertz (GHz) Intel Core i7 processor and 8 gigabytes (GB) random-access memory (RAM).

Table 4.1: Comparison of our different algorithms for $n = 5$

Method	p_0	Cost	Time(s)	Iteration step	Relative error
Alg 2	-0.95700	4.3888	183	3000	0,17%
Alg 3	-0.95873	2.3916	116	3000	0,0136 %

4.4.2 Some high-dimensional stochastic optimal control problems

From the results in Table 4.1, we can see that Algorithm 3 is the most effective among the three proposed algorithms with a relative error of 0,0136% and running time of 116 s.

And from both the relative errors and the running time, Algorithms 2 and 3 demonstrate much more effective performance on comparing with Algorithms 1. Therefore, when the optimal control $\tilde{u}$ is not explicit, Algorithms 2 or 3 will be a better choice for high dimensional cases. Otherwise Algorithms 1 could be chosen but it is not suitable for high dimensional cases.

Remark : *The results of Algorithm 1 are not presented in this table due to its particularly high execution time, which is estimated to be approximately 22 hours.*

4.4.2 Some high-dimensional stochastic optimal control problems

In this subsection, we show some numerical results for high-dimensional cases. As Algorithm 1 is not suitable for high-dimensional problems when the optimal control $\tilde{u}$ is not explicit, we mainly show the results of Algorithm 2 with $H_u(t; x; u; p; q) = 0$ and that of Algorithm 3 when we have an explicit expression of $\tilde{H}$.

A high-dimensional LQ stochastic optimal control problem

We first compute the LQ problem mentioned in last subsection for a $n = 100$ case. Let $x_0 = 1.0$; $T = 0.1$. Figure 4.2 shows the relative errors with different numbers of iteration steps for Algorithm 2 and 3, and the results of ODE45 are used as the benchmark for calculating the relative errors.

4.4.2 Some high-dimensional stochastic optimal control problems 57

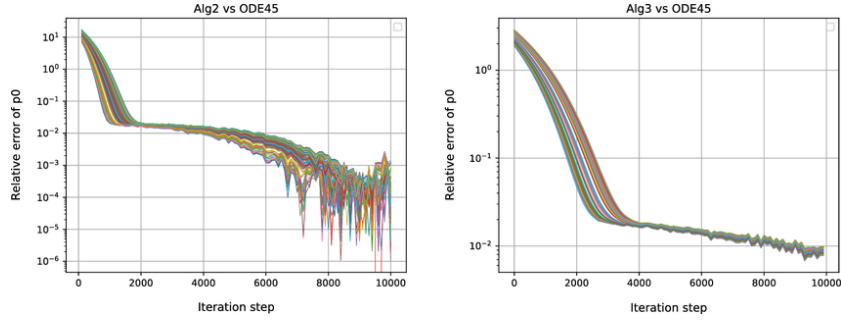


Figure 4.2: Relative error for $n = 100$

The left figure represents the relative errors for each dimension of p_0 with Algorithm 2 on comparing with the results of ODE45, and the right figure shows the relative errors for Algorithm 3.

We perform 10 independent runs and the means of the cost functional converges to 48.056. The curve of the cost functional is shown in Figure 4.3. [20].

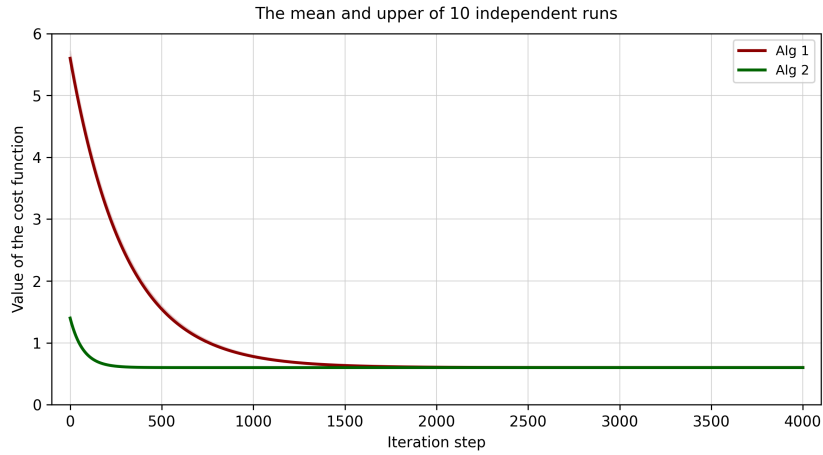


Figure 4.3: Mean and scope of the cost functional ($n = 100$)

For $n = 100$ and $\lambda = 0.05$, the figure shows the mean and scope of the cost functional among 10 independent runs. The green line and scope represent the results of Algorithm 2 with the constraint $H_u(t; x; u; p; q) = 0$. The red line and scope represent the value of cost functional for Algorithm

4.4.2 Some high-dimensional stochastic optimal control problems 58

3 when the optimal control $\tilde{u}$ has an explicit solution.

Now we consider a general case with the terminal $p_T^* = -\mathbb{I}_n x_T^*$, where $\mathbb{I}_n$ is an n -order matrix with all its elements equal to 1. In this case, the solution K_t of equation (4.9) at time 0 changes with the change of the dimension n .

Similarly, we first calculate the numerical solutions of p_0 with ODE45 and take the results as the benchmark. Then we compare our neural network solutions of p_0 with that of ODE45. Table 4.4.2 shows the comparison results for different dimensions and Figure 4.4.2 shows the curve of the cost functional for $n = 5$.

We notice that for $n = 20$, the results of Alg 3 are not very stable in different runs. In order to get more stable results, we use different neural network for different time points for $n = 20$ in Alg 3, but the cost is that more network parameters should be used which need more computation time.

		n=1	n=2	n=5	n=10	n=20
Solution with ODE45		-0.9586	-1.8275	-4.3638	-8.5306	-16.821
Solution with Neural Network	Alg 2	-0.9519	-1.8239	-4.3535	-8.4974	-16.730
	Alg 3	-0.9585	-1.8276	-4.3571	-8.4782	-16.663
Relative error	Alg 2	0.705%	0.197%	0.234%	0.390%	0.541%
	Alg 3	0.007%	0.003%	0.153%	0.615%	0.939%
Cost functional	Alg 2	0.4812	1.8375	10.863	42.850	169.80
	Alg 3	0.4892	1.8262	10.916	43.284	167.64

Table 4.2: Comparison between ODE45 and the algorithms for different dimensions

4.4.2 Some high-dimensional stochastic optimal control problems 59

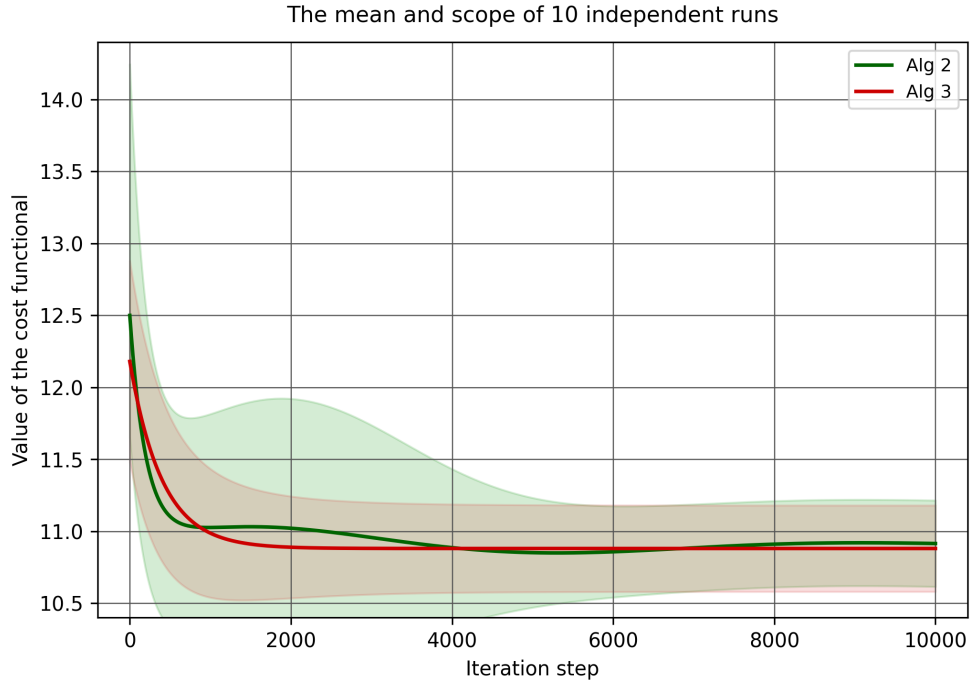


Figure 4.4: Mean and scope of the cost functional for 10 independent runs

The figure shows the mean and scope of the cost functional for 10 independent runs. After 1000 iterations, the value of the cost functional for Algorithm 2 is close to that of Algorithm 3, and the mean of the cost functional $J(0; x_0; \tilde{u}(\cdot))$ for Algorithm 2 is 10.863.

A 100–dim nonlinear control problem

Consider the following control system

$$\begin{cases} dx_t &= 2u_t dt + \sqrt{2}dW_t, \\ x(0) &= x_0, \end{cases} \quad (4.30)$$

and the corresponding cost functional

$$J(0; x_0; u(\cdot)) = E \left(\int_0^T \langle u_t, u_t \rangle dt + h(x_T) \right); \quad (4.31)$$

where $h(x) = \ln \left(\frac{1}{2}(1 + |x|^2) \right)$, $|x|^2 = \langle x, x \rangle$. W_t is a n -dimensional Brownian motion, and the process x_t is valued in $\mathbb{R}^n$.

4.4.2 Some high-dimensional stochastic optimal control problem 60

The control domain is $\mathbb{R}$ and the Hamiltonian H is

$$H(t; x; u; p; q) = \langle p; 2u \rangle + \text{tr}(\sqrt{2}q) - \langle u; u \rangle;$$

the optimal control u^* can be solved with $u^* = p$,

Then we have

$$\bar{H}(t; x; p; q) = \langle p; 2p \rangle + \text{tr}(\sqrt{2}q) - \langle p; p \rangle.$$

The corresponding Hamiltonian system is given as

$$\begin{cases} dx_t^* = 2u_t^* dt + \sqrt{2}dW_t, \\ -dp_t^* = -q_t^* dW_t, \\ x_0^* = x_0, \quad p_T^* = -h_x(x_T^*), \\ u_t^* = p_t^*, \end{cases} \quad (4.32)$$

where p^* and q^* are valued in $\mathbb{R}^n$ and $\mathbb{R}^{n \times n}$ respectively and $n = d = k$.

It is known that the cost functional of (4.31) can be obtained via the Hopf-Cole transformation, as shown in the following:

$$J(0; x_0; u^*(\cdot)) = -\ln \left(E \left[\exp(-h(x_0 + \sqrt{2}W_T)) \right] \right); \quad x_0 \in \mathbb{R}^n.$$

which can be approximated by Monte Carlo simulation. By setting $x_0 = 0$ and $T = 1.0$ with the dimension $n = 100$ and the approximate cost functional is 4.591 by performing the Monte Carlo simulation. Figure 4.4.2 shows the values of the cost functional and the relative errors of the cost functional on comparing with the results of the Monte Carlo simulation.

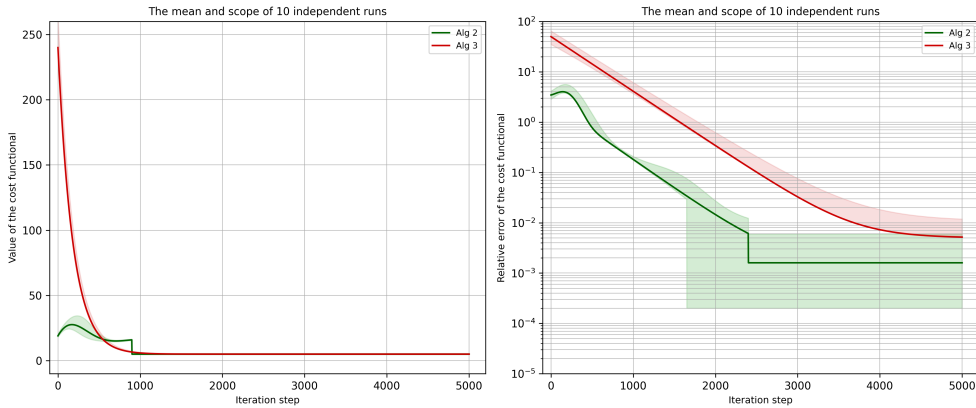


Figure 4.5: Mean and scope of relative errors of the cost functional

4.4.2 Some high-dimensional stochastic optimal control problem 61

For $n = 100$ and $\lambda = 0.05$. The left figure shows the cost functional curve and the right figure shows the mean and scope of relative errors of the cost functional $J(0; x_0; \tilde{u}(\cdot))$ among 10 independent runs. We can see that after 5000 iteration steps, the values of the cost functional are very close between Algorithm 2 and 3, and Algorithm 2 shows smaller mean of relative errors.

An example of the control $\tilde{u}$ without explicit solution

In this subsection, we show an example whose optimal control $\tilde{u}$ does not have an explicit form. Consider the following stochastic control problem,

$$\begin{cases} dx_t &= \sin(u_t)dt + x_t dW_t, \\ x(0) &= x_0, \end{cases} \quad (4.33)$$

with cost functional

$$J(u(\cdot)) = E \left(\int_0^T \langle u_t; u_t \rangle dt + \frac{1}{2} \langle x_T; x_T \rangle \right);$$

where the control domain is $U = \mathbb{R}^n$, and W_t is an n -dimensional Brownian motion. The Hamiltonian H is

$$H(t; x; u; p; q) = \langle p; \sin(u) \rangle + \langle q; x \rangle - \langle u; u \rangle;$$

which is a multi-dimensional transcendental equation and has not an explicit representation of both the optimal control $\tilde{u}$ and the function $\bar{H}$.

The derivative of the Hamiltonian H in u is given as

$$H_u(t; x; u; p; q) = p \cos(u) - 2u; \quad (4.34)$$

which is also a multi-dimensional transcendental equation. The corresponding Hamiltonian system with the constraint $H_u(t; x; u; p; q) = 0$ is

$$\begin{cases} dx_t^* &= \sin(u_t^*)dt + x_t^* dW_t, \\ -dp_t^* &= q_t^* dt - q_t^* dW_t, \\ x_0^* &= x_0, \quad p_T^* = -h_x(x_T^*), \\ p_t^* \cos(u_t^*) - 2u_t^* &= 0. \end{cases} \quad (4.35)$$

We set $x_0 = 1.0$, $T = 0.1$, $\lambda = 0.1$. As the optimal control $\tilde{u}$ has not an explicit solution and $\bar{H}$ does not have an explicit expression, thus Algorithm

4.4.2 Some high-dimensional stochastic optimal control problem 62

3 is not suitable for this example. Therefore we mainly give the numerical results of Algorithm 2 in this example. In order to show the convergence of our algorithm, we implement the algorithm of Han and E [17] which solves the stochastic optimal control problems directly via deep learning, then compare our algorithm with that of Han and E. The comparing results are shown in Figure 4.4.2. We can see that the value of the cost functional for the two algorithms are very close, but the advantage of our algorithm is that it provides a criterion to decide whether the state-control pair $(\tilde{x}(\cdot); \tilde{u}(\cdot))$ is an optimal pair, that is, whether the value of the loss function equals to 0.

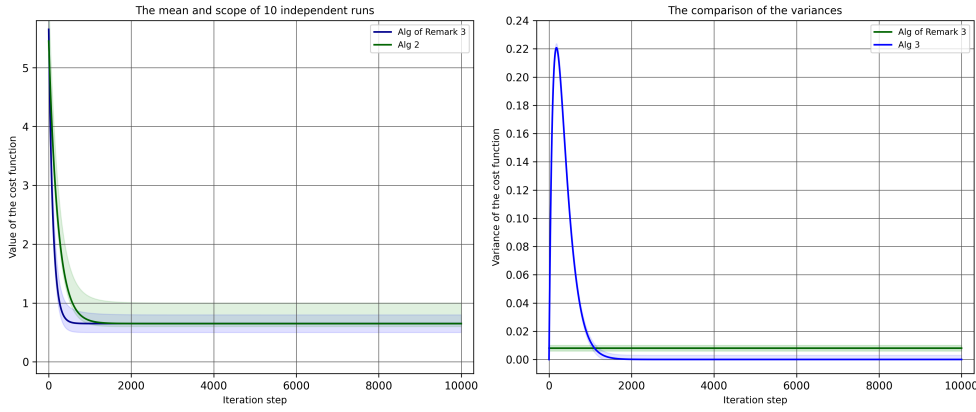


Figure 4.6: mean and scope of the cost functional among 10 independent runs (variances comparison)

Case $n = 100$ and $\lambda = 0.1$. The left figure shows the comparison between our algorithm and that of Han and E on the mean and scope of the cost functional among 10 independent runs. The right figure shows the variances comparison. We can see that the value of the cost functional for the two algorithms are very close from the left figure and our algorithm has lower variances from the right figure.

Note:

The figures presented in this thesis have been adapted from Peng et al [20].

4.5 Conclusion

We have seen how to solve the stochastic optimal control problem from the view of the stochastic maximum principle and proposed three different algorithms via deep learning. By comparing three proposed algorithms through numerical results and pointed out their applicative situations. The numerical results for different examples demonstrate the effectiveness of these proposed algorithms.

Bibliography

- [1] V. Barbu, G. Da Prato, and C. Popa, “Existence and uniqueness of the dynamic programming equation in Hilbert space,” *Nonlinear Analysis*, 7 (1983), 283–299.
- [2] M. Bardi and I. Capuzzo-Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkhäuser, Boston, 1997.
- [3] V. Barbu and G. Da Prato, *Hamilton-Jacobi Equations in Hilbert Spaces*, Pitman, London, 1983.
- [4] V. Barbu and G. Da Prato, “Hamilton-Jacobi equation in Hilbert spaces: variational and semigroup approach,” *Ann. Math. Pura. Appl.*, 142 (1985), 303–349.
- [5] G. Barles, “Deterministic impulse control problems,” *SIAM J. Control & Optim.*, 23 (1985), 419–432.
- [6] E. N. Barron and R. Jensen, “The Pontryagin maximum principle from dynamic programming and viscosity solutions to first-order partial differential equations,” *Trans. Amer. Math. Soc.*, 298 (1986), 635–641.
- [7] A. Bensoussan, “Some solvable stochastic control problems,” *Stochastics*, 4 (1980), 39–83.
- [8] L. D. Berkovitz, *The Hamilton-Jacobi Equation: A Global Approach*, Academic Press, New York, 1977.
- [9] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, New York, Inc., 2006.

-
- [10] A. Cadenillas and I. Karatzas, “The stochastic maximum principle for a singular control problem,” *Stoch. & Stoch. Rep.*, 49 (1994), 211–237.
 - [11] P. Cannarsa, “The stochastic maximum principle for linear convex optimal control with random coefficients,” *SIAM J. Control & Optim.*, 33 (1995), 590–624.
 - [12] P. Cannarsa and G. Da Prato, “Regularity properties of solutions to Hamilton-Jacobi equations in infinite dimensions and nonlinear optimal control,” *Dif. Int. Eqs.*, 2 (1989), 479–493.
 - [13] T. Chi, J. J. Liu, and H. Chen, “Optimal stochastic controls,” *Sankhya, Ser. A*, 30 (1968), 221–252.
 - [14] F. H. Clarke, “The maximum principle under minimal hypotheses,” *SIAM J. Control & Optim.*, 14 (1976), 1078–1091.
 - [15] R. J. Elliott, “The optimal control of a stochastic system,” *SIAM J. Control & Optim.*, 15 (1977), 756–778.
 - [16] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, Springer, 2006.
 - [17] J. Han and W. E., “Deep learning approximation for stochastic control problems,” *Deep Reinforcement Learning Workshop*, 2016.
 - [18] J. Han, A. Jentzen, and W. E., “Solving high-dimensional partial differential equations using deep learning,” *Proceedings of the National Academy of Sciences*, 2018.
 - [19] M. Hu, S. Ji, and X. Xue, “A global stochastic maximum principle for fully coupled forward-backward stochastic systems,” *SIAM Journal on Control and Optimization*, 56(6) (2018), 4309–4335.
 - [20] S. Ji, S. Peng, Y. Peng, and X. Zhang, “Solving stochastic optimal control problem via stochastic maximum principle with deep learning method,” *Journal of Scientific Computing*, 93(1) (2022), 30.

- [21] S. Ji, S. Peng, Y. Peng, and X. Zhang, “Three algorithms for solving high-dimensional fully coupled FBSDEs through deep learning,” *IEEE Intelligent Systems*, 35(3) (2020), 71–84.
- [22] S. Ji, S. Peng, Y. Peng, and X. Zhang, “Three algorithms for solving high-dimensional fully coupled FBSDEs through deep learning,” *IEEE Intelligent Systems*, 35(3) (2020), 71–84.
- [23] D. C. Liu and J. Nocedal, “On the limited memory BFGS method for large scale optimization,” *Mathematical Programming*, 45(1–3) (1989), 503–528.
- [24] B. Oksendal, *Stochastic Differential Equations: An Introduction with Applications*, Springer, 2003.
- [25] S. Peng, “Problem of eigenvalues of stochastic Hamiltonian systems with boundary conditions,” *Stochastic Processes & Their Applications*, 88(2) (2000), 259–290.
- [26] S. Peng, “Problem of eigenvalues of stochastic Hamiltonian systems with boundary conditions,” *Stochastic Processes & Their Applications*, 88(2) (2000), 259–290.
- [27] S. Peng, “Backward stochastic differential equations and applications to optimal control,” *Applied Mathematics and Optimization*, 27(2) (1993), 125–144.
- [28] S. Peng, *Nonlinear Expectations and Stochastic Calculus under Uncertainty – with Robust CLT and G-Brownian Motion*, Springer, 2019.
- [29] 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) (1999), 825–843.
- [30] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, 61 (2015), 85–117.
- [31] J. Yong and X. Y. Zhou, *Stochastic Controls: Hamiltonian Systems and HJB Equations*, Springer, New York, 1999.

-
- [32] J. Yong and X. Y. Zhou, *Stochastic Controls: Hamiltonian Systems and HJB Equations*, Springer, 1999.