3-25-2018

General Stochastic Integral and Itô Formula with Application to Stochastic Differential Equations and Mathematical Finance

Jiayu Zhai
Louisiana State University and Agricultural and Mechanical College, zhaijiayumath@hotmail.com

GENERAL STOCHASTIC INTEGRAL AND ITÔ FORMULA WITH APPLICATIONS TO STOCHASTIC DIFFERENTIAL EQUATIONS AND MATHEMATICAL FINANCE

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Mathematics

by
Jiayu Zhai
B.S., Ludong University, 2009
M.S., Shanghai University, 2012
M.S., Louisiana State University, 2014
May 2018

Acknowledgements

This dissertation would be impossible without the contribution and support from several people. I would like to express my deepest appreciation to my advisers, Professor Hui-Hsiung Kuo and Professor Xiaoliang Wan, for their guidance and encouragement throughout my PhD study in Louisiana State University. They set models for me to follow.

Meanwhile, I wish to thank Professors Hongchao Zhang and Professor Padmanabhan Sundar for their advises in my research, particularly in optimization theory and stochastic differential equations, respectively. I also want to thank Professor Daniel Cohen and the Dean's representative Professor Parampreet Singh for their help and instruction during my graduate study and dissertation writing. I appreciate them being the committee members of my general exam and thesis defence.

I want to thank the Department of Mathematics at Louisiana State University and its faculty, staff and graduate students for providing me a pleasant environment to study, discuss and work.

I also want to thank the programs that financially supported my research directly. The work on Part II of my dissertation was supported by AFOSR grant FA9550-15-1-0051 and NSF Grant DMS-1620026. The work on Chapter 8 was supported in part by an appointment with the NSF Mathematical Sciences Summer Internship Program sponsored by the National Science Foundation, Division of Mathematical Sciences (DMS). This program is administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and NSF. ORISE is managed by ORAU under DOE contract number DE-SC0014664.

I am extremely grateful to my mother for standing behind me. Her unconditional love and help provide me an inexhaustible resource of energy to continue my research. Some of the credit for my success must go to my friends, who are always there whenever I need them.

This dissertation is particularly dedicated to my father Zhigang Zhai, who passed away in May 2017. I would never have achieved my success without his support and love, especially when I make critical decisions. Thank you and love you, Dad! May you rest in peace!

Table of Contents

Acknowledgements ii

Abstract vi

Part I: General Stochastic Integral and Its Applications 1

Chapter 1: Introduction 2
- 1.1 Background 2
- 1.2 Stochastic Process and Brownian Motion 4
- 1.3 Conditional Expectation and Martingale 5
- 1.4 Itô Integral and the Itô Formula 8
- 1.5 Girsanov Theorem 12

Chapter 2: General Stochastic Integral and General Itô Formula 14
- 2.1 Stochastic Integral for Instantly Independent Stochastic Processes 14
- 2.2 General Stochastic Integral 25
- 2.3 Itô Isometry for General Stochastic Integral 30
- 2.4 A General Itô Formula 30

Chapter 3: Near-Martingale Property of Anticipating Stochastic Integration 34
- 3.1 Near-Martingale Property 34
- 3.2 Doob–Meyer's Decomposition for Near-Submartingales 37
- 3.3 General Girsanov Theorem 39

Chapter 4: Application to Stochastic Differential Equations 41
- 4.1 Stochastic Differential Equations for Exponential Processes 41
- 4.2 Stochastic Differential Equations with Anticipative Coefficients 43
- 4.3 Stochastic Differential Equations with Anticipative Initial Condition 46
- 4.4 Conditional Expectation of the Solution of Theorem 4.3.2 48

Chapter 5: A Black–Scholes Model with Anticipative Initial Conditions 56
- 5.1 The Classical Black–Scholes Model 56
- 5.2 A Black–Scholes Model with Anticipative Initial Conditions 58
- 5.3 Arbitrage-Free Property 59
- 5.4 Completeness of the Market 62
- 5.5 Option Pricing Formula 64
- 5.6 Hedging Portfolio 69

References 70

Part II: Temporal Minimum Action Method for Rare Events in Stochastic Dynamic Systems 73

Chapter 6: Introduction 74
6.1 Large Deviation Principle 74
6.2 Statistic Physics 76
6.3 Freidlin-Wentzell Theory 77

Chapter 7: Temporal Minimum Action Method 80
7.1 Problem Description 80
7.2 Introduction to the Temporal Minimum Action Method 82
7.3 Reformulation and Finite Element Discretization of the Problem 83
7.4 Problem I with a Fixed T 86
7.5 Problem II with a Finite T^* 91
7.6 Problem II with an Infinite T^* 96
7.7 A Priori Error Estimate for a Linear ODE System 101
7.8 Numerical Experiments 104

Chapter 8: Temporal Minimum Action Method for Time Delayed System 109
8.1 Introduction of Large Deviation Principle for Stochastic Differential Delay Equations 109
8.2 Penalty Method 110
8.3 Variational Calculus with Delay for Finite Element Approximation 112
8.4 Computation of the Gradient of $\hat{S}$ 114
8.5 Numerical Examples 116

References 119

Vita 122

Abstract

A general stochastic integration theory for adapted and instantly independent stochastic processes arises when we consider anticipative stochastic differential equations. In Part I of this thesis, we conduct a deeper research on the general stochastic integral introduced by W. Ayed and H.-H. Kuo in 2008. We provide a rigorous mathematical framework for the integral in Chapter 2, and prove that the integral is well-defined. Then a general Itô formula is given. In Chapter 3, we present an intrinsic property, near-martingale property, of the general stochastic integral, and Doob–Meyer's decomposition for near-submartigales. We apply the new stochastic integration theory to several kinds of anticipative stochastic differential equations and discuss their properties in Chapter 4. In Chapter 5, we apply our results to a general Black–Scholes model with anticipative initial conditions to obtain the properties of the market with inside information and the pricing strategy, which can help us better understand inside trading in mathematical finance.

In dynamical systems, small random noises can make transitions between equilibriums happen. This kind of transitions happens rarely but can have crucial impact on the system, which leads to the numerical study of it. In Part II of this thesis, we emphasize on the temporal minimum action method introduced by X. Wan, using optimal linear time scaling and finite element approximation to find the most possible transition paths, also known as the minimal action paths, numerically. We show the method and its numerical analysis in Chapter 7. In Chapter 8, we apply this method to a stochastic dynamic system with time delay using penalty method.

Part I

General Stochastic Integral and Its Applications

Chapter 1
Introduction

1.1 Background

The historical background of this work can be traced back to early 19th century. In 1827, Robert Brown published in [6] his observation that pollen grains suspended in water move continuously in a random way. However, Brown did not explain the source of the motion, which was later on named after him, a *Brownian motion*, in physics or set up the model in mathematics.

The application of Brownian motion in mathematical modelling goes earlier than its theoretical framework. It was used separately in two subjects: (1) In mathematical finance, Louis Bachelier used Brownian motion to model stock market in his doctoral thesis in 1900; and (2) In physics, Albert Einstein [9] used Brown motion to model atoms in 1905.

By using normal distribution, Einstein's description of Brownian motion is close to its rigorous definition in modern mathematics. In 1923, Norbert Wiener [33] provided a definition of Brownian motion in mathematics using real analysis, and so it is also called a *Wiener process* now. Based on this definition, the whole theory of stochastic process and its analysis are built up.

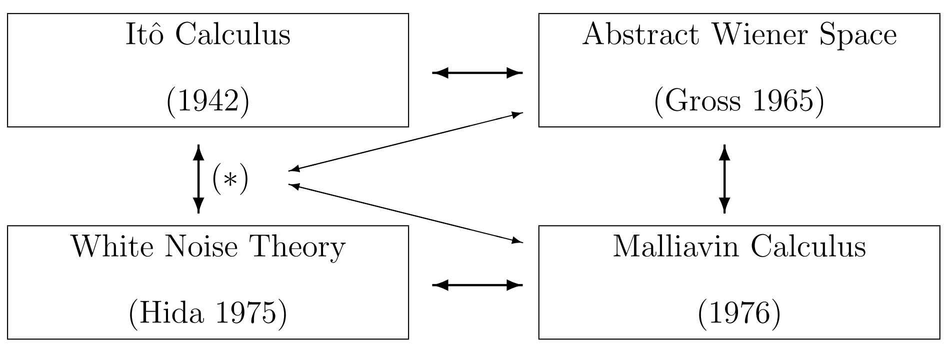
The Itô integral introduced by K. Itô in 1942 [16], as a generalization of Wiener integral, becomes a strong and standard tool in the study of statistic physics, control engineering, biological system, mathematical finance, and so on. However, the classic Itô integration theory relies on the requirement of adapted integrand, yet even in solving adapted problems, for example hedging portfolio in Black-Scholes model, there can be non-adapted SDEs involved. For instance, Itô integral is not enough to solve an even simple SDE

$$dX_t = X_t \, dB(t), \quad Y_0 = B(1), \quad t \geq 0, \tag{1.1.1}$$

since the initial condition is anticipating. Wiener's work inspired L. Gross to introduce abstract Wiener space in 1965 [12], based on which the white noise distribution theory was built up by T. Hida in 1975 [13]. In 1976, P. Malliavin provided a proof of the existence of transition probability (see [30]). In the white noise distribution theory, the Hitsuda–Skorokhod integral provides a certain generalization of Itô integral to define the stochastic integral $\int_a^t B(1)\, dB(s)$, which is used in solving (1.1.1). However, this generalization has two problems:

1. $\int_a^t f(t)\, dB(s) := \int_a^t \partial_t^* f(t)\, dt$ is defined as a generalized function in the space $(\mathcal{S})^*$ and lacks probabilistic meaning.

2. The white noise integral $\int_a^t \partial_t^* f(t)\, dt$ is a Hitsuda–Skorokhod integral if it is a random variable in $L^p(\mathcal{S}'(\mathbb{R}), \mu)$ for some $p > 1$. However, there is no theorem that characterizes a generalized function in $(\mathcal{S})^*$ to be in $L^p(\mathcal{S}'(\mathbb{R}), \mu)$.



E. Pardoux and Protter, P. in [32] defined the forward and backward stochastic integrals separately in order to solve the forward-backward SDEs. This provides an explicit expression of the solution of some kinds of backward SDEs. However, it cannot define a simple stochastic integral like $\int_0^t B(1) - B(s)\, dB(s), t \in [0,1]$. The disadvantage of the existing theories and the need in solving general SDEs motivated us to develop a new stochastic integral. As shown in the above diagram (see [21]) of relationships among the four subjects of stochastic analysis, our work is at $(*)$.

In the following sections, the definitions of the preliminary mathematical concepts and their properties are given.

1.2 Stochastic Process and Brownian Motion

Definition 1.2.1 (Stochastic Process)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space, where Ω is a sample space, $\mathcal{F}$ is a σ-field and P is a probability measure. A function $X(t, \omega) : [0, +\infty) \times \Omega \to \mathbb{R}$ is called a *stochastic process*, if it satisfies the following two conditions.

1. For each $t \in [0, +\infty)$, $f(t, \omega) : \Omega \to \mathbb{R}$ is a random variable.

2. For each $\omega \in \Omega$, $f(t, \omega) : [0, +\infty) \to \mathbb{R}$ is a measurable function of t.

Definition 1.2.2 (Brownian Motion)**.** A stochastic process $B(t, \omega) : [0, \infty) \times \Omega \to \mathbb{R}$ is called a *Brownian motion* or a *Wiener process*, if it satisfies the following four conditions.

i. For any $0 \le s < t$, $B(t, \omega) - B(s, \omega)$ is a random variable normally distributed with expectation 0 and variance $t - s$;

ii. $B(t)$ starts at 0 almost surely, i.e.,

$$P\{\omega : B(0, \omega) = 0\} = 1;$$

iii. For any $0 \le t_1 < \ldots < t_n$, the random variables $B(t_1, \omega)$, $B(t_2, \omega) - B(t_1, \omega), \ldots,$ $B(t_n, \omega) - B(t_{n-1}, \omega)$ are independent;

iv. Almost all sample paths of $B(t)$ are continuous, i.e.,

$$P\{\omega : B(t, \omega) \text{ is continuous in } [0, \infty)\} = 1.$$

1.3 Conditional Expectation and Martingale

Definition 1.3.1 (Conditional Expectation)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G}$ a sub-σ-field of $\mathcal{F}$, and X a random variable in $L^1(\Omega)$. Then the *conditional expectation* of X on $\mathcal{G}$, denoted by $E[X|\mathcal{G}]$, is a random variable Y satisfying the following conditions:

1. Y is measurable with respect to $\mathcal{G}$;

2. For any Borel set $A \in \mathcal{G}$, we have

$$\int_A Y\,dP = \int_A X\,dP.$$

Remark 1.3.2. The conditional expectation $E[X|\mathcal{G}]$ is the best approximation of X with information of $\mathcal{G}$.

Theorem 1.3.3 (Properties of Conditional Expectation)**.** *Let X and Y be random variables on Ω. Assume that $\mathcal{G}$ is a sub-σ-field of $\mathcal{F}$. Then we have*

1. $E\big[E[X|\mathcal{G}]\big] = E[X]$.

2. $E[X+Y|\mathcal{G}] = E[X|\mathcal{G}] + E[Y|\mathcal{G}]$.

3. *If X is measurable with respect to $\mathcal{G}$, then $E[X|\mathcal{G}] = X$.*

4. *If X is independent of $\mathcal{G}$, then $E[X|\mathcal{G}] = E[X]$.*

5. *If Y is measurable with respect to $\mathcal{G}$ and $XY \in L^1(\Omega)$, then*

$$E[XY|\mathcal{G}] = Y E[X|\mathcal{G}].$$

6. *If $\mathcal{H}$ is a sub-σ-field of $\mathcal{G}$, then $E\big[E[X|\mathcal{G}]\big|\mathcal{H}\big] = E[X|\mathcal{H}]$.*

Definition 1.3.4 (Filtration)**.** A family of σ-fields $\{\mathcal{F}_t; 0 \leq t < \infty\}$ on $(\Omega, \mathcal{F}, P)$ is called a *filtration*, if $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ for any $0 \leq s < t$.

Definition 1.3.5 (Adaptedness)**.** Let $X(t, \omega)$ be a stochastic process and $\{\mathcal{F}_t; 0 \leq t < \infty\}$ be a filtration. We say that X is *adapted* to $\{\mathcal{F}_t\}$, if for each $t \geq 0$, $X(t, \omega)$ is measurable with respect to $\mathcal{F}_t$.

Remark 1.3.6. From now on, we will also use B_t and X_t to denote a Brownian motion and a stochastic process, respectively. We say that $\{\mathcal{F}_t\}$ is the natural filtration, if it is generated from the corresponding Brownian motion, i.e., for each $t \geq 0$, $\mathcal{F}_t = \sigma\{B(s); 0 \leq s \leq t\}$.

Example 1.3.7. A Brownian motion $B(t)$ is adapted to its natural filtration $\{\mathcal{F}_t\}$.

Definition 1.3.8 (Martingale)**.** Let $\{\mathcal{F}_t\}$ be a filtration and $X(t)$ be a stochastic process adapted to $\{\mathcal{F}_t\}$ with $E|X_t| < \infty$ for all $t \geq 0$. We call $X(t)$ a *martingale* if for each $0 \leq s < t$,

$$E[X(t)|\mathcal{F}_s] = X(s), \quad \text{almost surely.}$$

Remark 1.3.9. Similar to Remark 1.3.2, a martingale is a stochastic process X_t whose best approximation with the information prior to time s is the process X_s itself at s.

Example 1.3.10. A Brownian motion $B(t)$ is a martingale with respect to its natural filtration $\{\mathcal{F}_t\}$. In fact, by Definition 1.2.2 and Theorem 1.3.3,

$$\begin{aligned} E[B(t)|\mathcal{F}_s] =& E[B(t) - B(s) + B(s)|\mathcal{F}_s] \\ =& E[B(t) - B(s)|\mathcal{F}_s] + E[B(s)|\mathcal{F}_s] \\ =& E[B(t) - B(s)] + B(s) \\ =& B(s). \end{aligned} \tag{1.3.1}$$

Many important stochastic processes are not martingales, but they are worth to study, e.g., $B(t)^2$ in Example 1.3.13 below. Although they are not martingales, a similar property is satisfied.

Definition 1.3.11 (Submartingale)**.** A stochastic process $X_t, a \leq t \leq b$, with $E|X_t| < \infty$ for all t, is called a *submartingale* with respect to a filtration $\{\mathcal{F}_t\}$, if for all $a \leq s < t \leq b$, we have

$$E[X_t|\mathcal{F}_s] \geq X_s, \quad \text{almost surely.} \tag{1.3.2}$$

Definition 1.3.12 (Supermartingale)**.** A stochastic process $X_t, a \leq t \leq b$, with $E|X_t| < \infty$ for all t, is called a *supermartingale* with respect to a filtration $\{\mathcal{F}_t\}$, if for all $a \leq s < t \leq b$, we have

$$E[X_t|\mathcal{F}_s] \leq X_s, \quad \text{almost surely.} \tag{1.3.3}$$

Example 1.3.13. Let $X_t = B(t)^2$. Then for $t > s$,

$$\begin{aligned} E[X_t|\mathcal{F}_s] =& E[B(t)^2|\mathcal{F}_s] \\ =& E\big[\big(B(t) - B(s) + B(s)\big)^2|\mathcal{F}_s\big] \\ =& E\big[\big(B(t) - B(s)\big)^2|\mathcal{F}_s\big] + 2B(s)E[B(t) - B(s)|\mathcal{F}_s] + E[B(s)^2|\mathcal{F}_s] \\ =& E\big[\big(B(t) - B(s)\big)^2\big] + B(s)^2 \\ =& B(s)^2 + t - s \geq B(s)^2 = X_s. \end{aligned}$$

So X_t is a submartingale.

A crucial and useful property of submartingale processes is Doob–Meyer's decomposition.

Theorem 1.3.14 (Doob–Meyer's Decomposition, [20])**.** *Let $X_t, a \leq t \leq b$, be a continuous submartingale with respect to a continuous filtration $\{\mathcal{F}_t; a \leq t \leq b\}$. Then X_t has a unique decomposition*

$$X_t = M_t + A_t, \quad a \leq t \leq b, \tag{1.3.4}$$

where M_t is a continuous martingale with respect to $\{\mathcal{F}_t\}$, and A_t is a continuous stochastic process satisfying the conditions:

1. $A_a = 0$;

2. A_t *is increasing in* t *almost surely;*

3. A_t *is adapted to* $\{\mathcal{F}_t\}$.

Definition 1.3.15 (Compensator)**.** The continuous stochastic process A_t in Theorem 1.3.14 is called the *compensator* of the submartingale X_t.

1.4 Itô Integral and the Itô Formula

From now on we will consider, without loss of generality, all the stochastic processes involved in a time interval $[a, b]$ for arbitrary $0 \leq a < b$, or $[0, T]$ for an arbitrary $T > 0$.

We use $L^2_{\mathrm{ad}}([a, b] \times \Omega)$ to denote the space of square integrable stochastic processes $f(t)$ adapted to $\{\mathcal{F}_t\}$, that is, $\int_a^b E\big(|f(t)|^2\big)\, dt < \infty$.

Definition 1.4.1 (Itô Integral)**.** Let $B(t)$ be a Brownian motion, and $f(t)$ be a stochastic process in $L^2_{\mathrm{ad}}([a, b] \times \Omega)$. The *Itô integral* of f is a random variable denoted in stochastic integral form

$$\int_a^b f(t)\, dB(t),$$

and defined in the following steps.

Step 1: *Step Processes.*

If $f(t)$ is an adapted step stochastic process

$$f(t) = \sum_{i=1}^{n} \xi_{i-1} 1_{[t_{i-1}, t_i)},$$

where $a = t_0 < t_1 < \cdots < t_n = b$ is a partition of $[a, b]$, then we define its Itô integral as

$$\int_a^b f(t)\, dB(t) = \sum_{i=1}^{n} \xi_{i-1}(B(t_i) - B(t_{i-1})).$$

Step 2: *Square Integrable Processes Approximated by Step Processes.*

For any process $f(t) \in L^2_{\text{ad}}([0,T] \times \Omega)$, we can prove that there is a sequence of adapted step stochastic processes $\{f_n(t)\}_{n=1}^{\infty}$ that converges to $f(t)$ in $L^2([0,T]\times\Omega)$.

Step 3: *Square Integrable Processes.*

Define the Itô integral of f by

$$\int_a^b f(t)\,dB(t) = \lim_{n\to\infty} \int_a^b f_n(t)\,dB(t), \quad \text{in } L^2(\Omega).$$

One can prove that the stochastic integral above is well-defined. For details, see [20]. Although Itô integral is defined in the rigorous framework of Hilbert space, its intuition is just a similarity of Riemann integral. Its consistency with Riemann integral is proved in the following theorem.

Theorem 1.4.2. *Assume that $f(t)$ is left continuous and adapted to $\{\mathcal{F}_t\}$. Then the limit of its Riemman type sum exists and is equal to its Itô integral, i.e.,*

$$\int_a^b f(t)\,dB(t) = \lim_{\|\Delta_n\|\to 0} \sum_{i=1}^{n} f(t_{i-1})(B(t_i) - B(t_{i-1})). \tag{1.4.1}$$

Remark 1.4.3. Notice that in the right hand side of Equation (1.4.1), as the limit of Riemann type sum, the evaluation point of f on the subinterval $[t_{i-1}, t_i)$ is the left endpoint t_{i-1}. We will see that this makes Itô process a martingale, and this is the key characteristic of Itô integration.

Now, we show the most important properties of Itô integral. They are used in, for example, the proof of existence and uniqueness of stochastic differential equations, Freidlin-Wendzell property, the Girsanov theorem and so on. Among others, the Itô formula palys the central role in modern stochastic analysis, and its applications to physics, biology, engineering and finance. For their proofs, refer to [20].

Theorem 1.4.4 (Itô Isometry)**.** *Let $f \in L^2_{ad}([0,T] \times \Omega)$. Then its Itô integral $\int_0^T f(t)\,dB(t)$ is a random variable satisfying*

$$E\Big(\int_0^T f(t)\,dB(t)\Big) = 0,$$

and

$$E\Big(\Big|\int_0^T f(t)\,dB(t)\Big|^2\Big) = \int_0^T (E[f(t)^2])\,dt.$$

In addition, if $f(t)$ is deterministic, then its Itô integral is normally distributed.

Theorem 1.4.5 (Martingale Property)**.** *Let $X_t, 0 \leq t \leq T$, be the stochastic process*

$$X_t = \int_0^t f(s)\,dB(s),$$

where $f \in L^2_{ad}([0,T] \times \Omega)$. Then X_t is a continuous martingale.

Definition 1.4.6 (Quadratic Variation)**.** Let $f(t)$ be a stochastic process and $\Delta = \{a = t_0 < t_1 < \ldots < t_{n-1} < t_n = b\}$ be a partition of $[a,b]$. Then the *quadratic variation* of f on $[a,b]$ is defined by

$$\langle f \rangle_{[a,b]} = \lim_{\|\Delta\| \to 0} \sum_{i=1}^{n} (f(t_i) - f(t_{i-1}))^2,$$

provided the limit exists in probability.

Example 1.4.7. The quadratic variation $\langle B \rangle_{[a,b]}$ of a Brownian motion $B(t)$ is $b - a$ for any $b > a > 0$.

Definition 1.4.8 (Itô Process)**.** Let $f(t) \in \mathcal{L}_{\text{ad}}(\Omega, L^2[0,T])$, $g(t) \in \mathcal{L}_{\text{ad}}(\Omega, L^1[0,T])$. Then we call the stochastic process $X(t)$ defined by

$$X(t) = \int_0^t f(t)\,dB(t) + \int_0^t g(t)\,dt \tag{1.4.2}$$

an *Itô process.*

Remark 1.4.9. In Equation (1.4.2), the first term is defined as an Itô integral, and the second term is defined as a Riemann integral for almost every $\omega \in \Omega$. By Theorem 1.4.5, X_t is continuous. Equation (1.4.2) is often written in stochastic differential

$$dX(t) = f(t)\, dB(t) + g(t)\, dt.$$

Theorem 1.4.10 (Itô's Formula, [17]). *Let $\{X_t^{(i)}\}_{i=1}^n$ be Itô processes as (1.4.2) for $f_i, g_i, i = 1, 2, \ldots, n$, and $\varphi(x_1, x_2, \ldots, x_n)$ be a twice continuously differentiable real function on $\mathbb{R}^n$. Then*

$$\begin{aligned}\varphi\big(X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(n)}\big) =&\varphi\big(X_0^{(1)}, X_0^{(2)}, \ldots, X_0^{(n)}\big) \\ &+ \sum_{i=1}^n \int_0^t \frac{\partial}{\partial x_i}\varphi\big(X_s^{(1)}, X_s^{(2)}, \ldots, X_s^{(n)}\big)\, dX_s^{(i)} \\ &+ \frac{1}{2}\sum_{i,j=1}^n \int_0^t \frac{\partial^2}{\partial x_i \partial x_j}\varphi\big(X_s^{(1)}, X_s^{(2)}, \ldots, X_s^{(n)}\big)\, \big(dX_s^{(i)}\big)\big(dX_s^{(j)}\big),\end{aligned} \tag{1.4.3}$$

or in stochastic differential

$$\begin{aligned}d\varphi\big(X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(n)}\big) =& \sum_{i=1}^n \frac{\partial}{\partial x_i}\varphi\big(X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(n)}\big)\, dX_t^{(i)} \\ &+ \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j}\varphi\big(X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(n)}\big)\, \big(dX_t^{(i)}\big)\big(dX_t^{(j)}\big),\end{aligned} \tag{1.4.4}$$

where $dX_t^{(i)}$ are the stochastic differentials for the Itô processes $X_t^{(i)}$ and their multiplications follow the rule shown in the following Itô table:

$\times$	$dB(t)$	dt
$dB(t)$	dt	0
dt	0	0

.

Example 1.4.11. By applying Theorem 1.4.10 to $X(t) = B(t)$ ($f(t) = 1$ and $g(t) = 0$) and $\varphi(x) = x^2$. Then $\varphi'(x) = 2x$ and $\varphi''(x) = 2$, and we have

$$B(t)^2 = 2\int_0^t B(s)\,dB(s) + t, \tag{1.4.5}$$

or equivalently,

$$\int_0^t B(s)\,dB(s) = \frac{1}{2}(B(t)^2 - t). \tag{1.4.6}$$

This calculates the Itô integral of $B(t)$ on $[0, t]$.

Example 1.4.12 (Exponential Process)**.** Let $B(t)$ be a Brownian motion and $\{\mathcal{F}_t\}$ be its natural filtration. Let $f(t)$ be a stochastic process in $L^2_{\mathrm{ad}}([0,T]\times\Omega)$. Define

$$\mathcal{E}_f(t) = \exp\left\{\int_0^t f(s)\,dB(s) - \frac{1}{2}\int_0^t f^2(s)\,ds\right\}.$$

Then by Theorem 1.4.10, we have

$$d\mathcal{E}_f(t) = f(t)\mathcal{E}_f(t)\,dB(t).$$

This is in fact an analogue of the differential of an exponential function, so we call $\mathcal{E}_f(t)$ the *exponential process* of f.

1.5 Girsanov Theorem

The Girsanov theorem shows that a certain class of transitions of a Brownian motion are also Brownian motions with respect to equivalent probability measures. As an important corollary of the Itô formula, it has a wide range of applications.

Theorem 1.5.1 (Girsanov Theorem, [11])**.** *Let $B(t)$ be a Brownian motion and $\{\mathcal{F}_t\}$ be its natural filtration. Let $f(t)$ be a square integrable $\{\mathcal{F}_t\}$-adapted stochastic process such that $E_P\big[\mathcal{E}_f(t)\big] < \infty$ for all $t \in [0,T]$. Then the stochastic process given by*

$$\widetilde{B}(t) = B(t) - \int_0^t f(s)\,ds$$

is a Brownian motion with respect to the probability measure Q given by

$$dQ = \mathcal{E}_f(T)\, dP = \exp\left\{ \int_0^T f(t)\, dB(t) - \frac{1}{2}\int_0^T f^2(t)\, dt \right\} dP, \tag{1.5.1}$$

which is equivalent to P.

Chapter 2

General Stochastic Integral and General Itô Formula

As indicated in Section 1.1, for non-adapted or anticipative stochastic processes, the classical Itô integration theory is not enough. So we need to define a general stochastic integration to deal with both adapted and non-adapted cases. In this chapter, we will give the definition of a general stochastic integral and prove its Itô formula. It is worth to point out that the Itô formula coincides with the one proved in the white noise distribution theory, which resolves the problem stated in Section 1.1. However, as indicated there, the conclusions from the white noise distribution theory are too abstract to have a realistic meaning.

We start with the definition of the stochastic integral for non-adapted stochastic processes (or so-called *counter part*).

2.1 Stochastic Integral for Instantly Independent Stochastic Processes

Definition 2.1.1 (Counter Filtration)**.** A family $\{\mathcal{G}^{(t)}; a \leq t \leq b\}$ of complete σ-fields is called a *counter* $\{\mathcal{F}_t\}$*-filtration* if it satisfies the following conditions:

1. $\{\mathcal{G}^{(t)}; a \leq t \leq b\}$ is instantly independent with respect to $\{\mathcal{F}_t; a \leq t \leq b\}$, i.e., for each $t \in [a, b]$, $\mathcal{G}^{(t)}$ and $\mathcal{F}_t$ are independent;

2. $\{\mathcal{G}^{(t)}; a \leq t \leq b\}$ is decreasing, i.e., for $t_1 < t_2$, $\mathcal{G}^{(t_1)} \supset \mathcal{G}^{(t_2)}$;

3. $\widetilde{\mathcal{G}}^{(t)} := \sigma\{B(b) - B(s) : t \leq s \leq b\} \subset \mathcal{G}^{(t)}$ for each $t \in [a, b]$.

Remark 2.1.2. Note that

1. A family $\{\mathcal{G}^{(t)}; a \leq t \leq b\}$ that is instantly independent with respect to $\{\mathcal{F}_t; a \leq t \leq b\}$ may not be decreasing. For example, $\mathcal{G}^{(t)} = \sigma\{B(b) - B(t)\}$ is instantly independent with respect to $\{\mathcal{F}_t\}$, but not decreasing.

2. The backward filtration $\widetilde{\mathcal{G}}^{(t)}$ is a counter $\{\mathcal{F}_t\}$-filtration. But it is not the only one. For example,

$$\mathcal{G}^{(t)} := \sigma\{B(b) - B(s), B(2b) - B(s) : a \leq t \leq b\},$$

is also a counter $\{\mathcal{F}_t\}$-filtration.

3. If X_t and $Y^{(t)}$ are processes that are adapted to $\{\mathcal{F}_t\}$ and $\{\mathcal{G}^{(t)}\}$, respectively, where $\{\mathcal{F}_t\}$ and $\{\mathcal{G}^{(t)}\}$ are instantly independent, then X_t and $Y^{(t)}$ are instantly independent. In the sense of the Hilbert space $L^2(\Omega)$, X_t and $Y^{(t)}$ are orthogonal for each t. Thus, we can consider $L^2(\mathcal{G}^{(t)})$ as a subspace that is orthogonal to $L^2(\mathcal{F}_t)$.

Proposition 2.1.3. *If $\{\mathcal{G}^{(t)}\}$ is a counter $\{\mathcal{F}_t\}$-filtration, and a stochastic process X_t is adapted to $\{\mathcal{G}^{(t)}\}$, then X_t is instantly independent to $\{\mathcal{F}_t\}$.*

Proof. For any Borel set A and $t \in [a, b]$, $X_t^{-1}(A) \in \mathcal{G}^{(t)}$. So $X_t^{-1}(A)$ is independent of $\{\mathcal{F}_t\}$. Thus, X_t is independent of $\{\mathcal{F}_t\}$ for each t. □

From now on, we use $L^2_{\text{ct}}([a, b] \times \Omega)$ to denote the space of all stochastic processes $g(t, \omega)$, $a \leq t \leq b$, $\omega \in \Omega$, satisfying the following conditions:

1. $g(t, \omega)$ is adapted to the counter $\{\mathcal{F}_t\}$-filtration $\{\mathcal{G}^{(t)}\}$;

2. $\int_a^b E\left[|g(t)|^2\right] dt < \infty$.

Remark 2.1.4. In fact, in $L^2([a, b] \times \Omega)$, $L^2_{\text{ct}}([a, b] \times \Omega)$ is a subspace of the subspace that is orthogonal to $L^2_{\text{ad}}([a, b] \times \Omega)$, i.e., $L^2_{\text{ad}}([a, b] \times \Omega) \perp L^2_{\text{ct}}([a, b] \times \Omega)$.

Now, we can define the stochastic integral $\int_a^b g(t)\,dB(t)$, and then $\int_a^b f(t)g(t)\,dB(t)$, for $g \in L^2_{\text{ct}}([a,b] \times \Omega)$ and $f \in L^2_{\text{ad}}([a,b] \times \Omega)$. We divide the definition into the following several steps.

Step 1: *g is a step stochastic process in $L^2_{\text{ct}}([a,b] \times \Omega)$.*

Suppose g is a step stochastic process given by

$$g(t,\omega) = \sum_{i=1}^{n} \eta_i(\omega) 1_{[t_{i-1},t_i)}(t),$$

where η_i is $\mathcal{G}^{(t_i)}$-measurable and $E[\eta_i^2] < \infty$. Then $g \in L^2_{\text{ct}}([a,b] \times \Omega)$. Define

$$I(g) = \sum_{i=1}^{n} \eta_i \left(B(t_i) - B(t_{i-1})\right).$$

Then $I(f) + I(g) = I(f+g)$, for step functions f and g in $L^2_{\text{ct}}([a,b] \times \Omega)$.

Lemma 2.1.5. *$E(I(g)) = 0$ and $E\left(|I(g)|^2\right) = \int_a^b E\left(|g(t)|^2\right) dt$.*

Proof. For each $1 \leq i \leq n$,

$$\begin{aligned} E\left(\eta_i \left(B(t_i) - B(t_{i-1})\right)\right) &= E\left[E\left(\eta_i \left(B(t_i) - B(t_{i-1})\right) |\mathcal{G}^{(t_i)}\right)\right] \\ &= E\left[\eta_i E\left(B(t_i) - B(t_{i-1})\right) |\mathcal{G}^{(t_i)}\right] \\ &= E\left[\eta_i E\left(B(t_i) - B(t_{i-1})\right)\right] \\ &= E(\eta_i \cdot 0) = 0. \end{aligned}$$

So $E(I(g)) = \sum_{i=1}^n E\left(\eta_i \left(B(t_i) - B(t_{i-1})\right)\right) = 0$.

$$|I(g)|^2 = \sum_{i,j=1}^{n} \eta_i \eta_j \left(B(t_i) - B(t_{i-1})\right)\left(B(t_j) - B(t_{j-1})\right).$$

For $i < j$,

$$\begin{aligned}
&E\left[\eta_i\eta_j\left(B(t_i)-B(t_{i-1})\right)\left(B(t_j)-B(t_{j-1})\right)\right] \\
=&E\left\{E\left[\eta_i\eta_j\left(B(t_i)-B(t_{i-1})\right)\left(B(t_j)-B(t_{j-1})\right)|\mathcal{G}^{(t_i)}\right]\right\} \\
=&E\left\{\eta_i\eta_j\left(B(t_j)-B(t_{j-1})\right)E\left[\left(B(t_i)-B(t_{i-1})\right)|\mathcal{G}^{(t_i)}\right]\right\} \\
=&E\left[\eta_i\eta_j\left(B(t_j)-B(t_{j-1})\right)E\left(B(t_i)-B(t_{i-1})\right)\right] \\
=&0.
\end{aligned}$$

So when $i \neq j$, $E\left[\eta_i\eta_j\left(B(t_i)-B(t_{i-1})\right)\left(B(t_j)-B(t_{j-1})\right)\right]=0$. Then

$$\begin{aligned}
E\left[\eta_i^2\left(B(t_i)-B(t_{i-1})\right)^2\right] =&E\left\{E\left[\eta_i^2\left(B(t_i)-B(t_{i-1})\right)^2|\mathcal{G}^{(t_i)}\right]\right\} \\
=&E\left\{\eta_i^2 E\left[\left(B(t_i)-B(t_{i-1})\right)^2|\mathcal{G}^{(t_i)}\right]\right\} \\
=&E\left\{\eta_i^2 E\left[\left(B(t_i)-B(t_{i-1})\right)^2\right]\right\} \\
=&E\left[\eta_i^2(t_i-t_{i-1})\right] \\
=&E\left(\eta_i^2\right)(t_i-t_{i-1}).
\end{aligned}$$

Thus,

$$\begin{aligned}
E\left(|I(g)|^2\right) &= E\left[\sum_{i,j=1}^{n}\eta_i\eta_j\left(B(t_i)-B(t_{i-1})\right)\left(B(t_j)-B(t_{j-1})\right)\right] \\
&= E\left[\sum_{i=1}^{n}\eta_i^2\left(B(t_i)-B(t_{i-1})\right)^2\right] \\
&= \sum_{i=1}^{n}E\left(\eta_i^2\right)(t_i-t_{i-1}) \\
&= \int_a^b E\left(|g(t)|^2\right)\,dt.
\end{aligned}$$

$\square$

Step 2: *An approximation lemma.*

Lemma 2.1.6. *Suppose $g \in L^2_{ct}([a,b]\times\Omega)$. Then there exists a sequence $\{g_n(t); n \geq 1\}$ of step stochastic processes in $L^2_{ct}([a,b]\times\Omega)$ such that*

$$\lim_{n\to\infty} \int_a^b E\left[|g(t) - g_n(t)|^2\right] dt = 0.$$

Proof. **Case I:** *$E[g(t)g(s)]$ is a continuous function of $(t,s) \in [a,b]^2$.*

Let $a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b$ be a partition of $[a,b]$ and define a stochastic process $g_n(t,\omega)$ by

$$g_n(t,\omega) = g(t_i,\omega), \quad t_{i-1} \leq t < t_i.$$

Then $\{g_n\}$ is a sequence of $\mathcal{G}(t)$-adapted processes. By the continuity of $E[g(t)g(s)]$ on $[a,b]^2$, we have

$$\lim_{s\to t} E[|g(t) - g(s)|^2] = \lim_{s\to t} [E[g(t)g(t)] - 2E[g(t)g(s)] + E[g(s)g(s)]] = 0,$$

which implies that for each $t \in [a,b]$,

$$\lim_{n\to\infty} E\left[|g(t) - g_n(t)|^2\right] = 0.$$

Moreover,

$$|g(t) - g_n(t)|^2 \leq 2\left[|g(t)|^2 + |g_n(t)|^2\right].$$

So for all $a \leq t \leq b$,

$$\begin{aligned} E[|g(t) - g_n(t)|^2] &\leq 2\left[E(|g(t)|^2) + E(|g_n(t)|^2)\right] \\ &\leq 4 \sup_{a\leq s\leq b} E(|g(s)|^2). \end{aligned}$$

The right hand side is finite by the continuity assumption. Thus, by the Lebesgue Dominated Convergence Theorem, we have

$$\lim_{n\to\infty} \int_a^b E\left[|g(t) - g_n(t)|^2\right] dt = \int_a^b \left[\lim_{n\to\infty} E(|g(t) - g_n(t)|^2)\right] dt = 0.$$

Case II: *g is bounded: $|g| \leq M$ for some $M > 0$.*

Define $\tilde{g}_n(t,\omega) = \int_0^{n(b-t)} e^{-\tau} g(t+n^{-1}\tau,\omega)\,d\tau$. Then $\tilde{g}_n$ is adapted to $\mathcal{G}^{(t)}$ since the integral is a pointwise limit of a sum of stochastic processes that are measurable with respect to $\mathcal{G}^{(t)}$. Moreover,

$$\begin{aligned}
\int_a^b E\left[|\tilde{g}_n(t)|^2\right]\,dt &= \int_a^b E\Big[\Big|\int_0^{n(b-t)} e^{-\tau} g(t+n^{-1}\tau,\omega)\,d\tau\Big|^2\Big]\,dt \\
&\le \int_a^b E\Big[\Big|\int_0^{n(b-t)} M\,d\tau\Big|^2\Big]\,dt \\
&= \int_a^b M^2 n^2 (b-t)^2\,dt < \infty.
\end{aligned}$$

And we have the following properties of g_n.

∘ Claim (a): *For each n, $E\left(\tilde{g}_n(t)\tilde{g}_n(s)\right)$ is a continuous function of (t,s).*

Take the substitution $u = t + n^{-1}\tau$. Then

$$\tilde{g}_n(t,\omega) = \int_t^b n e^{n(t-u)} g(u,\omega)\,du.$$

Thus,

$$\begin{aligned}
0 &\le \lim_{t\to s} E\left(|\tilde{g}_n(t) - \tilde{g}_n(s)|^2\right) \\
&= \lim_{t\to s} E\Big(\Big|\int_t^s n e^{n(t-u)} g(u,\omega)\,du\Big|^2\Big) \\
&\le \lim_{t\to s} n^2 M^2 (s-t)^2 = 0,
\end{aligned}$$

which implies

$$\lim_{t\to s} E\left(|\tilde{g}_n(t) - \tilde{g}_n(s)|^2\right) = 0.$$

So

$$\begin{aligned}
0 &\le |E\left(\tilde{g}_n(t)\tilde{g}_n(s)\right) - E\left(\tilde{g}_n(t_0)\tilde{g}_n(s_0)\right)| \\
&\le |E\left(\left(\tilde{g}_n(t) - \tilde{g}_n(t_0)\right)\tilde{g}_n(s)\right) - E\left(\tilde{g}_n(t_0)\left(\tilde{g}_n(s) - \tilde{g}_n(s_0)\right)\right)| \\
&\le \sqrt{E\left(|\tilde{g}_n(t) - \tilde{g}_n(t_0)|^2\right) E\left(|\tilde{g}_n(s)|^2\right)} - \sqrt{E\left(|\tilde{g}_n(t_0)|^2\right) E\left(|\tilde{g}_n(s) - \tilde{g}_n(s_0)|^2\right)} \\
&\to 0,
\end{aligned}$$

as $t \to t_0$ and $s \to s_0$. The claim follows.

∘ Claim (b): $\int_a^b E\left(|g(t) - \tilde{g}_n(t)|^2\right) dt \to 0,\ as\ n \to \infty.$

$$g(t) - \tilde{g}_n(t) = \int_0^\infty e^{-\tau}\left(g(t) - g(t + n^{-1}\tau)\right) d\tau,$$

where $g(t)$ is understood to be zero when $t > b$. Since $e^{-\tau}$ is a probability measure on $[0, \infty)$, we can apply the Schwartz inequality to get

$$|g(t) - \tilde{g}_n(t)|^2 \leq \int_0^\infty e^{-\tau}|g(t) - g(t + n^{-1}\tau)|^2 \, d\tau.$$

Then

$$\begin{aligned}
\int_a^b E\left(|g(t) - \tilde{g}_n(t)|^2\right) dt &\leq \int_a^b E\Big(\int_0^\infty e^{-\tau}|g(t) - g(t + n^{-1}\tau)|^2 \, d\tau\Big) \, dt \\
&= \int_a^b \int_0^\infty e^{-\tau} E\left(|g(t) - g(t + n^{-1}\tau)|^2\right) d\tau \, dt \\
&= \int_0^\infty e^{-\tau}\Big(\int_a^b E\big(|g(t) - g(t + n^{-1}\tau)|^2\big) \, dt\Big) \, d\tau \\
&= \int_0^\infty e^{-\tau} E\Big(\int_a^b |g(t) - g(t + n^{-1}\tau)|^2 \, dt\Big) \, d\tau \\
&\leq 4M^2(b - a) < \infty.
\end{aligned}$$

Thus, by the Lebesgue Dominated Convergence Theorem, we get the claim.

Now, by Claim (a) we can apply Case I to $\tilde{g}_n$ for each n to get a $\left\{\mathcal{G}^{(t)}\right\}$-adapted step stochastic process $g_n(t, \omega)$ such that

$$\int_a^b E\left(|g_n(t) - \tilde{g}_n(t)|^2\right) dt < \frac{1}{n}.$$

Then by Claim (b), we have

$$\lim_{n \to \infty} \int_a^b E\left(|g(t) - g_n(t)|^2\right) dt = 0.$$

Case III: *The general case for* $g \in L^2_{ct}\left([a, b] \times \Omega\right)$.

Let $g \in L^2_{\text{ct}}([a,b] \times \Omega)$. For each n, define

$$\bar{g}_n(t,\omega) = \begin{cases} g(t,\omega), & \text{if } |g(t,\omega)| \le n, \\ 0, & \text{if } |g(t,\omega)| > n. \end{cases}$$

Then

$$|g(t) - \bar{g}_n(t)|^2 \le 2\left(|g(t)|^2 + |\bar{g}_n(t)|^2\right) \le 4|g(t)|^2.$$

By the Lebesgue Dominated Convergence Theorem and $\int_a^b E\left(|g(t)|^2\right) dt < \infty$, we have

$$\lim_{n\to\infty} \int_a^b E\left(|g(t) - \bar{g}_n(t)|^2\right) dt = 0.$$

Now, for each $n \ge 1$, we apply Case II to $\bar{g}_n$ to get a $\left\{\mathcal{G}^{(t)}\right\}$-adapted step stochastic process $g_n(t,\omega)$ such that

$$\int_a^b E\left(|g_n(t) - \bar{g}_n(t)|^2\right) dt < \frac{1}{n}.$$

Hence,

$$\lim_{n\to\infty} \int_a^b E\left(|g(t) - g_n(t)|^2\right) dt = 0,$$

by the two equations above.

□

Step 3: *Stochastic integral* $\int_a^b g(t)\, dB(t)$ *for* $g \in L^2_{ct}([a,b] \times \Omega)$.

By Lemma 2, we can find a sequence $\{g_n(t,\omega); n \ge 1\}$ of $\left\{\mathcal{G}^{(t)}\right\}$-adapted step stochastic processes such that

$$\lim_{n\to\infty} \int_a^b E\left(|g(t) - g_n(t)|^2\right) dt = 0.$$

For each n, $I(g_n)$ is defined as in Step 1. By Lemma 1, we get

$$E\left(|I(g_n) - I(g_m)|^2\right) = \int_a^b E\left(|g_n(t) - g_m(t)|^2\right) dt \to 0, \quad \text{as } m, n \to \infty,$$

since $I(g_n) - I(g_m) = I(g_n - g_m)$ whereas $g_n - g_m$ is still a $\{\mathcal{G}^{(t)}\}$-adapted step stochastic process. Hence the sequence $\{I(g_n)\}$ is a Cauchy sequence in $L^2(\Omega)$. Define

$$I(g) = \lim_{n\to\infty} I(g_n), \quad \text{in } L^2(\Omega). \tag{2.1.1}$$

The stochastic integral $I(g)$ is well-defined. In fact, if there is another sequence $\{g_n\}$ of $\{\mathcal{G}^{(t)}\}$-adapted step stochastic processes such that

$$\lim_{n\to\infty} \int_a^b E\left(|g(t) - f_n(t)|^2\right)\, dt = 0,$$

then

$$\begin{aligned}
E\left(|I(f_n) - I(g_m)|^2\right) &= E\left(|I(f_n - g_m)|^2\right) \\
&= \int_a^b E\left(|f_n(t) - g_m(t)|^2\right)\, dt \\
&\leq 2\int_a^b E\left(|f_n(t) - g(t)|^2\right)\, dt + 2\int_a^b E\left(|g(t) - g_m(t)|^2\right)\, dt \\
&\to 0, \quad \text{as } m, n \to \infty.
\end{aligned}$$

Thus,

$$\lim_{n\to\infty} I(f_n) = \lim_{n\to\infty} I(g_m), \quad \text{in } L^2(\Omega).$$

Now, we define the stochastic integral of the counter part as follows.

Definition 2.1.7. The limit $I(g)$ defined in Equation (2.1.1) is called the *general stochastic integral* of $g \in L^2_{\mathrm{ct}}([a,b]\times\Omega)$ and is denoted by $\int_a^b g(t)\, dB(t)$.

It is obvious that I is linear, namely, for any $\alpha, \beta \in \mathbb{R}$ and $f, g \in L^2_{\mathrm{ct}}([a,b]\times\Omega)$, we have $I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)$.

Theorem 2.1.8 (Isometry)**.** *Suppose $g \in L^2_{ct}\big([a,b]\times\Omega\big)$. Then $\int_a^b g(t)\, dB(t)$ is a random variable with $E\left(\int_a^b g(t)\, dB(t)\right) = 0$ and*

$$E\Big(\Big|\int_a^b g(t)\, dB(t)\Big|^2\Big) = \int_a^b E\left(|g(t)|^2\right)\, dt.$$

Proof. By the definition of $\int_a^b g(t)\,dB(t)$, there exists a sequence of $\{\mathcal{G}^{(t)}\}$-adapted step stochastic processes g_n, such that

$$\lim_{n\to\infty}\int_a^b E\left(|g(t)-g_n(t)|^2\right)\,dt=0, \tag{2.1.2}$$

and

$$\int_a^b g(t)\,dB(t)=\lim_{n\to\infty}\int_a^b g_n(t)\,dB(t),\quad \text{in } L^2(\Omega). \tag{2.1.3}$$

But by Lemma 2.1.5, we have

$$E\Big(\int_a^b g_n(t)\,dB(t)\Big)=0 \quad\text{and}\quad E\Big(\Big|\int_a^b g_n(t)\,dB(t)\Big|^2\Big)=\int_a^b E\left(|g_n(t)|^2\right)\,dt.$$

Therefore,

$$\begin{aligned}\Big|E\Big(\int_a^b g(t)\,dB(t)\Big)\Big| &= \Big|E\Big(\int_a^b g(t)\,dB(t)\Big)-E\Big(\int_a^b g_n(t)\,dB(t)\Big)\Big|\\ &=\Big|E\Big(\int_a^b g(t)-g_n(t)\,dB(t)\Big)\Big|\\ &\le\Big(E\Big[\Big|\int_a^b g(t)-g_n(t)\,dB(t)\Big|^2\Big]\Big)^{\frac{1}{2}}\\ &=\int_a^b E\left(|g(t)-g_n(t)|^2\right)\,dt\to 0,\quad \text{as } n\to\infty.\end{aligned}$$

The left hand side is independent of n. So $E\big(\int_a^b g(t)\,dB(t)\big)=0$.

By Equation (2.1.2) and (2.1.3), we have

$$\Big\|\int_a^b g(t)\,dB(t)\Big\|_{L^2(\Omega)}=\lim_{n\to\infty}\Big\|\int_a^b g_n(t)\,dB(t)\Big\|_{L^2(\Omega)},$$

and

$$\|g\|_{L^2_{\mathrm{ct}}([a,b]\times\Omega)}=\lim_{n\to\infty}\|g_n\|_{L^2_{\mathrm{ct}}([a,b]\times\Omega)}.$$

So

$$
\begin{aligned}
E\Big(\Big|\int_a^b g(t)\,dB(t)\Big|^2\Big) &= \Big\|\int_a^b g(t)\,dB(t)\Big\|_{L^2(\Omega)} = \lim_{n\to\infty}\Big\|\int_a^b g_n(t)\,dB(t)\Big\|_{L^2(\Omega)} \\
&= \lim_{n\to\infty} E\Big(\Big|\int_a^b g_n(t)\,dB(t)\Big|^2\Big) = \lim_{n\to\infty}\int_a^b E\left(|g_n(t)|^2\right)\,dt \\
&= \lim_{n\to\infty}\|g_n\|_{L^2_{\mathrm{ct}}([a,b]\times\Omega)} = \|g\|_{L^2_{\mathrm{ct}}([a,b]\times\Omega)} \\
&= \int_a^b E\left(|g(t)|^2\right)\,dt.
\end{aligned}
$$

□

By this theorem, the stochastic integration operator $I : L^2_{\mathrm{ct}}([a,b]\times\Omega) \to L^2(\Omega)$ is an isometry.

From Case I of the proof of Lemma 2.1.6, we have the following theorem:

Theorem 2.1.9. *Suppose $g \in L^2_{ct}([a,b]\times\Omega)$ and assume that $E(g(t)g(s))$ is a continuous function of t and s. Then*

$$
\int_a^b g(t)\,dB(t) = \lim_{n\to\infty}\sum_{i=1}^n g(t_i)\left(B(t_i) - B(t_{i-1})\right), \quad \text{in } L^2(\Omega).
$$

We can rewrite the equation in Theorem 2.1.9 as

$$
\lim_{n\to\infty}\sum_{i=1}^n \Big[g(t_i)\left(B(t_i) - B(t_{i-1})\right) - \int_{t_{i-1}}^{t_i} g(t)\,dB(t)\Big] = 0, \quad \text{in } L^2(\Omega),
$$

so is true in probability. In this sense, we can write in an asymptotic and symbolic way that $\Delta Y^i \approx -g(t_i)\Delta B_i$, where the stochastic process $Y^{(t)} := \int_t^T g(s)\,dB(s)$, and $\Delta Y^i = Y^{(t_i)} - Y^{(t_{i-1})}$, $\Delta B_i = B_{t_i} - B_{t_{i-1}}$.

The continuity of the stochastic process $Y^{(t)}$ defined above can be proved similarly as in [20] for the classical Itô stochastic integral.

Now, we can define the stochastic integral for the combination of adapted and instantly independent processes. From now on, if it is not specified, then the statement "$\varphi(t)$ is instantly independent of $\{\mathcal{F}_t\}$" means "$\varphi(t)$ is adapted to a counter $\{\mathcal{F}_t\}$-filtration".

2.2 General Stochastic Integral

The general stochastic integral for both adapted and non-adapted processes was firstly introduced in [1] by W. Ayed and H.-H. Kuo.

Definition 2.2.1 (Ayed–Kuo [1])**.** The *stochastic integral* of $f(t)\varphi(t)$, where f is adapted and φ is instantly independent, is defined by

$$\int_a^b f(t)\varphi(t)\,dB(t) = \lim_{\|\Delta_n\|\to 0} \sum_{j=1}^n f(t_{j-1})\varphi(t_j)\big(B(t_j) - B(t_{j-1})\big),$$

provided that the limit in probability exists.

Definition 2.2.2 (Hwang–Kuo–Saitô–Zhai [14])**.** Suppose $\Phi(t), a \le t \le b$, is a stochastic process of the following form

$$\Phi(t) = \sum_{i=1}^m f_i(t)\varphi_i(t), \tag{2.2.1}$$

where $f_i(t)$'s are $\{\mathcal{F}_t\}$-adapted continuous stochastic processes and $\varphi_i(t)$'s are continuous stochastic processes being instantly independent of $\{\mathcal{F}_t\}$. The *stochastic integral* of $\Phi(t)$ is defined by

$$\int_a^b \Phi(t)\,dB(t) = \sum_{i=1}^m \int_a^b f_i(t)\varphi_i(t)\,dB(t),$$

where the integrals $\int_a^b f_i(t)\varphi_i(t)\,dB(t)$ are defined by

$$\int_a^b f_i(t)\varphi_i(t)\,dB(t) = \lim_{\|\Delta_n\|\to 0} \sum_{j=1}^n f_i(t_{j-1})\varphi_i(t_j)(B(t_j) - B(t_{j-1})),$$

provided that the limit in probability exists.

Intuitively, the new stochastic integral uses respectively the left endpoint for the Itô (adapted) part and the right endpoint for the counter (instantly independent) part as the evaluation points of integration. In comparison, the Itô integral only has adapted part, and thus uses the left endpoint. However, it is not proved in [1] that the new integral is well-defined. We filled this gap in the following lemma.

Lemma 2.2.3 (Hwang–Kuo–Saitô–Zhai [14]). *Let $f_i(t), 1 \leq i \leq m$, $g_j(t), 1 \leq j \leq n$, be $\{\mathcal{F}_t\}$-adapted continuous stochastic processes and let $\varphi_i(t), 1 \leq i \leq m$, $\xi_j(t), 1 \leq j \leq n$, be continuous stochastic processes being instantly independent of $\{\mathcal{F}_t\}$. Suppose the stochastic integrals $\int_a^b f_i(t)\varphi_i(t)\, dB(t)$ and $\int_a^b g_j(t)\xi_j(t)\, dB(t)$ exist for $1 \leq i \leq m, 1 \leq j \leq n$. Assume that*

$$\sum_{i=1}^{m} f_i(t)\varphi_i(t) = \sum_{j=1}^{n} g_j(t)\xi_j(t), \quad a \leq t \leq b.$$

Then the following equality holds:

$$\sum_{i=1}^{m} \int_a^b f_i(t)\varphi_i(t)\, dB(t) = \sum_{j=1}^{n} \int_a^b g_j(t)\xi_j(t)\, dB(t). \tag{2.2.2}$$

Remark 2.2.4. When $f(t)\varphi(t) = g(t)\xi(t)$ (for Case 1 in the proof below), we do not need to assume the existence of $\int_a^b f(t)\varphi(t)\, dB(t)$ and $\int_a^b g(t)\xi(t)\, dB(t)$ as limits in probability. Equation (2.2.2) for this case is understood to mean that if the limit in probability on one side of the equality exists, then it also exists on the other side and the equality holds.

Proof. We divide the proof into several cases:

Case 1: $m = n = 1$.

Suppose $f(t)\varphi(t) = g(t)\xi(t), a \leq t \leq b$. Without loss of generality, we assume that $f(t), \varphi(t), g(t), \xi(t)$ are non-zero. Then

$$\frac{f(t)}{g(t)} = \frac{\xi(t)}{\varphi(t)}.$$

Note that in the above equation, the left-hand side is $\{\mathcal{F}_t\}$-adapted and the right-hand side is instantly independent of $\{\mathcal{F}_t\}$. However, a continuous stochastic process that is both adapted to and instantly independent of $\{\mathcal{F}_t\}$ must be a continuous deterministic function. Denote this function as $a(t)$. Then

$$\frac{f(t)}{g(t)} = \frac{\xi(t)}{\varphi(t)} = a(t).$$

So

$$f(t) = g(t)a(t) \quad \text{and} \quad \varphi = \frac{1}{a(t)}\xi(t).$$

Thus, for any partition $a = t_0 < t_1 < \cdots < t_l = b$,

$$\begin{aligned}
\int_a^b f(t)\varphi(t)\, dB(t) &\approx \sum_{k=1}^{l} f(t_{i-1})\varphi(t_i)\big(B(t_i) - B(t_{i-1})\big) \\
&\approx \sum_{k=1}^{l} \big(a(t_{i-1})g(t_{i-1})\big)\Big(\frac{1}{a(t_i)}\varphi(t_i)\Big)\big(B(t_i) - B(t_{i-1})\big) \\
&\approx \sum_{k=1}^{l} g(t_{i-1})\xi(t_i)\big(B(t_i) - B(t_{i-1})\big) \\
&\approx \int_a^b g(t)\xi(t)\, dB(t),
\end{aligned}$$

which proves the theorem for this case.

Case 2: $m = 1, n = 2$.

Assume

$$f(t)\varphi(t) = g_1(t)\xi_1(t) + g_2(t)\xi_2(t). \tag{2.2.3}$$

Let $a(t) = E\varphi(t), b_1(t) = E\xi_1(t), b_2(t) = E\xi_2(t)$ and assume without loss of generality that $a(t), b_1(t), b_2(t)$ are non-zero. Then by taking conditional expectation of Equation (2.2.3) with respect to $\{\mathcal{F}_t\}$, we get

$$f(t)a(t) = g_1(t)b_1(t) + g_2(t)b_2(t). \tag{2.2.4}$$

Let $\tilde{a}(t) = a(t)/b_2(t), \tilde{b}_1(t) = b_1(t)/b_2(t)$. Then we have

$$g_2(t) = f(t)\tilde{a}(t) - g_1(t)\tilde{b}_1(t). \tag{2.2.5}$$

Substitute $g_2(t)$ in Equation (2.2.5) into Equation (2.2.3) to get

$$f(t)\big[\varphi(t) - \tilde{a}(t)\xi_2(t)\big] = g_1(t)\big[\xi_1(t) - \tilde{b}_1(t)\xi_2(t)\big]. \tag{2.2.6}$$

By Remark 2.2.4, we can apply Case 1. Hence

$$\int_a^b f(t)\big[\varphi(t) - \tilde{a}(t)\xi_2(t)\big]\, dB(t) = \int_a^b g_1(t)\big[\xi_1(t) - \tilde{b}_1(t)\xi_2(t)\big]\, dB(t). \tag{2.2.7}$$

But we can use Definition 2.2.1 to see that

$$\int_a^b f(t)\big[\varphi(t) - \tilde{a}(t)\xi_2(t)\big]\,dB(t) = \int_a^b f(t)\varphi(t)\,dB(t) - \int_a^b f(t)\tilde{a}(t)\xi_2(t)\,dB(t), \tag{2.2.8}$$

and

$$\int_a^b g_1(t)\big[\xi_1(t) - \tilde{b}_1(t)\xi_2(t)\big]\,dB(t) = \int_a^b g_1(t)\xi_1(t)\,dB(t) - \int_a^b g_1(t)\tilde{b}_1(t)\xi_2(t)\,dB(t), \tag{2.2.9}$$

since both sides have the same limit in probability. It follows from Equations (2.2.5), (2.2.7), (2.2.8), and (2.2.9) that

$$\begin{aligned}
&\int_a^b f(t)\varphi(t)\,dB(t) \\
&= \int_a^b g_1(t)\xi_1(t)\,dB(t) + \int_a^b \big[f(t)\tilde{a}(t) - g_1(t)\tilde{b}_1(t)\big]\xi_2(t)\,dB(t) \\
&= \int_a^b g_1(t)\xi_1(t)\,dB(t) + \int_a^b g_2(t)\xi_2(t)\,dB(t),
\end{aligned}$$

which proves Case 2.

Case 3: $m = 1, n \geq 3$.

Assume

$$f(t)\varphi(t) = \sum_{j=1}^{n} g_j(t)\xi_j(t).$$

Then similar to Equation (2.2.5), we have

$$g_n(t) = f(t)\tilde{a}(t) - \sum_{j=1}^{n-1} g_j(t)\tilde{b}_j(t),$$

where $\tilde{a}(t) = E\varphi(t)/E\xi_n(t)$ and $\tilde{b}_j(t) = E\xi_j(t)/E\xi_n(t), 1 \leq j \leq n-1$. Then, similar to Equation (2.2.7), we have

$$f(t)\big[\varphi(t) - \tilde{a}(t)\xi_n(t)\big] = \sum_{j=1}^{n-1} g_j(t)\big[\xi_j(t) - \tilde{b}_j(t)\xi_n(t)\big].$$

Thus, this case has been reduced to the case for $n-1$. Hence we can use the same arguments as those in Case 2 and the induction to prove the lemma for this case.

Case 4: $m \geq 2, n \geq 1$.

Assume

$$\sum_{i=1}^{m} f_i(t)\varphi_i(t) = \sum_{j=1}^{n} g_j(t)\xi_j(t).$$

which can be rewritten as

$$f_1(t)\varphi_1(t) = \sum_{j=1}^{n} g_j(t)\xi_j(t) - \sum_{i=2}^{m} f_i(t)\varphi_i(t).$$

Therefore, this case is reduced to Case 3 and the lemma is proved for general case. $\square$

In general, if a stochastic process $\Phi(t)$ can be written as a series of the form $f(t)\varphi(t)$ in $L^2(\Omega)$, we define the stochastic integral of $\Phi(t)$ to be the sum of the series in probability.

Definition 2.2.5 (Hwang–Kuo–Saitô–Zhai [14])**.** Suppose $\Phi(t), a \leq t \leq b$, is a stochastic process and there exists a sequence $\{\Phi_n(t)\}_{n=1}^{\infty}$ of stochastic processes of the form in Equation (2.2.1) satisfying the conditions:

1. $\int_a^b |\Phi(t) - \Phi_n(t)|^2 \, dt \to 0$ almost surely.

2. $\int_a^b \Phi_n(t) \, dB(t)$ converges in probability.

Then the *stochastic integral* of $\Phi(t)$ is defined by

$$\int_a^b \Phi(t) \, dB(t) = \lim_{n \to \infty} \int_a^b \Phi_n(t) \, dB(t), \quad \text{in probability.} \tag{2.2.10}$$

Example 2.2.6. We have

$$\int_0^t B(1) \, dB(s) = \int_0^t \big(B(s) + \big(B(1) - B(s)\big)\big) \, dB(s) = B(1)B(t) - t, \tag{2.2.11}$$

for $0 \leq t \leq 1$.

2.3 Itô Isometry for General Stochastic Integral

It is mentioned in Theorem 1.4.4 that the classical Itô integration operator is an isometry. In [25], an isometry equality for general stochastic integral is proved, and it is same to the isometry equality obtained in the white noise distribution theory.

Theorem 2.3.1 ([25]). *Let $f(t)$ and $\varphi(t)$ be stochastic processes adapted to and instantly independent of $\{\mathcal{F}_t\}$, respectively. Then we have,*

$$E\Big[\int_0^T f(t)\varphi(t)\,dB(t)\Big] = 0.$$

Theorem 2.3.2 ([25]). *Let $f(t)$ and $\varphi(x)$ be as above. In addition, assume that $f(t)$ and $\varphi(x)$ are equal to their Maclaurin series on $\mathbb{R}$. Let $B(t)$ be a Brownian motion. Then we have*

$$\begin{aligned} &E\Big[\Big(\int_0^T f(B(t))\varphi(B(T)-B(t))\,dB(t)\Big)^2\Big] \\ &= \int_0^T E\big[\big(f(B(t))\varphi(B(T)-B(t))\big)^2\big]\,dt \\ &\quad + 2\int_0^T\!\!\int_0^t E\big[f(B(s))\varphi'(B(T)-B(s))f'(B(t))\varphi(B(T)-B(t))\big]\,dsdt. \end{aligned} \tag{2.3.1}$$

2.4 A General Itô Formula

The Itô formula plays the central role in both theory and application of Itô calculus. For the general stochastic integral defined in Section 2.2, we proved its Itô formula in [14]. Consider the following stochastic processes

$$X_t = X_a + \int_a^t g(s)\,dB(s) + \int_a^t h(s)\,ds, \tag{2.4.1}$$

$$Y^{(t)} = Y^{(b)} + \int_t^b \xi(s)\,dB(s) + \int_t^b \eta(s)\,ds, \tag{2.4.2}$$

where $g(t)$ and $h(t)$ are $\{\mathcal{F}_t\}$-adapted so that X_t is an Itô process, and $\xi(t)$ and $\eta(t)$ are instantly independent of $\{\mathcal{F}_t\}$ such that $Y^{(t)}$ is also instantly independent of $\{\mathcal{F}_t\}$. Thus $Y^{(t)}$ is a stochastic process in the counter part. Then we have the following theorem and we use the proof from [14].

Theorem 2.4.1. *Let $X_t, a \leq t \leq b$, be an Itô process given by Equation* (2.4.1) *and $Y^{(t)}, a \leq t \leq b$, an instantly independent process given by Equation* (2.4.2). *Suppose $\theta(x, y)$ is a real-valued C^2-function on $\mathbb{R}^2$. Then the following equality holds for $a \leq t \leq b$:*

$$\begin{aligned}\theta(X_t, Y^{(t)}) =& \theta(X_a, Y^{(a)}) + \int_a^t \theta_x(X_s, Y^{(s)})\, dX_s + \frac{1}{2}\int_a^t \theta_{xx}(X_s, Y^{(s)})(dX_s)^2 \\ &+ \int_a^t \theta_y(X_s, Y^{(s)})\, dY^{(s)} - \frac{1}{2}\int_a^t \theta_{yy}(X_s, Y^{(s)})(dY^{(s)})^2,\end{aligned}$$

which can be expressed symbolically in terms of stochastic differentials as

$$d\theta(X_t, Y^{(t)}) = \theta_x\, dX_t + \frac{1}{2}\theta_{xx}(dX_t)^2 + \theta_y\, dY^{(t)} - \frac{1}{2}\theta_{yy}(dY^{(t)})^2. \tag{2.4.3}$$

Proof. Since $\theta(x, y)$ is a real-valued C^2-function, we have the estimate

$$\begin{aligned}\theta(x, y) \approx& \theta(x_0, y_0) + \theta_x(x_0, y_0)(x - x_0) + \theta_y(x_0, y_0)(y - y_0) \\ &+ \frac{1}{2}\theta_{xx}(x_0, y_0)(x - x_0)^2 + \frac{1}{2}\theta_{yy}(x_0, y_0)(y - y_0)^2 \\ &+ \theta_{xy}(x_0, y_0)(x - x_0)(y - y_0).\end{aligned} \tag{2.4.4}$$

Now, we calculate the stochastic differential $d\theta(X_t, Y^{(t)})$. For any partition $a = t_0 < t_1 < \cdots < t_n = t$ of the interval $[a, t]$, we have

$$\theta(X_t, Y^{(t)}) = \theta(X_a, Y^{(a)}) + \sum_{i=1}^{n} \Big[\theta(X_{t_i}, Y^{(t_i)}) - \theta(X_{t_{i-1}}, Y^{(t_{i-1})})\Big]. \tag{2.4.5}$$

Then for each $i = 1, 2, \ldots, n$, by Equation (2.4.4), we get

$$\begin{aligned}&\theta(X_{t_i}, Y^{(t_i)}) - \theta(X_{t_{i-1}}, Y^{(t_{i-1})}) \\ &\approx \theta_x(X_{t_{i-1}}, Y^{(t_{i-1})})(X_{t_i} - X_{t_{i-1}}) + \theta_y(X_{t_{i-1}}, Y^{(t_{i-1})})(Y^{(t_i)} - Y^{(t_{i-1})}) \\ &\quad + \frac{1}{2}\theta_{xx}(X_{t_{i-1}}, Y^{(t_{i-1})})(X_{t_i} - X_{t_{i-1}})^2 + \frac{1}{2}\theta_{yy}(X_{t_{i-1}}, Y^{(t_{i-1})})(Y^{(t_i)} - Y^{(t_{i-1})})^2 \\ &\quad + \theta_{xy}(X_{t_{i-1}}, Y^{(t_{i-1})})(X_{t_i} - X_{t_{i-1}})(Y^{(t_i)} - Y^{(t_{i-1})}) \\ &= \mathrm{I}_i + \mathrm{II}_i + \mathrm{III}_i + \mathrm{IV}_i + \mathrm{V}_i,\end{aligned} \tag{2.4.6}$$

where $\mathrm{I}_i, \mathrm{II}_i, \mathrm{III}_i, \mathrm{IV}_i, \mathrm{V}_i$ are defined term by term in the corresponding order, respectively.

Now, we consider the summations of these terms over i. For the first term in Equation (2.4.6), we have

$$\begin{aligned}\sum_{i=1}^{n} \mathrm{I}_i &= \sum_{i=1}^{n} \theta_x(X_{t_{i-1}}, Y^{(t_{i-1})})(X_{t_i} - X_{t_{i-1}}) \\ &\approx \sum_{i=1}^{n} \left[\theta_x(X_{t_{i-1}}, Y^{(t_i)}) + \theta_{xy}(X_{t_{i-1}}, Y^{(t_i)})(Y^{(t_i)} - Y^{(t_{i-1})})\right](X_{t_i} - X_{t_{i-1}}) \\ &\to \int_a^t \theta_x(X_s, Y^s)\, dX_s - \int_a^t \theta_{xy}(X_x, Y^x)\, (dX_s)(dY^{(s)}). \end{aligned} \tag{2.4.7}$$

For the second term in Equation (2.4.6), we have

$$\begin{aligned}\sum_{i=1}^{n} \mathrm{II}_i &= \sum_{i=1}^{n} \theta_y(X_{t_{i-1}}, Y^{(t_{i-1})})(Y^{(t_i)} - Y^{(t_{i-1})}) \\ &\approx \sum_{i=1}^{n} \left[\theta_y(X_{t_{i-1}}, Y^{(t_i)}) + \theta_{yy}(X_{t_{i-1}}, Y^{(t_i)})\right](Y^{(t_i)} - Y^{(t_{i-1})}) \\ &\to \int_a^t \theta_y(X_s, Y^s)\, dY^{(s)} - \int_a^t \theta_{yy}(X_x, Y^x)\, (dY^{(s)})^2. \end{aligned} \tag{2.4.8}$$

For the third term in Equation (2.4.6), we have

$$\begin{aligned}\sum_{i=1}^{n} \mathrm{III}_i &= \frac{1}{2}\sum_{i=1}^{n} \theta_{xx}(X_{t_{i-1}}, Y^{(t_{i-1})})(X_{t_i} - X_{t_{i-1}})^2 \\ &\to \frac{1}{2}\int_a^t \theta_{xx}(X_s, Y^s)\, (dX_s)^2. \end{aligned} \tag{2.4.9}$$

Note that we do not have to change $\theta_{xx}(X_{t_{i-1}}, Y^{(t_{i-1})})$ to $\theta_{xx}(X_{t_{i-1}}, Y^{(t_i)})$ since the integrator $(dX_s)^2 = g(s)^2\, ds$ from Equation (2.4.1). Similarly, for the fourth term in Equation (2.4.6), we have

$$\begin{aligned}\sum_{i=1}^{n} \mathrm{IV}_i &= \frac{1}{2}\sum_{i=1}^{n} \theta_{yy}(X_{t_{i-1}}, Y^{(t_{i-1})})(Y^{(t_i)} - Y^{(t_{i-1})})^2 \\ &\to \frac{1}{2}\int_a^t \theta_{yy}(X_s, Y^s)\, (dY^{(s)})^2. \end{aligned} \tag{2.4.10}$$

For the fifth term in Equation (2.4.6), we have

$$\begin{aligned}\sum_{i=1}^{n} \mathrm{V}_i &= \sum_{i=1}^{n} \theta_{xy}(X_{t_{i-1}}, Y^{(t_{i-1})})(X_{t_i} - X_{t_{i-1}})(Y^{(t_i)} - Y^{(t_{i-1})}) \\ &\to \int_a^t \theta_{xy}(X_s, Y^s)\,(dX_s)(dY^{(s)}). \end{aligned} \tag{2.4.11}$$

Finally, we sum up Equations (2.4.7)–(2.4.11) and get the stochastic differential of $\theta(X_t, Y^{(t)})$. And then, Equation (2.4.3) and the theorem follow. □

It is easy to generate the Itô formula for multiple stochastic processes.

Theorem 2.4.2. *Let $X_t^{(i)}, 1 \le i \le n$, be Itô processes given by Equation* (2.4.1) *and $Y_j^{(t)}, 1 \le j \le m$, instantly independent processes given by Equation* (2.4.2)*. Suppose $\theta(t, x_1, \ldots, x_n, y_1, \ldots, y_m)$ is a real-valued function being C^1 in t and C^2 in x_i and y_j. Then the stochastic differential of $\theta(t, X_t^{(1)}, \ldots, X_t^{(n)}, Y_1^{(t)}, \ldots, Y_m^{(t)})$ is given by*

$$\begin{aligned}&\theta(t, X_t^{(1)}, \ldots, X_t^{(n)}, Y_1^{(t)}, \ldots, Y_m^{(t)}) \\ &= \theta_t\,dt + \sum_{i=1}^{n} \theta_{x_i}\,dX_t^{(i)} + \frac{1}{2}\sum_{i,k=1}^{n} \theta_{x_i x_k}(dX_t^{(i)})(dX_t^{(k)}) \\ &\quad + \sum_{j=1}^{m} \theta_{y_j}\,dY_j^{(t)} - \frac{1}{2}\sum_{j,l=1}^{m} \theta_{y_j y_l}(dY_j^{(t)})(dY_l^{(t)}). \end{aligned} \tag{2.4.12}$$

Corollary 2.4.3. *Let X_t, Y_t be Itô processes given by Equation* (2.4.1)*, and $X^{(t)}, Y^{(t)}$ instantly independent processes given by Equation* (2.4.2)*. Then we have the following product rules:*

$$\begin{aligned} d(X_t Y_t) &= X_t\,dY_t + Y_t\,dX_t + (dX_t)(dY_t), \\ d(X^{(t)} Y^{(t)}) &= X^{(t)}\,dY^{(t)} + Y^{(t)}\,dX^{(t)} - (dX^{(t)})(dY^{(t)}), \\ d(X_t Y^{(t)}) &= X_t\,dY^{(t)} + Y^{(t)}\,dX_t. \end{aligned}$$

Chapter 3

Near-Martingale Property of Anticipating Stochastic Integration

3.1 Near-Martingale Property

In Section 1.4, we have seen from Theorem 1.4.5 the martingale property of a stochastic process defined by the classical Itô integral. However, one would not expect this property for a non-adapted process, not to say its general stochastic integral. Nevertheless, a similar property holds for general stochastic integrals. We discuss this property in this chapter. Recall the definition of martingale property. It is equivalent to the following three conditions:

1. For each $t \in [0, T]$, $E\,|X_t| < \infty$;
2. X_t is adapted to $\{\mathcal{F}_t\}$;
3. $E[X_t - X_s|\mathcal{F}_s] = 0$, for all $0 \le s < t \le T$.

The main reason that a stochastic process defined by general stochastic integral is not able to be a martingale is that the Condition 2 above does not hold for the counter part. So we expect a property that satisfies Condition 1 and 3. It is firstly introduced in [24] by H.-H. Kuo, A. Sae-Tang and B. Szozda.

Definition 3.1.1 (Kuo–Sae-Tang–Szozda [24])**.** A stochastic process X_t is called a *near-martingale* with respect to a filtration $\{\mathcal{F}_t\}$ if $E|X_t| < \infty$ for all $t \in [a, b]$ and $E[X_t - X_s|\mathcal{F}_s] = 0$ for all $a \le s < t \le b$.

The following theorem gives an intrinsic characterization of a near-martingale.

Theorem 3.1.2 (Hwang–Kuo–Saitô–Zhai [15])**.** *Let $X_t, a \le t \le b$ be a stochastic process with $E|X_t| < \infty$ for each $t \in [a, b]$ and let $Y_t = E[X_t|\mathcal{F}_t]$. Then X_t is a near-martingale if and only if Y_t is a martingale.*

Proof. First assume that X_t is a near-martingale. Then for any $s \leq t$ we have

$$E[Y_t|\mathcal{F}_s] = E\big[E[X_t|\mathcal{F}_t]|\mathcal{F}_s\big] = E[X_t|\mathcal{F}_s] = E[X_s|\mathcal{F}_s] = Y_s.$$

So Y_t is a martingale. On the other hand, if Y_t is a martingale, then for any $s \leq t$, we have

$$E[X_t|\mathcal{F}_s] = E\big[E[X_t|\mathcal{F}_t]|\mathcal{F}_s\big] = E[Y_t|\mathcal{F}_s] = Y_s = E[X_s|\mathcal{F}_s],$$

which implies that X_t is a near-martingale. $\square$

Example 3.1.3. From Example 2.2.6, we have seen the general stochastic integral

$$X_t = \int_0^t B(1)\, dB(s) = B(1)B(t) - t.$$

Then

$$Y_t = E[X_t|\mathcal{F}_t] = E[B(1)B(t) - t|\mathcal{F}_t] = B(t)E[B(1)|\mathcal{F}_t] - t = B(t)^2 - t.$$

But

$$\begin{aligned}
E[Y_t|\mathcal{F}_s] &= E[B(t)^2 - t|\mathcal{F}_s] \\
&= E\big[(B(t) - B(s) + B(s))^2|\mathcal{F}_s\big] - t \\
&= E\big[(B(t) - B(s))^2|\mathcal{F}_s\big] + 2E\big[(B(t) - B(s))B(s)|\mathcal{F}_s\big] + E\big[B(s)^2|\mathcal{F}_s\big] - t \\
&= E\big[(B(t) - B(s))^2\big] + 2B(s)E\big[(B(t) - B(s))\big] + B(s)^2 - t \\
&= t - s + 0 + B(s)^2 - t \\
&= B(s)^2 - s = Y_s,
\end{aligned}$$

which means Y_t is a martingale. By Theorem 3.1.2, X_t is a near-martingale.

Example 3.1.3 gives us an intuition that, similar to the martingale property of Itô integral shown in Theorem 1.4.5, we can expect that a stochastic process defined by the general stochastic integral is a near-martingale. In [24], the authors proved this conjecture.

Theorem 3.1.4 (Kuo–Sae-Tang–Szozda [24]). *The stochastic process X_t defined in Definition 2.2.1, i.e.,*

$$X(t) = \int_a^t f(s)\varphi(s)\,dB(s) \tag{3.1.1}$$

is a near-martingale with respect to $\{\mathcal{F}_t\}$.

Moreover, X_t defined in Theorem 3.1.4 is also a near-martingale with respect to a counter $\{\mathcal{F}_t\}$-filtration $\{\mathcal{G}^{(t)}\}$.

Theorem 3.1.5 (Kuo–Sae-Tang–Szozda [24]). *The stochastic process X_t defined in Theorem 3.1.4 is a near-martingale with respect to a counter $\{\mathcal{F}_t\}$-filtration $\{\mathcal{G}^{(t)}\}$.*

It is natural to extend the submartingale and supermartingale properties (Definition 1.3.11 and 1.3.12) to general stochastic integrals.

Definition 3.1.6 (Hwang–Kuo–Saitô–Zhai [15]). A stochastic process $X_t, a \leq t \leq b$, with $E|X_t| < \infty$ for all t, is called a *near-submartingale* with respect to a filtration $\{\mathcal{F}_t\}$, if for all $a \leq s < t \leq b$, we have

$$E[X_t|\mathcal{F}_s] \geq E[X_s|\mathcal{F}_s], \quad \text{almost surely,}$$

or equivalently,

$$E[X_t - X_s|\mathcal{F}_s] \geq 0, \quad \text{almost surely.} \tag{3.1.2}$$

Definition 3.1.7 (Hwang–Kuo–Saitô–Zhai [15]). A stochastic process $X_t, a \leq t \leq b$, with $E|X_t| < \infty$ for all t, is called a *near-supermartingale* with respect to a filtration $\{\mathcal{F}_t\}$, if for all $a \leq s < t \leq b$, we have

$$E[X_t|\mathcal{F}_s] \leq E[X_s|\mathcal{F}_s], \quad \text{almost surely,}$$

or equivalently,

$$E[X_t - X_s|\mathcal{F}_s] \leq 0, \quad \text{almost surely.} \tag{3.1.3}$$

By similar arguments, the intrinsic characterization of near-submartingale and near-supermartingale can be proved.

Theorem 3.1.8 (Hwang–Kuo–Saitô–Zhai [15])**.** *Let $X_t, a \le t \le b$ be a stochastic process with $E|X_t| < \infty$ for each $t \in [a,b]$ and let $Y_t = E[X_t|\mathcal{F}_t]$. Then X_t is a near-submartingale or near-supermartingale, if and only if Y_t is a submartingale or supermartingale, respectively.*

3.2 Doob–Meyer's Decomposition for Near-Submartingales

In Theorem 1.3.14, we have seen a unique decomposition of a submartingale. Its analogue near-submartingale also has a similar decomposition. The Doob–Meyer's decomposition for a near-submartingale random sequence $\{X_n\}_{n=1}^{\infty}$ with respect to a sequence filtration $\{\mathcal{F}_n\}_{n=1}^{\infty}$ is given in [27]. In this section, we give the Doob–Meyer's decomposition for near-submartingale processes.

Theorem 3.2.1 (Hwang–Kuo–Saitô–Zhai [15])**.** *Let $X_t, a \le t \le b$, be a continuous near-submartingale with respect to a continuous filtration $\{\mathcal{F}_t; a \le t \le b\}$. Then X_t has a unique decomposition*

$$X_t = M_t + A_t, \quad a \le t \le b, \tag{3.2.1}$$

where M_t is a continuous near-martingale with respect to $\{\mathcal{F}_t\}$, and A_t is a continuous stochastic process satisfying the conditions:

1. *$A_a = 0$;*

2. *A_t is increasing in t almost surely;*

3. *A_t is adapted to $\{\mathcal{F}_t\}$.*

Remark 3.2.2. Notice that the three conditions in the theorem for A_t are the same to the conditions for the compensator in the Doob–Meyer's decomposition for submartingales and that A_t indeed plays the same role in the decompositions. So we also call A_t the *compensator* of the near-submartingale X_t.

Proof. We prove the uniqueness first. Suppose

$$X_t = M_t + A_t = \widetilde{M}_t + \widetilde{A}_t,$$

where $M_t, \widetilde{M}_t$ are continuous near-martingales and $A_t, \widetilde{A}_t$ are continuous stochastic processes satisfying the three conditions in the theorem. By taking conditional expectation, we have

$$E[X_t|\mathcal{F}_t] = E[M_t|\mathcal{F}_t] + A_t = E[\widetilde{M}_t|\mathcal{F}_t] + \widetilde{A}_t.$$

By Theorem 3.1.2, $E[M_t|\mathcal{F}_t]$ and $E[\widetilde{M}_t|\mathcal{F}_t]$ are martingales, and by Theorem 3.1.8, $E[X_t|\mathcal{F}_t]$ is a submartingale. Thus, the above equation provides two Doob–Meyer's decompositions of the submartingale $E[X_t|\mathcal{F}_t]$. So $A_t = \widetilde{A}_t$ by Theorem 1.3.14, and then $M_t = \widetilde{M}_t$ by the decomposition of X_t. The uniqueness is proved.

To prove the existence, let $Y_t = E[X_t|\mathcal{F}_t]$. By Theorem 3.1.8, Y_t is a submartingale. By Theorem 1.3.14, the Doob–Meyer's decomposition for submartingales, we have $Y_t = N_t + A_t$, where N_t is a continuous martingale and A_t is a continuous stochastic process satisfying the three conditions in the theorem. Define

$$M_t = X_t - E[X_t|\mathcal{F}_t] + N_t.$$

Then we have

$$X_t = M_t + E[X_t|\mathcal{F}_t] - N_t = M_t + Y_t - N_t = M_t + A_t, \tag{3.2.2}$$

and

$$E[M_t|\mathcal{F}_t] = E[X_t|\mathcal{F}_t] - E\big[E[X_t|\mathcal{F}_t]|\mathcal{F}_t\big] + E[N_t|\mathcal{F}_t] = E[N_t|\mathcal{F}_t] = N_t.$$

But N_t is a martingale. So M_t is a near-martinagle by Theorem 3.1.2. Therefore, Equation (3.2.2) provides a Doob–Meyer's decomposition for the near-submartingale X_t. We have proved the existence and finalized the proof. □

3.3 General Girsanov Theorem

The most important application of the Itô formula is the Girsanov theorem (see Section 1.5). It plays a crucial role in transform of probability measure and application in the mathematical finance. For general stochastic integral, H.-H. Kuo, Y. Peng and B. Szozda proved in [23] a general Girsanov theorem by applying the general Itô formula. Recall that in the proof of the Girsanov theorem, the key tool is Lévy's Characterization Theorem that characterizes a Brownian motion.

Theorem 3.3.1 (Lévy's Characterization Theorem)**.** *A stochastic process X_t is a Brownian motion if and only if there exists a probability measure Q and a filtration $\{\mathcal{F}_t\}$ such that*

1. *X_t is a continuous martingale with respect to $\{\mathcal{F}_t\}$ under Q;*
2. *$Q(X_0 = 0) = 1$;*
3. *The quadratic variation $\langle X\rangle_t$ with respect to Q of X_t on the interval $[0, t]$ is equal to t almost surely.*

As discussed in Section 3.1, we can not expect the martingale property for non-adapted case. So Condition 1 must be corrected to the near-martingale property. The following theorem shows that the similar conclusion and gives the construction of the corresponding probability measure Q.

Theorem 3.3.2 (Kuo–Peng–Szozda [23])**.** *Suppose that $B(t)$ is a Brownian motion and $\{\mathcal{F}_t\}$ is its natural filtration on a probability space $(\Omega, \mathcal{F}, P)$. Let $f(t)$ and $g(t)$ be continuous square-integrable stochastic processes such that $f(t)$ is adapted to $\{\mathcal{F}_t\}$ and $g(t)$ is adapted to a counter $\{\mathcal{F}_t\}$-filtration $\{\mathcal{G}^{(t)}\}$, such that $E[\mathcal{E}_f(t)] < \infty$ and $E[\mathcal{E}^g(t)] < \infty$ for all $t > 0$, where*

$$\mathcal{E}_f(t) = \exp\Big\{ \int_0^t f(s)\, dB(s) - \frac{1}{2} \int_0^t f^2(s)\, ds \Big\},$$

and

$$\mathcal{E}^g(t) = \exp\Big\{ - \int_t^T g(s)\, dB(s) - \frac{1}{2}\int_t^T g^2(s)\, ds\Big\}.$$

Let

$$\widetilde{B}_t = B_t + \int_0^t \big(f(s) + g(s)\big)\, ds.$$

Then $\widetilde{B}_t$ *is a near-martingale with respect to* $(\Omega, \mathcal{F}, Q)$*, where*

$$dQ = \exp\,\{-\int_0^T (f(t) + g(t))\, dB_t - \frac{1}{2}\int_0^T (f(t) + g(t))^2\, dt\}\, dP. \tag{3.3.1}$$

By the construction of $\widetilde{B}$, Condition 2 in Lévy's Characterization Theorem also holds. The following theorem gives the analogue of Condition 3.

Theorem 3.3.3 (Kuo–Peng–Szozda [23]). *Suppose that the assumptions of Theorem 3.3.2 hold. Then the Q-quadratic variation of $\widetilde{B}$ on the interval $[0, t]$ is equal to t.*

Chapter 4

Application to Stochastic Differential Equations

4.1 Stochastic Differential Equations for Exponential Processes

As shown in Example 1.4.12, the exponential processes play a central role in transform of probability measures and the Girsanov theorem. Although the general Girsanov theorem shown in Section 3.3 is proved, its proof doesn't use exponential processes but uses the classical Girsanov theorem. This means that the full understanding of the measure transform and exponential process is not clear. In this section, we give some results in the study of non-adapted exponential processes using the general Itô formula.

Recall Example 1.4.12, we have that the exponential process for $f \in \mathcal{L}_{\mathrm{ad}}(\Omega, L^2[0,1])$

$$\mathcal{E}_f(t) = \exp\Big\{ \int_0^t f(s)\, dB(s) - \frac{1}{2}\int_0^t f^2(s)\, ds \Big\}$$

satisfies the stochastic differential equation

$$\begin{cases} dX(t) = f(t)X(t)\, dB(t), \quad 0 \le t \le 1, \\ X(0) = 1. \end{cases} \tag{4.1.1}$$

However, if f is not adapted, e.g., $f(t) = B(1)$, then we can prove that the exponential process defined as above does not satisfy the stochastic differential equation (4.1.1). In fact, by Equation (2.2.11), we have

$$\exp\Big\{ \int_0^t B(1)\, dB(s) - \frac{1}{2}\int_0^t B(1)^2\, ds \Big\} = \exp\big(B(1)B(t) - t - \frac{1}{2}B(1)^2 t \big).$$

Then by applying Theorem 2.4.2 to $X_t = B(t), Y^{(t)} = B(1) - B(t)$ and $\theta(t,x,y) = \exp\big((x+y)x - t - \frac{1}{2}(x+y)^2 t\big)$, we can see that Equation (4.1.1) does not hold. Nevertheless, if we define for $B(1)$ the exponential process as

$$\mathcal{E}(t) = \exp\Big[B(1)\int_0^t e^{-(t-s)}\, dB(s) - \frac{1}{4}B(1)^2\big(1 - e^{-2t}\big) - t \Big], \tag{4.1.2}$$

then we have

Theorem 4.1.1. *The stochastic process $\mathcal{E}(t)$ defined in Equation* (4.1.2) *is the solution to the stochastic differential equation*

$$\begin{cases} d\mathcal{E}(t) = B(1)\mathcal{E}(t)\, dB(t), \quad 0 \le t \le 1, \\ \mathcal{E}(0) = 1. \end{cases} \tag{4.1.3}$$

Remark 4.1.2. This stochastic differential equation was first studied by Buckdahn [7] by a different method. The white noise version is a special case of Theorem 13.34 in the book [19].

In fact, we can define a general exponential process for both adapted and anticipative processes. Let X_t be an Itô process

$$X_t = \int_a^t g(s)\, dB(s) - \frac{1}{2}\int_a^t g(s)^2\, ds, \quad a \le t \le b, \tag{4.1.4}$$

where $g \in L^2_{\mathrm{ad}}([a, b] \times \Omega)$, and $Y^{(t)}$ an instantly independent process

$$Y^{(t)} = -\int_t^b h(s)\, dB(s) - \frac{1}{2}\int_t^b h(s)^2\, ds, \quad a \le t \le b, \tag{4.1.5}$$

where $h \in L^2_{\mathrm{ct}}([a, b] \times \Omega)$. Define the *exponential process* $\mathcal{E}(t)$ associated with g and h by

$$\mathcal{E}(t) = e^{X_t} e^{Y^{(t)}}, \quad a \le t \le b. \tag{4.1.6}$$

Apply Theorem 2.4.2 to $X_t, Y^{(t)}$ and $\theta(x, y) = e^x e^y$. Then we obtain

$$d\theta\big(X_t, Y^{(t)}\big) = \big(g(t) + h(t)\big)\theta\big(X_t, Y^{(t)}\big)\, dB(t),$$

i.e.,

$$d\mathcal{E}(t) = \big(g(t) + h(t)\big)\mathcal{E}(t)\, dB(t).$$

The above argument proves

Theorem 4.1.3 (Hwang–Kuo–Saitô–Zhai [14]). *Let $g \in L^2_{ad}([a,b]\times\Omega)$ and $h \in L^2_{ct}([a,b]\times\Omega)$. Then the exponential process*

$$\mathcal{E}(t) = \exp\Big[\int_a^t g(s)\,dB(s) - \frac{1}{2}\int_a^t g(s)^2\,ds - \int_t^b h(s)\,dB(s) - \frac{1}{2}\int_t^b h(s)^2\,ds\Big] \tag{4.1.7}$$

is the solution to the stochastic differential equation

$$d\mathcal{E}(t) = \big(g(t) + h(t)\big)\mathcal{E}(t)\,dB(t).$$

Remark 4.1.4. The exponential process (4.1.7) involves two special cases of stochastic differential equations:

1. When $h = 0$, then

   $$\mathcal{E}(t) = \exp\Big[\int_a^t g(s)\,dB(s) - \frac{1}{2}\int_a^t g(s)^2\,ds\Big]$$

   solves the forward equation

   $$\begin{cases} d\mathcal{E}(t) = g(t)\mathcal{E}(t)\,dB(t), & a \le t \le b, \\ \mathcal{E}(a) = 1. \end{cases} \tag{4.1.8}$$

2. When $g = 0$, then

   $$\mathcal{E}(t) = \exp\Big[-\int_t^b h(s)\,dB(s) - \frac{1}{2}\int_t^b h(s)^2\,ds\Big]$$

   solves the backward equation

   $$\begin{cases} d\mathcal{E}(t) = h(t)\mathcal{E}(t)\,dB(t), & a \le t \le b, \\ \mathcal{E}(b) = 1. \end{cases} \tag{4.1.9}$$

4.2 Stochastic Differential Equations with Anticipative Coefficients

In the white noise distribution theory (see [19]), the following theorem gives the solution of a stochastic integral equation with anticipative coefficients:

Theorem 4.2.1. *Let $0 \leq a < b < \infty$ and let $f \in L^\infty([a,b])$. Then the stochastic integral equation*

$$X(t) = x_0 + \int_0^t \partial_s^*(f(s)B(b)X(s))\, ds, \quad a \leq t \leq b, \ (x_0 \in \mathbb{R}),$$

has a unique solution in $L^2([a,b];(L^2))$ given by

$$X(t)$$
$$= x_0 \exp\Big[B(b)\int_a^t f(s)e^{-\int_s^t f(\tau)\,d\tau}\,dB(s) - \frac{1}{2}B(b)^2\int_a^t f(s)^2 e^{-2\int_s^t f(\tau)\,d\tau}\,ds - \int_a^t f(s)\,ds\Big]. \tag{4.2.1}$$

The stochastic integral equation can be written in stochastic differential

$$\begin{cases} dX_t = f(t)B(b)X_t\,dB_t, & t \in [a,b], \\ X_a = x_0. \end{cases}$$

In particular, if $f(t) \equiv 1, a = 0, b = 1$, and $x_0 = 1$, then we get that the solution of

$$\begin{cases} dX_t = B(1)X_t\,dB_t, & t \in [0,1], \\ X_0 = 1, \end{cases}$$

is

$$X(t) = \exp\Big[B(1)\int_0^t f(s)e^{-(t-s)}\,dB(s) - \frac{1}{4}B(1)^2\left(1 - e^{-2t}\right) - t\Big].$$

The following theorem gives the solution of the stochastic differential equation with $B(1)$ replaced by a general anticipative coefficient $\int_0^1 h(t)\,dB(t)$, for $h \in L^2([0,1])$.

Theorem 4.2.2. *Let $h \in L^2([0,1])$. Then the stochastic differential equation*

$$\begin{cases} dX_t = \big(\int_0^1 h(t)\,dB(t)\big)X_t\,dB_t, & t \in [0,1], \\ X_0 = 1, \end{cases}$$

has a unique solution in $L^2([0,1];(L^2))$ given by

$$\begin{aligned} X(t) = \exp\Big[&\int_0^1 h(t)\,dB(t)\int_0^t e^{-\left(H(t)-H(s)\right)}h(s)\,dB(s) \\ &- \frac{1}{4}\Big(\int_0^1 h(t)\,dB(t)\Big)^2\left(1 - e^{-2H(t)}\right) - H(t)\Big], \end{aligned} \tag{4.2.2}$$

where $H(t) = \int_0^t h(s)^2\,ds$.

Proof. Let

$$Y^{(t)} = \int_t^1 h(s)\,dB(s), \quad \text{and} \quad Z_t = \int_0^t h(s)e^{H(s)}\,dB(s).$$

Then

$$dY^{(t)} = -h(t)\,dB(t), \quad \text{and} \quad dZ_t = h(t)e^{H(t)}\,dB(t).$$

Moreover, $dH(t) = h(t)^2\,dt$.

Now,

$$\begin{aligned} X(t) &= \exp\left[Y^{(0)}Z_t e^{-H(t)} - \frac{1}{4}\left(Y^{(0)}\right)^2\left(1 - e^{-2H(t)}\right) - H(t)\right] \\ &= \theta\left(Y^{(0)}, Z_t, t\right), \end{aligned}$$

where $\theta(y,z,t) = \exp\left[yze^{-H(t)} - \frac{1}{4}y^2\left(1 - e^{-2H(t)}\right) - H(t)\right]$. For this θ, we have

$$\frac{\partial\theta}{\partial z} = \left(ye^{-H(t)}\right)\theta,$$

$$\frac{\partial^2\theta}{\partial z^2} = \left(ye^{-H(t)}\right)^2\theta = y^2e^{-2H(t)}\theta,$$

$$\frac{\partial^2\theta}{\partial z\partial y} = \left(ze^{-H(t)} - \frac{1}{2}y\left(1 - e^{-2H(t)}\right)\right)\left(ye^{-H(t)}\right)\theta + e^{-H(t)}\theta,$$

$$\frac{\partial\theta}{\partial t} = \left(-yze^{-H(t)}h(t)^2 - \frac{1}{2}y^2e^{-2H(t)}h(t)^2 - h(t)^2\right)\theta.$$

By applying the general Itô formula in Theorem 2.4.2 to θ above, we get

$$
\begin{aligned}
dX_t =& d\theta\left(Y^{(0)}, Z_t, t\right) \\
=& \frac{\partial\theta}{\partial z}\, dZ_t + \frac{1}{2}\frac{\partial^2\theta}{\partial z^2}(dZ_t)^2 - \frac{\partial^2\theta}{\partial z\partial y}(dZ_t)\left(dY^{(t)}\right) + \frac{\partial\theta}{\partial t}\, dt \\
=& \theta \times \Big\{ Y^{(0)}e^{-H(t)}e^{H(t)}h(t)\, dB_t + \frac{1}{2}\left(Y^{(0)}\right)^2 e^{-2H(t)}e^{2H(t)}h(t)^2\, dt \\
& \quad + \Big(Z_t e^{-H(t)}Y^{(0)}e^{-H(t)} - \frac{1}{2}\big(Y^{(0)}\big)^2 e^{-H(t)} + \frac{1}{2}\big(Y^{(0)}\big)^2 e^{-3H(t)} + e^{-H(t)} \Big) \\
& \qquad \times e^{H(t)}h(t)^2\, dt \\
& \quad + \Big(- Y^{(0)}Z_t e^{-H(t)}h(t)^2 - \frac{1}{2}\big(dY^{(0)}\big)^2 e^{-2H(t)}h(t)^2 - h(t)^2 \Big)\, dt \Big\} \\
=& \theta \times \Big\{ Y^{(0)}h(t)\, dB_t + \frac{1}{2}\big(Y^{(0)}\big)^2 h(t)^2\, dt + Z_t Y^{(0)}e^{-H(t)}h(t)^2\, dt - \frac{1}{2}\big(Y^{(0)}\big)^2 h(t)^2\, dt \\
& \quad + \frac{1}{2}\big(Y^{(0)}\big)^2 e^{-2H(t)}h(t)^2\, dt + h(t)^2\, dt - Y^{(0)}Z_t e^{-H(t)}h(t)^2\, dt \\
& \quad - \frac{1}{2}\big(Y^{(0)}\big)^2 e^{-2H(t)}h(t)^2\, dt - h(t)^2\, dt \Big\} \\
=& \theta Y^{(0)}h(t)\, dB_t \\
=& X_t Y^{(0)}h(t)\, dB_t,
\end{aligned}
$$

where, for simplicity, we omit the variables of θ and its partial derivatives. Thus, X_t satisfies the stochastic differential equation

$$
\begin{cases}
dX_t = X_t Y^{(0)}h(t)\, dB_t, \quad t \in [0,1], \\
X_0 = 1.
\end{cases}
$$

□

4.3 Stochastic Differential Equations with Anticipative Initial Condition

In [18], the authors gave the solution to stochastic differential equations with anticipative initial condition in the following theorem:

Theorem 4.3.1 (Khalifa–Kuo–Ouerdiane–Szozda [18]). *Suppose that $\alpha \in L^2([a,b])$ and $\beta \in L^2_{ad}([a,b]\times\Omega)$. Suppose also that $\rho \in \mathcal{M}^\infty \bigcap \mathcal{S}(\mathbb{R})$. Then the stochastic* *differential equation*

$$\begin{cases} dX_t = \alpha(t)X_t\,dB_t + \beta(t)X_t\,dt, & t \in [a,b], \\ X_a = \rho(B_b - B_a). \end{cases}$$

has a unique solution given by

$$X_t = [\rho(B_b - B_a) - \xi(t, B_b - B_a)]Z_t,$$

where

$$\xi(t,y) = \int_a^t \alpha(s)\rho'\left(y - \int_s^t \alpha(u)\,du\right)\,ds,$$

and

$$Z_t = \exp\left\{\int_a^t \alpha(s)\,dB_s + \int_a^t \left(\beta(s) - \frac{1}{2}\alpha(s)^2\right)\,ds\right\}.$$

However, this theorem is not clear and its proof does not show the essential structure of the stochastic differential equation. In [14], we have proved a simplified form of Theorem 4.3.1, and its proof is also intrinsic. In fact, by the general Itô formula in Theorem 2.4.2, we can give an explicit formula for the solution of a stochastic differential equation with anticipative initial condition:

Theorem 4.3.2 (Hwang–Kuo–Saitô–Zhai [14])**.** *Let $\alpha(t)$ be a deterministic function in $L^2([a,b])$, $\beta(t)$ an adapted stochastic process such that $E\int_a^b |\beta(t)|^2\,dt < \infty$, and ρ a continuous function on $\mathbb{R}$. Then the stochastic differential equation*

$$\begin{cases} dX_t = \alpha(t)X_t\,dB(t) + \beta(t)X_t\,dt, & a \le t \le b, \\ X_a = \rho\big(B(b) - B(a)\big), \end{cases} \tag{4.3.1}$$

has a unique solution given by

$$X_t = \rho\big(B(b) - B(a) - \int_a^t \alpha(s)\,ds\big)\exp\Big[\int_a^t \alpha(s)\,dB(s) + \int_a^t \big(\beta(s) - \frac{1}{2}\alpha(s)^2\big)\,ds\Big]. \tag{4.3.2}$$

This kind of equations has strong practical meaning. As we will see in Section 5.2, we will use this model to represent the stock market with inside information.

In fact, they were considered in [8] and [10] in the white noise distribution theory. However, as mentioned in Section 1.1, these results have difficulty to have probabilistic meaning, and so to apply to the analysis of real market and option pricing. The solution expression in Theorem 4.3.2 provides the explicit way of calculating and simulating of the market, and can offer an option pricing formula (see Section 5.2, 5.5 and 5.6).

4.4 Conditional Expectation of the Solution of Theorem 4.3.2

Now, let $Y_t = E[X_t|\mathcal{F}_t]$ be the conditional expectation of the solution process X_t in Equation (4.3.2). Then Y_t becomes adapted to the filtration $\{\mathcal{F}_t\}$. It is natural to conjecture that Y_t is the solution to the adapted version of the stochastic differential equation (4.3.1), namely,

$$\begin{cases} dX_t = \alpha(t)X_t\,dB_t + \beta(t)X_t\,dt, \quad t \in [a,b], \\ X_a = E\big(\rho(B_b - B_a)\big). \end{cases}$$

However, this is not the fact and we have the following result:

Theorem 4.4.1 (Kuo–Zhai [29]). *Suppose $\alpha(t)$ and $\beta(t)$ are as in Theorem 4.3.2. Moreover, assume that ρ is a smooth function whose Maclaurin series converges to itself. Suppose also that $X_1(t)$ and $X_2(t)$ are the solutions of the same linear stochastic differential equation*

$$dX_t = \alpha(t)X_t\,dB_t + \beta(t)X_t\,dt, \quad t \in [a,b],$$

with different initial conditions

$$X_1(a) = \rho(B_b - B_a) \quad \text{and} \quad X_2(a) = \rho'(B_b - B_a),$$

respectively. Let $Y_1(t) = E[X_1(t)|\mathcal{F}_t]$ and $Y_2(t) = E[X_2(t)|\mathcal{F}_t]$. Then $Y_1(t)$ satisfies the following stochastic differential equation

$$\begin{cases} dY_1(t) = \alpha(t)Y_1(t)\,dB_t + \beta(t)Y_1(t)\,dt + Y_2(t)\,dB_t, \quad t \in [a,b], \\ Y_1(a) = E\big(\rho(B_b - B_a)\big). \end{cases} \tag{4.4.1}$$

Proof. Define

$$Z_t = \exp\Big\{\int_a^t \alpha(s)\,dB(s) + \int_a^t \big(\beta(s) - \frac{1}{2}\alpha(s)^2\big)\,ds\Big\}. \tag{4.4.2}$$

By the assumption, we can write $X_1(t)$ as

$$\begin{aligned} X_1(t) &= \rho\big(B_b - B_a - \int_a^t \alpha(s)\,ds\big)Z_t \\ &= Z_t \sum_{k=0}^{\infty} \frac{1}{k!}\rho^{(k)}\big(B_t - B_a - \int_a^t \alpha(s)\,ds\big)(B_b - B_t)^k. \end{aligned}$$

Then

$$\begin{aligned} Y_1(t) &= E[X_1(t)|\mathcal{F}_t] \\ &= Z_t \sum_{k=0}^{\infty} \frac{1}{k!}\rho^{(k)}\big(B_t - B_a - \int_a^t \alpha(s)\,ds\big)E\left[(B_b - B_t)^k|\mathcal{F}_t\right] \\ &= Z_t \sum_{k=0}^{\infty} \frac{1}{(2k)!}\rho^{(2k)}\big(B_t - B_a - \int_a^t \alpha(s)\,ds\big)(b-t)^k(2k-1)!! \\ &= Z_t \sum_{k=0}^{\infty} \frac{1}{(2k)!!}\rho^{(2k)}\big(B_t - B_a - \int_a^t \alpha(s)\,ds\big)(b-t)^k \\ &= Z_t \sum_{k=0}^{\infty} \frac{1}{(2k)!!}V_t^k, \end{aligned}$$

where $V_t^k = \rho^{(2k)}\big(B_t - B_a - \int_a^t \alpha(s)\,ds\big)(b-t)^k$. Similarly, we have

$$Y_2(t) = Z_t \sum_{k=0}^{\infty} \frac{1}{(2k)!!}\rho^{(2k+1)}\big(B_t - B_a - \int_a^t \alpha(s)\,ds\big)(b-t)^k.$$

Now, both V_t^k and Z_t are adapted. In order to find $dY_1(t)$, we first find dV_t^k and dZ_t. By the definition of Z_t,

$$dZ_t = \alpha(t)Z_t\,dB_t + \beta(t)Z_t\,dt.$$

By Itô's formula (2.4.12), we have

$$\begin{aligned} dV_t^k =& \rho^{(2k+1)}\big(B_t - B_a - \int_a^t \alpha(s)\,ds\big)(b-t)^k(dB_t - \alpha(t)\,dt) \\ &+ \frac{1}{2}\rho^{(2k+2)}\big(B_t - B_a - \int_a^t \alpha(s)\,ds\big)(b-t)^k\,dt \\ &- \rho^{(2k)}\big(B_t - B_a - \int_a^t \alpha(s)\,ds\big)k(b-t)^{k-1}\,dt. \end{aligned}$$

So

$$
\begin{aligned}
&d\big(Y_1(t)^k\big)\\
&=\sum_{k=0}^{\infty}\frac{1}{(2k)!!}\left[V_t^k\,dZ_t+Z_t\,dV_t^k+(dV_t^k)(dZ_t)\right]\\
&=\sum_{k=0}^{\infty}\frac{1}{(2k)!!}\rho_{2k}(b-t)^k(\alpha(t)Z_t\,dB_t+\beta(t)Z_t\,dt)\\
&\quad+\sum_{k=0}^{\infty}\frac{1}{(2k)!!}Z_t\rho_{2k+1}(b-t)^k(dB_t-\alpha(t)\,dt)+\frac{1}{2}\sum_{k=0}^{\infty}\frac{1}{(2k)!!}Z_t\rho_{2k+2}(b-t)^k\,dt\\
&\quad-\sum_{k=0}^{\infty}\frac{1}{(2k)!!}Z_t\rho_{2k}k(b-t)^{k-1}\,dt+\sum_{k=0}^{\infty}\frac{1}{(2k)!!}\alpha(t)Z_t\rho_{2k+1}(b-t)^k\,dt\\
&=\sum_{k=0}^{\infty}\frac{1}{(2k)!!}\rho_{2k}(b-t)^k\alpha(t)Z_t\,dB_t+\sum_{k=0}^{\infty}\frac{1}{(2k)!!}\rho_{2k}(b-t)^k\beta(t)Z_t\,dt\\
&\quad+\sum_{k=0}^{\infty}\frac{1}{(2k)!!}Z_t\rho_{2k+1}(b-t)^k\,dB_t-\sum_{k=0}^{\infty}\frac{1}{(2k)!!}Z_t\rho_{2k+1}(b-t)^k\alpha(t)\,dt\\
&\quad+\frac{1}{2}\sum_{k=0}^{\infty}\frac{1}{(2k)!!}Z_t\rho_{2k+2}(b-t)^k\,dt-\frac{1}{2}\sum_{k=0}^{\infty}\frac{1}{(2(k-1))!!}Z_t\rho_{2k}(b-t)^{k-1}\,dt\\
&\quad+\sum_{k=0}^{\infty}\frac{1}{(2k)!!}\alpha(t)Z_t\rho_{2k+1}(b-t)^k\,dt\\
&=\sum_{k=0}^{\infty}\frac{1}{(2k)!!}\rho_{2k}(b-t)^k\alpha(t)Z_t\,dB_t+\sum_{k=0}^{\infty}\frac{1}{(2k)!!}\rho_{2k}(b-t)^k\beta(t)Z_t\,dt\\
&\quad+\sum_{k=0}^{\infty}\frac{1}{(2k)!!}Z_t\rho_{2k+1}(b-t)^k\,dB_t\\
&=\alpha(t)Y_1(t)\,dB_t+\beta(t)Y_1(t)\,dt+Y_2(t)\,dB_t,
\end{aligned}
$$

where, for simplicity, ρ_n denotes $\rho^{(n)}\big(B_t-B_a-\int_a^t\alpha(s)\,ds\big)$, and the forth equality is by the cancelation of the last four terms. $\square$

Remark 4.4.2. If we consider what we have seen in the near-martingale property that a general stochastic integral satisfies, the appearance of the extra term $Y_2(t)\,dB_t$ in the stochastic differential equation (4.4.1) is not unexpected. Due to the existence of the anticipative part, the stochastic differential equation satisfied by the conditional expectation $Y_1(t)$ of $X_1(t)$ is not just simply the solution of the stochastic differential equation (4.3.2). Moreover, the derivative ρ' appearing in the initial condition satisfied by $X_2(t)$ is due to the extra term in the Itô isometry for general stochastic integral (see Theorem 2.3.2), which essentially, comes from the definition of general stochastic integration and its Itô formula.

A case of most interest of ρ in Theorem 4.4.1 is when it is a Hermite polynomial. Recall that the Hermite polynomial of degree n with parameter η is defined by (see, e.g., [20])

$$H_n(x;\eta) = (-\eta)^n e^{x^2/2\eta} D_x^n e^{-x^2/2\eta}, \tag{4.4.3}$$

where D_x is the differential operator with respect to the variable x. Hermite polynomial H_n have the following properties:

$$D_x H_n(x;\eta) = nH_{n-1}(x;\eta), \tag{4.4.4}$$

$$D_\eta H_n(x;\eta) = -\frac{1}{2} D_x^2 H_n(x;\eta), \tag{4.4.5}$$

$$H_n(x+y;\eta) = \sum_{k=0}^{n} \binom{n}{k} H_{n-k}(x;\eta) y^k. \tag{4.4.6}$$

Lemma 4.4.3 ([20]). *The stochastic process $X_t = H_n(B(t)-B(a);t-a), a \le t \le b$ is a martingale.*

Proof. Apply Itô's formula to X_t. Then we obtain

$$\begin{aligned}
dX_t =& dH_n(B(t)-B(a);t-a) \\
=& D_x H_n(B(t)-B(a);t-a)\, dB(t) + \frac{1}{2} D_x^2 H_n(B(t)-B(a);t-a)\, dt \\
&+ D_\eta H_n(B(t)-B(a);t-a)\, dt \\
=& nH_{n-1}(B(t)-B(a);t-a)\, dB(t) + \frac{1}{2} D_x^2 H_n(B(t)-B(a);t-a)\, dt \\
&- \frac{1}{2} D_x^2 H_n(B(t)-B(a);t-a)\, dt \\
=& nH_{n-1}(B(t)-B(a);t-a)\, dB(t),
\end{aligned} \tag{4.4.7}$$

where we used Equation (4.4.4) and (4.4.5). So

$$X_t = H_n(B(t) - B(a); t - a) = \int_a^t nH_{n-1}(B(s) - B(a); s - a)\, dB(s),$$

which is an Itô integral of the adapted process $nH_{n-1}(B(t) - B(a); t - a)$, and thus is a martingale. $\square$

For Hermite Polynomial $\rho(x) = H_n(x; b - a)$, we have

Theorem 4.4.4 (Hwang–Kuo–Saitô–Zhai [15])**.** *Let $\alpha(t)$ be a deterministic function in $L^2([a, b])$, $\beta(t)$ an adapted stochastic process such that $E \int_a^b |\beta(t)|^2\, dt < \infty$, and n a fixed natural number. Let X_t be the solution of the stochastic differential equation*

$$\begin{cases} dX_t = \alpha(t)X_t\, dB(t) + \beta(t)X_t\, dt, \quad a \le t \le b, \\ X_a = H_n\big(B(b) - B(a); b - a\big), \end{cases} \tag{4.4.8}$$

and $Y_t = E[X_t|\mathcal{F}_t]$. Then we have

$$Y_t = H_n\Big(B(t) - B(a) - \int_a^t \alpha(s)\, ds; t - a\Big) Z_t, \quad a \le t \le b, \tag{4.4.9}$$

where Z_t is defined in Equation (4.4.2). Moreover, Y_t satisfies the stochastic differential equation

$$dY_t = \Big[\alpha(t)Y_t + \Big(D_x H_n\Big(B(b) - B(a) - \int_a^t \alpha(s)\, ds; t - a\Big)\Big) Z_t\Big]\, dB(t) + \beta(t)Y_t\, dt, \tag{4.4.10}$$

or equivalently

$$dY_t = \Big[\alpha(t)Y_t + nH_{n-1}\Big(B(t) - B(a) - \int_a^t \alpha(s)\, ds; t - a\Big) Z_t\Big]\, dB(t) + \beta(t)Y_t\, dt, \tag{4.4.10'}$$

for $a \le t \le b$, with initial condition $Y_a = 0$.

Proof. First, we prove Equation (4.4.9). By Theorem 4.3.2, we have for $a \le t \le b$,

$$X_t = H_n\Big(B(b) - B(a) - \int_a^t \alpha(s)\, ds; b - a\Big) Z_t.$$

Note that Z_t is adapted to $\{\mathcal{F}_t\}$. So

$$Y_t = E[X_t|\mathcal{F}_t] = E\big[H_n\big(B(b) - B(a) - \int_a^t \alpha(s)\,ds; b-a\big)\big|\mathcal{F}_t\big]Z_t. \tag{4.4.11}$$

By Equation (4.4.6) with $x = B(b) - B(a), y = -\int_a^t \alpha(s)\,ds$, and $\eta = b - a$, we get

$$H_n\big(B(b)-B(a)-\int_a^t \alpha(s)\,ds; b-a\big)=\sum_{k=0}^{n}\binom{n}{k}H_{n-k}(B(b)-B(a); b-a)\big(-\int_a^t \alpha(s)\,ds\big)^k.$$

Taking conditional expectation with respect to $\mathcal{F}_t$, we obtain

$$\begin{aligned}
&E\big[H_n\big(B(b) - B(a) - \int_a^t \alpha(s)\,ds; b-a\big)\big|\mathcal{F}_t\big] \\
&= \sum_{k=0}^{n}\binom{n}{k}E\big[H_{n-k}(B(b) - B(a); b-a)\big(-\int_a^t \alpha(s)\,ds\big)^k\big|\mathcal{F}_t\big] \\
&= \sum_{k=0}^{n}\binom{n}{k}E\big[H_{n-k}(B(b) - B(a); b-a)\big|\mathcal{F}_t\big]\big(-\int_a^t \alpha(s)\,ds\big)^k \\
&= \sum_{k=0}^{n}\binom{n}{k}H_{n-k}(B(t) - B(a); t-a)\big(-\int_a^t \alpha(s)\,ds\big)^k \\
&= H_n\big(B(t) - B(a) - \int_a^t \alpha(s)\,ds; t-a\big),
\end{aligned} \tag{4.4.12}$$

since $\int_a^t \alpha(s)\,ds$ is adapted to $\{\mathcal{F}_t\}$ and we have used Lemma 4.4.3. Combining Equation (4.4.11) and (4.4.12), we get Equation (4.4.9)

$$Y_t = E[X_t|\mathcal{F}_t] = H_n\big(B(t) - B(a) - \int_a^t \alpha(s)\,ds; t-a\big)Z_t, \quad a \le t \le b.$$

By Itô's formula, we have

$$\begin{aligned}
&dH_n\big(B(t) - B(a) - \int_a^t \alpha(s)\,ds; t-a\big) \\
&= D_xH_n\,dB(t) - D_xH_n\alpha(t)\,dt + \frac{1}{2}D_x^2H_n\,dt + D_\eta H_n\,dt \\
&= nH_{n-1}\,dB(t) - nH_{n-1}\alpha(t)\,dt + \frac{1}{2}D_x^2H_n\,dt - \frac{1}{2}D_x^2H_n\,dt \\
&= nH_{n-1}(B(t) - B(a) - \int_a^t \alpha(s)\,ds; t-a)\,dB(t) \\
&\quad - nH_{n-1}(B(t) - B(a) - \int_a^t \alpha(s)\,ds; t-a)\alpha(t)\,dt,
\end{aligned}$$

where we used Equation (4.4.7), and

$$dZ_t = \alpha(t) Z_t \, dB_t + \beta(t) Z_t \, dt,$$

and thus

$$\begin{aligned}
dY_t =& H_n \, dZ_t + Z_t \, dH_n + (dH_n)(dZ_t) \\
=& H_n \alpha(t) Z_t \, dB_t + H_n \beta(t) Z_t \, dt + n H_{n-1} \alpha(t) Z_t \, dB_t \\
=& \Big[\alpha(t) Y_t + n H_{n-1} \big(B(t) - B(a) - \int_a^t \alpha(s) \, ds; t-a \big) Z_t \Big] \, dB(t) + \beta(t) Y_t \, dt,
\end{aligned}$$

which proves Equation (4.4.10) and (4.4.10'). $\square$

Remark 4.4.5. By using Theorem 4.4.1, we can directly get Equation (4.4.10) and (4.4.10'). In fact, if we denote explicitly the fixed natural number n in Theorem 4.4.4 in X_t and Y_t as $X_t^{[n]}$ and $Y_t^{[n]}$, respectively. Note that for $\rho(x) = H_n(x; b-a)$, by Equation (4.4.4), $\rho'(x) = n H_{n-1}(x; b-a)$. Applying Theorem 4.4.1 to $X_1 = X_t^{[n]}$ and $X_2 = n X_t^{[n-1]}$ (and then $Y_1 = Y_t^{[n]}$ and $Y_2 = n Y_t^{[n-1]}$), we get

$$dY_t^{[n]} = (\alpha(t) Y_t^{[n]} + n Y_t^{[n-1]}) \, dB(t) + \beta(t) Y_t^{[n]} \, dt,$$

with $Y_a^{[n]} = 0$, which reaches the same conclusion of Theorem 4.4.4.

Remark 4.4.6. The spirit of Theorem 4.4.4 is that if ρ is a Hermite polynomial, then we can express the conditional expectation process $Y_t = E[X(t)|\mathcal{F}_t]$ in ρ. This observation is not reflected in Theorem 4.4.1, and leads to the following discussion.

It is well-known that the Hermite polynomials form an orthonormal basis of the Hilbert space $L^2(\mathbb{R}, w(x) \, dx)$, and thus a basis the Sobolev space $H^1(\mathbb{R}, w(x) \, dx)$, where the measure $w(x) \, dx$ is determined by the Gaussian weight function $w(x) = \frac{1}{\sqrt{2\pi\rho}} e^{-x^2/2\rho}$. So we can extend Theorem 4.4.4 from Hermite polynomials to functions in $H^1(\mathbb{R}, w(x) \, dx)$.

Theorem 4.4.7 (Kuo–Zhai [29]). *Suppose $\alpha(t)$ and $\beta(t)$ are as in Theorem 4.3.2, and $\rho \in H^1(\mathbb{R}, w(x)\,dx)$ that can be written as a Hermite series*

$$\rho(x) = \sum_{i=0}^{\infty} \alpha_n H_n(x; b-a).$$

Suppose also that $X_1(t)$ and $X_2(t)$ are the solutions of the same linear stochastic differential equation

$$dX_t = \alpha(t) X_t \, dB_t + \beta(t) X_t \, dt, \quad t \in [a, b],$$

with different initial conditions

$$X_1(a) = \rho(B_b - B_a) \quad \textit{and} \quad X_2(a) = \rho'(B_b - B_a),$$

respectively. Let $Y_1(t) = E[X_1(t)|\mathcal{F}_t]$ and $Y_2(t) = E[X_2(t)|\mathcal{F}_t]$. Then

$$Y_1(t) = \sum_{i=0}^{\infty} \alpha_n H_n(B(t) - B(a) - \int_a^t \alpha(s)\, ds; t-a) Z_t, \quad a \le t \le b,$$

and it satisfies the following stochastic differential equation

$$\begin{cases} dY_1(t) = \alpha(t) Y_1(t)\, dB_t + \beta(t) Y_1(t)\, dt + Y_2(t)\, dB_t, \quad t \in [a, b], \\ Y_1(a) = E\big(\rho(B_b - B_a)\big). \end{cases} \tag{4.4.13}$$

Remark 4.4.8. The proofs of Theorem 4.4.1 and 4.4.7 use Maclaurin expansion and Hermite expansion, respectively. It is natural to conjecture that these theorems can be extended to more function spaces for ρ.

Chapter 5

A Black–Scholes Model with Anticipative Initial Conditions

5.1 The Classical Black–Scholes Model

In mathematical finance, stochastic modelling has become a standard tool in pricing analysis. The Black–Scholes model introduced by Black and Scholes [5] and Merton [31] is widely used in market analysis and investment strategy making. The classical Black–Scholes model assumes adaptedness of the market, i.e., an adapted stochastic process $X_t = (X_t^{(0)}, X_t^{(1)})$ such that

$$\begin{cases} \text{(risk-free)} & dX_t^{(0)} = \gamma(t) X_t^{(0)}\, dt, & X_0^{(0)} = 1, \\ \text{(risky)} & dX_t^{(1)} = \alpha(t) X_t^{(1)}\, dB(t) + J(t) X_t^{(1)}\, dt, & X_0^{(1)} = 1. \end{cases} \tag{5.1.1}$$

This means that the model assumes that the investors only know the current information about the market based on which they make pricing decisions. However, in reality, some individuals in the market know some inside information (about the future), so that they can use it to make a better pricing decision. The inside information makes the market anticipative (to them). In this situation, we assume that the initial conditions of the market depend on the future, i.e., the dynamics of the system still go on as usual (adapted), but the decision maker grasps the inside information and uses it in initial invested wealth. So we consider the generalized Black–Scholes model with $X_0^{(0)} = q_0(B(T))$ and $X_0^{(1)} = q_1(B(T))$. This model has two useful aspects in realistic application:

- The first application is in the *initial public offering* (IPO). In IPO pricing, it is inevitable for someone to have inside information in the investment. If the understanding of the influence of inside information on the market is fully analysed, then the pricing strategy for IPO can help lower the risk for the investor, money lender and company so that it would attract more investor into the game. Also, it helps the strategy maker to give a clearer prediction or more credible promise of the profit by putting the future information into the initial allocation design.

- The second application is for the government or regulatory organizations to restrict the abuse of inside trading. For example, in the quiet periods in an IPO's history, insiders and the company are restricted in discussing, analysing or trading the offering. However, even the quiet periods are set up to prevent inside trading, these regulatory organizations can not fully guarantee that the inside information is not abused by the individuals involved, directly or indirectly. So it is of high importance for the regulatory organizations to have a deep understanding of the influence of inside information on the market and investment. This way, these organizations can provide an effective, open and fair scheme for IPO. Moreover, the scheme can be designed case by case to avoid rigidity. For instance, the quiet periods can be too long or too short for different cases so that it can harm the investment behaviors.

We will give a full analysis of the Black–Scholes market with anticipative initial condition in this chapter. We will talk about the arbitrage-free and completeness of this general Black–Scholes market in Section 5.3 and 5.4. The option pricing formula for this anticipative market will be given in Section 5.5, and the corresponding hedging portfolio will be given in Section 5.6.

Notice that the solution to this new anticipative market (see (5.2.2) and (5.2.3)) is no longer adapted, the classical framework of the classical Itô stochastic integration is not enough. We need a general stochastic integral and its properties for pricing analysis.

5.2 A Black–Scholes Model with Anticipative Initial Conditions

Consider the *market* determined by the following Black–Scholes model with anticipative initial conditions:

$$\begin{cases} dX_t^{(0)} = \gamma(t)X_t^{(0)}\,dt, & X_0^{(0)} = q_0(B(T)), \\ dX_t^{(1)} = \alpha(t)X_t^{(1)}\,dB(t) + J(t)X_t^{(1)}\,dt, & X_0^{(1)} = q_1(B(T)), \end{cases} \tag{5.2.1}$$

where q_0 and q_1 are continuously differentiable functions, and $\gamma(t), \alpha(t), J(t)$ are $\mathcal{F}_t$-adapted. $X_t^{(0)}$ refers to the unit price of the safe investment and $X_t^{(1)}$ refers to the unit price of the risky investment. Then by Theorem 4.3.2, the solutions to the above stochastic differential equations with anticipative initial conditions are

$$X_t^{(0)} = q_1\big(B(T)\big)e^{\int_0^t \gamma(s)\,ds}, \tag{5.2.2}$$

and

$$X_t^{(1)} = q_1\Big(B(T) - \int_0^t \alpha(s)\,ds\Big)\exp\Big\{\int_0^t \alpha(s)\,dB(s) + \int_0^t \Big(J(s) - \frac{1}{2}\alpha(s)^2\Big)\,ds\Big\}. \tag{5.2.3}$$

Let

$$h(t) = \frac{\gamma(t) - J(t)}{\alpha(t)}.$$

Define the normalized market

$$\widetilde{X}_t^{(0)} = q_0(B(T)),$$

and

$$\widetilde{X}_t^{(1)} = X_t^{(1)}\xi(t),$$

where

$$\xi(t) = \exp\Big\{-\int_0^t \gamma(s)\,ds\Big\}$$

is the normalization factor.

A *portfolio* in the market X_t is a stochastic process

$$p(t) = \big(p_0(t), p_1(t)\big),$$

where $p_i(t), i = 1, 2$, represents the number of units of the ith investment at time t. The *value* of a portfolio $p(t)$ is defined by

$$V_p(t) = p(t) \cdot X_t = p_0(t)X_t^{(0)} + p_1(t)X_t^{(1)}.$$

A portfolio $p(t)$ is *self-financing* if its value $V_p(t)$ satisfies the following property

$$V_p(t) = V_p(0) + \int_0^t p(s) \cdot dX_s,$$

or equivalently in stochastic differential

$$dV_p(t) = p(t) \cdot dX_t.$$

The self-financing property means that we can rearrange the assets, but there is no external new investment or withdrawal of money. In order to buy a new asset, we must sell an old one; meanwhile, if we sell an asset, we must use the money to buy a new one from the same market.

A self-financing portfolio is called *admissible* if its value $V_p(t)$ is lower bounded for almost all $(t, \omega) \in [0, T] \times \Omega$, i.e., there is a constant $C > 0$ such that

$$V_p(t, \omega) \geq -C, \quad \text{for almost all } (t, \omega) \in [0, T] \times \Omega.$$

This means that the creditors have their debt limit.

5.3 Arbitrage-Free Property

The first important property for a market is arbitrage-free property.

Definition 5.3.1. An admissible portfolio $p(t)$ is called an *arbitrage* in a market $X_t, 0 \leq t \leq T$, if its value $V_p(t)$ satisfies the conditions

$$V_p(0) = 0, \quad V_p(T) \geq 0, \quad P\big\{V_p(T) > 0\big\} > 0.$$

In Theorem 5.3.3, we will give a sufficient condition for a general Black–Scholes market to be arbitrage-free. We will use the following auxiliary lemma to facilitate our proofs of the theorems. Its meaning is that we can always consider a normalized market, i.e., the safe investment does not change and the risky investment changes (decreases) proportionally with the rate $\xi(t)$.

Lemma 5.3.2. *A portfolio $p(t)$ is an arbitrage in $X(t)$ if and only if it is an arbitrage in the normalized market $\widetilde{X}(t)$.*

Proof. If $p(t)$ is an arbitrage in $X(t)$, then

$$V_p(0) = 0, \quad V_p(T) \geq 0, \quad P\big\{V_p(T) > 0\big\} > 0.$$

So

$$\begin{aligned} \widetilde{V}_p(0) &= \xi(0)V_p(0) = 0, \\ \widetilde{V}_p(T) &= \xi(T)V_p(T) \geq 0, \\ P\big\{\widetilde{V}_p(T) > 0\big\} &= P\big\{V_p(T) > 0\big\} > 0. \end{aligned}$$

The other direction is similar. □

Theorem 5.3.3 (Arbitrage-Free Property)**.** *Assume the condition*

$$E \exp\left[\frac{1}{2}\int_0^T \Big(\frac{\gamma(t) - J(t)}{\alpha(t)}\Big)^2 dt\right] < \infty.$$

Then the market $X(t)$, $0 \leq t \leq T$, determined by (5.2.1) *has no arbitrage.*

Proof. Let

$$B_h(t) = B(t) - \int_0^t h(s)\, ds.$$

By the Girsanov theorem 1.5.1 and the $\mathcal{F}_t$-adaptedness of $h(t)$, $B_h(t)$ is a Brownian motion with respect to the probability measure Q determined by

$$dQ = \exp\Big\{\int_0^T h(t)\, dB(t) - \frac{1}{2}\int_0^T h(t)^2\, dt\Big\}\, dP.$$

Moreover,

$$\begin{aligned}dX_t^{(1)} &= \alpha(t)X_t^{(1)}\,dB(t) + J(t)X_t^{(1)}\,dt\\ &= \alpha(t)X_t^{(1)}\big(dB_h(t) + h(t)\,dt\big) + J(t)X_t^{(1)}\,dt\\ &= \alpha(t)X_t^{(1)}\,dB_h(t) + \gamma(t)X_t^{(1)}\,dt.\end{aligned}$$

By Itô's lemma, we get

$$\begin{aligned}d\widetilde{X}_t^{(1)} &= X_t^{(1)}\,d\xi(t) + \xi(t)\,dX_t^{(1)}\\ &= -X_t^{(1)}\gamma(t)\xi(t)\,dt + \xi(t)\big(\alpha(t)X_t^{(1)}\,dB_h(t) + \gamma(t)X_t^{(1)}\,dt\big)\\ &= \xi(t)\alpha(t)X_t^{(1)}\,dB_h(t)\\ &= \alpha(t)\widetilde{X}_t^{(1)}\,dB_h(t).\end{aligned}$$

Suppose $p(t) = (p_0(t), p_1(t))$ is an arbitrage in $X(t)$, so is in $\widetilde{X}(t)$. By the general Itô formula, we have $dq_0(B(T)) = 0$. Then

$$\begin{aligned}\widetilde{V}_p(t) &= \int_0^t p(s)\cdot\,d\widetilde{X}_s\\ &= \int_0^t p_0(t)\,dq_0(B(T)) + \int_0^t p_1(t)\,d\widetilde{X}_t^{(1)}\\ &= \int_0^t p_1(t)\alpha(t)\widetilde{X}_t^{(1)}\,dB_h(t)\end{aligned}$$

Thus, by Theorem 3.1.4, $\widetilde{V}_p(t)$ is a near-martingale with respect to Q. So

$$\begin{aligned}E_Q\big(\widetilde{V}_p(T)\big) &= E_Q\big(E_Q\{\widetilde{V}_p(T)\big|\mathcal{F}_0\}\big)\\ &= E_Q\big(E_Q\{\widetilde{V}_p(0)\big|\mathcal{F}_0\}\big)\\ &= E_Q\big(\widetilde{V}_p(0)\big) = 0.\end{aligned}$$

On the other hand, P and Q are equivalent. Then the conditions for $p(t)$ to be an arbitrage are equivalent to

$$V_p(0) = 0, \quad V_p(T) \geq 0,\ Q\text{-a.s.}, \quad \text{and } Q\big\{V_p(T) > 0\big\} > 0.$$

So $E_Q\big(\widetilde{V}_p(T)\big) > 0$, which is a contradiction. Therefore, there is no arbitrage in the market $X(t)$. □

5.4 Completeness of the Market

A lower bounded $\mathcal{F}_T$-measurable random variable Φ is called a *T-claim*. A T-claim is called *attainable* in $X_t, 0 \le t \le T$, if there exist a real number r and an admissible portfolio $p(t)$ such that

$$\Phi = V_p(T) = r + \int_0^T p(t) \cdot dX_t.$$

This portfolio is called a *hedging portfolio* for Φ.

Definition 5.4.1. A market is said to be *complete* if every bounded T-claim is attainable.

The completeness of a market means one can always find a hedging portfolio to realize every possible claim. The following theorem gives a sufficient condition for a general Black–Scholes market to be complete.

Theorem 5.4.2 (Completeness of a Market)**.** *Let X_t, $0 \le t \le T$, be a Black–Scholes model satisfying the assumption in Theorem 5.3.3. In addition, assume that*

$$\sigma\big\{B(s)|0 \le s \le t\big\} = \sigma\big\{B_h(s)|0 \le s \le t\big\}, \quad \forall 0 \le t \le T.$$

Then X_t is complete.

Proof. For a bounded T-claim Φ, we need to find a portfolio $p(t)$, such that,

$$\Phi = r + \int_0^T p(t) \cdot \, dX(t), \tag{5.4.1}$$

or

$$\xi(T)\Phi = \widetilde{V}_p(T) = r + \int_0^T d\widetilde{V}_p(t). \tag{5.4.1'}$$

But

$$
\begin{aligned}
d\widetilde{V}_p(t) &= p_0(t)\,dq_0(B(T)) + p_1(t)\,d\widetilde{(X)}_t^{(1)}\\
&= p_1(t)\,d\big(X_t^{(1)}\xi(t)\big)\\
&= p_1(t)\big(X_t^{(1)}\,d\xi(t) + \xi(t)\,dX_t^{(1)}\big)\\
&= p_1(t)X_t^{(1)}\xi(t)\gamma(t)\,dt + p_1(t)\xi(t)\big(\alpha(t)X_t^{(1)}\,dB(t) + J(t)X_t^{(1)}\,dt\big)\\
&= p_1(t)X_t^{(1)}\xi(t)\gamma(t)\,dt + p_1(t)\xi(t)\Big(\alpha(t)X_t^{(1)}\big(dB_h(t) + h(t)\,dt\big) + J(t)X_t^{(1)}\,dt\Big)\\
&= p_1(t)X_t^{(1)}\xi(t)\big(\gamma(t)\,dt + \alpha(t)\,dB_h(t) + \alpha(t)h(t)\,dt + J(t)\,dt\big)\\
&= p_1(t)X_t^{(1)}\xi(t)\alpha(t)\,dB_h(t).
\end{aligned}
$$

So

$$
V_p(t) = r + \int_0^t \alpha(t)p_1(t)X_t^{(1)}\,dB_h(t) \tag{5.4.2}
$$

is equivalent to

$$
\widetilde{V}_p(t) = r + \int_0^t \alpha(t)p_1(t)\widetilde{X}_t^{(1)}\,dB_h(t), \tag{5.4.2'}
$$

and $\widetilde{V}_p(t) = r + \int_0^t \alpha(t)p_1(t)\xi(t)X_t^{(1)}\,dB_h(t)$ is a near-martingale with respect to P and Q by Theorem 3.1.4. Then (5.4.1') is equivalent to

$$
\xi(T)\Phi = r + \int_0^T \alpha(t)p_1(t)\xi(t)X_t^{(1)}\,dB_h(t).
$$

On the other hand, $\xi(T)\Phi$ is bounded and $\mathcal{F}_T^{B_h}$-measurable. By the martingale representation theorem, there is a process $\theta \in L^2_{\mathrm{ad}}\big([0,T]\times\Omega\big)$, such that,

$$
\xi(T)\Phi = E_Q[\xi(T)\Phi] + \int_0^T \theta(t)\,dB_h(t).
$$

Let $r = E_Q[\xi(T)\Phi]$ and

$$
p_1(t) = \big(\alpha(t)\xi(t)X_t^{(1)}\big)^{-1}\theta(t),
$$

and

$$
p_0(t) = q_0(B(T))^{-1}\Big(E_Q[\xi(T)\Phi] + \xi(t)A(t) + \int_0^t A(s)\gamma(s)\xi(s)\,ds\Big),
$$

where

$$A(t) = \int_0^t p_1(t)\, dX_t^{(1)} - p_1(t) X_t^{(1)}.$$

Then $p(t) = \big(p_0(t), p_1(t)\big)$ is a hedging portfolio for Φ. Therefore, by the arbitrariness of Φ, X_t is a complete market. □

5.5 Option Pricing Formula

Let Φ be a T-claim. An (*European*) *option* on Φ is a guarantee to be paid the amount $\Phi(\omega)$ at time $t = T > 0$. Then

1. Buyer's price of the T-claim Φ is

$$\mathcal{P}_b(\Phi) = \sup\Big\{x : \exists p(t) \text{ such that } -x + \int_0^T p(t) \cdot dX_t + \Phi \geq 0, \text{a.s.}\Big\}. \quad (5.5.1)$$

2. Seller's price of Φ is

$$\mathcal{P}_s(\Phi) = \inf\Big\{y : \exists p(t) \text{ such that } y + \int_0^T p(t) \cdot dX_t \geq \Phi, \text{a.s.}\Big\}. \quad (5.5.2)$$

Remark 5.5.1. If the infimum in (5.5.2) does not exist, we assume $\mathcal{P}_s(\Phi) = \infty$.

Property 5.5.2. *For the prices $\mathcal{P}_b$ and $\mathcal{P}_s$ defined above, we have*

1. *essinf $\Phi \leq \mathcal{P}_b(\Phi)$.*

2. *If X_t satisfies the assumptions in Theorem 5.3.3, then we have*

$$\mathcal{P}_b(\Phi) \leq E_Q\big[\xi(T)\Phi(\omega)\big] \leq \mathcal{P}_s(\Phi). \quad (5.5.3)$$

Proof. Property 1 is trivial by taking $p = 0$. We prove Property 2 in the following.

Suppose $x \in \big\{x : \exists p(t) \text{ such that } - x + \int_0^T p(t) \cdot dX_t + \Phi \geq 0, \text{a.s.}\big\}$. Then there exists a portfolio $p(t)$ such that $-x + \int_0^T p(t) \cdot dX_t \geq -\Phi$, a.s.. So by (5.4.2), we have

$$-x + \int_0^T p_1(t)\alpha(t) X_t^{(1)} \xi(t)\, dB_h(t) \geq -\xi(T)\Phi, \quad \text{a.s.}. \quad (5.5.4)$$

Since $\int_0^t p_1(s)\alpha(s)X_s^{(1)}\xi(s)\,dB_h(s)$ is a near-martingale with respect to P and Q by Theorem 3.1.4, we have that for all $t \in [0,T]$,

$$\begin{aligned}
&E_Q\Big[\int_0^t p_1(s)\alpha(s)X_s^{(1)}\xi(s)\,dB_h(s)\Big] \\
&= E_Q\Big[E_Q\Big[\int_0^t p_1(s)\alpha(s)X_s^{(1)}\xi(s)\,dB_h(s)\Big|\mathcal{F}_0\Big]\Big] \\
&= E_Q\Big[E_Q\Big[\int_0^0 p_1(s)\alpha(s)X_s^{(1)}\xi(s)\,dB_h(s)\Big|\mathcal{F}_0\Big]\Big] \\
&= E_Q\Big[\int_0^0 p_1(s)\alpha(s)X_s^{(1)}\xi(s)\,dB_h(s)\Big] \\
&= E_Q[0] = 0.
\end{aligned}$$

Thus, taking expectation with respect to Q on (5.5.4), we get

$$x \le E_Q\big(\xi(T)\Phi\big).$$

Therefore, by the arbitrariness of x,

$$\mathcal{P}_b(\Phi) \le E_Q\big(\xi(T)\Phi\big).$$

Similarly, we assume $\mathcal{P}_s(\Phi) < \infty$. Then there exists y, such that for some $p(t)$,

$$y + \int_0^T p(t) \cdot dX_t \ge \Phi, \quad \text{a.s.}.$$

Then by (5.4.2),

$$y + \int_0^T p_1(t)\alpha(t)X_t^{(1)}\xi(t)\,dB_h(t) \ge \xi(T)\Phi, \quad \text{a.s.}.$$

Taking expectation with respect to Q, we get

$$y \ge E_Q\big(\xi(T)\Phi\big).$$

By the arbitrariness of y,

$$\mathcal{P}_s(\Phi) \ge E_Q\big(\xi(T)\Phi\big).$$

We then complete the proof. □

Definition 5.5.3. The price of a T-claim is said to *exist* if $\mathcal{P}_b(\Phi) = \mathcal{P}_s(\Phi)$. Their common value is called the *price* of the option on Φ at time $t = 0$, and is denoted by $\mathcal{P}(\Phi)$.

Now, we can give the the pricing formula for the option on any claim Φ.

Theorem 5.5.4 (Option Pricing Formula)**.** *Let X_t be a market satisfying the assumptions in Theorem 5.3.3 and 5.4.2. Then for any T-claim Φ with $E_Q\big[\xi(T)\Phi\big] < \infty$, the price of Φ at time $t = 0$ is given by*

$$\mathcal{P}(\Phi) = E_Q\big[\xi(T)\Phi\big]. \tag{5.5.5}$$

Proof. Since X_t satisfies the assumptions in Theorem 5.3.3 and 5.4.2, it is complete. Define

$$\Phi_n(\omega) = \begin{cases} \Phi(\omega), & \text{if } \Phi(\omega) \leq n, \\ n, & \text{if } \Phi(\omega) > n. \end{cases}$$

Then Φ_n is a bounded T-claim, since Φ is lower bounded.

Since X_t is complete, there is $y_n \in \mathbb{R}$ such that for some portfolio $p_n(t) = \big(p_n^{(0)}(t), p_n^{(1)}(t)\big)$, we have

$$y_n + \int_0^T p_n(t) \cdot dX_t = \Phi_n,$$

or equivalently,

$$y_n + \int_0^T p_n^{(1)}(t)\xi(t)\alpha(t)X_t^{(1)}\, dB_h(t) = \xi(T)\Phi_n.$$

By the completeness of X_t, Φ_n is attainable, so the integral in the above equation is a Q-near-martingale, so

$$y_n = E_Q\big[\xi(T)\Phi_n\big].$$

By the definition of $\mathcal{P}_s$, we have

$$\mathcal{P}_s(\Phi_n) \leq y_n,$$

and since $\Phi \geq \Phi_n$,

$$\mathcal{P}_s(\Phi_n) \geq \mathcal{P}_s(\Phi).$$

Thus,

$$\mathcal{P}_s(\Phi) \leq E_Q\big[\xi(T)\Phi\big].$$

Combining with the inequality (5.5.3) in Property (3), we obtain

$$\mathcal{P}_s(\Phi) = E_Q\big[\xi(T)\Phi\big].$$

Similarly, we have the other equality

$$\mathcal{P}_b(\Phi) = E_Q\big[\xi(T)\Phi\big].$$

Therefore, we have the price of the option

$$\mathcal{P}(\Phi) = E_Q\big[\xi(T)\Phi\big].$$

. $\square$

Now, we are going to find out the formula to calculate the option price. In order for the price to have a explicit formula, we need to assume that $\gamma(t)$ and $\alpha(t)$ are deterministic functions, and Φ is of the form

$$\Phi = F\big(X_T^{(1)}\big),$$

which means that Φ is not only $\mathcal{F}_T$-measurable, but also totally determined by the knowledge of the market at time T.

Theorem 5.5.5 (Calculation of the Option Price)**.** *The explicit formula of the price* (5.5.5) *is given by*

$$\begin{aligned}
\mathcal{P}(\Phi) =& E_Q\big[\xi(T)F\big(X_T^{(1)}\big)\big] \\
&= \exp\Big(- \int_0^T \gamma(t)\, dt\Big) \\
&\quad \times E_Q\Big[F\Big(q_1\Big(B(T) - \int_0^T \alpha(t)\, dt\Big) \exp\Big\{ \int_0^T \alpha(t)\, dB(t) + \int_0^T J(t) - \frac{1}{2}\alpha(t)^2\, dt\Big\}\Big)\Big].
\end{aligned}$$

On the other hand,

$$\begin{aligned}\int_0^T \alpha(t)\,dB(t)+\int_0^T J(t)-\frac{1}{2}\alpha(t)^2\,dt &= \int_0^T \alpha(t)\big(dB_h(t)+h(t)\,dt\big)+\int_0^T J(t)-\frac{1}{2}\alpha(t)^2\,dt\\ &= \int_0^T \alpha(t)\,dB_h(t)+\int_0^T \gamma(t)-\frac{1}{2}\alpha(t)^2\,dt.\end{aligned}$$

So

$$\begin{aligned}&\mathcal{P}(\Phi)\\ &= \exp\Big(-\int_0^T \gamma(t)\,dt\Big)\\ &\quad\times E_Q\Big[F\Big(q_1\big(B(T)-\int_0^T \alpha(t)\,dt\big)\exp\Big\{\int_0^T \alpha(t)\,dB_h(t)+\int_0^T \gamma(t)-\frac{1}{2}\alpha(t)^2\,dt\Big\}\Big)\Big]\\ &= \exp\Big(-\int_0^T \gamma(t)\,dt\Big)\\ &\quad\times E_Q\Big[F\Big(q_1\big(B_h(T)-\int_0^T \alpha(t)-h(t)\,dt\big)\exp\Big\{\int_0^T \alpha(t)\,dB_h(t)+\int_0^T \gamma(t)-\frac{1}{2}\alpha(t)^2\,dt\Big\}\Big)\Big].\end{aligned}$$

Furthermore, let $\alpha(t)=\alpha$ and $\gamma(t)=\gamma$ be constant functions, and $J(t)$ be a deterministic function. Then $h(t)$ is deterministic. And we have the following option pricing formula as an explicit integral

$$\begin{aligned}\mathcal{P}(\Phi) =& e^{-\gamma T}E_Q\Big[F\Big(q_1\big(B_h(T)-\alpha T+\int_0^T h(t)\,dt\big)\exp\Big\{\alpha B_h(T)+\big(\gamma-\frac{1}{2}\alpha^2\big)T\Big\}\Big)\Big]\\ =&\frac{e^{-\gamma T}}{\sqrt{2\pi T}}\int_{\mathbb{R}} F\Big(q_1\big(y-\alpha T+\int_0^T h(t)\,dt\big)\exp\Big\{\alpha y+\big(\gamma-\frac{1}{2}\alpha^2\big)T\Big\}\Big)e^{-\frac{y^2}{2T}}\,dy.\end{aligned} \tag{5.5.6}$$

Example 5.5.6. Suppose $q_1(x)=x$ and $F(x)=x^n$. Then the option price is

$$\begin{aligned}\mathcal{P}(\Phi) =&\frac{e^{-\gamma T}}{\sqrt{2\pi T}}\int_{\mathbb{R}}\big(y+\alpha T+\int_0^T h(t)\,dt\big)^n\exp\big\{\alpha n y+n(\gamma-\frac{1}{2}\alpha^2)T-\frac{y^2}{2T}\big\}\,dy\\ =&\frac{1}{\sqrt{2\pi T}}\exp\big\{(n-1)\gamma T-\frac{1}{2}n\alpha^2 T\big\}\\ &\times\int_{\mathbb{R}}\big(y+\alpha T+\int_0^T h(t)\,dt\big)^n\exp\big\{-\frac{1}{2T}(y+\alpha n T)^2+\frac{(\alpha n T)^2}{2T}\big\}\,d(y+\alpha n T)\\ =&\frac{1}{\sqrt{2\pi T}}\exp\big\{(n-1)\gamma T-\frac{1}{2}n\alpha^2 T(n-1)\big\}\\ &\times\int_{\mathbb{R}}\big(y+\alpha T-\alpha n T+\int_0^T h(t)\,dt\big)^n e^{-\frac{y^2}{2T}}\,dy.\end{aligned}$$

5.6 Hedging Portfolio

Section 5.5 gives the formula to help fixing a price for an arbitrary option. In order to realize the claim, we need to find out its hedging portfolio.

Suppose $\alpha(t) = \alpha$ and $\gamma(t) = \gamma$ are constant functions, and $J(t)$ is a deterministic function. From the proof of Theorem 5.4.2, the hedging portfolio corresponding to the option pricing formula in Theorem 5.5.5 is given by

$$p_0(t) = q_0\big(B(T)\big)^{-1}\Big(E_Q\big(\xi(T)F\big(X_T^{(1)}\big)\big) + \xi(t)A(t) + \int_0^t A(s)\gamma(s)\xi(s)\,ds\Big), \tag{5.6.1}$$

$$p_1(t) = \big(\alpha(t)\xi(t)X_t^{(1)}\big)^{-1}\theta(t), \tag{5.6.2}$$

where $A(t) = \int_0^t p_1(s)\,dX_s^{(1)} - p_1(t)X_t^{(1)}$ and $\theta(t)$ is a stochastic process satisfying

$$\xi(T)F\big(X_T^{(1)}\big) = E_Q\big[\xi(T)F\big(X_T^{(1)}\big)\big] + \int_0^T \theta(t)\,dB_h(t).$$

The above function $\theta(t)$ is calculated as

$$\theta(t) = E_Q\Big\{\frac{\delta}{\delta t}\Big(e^{-\gamma T}F\big(X_T^{(1)}\big)\Big)\Big|\mathcal{F}_t^{B_h}\Big\},$$

with the variational derivative (see [20]) calculated as

$$\begin{aligned}
&\frac{\delta}{\delta t}\Big(e^{-\gamma T}F\big(X_T^{(1)}\big)\Big)\\
&= \frac{\delta}{\delta t}\bigg(e^{-\gamma T}F\Big(q_1\big(B_h(T) + \alpha T + \int_0^T h(t)\,dt\big)\exp\{\alpha B_h(T) + (\gamma - \frac{1}{2}\alpha^2)T\}\Big)\bigg)\\
&= e^{\gamma T}F'\big(X_T^{(1)}\big)\exp\big(\alpha B_h(T) + (\gamma - \frac{1}{2}\alpha^2)T\big)\\
&\quad\times\Big[q_1'\Big(B_h(T) + \alpha T + \int_0^T h(t)\,dt\Big) + \alpha q_1(B(T) + \alpha T + \int_0^T h(t)\,dt)\Big].
\end{aligned}$$

So

$$\begin{aligned}
\theta(t) =& \frac{1}{\sqrt{2\pi(T-t)}}e^{\gamma T}\int_{\mathbb{R}}F'\Big(q_1\big(B_h(T)+\alpha T+\int_0^T h(t)\,dt\big)\exp\big(\alpha y+\alpha B_h(t)+(\gamma-\frac{1}{2}\alpha^2)T\big)\Big)\\
&\times\exp\big(\alpha y + \alpha B_h(t) + (\gamma - \frac{1}{2}\alpha^2)T - \frac{1}{2(T-t)}y^2\big)\,dy.
\end{aligned}$$

References

[1] Ayed, W. and Kuo, H.-H.: An extension of the Itô integral, *Communications on Stochastic Analysis* **2**, no. 3 (2008) 323–333.

[2] Ayed, W. and Kuo, H.-H.: An extension of the Itô integral: toward a general theory of stochastic integration, *Theory of Stochastic Processes* **16 (32)**, no. 1 (2010) 1–11.

[3] Basse-O'Connor, A., Graversen, S.-E., and Pedersen, J.: Stochastic integration on the real line, *Theory of Probability and Its Applications*, **58 (2)**, 193–215.

[4] Björk, T.: *Arbitrage Theory in Continuous Time* Oxford University Press, 2004.

[5] Black, F. and Scholes, M.: The pricing of options and corporate liabilities, *J. Political Economy* **81**, (1973) 637–659.

[6] Brown, R.: A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies, *Philosophical Magazine* **N.S.4**(1928), 161–173.

[7] Buckdahn, R.: Anticipating linear stochastic differential equations, *Lecture Notes in Control and Information Sciences* **136**, (1989) 18–23, Springer-Verlag.

[8] Buckdahn, R. and Nualart, D.: Linear stochastic differential equations and Wick products, *Probability Theory and Realted Fields*, **99**, no. 4 (1994) 501–526.

[9] Einstein, A.: Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen, *Annalen der Physik* **322**(1905), no. 8, 549–560.

[10] Esunge, J.: A class of anticipating linear stochastic differential equations, *Communications on Stochastic Analysis*, **3**, no. 1 (2009) 155–164.

[11] Girsanov, I. V.: On transforming a certain class of stochastic processes by absolutely continuous substitution of measures, *Theory of Probability & Its Applications* **5**, no. 3 (1960) 285–301.

[12] Gross, L.: Abstract Wiener spaces, *Proc. 5th Berkeley Symp. Math. Stat. and Probab.*, **2**, part 1 (1965), 31–42, University of California Press, Berkeley.

[13] Hida, T.: *Analysis of Brownian Functionals*, Carleton Mathematical Lecture Notes **13** 1975.

[14] Hwang, C.-R., Kuo, H.-H., Saitô, K., and Zhai, J.: A general Itô formula for adapted and instantly independent stochastic processes, *Communications on Stochastic Analysis* **10** (2016), no. 3, 341–362.

[15] Hwang, C.-R., Kuo, H.-H., Saitô, K., and Zhai, J.: Near-martingale property of anticipating stochastic integration, *Communications on Stochastic Analysis* **11** (2017), no. 4, 491–504.

[16] Itô, K.: Differential equations determining Markov processes, *Zenkoku Shijō Sūgaku Danwakai*, **244** (1942), no. 1077 1352–1400 (in Japanese).

[17] Itô, K.: Stochastic integral, *Proceedings of the Imperial Academy* **20**, no. 8 (1944) 519–524.

[18] Khalifa, N., Kuo, H.-H., Ouerdiane, H., and Szozda, B.: Linear stochastic differential equations with anticipating initial conditions, *Communications on Stochastic Analysis* **7**, no. 2 (2013) 245–253.

[19] Kuo, H.-H.: *White Noise Distribution Theory*, CRC Press, 1996.

[20] Kuo, H.-H.: *Introduction to Stochastic Integration*, Universitext, Springer, 2006.

[21] Kuo, H.-H.: The Itô calculus and white noise theory: a brief survey toward general stochastic integration, *Communication on Stochastic Analysis* **8**, no. 1 (2014) 111–139.

[22] Kuo, H.-H., Peng, Y., and Szozda, B.: Itô formula and Girsanov theorem for anticipating stochastic integrals, *Communication on Stochastic Analysis* **7**, no. 3 (2013) 441–458.

[23] Kuo, H.-H., Peng, Y., and Szozda, B.: Generalization of the anticipative Girsanov theorem, *Communications on Stochastic Analysis* **7**, no. 4 (2013) 573–589.

[24] Kuo, H.-H., Sae-Tang, A., and Szozda, B.: A stochastic integral for adapted and instantly independent stochastic processes, in "*Advances in Statistics, Probability and Actuarial Science*" Vol. I, Stochastic Processes, Finance and Control: A Festschrift in Honour of Robert J. Elliott (eds.: Cohen, S., Madan, D., Siu, T. and Yang, H.), World Scientific, 2012, 53–71.

[25] Kuo, H.-H., Sae-Tang, A., and Szozda, B.: An isometry formula for a new stochastic integral, In "*Proceedings of International Conference on Quantum Probability and Related Topics*," May 29–June 4, 2011, Levico, Italy, *QP–PQ: Quantum Probability and White Noise Analysis* **29** (2013) 222–232.

[26] Kuo, H.-H., Sae-Tang, A., and Szozda, B.: The Itô formula for a new stochastic integral, *Communications on Stochastic Analysis* **6**, no. 4 (2012) 603–614.

[27] Kuo, H.-H. and Saitô, K.: Doob's decomposition theorem for near-submartingales, *Communications on Stochastic Analysis* **9**, no. 4 (2015) 467–476.

[28] Kuo, H.-H., Sudip, S., and Zhai, J.: A Black–Scholes model with anticipating initial conditions, preprint.

[29] Kuo, H.-H. and Zhai, J.: On some anticipative stochastic differential equations, preprint.

[30] Malliavin, P.: Stochastic calculus of variation and hypoelliptic operators, *Proceedings of the International Symposium on Stochastic Differential Equations*, Tokyo, 1976, pp. 195–263, Wiley, New York-Chichester-Brisbane, 1978.

[31] Merton, R.: Theory of rational option pricing, *Bell Journal of Economics and Management Science* **4**, no. 2 (1973) 141–183.

[32] Pardoux, E. and Protter, P.: A two-sided stochastic integral and its calculus, *Probab Theory Related Fields*, **76** (1987), 15–49.

[33] Wiener, N: Differential space, *Journal of Mathematical Physics* **2**(1923), 131–174.

Part II

Temporal Minimum Action Method for Rare Events in Stochastic Dynamic Systems

Chapter 6
Introduction

6.1 Large Deviation Principle

We start talking about the motivation of the large deviation theory from mathematics point of view. Recall the most important theorems in probability: law of large numbers and central limit theorem.

Theorem 6.1.1.

$$\bar{X}_n \to \mu, \quad \text{as } n \to \infty,$$

where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$*, and the limit is in the sense of convergence in probability for the weak law and almost surely for the strong law.*

This law of large numbers (weak or strong) claims the stability of mean value, that is, when n is large enough, the average value of events will approach its expectation. Or equivalently, the probability that the mean value is far away from expectation (large deviation happens) is very small.

Theorem 6.1.2.

$$\frac{n_X}{n} \to p_X, \quad \text{as } n \to \infty,$$

where n_X *and* p_X *are the number of occurrence and the probability of event* X *respectively, and the limit is in the sense of convergence in probability for Bernoulli's law and almost surely for Borel's law.*

This law of large numbers claims the stability of frequency, that is, when n is large enough, the frequency of occurrence of an event will approach its probability. Or equivalently, the probability that the frequency is far away from probability is very small.

Theorem 6.1.3 (Lindeberg-Lévy Central Limit Theorem).

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} X_n - \mu\Big) \to N(0,\sigma^2),$$

where the limit is in distribution.

The central limit theorem claims the normality of distribution, that is, the normalization of the distribution of a sequence of random variables approaches normal distribution. This fact is the foundation of the effectiveness of using Brownian motion or white noise when we simulate a stochastic noise.

The central limit theorem provides the asymptotic property of average value approaching expectation stated in the law of large numbers, that is, how small the probability is for having large deviation:

$$P(|S_n| \geq \delta) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} e^{-x^2/2}\, dx.$$

By applying L'Hôpital's Rule twice, we obtain:

$$\frac{1}{n}(\log P(|S_n| \geq \delta)) \to -\frac{\delta^2}{2}, \quad \text{as } n \to \infty. \tag{6.1.1}$$

More generally, we can consider the asymptotic property of the probability of S_n having any kind of large deviation, for example,

$$P(\sqrt{n}S_n \in A) \to \frac{1}{\sqrt{2\pi}} \int_A e^{-x^2/2}\, dx, \quad \text{as } n \to \infty.$$

This leads to the usual discussion of the asymptotic property of $\frac{1}{n}\log \mu_n(A)$, or more generally $\varepsilon \log \mu_\varepsilon(A)$, and

Definition 6.1.4 (Large Deviation Principle). Let $\{\mu_\varepsilon\}$ be a family of probability measures. We call it satisfies *large deviation principle* if the inequality

$$-\inf_{x\in A^\circ} I(x) \leq \liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A^\circ) \leq \limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(\bar{A}) \leq -\inf_{x\in \bar{A}} I(x)$$

holds for some rate function $I(x)$ and all Borel sets A, where A° and $\bar{A}$ denote the interior and the closure of A, respectively.

6.2 Statistic Physics

In physics, large deviation principle is also motivated from statistic mechanics. Modern physics believes that some macro physical phenomena are in fact the statistic appearance of the particles contained in the system, for example, the heat of the system is the average kinetic energy of the particles. And these phenomena sometimes show rare performance and they follow large deviation principle. The rate function here is relative entropy originally.

Consider a vector $\gamma = (\gamma_1, \ldots, \gamma_\alpha) \in \mathbb{R}^\alpha$ satisfying $\gamma_i \geq 0$ and $\sum_{i=1}^{\alpha} \gamma_i = 1$. We call γ a *probability vector*. The set of all the probability vectors is denoted by $\mathcal{P}_\alpha$, which is also called the *α-simplex* in the space of α dimension.

Definition 6.2.1 (Relative Entropy)**.** For a fixed $\rho \in \mathcal{P}_\alpha$, the *relative entropy* of $\gamma \in \mathcal{P}_\alpha$ with respect to ρ is defined by

$$I_\rho(\gamma) = \sum_{i=1}^{\alpha} \gamma_i \log \frac{\gamma_i}{\rho_i}.$$

Property 6.2.2. *The relative entropy satisfies the following properties:*

(1). $I_\rho(\gamma) \geq 0$, and $I_\rho(\gamma) = 0 \Leftrightarrow \rho = \gamma$.

(2). I_ρ is strictly convex.

Remark 6.2.3. Property 6.2.2 (1) explains the reason of using I_ρ, that is, it describes the difference (deviation) between two distributions.

Boltzmann discovered from statistic point of view that in a multi-particle system, the probability of empirical distribution ω being in the neighborhood of the distribution γ approaches 0 asymptotically as $\exp[-nI_\rho(\gamma)]$. The rate function is exactly the relative entropy I_ρ.

If we call ρ the *equilibrium* distribution, then it is in fact the minimizer of the rate function (relative entropy). The empirical distribution must converge to ρ. But if γ is another state that is different from the equilibrium ρ, then "the empirical distribution approaches γ" is a rare event, and the probability of this approach (large deviation happens) converges to 0 asymptotically as the exponential of rate $-nI_\rho$. This principle is called the *principle of maximum entropy* in physics.

6.3 Freidlin-Wentzell Theory

Consider a stochastic differential equation

$$dX_t^\varepsilon = b(X_t^\varepsilon)\,dt + \sqrt{\varepsilon}\,dB_t, \tag{6.3.1}$$

or more generally a degenerate version

$$dX_t^\varepsilon = b(X_t^\varepsilon)\,dt + \sqrt{\varepsilon}\sigma(X_t^\varepsilon)\,dB_t, \tag{6.3.2}$$

with the initial condition $X_0^\varepsilon = 0$.

Example 6.3.1 (Langevin Equation)**.** A multi-Brownian-particle system satisfies the Langevin equation

$$\begin{cases} dQ_i(t) = \frac{P_i(t)}{m}\,dt, \\ dP_i(t) = -\nabla V(Q_i(t))\,dt - \gamma\frac{P_i(t)}{m}\,dt - \sqrt{2\gamma k_B T}\,dW_i(t), \end{cases}$$

where γ is the damping, k_B is the Boltzmann constant, and T is the absolute temperature. k_BT here is ε, the magnitude of environmental heat noise.

Consider the system (6.3.1). Let $I(w) = \frac{1}{2}\int_0^T |w_t' - b(w_t)|^2\,dt$. Then the distribution of the solution process X_t^ε in the Banach space $\mathcal{C}_0[0,T]$ satisfies the large deviation principle with rate function I, namely,

$$-I(A^\circ) \le \liminf_{\varepsilon\to 0} \varepsilon \log P[X^\varepsilon \in A^\circ] \le \limsup_{\varepsilon\to 0} \varepsilon \log P[X^\varepsilon \in \bar{A}] \le -I(\bar{A}),$$

for $A \in \mathcal{B}(\mathcal{C}_0)$.

One of the most important consequence of Freidlin-Wentzell theory is that in comparison to the deterministic system of several equilibrium (attractor) without random noise

$$dX_t^\varepsilon = b(X_t^\varepsilon)\,dt, \tag{6.3.3}$$

a particle/state is not able to escape from one equilibrium and transit to another. But in a system with noise, even if the noise has very small altitude, it is still possible to escape from an equilibrium and move to another. Under the assumption of white noise, this movement is a rare event, or almost impossible event. While the possibility exists, it asymptotically converges to 0 with rate function I. Among all these rare events, there is one most possible (least unlikely) event. This event is the most possible way of moving from one equilibrium to another driven by noise. And it is in fact the minimizer of I. This minimum of I corresponds to the minimum action, so I is also called the *action functional* which is related to Lagrangian-Hamiltonian in physics. So the long-term behavior of the perturbed system (6.3.1) is characterized by the small-noise-induced transitions between the equilibriums of the unperturbed system (6.3.3). Although these transitions rarely occur, they can describe many critical phenomena in physical, chemical, and biological systems, such as non-equilibrium interface growth [12, 24], regime change in climate [36], switching in biophysical network [35], hydrodynamic instability [31, 30], etc. Therefore, to find out this minimum action and the minimizer is very important for studying the stability of the system in noise environment.

We describe our problem rigorously. In order to do so, from now on, we denote the action functional I as

$$S_T(\varphi) = I(\varphi) = \frac{1}{2}\int_0^T |\varphi' - b(\varphi)|^2\,dt. \tag{6.3.4}$$

Let x_1 and x_2 be two arbitrary points in the phase space. If we restrict the transition on a certain time interval $[0, T]$, we have

$$\lim_{\delta\downarrow 0}\lim_{\varepsilon\downarrow 0} -\varepsilon \log \Pr(\tau_\delta \le T) = \inf_{\substack{\varphi(0)=x_1,\\ \varphi(T)=x_2}} S_T(\varphi), \tag{6.3.5}$$

where τ_δ is the first entrance time of the δ-neighborhood of x_2 for the random process $X(t)$ starting from x_1. The path variable φ connecting x_0 and x_1, over which the action functional is minimized, is called a *transition path*. If the time scale is not specified, the transition probability can be described with respect to the quasi-potential $V(x_1, x_2)$ from x_1 to x_2:

$$V(x_1, x_2) := \inf_{T \in \mathbb{R}^+} \inf_{\substack{\varphi(0)=x_1, \\ \varphi(T)=x_2}} S_T(\varphi). \tag{6.3.6}$$

The probability meaning of $V(x_1, x_2)$ is

$$V(x_1, x_2) = \inf_{T \in \mathbb{R}^+} \lim_{\delta \downarrow 0} \lim_{\varepsilon \downarrow 0} -\varepsilon \log \Pr(\tau_\delta \leq T). \tag{6.3.7}$$

We in general call the asymptotic results given in Equation (6.3.5) and (6.3.7) *large deviation principle* (*LDP*). We use φ^* to indicate transition path that minimizes the action functional in Equation (6.3.5) or (6.3.6), which is also called the *minimal action path* (*MAP*) [9], or the instanton in physical literature related to path integral. The MAP φ^* is the most probable transition path from x_1 to x_2. For the quasi-potential, we let T^* indicate the optimal integration time, which can be either finite or infinite depending on x_1 and x_2. The importance of LDP is that it simplifies the computation of transition probability, which is a path integral in a function space, to seeking the minimizers φ^* or (T^*, φ^*).

Chapter 7
Temporal Minimum Action Method

7.1 Problem Description

As we have seen from Section 6.3, the results (6.3.5) and (6.3.6) from the Freidlin-Wentzell (F-W) theory of large deviations lead to the following two problems

$$\text{Problem I :} \quad S_T(\varphi^*) = \inf_{\substack{\varphi(0)=x_1,\\ \varphi(T)=x_2}} S_T(\varphi), \tag{7.1.1}$$

and

$$\text{Problem II :} \quad S_{T^*}(\varphi^*) = \inf_{T\in\mathbb{R}^+} \inf_{\substack{\varphi(0)=x_1,\\ \varphi(T)=x_2}} S_T(\varphi). \tag{7.1.2}$$

Here $\varphi(t)$ is a path connecting x_1 and x_2 in the phase space on the time interval $[0,T]$. The minima and minimizers of Problems I and II characterize the difficulty of the small-noise-induced transition from x_1 to the vicinity of x_2, see Equation (6.3.5) and (6.3.7). In Problem I, the transition is restricted to a certain time scale T, which is relaxed in Problem II.

To analyze the convergence of numerical approximations for Problem I and II, we need the following assumptions for $b(x)$.

Assumption 7.1.1. (1) $b(x)$ is Lipschitz continuous in a big ball, i.e., there exist constants $K>0$ and $R_1>0$, such that

$$|b(x)-b(y)| \le K|x-y|, \quad \forall\ x,y\in B_{R_1}(0), \tag{7.1.3}$$

where $|\cdot|$ denotes the ℓ_2 norm of a vector in $\mathbb{R}^n$;

(2) There exist positive numbers β, R_2, such that

$$\langle b(x), x\rangle \le -\beta|x|^2, \quad \forall\ |x|\ge R_2, \tag{7.1.4}$$

where $R_2^2 \le R_1^2 - \frac{S^*}{\beta}$, and

$$S^* = \max_{x,y \in B_{R_2}(0)} \frac{1}{2} \int_0^1 \left| y - x - b\big(x + (y-x)t\big) \right|^2 dt.$$

(3) The solution points of $b(x) = 0$ are isolated.

Lemma 7.1.2. *Let Assumption* (7.1.4) *hold. If both the starting and ending points of a MAP $\varphi(t)$ are inside $B_{R_2}(0)$, then $\varphi(t)$ is located within $B_{R_1}(0)$ for any t.*

Proof. Suppose that $\varphi(t)$ is a MAP outside of $B_{R_2}(0)$ but connecting two points x and y on the the surface of $B_{R_2}(0)$. Let $w(t) = \varphi' - b(\varphi)$. We have

$$\varphi' = b(\varphi) + w. \tag{7.1.5}$$

Taking inner product on both sides of the above equation with 2φ, we get

$$\frac{d|\varphi|^2}{dt} = 2\langle b(\varphi), \varphi \rangle + 2\langle w, \varphi \rangle. \tag{7.1.6}$$

Then by using Cauchy's inequality with β, and Assumption (7.1.4), we get

$$\frac{d|\varphi|^2}{dt} \le -2\beta|\varphi|^2 + \frac{1}{2\beta}|w|^2 + 2\beta|\varphi|^2 = \frac{1}{2\beta}|w|^2. \tag{7.1.7}$$

Taking integration and using the definition of minimum action, we obtain a bound for any t along the MAP:

$$|\varphi|^2 \le |x|^2 + \int_0^t \frac{1}{2\beta}|w|^2\, dt \le R_2^2 + \frac{1}{\beta} S_{T^*}(\varphi) \le R_2^2 + \frac{1}{\beta} S^* \le R_1^2, \tag{7.1.8}$$

which means that the whole MAP is located within $B_{R_1}(0)$. $\square$

Remark 7.1.3. The Assumptions (7.1.3) and (7.1.4) allow most of the physically relevant smooth nonlinear dynamics. It is seen from Lemma 7.1.2 that Assumption (7.1.4) is used to restrict all MAPs of interest inside $B_{R_1}(0)$. For simplicity and without loss of generality, we will assume from now on that the Lipschitz continuity of $b(x)$ is global, namely, $R_1 = \infty$. For the general case given in Assumption 7.1.1, one can achieve all the conclusions by restricting the theorems and proofs to $B_{R_1}(0)$.

For simplicity, we will use the following notations in this chapter. For $\varphi(t) \in \mathbb{R}^n$ defined on $\Gamma_T = [0, T]$, denote $|\varphi|^2 = \sum_{i=1}^n |\varphi_i|^2$ and $|\varphi|^2_{m,\Gamma_T} = \sum_{i=1}^n |\varphi_i|^2_{m,\Gamma_T}$, where φ_i is the i-th component of φ and $|\varphi_i|^2_{m,\Gamma_T} = \int_{\Gamma_T} |\varphi_i^{(m)}|^2\, dt$. Denote $\|\varphi\|^2_{m,\Gamma_T} = \sum_{i=1}^n \|\varphi_i\|^2_{m,\Gamma_T}$, where $\|\varphi_i\|^2_{m,\Gamma_T} = \sum_{k\le m} \int_{\Gamma_T} |\varphi_i^{(k)}|^2\, dt$. For $f(t), g(t) : \Gamma_T \to \mathbb{R}^n$, we denote the inner products $\langle f, g\rangle = \sum_{i=1}^n f_i g_i$ and $\langle f, g\rangle_{\Gamma_T} = \int_{\Gamma_T} \left(\sum_{i=1}^n f_i g_i\right) dt$.

7.2 Introduction to the Temporal Minimum Action Method

Starting from [8], the large deviation principle given by the F-W theory has been approximated numerically, especially for non-gradient systems, and the numerical methods are, in general, called *minimum action method* (*MAM*).

The numerical difficulty is to separate the slow dynamics around critical points from fast dynamics elsewhere. In fact, since the noise is small, it takes infinite time to escape from a small neighborhood of a critical point. So the MAP will be mainly captured by the fast dynamics subject to a finite time, but it will take infinite time to pass a critical point. To overcome this difficulty, there exist two basic techniques: (1) non-uniform temporal discretization including moving mesh technique and adaptive finite element method (aMAM), and (2) reformulation of the action functional with respect to arc length (gMAM). Both techniques aim to solve Problem II with $T^* = \infty$. In aMAM, a finite but large T is used while in gMAM, the infinite T^* is mapped to a finite arc length. So, aMAM is not able to deal with Problem II subject to a finite T^* since a fixed T is required. Since T has been removed, gMAM is not able to deal with Problem I.

To deal with both Problem I and II in a consistent way, we have developed a minimum action method with optimal linear time scaling (*temporal minimum action method*, or *tMAM*) coupled with finite element discretization. This method is based on two observations:

1. For any given transition path, there exists a unique T to minimize the action functional subject to a linear scaling of time;

2. For transition paths defined on a finite element approximation space, the optimal integration time T^* is always finite but increases to infinity as the approximation space is refined.

The first observation removes the parameter T in Problem II and allows us to always discretize only on the unit interval Γ_1. The second observation guarantees that the discrete problem of Problem II is well-posed after T is removed. In addition, Problem I becomes a special case of our reformation of Problem II. Then we achieve our aim to deal with both Problem I and II in a consistent way.

Although many techniques have been developed from the algorithm point of view, few numerical analysis has been done for minimum action method. We fill this gap partially in this chapter. We start with the idea of tMAM.

7.3 Reformulation and Finite Element Discretization of the Problem

By Maupertuis' principle of least action for the minimizer (T^*, φ^*) of Problem II, we have a necessary condition:

Lemma 7.3.1 ([17]). *Let (T^*, φ^*) be the minimizer of Problem II. Then φ^* is located on the surface $H(\varphi, \frac{\partial L}{\partial \varphi'}) = 0$, where H is the Hamiltonian given by the Legendre transform $L(\varphi, \varphi') := \frac{1}{2}|\varphi' - b(\varphi)|^2$. More specifically, for Equation* (6.3.1)

$$H(\varphi, \frac{\partial L}{\partial \varphi'}) = 0 \quad \Longleftrightarrow \quad |\varphi'(t)| = |b(\varphi(t))|, \quad \forall\, t. \tag{7.3.1}$$

We call Equation (7.3.1) the *zero-Hamiltonian constraint*. The zero-Hamiltonian constraint provides a nonlinear mapping between the arc length of the geometrically fixed lines on surface $H = 0$ and time t (see Section 7.6 for more details).

Here we consider a linear time scaling on Γ_T, which is simpler and more flexible for numerical approximation. For any given transition path φ and a fixed T, we consider the change of variable $s = t/T \in [0,1] = \Gamma_1$. Let $\varphi(t) = \varphi(sT) =: \bar{\varphi}(s)$. Then $\bar{\varphi}'(s) = \varphi'(t)T$, and we can rewrite the action functional as

$$S_T(\varphi(t)) = S_T(\bar{\varphi}(s)) = \frac{T}{2}\int_0^1 \left|T^{-1}\bar{\varphi}'(s) - b(\bar{\varphi}(s))\right|^2 ds =: S(T, \bar{\varphi}). \tag{7.3.2}$$

Now, the two variables $\bar{\varphi}$ and T are independent, since $\bar{\varphi}$ is just an arbitrary path defined on $[0,1]$ connecting x_0 and x_1. So for a fixed path and using optimality condition, we can find the minimizer of S, which is simply a function of T.

Lemma 7.3.2. *For any given transition path $\bar{\varphi}$, we have*

$$\hat{S}(\bar{\varphi}) := S(\hat{T}(\bar{\varphi}), \bar{\varphi}) = \inf_{T\in\mathbb{R}^+} S(T, \bar{\varphi}), \tag{7.3.3}$$

if $\hat{T}(\bar{\varphi}) < \infty$, where

$$\hat{T}(\bar{\varphi}) = \frac{|\bar{\varphi}'|_{0,\Gamma_1}}{|b(\bar{\varphi})|_{0,\Gamma_1}}. \tag{7.3.4}$$

Proof. It is easy to verify that the functional $\hat{T}(\bar{\varphi})$ is nothing but the unique solution of the optimality condition $\partial_T S(T, \bar{\varphi}) = 0$. □

Corollary 7.3.3. *Let (T^*, φ^*) be the minimizer of Problem II. If $T^* < \infty$, we have $T^* = \hat{T}(\bar{\varphi}^*)$, where $\bar{\varphi}^*(s) := \varphi^*(sT^*)$.*

Proof. By the zero-Hamiltonian constraint (7.3.1) and the definition of $\bar{\varphi}$, we have

$$|(\bar{\varphi}^*)'| = |(\varphi^*)'|T^* = |b(\varphi^*)|T^* = |b(\bar{\varphi}^*)|T^*.$$

Integrating both sides on Γ_1, we have the conclusion. □

For any absolutely continuous path φ, it is shown in Theorem 5.6.3 in [6] that S_T can be written as

$$S_T(\varphi) = \begin{cases} S_T(\varphi), & \varphi \in H^1(\Gamma_T; \mathbb{R}^n), \\ \infty, & \text{otherwise.} \end{cases} \tag{7.3.5}$$

Thus, we can solve the problem in the framework of Sobolev space $H^1(\Gamma_T;\mathbb{R}^n)$. From now on, we use $H^1(\Gamma_T)$ to indicate $H^1(\Gamma_T;\mathbb{R}^n)$ for simplicity. The same rule applies to other spaces such as $H_0^1(\Gamma;\mathbb{R}^n)$ and $L^2(\Gamma;\mathbb{R}^n)$.

Define the following two admissible sets consisting of transition paths:

$$\mathcal{A}_T = \big\{\varphi \in H^1(\Gamma_T) : \varphi(0) = 0,\, \varphi(T) = x\big\}, \tag{7.3.6}$$

$$\mathcal{A}_1 = \big\{\bar{\varphi} \in H^1(\Gamma_1) : \bar{\varphi}(0) = 0,\, \bar{\varphi}(1) = x\big\}, \tag{7.3.7}$$

where we let $x_1 = 0$ and $x_2 = x$ just for convenience.

Lemma 7.3.4. *If $T^* < \infty$, we have*

$$S_{T^*}(\varphi^*) = \hat{S}(\bar{\varphi}^*) = \inf_{\bar{\varphi}\in\mathcal{A}_1} \hat{S}(\bar{\varphi}). \tag{7.3.8}$$

and $T^* = \hat{T}(\bar{\varphi}^*)$ (see Equation (7.3.4)), where $\varphi^*(t) = \bar{\varphi}^*(t/T^*)$ (or $\bar{\varphi}^*(s) = \varphi^*(sT^*)$).

Proof. If (T^*, φ^*) is a minimizer of $S_T(\varphi)$ and $T^* < \infty$, then

$$S_{T^*}(\varphi^*) = \inf_{T\in\mathbb{R}^+}\inf_{\varphi\in\mathcal{A}_T} S_T(\varphi) = \inf_{\bar{\varphi}\in\mathcal{A}_1} \hat{S}(\bar{\varphi}) \le \hat{S}(\bar{\varphi}^*),$$

and

$$\hat{S}(\bar{\varphi}^*) = \inf_{T\in\mathbb{R}^+} S(T,\bar{\varphi}^*) = \inf_{T\in\mathbb{R}^+} S_T(\varphi^*) \le S_{T^*}(\varphi^*).$$

Thus, $S_{T^*}(\varphi^*) = S(T^*,\bar{\varphi}^*) = \hat{S}(\bar{\varphi}^*)$, that is, $\bar{\varphi}^*$ is a minimizer of $\hat{S}(\bar{\varphi})$ for $\bar{\varphi} \in \mathcal{A}_1$, and $T^* = \hat{T}(\bar{\varphi}^*)$ from Corollary 7.3.3.

Conversely, if $\bar{\varphi}^*$ is a minimizer of $\hat{S}(\bar{\varphi})$, we let $T^* = \hat{T}(\bar{\varphi}^*)$, and $\varphi^*(t) = \bar{\varphi}^*(\frac{t}{T^*})$, for $t \in [0, T^*]$. We have

$$S_{T^*}(\varphi^*) = S(\hat{T}(\bar{\varphi}^*),\bar{\varphi}^*) = \hat{S}(\bar{\varphi}^*) = \inf_{\bar{\varphi}\in\mathcal{A}_1} \hat{S}(\bar{\varphi}) = \inf_{T\in\mathbb{R}^+}\inf_{\varphi\in\mathcal{A}_T} S_T(\varphi),$$

when $T^* < \infty$. Then (T^*, φ^*) is a minimizer of $S_T(\varphi)$. So the minimizers of $\hat{S}(\bar{\varphi})$ and $S_T(\varphi)$ have a one-to-one correspondence when the optimal integral time is finite. $\square$

Lemma 7.3.4 shows that for a finite T^* we can use Equation (7.3.8) instead of Problem II to approximate the quasi-potential such that the optimization parameter T is removed and we obtain a new problem

$$\hat{S}(\bar{\varphi}^*) = \inf_{\substack{\bar{\varphi}(0)=x_1,\\ \bar{\varphi}(1)=x_2}} \hat{S}(\bar{\varphi}) \tag{7.3.9}$$

that is equivalent to Problem II.

Now, we discretize the problems and solve it numerically. Let $\mathcal{T}_h$ and $\overline{\mathcal{T}}_h$ be partitions of Γ_T and Γ_1, respectively. We define the following approximation spaces given by linear finite elements:

$$\mathcal{B}_h = \big\{\varphi_h \in \mathcal{A}_T : \varphi_h|_I \text{ is affine for each } I \in \mathcal{T}_h\big\},$$

$$\overline{\mathcal{B}}_h = \big\{\bar{\varphi}_h \in \mathcal{A}_1 : \bar{\varphi}_h|_I \text{ is affine for each } I \in \overline{\mathcal{T}}_h\big\}.$$

For any h, we define the following discretized action functionals:

$$S_{T,h}(\varphi_h) = \begin{cases} \frac{1}{2}\int_0^T |\varphi_h' - b(\varphi_h)|^2\, dt, & \text{if } \varphi_h \in \mathcal{B}_h, \\ \infty, & \text{if } \varphi_h \notin \mathcal{B}_h, \end{cases} \tag{7.3.10}$$

and

$$\hat{S}_h(\bar{\varphi}_h) = \begin{cases} \frac{\hat{T}(\bar{\varphi}_h)}{2}\int_0^1 |\frac{1}{\hat{T}(\bar{\varphi}_h)}\bar{\varphi}_h' - b(\bar{\varphi}_h)|^2\, dt, & \text{if } \bar{\varphi}_h \in \overline{\mathcal{B}}_h, \\ \infty, & \text{if } \bar{\varphi}_h \notin \overline{\mathcal{B}}_h. \end{cases} \tag{7.3.11}$$

7.4 Problem I with a Fixed T

For this case, our main results are summarized in the following theorem:

Theorem 7.4.1. *For Problem I with a fixed T, we have*

$$\min_{\varphi \in \mathcal{A}_T} S_T(\varphi) = \lim_{h\to 0} \inf_{\varphi_h \in \mathcal{B}_h} S_{T,h}(\varphi_h),$$

namely, the minima of $S_{T,h}$ converge to the minimum of $S_T(\varphi)$ as $h \to 0$. Moreover, if $\{\varphi_h\} \subset \mathcal{B}_h$ is a sequence of minimizers of $S_{T,h}$, then there is a subsequence that converges weakly in $H^1(\Gamma_T)$ to some $\varphi \in \mathcal{A}_T$, which is a minimizer of S_T.

We prove this theorem in three steps:

Step 1: *The existence of the minimizer of $S_T(\varphi)$ in $\mathcal{A}_T$.*

Lemma 7.4.2 (Solution Existence)**.** *There exists at least one function $\varphi^* \in \mathcal{A}_T$ such that*

$$S_T(\varphi^*) = \min_{\varphi \in \mathcal{A}_T} S_T(\varphi).$$

Proof. We first prove the coerciveness of $S_T(\varphi) = \frac{1}{2}\int_0^T |\varphi' - b(\varphi)|^2\, dt$. Define an auxiliary function g by

$$g(t) = \varphi(t) - \int_0^t b(\varphi(u))\, du.$$

Then $g' = \varphi' - b(\varphi)$ and $g(0) = 0$. Since b is globally Lipschitz continuous, we have

$$\begin{aligned} |\varphi'(t)| &\leq |b(\varphi(t)) - b(0)| + |b(0)| + |g'| \\ &\leq K|\varphi| + |b(0)| + |g'(t)| \\ &\leq K \int_0^t |\varphi'(s)|\, ds + |b(0)| + |g'(t)|. \end{aligned}$$

By Gronwall's inequality, we have

$$|\varphi'(t)| \leq K \int_0^t (|b(0)| + |g'(s)|) e^{K(t-s)}\, ds + |b(0)| + |g'(t)|,$$

from which we obtain

$$|\varphi|_{1,\Gamma_T} \leq C_1 |b(0)|^2 + C_2 |g|^2_{1,\Gamma_T},$$

where C_1 and C_2 are two positive constants depending on K and T. Thus,

$$S_T(\varphi) = \frac{1}{2}|g|^2_{1,\Gamma_T} \geq \frac{1}{2} C_2^{-1} |\varphi|^2_{1,\Gamma_T} - \frac{1}{2} C_1 C_2^{-1} |b(0)|^2.$$

The coerciveness of the action functional follows. On the other hand, the integrand $|\varphi' - b(\varphi)|^2$ is bounded below by 0, and convex in φ'. By Theorem 2 on Page 448 in [11], $S_T(\varphi)$ is weakly lower semicontinuous on $H^1(\Gamma_T)$.

For any minimizing sequence $\{\varphi_k\}_{k=1}^{\infty}$, from the coerciveness, we have

$$\sup_k |\varphi_k|_{1,\Gamma_T} < \infty.$$

Let $\varphi_0 \in \mathcal{A}_T$ be any fixed function, e.g., the linear function on Γ_T from 0 to x. Then $\varphi_k - \varphi_0 \in H_0^1(\Gamma_T)$, and

$$|\varphi_k|_{0,\Gamma_T} \leq |\varphi_k - \varphi_0|_{0,\Gamma_T} + |\varphi_0|_{0,\Gamma_T} \leq C_p|\varphi_k - \varphi_0|_{1,\Gamma_T} + |\varphi_0|_{0,\Gamma_T} < \infty,$$

by the Poincaré's Inequality. Thus $\{\varphi_k\}_{k=1}^{\infty}$ is bounded in $H^1(\Gamma_T)$. Then there exists a subsequence $\{\varphi_{k_j}\}_{j=1}^{\infty}$ converging weakly to some φ^* in $H^1(\Gamma_T)$. Then $\varphi_{k_j} - \varphi_0$ converges to $\varphi^* - \varphi_0$ weakly in $H_0^1(\Gamma_T)$. By Mazur's Theorem [11], $H_0^1(\Gamma_T)$ is weakly closed. So $\varphi^* - \varphi_0 \in H_0^1(\Gamma_T)$, i.e., $\varphi^* \in \mathcal{A}_T$.

Therefore, $S_T(\varphi^*) \leq \liminf_{j\to\infty} S_T(\varphi_{k_j}) = \inf_{\varphi\in\mathcal{A}_T} S_T(\varphi)$. Since $\varphi^* \in \mathcal{A}_T$, we reach the conclusion. □

Step 2: *The Γ-convergence of $S_{T,h}$ to S_T as $h \to 0$.*

The following property will be used frequently.

Property 7.4.3. *For any sequence $\{\varphi_h\} \subset \mathcal{B}_h$ converging weakly to $\varphi \in H^1(\Gamma_T)$, we have*

$$\lim_{h\to 0} |b(\varphi_h) - b(\varphi)|_{0,\Gamma_T} = 0.$$

Proof. Since φ_h converges weakly to φ in $H^1(\Gamma_T)$, $\varphi_h \to \varphi$ in $L^2(\Gamma_T)$, i.e., φ_h converges strongly to φ in the L^2 sense. By the Lipschitz continuity of b, we reach the conclusion. □

Lemma 7.4.4 (Γ-convergence of $S_{T,h}$)**.** *Let $\{\mathcal{T}_h\}$ be a sequence of finite element meshes with $h \to 0$. For every $\varphi \in \mathcal{A}_T$, the following two properties hold:*

- *Lim-inf inequality: for every sequence $\{\varphi_h\}$ converging weakly to φ in $H^1(\Gamma_T)$, we have*

$$S_T(\varphi) \leq \liminf_{h\to 0} S_{T,h}(\varphi_h). \tag{7.4.1}$$

- *Lim-sup inequality: there exists a sequence $\{\varphi_h\} \subset \mathcal{B}_h$ converging weakly to φ in $H^1(\Gamma_T)$, such that*

$$S_T(\varphi) \geq \limsup_{h\to 0} S_{T,h}(\varphi_h). \tag{7.4.2}$$

Proof. We only need to consider a sequence $\{\varphi_h\} \subset \mathcal{B}_h$ for the lim-inf inequality, since otherwise, (7.4.1) is trivial by the definition of $S_{T,h}(\varphi)$. Let $\{\varphi_h\} \subset \mathcal{B}_h$ be an arbitrary sequence converging weakly to φ in $H^1(\Gamma_T)$. The discretized action functional can be rewritten as

$$\begin{aligned}
&\int_0^T |\varphi_h' - b(\varphi_h)|^2\, dt \\
&= \int_0^T |\varphi_h'|^2\, dt + \int_0^T |b(\varphi_h)|^2\, dt - 2\int_0^T \langle \varphi_h', b(\varphi_h)\rangle\, dt \\
&= I_1 + I_2 + I_3.
\end{aligned} \tag{7.4.3}$$

The first summand functional I_1 is obviously weakly lower semicontinuous in $H^1(\Gamma_T)$ since the integrand is convex with respect to φ'.

For I_2 in Equation (7.4.3), using Property 7.4.3, we have

$$\lim_{h\to 0} |b(\varphi_h)|_{0,\Gamma_T} = |b(\varphi)|_{0,\Gamma_T},$$

For I_3 in Equation (7.4.3), we have

$$\begin{aligned}
&\left|\int_0^T \langle \varphi_h', b(\varphi_h)\rangle\, dt - \int_0^T \langle \varphi', b(\varphi)\rangle\, dt\right| \\
&= \left|\int_0^T \langle \varphi_h', b(\varphi_h) - b(\varphi)\rangle\, dt + \int_0^T \langle \varphi_h' - \varphi', b(\varphi)\rangle\, dt\right| \\
&\leq |\varphi_h|_{1,\Gamma_T} |b(\varphi_h) - b(\varphi)|_{0,\Gamma_T} + |\langle \varphi_h' - \varphi', b(\varphi)\rangle_{\Gamma_T}|.
\end{aligned}$$

Using Property 7.4.3 and the fact that $\sup_h |\varphi_h|_{1,\Gamma_T} < \infty$, we have that the first term of the above inequality converges to 0. Moreover, the second term also converges to 0 due to the weak convergence of φ_h to φ in $H^1(\Gamma_T)$. Thus,

$$\lim_{h\to 0} \int_0^T \langle \varphi_h', b(\varphi_h)\rangle\, dt = \int_0^T \langle \varphi', b(\varphi)\rangle\, dt.$$

Combining the results for I_1, I_2 and I_3, we obtain the lim-inf inequality:

$$\begin{aligned}
&\liminf_{h\to 0} \int_0^T |\varphi_h' - b(\varphi_h)|^2\, dt \\
&= \liminf_{h\to 0} \left[\int_0^T |\varphi_h'|^2\, dt + \int_0^T |b(\varphi_h)|^2\, dt - 2\int_0^T \langle \varphi_h', b(\varphi_h)\rangle\, dt \right] \\
&= \liminf_{h\to 0} \int_0^T |\varphi_h'|^2\, dt + \lim_{h\to 0} \int_0^T |b(\varphi_h)|^2\, dt - 2\lim_{h\to 0}\int_0^T \langle \varphi_h', b(\varphi_h)\rangle\, dt \\
&\geq \int_0^T |\varphi'|^2\, dt + \int_0^T |b(\varphi)|^2\, dt - 2\int_0^T \langle \varphi', b(\varphi)\rangle\, dt \\
&= \int_0^T |\varphi' - b(\varphi)|^2\, dt.
\end{aligned}$$

Now, we address the lim-sup inequality. Since $H^2(\Gamma_T)$ is dense in $H^1(\Gamma_T)$, for any $\varphi \in H^1(\Gamma_T)$ and $\varepsilon > 0$, there exists a non-zero $u_\varepsilon \in H^2(\Gamma_T)$, such that $\|\varphi - u_\varepsilon\|_{1,\Gamma_T} < \varepsilon$. We have

$$|\mathcal{I}_h u_\varepsilon - u_\varepsilon|_{1,\Gamma_T} \leq ch|u_\varepsilon|_{2,\Gamma_T} \leq c\varepsilon,$$

by letting

$$h = h(\varepsilon) = \min\{\frac{\varepsilon}{|u_\varepsilon|_{1,\Gamma_T}}, \frac{\varepsilon}{|u_\varepsilon|_{2,\Gamma_T}}, \varepsilon\},$$

where $\mathcal{I}_h$ is an interpolation operator defined by linear finite elements. Let $\varphi_h = \mathcal{I}_h u_\varepsilon$. Then we have $\varphi_h \in \mathcal{B}_h$, and

$$\begin{aligned}
|\varphi_h - \varphi|_{1,\Gamma_T} \leq& |\varphi_h - u_\varepsilon|_{1,\Gamma_T} + |u_\varepsilon - \varphi|_{1,\Gamma_T} \\
=& |\mathcal{I}_h u_\varepsilon - u_\varepsilon|_{1,\Gamma_T} + |u_\varepsilon - \varphi|_{1,\Gamma_T} \\
<& c\varepsilon + \varepsilon \to 0,
\end{aligned}$$

and

$$\begin{aligned}
|\varphi_h - \varphi|_{0,\Gamma_T} \leq& |\varphi_h - u_\varepsilon|_{0,\Gamma_T} + |u_\varepsilon - \varphi|_{0,\Gamma_T} \\
=& |\mathcal{I}_h u_\varepsilon - u_\varepsilon|_{0,\Gamma_T} + |u_\varepsilon - \varphi|_{0,\Gamma_T} \\
\leq& ch|u_\varepsilon|_{1,\Gamma_T} + \varepsilon \\
<& c\varepsilon + \varepsilon \to 0,
\end{aligned}$$

as $\varepsilon \to 0$. So φ_h converges to φ in $H^1(\Gamma_T)$, and also converges weakly in $H^1(\Gamma_T)$. By Property 7.4.3, we know that $b(\varphi_h) \to b(\varphi)$ in $L_2(\Gamma_T)$. Thus,

$$\lim_{h\to 0} S_{T,h}(\varphi_h) = \lim_{h\to 0} \frac{1}{2}|\varphi_h' - b(\varphi_h)|^2_{0,\Gamma_T} = S_T(\varphi),$$

which proves the lim-sup equality. □

Step 3: *The proof of Theorem 7.4.1.*

Proof. With the solution existence and the Γ-convergence, we only need the equi-coerciveness of $S_{T,h}$ for the final conclusion. For any $\varphi_h \in \mathcal{B}_h$, we have $S_{T,h}(\varphi_h) = S_T(\varphi_h)$. Then the equi-coerciveness of $S_{T,h}$ in $\mathcal{B}_h$ follows from the coerciveness of $S_T(\varphi_h)$ restricted to $\mathcal{B}_h \subset \mathcal{A}_T$ (see the first part of the proof of Lemma 7.4.2). □

7.5 Problem II with a Finite T^*

For this case, we consider the reformulation of S_T given in Section 7.3. From Lemma 7.3.4, we know that Problem II with a finite T^* is equivalent to minimizing $\hat{S}$ in $\mathcal{A}_1$ (see Equation (7.3.8)). Our main results are summarized in the following theorem:

Theorem 7.5.1. *For Problem II with a finite T^*, we have*

$$\min_{\bar{\varphi}\in\mathcal{A}_1} \hat{S}(\bar{\varphi}) = \lim_{h\to 0} \inf_{\bar{\varphi}_h\in\overline{\mathcal{B}}_h} \hat{S}_h(\bar{\varphi}_h),$$

namely, the minima of $\hat{S}_h$ converge to the minimum of $\hat{S}$ as $h \to 0$. Moreover, if $\{\bar{\varphi}_h\} \subset \overline{\mathcal{B}}_h$ is a sequence of minimizers of $\hat{S}_h$, then there is a subsequence that converges weakly in $H^1(\Gamma_1)$ to some $\bar{\varphi} \in \mathcal{A}_1$, which is a minimizer of $\hat{S}$.

Similar to Problem I with a fixed T, we prove this theorem in four steps:

Step 1: *The strong positiveness of the functional $\hat{T}$.*

Property 7.5.2. *There exists a constant $C_{\hat{T}} > 0$ such that*

$$\hat{T}(\bar{\varphi}) \geq C_{\hat{T}}, \tag{7.5.1}$$

for any $\bar{\varphi} \in \mathcal{A}_1$.

Proof. For any $\bar{\varphi} \in \mathcal{A}_1$, let $\bar{\varphi} = \bar{\varphi}_0 + \bar{\varphi}_L$, where $\bar{\varphi}_0 \in H_0^1(\Gamma_1)$ and $\bar{\varphi}_L(s) = xs, s \in [0,1]$ is the linear path connecting 0 and x. We have

$$\begin{aligned}
\hat{T}(\bar{\varphi}) &= \frac{|\bar{\varphi}_0' + x|_{0,\Gamma_1}}{|b(\bar{\varphi}_0 + \bar{\varphi}_L)|_{0,\Gamma_1}} \\
&\geq \frac{|\bar{\varphi}_0' + x|_{0,\Gamma_1}}{|b(\bar{\varphi}_0 + \bar{\varphi}_L) - b(\bar{\varphi}_L)|_{0,\Gamma_1} + |b(\bar{\varphi}_L)|_{0,\Gamma_1}} \\
&\geq \frac{|\bar{\varphi}_0' + x|_{0,\Gamma_1}}{K|\bar{\varphi}_0|_{0,\Gamma_1} + |b(\bar{\varphi}_L)|_{0,\Gamma_1}} \\
&\geq \frac{|\bar{\varphi}_0' + x|_{0,\Gamma_1}}{KC_p|\bar{\varphi}_0'|_{0,\Gamma_1} + |b(\bar{\varphi}_L)|_{0,\Gamma_1}},
\end{aligned}$$

where C_p is the constant for Poincaré's Inequality. So

$$\begin{aligned}
\hat{T}(\bar{\varphi})^2 &\geq \frac{|\bar{\varphi}_0' + x|_{0,\Gamma_1}^2}{2K^2C_p^2|\bar{\varphi}_0'|_{0,\Gamma_1}^2 + 2|b(\bar{\varphi}_L)|_{0,\Gamma_1}^2} \\
&= \frac{|\bar{\varphi}_0' + x|_{0,\Gamma_1}^2}{C_1|\bar{\varphi}_0'|_{0,\Gamma_1}^2 + C_2} \\
&=: J(\bar{\varphi}_0) > 0,
\end{aligned}$$

where $C_1 = 2K^2C_p^2 > 0$, and $C_2 = 2|b(\bar{\varphi}_L)|_{0,\Gamma_1}^2 > 0$.

Let $\delta\bar{\varphi} \in H_0^1(\Gamma_1)$ be a perturbation function with $\delta\bar{\varphi}(0) = \delta\bar{\varphi}(1) = 0$. Then

$$\begin{aligned}
&J(\bar{\varphi}_0 + \delta\bar{\varphi}) - J(\bar{\varphi}_0) \\
&= \frac{|\bar{\varphi}_0' + x + \delta\bar{\varphi}'|_{0,\Gamma_1}^2}{C_1|\bar{\varphi}_0' + \delta\bar{\varphi}'|_{0,\Gamma_1}^2 + C_2} - \frac{|\bar{\varphi}_0' + x|_{0,\Gamma_1}^2}{C_1|\bar{\varphi}_0'|_{0,\Gamma_1}^2 + C_2} \\
&= \frac{|\bar{\varphi}_0' + x + \delta\bar{\varphi}'|_{0,\Gamma_1}^2(C_1|\bar{\varphi}_0'|_{0,\Gamma_1}^2 + C_2) - |\bar{\varphi}_0' + x|_{0,\Gamma_1}^2(C_1|\bar{\varphi}_0' + \delta\bar{\varphi}'|_{0,\Gamma_1}^2 + C_2)}{(C_1|\bar{\varphi}_0' + \delta\bar{\varphi}'|_{0,\Gamma_1}^2 + C_2)(C_1|\bar{\varphi}_0'|_{0,\Gamma_1}^2 + C_2)} \\
&= \frac{2\langle\bar{\varphi}_0' + x, \delta\bar{\varphi}'\rangle_{\Gamma_1}(C_1|\bar{\varphi}_0'|_{0,\Gamma_1}^2 + C_2) - 2C_1\langle\bar{\varphi}_0', \delta\bar{\varphi}'\rangle_{\Gamma_1}|\bar{\varphi}_0' + x|_{0,\Gamma_1}^2}{(C_1|\bar{\varphi}_0'|_{0,\Gamma_1}^2 + C_2)^2} + R(\bar{\varphi}_0', x, \delta\bar{\varphi}'),
\end{aligned}$$

where R is the remainder term of $O(|\delta\bar{\varphi}|_{1,\Gamma_1}^2)$. So the first-order variation of J is

$$\delta J = \frac{2\langle\bar{\varphi}_0', \delta\bar{\varphi}'\rangle_{\Gamma_1}(C_1|\bar{\varphi}_0'|_{0,\Gamma_1}^2 + C_2 - C_1|\bar{\varphi}_0' + x|_{0,\Gamma_1}^2)}{(C_1|\bar{\varphi}_0'|_{0,\Gamma_1}^2 + C_2)^2}.$$

Then the optimality condition $\delta J = 0$ is satisfied only in two cases: $\bar{\varphi}_0' = 0$ or $C_1|\bar{\varphi}_0'|_{0,\Gamma_1}^2 + C_2 = C_1|\bar{\varphi}_0' + x|_{0,\Gamma_1}^2$. For the first case, $\bar{\varphi}_0$ is a constant. But $\bar{\varphi}_0 \in H_0^1(\Gamma_1)$, so $\bar{\varphi}_0 = 0$. Then $J(0) = \frac{|x|^2}{C_2} > 0$. For the second case, $J(\bar{\varphi}_0) = \frac{1}{C_1} > 0$. Thus,

$$\hat{T}^2(\bar{\varphi}) \geq \min\{\frac{|x|^2}{C_2}, \frac{1}{C_1}\},$$

or more specifically,

$$\hat{T}(\bar{\varphi}) \geq C_{\hat{T}} := \min\{\frac{|x|}{\sqrt{2}|b(\bar{\varphi}_L)|^2_{0,\Gamma_1}}, \frac{1}{\sqrt{2}KC_p}\} > 0,$$

which proves the property. □

Step 2: *The existence of the minimizer of $\hat{S}(\bar{\varphi})$ in $\mathcal{A}_1$.*

Lemma 7.5.3 (Solution Existence)**.** *If the optimal integral time T^* for Problem II is finite, there exists at least one function $\varphi^* \in \mathcal{A}_T$ such that*

$$S_{T^*}(\varphi^*) = \min_{\substack{T\in\mathbb{R}^+,\\ \varphi\in\mathcal{A}_T}} S_T(\varphi) = \min_{\bar{\varphi}\in\mathcal{A}_1} \hat{S}(\bar{\varphi}).$$

Proof. We first prove the weakly lower semi-continuity of $\hat{S}(\bar{\varphi})$ in $H^1(\Gamma_1)$. Rewrite $\hat{S}(\bar{\varphi})$ by substituting (7.3.4). Then we get

$$\begin{aligned}\hat{S}(\bar{\varphi}) &= \frac{\hat{T}(\bar{\varphi})}{2}\int_0^1 \left|\hat{T}^{-1}(\bar{\varphi})\bar{\varphi}' - b(\bar{\varphi})\right|^2 dt \\ &= |\bar{\varphi}'|_{0,\Gamma_1}|b(\bar{\varphi})|_{0,\Gamma_1} - \langle\bar{\varphi}', b(\bar{\varphi})\rangle_{\Gamma_1}.\end{aligned}$$

For any sequence $\bar{\varphi}_k$ converging weakly to $\bar{\varphi}$ in $H^1(\Gamma_1)$, $\{\bar{\varphi}'_k\}$ is bounded in $L^2(\Gamma_1)$ and $\bar{\varphi}_k \to \bar{\varphi}$ in $L^2(\Gamma_1)$. Coupling with the global Lipschitz continuity of b, we can obtain

$$\lim_{k\to\infty} |b(\bar{\varphi}_k)|^2_{0,\Gamma_1} = |b(\bar{\varphi})|^2_{0,\Gamma_1},$$

$$\lim_{k\to\infty} \langle\bar{\varphi}'_k, b(\bar{\varphi}_k)\rangle_{\Gamma_1} = \langle\bar{\varphi}', b(\bar{\varphi})\rangle_{\Gamma_1}.$$

The weakly lower semicontinuity of $|\bar{\varphi}'|_{0,\Gamma_1}$ yields that

$$\liminf_{k\to\infty} |\bar{\varphi}'_k|_{0,\Gamma_1} \geq |\bar{\varphi}'|_{0,\Gamma_1}. \tag{7.5.2}$$

Combining the above results, we have

$$
\begin{aligned}
&\liminf_{k\to\infty} \hat{S}_k(\bar{\varphi}_k)\\
&= \liminf_{k\to\infty} (|\bar{\varphi}_k'|_{0,\Gamma_1}|b(\bar{\varphi}_k)|_{0,\Gamma_1} - \langle\bar{\varphi}_k', b(\bar{\varphi}_k)\rangle_{\Gamma_1})\\
&= \liminf_{k\to\infty} |\bar{\varphi}_k'|_{0,\Gamma_1}|b(\bar{\varphi}_k)|_{0,\Gamma_1} - \lim_{k\to\infty}\langle\bar{\varphi}_k', b(\bar{\varphi}_k)\rangle_{\Gamma_1}\\
&\geq |\bar{\varphi}'|_{0,\Gamma_1}|b(\bar{\varphi})|_{0,\Gamma_1} - \langle\bar{\varphi}', b(\bar{\varphi})\rangle_{\Gamma_1}\\
&= \hat{S}(\bar{\varphi}),
\end{aligned}
$$

that is, $\hat{S}(\bar{\varphi})$ is weakly lower semicontinuous in $H^1(\Gamma_1)$.

Now, we establish the coerciveness of $\hat{S}(\bar{\varphi})$. Since T^* is finite, there exists $M \in (T^*, \infty)$, such that

$$\inf_{\bar{\varphi}\in\mathcal{A}_1} \hat{S}(\bar{\varphi}) = \inf_{\substack{\bar{\varphi}\in\mathcal{A}_1,\\ \hat{T}(\bar{\varphi})<M}} \hat{S}(\bar{\varphi}).$$

In fact, by Lemma 7.3.4, a minimizing sequence $\{\bar{\varphi}_k\}$ of $\hat{S}(\bar{\varphi})$ defines a minimizing sequence $\{(\hat{T}(\bar{\varphi}_k), \bar{\varphi}_k)\}$ of $S(T, \bar{\varphi})$, which also corresponds to a minimizing sequence of $S_T(\varphi)$. The assumption of $T^* < \infty$ allows us to add the condition that $\sup_k \hat{T}(\bar{\varphi}_k) < M$. Otherwise, $\hat{T}(\bar{\varphi}_k)$ must go to infinity. The continuity of $S(T, \bar{\varphi})$ with respect to T yields that $T^* = \infty$, which contradicts our assumption that $T^* < \infty$. Now, let $\hat{T}^{-1}(\bar{\varphi})\bar{\varphi}'(s) - b(\bar{\varphi}(s)) = \bar{g}'(s)$. Then for any $\bar{\varphi} \in \mathcal{A}_1$ with $\hat{T}(\bar{\varphi}) < M$,

$$
\begin{aligned}
|\bar{\varphi}'| &\leq |\hat{T}(\bar{\varphi})||b(\bar{\varphi})| + |\hat{T}(\bar{\varphi})||g'|\\
&\leq M|b(\bar{\varphi})| + M|\bar{g}'|\\
&\leq MK|\bar{\varphi}| + M|b(0)| + M|\bar{g}'|\\
&\leq MK\int_0^s |\bar{\varphi}'(u)|\,du + M|b(0)| + M|\bar{g}'|.
\end{aligned}
$$

By Gronwall's Inequality, we have

$$|\bar{\varphi}'(s)| \leq \int_0^s M^2K(|\bar{g}'(u)| + |b(0)|)e^{KM(s-u)}\,du + M|b(0)| + M|\bar{g}'(s)|,$$

which yields that

$$|\bar{\varphi}'|_{0,\Gamma_1}^2 \leq C_1|b(0)|^2 + C_2|\bar{g}'|_{0,\Gamma_1}^2, \tag{7.5.3}$$

where $C_1, C_2 \in (0,\infty)$ only depend on M and K. So

$$\begin{aligned}
\hat{S}(\bar{\varphi}) &= \frac{\hat{T}(\bar{\varphi})}{2}\int_0^1 \left|\hat{T}^{-1}(\bar{\varphi})\bar{\varphi}'(s) - b(\bar{\varphi}(s))\right|^2 ds \\
&= \frac{\hat{T}(\bar{\varphi})}{2}|\bar{g}'|_{0,\Gamma_1}^2 \\
&\geq \frac{C_{\hat{T}}}{2}\left(\frac{1}{C_2}|\bar{\varphi}'|_{0,\Gamma_1}^2 - \frac{C_1}{C_2}|b(0)|^2\right),
\end{aligned}$$

where we used Property 7.5.2 in the last step. The coerciveness is proved.

For any minimizing sequence $\{\bar{\varphi}_k\}_{k=1}^\infty$ of $\hat{S}(\bar{\varphi})$, we have

$$\sup_k |\bar{\varphi}_k'|_{0,\Gamma_1} \leq \frac{2C_1}{C_T}|b(0)|^2 + \frac{2C_2}{C_T}\sup_k\{\hat{S}(\bar{\varphi}_k)\} < \infty.$$

Let $\bar{\varphi}_0 \in \mathcal{A}_1$. Then

$$\begin{aligned}
|\bar{\varphi}_k|_{0,\Gamma_1} &\leq |\bar{\varphi}_k - \bar{\varphi}_0|_{0,\Gamma_1} + |\bar{\varphi}_0|_{0,\Gamma_1} \\
&\leq C_p|\bar{\varphi}_k' - \bar{\varphi}_0'|_{0,\Gamma_1} + |\bar{\varphi}_0|_{0,\Gamma_1} < \infty,
\end{aligned}$$

by Poincaré's Inequality. Thus, $\{\bar{\varphi}_k\}_{k=1}^\infty$ is bounded in $H^1(\Gamma_1)$. Then there is a subsequence $\{\bar{\varphi}_{k_j}\}_{j=1}^\infty$ converging to some $\bar{\varphi}^* \in H^1(\Gamma_1)$ weakly in $H^1(\Gamma_1)$. So $\bar{\varphi}_{k_j} - \bar{\varphi}_0$ converges weakly to $\bar{\varphi}^* - \bar{\varphi}_0$ in $H_0^1(\Gamma_1)$. By Mazur's Theorem, $H_0^1(\Gamma_1)$ is weakly closed. So $\bar{\varphi}^* - \bar{\varphi}_0 \in H_0^1(\Gamma_1)$, and $\bar{\varphi}^* \in \mathcal{A}_1$. By Lemma 7.3.4, $\varphi^* \in \mathcal{A}_T$ corresponding to $\bar{\varphi}^* \in \mathcal{A}_1$ is a minimizer of $S_T(\varphi)$ and $T^* = \hat{T}(\bar{\varphi}^*)$. $\square$

Step 3: *The Γ-convergence of $\hat{S}_h$ to $\hat{S}$ as $h \to 0$.*

Lemma 7.5.4 (Γ-convergence of $\hat{S}_h$)**.** *Let $\{\mathcal{T}_h\}$ be a sequence of finite element meshes. For every $\bar{\varphi} \in \mathcal{A}_1$, the following two properties hold:*

- *Lim-inf inequality: for every sequence $\{\bar{\varphi}_h\}$ converging weakly to $\bar{\varphi}$ in $H^1(\Gamma_1)$, we have*

$$\hat{S}(\bar{\varphi}) \leq \liminf_{h\to 0} \hat{S}_h(\bar{\varphi}_h). \tag{7.5.4}$$

- *Lim-sup inequality: there exists a sequence* $\{\bar{\varphi}_h\} \subset \hat{\mathcal{B}}_h$ *converging weakly to* $\bar{\varphi}$ *in* $H^1(\Gamma_1)$*, such that*

$$\hat{S}(\bar{\varphi}) \geq \limsup_{h \to 0} \hat{S}_h(\bar{\varphi}_h). \tag{7.5.5}$$

Proof. We only need to consider a sequence $\{\bar{\varphi}_h\} \subset \mathcal{B}_h$ for the lim-inf inequality, since otherwise, the inequality is trivial. Similar to the proof of Lemma 7.5.3, rewrite the discretized action functional as

$$\begin{aligned}
\hat{S}_h(\bar{\varphi}_h) &= \frac{\hat{T}(\bar{\varphi}_h)}{2} \int_0^1 \left| \hat{T}^{-1}(\bar{\varphi}_h)\bar{\varphi}_h' - b(\bar{\varphi}_h) \right|^2 dt \\
&= |\bar{\varphi}_h'|_{0,\Gamma_1} |b(\bar{\varphi}_h)|_{0,\Gamma_1} - \langle \bar{\varphi}_h', b(\bar{\varphi}_h) \rangle_{\Gamma_1}.
\end{aligned}$$

By the same argument as in the proof of Lemma 7.4.4, we have

$$\begin{aligned}
\liminf_{h \to 0} |\bar{\varphi}_h'|_{0,\Gamma_1} &\geq |\bar{\varphi}'|_{0,\Gamma_1}, \\
\lim_{h \to 0} |b(\bar{\varphi}_h)|_{0,\Gamma_1} &= |b(\bar{\varphi})|_{0,\Gamma_1}, \\
\lim_{h \to 0} \langle \bar{\varphi}_h', b(\bar{\varphi}_h) \rangle_{\Gamma_1} &= \langle \bar{\varphi}', b(\bar{\varphi}) \rangle_{\Gamma_1}.
\end{aligned}$$

Combining these results, we have the lim-inf inequality. The lim-sup inequality can be obtained by the same argument as in the proof of Lemma 7.4.4 □

Step 4: *The proof of Theorem 7.5.1.*

Proof. Similar to the proof of Theorem 7.4.1, the only thing left is to verify the equi-coerciveness of $\hat{S}_h(\bar{\varphi}_h)$, which can be obtained directly from the coerciveness of $\hat{S}(\bar{\varphi})$ restricted onto $\bar{\mathcal{B}}_h \subset \mathcal{A}_1$ (see the second step in the proof of Lemma 7.5.3). So the proof is complete. □

7.6 Problem II with an Infinite T^*

In this case, the integration domain becomes the whole positive real line, which is not able to be rescaled to the unit interval. The geometric MAM (gMAM) [17] rescales this infinite domain by arc length by assuming, instead of finite transition time, finite total arc length. This way the zero-Hamiltonian constraint (7.3.1) can also remove the parameter T. However, since the Jacobian of the transform between time and arc length variables is singular at critical points, the numerical accuracy will deteriorate when unknown critical points exist along the MAP.

So we still use the reformulation in time. But we use a large but finite integration time to deal with the infinite T^*. We simplify the scenario, but reserve the numerical difficulties. Let $0 \in D$ be an asymptotically stable equilibrium point, D is contained in the basin of attraction of 0, and $\langle b(y), n(y)\rangle < 0$ for any $y \in \partial D$, where $n(y)$ is the exterior normal to the boundary ∂D. Then starting from any point in D, a trajectory of system (6.3.3) will converge to 0. We assume that the ending point x of Problem II is located on ∂D.

If we consider the change of variable in general, say $\alpha = \alpha(t)$, we have (see Lemma 3.1, Chapter 4 in [13])

$$S_T(\varphi) \geq S(\tilde{\varphi}) = \int_{\alpha(0)}^{\alpha(T)} (|\tilde{\varphi}'||b| - \langle \tilde{\varphi}', b\rangle) d\alpha, \tag{7.6.1}$$

where $\tilde{\varphi}(\alpha) = \varphi(t(\alpha))$, $\tilde{\varphi}'$ is the derivative with respect to α, and the equality holds when the zero-Hamiltonian constraint (7.3.1) is satisfied. In terms of α, the zero-Hamiltonian constraint can be written as

$$|\tilde{\varphi}'|\dot{\alpha}(t) = |b(\tilde{\varphi})|, \tag{7.6.2}$$

which yields

$$t = \int_0^{\alpha(t)} \frac{|\tilde{\varphi}'|}{|b(\tilde{\varphi})|} d\alpha. \tag{7.6.3}$$

If $|\tilde{\varphi}'| \equiv cnst$, the variable α just the rescaled arc length. Assuming that the length of the optimal curve is finite, we can rescale the total arc length to one, i.e., $\alpha(T) = 1$, which yields gMAM.

For any transition path $\tilde{\varphi}(\alpha) = \varphi(t(\alpha))$ that satisfies the zero-Hamiltonian constraint, if α is the arc length variable with $\tilde{\varphi}(0) = 0$, then $|\tilde{\varphi}'| = 1$. Let y be an arbitrary point on $\tilde{\varphi}$. Then the integration time from 0 to y is

$$t = \int_0^{\alpha_y} \frac{1}{|b(\tilde{\varphi})|} d\alpha,$$

where α_y is the arc length of the curve connecting 0 and y. Then the total arc length from the equilibrium 0 to y along $\bar{\varphi}$ is small if y is in a small neighborhood of 0. However, from

$$|b(\tilde{\varphi})| = |b(\tilde{\varphi}) - b(0)| \leq K|\tilde{\varphi}| \leq K\alpha,$$

we get

$$t = \int_0^{\alpha_y} \frac{1}{|b(\tilde{\varphi})|} d\alpha \geq \int_0^{\alpha_y} \frac{1}{K\alpha} d\alpha = \infty,$$

as long as $\alpha_y > 0$. Thus, $T^* = \infty$, since 0 is a critical point.

For clarity, we present the starting and ending points of a transition path in its notation. Let $\varphi^*_{y,x}$ denote the minimizer of Problem II with starting point y and ending point x, and $T^*_{y,x}$ the corresponding optimal integration time. Then for any y on $\varphi^*_{0,x}$, $T^*_{0,y} = \infty$ and $T^*_{y,x} < \infty$ whenever the path $\varphi^*_{y,x}$ has finite arc length.

We can construct a minimizing sequence for $\hat{S}(\varphi)$. In fact, let $\varphi^L_{0,y} = yt$ be the linear path from 0 to y in one time unit $T = 1$. Then

$$\begin{aligned} S_{T^*_{0,y}}(\varphi^*_{0,y}) \leq S_T(\varphi^L_{0,y}) &= \frac{1}{2}\int_0^1 |y - b(yt)|^2 \, dt \\ &\leq \int_0^1 (|y|^2 + K^2|y|^2t^2) \, dt \leq C(K)\rho^2, \end{aligned}$$

where $|y| \leq \rho$. So the action $S_{T^*_{0,y}}(\varphi^*_{0,y})$ decreases to zero with respect to ρ does, even if $T^*_{0,y} = \infty$ for any finite ρ. Consider a sequence of optimization problems

$$\hat{S}(\bar{\varphi}^{*,n}_{0,x}) = \inf_{\substack{\hat{T}(\bar{\varphi}) \leq n, \\ \bar{\varphi} \in \bar{\mathcal{A}}_1}} \hat{S}(\bar{\varphi}), \quad n = 1, 2, 3, \ldots, \tag{7.6.4}$$

by adding the constraint $\hat{T}(\bar{\varphi}) \leq n$. We have

Lemma 7.6.1. $\{\bar{\varphi}_{0,x}^{*,n}\}_{n=1}^{\infty}$ *is a minimizing sequence of* (7.3.9).

Proof. First of all, $\hat{S}(\varphi_{0,x}^{*,n})$ is decreasing as n increases. Choose ρ such that $x \notin B_\rho(0)$ and define $\rho_k = 2^{-k}\rho$, $k = 1, 2, \ldots$. Let y_k be the first intersection point of $\varphi_{0,x}^*$ and $B_{\rho_k}(0)$ when travelling along $\varphi_{0,x}^*$ from x to 0. Then $|y_k| = \rho_k$ and the optimal transition time $T_{y_k,x}^* < \infty$. For each k, we construct a path from 0 to x by:

$$\varphi_k = \begin{cases} \varphi_{0,y_k}^L, t \in [-T_{y_k,x}^* - 1, -T_{y_k,x}^*], \\ \varphi_{y_k,x}^*, t \in [-T_{y_k,x}^*, 0]. \end{cases}$$

Due to the additivity, we know that $\varphi_{y_k,x}^*$ is located on $\varphi_{0,x}^*$ since $y_k \in \varphi_{0,x}^*$. Then $\{(T_k = T_{y_k,x}^* + 1, \varphi_k)\}$ is a minimizing sequence as ρ_k decreases, and

$$S_{T_k}(\varphi_k) \leq S_{T_{0,x}^*}(\varphi_{0,x}^*) + C(K)\rho_k^2.$$

Consider $n = \lceil T_k \rceil$. We have

$$\hat{S}(\varphi_{0,x}^{*,n}) \leq S_{T_k}(\varphi_k) \leq S_{T_{0,x}^*}(\varphi_{0,x}^*) + C(K)\rho_k^2,$$

where the first inequality is because $\hat{T}(\varphi_k)$ is not equal to T_k in general. We then reach the conclusion. $\square$

When $T^* = \infty$, we have $\hat{T}(\bar{\varphi}) \to \infty$ as $\bar{\varphi}$ gets close to the minimizer, which implies $|\bar{\varphi}'|_{0,\Gamma_1} \to \infty$. Thus for this case, it is not able to conduct the convergence analysis in $H^1(\Gamma_1)$. So we use a larger space $\bar{C}_{x_1}^{x_2}(0, T)$, the space of absolutely continuous functions connecting x_1 and x_2 on $[0, T]$ with $0 < T \leq \infty$. To prove the convergence of the minimizing sequence in Lemma 7.6.1, we use the following proposition:

Proposition 7.6.2 (Proposition 2.1 in [17])**.** *Assume that the sequence* $\big((T_k, \varphi_k)\big)_{k \in \mathbb{N}}$ *with* $T_k > 0$ *and* $\varphi_k \in \bar{C}_{x_1}^{x_2}(0, T_k)$ *for every* $k \in \mathbb{N}$*, is a minimizing sequence of* (6.3.6) *and that the lengths of the curves of* φ_k *are uniformly bounded, i.e.*

$$\lim_{k\to\infty} S_{T_k}(\varphi_k) = V(x_1, x_2) \quad \textit{and} \quad \sup_{k\in\mathbb{N}} \int_0^{T_k} |\dot{\varphi}_k(t)|\, dt < \infty.$$

Then the the action functional $\hat{S}$ has a minimizer φ^*, and for some subsequence $(\varphi_{k_l})_{l\in\mathbb{N}}$ we have that

$$\lim_{l\to\infty} d(\varphi_{k_l}, \varphi^*) = 0,$$

where d denotes the Fréchet distance.

Theorem 7.6.3. *Assume that the lengths of the curves $\varphi_{0,x}^{*,n}$ are uniformly bounded. Then there exists a subsequence $\varphi_{0,x}^{*,n_l}$ that converges to a minimizer $\varphi^* \in \bar{C}_0^x(0,T)$ with respect to Fréchet distance.*

Proof. By Lemma 7.3.4, we have a one-to-one correspondence between $\{\bar{\varphi}_{0,x}^{*,n}\}_{n=1}^{\infty}$ and $\{\varphi_{0,x}^{*,n}\}_{n=1}^{\infty}$. So by Lemma 7.6.1, $\{(n, \varphi_{0,x}^{*,n})\}_{n=1}^{\infty}$ defines a minimizing sequence of Problem II. The theorem then is a direct application of Proposition 7.6.2. $\square$

Although we just constructed $\{\varphi\}_{0,x}^{*,n}$ for an exit problem around the neighborhood of an asymptotically stable equilibrium, it is easy to see that the idea can be applied to a global transition, say both x_1 and x_2 are asymptotically stable equilibrium, as long as there exist finitely many critical points along the MAP.

In Equation (7.6.4), we introduced an extra constraint $\hat{T}(\bar{\varphi}) \leq n$, which implies that the infimum may be reached at the boundary $\hat{T}(\bar{\varphi}) = n$. However, such a box-type constraint is not favorable from the optimization point of view. In the next lemma, we show that this constraint is not needed for the discrete problems, which is a key observation based on which tMAM is developed (See Section 7.2).

Lemma 7.6.4. *If $\bar{\varphi}_h^*$ is the minimizer of $\hat{S}_h(\bar{\varphi}_h)$ over $\overline{\mathcal{B}}_h$, then $\hat{T}(\bar{\varphi}_h^*) \leq C_h < \infty$, for any given h.*

Proof. Note that

$$\hat{T}^2(\bar{\varphi}_h^*) = \frac{|(\bar{\varphi}_h^*)'|_{0,\Gamma_1}^2}{|b(\bar{\varphi}_h^*)|_{0,\Gamma_1}^2} \leq C \frac{|\bar{\varphi}_h^*|_{0,\Gamma_1}^2}{|b(\bar{\varphi}_h^*)|_{0,\Gamma_1}^2},$$

where the last inequality is from the inverse inequality of finite element discretization and the constant C only depends on mesh [4]. Suppose $\hat{T}(\bar{\varphi}_h^*) = \infty$. Then either $|b(\bar{\varphi}_h^*)|_{0,\Gamma_1} = 0$ or $|\bar{\varphi}_h^*|_{0,\Gamma_1} = \infty$. The first case implies that $b(\bar{\varphi}_h^*(s)) = 0$ for all $s \in \Gamma_1$, which contradicts Assumption 7.1.1 (3). The second case implies that $\bar{\varphi}_h^*$ contradicts Lemma 7.1.2 due to the continuity. □

Lemma 7.6.4 means that for a discrete problem the constraint $\hat{T}(\bar{\varphi}) \leq n$ in Equation (7.6.4) is not necessary in the sense that there always exists a number n such that $\hat{T}(\bar{\varphi}_h^*) < n$. We can then consider a sequence $\{\overline{\mathcal{T}}_h\}$ of finite element meshes and treat the minimization of $\hat{S}_h(\bar{\varphi}_h)$ exactly in the same way as the case that T^* is finite. Simply speaking, $\{(\hat{T}(\bar{\varphi}_h^*), \bar{\varphi}_h^*)\}$ defines a minimizing sequence as $h \to 0$ no matter that T^* is finite or infinite. The only difference between $T^* < \infty$ and $T^* = \infty$ is that we address the convergence of $\bar{\varphi}_h^*$ in $H^1(\Gamma_1)$ for $T^* < \infty$ and in $\bar{C}_{x_1}^{x_2}$ for $T^* = \infty$.

7.7 A Priori Error Estimate for a Linear ODE System

In this section we apply our strategy to a linear ODE system with $b(x) = -Ax$, where A is a symmetric positive definite matrix. Then $x = 0$ is a global attractor. The E-L equation associated with $\hat{S}(\bar{\varphi})$ becomes

$$-\hat{T}^{-2}(\bar{\varphi})\bar{\varphi}'' + A^2\bar{\varphi} = 0, \tag{7.7.1}$$

which is a nonlocal elliptic problem of Kirchhoff type. For Problem I with a fixed T, i.e., $\hat{T}(\varphi) = T$, Equation (7.7.1) becomes a standard diffusion-reaction equation.

Let $\bar{\varphi}^* = \bar{\varphi}_0^* + \varphi_L \in \mathcal{A}_1$ be the minimizer of $\hat{S}(\bar{\varphi})$, and $\bar{\varphi}_h^* = \bar{\varphi}_{h,0}^* + \varphi_L \in \overline{\mathcal{B}}_h$ the minimizer of $\hat{S}_h(\bar{\varphi}_h)$, where $\varphi_L = x_1 + (x_2 - x_1)s$ is a linear function connecting x_1 and x_2 on Γ_1. Let

$$V_h = \{v : v|_I \text{ is affine on all } I \in \overline{\mathcal{T}}_h,\ v(0) = v(1) = 0\} \subset H_0^1(\Gamma_1).$$

For a fixed T, Equation (7.7.1) has a unique solution and the standard argument shows that

$$|\bar{\varphi}^* - \bar{\varphi}_h^*|_{1,\Gamma_1} = \left|\bar{\varphi}_0^* - \bar{\varphi}_{h,0}^*\right|_{1,\Gamma_1} \leq CT^2 \inf_{w \in V_h} |\bar{\varphi}_0^* - w|_{1,\Gamma_1}. \tag{7.7.2}$$

If T is large enough, Equation (7.7.1) can be regarded as a singularly perturbed problem, the best approximation given by a uniform mesh cannot reach the optimal convergence rate due to the existence of boundary layer.

We now consider Problem II with a finite T^*. The minimizer $\bar{\varphi}^*$ of $\hat{S}(\bar{\varphi})$ satisfies the weak form of Equation (7.7.1):

$$\left\langle (\bar{\varphi}^*)', v' \right\rangle_{\Gamma_1} = -\hat{T}^2 (\bar{\varphi}^*) \langle A\bar{\varphi}^*, Av \rangle_{\Gamma_1}, \quad \forall\, v \in H_0^1(\Gamma_1). \tag{7.7.3}$$

The minimizer $\bar{\varphi}_h^*$ of $\hat{S}_h(\bar{\varphi}_h)$ satisfies the discrete weak form:

$$\left\langle (\bar{\varphi}_h^*)', v' \right\rangle_{\Gamma_1} = -\hat{T}^2 (\bar{\varphi}_h^*) \langle A\bar{\varphi}_h^*, Av \rangle_{\Gamma_1}, \quad \forall\, v \in V_h. \tag{7.7.4}$$

We have the following a priori error estimate for Problem II with a finite T^*:

Proposition 7.7.1. *Consider a subsequence $\bar{\varphi}_h^*$ converging weakly to $\bar{\varphi}^*$ in $H^1(\Gamma_1)$ as $h \to 0$. Assume that $\bar{\varphi}^*$ and $\bar{\varphi}_h^*$ satisfy Equations (7.7.3) and (7.7.4), respectively. For problem II with a finite T^*, there exists a constant $C \sim (T^*)^2$ such that*

$$|\bar{\varphi}^* - \bar{\varphi}_h^*|_{1,\Gamma_1} = \left|\bar{\varphi}_0^* - \bar{\varphi}_{h,0}^*\right|_{1,\Gamma_1} \leq C \inf_{w \in V_h} |\bar{\varphi}_0^* - w|_{1,\Gamma_1}, \tag{7.7.5}$$

when h is small enough.

Proof. Let η be the best approximation of $\bar{\varphi}^*$ on $V_h \oplus \varphi_L$, i.e.,

$$|\bar{\varphi}^* - \eta|_{1,\Gamma_1} = \inf_{w \in V_h \oplus \varphi_L} |\bar{\varphi}^* - w|_{1,\Gamma_1}.$$

Then we have

$$\langle (\bar{\varphi}^* - \eta)', w' \rangle = 0, \quad \forall\, w \in V_h,$$

where $\bar{\varphi}^* - \eta \in H_0^1(\Gamma_1)$. Consider

$$
\begin{aligned}
|\bar{\varphi}_h^* - \eta|_{1,\Gamma_1}^2 &= \langle (\bar{\varphi}_h^* - \bar{\varphi}^*)', (\bar{\varphi}_h^* - \eta)' \rangle_{\Gamma_1} + \langle (\bar{\varphi}^* - \eta)', (\bar{\varphi}_h^* - \eta)' \rangle_{\Gamma_1} \\
&= \langle (\bar{\varphi}_h^* - \bar{\varphi}^*)', (\bar{\varphi}_h^* - \eta)' \rangle_{\Gamma_1} \\
&= - \left\langle (\hat{T}^2(\bar{\varphi}_h^*)\bar{\varphi}_h^* - \hat{T}^2(\bar{\varphi}^*)\bar{\varphi}^*), A^2(\bar{\varphi}_h^* - \eta) \right\rangle_{\Gamma_1} \\
&= - \left\langle (\hat{T}^2(\bar{\varphi}_h^*)\bar{\varphi}_h^* - \hat{T}^2(\eta)\eta), A^2(\bar{\varphi}_h^* - \eta) \right\rangle_{\Gamma_1} \\
&\quad - \left\langle (\hat{T}^2(\eta)\eta - \hat{T}^2(\bar{\varphi}^*)\bar{\varphi}^*), A^2(\bar{\varphi}_h^* - \eta) \right\rangle_{\Gamma_1} \\
&= I_1 + I_2.
\end{aligned} \tag{7.7.6}
$$

We look at I_2 first. Note that

$$
\begin{aligned}
&|\hat{T}^2(\eta) - \hat{T}^2(\bar{\varphi}^*)| \\
&= \left| \frac{|\eta|_{1,\Gamma_1}^2}{|A\eta|_{0,\Gamma_1}^2} - \frac{|\bar{\varphi}^*|_{1,\Gamma_1}^2}{|A\bar{\varphi}^*|_{0,\Gamma_1}^2} \right| \\
&= \left| \frac{|\eta|_{1,\Gamma_1}^2 - |\bar{\varphi}^*|_{1,\Gamma_1}^2}{|A\eta|_{0,\Gamma_1}^2} + \frac{|\bar{\varphi}^*|_{1,\Gamma_1}^2 (|A\bar{\varphi}^*|_{0,\Gamma_1}^2 - |A\eta|_{0,\Gamma_1}^2)}{|A\eta|_{0,\Gamma_1}^2 |A\bar{\varphi}^*|_{0,\Gamma_1}^2} \right| \\
&\leq C_{\hat{T}}(\eta, \bar{\varphi}^*) |\eta - \bar{\varphi}^*|_{1,\Gamma_1},
\end{aligned} \tag{7.7.7}
$$

where

$$
C_{\hat{T}}(\eta, \bar{\varphi}^*) = \frac{|\eta|_{1,\Gamma_1} + |\bar{\varphi}^*|_{1,\Gamma_1}}{|A\eta|_{0,\Gamma_1}^2} + \frac{|\bar{\varphi}^*|_{1,\Gamma_1}^2 (|A\eta|_{0,\Gamma_1} + |A\bar{\varphi}^*|_{0,\Gamma_1}) \|A\| C_p}{|A\eta|_{0,\Gamma_1}^2 |A\bar{\varphi}^*|_{0,\Gamma_1}^2},
$$

and C_p is the Poincaré constant. Then we have

$$
\begin{aligned}
|I_2| &= \left| \left\langle (\hat{T}^2(\eta)\eta - \hat{T}^2(\bar{\varphi}^*)\bar{\varphi}^*), A^2(\bar{\varphi}_h^* - \eta) \right\rangle_{\Gamma_1} \right| \\
&\leq \left| \left\langle \hat{T}^2(\eta)(\eta - \bar{\varphi}^*), A^2(\bar{\varphi}_h^* - \eta) \right\rangle_{\Gamma_1} \right| \\
&\quad + \left| \left\langle (\hat{T}^2(\eta) - \hat{T}^2(\bar{\varphi}^*))\bar{\varphi}^*, A^2(\bar{\varphi}_h^* - \eta) \right\rangle_{\Gamma_1} \right| \\
&\leq (\hat{T}^2(\eta) C_p^2 \|A\|^2 + C_{\hat{T}}(\eta, \bar{\varphi}^*) |A\bar{\varphi}^*|_{0,\Gamma_1} \|A\| C_p) |\eta - \bar{\varphi}^*|_{1,\Gamma_1} |\bar{\varphi}_h^* - \eta|_{1,\Gamma_1} \\
&= C_{I_2}(\eta, \bar{\varphi}^*) |\eta - \bar{\varphi}^*|_{1,\Gamma_1} |\bar{\varphi}_h^* - \eta|_{1,\Gamma_1}.
\end{aligned} \tag{7.7.8}
$$

By the definition of η, we have

$$\lim_{h\to 0} |\eta|_{1,\Gamma_1} = |\bar{\varphi}^*|_{1,\Gamma_1}, \quad \lim_{h\to 0} |A\eta|_{0,\Gamma_1} = |A\bar{\varphi}^*|_{0,\Gamma_1}.$$

We then have

$$\lim_{h\to 0} C_{I_2}(\eta, \bar{\varphi}^*) = 2M + 3M^2,$$

where $M = \|A\| C_p T^*$.

We now look at I_1. Since $\lim_{h\to 0} \hat{T}(\bar{\varphi}_h^*) = \lim_{h\to 0} \hat{T}(\eta) = T^*$ and $T^* < \infty$, we know that when h is small enough, $I_1 \sim -(T^*)^2 \left\langle (\bar{\varphi}_h^* - \eta), A^2(\bar{\varphi}_h^* - \eta) \right\rangle_{\Gamma_1} < 0$. Combining this fact with Equation (7.7.6) and (7.7.8), we have that for h small enough there exists a constant $C > 2M + 3M^2$ such that

$$\left|\bar{\varphi}_h^* - \eta\right|_{1,\Gamma_1} \leq C \left|\bar{\varphi}^* - \eta\right|_{1,\Gamma_1} = C \inf_{w \in V_h \oplus \varphi_L} \left|\bar{\varphi}^* - w\right|_{1,\Gamma_1}.$$

The proof is complete. □

To this end, we obtain a similar a priori error estimate to that for Problem I with a fixed T. Since T^* can be arbitrarily large, we know that the optimal convergence rate may degenerate when a boundary layer exists. Using Proposition 7.7.1, we can easily obtain the optimal convergence rate with respect to the error of action functional:

$$|\hat{S}(\bar{\varphi}^*) - \hat{S}(\bar{\varphi}_h^*)| \sim |\delta^2 \hat{S}(\bar{\varphi}^*)| \sim |\bar{\varphi}_h^* - \bar{\varphi}^*|_{1,\Gamma_1}^2, \tag{7.7.9}$$

where the second-order variation can be obtained with respect to the perturbation function $\delta\bar{\varphi} = \bar{\varphi}^* - \bar{\varphi}_h^*$.

7.8 Numerical Experiments

We will use the following simple linear stochastic ODE system to demonstrate our analysis results

$$dX(t) = AX(t)\,dt + \sqrt{\varepsilon}\,dW(t), \tag{7.8.1}$$

where

$$A = B^{-1}JB = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} a & b \\ -b & a \end{bmatrix},$$

with $a = 1/3$, $b = \sqrt{8}/3$, $\lambda_1 = -10$, and $\lambda_2 = -2$. Then $z = (0,0)^{\mathsf{T}}$ is a stable fixed point. For the corresponding deterministic system, namely when $\varepsilon = 0$, and any given point $X(0) = x \neq z$ in the phase space, the trajectory $X(t) = e^{tA}x$ converges to z as $t \to \infty$. When noise exists, this trajectory is also the minimal action path φ^* from x to $e^{tA}x$ with $T^* = t$, since $V(x, e^{tA}x) = 0$. Moreover, if the ending point is z, $T^* = \infty$. This obviously is not an exit problem, which is a typical application of MAM. However, it includes most of the numerical difficulties of MAM, and the trajectory can serve as an exact solution, which simplifies the discussions.

Consider the minimal action path from x $(\neq z)$ to $e^{tA}x$ such that $T^* = t$. Since the minimal action path corresponds to a trajectory, we can use the value of action functional as the measure of error with an optimal rate $O(h^2) \sim O(N^{-2})$ (see Equation (7.7.9)), where N is the number of elements. We will look at the following two cases:

(i). T^* is finite and small. According to Theorem 7.5.1, $\bar{\varphi}_h$ converges to $\bar{\varphi}^*$. Since T^* is small, according to Proposition 7.7.1, we expect optimal convergence rate of $\bar{\varphi}_h$ as $h \to 0$.

(ii). $T^* = \infty$. We will compare the convergence behavior between tMAM and MAM with a fixed large T. According to Theorem 7.6.3, we have the convergence in $\bar{C}_x^z$. However, the convergence of this case is similar to that for a finite but large T^*, where we expect a deteriorated convergence rate.

Case (i):

Let $x = (1,1)$. We use $e^{A}x$ as the ending point such that $T^* = 1$. In figure 7.1 we plot the convergence behavior of tMAM with uniform linear finite element

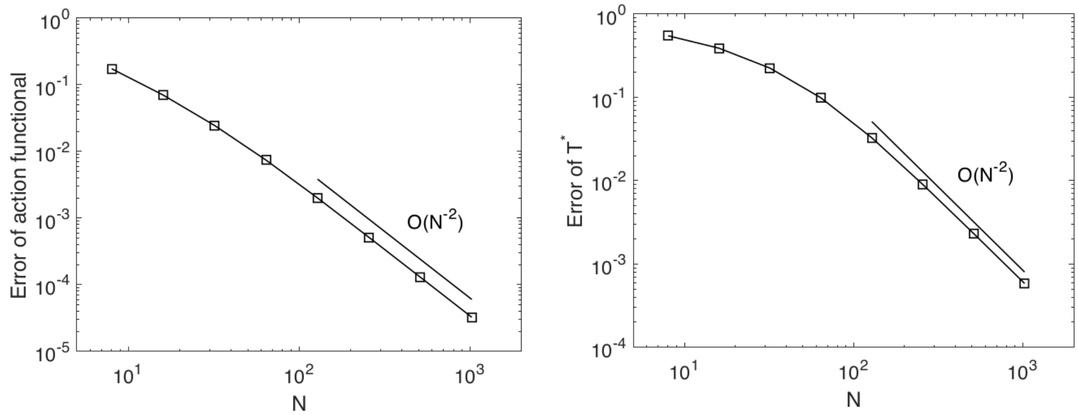


FIGURE 7.1. [32] Convergence behavior of tMAM for Case (i). Left: errors of action functional; Right: errors of T^*.

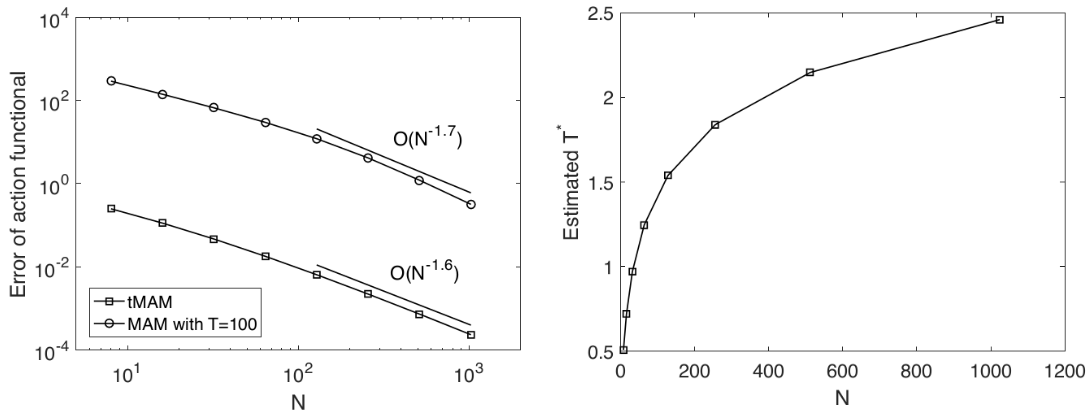


FIGURE 7.2. [32] Convergence behavior of tMAM and MAM with a fixed T for Case (ii). Left: errors of action functional; Right: estimated T^* of tMAM, i.e., $\hat{T}(\bar{\varphi}_h^*)$.

discretization. It is seen that the optimal convergence rate is reached for both action functional and T^* estimated by $\hat{T}(\bar{\varphi}_h)$.

Case (ii):

For this case, we still use $x = (1, 1)$ as the starting point. The ending point is chosen as $a = (0, 0)^\mathsf{T}$ such that $T^* = \infty$. Except tMAM, we use MAM with a fixed T to approximate this case, where T is supposed to be large. In general, we do not have a criterion to define how large is enough because the accuracy is affected by two competing issues: 1) The fact that $T^* = \infty$ favors a large T; but 2) a fixed discretization favors a small T. This implies that for any given h, an "optimal" finite T exists. For the purpose of demonstration, we choose $T = 100$,

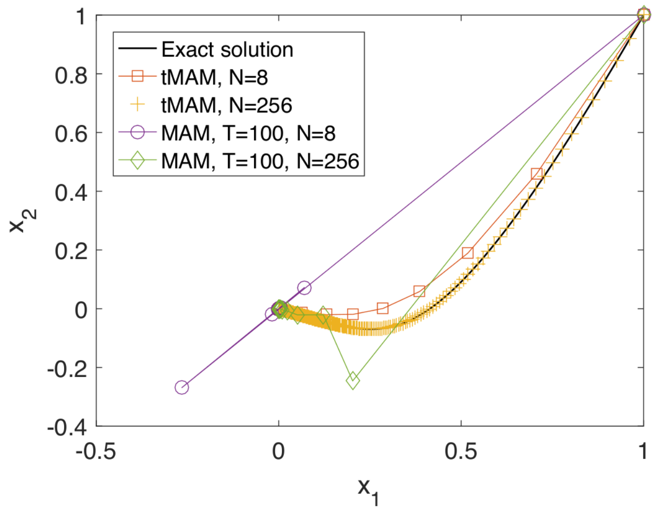


FIGURE 7.3. [32] Approximate MAPs given by tMAM and MAM with a fixed T for Case (ii).

which is actually too large from the accuracy point of view. Let $\varphi_T^*(t)$ be the approximate MAP given by MAM with a fixed T. We know that $\bar{\varphi}_T^*(s) = \varphi_T^*(t/T)$ yields a smaller action with the integration time $\hat{T}(\bar{\varphi}_T^*(s))$. In this sense, no matter what T is chosen, for the same discretization tMAM will always provide a better approximation than MAM with a fixed T. The reason we use an overlarge T is to demonstrate the deterioration of convergence rate. In figure 7.2, we plot the convergence behavior of tMAM and MAM with $T = 100$ on the left, and the estimated T^* given by tMAM on the right. It is seen that the convergence is slower than $O(N^{-2})$ as we have analyzed in section 7.7. For the same discretization, tMAM has an accuracy that is several orders of magnitude better than MAM with $T = 100$. In the right plot of figure 7.2, we see that the optimal integration time for a certain discretization is actually not large at all. This implies that MAM with a fixed T for Case (ii) is actually not very reliable. In figure 7.3, we compare the MAPs given by tMAM and MAM with the exact solution $e^{tA}x$, where all symbols indicate the nodes of finite element discretization. First of all, we note that the number of effective nodes in MAM is small because of the scale separation of fast dynamics and small dynamics. Most nodes are clustered around the fixed point. This is called a problem of clustering (see [25, 33] for the discussion of this issue). Second, if the chosen T is too large, oscillation is observed in the paths given by MAM especially when the resolution is relatively low; on the other hand, tMAM does not suffer such an oscillation by adjusting the integration time according to the resolution. Third, although tMAM is able to provide a good approximation even with a coarse discretization, more than enough nodes are put into the region around the fixed point, which corresponds to the deterioration of convergence rate. To recover the optimal convergence rate, we need to resort to adaptivity (see [28, 33] for the construction of the algorithm).

Chapter 8

Temporal Minimum Action Method for Time Delayed System

8.1 Introduction of Large Deviation Principle for Stochastic Differential Delay Equations

In reality, some systems in physics, chemistry, biology, medicine, economics, engineering, etc., depend not only on the current state, but also on the past states (see e.g. [21]). In addition to the properties of transitions as rare events in stochastic delayed systems, the study of LDP provides us insight of the impact of delays on the systems.

The delay of a system can have memorial influence on the system in different ways, for example, discrete delays at finite previous points and distributed delay along a previous time interval. In this chapter, we will apply tMAM to the system with discrete delay. Without loss of generality, we consider delay at one previous point, i.e., the stochastic differential delay equation (SDDE)

$$\begin{cases} dX_t = b(X_t, X_{t-\tau})\,dt + \sqrt{\varepsilon}\,dW_t, & t \in (0, T], \\ X(t) = f(t), & t \in [-\tau, 0], \end{cases} \tag{8.1.1}$$

where τ is a positive constant, $b(x, y)$ is Lipschitz continuous with respect to both x and y, and $X_t \in \mathbb{R}^n$. The solution of a time delayed system is not uniquely defined by the sole knowledge of the pointwise initial condition at $t = 0$ but by a functional initial condition $f(\cdot)$ defined over the interval $[-\tau, 0]$ [16]. In some literature, this is also referred to as a memory effect. Due to the dependence on a function instead of a point, Equation (8.1.1) is not a finite-dimensional system, but an infinite-dimensional one.

The large deviation principle for this SDDE is proved in [22] with the action functional

$$S_T(\varphi) = \frac{1}{2}\int_0^T |\varphi'(t) - b(\varphi(t), \varphi(t-\tau))|^2 \, dt. \tag{8.1.2}$$

To apply tMAM, we need to scale the time linearly, then rewrite $S_T(\varphi)$ as

$$S(T, \bar{\varphi}) = \frac{T}{2}\int_0^1 |T^{-1}\bar{\varphi}'(s) - b(\bar{\varphi}(s), \bar{\varphi}(s-\tau/T))|^2 \, dt, \tag{8.1.3}$$

and seek the optimal linear time scaling for φ, where $\bar{\varphi}(s)$ is the linearly rescaled path of φ on $[0,1]$. However, we notice that comparing to a simple quadratic equation as in tMAM for systems without delays, T is coupled with the delay such that the optimality condition $\partial_T S = 0$ yields a complicated nonlinear equation

$$\partial_T S = \frac{1}{2}\langle b, b\rangle - \frac{1}{2}T^{-2}\langle \bar{\varphi}'_s, \bar{\varphi}'_s\rangle - \langle \nabla_2 b \bar{\varphi}'_{\hat{s}} T^{-1}\tau, T^{-1}\bar{\varphi}'_s - b\rangle = 0, \tag{8.1.4}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in $L^2([0,1])$, $\bar{\varphi}_{s-\tau/T} = \bar{\varphi}_{\hat{s}}$ and $\nabla_2 b$ indicates the gradient of b with respect to $\bar{\varphi}_{\hat{s}}$. Although a root-finding algorithm is always possible, we expect to control T more explicitly, which is important for the development and understanding of the algorithm.

Note that if there is no delay, i.e., $\tau = 0$, we have only one positive solution

$$T = \frac{\langle \bar{\varphi}'_s, \bar{\varphi}'_s\rangle^{1/2}}{\langle b, b\rangle^{1/2}}, \tag{8.1.5}$$

which is the same to the result in Chapter 7 as expected. We will use this rescaling time for our penalty method introduced in the next section, and we will see at the end of Section 8.4 that it is a good approximation for the exact optimal time T^*.

8.2 Penalty Method

We investigate the penalty method by introducing an auxiliary unknown path $\bar{\psi}$ for the delayed function $\bar{\varphi}(s-\tau/T)$. The consistency between $\varphi(t)$ and $\varphi(t-\tau)$ will be maintained through a penalty term. This strategy will give us a more explicit control on T.

We rewrite the action functional (8.1.3) as

$$S(T,\bar{\varphi}_s,\bar{\psi}_s) = \frac{T}{2}\int_0^1 |T^{-1}\bar{\varphi}'_s - b(\bar{\varphi}_s,\bar{\psi}_s)|^2\,ds, \tag{8.2.1}$$

where we replace $\bar{\varphi}_{s-\tau/T}$ by $\bar{\psi}_s$. We introduce a point-wise constraint that $\bar{\psi}_s = \bar{\varphi}_{\hat{s}} = \bar{\varphi}(s-\tau/T)$. We will use the penalty method to deal with this constraint. In other words, we define

$$\hat{S}(T,\bar{\varphi}_s,\bar{\psi}_s) = \frac{T}{2}\int_0^1 |T^{-1}\bar{\varphi}'_s - b(\bar{\varphi}_s,\bar{\psi}_s)|^2\,ds + \frac{\beta^2 T}{2}\int_0^1 |\bar{\psi}_s - \bar{\varphi}_{\hat{s}}|^2\,ds, \tag{8.2.2}$$

where $0 \neq \beta \in \mathbb{R}$. This way, we make $\bar{\varphi}_s$ and $\bar{\psi}_s$ independent and link them through the penalty term.

Remark 8.2.1. In fact, there are two kinds of penalty term:

$$\frac{\beta^2}{2}\int_0^T |\psi_t - \varphi_{t-\tau}|^2\,dt = \frac{\beta^2 T}{2}\int_0^1 |\bar{\psi}_s - \bar{\varphi}_{\hat{s}}|^2\,ds \quad \text{and} \quad \frac{\beta^2}{2}\int_0^1 |\bar{\psi}_s - \bar{\varphi}_{\hat{s}}|^2\,ds.$$

Here in this dissertation, we use the first one.

Assuming that $\bar{\varphi}_s$ and $\bar{\psi}_s$ in Equation (8.2.1) are independent, we can obtain an optimal linear time scaling in a similar way to the delay-free case. Taking the partial derivative of (8.2.1) with respect to the time variable T, we have

$$\frac{\partial S(T,\bar{\varphi}_s,\bar{\psi}_s)}{\partial T} = -\frac{1}{2T^2}\int_0^1 |\bar{\varphi}'_s|^2\,ds + \frac{1}{2}\int_0^1 |b(\bar{\varphi}_s,\bar{\psi}_s)|^2\,ds = 0.$$

Then we get the optimal time $\hat{T}$ of (8.2.1) as a functional of $\bar{\varphi}_s$ and $\bar{\psi}_s$

$$\hat{T}(\bar{\varphi}_s,\bar{\psi}_s) = \frac{\sqrt{\int_0^1 |\bar{\varphi}'_s|^2\,ds}}{\sqrt{\int_0^1 |b(\bar{\varphi}_s,\bar{\psi}_s)|^2\,ds}} = \frac{\|\bar{\varphi}'\|_{L^2([0,1])}}{\|b(\bar{\varphi},\bar{\psi})\|_{L^2([0,1])}}. \tag{8.2.3}$$

Substituting this optimal time to (8.2.2), we get the TMAM action functional on $L^2([0,1])$

$$\begin{aligned}
&\hat{S}(\bar{\varphi}_s, \bar{\psi}_s) \\
&= \frac{\hat{T}(\bar{\varphi}_s, \bar{\psi}_s)}{2} \int_0^1 |\hat{T}(\bar{\varphi}_s, \bar{\psi}_s)^{-1} \bar{\varphi}_s' - b(\bar{\varphi}_s, \bar{\psi}_s)|^2 \, ds + \frac{\beta^2 \hat{T}(\bar{\varphi}_s, \bar{\psi}_s)}{2} \int_0^1 |\bar{\psi}_s - \bar{\varphi}_{\hat{s}}|^2 \, ds \\
&= \|\bar{\varphi}'\| \|b(\bar{\varphi}, \bar{\psi})\| - \langle \bar{\varphi}', b(\bar{\varphi}, \bar{\psi}) \rangle + \frac{\beta^2 \|\bar{\varphi}'\|}{2\|b(\bar{\varphi}, \bar{\psi})\|} \|\bar{\psi}_s - \bar{\varphi}_{\hat{s}}\|^2,
\end{aligned} \tag{8.2.4}$$

where $\|\cdot\|$ denotes the norm in $L^2([0,1])$, and $\hat{s} = s - \tau / \hat{T}(\bar{\varphi}_s, \bar{\psi}_s)$ is a functional of $\bar{\varphi}_s$ and $\bar{\psi}_s$.

Now, we have reformulated the optimization problem

$$\min_{T>0} \min_{\substack{\varphi(t) \in H^1([0,T]), \\ \varphi(0)=x_0, \varphi(T)=x_1}} S_T(\varphi(t)) \tag{8.2.5}$$

into a new optimization problem

$$\min_{\substack{\bar{\varphi}(s), \bar{\psi}(s) \in H^1([0,1]), \\ \bar{\varphi}(0)=x_0, \bar{\varphi}(1)=x_1}} \hat{S}(\bar{\varphi}(s), \bar{\psi}(s)). \tag{8.2.6}$$

8.3 Variational Calculus with Delay for Finite Element Approximation

As discussed in [32], the Euler–Lagrange equation of this kind of action functional is a high-order nonlinear and singular PDE, which is very difficult to solve. So we still consider the minimization problem itself directly and use a variational approach. In [32], we gave a finite element approximation framework and provide the analysis of its convergence. We showed that this framework is efficient for a system without delay no matter that the optimal transition time is finite or infinite. We still use the same framework for discretizing the action functional (8.1.2) and its reformulation (8.2.4).

Suppose that we have a finite dimensional approximation space $V_h \subset H^1([0,1])$ spanned by $\{\bar{\varphi}_i(s), \bar{\psi}_l(s); 1 \le i \le M, 1 \le l \le N\}$ such that

$$\bar{\varphi}(s) = \sum_{i=1}^{M} x_i \bar{\varphi}_i(s), \quad \bar{\psi}(s) = \sum_{l=1}^{N} y_l \bar{\psi}_l(s),$$

where $x_i, y_l \in \mathbb{R}^n$. We then have the discretized version of Problem (8.2.6)

$$\min_{\substack{\bar{\varphi}(s), \bar{\psi}(s) \in V_h, \\ \bar{\varphi}(0)=x_0, \bar{\varphi}(1)=x_1}} \hat{S}(\bar{\varphi}(s), \bar{\psi}(s)). \tag{8.3.1}$$

Since most modern optimization algorithms require the gradient, we now compute the gradient of the discretized action functional in (8.3.1) using a variational approach. Let

$$\delta\bar{\varphi}(t) = \sum_{i=1}^{M} \delta x_i \bar{\varphi}_i(s), \quad \delta\bar{\psi}(s) = \sum_{l=1}^{N} \delta y_l \bar{\psi}_l(s),$$

where $\delta x_i, \delta y_i \in \mathbb{R}^n$. Then

$$\begin{aligned}\delta\hat{S}(\bar{\varphi}, \bar{\psi}) &= \langle \frac{\delta \hat{S}}{\delta \bar{\varphi}}, \delta\bar{\varphi} \rangle + \langle \frac{\delta \hat{S}}{\delta \bar{\psi}}, \delta\bar{\psi} \rangle \\ &= \sum_{i=1}^{M} \langle \frac{\delta \hat{S}}{\delta \bar{\varphi}}, \delta x_i \bar{\varphi}_i(s) \rangle + \sum_{l=1}^{N} \langle \frac{\delta \hat{S}}{\delta \bar{\psi}}, \delta y_l \bar{\psi}_l(s) \rangle.\end{aligned}$$

Consider a particular choice of $\delta\bar{\varphi}$, whose coefficients are equal to zero except the jth component $\delta x_{i,j}$ of δx_i. Then we have

$$\delta\hat{S} = \langle \frac{\delta \hat{S}}{\delta \bar{\varphi}}, \bar{\varphi}_i(s) \mathbf{e}_j \rangle \delta x_{i,j},$$

which implies that

$$(\nabla \hat{S}(\bar{\varphi}, \bar{\psi}))_{k(i,j)} = \frac{\partial \hat{S}}{\partial x_{i,j}} = \langle \frac{\delta \hat{S}}{\delta \bar{\varphi}}, \bar{\varphi}_i(s) \mathbf{e}_j \rangle, \tag{8.3.2}$$

where $\mathbf{e}_j$ is the unit vector in $\mathbb{R}^n$ with its jth component being 1 and the rest being 0, and $k(i,j)$ is a global index for the variable $\bar{\varphi}$ uniquely determined by $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, n$. The same is for $\frac{\partial \hat{S}}{\partial y_{l,j}}$:

$$(\nabla \hat{S}(\bar{\varphi}, \bar{\psi}))_{\tilde{k}(l,j)} = \frac{\partial \hat{S}}{\partial y_{l,j}} = \langle \frac{\delta \hat{S}}{\delta \bar{\psi}}, \bar{\psi}_l(s) \mathbf{e}_j \rangle, \tag{8.3.3}$$

where $\tilde{k}(l,j)$ is a global index for the variable $\bar{\psi}$.

Now, we introduce the finite element approximation space. Note that the variable $\bar{\varphi} \in H^1([0,1])$ is subject to the boundary condition $\bar{\varphi}(0) = x_0$ and $\bar{\varphi}(1) = x_1$. Then $\bar{\varphi}(s) - l(s) \in H_0^1([0,1])$, where $l(s) = x_0 + s(x_1 - x_0)$ is the linear path connecting x_0 and x_1. The variable $\bar{\psi}$ that replaces the delay part of $\bar{\varphi}$ doesn't need to satisfy the boundary condition. So $\bar{\varphi} \in H_0^1([0,1]) + l$ and $\psi \in H^1([0,1])$. While the higher order techniques can be considered, we use the linear finite elements here. Let

$$\mathcal{T}_h : 0 = s_0 < s_1 < \cdots < s_N = 1$$

be a partition of the interval $[0,1]$. We define the linear finite element spaces

$$U_h \subset H_0^1([0,1]), \quad W_h \subset H^1([0,1]),$$

and the admissible set

$$V_h = (U_h \times W_h) + (l, 0)$$

for $(\bar{\varphi}, \bar{\psi})$. Let $\{\bar{\varphi}_i\}_{i=1}^{N-1}$ be the basis for U_h and $\{\bar{\psi}_i\}_{i=1}^{N+1}$ be the basis for W_h. Then the variables $(\bar{\varphi}, \bar{\psi})$ can be approximated as

$$(\bar{\varphi}, \bar{\psi}) \approx (\bar{\varphi}_h, \bar{\psi}_h) = (\sum_{i=1}^{N-1} x_i \bar{\varphi}_i + l, \sum_{i=1}^{N+1} y_i \bar{\psi}_i).$$

Then the discretized action functional in (8.3.1) has the form

$$\begin{aligned}
&\hat{S}(\bar{\varphi}_h, \bar{\psi}_h) \\
&= \|\bar{\varphi}_h'\| \|b(\bar{\varphi}_h, \bar{\psi}_h)\| - \langle \bar{\varphi}_h', b(\bar{\varphi}_h, \bar{\psi}_h) \rangle + \frac{\beta^2 \|\bar{\varphi}_h'\|}{2\|b(\bar{\varphi}_h, \bar{\psi}_h)\|} \|\bar{\psi}_h(s) - \bar{\varphi}_h(s - \frac{\tau}{\hat{T}(\bar{\varphi}_h, \bar{\psi}_h)})\|^2.
\end{aligned}$$

8.4 Computation of the Gradient of $\hat{S}$

Let $(\bar{\varphi}_h^*, \bar{\psi}_h^*) = (\sum_{i=1}^{N-1} x_i^* \bar{\varphi}_i(s) + l(s), \sum_{l=1}^{N+1} y_l^* \bar{\psi}_l(s))$ be the finite element approximation of the minimizer $(\bar{\varphi}^*, \bar{\psi}^*)$ for the minimal action path $\bar{\varphi}^*$. Then the finite element coefficients $x_{i,j}$ and $y_{l,j}$ will satisfy the problem

$$\hat{S}(\bar{\varphi}_h^*, \bar{\psi}_h^*) = \min_{(\bar{\varphi}_h, \bar{\psi}_h) \in V_h} \hat{S}(\bar{\varphi}_h, \bar{\psi}_h) = \min_{x_{i,j}, y_{l,j} \in \mathbb{R}} \hat{S}(x_{i,j}, y_{l,j}). \tag{8.4.1}$$

Now, we derive the partial derivatives $\partial \hat{S}/\partial x_{i,j}$ and $\partial \hat{S}/\partial y_{l,j}$. First, we have

$$b(\bar{\varphi}+\delta\bar{\varphi},\bar{\psi}+\delta\bar{\psi}) = b(\bar{\varphi},\bar{\psi}) + \nabla_1 b(\bar{\varphi},\bar{\psi})\delta\bar{\varphi} + \nabla_2 b(\bar{\varphi},\bar{\psi})\delta\bar{\psi} + O(\|\delta\bar{\varphi}\|^2 + \|\delta\bar{\psi}\|^2),$$

where $\delta\bar{\varphi}$ and $\delta\bar{\psi}$ are small perturbations of $\bar{\varphi}$ and $\bar{\psi}$, respectively, and $\nabla_1 b$ and $\nabla_2 b$ are the gradients of b with respect to the first and second variables. Then we get

$$\begin{aligned}
&\hat{S}(\bar{\varphi}+\delta\bar{\varphi},\bar{\psi}) - \hat{S}(\bar{\varphi},\bar{\psi}) \\
&= \big(\|(\bar{\varphi}+\delta\bar{\varphi})'\| - \|\bar{\varphi}'\|\big)\|b(\bar{\varphi},\bar{\psi})\| + \|\bar{\varphi}'\|\big(\|b(\bar{\varphi}+\delta\bar{\varphi},\bar{\psi})\| - \|b(\bar{\varphi},\bar{\varphi})\|\big) \\
&\quad - \langle \delta\bar{\varphi}', b(\bar{\varphi},\bar{\psi})\rangle - \langle \bar{\varphi}', b(\bar{\varphi}+\delta\bar{\varphi},\bar{\psi}) - b(\bar{\varphi},\bar{\psi})\rangle \\
&\quad + \frac{\beta^2(\|(\bar{\varphi}+\delta\bar{\varphi})'\| - \|\bar{\varphi}'\|)}{2\|b(\bar{\varphi},\bar{\psi})\|}\|\bar{\psi} - \bar{\varphi}(s-\tau/\hat{T}(\bar{\varphi},\bar{\psi}))\|^2 \\
&\quad + \frac{\beta^2\|\bar{\varphi}'\|}{2\|b(\bar{\varphi},\bar{\psi})\|^2}(\|b(\bar{\varphi}+\delta\bar{\varphi},\bar{\psi})\| - \|b(\bar{\varphi},\bar{\psi})\|)\|\bar{\psi} - \bar{\varphi}(s-\tau/\hat{T}(\bar{\varphi},\bar{\psi}))\|^2 \\
&\quad + \frac{\beta^2\|\bar{\varphi}'\|}{2\|b(\bar{\varphi}+\delta\bar{\varphi},\bar{\psi})\|}\frac{\delta\|\bar{\psi} - \bar{\varphi}(s-\tau/\hat{T})\|^2}{\delta\hat{T}}(\hat{T}(\bar{\varphi}+\delta\bar{\varphi},\bar{\psi}) - \hat{T}(\bar{\varphi},\bar{\psi})) \\
&\quad + O(\|\delta\bar{\varphi}\|^2 + \|\delta\bar{\varphi}'\|^2),
\end{aligned}$$

with

$$\|(\bar{\varphi}+\delta\bar{\varphi})'\| - \|\bar{\varphi}'\| = \frac{1}{\|\bar{\varphi}'\|}\langle \bar{\varphi}', \delta\bar{\varphi}'\rangle + O(\|\delta\bar{\varphi}'\|^2),$$

$$\|b(\bar{\varphi}+\delta\bar{\varphi},\bar{\psi})\| - \|b(\bar{\varphi},\bar{\psi})\| = \frac{1}{\|b(\bar{\varphi},\bar{\psi})\|}\langle b(\bar{\varphi},\bar{\psi}), \nabla_1 b(\bar{\varphi},\bar{\psi})\delta\bar{\varphi}\rangle + O(\|\delta\bar{\varphi}\|^2),$$

$$\langle \bar{\varphi}', b(\bar{\varphi}+\delta\bar{\varphi},\bar{\psi}) - b(\bar{\varphi},\bar{\psi})\rangle = \langle \bar{\varphi}', \nabla_1 b(\bar{\varphi},\bar{\psi})\delta\bar{\varphi}\rangle + O(\|\delta\bar{\varphi}\|^2),$$

$$\begin{aligned}
&\hat{T}(\bar{\varphi}+\delta\bar{\varphi},\bar{\psi}) - \hat{T}(\bar{\varphi},\bar{\psi}) \\
&= \frac{1}{\|b(\bar{\varphi},\bar{\psi})\|}(\|(\bar{\varphi}+\delta\bar{\varphi})'\| - \|\bar{\varphi}'\|) - \frac{\|\bar{\varphi}'\|}{\|b(\bar{\varphi},\bar{\psi})\|^2}(\|b(\bar{\varphi}+\delta\bar{\varphi},\bar{\psi})\| - \|b(\bar{\varphi},\bar{\psi})\|),
\end{aligned}$$

$$\frac{\delta\|\bar{\psi} - \bar{\varphi}(s-\tau/\hat{T})\|^2}{\delta\hat{T}} = -\frac{\tau}{\hat{T}^2}\Big\langle \bar{\psi} - \bar{\varphi}(s-\frac{\tau}{\hat{T}}), \bar{\varphi}'(s-\frac{\tau}{\hat{T}})\Big\rangle.$$

Combining the equations above and simplifying, we have

$$\begin{aligned}
&\frac{\partial \hat{S}}{\partial x_{i,j}} \\
&= \hat{T}\Big\langle \frac{1}{\hat{T}}\bar{\varphi}' - b(\bar{\varphi},\bar{\psi}), \frac{1}{\hat{T}}\bar{\varphi}_i'\mathbf{e}_j - \nabla_1 b(\bar{\varphi},\bar{\psi})\bar{\varphi}_i\mathbf{e}_j\Big\rangle - \beta^2\Big\langle \bar{\psi} - \bar{\varphi}(s-\frac{\tau}{\hat{T}}), \bar{\varphi}_i(s-\frac{\tau}{\hat{T}})\mathbf{e}_j\Big\rangle \\
&\quad - \frac{\tau\beta^2}{\|\bar{\varphi}'\|^2}\Big\langle \bar{\psi} - \bar{\varphi}(s-\frac{\tau}{\hat{T}}), \bar{\varphi}'(s-\frac{\tau}{\hat{T}})\Big\rangle\Big(\frac{1}{\hat{T}}\langle\bar{\varphi}', \bar{\varphi}_i'\mathbf{e}_j\rangle - \hat{T}\Big\langle b(\bar{\varphi},\bar{\psi}), \nabla_1 b(\bar{\varphi},\bar{\psi})\bar{\varphi}_i\mathbf{e}_j\Big\rangle\Big).
\end{aligned} \tag{8.4.2}$$

Similarly, we have

$$\begin{aligned}
\frac{\partial \hat{S}}{\partial y_{l,j}} =& -\hat{T}\Big\langle \frac{1}{\hat{T}}\bar{\varphi}' - b(\bar{\varphi},\bar{\psi}), \nabla_2 b(\bar{\varphi},\bar{\psi})\bar{\psi}_l\mathbf{e}_j\Big\rangle + \beta^2\Big\langle \bar{\psi} - \bar{\varphi}(s-\frac{\tau}{\hat{T}}), \bar{\psi}_l\mathbf{e}_j\Big\rangle \\
&+ \frac{\tau\beta^2\hat{T}}{\|\bar{\varphi}'\|^2}\Big\langle \bar{\psi} - \bar{\varphi}(s-\frac{\tau}{\hat{T}}), \bar{\varphi}'(s-\frac{\tau}{\hat{T}})\Big\rangle\Big\langle b(\bar{\varphi},\bar{\psi}), \nabla_2 b(\bar{\varphi},\bar{\psi})\bar{\psi}_l\mathbf{e}_j\Big\rangle.
\end{aligned} \tag{8.4.3}$$

With the computation of the gradient, we can use a gradient-type optimization algorithm, such as conjugate gradient, stochastic gradient method or BFGS, to solve the dicretized problem (8.3.1) (or (8.4.1)). Once the minimizer $(\bar{\varphi}_h^*, \bar{\psi}_h^*)$ or the minimal action path $\bar{\varphi}_h^*$ is obtained, the optimal integration time can be approximated by

$$\hat{T}_h^* \approx \hat{T}(\bar{\varphi}_h^*, \bar{\varphi}_h^*(s-\tau/\hat{T})) = \frac{\|\bar{\varphi}'\|}{\|b(\bar{\varphi}_h^*, \bar{\varphi}_h^*(s-\tau/\hat{T}))\|},$$

or

$$\hat{T}_h^* \approx \hat{T}(\bar{\varphi}_h^*, \bar{\psi}_h^*) = \frac{\|\bar{\varphi}'\|}{\|b(\bar{\varphi}_h^*, \bar{\psi}_h^*)\|}.$$

8.5 Numerical Examples

We verify our method using the same system with different delays. We consider in this section the stochastic delayed system (8.1.1) with $b(x,y) = Ax + By$, where

$$A = \begin{bmatrix} -1 & 0 \\ 0 & -5 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

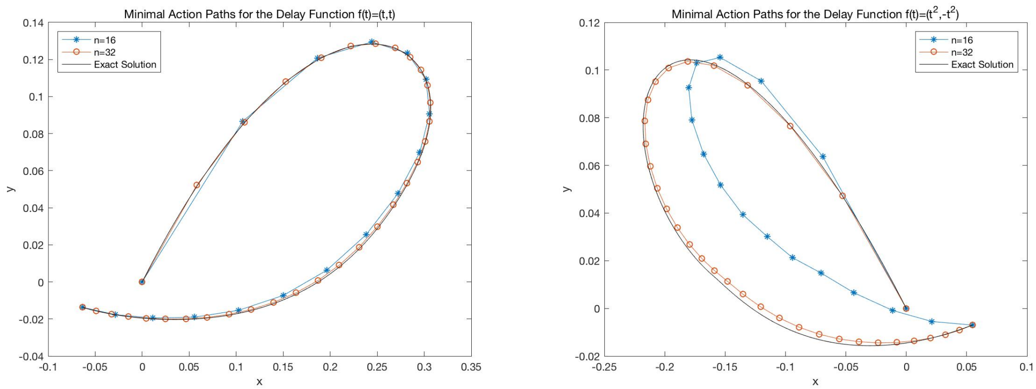


FIGURE 8.1. Approximate MAPs for different delayed systems.

Example 8.5.1. Suppose $\tau = 1, T = 2, f(t) = (t, t)$, i.e., the delay is simply linear. We use our method in Section 8.2 to find the minimal action path numerically. The left plot in Figure 8.1 shows the approximate minimal action path with $n = 16$ and $n = 32$ nodes. For this simple case, the solution is already close to the exact solution when $n = 16$.

The approximate time scaling are $\hat{T}_{16} \approx 1.90$ and $\hat{T}_{32} \approx 1.97$. The corresponding minimum actions are $\hat{S}_{16} \approx 4.90e-3$ and $\hat{S}_{32} \approx 1.23e-3$, respectively. This means that as the mesh refines, the approximate optimal time scaling converges to the optimal time scaling $T^* = 2$, and the discrete action decreases to the minimum action $S^* = 0$.

Example 8.5.2. Suppose $\tau = 1, T = 2, f(t) = (t^2, -t^2)$, i.e., the delay is nonlinear. We apply our method in Section 8.2 to solve the problem numerically. The right plot in Figure 8.1 shows the approximate minimal action path with $n = 16$ and $n = 32$ nodes. For this case, we can see that the solution for $n = 16$ is still far away from the exact solution. But when $n = 32$, the solution is close to the exact solution. This shows an obvious comparison between different numbers of nodes.

The approximate time scaling are $\hat{T}_{16} \approx 1.44$ and $\hat{T}_{32} \approx 1.85$. This also shows an obvious convergence of the approximate optimal time scaling to the optimal time scaling $T^* = 2$ as the mesh refines. The corresponding minimum actions are $\hat{S}_{16} \approx 8.75e - 3$ and $\hat{S}_{32} \approx 1.36e - 3$, respectively. This presents the decrease of discrete action to the minimum action $S^* = 0$.

Remark 8.5.3. It is worth to point out that the convergence rate of the method with uniform mesh and linear elements is very slow. So we need to develop the algorithm by adaptive method (see [28, 33], and also see the discussion at the end of Section 7.8).

References

[1] Braides, A.: Γ-*Convergence for Beginners*, Oxford University Press, 2002.

[2] Braides, A.: *Local Minimization, Variational Evolution and* Γ-*Convergence*, Lecture Notes in Mathematics, vol 2094, Springer, Berlin, 2014.

[3] Cerrai, S. and Freidlin, M.: Large deviations for the Langevin equation with strong damping. *Appl. Math. Optim.*, **161** (2015), 859–875.

[4] Ciarlet, P. G.: *The Finite Element Method for Elliptic Problems*, SIAM, 2002.

[5] Dal Maso, G.: *An Introduction to* Γ-*convergence*, Birkhauser, Boston, 1993.

[6] Dembo, A. and Zeitouni, O.: *Large Deviations Techniques and Applications*, 2nd ed., Springer, 1998.

[7] Duong, M. H., Peletier, M. A., and Zimmer, J.: Conservative-dissipative approximation schemes for a generalized Kramers equation. *Mathematical Methods in the Applied Sciences*. **37** (2014), 2517–2540.

[8] E, W., Ren, W., and Vanden-Eijnden, E.: *String method for the study of rare events*, Phys. Rev. B, **66** (2002), 052301.

[9] E, W., Ren, W., and Vanden-Eijnden, E.: *Minimum action method for the study of rare events*, Commun. Pure Appl. Math., **57** (2004), 637–656.

[10] E, W., Ren, W., and Vanden-Eijnden, E.: *Simplified and improved string method for computing the minimum energy paths in barrier-crossing events*, J. Chem. Phys., **126** (2007), 164103.

[11] Evans, L. C.: *Partial Differential Equations*, American Mathematical Society, 2nd Edition, 2010.

[12] Fogedby, H. C. and Ren, W.: *Minimum action method for the Kardar-Parisi-Zhang equation*, Phys. Rev. E, **80** (2009), 041116.

[13] Freidlin, M. and Wentzell, A.: *Random Perturbations of Dynamical Systems*, second ed., Springer-Verlag, New York, 1998.

[14] Grafke, T., Grauer, R., Schäfer, T., and Vanden-Eijnden, E.: *Arclength parametrized Hamilton's equations for the calculation of instantons*, SIAM, Multiscale Model. Simul., **12(2)** (2014), 566–580.

[15] Grafke, T., Schäfer, T., and Vanden-Eijnden, E.: *Long-term effects of small random perturbations on dynamical systems: theoretical and computational tools*, In: R. Melnik, R. Makarov and J. Belair (eds), Recent Progress and Modern Challenges in Applied Mathematics, Modeling and Computational Science, Fields Institute Communications, vol 79. Springer, New York, NY.

[16] Hale, J. K. and Verduyn Lunel, S. M.: *Introduction to Functional Differential Equations*, Springer, New York, 1991.

[17] Heymann, M. and Vanden-Eijnden, E.: *The geometric minimum action method: A least action principle on the space of curves*, Commun. Pure Appl. Math., **61** (2008), 1052–1117.

[18] Jònsson, H., Mills, G., and Jacobsen, K.: *Nudged elastic band method for finding minimum energy paths of transitions*, Classical and Quantum Dynamics in Condensed Phase Simulations, Ed. B. Berne, G. Ciccotti and D. Coker, 1998.

[19] Kesavan, S.: *Topics in Functional Analysis and Applications*, John Wiley & Sons Inc., 1989.

[20] Lei, H., Baker, N. A., and Li, X.: Data-driven parameterization of the generalized Langevin equation. *PNAS*. **113** (2016), 14183–14188.

[21] Mao, X.: *Stochastic Differential Equations and Applications*, 2nd ed. Horwood, England, 1997.

[22] Mo, C. and Luo, J.: Large deviations for stochastic differential delay equations. *Nonlinear Analysis*. **80** (2013), 202–210.

[23] Nocedal, J. and Wright, S.: *Numerical Optimization*, Springer Series in Operations Research, Springer, New York, 1999.

[24] Smith, N. R., Meerson, B., and Sasorov, P. V.: *Local average height distribution of fluctuating interfaces*, Phys. Rev. E, **95** (2017), 012134.

[25] Sun, Y. and Zhou, X.: *An improved adaptive minimum action method for the calculation of transition path in non-gradient systems*, arXiv:1701.04044.

[26] Wan, X.: *An adaptive high-order minimum action method*, J. Compt. Phys., **230** (2011), 8669–8682.

[27] Wan, X.: *A minimum action method for small random perturbations of two-dimensional parallel shear flows*, J. Compt. Phys., **235** (2013), 497–514.

[28] Wan, X.: *A minimum action method with optimal linear time scaling*, Comm. Compt. Phys., **18(5)** (2015), 1352–1379.

[29] Wan, X. and Lin, G.: *Hybrid parallel computing of minimum action method*, Parallel Computing, **39** (2013), 638–651.

[30] Wan, X. and Yu, H.: *A dynamic-solver-consistent minimum action method: With an application to 2D Navier-Stokes equations*, J. Comput. Phys., **33** (2017), 209–226.

[31] Wan, X., Yu, H., and E, W.: *Model the nonlinear instability of wall-bouned shear flows as a rare event: A study on two-dimensional Poiseuille flows*, Nonlinearity, **28** (2015), 1409–1440.

[32] Wan, X., Yu, H., and Zhai, J.: *Convergence analysis of finite element approximation of large deviation principle*, Accepted (2018), arXiv:1710.03471.

[33] Wan, X., Zheng, B., and Lin, G.: *An hp-adaptive of minimum action method based on a posteriori error estimate*, *Communications in Computational Physics*, **23(2)** (2018), 408–439.

[34] Wan, X., Zhou, X., and E, W.: *Study of the noise-induced transition and the exploration of the configuration space for the Kuramoto-Sivashinsky equation using the minimum action method*, Nonlinearity, **23** (2010), 475–493.

[35] Wells, D. K., Kath, W. L., and Motter, A. E.: *Control of stochastic and induced switching in biophysical networks*, Phys. Rev. X **5** (2015), 031036.

[36] Yao, W. and Ren, W.: *Noise-induced transition in barotropic flow over topography and application to Kuroshio*, J. Comput. Phys. **300** (2015), 352–364.

[37] Zhou, X. and E, W.: *Study of noise-induced transitions in the Lorenz system using the minimum action method*, Comm. Math. Sci., **8(2)** (2010), 341–355.

[38] Zhou, X., Ren, W., and E, W.: *Adaptive minimum action method for the study of rare events*, J. Chem. Phys., **128** (2008), 104111.

Vita

Jiayu Zhai was born in June 1986, in Weifang, Shandong Province, China. He finished his undergraduate studies at Ludong University in July 2009. He earned a master of science degree in mathematics from Shanghai University in July 2012. In August 2012, he came to Louisiana State University to pursue graduate studies in mathematics. He earned a master of science degree in mathematics from Louisiana State University in May 2014. He is currently a candidate for the degree of Doctor of Philosophy in mathematics, which will be awarded in May 2018.