Lecture notes for MCS.T419

Stochastic Differential Equations

Yumiharu Nakano

November 25, 2025

Preface

These lecture notes have been prepared for the course MCS.T419: Stochastic Differential Equations at the Institute of Science Tokyo. The purpose of the notes is to provide an elementary yet rigorous introduction to stochastic differential equations (SDEs), together with the foundations of stochastic control and modern numerical methods for nonlinear partial differential equations (PDEs). The notes are intended to offer a systematic pathway from basic stochastic processes to controlled diffusions, viscosity solutions, and kernel-based numerical schemes.

Chapter 1 is devoted to some preliminaries for handling continuous-time stochastic processes. In particular, we need to introduce the notion of measurability that describes predictabilities of random motions. This theory is often bothersome to application-oriented students, but is indispensable for a rigorous analysis of stochastic processes. Brownian motion is introduced as the canonical model of continuous-time randomness.

Chapters 2 and 3 develop the basic theory of stochastic calculus. Chapter 2 introduces stochastic integrals and Itô's formula, followed by change-of-measure techniques such as the Girsanov–Maruyama theorem and the martingale representation theorem. Chapter 3 then formulates stochastic differential equations, covering existence and uniqueness, explicit solutions, numerical approximations, statistical inference, weak solutions, and time reversal of diffusions. The theory of time reversal provides the mathematical foundation for recent generative models such as denoising diffusion probabilistic models (DDPMs), where the reverse time dynamics of diffusions play a central role.

Chapters 4 and 5 introduce stochastic control theory for controlled diffusions. Chapter 4 presents a basic framework of stochastic controls, continuous-time optimization problems. Then we give a characterization of the stochastic control problems by Hamilton–Jadobi–Bellman (HJB) equations, through verification theorem. Further we study the stochastic control problems with terminal constraints. These terminal-constraint formulations encompass the classical Schrödinger problem (or Schrödinger bridge problem), which can be regarded as an entropy-regularized stochastic control problem connecting prescribed initial and terminal distributions. Chapter 5 then develops the theory of the viscosity solutions, which are the most useful and elegant notion for weak solutions of nonlinear elliptic and parabolic partial differential equations, as well as open up the possibility of rigorous numerical analysis of HJB equations whose classical solutions might not exist.

Chapter 6 turns to numerical methods for nonlinear PDEs. While classical finite-difference methods are powerful in one dimension, their applicability is limited by the curse of dimensionality and stringent regularity requirements. As an alternative, this chapter introduces kernel-based collocation methods, which rely on reproducing kernel Hilbert spaces and have recently attracted attention for multi-dimensional nonlinear PDEs.

Several important topics necessarily fall outside the scope of these notes, including advanced properties of Brownian motion and diffusion processes, stochastic integration with respect to discontinuous semimartingales, Rough paths, backward stochastic differential equations, optimal filtering, infinite-horizon control, optimal stopping, and applications to mathematical finance.

These topics may be treated in future versions.

To the Reader: The reader of these notes is expected to have knowledge of measure-theoretic probability theory and of functional analysis at an introductory level. Several technical parts can be skipped on a first reading, which are explicitly indicated. In particular, the proofs of mathematical statements with the caption "Proof*" can be skipped on a first reading.

Contents

Convention and Notation

Convention

- Throughout these notes except for the appendix, we work on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In particular, all random variables appeared in Chapters 1–5 are assumed to be defined on the measurable space $(\Omega, \mathcal{F})$.
- All stochastic processes appeared in Chapters 2–5 are assumed to be measurable.

Notation

- $\mathbb{N} = \{1, 2, \ldots\}$.
- $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$.
- $\mathbb{R}^d$: d-dimensional Euclidean space.
- $\mathbb{Z}^d = \{(x^1, \ldots, x^d) : x^i \in \mathbb{Z},\ \ 1 \leq i \leq d\}$.
- $\mathbb{R}^{m \times d}$: the totality of real $m \times d$-matrices.
- $\mathbb{S}^d$: the set of all $d \times d$ real symmetric matrices.
- $\mathbb{C}$: the set of complex numbers.
- $|x|$: the standard Euclidean norm of $x \in \mathbb{R}^d$.
- $|a| = (\sum_{i,j} |a_{ij}|^2)^{1/2}$ for any real matrix $a = (a_{ij})$.
- $x^+ = \max\{x, 0\}$, $x \in \mathbb{R}$.
- $x^- = \max\{-x, 0\}$, $x \in \mathbb{R}$.
- a^T: the transposition of a real vector or matrix a.
- A^c: the complement of a set A.
- 1_A: the indicator function for a set A.
- $\mathbb{E}[X]$: the expectation of a random variable X under $\mathbb{P}$.
- $\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$: the variance of X under $\mathbb{P}$.
- $\mathbb{E}_\mathbb{Q}[X]$: the expectation of a random variable X under a probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$.
- I_d: the identity matrix in $\mathbb{R}^{d \times d}$.
- $L^p = L^p(\Omega, \mathcal{F}, \mathbb{P})$ for $p \in [1, \infty]$.

- $\partial_\xi f = \partial f/\partial\xi$ and $\partial^2_{\xi\eta} f = \partial^2 f/\partial\xi\partial\eta$ if the partial derivatives exist for any function f defined on a subset of an Euclidean space.
- For every multi-index $\alpha = (\alpha_1, \ldots, \alpha_d)$ with $|\alpha|_1 := \alpha_1 + \cdots + \alpha_d$, the differential operator D^α is defined as usual by

$$D^\alpha f(x_1, \ldots, x_d) = \frac{\partial^{|\alpha|_1}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} f(x_1, \ldots, x_d).$$

- $C(U)$: the set of all continuous functions on $U \subset \mathbb{R}^n$.
- $C_b(U)$: the set of all bounded continuous functions on $U \subset \mathbb{R}^n$.
- $C_0^\infty(\mathbb{R}^n)$: the set of all infinitely differentiable functions on $\mathbb{R}^n$ having compact supports.
- $C^{1,2}([0,T]\times\mathbb{R}^n)$: the set of all functions $f : [0,T]\times\mathbb{R}^n \to \mathbb{R}$ such that the partial derivatives $\partial_t f$, $\partial_{x_i} f$, $\partial^2_{x_i x_j} f$, $i, j = 1, \ldots, n$, exist and continuous on $[0,T] \times \mathbb{R}^d$.
- $C(U;\mathbb{R}^d)$: the set of all $\mathbb{R}^d$-valued continuous function on U.
- $B_r(x) = \{y \in \mathbb{R}^n : |y - x| < r\}$ for $x \in \mathbb{R}^n$ and $r > 0$.

CHAPTER 1

Preliminaries for Continuous-Time Stochastic Processes

In the theory of stochastic differential equations, *martingales* play a fundamental role. So we first review the abstract notion of *conditional expectation* on which martingale theory is built. Next, we discuss several kinds of *measurability* which are indispensable for handling unpredictable motions of dynamical systems. Then, we deal with *Brownian motions*, which is a basic model of a source of purely random fluctuations.

1.1 Conditional Expectation

For $A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, we call

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

the *conditional probablity of A given B.*

Similarly, for random variable X and $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, we call

$$\mathbb{E}[X|B] := \frac{\mathbb{E}[X 1_B]}{\mathbb{P}(B)}$$

the *conditional expectation of X given B.*

The case of finite σ-algebras

Definition 1.1. A sub σ-algebra $\mathcal{G}$ in $\mathcal{F}$ is said to be *finite* if there exist $A_1, \ldots, A_n \in \mathcal{F}$ such that $\Omega = \cup_{k=1}^n A_k$, $A_i \cap A_j = \emptyset$ $(i \neq j)$ and $\mathcal{G} = \sigma(A_1, \ldots, A_n)$.

- We call $\{A_k\}_{k=1}^n$ in Definition 1.1 a *partition of* Ω.
- The σ-algebra $\mathcal{G}$ in Definition 1.1 is said to be *generated by the partition* $\{A_k\}$.

Definition 1.2. Let $X \in L^1$ and $\mathcal{G}$ be the σ-algebra generated by the partition $\{A_k\}_{k=1}^n$. Then,

$$\mathbb{E}[X|\mathcal{G}] := \sum_{k=1}^{n} \mathbb{E}[X|A_k] 1_{A_k}$$

is said to be the *conditional expectation of X given $\mathcal{G}$.* Here, we set an arbitrary value for $\mathbb{E}[X|A_k]$ if $\mathbb{P}(A_k) = 0$.

- Roughly speaking, $\mathbb{E}[X|\mathcal{G}]$ is the expectation of X computed provided that we know information of $\mathcal{G}$.
- Note that $\mathbb{E}[X|\mathcal{G}]$ is also a random variable. In particular, it is a $\mathcal{G}$-measurable random variable.
- We often write $\mathbb{E}[X|\mathcal{G}](\omega)$ to emphasize that it is a function of $\omega \in \Omega$.
- Since $\{A_k\}$ is a partition of Ω, the quantity $\mathbb{E}[X|\mathcal{G}](\omega)$ gives the conditional expectation of X given the events of which ω belongs to.
- For random variables X, Y, we often write $\mathbb{E}[X|Y]$ for $\mathbb{E}[X|\sigma(Y)]$.

Problem 1.3. Let $p \in (0,1)$ and $0 < d < 1 < u$. Consider the random variables S_i, $i = 0, 1, 2$, defined by

$$S_{i+1} = D_{i+1} S_i, \quad i = 0, 1,$$

where D_1, D_2 are IID with $\mathbb{P}(D_1 = u) = 1 - \mathbb{P}(D_1 = d) = p$ and S_0 is assumed to be a positive constant.

(i) Show that $\sigma(S_1)$ is finite.

(ii) Prove that

$$\mathbb{E}[S_2|S_1] = (up + d(1-p))S_1.$$

General definition

Next consider the case where σ-filed is not necessarily finite. Then of course Definition 1.2 is no longer available. Our idea is to derive a good implication that can be described without the definition of finite σ-fields, and to adopt it as the definition of general conditional expectations.

Proposition 1.4

Let $X \in L^1$ and $\mathcal{G}$ a finite σ-field. Then, for $A \in \mathcal{G}$ we have $\mathbb{E}[X 1_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]1_A]$.

Proof. Let $\{B_k\}_{k=1}^n$ be a partition of Ω satisfying $\mathcal{G} = \sigma(B_1, \ldots, B_n)$.

First notice that the proposition immediately follows if $A \in \mathcal{G}$ is empty. Thus assume that $A \in \mathcal{G}$ is nonempty. Then, $A = \cup_{k=1}^m B_{i_k}$ for some $i_1, \ldots, i_m \in \{1, \ldots, n\}$, and so

$$\begin{aligned}\mathbb{E}[\mathbb{E}[X|\mathcal{G}]1_A] &= \sum_{k=1}^m \mathbb{E}[\mathbb{E}[X|\mathcal{G}]1_{B_{i_k}}] = \sum_{k=1}^m \mathbb{E}[\mathbb{E}[X|B_{i_k}]1_{B_{i_k}}] \\ &= \sum_{k=1}^m \mathbb{E}[X|B_{i_k}]\mathbb{P}(B_{i_k}) = \sum_{k=1}^m \mathbb{E}[X 1_{B_{i_k}}] = \mathbb{E}[X 1_A].\end{aligned}$$

□

Proposition 1.4 means that if $\mathcal{G}$ is finite, then $Y = \mathbb{E}[X|\mathcal{G}]$ is a $\mathcal{G}$-measurable random variable such that $\mathbb{E}[X 1_A] = \mathbb{E}[Y 1_A]$, $A \in \mathcal{G}$. A random variable Y with this property exists when σ-algebra is not necessarily finite, and this existence is unique.

Theorem 1.5

Let $X \in L^1$ and $\mathcal{G}$ a sub σ-algebra in $\mathcal{F}$. Then there exists a random variable Y satisfying the following:

(i) Y is $\mathcal{G}$-measurable.

(ii) $Y \in L^1$.

(iii) $\mathbb{E}[1_A Y] = \mathbb{E}[1_A X]$, $A \in \mathcal{G}$.

Moreover, this existence is almost surely unique, i.e., for $\tilde{Y}$ with the three properties above, we have $Y = \tilde{Y}$ a.s.

Proof. We use the representation $X = X^+ - X^-$. For each X^+ and X^-, we define the probability measure $\mathbb{Q}^\pm$ on $(\Omega, \mathcal{G})$ by

$$\mathbb{Q}^\pm(A) = \int_A \frac{X^\pm + 1}{\mathbb{E}[X^\pm + 1]} d\mathbb{P}, \quad A \in \mathcal{G},$$

respectively. Since $\mathbb{Q}^+$ and $\mathbb{Q}^-$ are both absolutely continuous with respect to $\mathbb{P}$, by Radon-Nikodym theorem (see Theorem A.38), there exist nonnegative, integrable, and $\mathcal{G}$-measurable random variables $Z^\pm$ such that $\mathbb{Q}^\pm(A) = \mathbb{E}[1_A Z^\pm]$, $A \in \mathcal{G}$. Hence, the $\mathcal{G}$-measurable random variable

$$Y = \mathbb{E}[X^+ + 1]Z^+ - \mathbb{E}[X^- + 1]Z^-$$

satisfies (ii) and (iii) in the statement of the theorem.

Next we will show the uniqueness. Suppose that Y and $\tilde{Y}$ satisfy (i)–(iii) in the statement of the theorem and $\mathbb{P}(Y > \tilde{Y}) > 0$. Then, since $\lim_{n\to\infty} \mathbb{P}(Y > \tilde{Y} + 1/n) = \mathbb{P}(Y > \tilde{Y})$, we have $\mathbb{P}(Y > \tilde{Y} + 1/n) > 0$ for some $n \in \mathbb{N}$. It follows from this that

$$\mathbb{E}[(Y - \tilde{Y})1_{\{Y - \tilde{Y} > 1/n\}}] \geq \frac{1}{n}\mathbb{P}(Y > \tilde{Y} + 1/n) > 0.$$

On the other hand, the conditions (ii) and (iii) imply that $A := \{Y > \tilde{Y} + 1/n\} \in \mathcal{G}$ and $\mathbb{E}[Y 1_A] = \mathbb{E}[\tilde{Y} 1_A]$, which lead a contradiction. Thus $Y \leq \tilde{Y}$ a.s. By a similar argument, we see $Y \geq \tilde{Y}$ a.s. Hence $Y = \tilde{Y}$ a.s. □

Therefore, the conditional expectations with respect to finite σ-algebras are completely characterized by the three properties in Theorem 1.5. Then we define the conditional expectations with respect to general σ-algebras by these properties.

Definition 1.6. For $X \in L^1$ and any sub σ-algebra $\mathcal{G}$ in $\mathcal{F}$, we call the unique random variable Y as in Theorem 1.5 the *conditional expectation of X given $\mathcal{G}$*, and write $Y = \mathbb{E}[X|\mathcal{G}]$.

- If you want to confirm that $Y = \mathbb{E}[X|\mathcal{G}]$ a.s., then you only need to check that Y satisfies the properties (i)–(iii) in Theorem 1.5.

We collect basic properties of the conditional expectations given σ-algebras.

Proposition 1.7

Let $X, Y \in L^1$ and let $\mathcal{G}, \mathcal{H}$ be σ-algebras. Then the following hold:

(i) If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X|\mathcal{G}] = X$ a.s.

(ii) $\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}]$ a.s. for $a, b \in \mathbb{R}$.

(iii) If $X \geq 0$ a.s., then $\mathbb{E}[X|\mathcal{G}] \geq 0$ a.s.

(iv) For a sequence $\{X_n\}_{n=1}^\infty$ of random variables such that $0 \leq X_n \leq X_{n+1} \leq \cdots$ a.s. and $X_n \to X$ a.s., then $\mathbb{E}[X_n|\mathcal{G}] \nearrow \mathbb{E}[X|\mathcal{G}]$ a.s.

(v) For a sequence $\{X_n\}_{n=1}^\infty$ of random variables such that $|X_n| \leq Z$ $(\forall n)$ a.s. for some nonnegative random variable $Z \in L^1$ and $\lim_{n\to\infty} X_n = X$ a.s., then

$$\lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}] \text{ a.s.}$$

(vi) If $\mathcal{H} \subset \mathcal{G}$ then $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$ a.s.

(vii) $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$.

(viii) If X is $\mathcal{G}$-measurable and $XY \in L^1$, then $\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}]$ a.s.

(ix) If $\mathcal{H}$ is independent of $\sigma(X, \mathcal{G})$, then $\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X|\mathcal{G}]$ a.s.

(x) If X is independent of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ a.s.

(xi) For $\mathbb{R}$-valued convex function g on $\mathbb{R}$ such that $g(X) \in L^1$, we have $\mathbb{E}[g(X)|\mathcal{G}] \geq g(\mathbb{E}[X|\mathcal{G}])$ a.s.

Proof. (i). The random variable X itself satisfies (i)–(iii) in Theorem 1.5. By the uniqueness, $X = \mathbb{E}[X|\mathcal{G}]$ a.s.
(ii). By the linearity of $\mathbb{E}[\cdot]$, for $A \in \mathcal{G}$,

$$\begin{aligned}\mathbb{E}[(aX + bY)1_A] &= a\mathbb{E}[X1_A] + b\mathbb{E}[Y1_A] = a\mathbb{E}[\mathbb{E}[X|\mathcal{G}]1_A] + b\mathbb{E}[\mathbb{E}[Y|\mathcal{G}]1_A] \\ &= \mathbb{E}[(a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}])1_A].\end{aligned}$$

The uniqueness of $\mathbb{E}[aX + bY|\mathcal{G}]$ means $\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}]$ a.s.
(iii). It follows from $X \geq 0$ and Theorem 1.5 (iii) that $\mathbb{E}[1_A\mathbb{E}[X|\mathcal{G}]] \geq 0$ for $A \in \mathcal{G}$. Hence $\mathbb{E}[X|\mathcal{G}] \geq 0$ a.s.
(iv). From (iii) the sequence $\{Y_n\}$ defined by $Y_n := \mathbb{E}[X_n|\mathcal{G}]$ is almost surely nonnegative and nondecreasing. Thus $Y(\omega) := \limsup_{n\to\infty} Y_n(\omega)$ satisfies $Y_n \nearrow Y$ a.s. Then the monotone convergence theorem for the expectation (see Theorem A.36) yields

$$\mathbb{E}[Y1_A] = \lim_{n\to\infty} \mathbb{E}[Y_n1_A] = \lim_{n\to\infty} \mathbb{E}[X_n1_A] = \mathbb{E}[X1_A], \quad A \in \mathcal{G}.$$

This means that Y satisfies the conditions (i)–(iii) in Theorem 1.5.
(v). Use an argument similar to that in the proof of (iv).
(vi). Let $A \in \mathcal{H}$. Since $A \in \mathcal{G}$, we have $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]1_A] = \mathbb{E}[X1_A]$.
(vii). This follows from the property (iii) in Theorem 1.5 for $A = \Omega$.
(viii). For $B \in \mathcal{G}$ we see

$$\mathbb{E}[1_B\mathbb{E}[Y|\mathcal{G}]1_A] = \mathbb{E}[\mathbb{E}[Y|\mathcal{G}]1_{B\cap A}] = \mathbb{E}[Y1_{B\cap A}] = \mathbb{E}[(1_BY)1_A], \quad A \in \mathcal{G}.$$

Thus, the claim follows for $X = 1_B$. For general X, approximate it with simple random variables and then use a convergence theorem.
(ix). We may assume that $X \geq 0$ a.s. without loss of generality. The claim is trivial when $X = 0$ a.s. Thus we further assume $\mathbb{E}[X] > 0$. Set $Y = \mathbb{E}[X|\mathcal{G}]$. Then we will show that the two probability measures

$$\mu_1(A) = \mathbb{E}[X1_A]/\mathbb{E}[X], \quad \mu_2(A) = \mathbb{E}[Y1_A]/\mathbb{E}[Y], \quad A \in \mathcal{F}$$

coincide with each other on $\sigma(\mathcal{G}, \mathcal{H})$.

Indeed, for $A \in \mathcal{G}$ and $B \in \mathcal{H}$, since $X1_A$ and $Y1_A$ are independent of B, we find

$$\mathbb{E}[X1_{A\cap B}] = \mathbb{E}[X1_A]\mathbb{P}(B) = \mathbb{E}[Y1_A]\mathbb{P}(B) = \mathbb{E}[Y1_{A\cap B}].$$

Hence $\mu_1 = \mu_2$ on $\mathcal{C} := \{A \cap B : A \in \mathcal{G},\ B \in \mathcal{H}\}$. Lemma A.44 now implies that $\mu_1 = \mu_2$ on $\sigma(\mathcal{G}, \mathcal{H}) = \sigma(\mathcal{C})$.
(x). Take $\mathcal{G} = \{\emptyset, \Omega\}$ in (ix).
(xi). We will prove the claim in the case where $\mathcal{G}$ is finite, i.e., it is generated by a partition $\{A_k\}_{k=1}^n$. For general cases we refer to, e.g., [39]. In the present case, $\mathbb{E}[X|\mathcal{G}] = \sum_{k=1}^n \mathbb{E}[X|A_k]1_{A_k}$. Then notice that $\mathbb{E}[X|A_k] = \mathbb{E}^{\mathbb{Q}}[X]$, where $\mathbb{Q}$ is the probability measure defined by $d\mathbb{Q}/d\mathbb{P} = 1_{A_k}/\mathbb{P}(A_k)$. Thus by Jensen's inequality (Proposition A.27),

$$g(\mathbb{E}[X|\mathcal{G}]) = \sum_{k=1}^n g(\mathbb{E}[X|A_k])1_{A_k} \leq \sum_{k=1}^n \mathbb{E}[g(X)|A_k]1_{A_k} = \mathbb{E}[g(X)|\mathcal{G}],$$

as required. □

The conditional expectation $\mathbb{E}[X|\mathcal{G}]$ can be interpreted as the least square estimates of X over $\mathcal{G}$-measurable random variables.

Proposition 1.8

For $X \in L^2$, the conditional expectation $\mathbb{E}[X|\mathcal{G}]$ is almost surely unique $\mathcal{G}$-measurable random variable such that

$$\mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])^2] = \min\{\mathbb{E}[(X - Y)^2] : Y \in L^2,\ \mathcal{G}\text{-measurable}\}.$$

Proof. First notice that for $Y \in L^2$, Cauchy-Schwartz inequality (see Proposition A.28 (i)) yields $|\mathbb{E}[XY]| < \infty$. Thus $(X - Y)^2 \in L^1$. Next, setting $Z = \mathbb{E}[X|\mathcal{G}] - Y$, we have

$$(X - Y)^2 = (X - \mathbb{E}[X|\mathcal{G}] + \mathbb{E}[X|\mathcal{G}] - Y)^2 = (X - \mathbb{E}[X|\mathcal{G}])^2 + 2(X - \mathbb{E}[X|\mathcal{G}])Z + Z^2.$$

If Y is $\mathcal{G}$-measurable, so is Z. Thus By Proposition 1.7,

$$\begin{aligned}\mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])Z] &= \mathbb{E}[\mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])Z|\mathcal{G}]] = \mathbb{E}[Z\mathbb{E}[X - \mathbb{E}[X|\mathcal{G}]|\mathcal{G}]] \\ &= \mathbb{E}[Z(\mathbb{E}[X|\mathcal{G}] - \mathbb{E}[X|\mathcal{G}])] = 0.\end{aligned}$$

This implies

$$\mathbb{E}[(X - Y)^2] = \mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])^2] + \mathbb{E}[Z^2]$$

for any $\mathcal{G}$-measurable $Y \in L^2$. Therefore $\mathbb{E}[Z^2]$ attains the minimum 0 only when $Y = \mathbb{E}[X|\mathcal{G}]$ a.s., which leads to the claim. □

Let $\mathcal{N}$ be the collection of all $\mathbb{P}$-null sets from $\mathcal{F}$. Then, $\sigma(\mathcal{N}) = \{A \in \mathcal{F} : \mathbb{P}(A) = 1 \text{ or } \mathbb{P}(A) = 0\}$. The following is a generalization of Theorem A.17:

Theorem 1.9

Let $(E, \mathcal{E})$ be a measurable space, $Y : \Omega \to E$, and $X : \Omega \to \mathbb{R}$ a $\sigma(\mathcal{N} \cup \sigma(Y))$-measurable random variable. Then, there exists an $\mathcal{E}$-measurable function $f : E \to \mathbb{R}$ such that $X = f(Y)$ a.s.

Proof.* We may assume that X is bounded. Otherwise, it suffices to consider $\arctan(X)$. We also assume that $X \geq 0$ a.s. and $\mathbb{P}(X > 0) > 0$ without loss of generality. Then, define

$$\tilde{X}(\omega) = \mathbb{E}[X|\sigma(Y)](\omega), \quad \omega \in \Omega.$$

By Theorem A.17, $\tilde{X}(\omega) = f(Y(\omega))$, $\omega \in \Omega$, for some $\mathcal{E}$-measurable f. We will show that $X = \tilde{X}$ a.s. To this end, first note that $\mathcal{G} := \sigma(\mathcal{N} \cup \sigma(Y)) = \sigma(\sigma(\mathcal{N}) \cap \sigma(Y))$ and $\sigma(\mathcal{N}) \cap \sigma(Y)$ is a π-system. For any $A \in \sigma(\mathcal{N})$ and $B \in \sigma(Y)$ we have

$$\mathbb{E}[\tilde{X}1_{A\cap B}] = \mathbb{E}[\tilde{X}1_B] = \mathbb{E}[X1_B] = \mathbb{E}[X1_{A\cap B}]$$

if $\mathbb{P}(A) = 1$. Otherwise, $\mathbb{E}[\tilde{X}1_{A\cap B}] = 0 = \mathbb{E}[X1_{A\cap B}]$. Thus, the two probability measures $\mathbb{Q}$ and $\tilde{\mathbb{Q}}$ on $(\Omega, \mathcal{G})$ defined respectively by

$$\mathbb{Q}(A) = \frac{\mathbb{E}[X1_A]}{\mathbb{E}[X]}, \quad \tilde{\mathbb{Q}}(A) = \frac{\mathbb{E}[\tilde{X}1_A]}{\mathbb{E}[X]}, \quad A \in \mathcal{G},$$

agree with each other on $\sigma(\mathcal{N}) \cap \sigma(Y)$. Then, applying Lemma A.44, we find that $\mathbb{E}[X1_A] = \mathbb{E}[\tilde{X}1_A]$, $A \in \mathcal{G}$, whence

$$\mathbb{E}[XZ] = \mathbb{E}[\tilde{X}Z]$$

for any bounded $\mathcal{G}$-measurable random variable Z. Therefore, for any $A \in \mathcal{F}$,

$$\mathbb{E}[X1_A] = \mathbb{E}[X\mathbb{E}[1_A|\mathcal{G}]] = \mathbb{E}[\tilde{X}\mathbb{E}[1_A|\mathcal{G}]] = \mathbb{E}[\tilde{X}1_A].$$

This means $X = \tilde{X}$ a.s., as wanted. $\square$

By Theorem A.17, there exists a measurable function f such that $\mathbb{E}[X|\sigma(Y)] = f(Y)$. Thus, $f(x)$ can be interpreted as the "conditional expectation" $\mathbb{E}[X|Y = x]$. Rigorously, this quantity has no meaning when $\mathbb{P}(Y = x) = 0$. The next theorem gives a valid version of the conditional expectation given $Y = x$. A proof can be found in [28].

Theorem 1.10

Suppose that Ω is a complete separable metric space and $\mathcal{F} = \mathcal{B}(\Omega)$. Let S be a separable metrizable space, and $Y : \Omega \to S$ a Borel measurable map. Denote by μ the law of Y. Then, there exists a family $\{P_x\}_{x\in S}$ of probability measures on $(S, \mathcal{B}(S))$ such that

(i) $S \ni x \mapsto P_x(A)$ is Borel measurable for any $A \in \mathcal{F}$;

(ii) $P_x(A \setminus \{Y = x\}) = 0$ for μ-almost all $x \in S$;

(iii) for any $X \in L^1$,

$$E_{Y(\omega)}[X] = \mathbb{E}[X|\sigma(Y)](\omega)$$

for almost every $\omega \in \Omega$, where E_x is the expectation operator with respect to P_x.

Theorem 1.10 means that for any $X \in L^1$ and bounded measurable f,

$$\mathbb{E}[Xf(Y)] = \int_S f(x)E_x[X]\mu(dx). \tag{1.1.1}$$

In particular, we have the *disintegration formula*

$$\mathbb{P}(A) = \int_S P_x(A)\mu(dy), \quad A \in \mathcal{F}. \tag{1.1.2}$$

To give the interpretation of E_x mentioned above, let $x \in S$ be fixed and $\varepsilon \geq 0$. Assume that $y \mapsto E_y[X]$ is bounded on the open ball $B_{\varepsilon,x}$ at x with radius ε . Then by (1.1.1) and the mean value theorem for Lebesgue integral (see, e.g., [43, 定理 12.5]),

$$\mathbb{E}[X1_{\{Y\in B_{\varepsilon,x}\}}] = \int_{B_{\varepsilon,x}} E_y[X]\mu(dy) = c\mathbb{P}(Y \in B_{\varepsilon,x})$$

for some $c \in [\inf_{y\in B_{\varepsilon,x}} E_y[X], \sup_{y\in B_{\varepsilon,x}} E_y[X]]$. Therefore, if $\mathbb{P}(Y = x) > 0$, then considering $\varepsilon = 0$ we obtain

$$\mathbb{E}[X|Y = x] = E_x[X].$$

In the case of $\mathbb{P}(Y = x) = 0$, by assuming the continuity of $y \mapsto E_y[X]$, we get

$$\lim_{\varepsilon\to 0} \mathbb{E}[X|Y \in B_{\varepsilon,x}] = E_x[X].$$

Thus, we conclude that $E_x[X]$ can be interpreted as the conditional expectation of X given $Y = x$.

1.2 Filtration, Measurability, and Martingales

An $\mathbb{R}^d$-valued *stochastic process* is a family $\{X_t\}_{t\in\mathbb{T}}$ of random variables taking values in $\mathbb{R}^d$. The index generally represents a continuous or discrete time variable.

Definition 1.11. Let $\mathbb{T} = [0, \infty)$, $[0, T]$, $\mathbb{N}\cup\{0\}$, or $\{0, 1, \ldots, N\}$, where $T \in (0, \infty)$ and $N \in \mathbb{N}$. A family $\mathbb{F} = \{\mathcal{F}_t\}_{t\in\mathbb{T}}$ of sub σ-fields of $\mathcal{F}$ is said to be a *filtration* if $\mathcal{F}_s \subset \mathcal{F}_t$ for $s, t \in \mathbb{T}$ with $s \leq t$.

- $\mathcal{F}_t$ is interpreted as the information available at time t.
- The quadruplet $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is said to be a *filtered probability space.*

Definition 1.12. Let $\mathbb{T}$ be as in Definition 1.11, and let $\mathbb{F} = \{\mathcal{F}\}_{t\in\mathbb{T}}$ be a filtration. An $\mathbb{R}^d$-valued stochastic process $\{X_t\}_{t\in\mathbb{T}}$ is said to be $\mathbb{F}$-*adapted* if X_t is $\mathcal{F}_t$-measurable for any $t \in \mathbb{T}$.

- If $\{X_t\}$ is an adapted process, then the random variable X_t is realized up to time t.
- For an arbitrary process $\{X_t\}_{t\in\mathbb{T}}$, the family $\mathbb{F}^X = \{\mathcal{F}_t^X\}_{t\in\mathbb{T}}$ of sub σ-algebras defined by $\mathcal{F}_t^X = \sigma(X_s; s \in \mathbb{T},\ s \leq t)$ is said to be the *natural filtration* generated by $\{X_t\}_{t\in\mathbb{T}}$. Here, for a family $\{Z_\lambda\}_{\lambda\in\Lambda}$ of random variables,

$$\sigma(Z_\lambda; \lambda \in \Lambda) := \sigma\left(\bigcup_{\lambda\in\Lambda} \sigma(Z_\lambda)\right).$$

- Any stochastic process is adapted w.r.t. the natural filtration generated by itself.

In what follows, we work on a fixed filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$.

Definition 1.13. (i) A process $\{X_t\}_{t\geq 0}$ is said to be *measurable* if $X_\cdot : [0,\infty)\times\Omega\to\mathbb{R}$ is $\mathcal{B}[0,\infty)\times\mathcal{F}$-measurable.

(ii) A process $\{X_t\}_{t\geq 0}$ is said to be $\mathbb{F}$-*progressively measurable* if $X_\cdot : [0,t]\times\Omega\to\mathbb{R}$ is $\mathcal{B}[0,t]\times\mathcal{F}_t$-measurable for every $t\in[0,\infty)$.

- If $\{X_t\}$ is measurable, then for every t the random variable $Y_t := \int_0^t X_s ds$ is $\mathcal{F}$-measurable.
- If $\{X_t\}$ is progressively measurable, then $\{Y_t\}$ above is an adapted process.

Problem 1.14. Show that every progressively measurable process is measurable and adapted.

Hereafter, all processes appeared in these notes are assumed to be measurable.

Definition 1.15. We say that $\{X_t\}_{t\geq 0}$ is a *modification* of $\{Y_t\}_{t\geq 0}$ if $\mathbb{P}(X_t = Y_t) = 1$ for any $t\geq 0$. Moreover, $\{X_t\}$ and $\{Y_t\}$ are said to be *indistinguishable* if $\mathbb{P}(X_t = Y_t,\ t\geq 0) = 1$.

Example 1.16. Let τ be a $(0,\infty)$-valued random variable having a continuous density, say an exponentially distributed random variable. Set $X_t = 1_{\{\tau\leq t\}}$, $t\geq 0$ and consider the left-limit $X_{t-} = \lim_{s\nearrow t} X_t$. Then it is straightforward to see that $Y_t := X_t - X_{t-} = 1_{\{\tau=t\}}$, and that $\mathbb{P}(Y_t = 0) = \mathbb{P}(\tau\neq t) = 1$ for every $t\in[0,\infty)$. Hence, the process $Z_t\equiv 0$ is a modification of $\{Y_t\}$. On the other hand, we have $\mathbb{P}(Y_t = 0,\ t\geq 0) = \mathbb{P}(\tau\neq t,\ t\geq 0) = \mathbb{P}(\tau\notin[0,\infty)) = 0$, which implies that $\{Y_t\}$ and $\{Z_t\}$ are not indistinguishable.

Proposition 1.17

Suppose that $\{X_t\}_{t\geq 0}$ is adapted and $\{Y_t\}_{t\geq 0}$ is a modification of $\{X_t\}$. Suppose moreover that $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets that are $\mathcal{F}$-measurable, i.e., that $\mathcal{N}\subset\mathcal{F}_0$. Then $\{Y_t\}_{t\geq 0}$ is also adapted.

Proof. Fix $t\geq 0$ and set $N = \{X_t\neq Y_t\}$. Then observe that for $A\in\mathcal{B}(\mathbb{R})$,

$$\{Y_t\in A\} = (\{Y_t\in A\}\cap N)\cup(\{Y_t\in A\}\cap N^c) = (\{Y_t\in A\}\cap N)\cup(\{X_t\in A\}\cap N^c).$$

Since Y_t is $\mathcal{F}$-measurable and $N, N^c\in\mathcal{F}_0$, we have $\{Y_t\in A\}\in\mathcal{F}$ and $\mathbb{P}(\{Y_t\in A\}\cap N) = 0$. Hence $\{Y_t\in A\}\cap N\in\mathcal{F}_0$. This together with $\{X_t\in A\}\cap N^c\in\mathcal{F}_t$ means $\{Y_t\in A\}\in\mathcal{F}_t$. ☐

- We often assume $\mathcal{F}_0\supset\mathcal{N} = \{A\in\mathcal{F} : \mathbb{P}(A) = 0\}$ to use the convenient property above.
- The filtration $\sigma(\mathcal{F}_t^X\cup\mathcal{N})$, $t\geq 0$, is called the *augmented natural filtration generated by* X.

Problem 1.18. Suppose that $\mathcal{F}_0\supset\mathcal{N}$. Let $\{X_t^{(n)}\}_{t\geq 0}$ be a sequence of adapted processes such that $X_t^{(n)}$ converges to some X_t almost surely for any $t\geq 0$. Show that $\{X_t\}_{t\geq 0}$ is adapted.

In general, $t\mapsto X_t(\omega)$, $\omega\in\Omega$, is called a *sample path* of the process $\{X_t\}$ with respect to ω. We say that $\{X_t\}$ is a *continuous process* if every sample path of $\{X_t\}$ is continuous, i.e., $t\mapsto X_t(\omega)$ is continuous for every $\omega\in\Omega$. We also say that $\{X_t\}$ is a.s. continuous if $t\mapsto X_t(\omega)$ is continuous for almost all $\omega\in\Omega$.

Proposition 1.19

Let $\{X_t\}$ and $\{Y_t\}$ be continuous. If $\{X_t\}$ and $\{Y_t\}$ are modifications of each other, then the two processes are indistinguishable. Moreover, if $\{X_t\}$ is adapted, then it is progressively measurable.

Proof. Let $\omega \in \{X_t = Y_t \text{ for all } t \in \mathbb{Q} \cap [0, \infty)\}$. For any $t \geq 0$ there exists $\{t_n\} \subset \mathbb{Q} \cap [0, \infty)$ such that $t_n \to t$. Then, by the continuity of $\{X_t\}$, we have $X_t(\omega) = \lim_{n\to\infty} X_{t_n}(\omega) = \lim_{n\to\infty} Y_{t_n}(\omega) = Y_t(\omega)$. This implies $\mathbb{P}(X_t = Y_t, \ \forall t) = \cap_{s\in\mathbb{Q}\cap[0,\infty)} \mathbb{P}(X_s = Y_s) = 1$.

To prove the second claim, we consider a piece-wise linear function $[0, t] \ni s \mapsto X_s^{(n)}(\omega)$ satisfying $X_s^{(n)}(\omega) = X_s(\omega)$, $s = 0, 2^{-n}, \ldots, 2^{-n}\lfloor 2^n t\rfloor$. Here, $\lfloor x \rfloor$ denotes the greatest integer not exceeding $x \in \mathbb{R}$. Then, $X^{(n)}$ is $\mathcal{B}([0, t]) \times \mathcal{F}_t$-measurable. This together with $\lim_{n\to\infty} X_s^{(n)}(\omega) = X_s(\omega)$ for ω and $s \in [0, t]$ means that X_s, $s \leq t$, is also $\mathcal{B}([0, t]) \times \mathcal{F}_t$-measurable. □

The proposition above is generalized in the following sense:

Proposition 1.20

Every measurable and adapted process has a progressively measurable modification.

The proof of this result is found in [28].

Proposition 1.21

Suppose that $\mathcal{F}_0 \supset \mathcal{N}$. Let $\{X_t\}_{0\leq t\leq T}$ be an adapted process satisfying

$$\int_0^T |X_t| dt < \infty, \quad \text{a.s.}$$

Then, the process

$$Y_t = \int_0^t X_s ds, \quad 0 \leq t \leq T,$$

is progressively measurable. In particular, $\{Y_t\}$ is adapted.

Proof. By Proposition 1.20, the process $\{X_t\}$ has a progressively measurable modification $\{\tilde{X}_t\}$. Then, $\tilde{Y}_t := \int_0^t \tilde{X}_s ds$, $0 \leq t \leq T$, is adapted. By Fubini theorem,

$$\mathbb{E} \int_0^T 1_{\{X_s \neq \tilde{X}_s\}} ds = \int_0^T \mathbb{E}[1_{\{X_s \neq \tilde{X}_s\}}] ds = 0.$$

Thus, the Lebesgue measure of $\{s : X_s \neq \tilde{X}_s\}$ is zero almost surely, whence $Y_t = \tilde{Y}_t$ a.s., $t \in [0, T]$. Then Proposition 1.17 and Proposition 1.19 mean that $\{Y_t\}$ is adapted and so is progressively measurable due to the continuity. □

Problem 1.22. Prove that if $\{X_t\}_{t\geq 0}$ is continuous then $\sup_{t\geq 0} X_t$, $\inf_{t\geq 0} X_t$, $\limsup_{t\to\infty} X_t$, and $\liminf_{t\to\infty} X_t$ are all $\mathcal{F}$-measurable random variables.

Problem 1.23. Prove that if $\{X_t\}_{t\geq 0}$ is continuous then

$$\sigma(X_t; 0 \leq t \leq T) = \sigma(X_t; t \in \mathbb{T}' \cap [0, T])$$

for any dense subset $\mathbb{T}' \subset [0, \infty)$ and $T \in [0, \infty)$.

Definition 1.24. Let $\mathbb{T}$ be as in Definition 1.11, and let $\mathbb{F} = \{\mathcal{F}\}_{t\in\mathbb{T}}$ be a filtration. A real-valued process $\{X_t\}_{t\in\mathbb{T}}$ is said to be an $\mathbb{F}$-*martingale* if the following three conditions are satisfied:

(i) $X_t \in L^1$ for any $t \in \mathbb{T}$.

(ii) $\{X_t\}$ is $\mathbb{F}$-adapted.

(iii) $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ for $s, t \in \mathbb{T}$ with $s \leq t$.

Example 1.25 (Simple random walk). Let $X_0 \in \mathbb{R}$, and let $\{X_n\}_{n=1}^{\infty}$ be an IID sequence with $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = 1/2$, $n \in \mathbb{N}$. Then define $\{S_n\}_{n=0}^{\infty}$ by

$$S_n = \sum_{k=0}^{n} X_k, \quad n \in \mathbb{N}.$$

We say that the process $\{S_n\}_{n=0}^{\infty}$ is a *simple random walk* starting from X_0.

Now, let $\mathbb{F}$ be the natural filtration generated by $\{X_n\}$. Then it is straightforward to see from Proposition 1.7 that $\mathbb{E}[X_m|\mathcal{F}_n] = 0$ for $m > n$. This means that $\{S_n\}$ is an $\mathbb{F}$-martingale.

Example 1.26. Let $X \in L^1$. Then $X_t := \mathbb{E}[X|\mathcal{F}_t]$, $t \in \mathbb{T}$, gives the estimation of unrealized variable X based on the information available at time t. By Proposition 1.7, the process $\{X_t\}$ is a martingale.

In Example 1.26, if $\mathbb{T} = \mathbb{N} \cup \{0\}$, then one might expect that $X_n \to X$ as $n \to \infty$, which is guaranteed by the following result:

Theorem 1.27

Let $\mathcal{G}$ be a sub σ-field of $\mathcal{F}$, and $X \in L^2$ a $\mathcal{G}$-measurable random variable. Suppose that the filtration $\mathbb{G} = \{\mathcal{G}_n\}_{n\geq 0}$ saitisfies $\mathcal{G} = \sigma(\mathcal{G}_n : n \geq 0)$. Then $\mathbb{E}[X|\mathcal{G}_n]$ converges to X almost surely and in L^2.

The proof is omitted. An interested reader may refer to [39, Ch. 14].

Definition 1.28. Let $\mathbb{T}$ be as in Definition 1.11, and let $\mathbb{F} = \{\mathcal{F}\}_{t\in\mathbb{T}}$ be a filtration. Suppose that a real-valued process $\{X_t\}_{t\in\mathbb{T}}$ is $\mathbb{F}$-adapted and satisfies $X_t \in L^1$, $t \in \mathbb{T}$. We say that $\{X_t\}$ is an $\mathbb{F}$-*supermartingale* if

$$\mathbb{E}[X_t|\mathcal{F}_s] \leq X_s \text{ a.s.} \quad t \geq s,$$

and that $\{X_t\}$ is an $\mathbb{F}$-*submartingale* if

$$\mathbb{E}[X_t|\mathcal{F}_s] \geq X_s \text{ a.s.} \quad t \geq s.$$

- If $\{X_t\}$ is a supermartingale (resp. submartingale), then $\mathbb{E}[X_t]$ is nonincreasing (resp. nondecreasing).

Problem 1.29. Let $\{M_t\}_{t\in\mathbb{T}}$ be a martingale and $p \geq 1$. Show that if $\mathbb{E}|M_t|^p < \infty$ for every $t \in \mathbb{T}$ then the process $\{|M_t|^p\}_{t\in\mathbb{T}}$ is a submartingale.

Definition 1.30. Let $\mathbb{F} = \{\mathcal{F}_t\}_{t\in[0,\infty)}$ be a filtration. We say that $\tau : \Omega \to [0, \infty]$ is $\mathbb{F}$-*stopping time* if it satisfies $\{\tau \leq t\} \in \mathcal{F}_t$ for any $t \in [0, \infty)$.

- If τ_1 and τ_2 are $\mathbb{F}$-stopping times, then $\tau_1 \vee \tau_2$ and $\tau_1 \wedge \tau_2$ are also $\mathbb{F}$-stopping times. This follows from

$$\{\tau_1 \vee \tau_2 \leq t\} = \{\tau_1 \leq t\} \cap \{\tau_2 \leq t\},$$
$$\{\tau_1 \wedge \tau_2 \leq t\} = \{\tau_1 \leq t\} \cup \{\tau_2 \leq t\}.$$

A filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}$ is said to be *right-continuous* if $\mathcal{F}_t = \mathcal{F}_{t+} := \cap_{s>t}\mathcal{F}_s$ for any $t \geq 0$.

Proposition 1.31

Let $\mathbb{F}$ be a right-continuous filtration. Then the following (i)–(iv) are equivalent:

(i) τ is a stopping time.

(ii) $\{\tau < t\} \in \mathcal{F}_t$ for any $t \geq 0$.

(iii) $\{\tau > t\} \in \mathcal{F}_t$ for any $t \geq 0$.

(iv) $\{\tau \geq t\} \in \mathcal{F}_t$ for any $t \geq 0$.

Proof. If τ is a stopping time, then by definition $\{\tau < t\} = \cup_{n=1}^{\infty}\{\tau \le t - 1/n\} \in \mathcal{F}_t$. Thus the implication (i)⇒(ii) follows. Conversely, assume that (ii) holds. Then for $k \ge 1$ we have $\{\tau \le t\} = \cap_{n=k}^{\infty}\{\tau < t + 1/n\} \in \mathcal{F}_{t+1/k}$. This together with the right-continuity of $\mathbb{F}$ implies that (i) holds. The claims (i)⇔(iii) and (ii)⇔(iv) are trivial. Thus the proposition follows. □

Proposition 1.32

Let $\mathbb{F}$ be a right-continuous filtration and $\{X_t\}_{t\ge0}$ an $\mathbb{R}^d$-valued continuous $\mathbb{F}$-adapted process. If A is an open or a closed subset of $\mathbb{R}^d$, then the random variable

$$\tau_A(\omega) := \inf\{t > 0 : X_t(\omega) \in A\}$$

is an $\mathbb{F}$-stopping time. Here, by convention, $\inf\emptyset = +\infty$.

- τ_A is called the *hitting time* of $\{X_t\}$ to A or the first exit time of $\{X_t\}$ from A^c.
- We say that a filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\ge0}$ satisfies the *usual conditions* if it is right-continuous and $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets from $\mathcal{F}$.

For a filtration $\mathbb{G} = \{\mathcal{G}_t\}_{t\ge0}$ and a $\mathbb{G}$-stopping time τ, we define

$$\mathcal{G}_\tau := \{A \in \mathcal{G}_\infty : A \cap \{\tau \le t\} \in \mathcal{G}_t,\ \forall t \ge 0\}.$$

- Here, $\mathcal{G}_\infty := \sigma(\mathcal{G}_t : t \ge 0)$.
- Roughly speaking, $\mathcal{G}_\tau$ is the σ-algebra generated by events occurring before τ.
- If two stopping times σ and τ satisfies $\sigma(\omega) \le \tau(\omega)$ for all $\omega \in \Omega$, then we have $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.

Proposition 1.33

Suppose that $\mathbb{F}$ is right-continuous. Let $\{X_t\}_{t\ge0}$ be an $\mathbb{F}$-progressively measurable process, and let τ an $\mathbb{F}$-stopping time with $\tau < \infty$ a.s. Then X_τ is $\mathcal{F}_\tau$-measurable.

Proof. Fix $t \ge 0$. By the assumption, the mapping $(\omega, s) \mapsto X_s(\omega)$ is measurable from $(\Omega \times [0,t], \mathcal{F}_t \times \mathcal{B}[0,t])$ into $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Moreover, the mapping $\omega \mapsto (\omega, \tau(\omega) \wedge t)$ is measurable from $(\Omega, \mathcal{F}_t)$ into $(\Omega \times [0,t], \mathcal{F}_t \times \mathcal{B}[0,t])$. Hence $X_{\tau\wedge t}$ is $\mathcal{F}_t$-measurable. In addition, by Proposition 1.31, we have $\{\tau < t\}, \{\tau = t\} \in \mathcal{F}_t$. Therefore, for $B \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}\{X_\tau \in B\} \cap \{\tau \le t\} &= \{X_\tau \in B\} \cap \{\tau < t\} \cup \{X_\tau \in B\} \cap \{\tau = t\} \\ &= \{X_{\tau\wedge t} \in B\} \cap \{\tau < t\} \cup \{X_t \in B\} \cap \{\tau = t\} \in \mathcal{F}_t.\end{aligned}$$

Thus the proposition follows. □

The following inequality for continuous submartingales is frequently used.

Theorem 1.34: Doob's maximal inequality

Suppose that $\{X_t\}_{t\geq 0}$ is a nonnegative submartingale with continuous paths. Then, for every $T \geq 0$ and $\lambda > 0$,

$$\mathbb{P}\left(\sup_{0\leq t\leq T} X_t \geq \lambda\right) \leq \frac{1}{\lambda}\mathbb{E}[X_T].$$

Moreover, for any $p > 1$, if $\mathbb{E}[X_T^p] < \infty$ then we have

$$\mathbb{E}\left[\sup_{0\leq t\leq T} X_t^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[X_T^p\right].$$

Proof. Notice that by the continuity $\sup_{0\leq t\leq T} X_t$ is certainly $\mathcal{F}$-measurable (see Problem 1.22) and

$$\sup_{0\leq t\leq T} X_t = \sup_{n\geq 0} X_{t_n},$$

where $\{t_n\}_{n=0}^{\infty} = \mathbb{Q}\cap[0,T]$ such that $0 = t_0 < t_1 < \cdots$ and $\lim_{n\to\infty} t_n = T$. Then, we find that the event $A^{(n)} = \{\sup_{0\leq k\leq n} X_{t_k} \geq \lambda\}$ is represented as $A^{(n)} = \cup_{k=0}^n A_k$ with

$$A_0 = \{X_0 \geq \lambda\}, \quad A_k = \left\{X_{t_k} \geq \lambda, \max_{0\leq i\leq k-1} X_{t_i} < \lambda\right\}, \quad k = 1, 2, \ldots, n.$$

Since $A_k^{(n)}$'s are disjoint, by Chebyshev's inequality and the submartingale property we see

$$\begin{aligned}\mathbb{P}(A^{(n)}) &= \sum_{k=0}^n \mathbb{P}(A_k^{(n)}) \leq \frac{1}{\lambda}\sum_{k=0}^n \mathbb{E}[X_{t_k}1_{A_k^{(n)}}] \leq \frac{1}{\lambda}\sum_{k=0}^n \mathbb{E}[X_T 1_{A_k^{(n)}}] = \frac{1}{\lambda}\mathbb{E}[X_T 1_{A^{(n)}}] \\ &\leq \frac{1}{\lambda}\mathbb{E}[X_T 1_{\{\sup_{0\leq t\leq T} X_t \geq \lambda\}}] \leq \frac{1}{\lambda}\mathbb{E}[X_T]. \qquad (1.2.1)\end{aligned}$$

Letting $n \to \infty$, we obtain the first required inequality.

To show the second inequality, put $Y = \sup_{0\leq t\leq T} X_t$ and observe, for $K > 0$,

$$\begin{aligned}\mathbb{E}[(Y\wedge K)^p] &= p\int_0^\infty \lambda^{p-1}\mathbb{P}(Y\wedge K \geq \lambda)d\lambda \leq p\int_0^K \lambda^{p-1}\frac{1}{\lambda}\mathbb{E}[X_T 1_{\{Y\geq\lambda\}}]d\lambda \\ &= p\mathbb{E}\left[\int_0^{Y\wedge K} \lambda^{p-2}d\lambda X_T\right] = \frac{p}{p-1}\mathbb{E}[(Y\wedge K)^{p-1}X_T] \\ &\leq \frac{p}{p-1}\mathbb{E}[(Y\wedge K)^p]^{(p-1)/p}\mathbb{E}[X_T^p]^{1/p}.\end{aligned}$$

Here, we have used (1.2.1) with limit, Fubini's theorem, and Hölder's inequality. Thus,

$$\mathbb{E}[Y^p]^{1/p} = \lim_{K\to\infty}\mathbb{E}[(Y\wedge K)^p]^{1/p} \leq \frac{p}{p-1}\mathbb{E}[X_T^p]^{1/p},$$

as wanted. □

1.3 Brownian Motion

Consider the simple random walk $S_n = \sum_{k=1}^n X_k$, $n \geq 0$, starting from 0. To embed this into the continuous time framework, we use the normalized process

$$W_0^{(n)} := \frac{1}{\sqrt{n}}S_0 = 0, \quad W_{1/n}^{(n)} := \frac{1}{\sqrt{n}}S_1, \quad W_{2/n}^{(n)} := \frac{1}{\sqrt{n}}S_2, \cdots$$

of S_n by $\sqrt{n}$. Then we define the continuous time process $W_t^{(n)}$ by its linear interpolation, i.e.,

$$W_t^{(n)} = \frac{1}{\sqrt{n}} \left[S_{\lfloor nt \rfloor} + X_{\lfloor nt \rfloor + 1}(nt - \lfloor nt \rfloor) \right], \quad t \geq 0.$$

Figure 1.3.1: Sample paths of $W_t^{(n)}$. The cases of $n = 10$ (left), $n = 100$ (center), $n = 1000$ (right).

We shall consider a limit of $W_t^{(n)}$ as $n \to \infty$.

Proposition 1.35

Let $0 = t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_m$. Then the $\mathbb{R}^{m+1}$-valued random variable $(W_{t_0}^{(n)}, W_{t_1}^{(n)}, \ldots, W_{t_m}^{(n)})$ converges in distribution to an $\mathbb{R}^{m+1}$-valued random variable $(W_{t_0}, W_{t_1}, \ldots, W_{t_m})$ having the following properties:

(i) $W_{t_0} = 0$ a.s.

(ii) $W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_m} - W_{t_{m-1}}$ are independent.

(iii) For each k, the random variable $W_{t_k} - W_{t_{k-1}}$ has a Gaussian distribution with mean 0 and variance $t_k - t_{k-1}$.

Proof. We will prove the case of $m = 2$. The proof for the general case is similar. For simplicity set $s = t_1$ and $t = t_2$. We see

$$\left| W_t^{(n)} - \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor} \right| \leq \frac{1}{\sqrt{n}}$$

to obtain

$$\left| (W_s^{(n)}, W_t^{(n)}) - \frac{1}{\sqrt{n}} (S_{\lfloor sn \rfloor}, S_{\lfloor tn \rfloor}) \right| \to 0 \quad \text{a.s.}$$

Hence, it is sufficient to show that

$$\frac{1}{\sqrt{n}} \left(\sum_{j=1}^{\lfloor sn \rfloor} X_j, \sum_{j=1}^{\lfloor tn \rfloor} X_j \right) \to (W_s, W_t) \quad \text{in law.} \tag{1.3.1}$$

To this end, let i be the imaginary unit and $\alpha, \beta \in \mathbb{R}$. Then, by the IID property of $\{\xi_j\}$,

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(i\alpha\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor sn\rfloor}X_j + i\beta\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor tn\rfloor}X_j\right)\right] \\
&= \mathbb{E}\left[\exp\left(i(\alpha+\beta)\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor sn\rfloor}X_j + i\beta\frac{1}{\sqrt{n}}\sum_{j=\lfloor sn\rfloor+1}^{\lfloor tn\rfloor}X_j\right)\right] \\
&= \mathbb{E}\left[\exp\left(i(\alpha+\beta)\sqrt{\frac{\lfloor sn\rfloor}{n}}\frac{1}{\sqrt{\lfloor sn\rfloor}}\sum_{j=1}^{\lfloor sn\rfloor}X_j\right)\right] \\
&\quad \times \mathbb{E}\left[\exp\left(i\beta\sqrt{\frac{\lfloor tn\rfloor-\lfloor sn\rfloor}{n}}\frac{1}{\sqrt{\lfloor tn\rfloor-\lfloor sn\rfloor}}\sum_{j=1}^{\lfloor tn\rfloor-\lfloor sn\rfloor}X_j\right)\right].
\end{aligned}$$

It follows from $(sn-1)/n \le \lfloor sn\rfloor/n \le s$ that $\lfloor sn\rfloor/n \to s$. Further, by the central limit theorem, the distribution of $\frac{1}{\sqrt{\lfloor sn\rfloor}}\sum_{j=1}^{\lfloor sn\rfloor}X_j$ converges to the standard normal distribution. Therefore

$$\begin{aligned}
\mathbb{E}\left[\exp\left(i\alpha\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor sn\rfloor}X_j + i\beta\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor tn\rfloor}X_j\right)\right] &\to \mathbb{E}[e^{i(\alpha-\beta)W_s}]\mathbb{E}[e^{i\beta(W_{t-s})}] \\
&= \mathbb{E}[e^{i\alpha W_s + i\beta W_t}].
\end{aligned}$$

Thus (1.3.1) follows. □

This suggests that a process $\{W_t\}$ satisfying Proposition 1.35 (i)–(iii) can be seen as a limit of $\{W_t^{(n)}\}$. We shall call such process $\{W_t\}$ as *Brownian motion.*

Definition 1.36. A real-valued process $\{W_t\}_{t\ge 0}$ is said to be a *Brownian motion* if

(i) $W_0 = 0$ a.s.

(ii) Independent increments property: for $0 = t_0 < t_1 < \cdots < t_m$, the random variables $W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_m} - W_{t_{m-1}}$ are independent.

(iii) Stationary increments property: for $s \le t$, the random variable $W_t - W_s$ is a Gaussian random variable with mean 0 and variance $t - s$.

It should be noted that Proposition 1.35 *does not* guarantee the existence of a Brownian motion. The proposition means that if a Brownian motion exists then its distribution coincides with the limiting distribution of $\{W_t^{(n)}\}$.

To discuss the existence of a Brownian motion rigorously, we consider the measurable space $(C[0,\infty), \mathcal{B}(C[0,\infty)))$ defined by the totality of continuous functions on $[0,\infty)$. Then, the projection π_t defined by $\pi_t(\omega) = \omega(t)$, $\omega \in C[0,\infty)$ is a measurable function on $C[0,\infty)$. For each $\omega \in C[0,\infty)$ we can regard $\{\pi_t(\omega)\}_{t\ge 0} = \{\omega(t)\}_{t\ge 0}$ as the sample paths of a process. We call $\{\pi_t\}_{t\ge 0}$ as *coordinate process.*

Now suppose that a probability measure P on $(C[0,\infty), \mathcal{B}(C[0,\infty)))$ satisfies, for $0 = t_0 < t_1 < \cdots < t_m$ and $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$,

$$\begin{aligned}
&P(\omega : \omega(t_k) - \omega(t_{k-1}) \le \alpha_k,\ k = 1, \ldots, m) \\
&= \prod_{k=1}^{m}\frac{1}{\sqrt{2\pi(t_k - t_{k-1})}}\int_{-\infty}^{\alpha_k} e^{-u^2/2(t_k - t_{k-1})}du.
\end{aligned} \tag{1.3.2}$$

Then, the coordinate process $\{\pi_t\}$ on the probability space $(C[0,\infty), \mathcal{B}(C[0,\infty)), P)$ is a Brownian motion. Therefore, the existence problem of a Brownian motion is reduced to that of P. Let P_n be the distribution of $C[0,\infty)$-valued random variable $W^{(n)} := \{W_t^{(n)}\}$. Then P_n is a probability measure on $(C[0,\infty), \mathcal{B}(C[0,\infty)))$. If $\{P_n\}$ weakly converges to some P then it follows from Proposition 1.35 that P satisfies (1.3.2).

A general theory of weak convergence of probability measures tells us that if the two conditions in the statement of Theorem A.42 hold then there exists a subsequence $\{P_{n_k}\}$ that converges weakly. Indeed, we can prove that the two conditions do hold, and so a weak limit P satisfies (1.3.2). An interested reader may consult [21, Chapter 2] and [5, Chapter 2]. Consequently, under the weak limit P, the coordinate process $\{\pi_t\}$ satisfies the conditions in Definition 1.36.

The arguments above shows the following claim:

Theorem 1.37

There exists a Brownian motion on some probability space.

- P is called the *Wiener measure*.
- We also say that a process satisfying the requirements in Definition 1.36 is a *Wiener process*.
- An $\mathbb{R}^d$-valued process $W_t = (W_t^1, \dots, W_t^d)$, $t \geq 0$, is said to be a d-dimensional Brownian motion if each W_t^i is a Brownian motion and W_t^i and W_t^j are independent of each other for $i \neq j$.
- Let $P^{(i)}$, $i = 1, \dots, d$, be d copies of the Wiener measure on $(C[0,\infty), \mathcal{B}(C[0,\infty)))$. Then the product measure $P^0 := P^{(1)} \times \cdots \times P^{(d)}$ is called the *d-dimensional Wiener measure* on $(C([0,\infty); \mathbb{R}^d), \mathcal{B}(C([0,\infty); \mathbb{R}^d))$, and the coordinate process $W_t(\omega) := \omega(t)$, $t \geq 0$, is a d-dimensional Brownian motion under P^0.
- It is known that for any Brownian motion there exists a continuous modification (this follows from Kolmogorov's continuity theorem. See, e.g., [31, Chapter 2]). Hereafter, we always take this modification as a Brownian motion, i.e., any Brownian motion is assumed to be continuous.

Definition 1.38. Let $\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}$ be a filtration. We call $\{W_t\}_{t\geq 0}$ as a d-dimensional $\mathbb{F}$-Brownian motion if

(i) $\{W_t\}_{t\geq 0}$ is $\mathbb{F}$-adapted and a d-dimensional Brownian motion.

(ii) For $s \leq t$ the random variable $W_t - W_s$ is independent of $\mathcal{F}_s$.

- Let $\{W_t\}$ be a d-dimensional Brownian motion and consider the augmented natural filtration $\mathbb{G} = \{\mathcal{G}_t\}_{t\geq 0}$ generated by $\{W_t\}$, i.e., $\mathcal{G}_t := \sigma(\mathcal{F}_t^W \cup \mathcal{N})$, where $\mathcal{N}$ is the collection of all $\mathbb{P}$-null sets from $\mathcal{F}$. Then $\{W_t\}$ is also a d-dimensional $\mathbb{G}$-Brownian motion.
- It is known that the filtration $\mathbb{G}$ above satisfies the usual conditions (see, e.g., [21, Theorem 2.7.9]).

Problem 1.39. Let $\{W_t\}_{t\geq 0}$ be a d-dimensional Brownian motion. Show that

$$\sigma\left(\mathcal{F}_t^W \cup \mathcal{N}\right) = \sigma\left(\{\sigma(W_{t_1}, \dots, W_{t_n}) : 0 \leq t_1 < \cdots < t_n \leq t, \ \ n \geq 1\} \cup \mathcal{N}\right).$$

There are infinitely many Brownian motion on the same probability space, as seen in the following problem:

Problem 1.40. Let $\{W_t\}_{t\geq 0}$ be a Brownian motion. Then show that the processes defined by the following (i)–(iii) are all Brownian motions:

(i) $\{-W_t\}_{t\geq 0}$.

(ii) $\{W_{t+s} - W_s\}_{t\geq 0}$.

(iii) $\{cW_{t/(c^2)}\}_{t\geq 0}$.

Here $s > 0$ and $c \neq 0$.

Next we focus on an irregularity of the sample paths of a Brownian motion.

Theorem 1.41

Let $\{W_t\}$ be a Brownian motion. Then

$$\mathbb{P}(\{\omega \in \Omega : t \mapsto W_t(\omega) \text{ is not differentiable at } s \in \mathbb{Q} \cap [0,\infty)\}) = 1.$$

Proof. Fix $s \geq 0$, put $A_s = \{\omega : t \mapsto W_t(\omega) \text{ is differentiable at } s\}$, and take $\omega \in A_s$. Then the limit $\lim_{h\searrow 0}(W_{s+h}(\omega) - W_s(\omega))/h$ exists and is finite. In particular, there exist $\delta > 0$ and $h_0 > 0$ such that $|W_{s+h}(\omega) - W_s(\omega)|/h \leq \delta$, $0 < \forall h < h_0$. Hence $\sup_{n\geq 1} n|W_{s+1/n}(\omega) - W_s(\omega)| < \infty$, and so there exists $N \geq 1$ such that for $n \geq 1$ we have $n|W_{s+1/n}(\omega) - W_s(\omega)| \leq N$. This implies

$$A_s \subset \bigcup_{N\geq 1} \bigcap_{n\geq 1} \{n|W_{s+1/n} - W_s| \leq N\},$$

whence by the continuity of the probability measures

$$\mathbb{P}(A_s) \leq \lim_{N\to\infty} \inf_{n\geq 1} \mathbb{P}(n|W_{s+1/n} - W_s| \leq N).$$

Take $\xi \sim N(0,1)$ and use $W_{s+1/n} - W_s \sim N(0,1/n)$ to obtain

$$\inf_{n\geq 1} \mathbb{P}(n|W_{s+1/n} - W_s| \leq N) = \inf_{n\geq 1} \mathbb{P}(n\sqrt{1/n}|\xi| \leq N) = \inf_{n\geq 1} \mathbb{P}(|\xi| \leq Nn^{-1/2}) = 0.$$

Consequently we have $\mathbb{P}(A_s) = 0$. Therefore $\mathbb{P}(\cup_{s\in\mathbb{Q}\cap[0,\infty)} A_s) = 0$. $\square$

- Actually, we can show that the sample paths of a Brownian motion is not differentiable for any time almost surely We refer to [21, Theorem 2.9.18] for a proof.
- This fact suggests an unpredictability of Brownian motion in a pathwise way.

We shall see an irregularity of Brownian motions with a different criterion. To this end, we use the *total variation* of $\{W_t\}$ in $[0,t]$ for each $t > 0$, defined by

$$V_W([0,t]) := \sup_{k\geq 0} \sup_{\pi} \sum_{i=0}^{k} |W_{t_{i+1}} - W_{t_i}|,$$

where the second supremum is taken over the partitions $\pi : 0 = t_0 < t_1 < \cdots < t_k < t_{k+1} = t$ of $[0,t]$ having $k+1$ points.

Theorem 1.42

The total variation of $\{W_t\}$ is almost surely infinite, i.e., $\mathbb{P}(V_W([0,t]) = \infty,\ t > 0) = 1$.

Proof. First notice that for each partition π of $[0,t]$,

$$\mathbb{E}\sum_{t_i\in\pi}(W_{t_{i+1}}-W_{t_i})^2=\sum_{t_i\in\pi}(t_{i+1}-t_i)=t.$$

Then write $Z_i=(W_{t_{i+1}}-W_{t_i})^2-(t_{i+1}-t_i)$ and take $\xi\sim N(0,1)$. Clearly, $\{Z_i\}$ is independent and each Z_i has the same distribution as that of $(\xi^2-1)(t_{i+1}-t_i)$. Thus

$$\mathbb{E}\left[\left(\sum_{t_i\in\pi}(W_{t_{i+1}}-W_{t_i})^2-t\right)^2\right]=\mathbb{E}\sum_{t_i\in\pi}Z_i^2=\mathbb{E}[(\xi^2-1)^2]\sum_{t_i\in\pi}(t_{i+1}-t_i)^2.$$

Let π_n be a sequence of the partition such that $\Delta_n:=\sup_{t_i\in\pi_n}|t_{i+1}-t_i|\to 0$. Then the right-hand side of the equality just above is at most $t\mathbb{E}[(\xi^2-1)^2]\Delta_n$. Therefore,

$$Q_n:=\sum_{t_i\in\pi_n}(W_{t_{i+1}}-W_{t_i})^2\to t,\ \ n\to\infty,\ \ \text{in } L^2,$$

whence there exists a subsequence Q_{n_k} that converges almost surely.

Now, suppose that $\mathbb{P}(V_W([0,t])<\infty)>0$. By the continuity of Brownian sample paths, we have $\sup_{t_i\in\pi_{n_k}}|W_{t_{i+1}}-W_{t_i}|\to 0$, and so the probability of the event

$$t\le\lim_{k\to\infty}\left(\sup_{t_i\in\pi_{n_k}}|W_{t_{i+1}}-W_{t_i}|\right)\sum_{t_i\in\pi_{n_k}}|W_{t_{i+1}}-W_{t_i}|=0$$

is positive, which is impossible for $t>0$. Hence $\mathbb{P}(V_W([0,t])=\infty)$ for every $t>0$. Furthermore, since $V_W([0,s])\le V_W([0,t])$ for any $t>0$ and $s\in\mathbb{Q}$ with $s<t$, we have

$$1=\mathbb{P}(V_W([0,s])=\infty,\ s\in\mathbb{Q}\cap(0,\infty))\le\mathbb{P}(V_W([0,t])=\infty,\ t>0).$$

Thus the theorem follows. □

The proof of the theorem above implies that for each partition $0=t_0<t_1<\cdots<t_n<t_{n+1}=t$ such that $\Delta_n=\sup|t_{i+1}-t_i|\to 0$,

$$\langle W\rangle_t:=\lim\sum_{i=0}^{n}(W_{t_{i+1}}-W_{t_i})^2=t,\quad \text{in } L^2.$$

We call $\langle W\rangle_t$, $t\ge 0$, as the *quadratic variation* of $\{W_t\}$.

Definition 1.43. We say that an $\mathbb{R}^d$-valued $\mathbb{F}$-adapted process $\{X_t\}$ is an $\mathbb{F}$-*Markov process* if

$$\mathbb{E}[f(X_t)|\mathcal{F}_s]=\mathbb{E}[f(X_t)|X_s],\quad s\le t,$$

for any bounded Borel function f on $\mathbb{R}^d$.

- $\{X_t\}$ is simply called a Markov process if it is Markov with respect to $\{\mathcal{F}_t^X\}_{t\ge 0}$.

Theorem 1.44

Any d-dimensional $\mathbb{F}$-Brownian motion is $\mathbb{F}$-Markov.

Proof. Let $s\le t$. Since W_t-W_s is independent of $\mathcal{F}_s$, we can apply Lemma 1.45 below to obtain

$$\mathbb{E}[f(W_t)|\mathcal{F}_s]=\mathbb{E}[f(W_t-W_s+W_s)|\mathcal{F}_s]=g(W_s).$$

Here $g(y)=\mathbb{E}[f(W_t-W_s+y)]$.

On the other hand, $\sigma(W_s)\subset\mathcal{F}_s$ yields $\mathbb{E}[f(W_t)|\sigma(W_s)]=\mathbb{E}[\mathbb{E}[f(W_t)|\mathcal{F}_s]|\sigma(W_s)]=g(W_s)$, whence the claim follows. □

We have used the following lemma to show Theorem 1.44.

Lemma 1.45

Let $(S_i, \mathcal{S}_i)$, $i = 1, 2$, be measurable spaces. Suppose that an S_1-valued random variable X_1 is independet of a σ-algebra $\mathcal{G}$ and that an S_2-valued random variable X_2 is $\mathcal{G}$-measurable. Then for any bounded Borel function f on $(S_1 \times S_2, \mathcal{S}_1 \times \mathcal{S}_2)$ we have

$$\mathbb{E}[f(X_1, X_2)|\mathcal{G}] = \mathbb{E}[f(X_1, x)]|_{x=X_2}.$$

Proof. Let $A \in \mathcal{G}$. The assumption implies that $Z = (X_2, 1_A)$ is independent of X_1. So applying Theorem A.36, we have

$$\mathbb{E}[f(X, Y)1_A] = \int f(x, y)\xi\mu_{(X,Z)}(dx, dy, d\xi) = \int f(x, y)\xi\mu_X(dx)\mu_Z(dy, \xi),$$

where μ_V denotes the distribution of V. Thus by Fubini's theorem (Theorem A.35),

$$\mathbb{E}[f(X, Y)1_A] = \int \left[\int f(x, y)\xi\mu_X(dx)\right]\xi\mu_Z(dy, \xi) = \mathbb{E}[g(Y)1_A].$$

Since $A \in \mathcal{G}$ is arbitrary, we are done. □

Theorem 1.46: The strong Markov property for Brownian motions

Suppose that the filtration $\mathbb{F}$ is right-continuous. Let $\{W_t\}_{t\geq 0}$ be a d-dimensional $\mathbb{F}$-Brownian motion. Then, for any $\mathbb{F}$-stopping time τ and bounded Borel function f on $\mathbb{R}^d$, we heve

$$\mathbb{E}[1_{\{\tau<\infty\}}f(X_{\tau+t})|\mathcal{F}_\tau] = \mathbb{E}[1_{\{\tau<\infty\}}f(X_{\tau+t})|X_\tau], \quad t \geq 0.$$

Proof. First notice that for every bounded Borel measurable function f on $\mathcal{B}(\mathbb{R})$ there exists a sequence $\{f_n\}_{n=1}^\infty \subset C_b(\mathbb{R}^d)$ such that $f_n(x) \to f(x)$, $x \in \mathbb{R}^d$. To confirm this, recall that any Borel measurable function can be approximated by simple functions and the indicator function on $\prod_{i=1}^d (a_i, b_i]$ can be approximated by continuous functions. Thus, in view of this pointwise approximation and the dominated convergence theorem, we can assume $f \in C_b(\mathbb{R}^d)$ without loss of generality.

Let τ be a stopping time and put $\tau_n = (\lfloor n\tau \rfloor + 1)/n$, $n \in \mathbb{N}$. Fix $A \in \mathcal{F}_\tau$. Then,

$$\mathbb{E}\left[1_{\{\tau<\infty\}}f(W_{t+\tau_n})1_A\right] = \sum_{k=1}^\infty \mathbb{E}\left[f(W_{t+k/n})1_{A\cap\{\tau_n=k/n\}}\right].$$

Since $\tau \leq \tau_n$, we have $\mathcal{F}_\tau \subset \mathcal{F}_{\tau_n}$. Thus $A \cap \{\tau_n = k/n\} \in \mathcal{F}_{k/n}$. Then by Theorem 1.44,

$$\begin{aligned}\mathbb{E}[f(W_{t+\tau_n})1_{A\cap\{\tau_n=k/n\}}] &= \mathbb{E}[\mathbb{E}[f(W_{t+\tau_n})|\mathcal{F}_{k/n}]1_{A\cap\{\tau_n=k/n\}}] \\ &= \mathbb{E}[\mathbb{E}[f(W_{t+\tau_n})|W_{k/n}]1_{A\cap\{\tau_n=k/n\}}] \\ &= \mathbb{E}[\mathbb{E}[f(W_{t+k/n} - W_{k/n} + x)]|_{x=W_{k/n}}1_{A\cap\{\tau_n=k/n\}}].\end{aligned}$$

Therefore,

$$\mathbb{E}\left[1_{\{\tau<\infty\}}f(W_{t+\tau_n})1_A\right] = \mathbb{E}\left[1_{\{\tau<\infty\}}\mathbb{E}[f(W_t + x)]|_{x=W_{\tau_n}}1_A\right].$$

By the continuity of f and the dominated convergence theorem, letting $n \to \infty$, we obtain

$$\mathbb{E}\left[1_{\{\tau<\infty\}}f(W_{t+\tau})1_A\right] = \mathbb{E}\left[1_{\{\tau<\infty\}}\mathbb{E}[f(W_t + x)]|_{x=W_\tau}1_A\right]. \tag{1.3.3}$$

On the other hand, by Proposition 1.19, $\{W_t\}$ is progressively measurable. This together with Proposition 1.33 means that W_τ is $\mathcal{F}_\tau$-measurable. Thus, $\sigma(W_\tau) \subset \mathcal{F}_\tau$ and (1.3.3) holds for any event in $\sigma(W_\tau)$. Consequently,

$$\mathbb{E}\left[1_{\{\tau<\infty\}} f(W_{t+\tau}) \,|\, \mathcal{F}_\tau\right] = \left[1_{\{\tau<\infty\}} f(W_{t+\tau}) \,|\, \sigma(W_\tau)\right],$$

as required. □

Theorem 1.47

Suppose that the filtration $\mathbb{F}$ is right-continuous. Let $\{W_t\}_{t\geq 0}$ be a d-dimensional $\{\mathcal{F}_t\}_{t\geq 0}$-Brownian motion and θ an $\mathbb{F}$-stopping time with $\theta < \infty$, a.s. Then, $\tilde{W}_t := W_{t+\theta} - W_\theta$, $t \geq 0$, is also a d-dimensional Brownian motion with respect to $\{\mathcal{F}_{t+\theta}\}_{t\geq 0}$ and is independent of $\mathcal{F}_\theta$.

Proof. As in (1.3.3), we can show that

$$\mathbb{E}\left[e^{i\xi^\top(W_{t+\theta}-W_{s+\theta})} 1_A\right] = \mathbb{E}\left[\mathbb{E}[e^{i\xi^\top W_{t-s}}] 1_A\right], \quad t \geq s, \;\; A \in \mathcal{F}_{s+\theta}, \;\; \xi \in \mathbb{R}^d,$$

where $i = \sqrt{-1}$. Thus

$$\mathbb{E}\left[e^{i\xi^\top(W_{t+\theta}-W_{s+\theta})} \middle| \mathcal{F}_{s+\theta}\right] = e^{-(t-s)|\xi|^2/2}, \quad t \geq s, \;\; \xi \in \mathbb{R}^d.$$

This leads to the claims. □

Proposition 1.48

Let $\{W_t\}_{t\geq 0}$ be an $\mathbb{F}$-Brownian motion, and $\sigma \in \mathbb{R}$. Then the following three processes are $\mathbb{F}$-martingales.

(i) $\{W_t\}_{t\geq 0}$,

(ii) $\{W_t^2 - t\}_{t\geq 0}$,

(iii) $\{e^{\sigma W_t - (\sigma^2/2)t}\}_{t\geq 0}$.

Proof. Let $s \leq t$. (i). Since $W_t - W_s$ is independent of $\mathcal{F}_s$, we have

$$\mathbb{E}[W_t|\mathcal{F}_s] = \mathbb{E}[W_t - W_s + W_s|\mathcal{F}_s] = \mathbb{E}[W_t - W_s] + W_s = W_s.$$

(ii). We use the representation $W_t^2 - t = (W_t - W_s + W_s)^2 - t = (W_t - W_s)^2 - (t-s) + 2W_s(W_t - W_s) + W_s^2 - s$ to see

$$\mathbb{E}[W_t^2 - t|\mathcal{F}_s] = \mathbb{E}[(W_t - W_s)^2] - (t-s) + 2W_s\mathbb{E}[W_t - W_s] + W_s^2 - s = W_s^2 - s.$$

(iii). This follows from

$$\mathbb{E}[e^{\sigma W_t - (\sigma^2/2)t}|\mathcal{F}_s] = e^{\sigma W_s - (\sigma^2/2)s}\mathbb{E}[e^{\sigma(W_t - W_s) - (\sigma^2/2)(t-s)}] = e^{\sigma W_s - (\sigma^2/2)s}.$$

□

Problem 1.49. Apply Doob's maximal inequality to show that

$$\mathbb{E}\left[\exp\left(\sigma \sup_{0\leq t\leq T} |W_t|\right)\right] < \infty$$

for any $T > 0$ and $\sigma > 0$.

Let $\{W_t\}$ be a 1-dimensional Brownian motion. For $t \geq 0$ and $x \in \mathbb{R}$,

$$W_s^{t,x} := x + W_s - W_t, \quad s \geq t$$

is a Brownian motion starting at (t, x). Then, the probability density function

$$p(s, y \,|\, t, x) := \frac{\partial}{\partial y} \mathbb{P}(W_s^{t,x} \leq y) = \frac{e^{-|x-y|^2/2(s-t)}}{\sqrt{2\pi(s-t)}}, \quad s > t, \;\; y \in \mathbb{R}$$

of $W_s^{t,x}$ is called a *transition density* from (t, x) to (s, y). This satisfies second order parabolic partial differential equations

$$\partial_s p - \frac{1}{2}\partial_{yy}^2 p = 0, \tag{1.3.4}$$

$$\partial_t p + \frac{1}{2}\partial_{xx}^2 p = 0, . \tag{1.3.5}$$

The equation (1.3.4) is called the *forward Kolmogorov equation*, whereas (1.3.5) is called the *backward Kolmogorov equation.*

Let f be a bounded continuous function on $\mathbb{R}$. Then, by the backward Kolmogorov equation (1.3.5), the function

$$u(t, x) := \mathbb{E}[f(W_T^{t,x})], \quad (t, x) \in [0, T] \times \mathbb{R},$$

satisfies

$$\partial_t u(t, x) + \frac{1}{2}\partial_{xx}^2 u(t, x) = 0, \quad (t, x) \in [0, T) \times \mathbb{R},$$

and $u(T, x) = f(x)$, $x \in \mathbb{R}$.

CHAPTER 2

Stochastic Integrals

Standard textbooks for the contents of this chapter are, e.g., [31], [49], [45], [21].

In what follows, we fix a time maturity $T \in (0, \infty)$ and work on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0 \le t \le T}, \mathbb{P})$. For the technical reasons described in Chapter 1, we assume that $\mathbb{F}$ satisfies the usual conditions.

2.1 Construction

Let $\{W_t\}_{0 \le t \le T}$ be a one-dimensional $\{\mathcal{F}_t\}_{0 \le t \le T}$-standard Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$. As seen in Chapter 1, Brownian motions can be a mathematical model for unpredictable motions. One might expect that an infinitesimal analysis for Brownian motions can be available as in the case of the classical calculus. However, by Theorem 1.41, the sample paths of Brownian motions are not differentiable. Therefore, to say nothing of a differentiation, an integral $\int_0^t f_s dW_s$ cannot be defined via the classical change of variation formula $\int_0^t f_s (dW_s/ds) ds$. Moreover, since the total variation of any Brownian motion diverges (Theorem 1.42), an integral $\int_0^t f_s dW_s$ cannot also be defined by the so-called Lebesgue-Stieltjes integrals.

The case of simple processes

As in the case where the definition of the expectation, we start with the case of simple integrands.

Definition 2.1. We say that $\{\phi_t\}_{0 \le t \le T}$ is a *simple process* if there exist a partition $0 = t_0 < t_1 < \cdots < t_n < t_{n+1} = T$ of $[0, T]$, $\mathcal{F}_0$-measurable $\psi_0 \in L^2$ and $\mathcal{F}_{t_i}$-measurable $\varphi_i \in L^2$, $i = 0, \ldots, n$, such that

$$\phi_t(\omega) = \psi_0(\omega) 1_{\{0\}}(t) + \sum_{i=0}^{n} \varphi_i(\omega) 1_{(t_i, t_{i+1}]}(t), \quad (t, \omega) \in [0, T] \times \Omega. \tag{2.1.1}$$

For simple processes $\{\phi_t\}$, we define the *stochastic integral* or *Itô integral* on $[0, T]$ of $\{\phi_t\}$ with respect to $\{W_t\}$ by

$$I(\phi) \equiv \int_0^T \phi_s dW_s := \sum_{i=0}^{n} \varphi_i (W_{t_{i+1}} - W_{t_i}). \tag{2.1.2}$$

It should be mentioned that the values at leftmost point in $[t_i, t_{i+1}]$ are adopted for the integrals, which differs from the arbitrariness in the case of Riemann integrals.

Our first task is to confirm that the definition (2.1.2) is well-defined, i.e., (2.1.2) is independent of the representation (2.1.1) of ϕ_t as a simple process. Suppose that $\{X_t\}$ is represented as

$$\phi_t = \psi_0 1_{\{0\}}(t) + \sum_{i=1}^{m} \varphi_i' 1_{(s_i, s_{i+1}]}(t)$$

for some partition $0 = s_0 < s_1 < \cdots < s_m < s_{m+1} = T$, and $\mathcal{F}_{s_i}$-measurable $\varphi_i' \in L^2$, $i = 1, \ldots, m$. Then, with the common partition $0 = u_0 < u_1 < \cdots < u_k < u_{k+1} = T$, we see $\phi_t = \psi_0'' 1_{\{0\}}(t) + \sum_{i=0}^{k} \varphi_i'' 1_{(u_i, u_{i+1}]}(t)$, where φ_i'' is given by $\varphi_i'' = \varphi_j = \varphi_\ell'$ for appropriate j and ℓ. Since the interval $(u_i, u_{i+1}]$ is a subdivision of $(t_j, t_{j+1}]$ for some j, we have $(t_j, t_{j+1}] = \cup_{i=i_0}^{i_1} (u_i, u_{i+1}]$ for some $i_0 \leq i_1$. Hence, $\varphi_j(W_{t_{j+1}} - W_{t_j}) = \sum_{i=i_0}^{i_1} \varphi_j''(W_{u_{i+1}} - W_{u_i})$. A similar relation is obtained for the representation of $\varphi_\ell'(W_{s_{\ell+1}} - W_{s_\ell})$. Therefore,

$$I(\phi) = \sum_{j=0}^{n} \varphi_j(W_{t_{j+1}} - W_{t_j}) = \sum_{i=0}^{k} \varphi_i''(W_{u_{i+1}} - W_{u_i}) = \sum_{\ell=0}^{m} \varphi_\ell'(W_{s_{\ell+1}} - W_{s_\ell}).$$

This shows that (2.1.2) is well-defined.

Now, we shall define the Itô integrals for general integrands by extending the definition (2.1.2) in a natural way. To this end, we focus on the following fact:

Proposition 2.2

If $\{\phi_t\}$ is a simple process, then

$$\mathbb{E}\left[\left(\int_0^T \phi_s dW_s\right)^2\right] = \mathbb{E}\left[\int_0^T \phi_s^2 ds\right]. \tag{2.1.3}$$

Proof. Suppose that ϕ_t is represented as in (2.1.1). Then,

$$\begin{aligned}\left(\int_0^T \phi_t dW_t\right)^2 &= \sum_{i,j} \varphi_i \varphi_j (W_{t_{i+1}} - W_{t_i})(W_{t_{j+1}} - W_{t_j}) \\ &= \sum_{i=0}^{n} \varphi_i^2 (W_{t_{i_1}} - W_{t_i})^2 + 2\sum_{j>i} \varphi_i \varphi_j (W_{t_{i+1}} - W_{t_i})(W_{t_{j+1}} - W_{t_j}).\end{aligned}$$

By the independent increments property of $\{W_t\}$, for $j > i$ we have

$$\mathbb{E}[\varphi_i \varphi_j (W_{t_{i+1}} - W_{t_i})(W_{t_{j+1}} - W_{t_j})] = \mathbb{E}[\varphi_i (W_{t_{i+1}} - W_{t_i}) \varphi_j \mathbb{E}[W_{t_{j+1}} - W_{t_j} | \mathcal{F}_{t_j}]] = 0,$$

whence

$$\mathbb{E}\left[\left(\int_0^T \phi_t dW_t\right)^2\right] = \sum_{i=0}^{n} \mathbb{E}\left[\phi_i^2\right](t_{i+1} - t_i) = \mathbb{E}\int_0^T \phi_t^2 dt.$$

□

- The property (2.1.3) is called as the *isometry* of the Itô integrals.
- Proposition 2.2 means that for two simple processes $\{\phi_t\}$ and $\{\psi_t\}$

$$\mathbb{E}[(I(\phi) - I(\psi))^2] = \mathbb{E}\int_0^T (\phi_t - \psi_t)^2 dt.$$

  Thus, the L^2-error between $I(\phi)$ and $I(\psi)$ is equal to the mean squared error $\mathbb{E}\int_0^T (\phi_t - \psi_t)^2 dt$ of the two stochastic processes $\{\phi_t\}$ and $\{\psi_t\}$.

The case of square integrable processes

The preceding argument suggests that for a general process $\{\phi_t\}$ having approximate sequence $\{\phi_t^{(n)}\}$ of simple processes, the L^2-limit of $I(\phi^{(n)})$ is meaningful and can be defined as an integral of $\{\phi_t\}$.

We consider the class

$$\mathcal{L}_2 = \left\{ \{\phi_t\}_{0\le t\le T} : \mathbb{F}\text{-adapted, } \mathbb{E}\int_0^T \phi_t^2 dt < \infty \right\}.$$

Then we have the following:

Lemma 2.3

For any $\{\phi_t\} \in \mathcal{L}_2$, there exists a sequence $\{\phi_t^{(n)}\}$, $n \ge 1$, of simple processes such that

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^T |\phi_t - \phi_t^{(n)}|^2 dt\right] = 0.$$

Proof.* First, consider the case where $\phi_t(\omega)$ is continuous for any $\omega \in \Omega$ and uniformly bounded, i.e., $\sup_{(t,\omega)\in[0,T]\times\Omega} |\phi_t(\omega)| < \infty$. Then, the sequence

$$\phi_t^{(n)} := \phi_{k2^{-n}T} \quad t \in [k2^{-nT}, (k+1)2^{-n}T), \;\; k = 0, \dots, 2^n - 1, \quad n = 1, 2, \dots$$

of simple processes converges to $\phi_t(\omega)$ for any (t, ω). Further, it follows that $|\phi_t^{(n)} - \phi_t| \le 2\sup_{s,\omega} |\phi_s(\omega)| < \infty$, whence, by the dominated convergence theorem, $\mathbb{E}\int_0^T |\phi_t^{(n)} - \phi_t|^2 dt \to 0$.

Second, consider the case where $\{\phi_t\}$ is adapted and uniformly bounded. Then, by Proposition 1.21, the process

$$\phi_t^{(\varepsilon)} := \frac{1}{\varepsilon}\int_{t-\varepsilon}^t \phi_{s\vee 0} ds, \quad 0 \le t \le T$$

is adapted, uniformly bounded, and continuous. By [43, 定理 19.3], we have $\phi_t^{(\varepsilon)} \to \phi_t$ as $\varepsilon \to 0$ for almost every t. Moreover, there exists a sequence $\{\phi_t^{(n,\varepsilon)}\}$ of simple processes that approximate $\{\phi_t^{(\varepsilon)}\}$ for every $\varepsilon > 0$. Therefore, applying the dominated convergence theorem, we obtain

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty} \mathbb{E}\int_0^T |\phi_t^{(n,\varepsilon)} - \phi_t|^2 dt = 0.$$

Thus we have $\mathbb{E}\int_0^T |\phi_t^{(n,\varepsilon_n)} - \phi_t|^2 dt \to 0$ for some subsequence $\varepsilon_n \to 0$.

Third, consider the case where $\{\phi_t\}$ is adapted and is not necessarily bounded. Then, the process $\phi_t^{(N)} := \phi_t 1_{\{|\phi_t|\le N\}}$, $0 \le t \le T$, is adapted and bounded, and satisfies

$$\lim_{N\to\infty} \mathbb{E}\int_0^T |\phi_t^{(N)} - \phi_t|^2 dt = \lim_{N\to\infty} \mathbb{E}\int_0^T (\phi_t)^2 1_{\{|\phi_t|>N\}} dt = 0. \tag{2.1.4}$$

Hence, there exists a sequence $\{\phi_t^{(n,N)}\}$ of simple processes that approximate $\{\phi_t^{(N)}\}$ for every $N \ge 1$. This together with (2.1.4) implies that $\mathbb{E}\int_0^T |\phi_t^{(n,N_n)} - \phi_t|^2 dt \to 0$ for some subsequence $N_n \to \infty$. □

By Proposition 2.2 and Lemma 2.3, for any $\{\phi_t\} \in \mathcal{L}_2$ there exists a sequence $\{\phi_t^{(n)}\}_{0\le t\le T}$ of simple processes such that

$$\mathbb{E}|I(\phi^{(n)}) - I(\phi^{(m)})|^2 = \mathbb{E}\int_0^T |\phi_t^{(n)} - \phi_t^{(m)}|^2 dt, \quad m, n \in \mathbb{N},$$

$$\lim_{n\to\infty} \mathbb{E}\int_0^T |\phi_t - \phi_t^{(n)}|^2 dt = 0.$$

This shows that $\{I(\phi^{(n)})\}_{n=1}^{\infty}$ is a Cauchy sequence in L^2, whence there exists a limit $I(\phi) \in L^2$. Moreover, $I(\phi)$ does not depend on the choice of approximating simple processes $\{\phi_t^{(n)}\}$. Indeed, if $\{\psi_t^{(n)}\}_{0\le t\le T}$, $n \in \mathbb{N}$, are another simple processes such that $\mathbb{E}\int_0^T |\phi_t - \psi_t^{(n)}|^2 dt \to 0$, then

$$\begin{aligned}\mathbb{E}|I(\phi) - I(\psi^{(n)})|^2 &\le 2\mathbb{E}|I(\phi^{(n)}) - I(\psi^{(n)})|^2 + 2\mathbb{E}|I(\phi^{(n)}) - I(\phi)|^2 \\ &= 2\mathbb{E}\int_0^T |\phi_t^{(n)} - \psi_t^{(n)}|^2 dt + 2\mathbb{E}|I(\phi^{(n)}) - I(\phi)|^2 \to 0.\end{aligned}$$

The arguments above justify the following definition:

Definition 2.4. Let $\{\phi_t\} \in \mathcal{L}_2$ and $\{\phi_t^{(n)}\}$ be as in Lemma 2.3. Then we define the Itô integral $I(\phi) = \int_0^T \phi_t dW_t$ of $\{\phi_t\}$ by the L^2-limit of $I(\phi^{(n)})$.

Example 2.5. Let us compute $\int_0^T W_t dW_t$. In this case,

$$\phi_t^{(n)} = \sum_{j=0}^{2^n-1} W_{j2^{-n}T} 1_{(j2^{-n}T,(j+1)2^{-n}T]}(t), \quad 0 \le t \le T, \quad n = 1, 2, \dots$$

is an approximate sequence of $\{W_t\}$. Indeed,

$$\begin{aligned}&\mathbb{E}\left[\int_0^T (\phi_t^{(n)} - W_t)^2 dt\right] = \mathbb{E}\left[\sum_{j=0}^{2^n-1}\int_{j2^{-n}T}^{(j+1)2^{-n}T} (W_t - W_{j2^{-n}T})^2 dt\right] \\ &= \sum_{j=0}^{2^n-1}\int_{j2^{-n}T}^{(j+1)2^{-n}T} (t - j2^{-n}T)dt = \sum_{j=0}^{2^n-1} 2^{-1}((j+1)2^{-n}T - j2^{-n}T)^2 \to 0.\end{aligned}$$

Thus,

$$\int_0^T W_t dW_t = \lim_{n\to\infty}\sum_{j=0}^{2^n-1} W_{j2^{-n}T}(W_{(j+1)2^{-n}T} - W_{j2^{-n}T}) \ \text{ in } L^2.$$

Using $2y(x-y) = x^2 - y^2 - (x-y)^2$, we see

$$2\sum_{j=0}^{2^n-1} W_{j2^{-n}T}(W_{(j+1)2^{-n}T} - W_{j2^{-n}T}) = W_T^2 - \sum_{j=0}^{2^n-1}(W_{(j+1)2^{-n}T} - W_{j2^{-n}T})^2.$$

Further, the second term of the right-hand side in the equality just above converges to T in L^2. Therefore,

$$\int_0^T W_t dW_t = \frac{1}{2}W_T^2 - \frac{T}{2}.$$

Itô integral as stochastic processes

We shall define the stochastic integrals on $[0, t]$ for each $t \in [0, T]$, and then construct the processes of the integrals. For the simple process $\{\phi_t\}$ with representation (2.1.1),

$$I_t(\phi) = \int_0^t \phi_s dW_s := \int_0^T \phi_s 1_{\{s\le t\}} dW_s = \sum_{k=0}^{n} \varphi_k (W_{t_{k+1}\wedge t} - W_{t_k\wedge t}), \quad 0 \le t \le T.$$

That is, for $t \in (t_i, t_{i+1}]$, $I_t(\phi) = \sum_{k=0}^{i-1}\varphi_k(W_{t_{k+1}} - W_{t_k}) + \varphi_i(W_t - W_{t_i})$. The sample paths of $\{I_t(\phi)\}$ is clearly continuous almost surely.

Next, we introduce the class

$$\mathcal{M}_2 := \{\{M_t\}_{0\le t\le T} : \text{a.s. continuous, } \mathbb{F}\text{-martingales, } M_0 = 0,\ \mathbb{E}|M_T|^2 < \infty\}$$

of processes. Then we have the following fundamental result:

Proposition 2.6

For any simple process $\{\phi_t\}_{0\le t\le T}$, the process $\{I_t(\phi)\}_{0\le t\le T}$ is an $\mathbb{F}$-martingale, i.e., $\{I_t(\phi)\}\in\mathcal{M}_2$.

Proof. Let $\{\phi_t\}$ be given by (2.1.1). Then, for $t>s$,

$$\begin{aligned}\mathbb{E}[I_t(\phi)|\mathcal{F}_s] &= \sum_{k:t_k\le s}\varphi_k\mathbb{E}[W_{t_{k+1}\wedge t}-W_{t_k\wedge t}|\mathcal{F}_s]+\sum_{k:t_k>s}\mathbb{E}\left[\varphi_k\mathbb{E}[W_{t_{k+1}\wedge t}-W_{t_k\wedge t}|\mathcal{F}_{t_k}]|\mathcal{F}_s\right]\\ &= \sum_{k:t_k\le s}\varphi_k(W_{t_{k+1}\wedge s}-W_{t_k\wedge s})=I_s(\phi).\end{aligned}$$

□

For $t\in[0,T]$ and for $\{\phi_s\}_{0\le s\le t}\in\mathcal{L}_2$, we define $I_t(\phi)$ by the L^2-limit of the stochastic integrals $I_t(\phi^{(n)})$ of an approximating simple processes $\{\phi_s^{(n)}\}_{0\le s\le t}$. Then we have the following:

Theorem 2.7

For any $\{\phi_t\}_{0\le t\le T}\in\mathcal{L}_2$ there exists a modification process $\{J_t\}\in\mathcal{M}_2$ of $\{I_t(\phi)\}_{0\le t\le T}$. Namely, $\{J_t\}$ is a continuous $\mathbb{F}$-martingale and satisfies $\mathbb{P}(J_t=I_t(\phi))=1$ for $t\in[0,T]$.

Proof.* By Doob's maximal inequality (Theorem 1.34), for any fixed $\varepsilon>0$,

$$\begin{aligned}\mathbb{P}\left(\sup_{0\le t\le T}|I_t(\phi^{(n)})-I_t(\phi^{(m)})|>\varepsilon\right) &\le \frac{1}{\varepsilon^2}\mathbb{E}\left[|I_T(\phi^{(n)})-I_T(\phi^{(m)})|^2\right]\\ &= \frac{1}{\varepsilon^2}\mathbb{E}\int_0^T|\phi_t^{(n)}-\phi_t^{(m)}|^2dt\to 0\end{aligned}$$

as $m,n\to\infty$. Hence there exists a subsequence $n_k\nearrow\infty$ such that

$$\mathbb{P}\left(\sup_{0\le t\le T}|I_t(\phi^{(n_{k+1})})-I_t(\phi^{(n_k)})|>2^{-k}\right)\le 2^{-k}.$$

Then we apply Borel-Cantelli lemma (Lemma A.12) to obtain

$$\mathbb{P}\left(\bigcup_{k\ge1}\bigcap_{j\ge k}\left\{\sup_{0\le t\le T}|I_t(\phi^{(n_{k+1})})-I_t(\phi^{(n_k)})|>2^{-k}\right\}\right)=1.$$

From this, for almost every $\omega\in\Omega$ there exists $k_0(\omega)$ such that

$$\sup_{0\le t\le T}|I_t(\phi^{(n_{k+1})})(\omega)-I_t(\phi^{(n_k)})(\omega)|\le 2^{-k},\quad k\ge k_0(\omega).$$

This implies that for almost every ω the sequence $I_t(\phi^{(n_k)})(\omega)$ of functions converges to some $J_t(\omega)$ uniformly on $[0,T]$. We set $J_t(\omega)=0$ for ω such that the limit $I_t(\phi^{(n_k)})(\omega)$ does not exist. Then $\{J_t\}$ is continuous almost surely and a modification of $\{I_t(\phi)\}$. Indeed, by Fatou's lemma,

$$\mathbb{E}[(J_t-I_t(\phi))^2]=\mathbb{E}[\lim_{n\to\infty}(I_t(\phi^{(n)})-I_t(\phi))^2]\le\liminf_{n\to\infty}\mathbb{E}[(I_t(\phi^{(n)})-I_t(\phi))^2]=0$$

whence $J_t=I_t(\phi)$ a.s.

Next we will show that $\{J_t\}$ is a martingale. It is clear that $J_t \in L^1$ for every t. By Problem 1.18 and Proposition 1.17, $\{I_t(\phi)\}$ and $\{J_t\}$ are adapted. Moreover, for $s \le t$, the inequality $(a+b)^2 \le 2(a^2+b^2)$ for $a, b \in \mathbb{R}$ and Jensen's inequality for conditional expectations yield

$$\begin{aligned}\mathbb{E}\left|\mathbb{E}[J_t|\mathcal{F}_s] - J_s\right|^2 &\le 2\mathbb{E}\left|\mathbb{E}[J_t|\mathcal{F}_s] - \mathbb{E}[I_t(\phi^{(n)})|\mathcal{F}_s]\right|^2 + 2\mathbb{E}\left|I_s^{(n)} - J_s\right|^2 \\ &\le 2\left(\mathbb{E}\left|I_t^{(n)} - J_t\right|^2 + \mathbb{E}\left|I_s^{(n)} - J_s\right|^2\right) \\ &\to 0,\end{aligned}$$

whence $\mathbb{E}[J_t|\mathcal{F}_s] = J_s$. Therefore we have $\{J_t\} \in \mathcal{M}$. □

- In what follows, the process $I_t(\phi) = \int_0^t \phi_s dW_s$, $0 \le t \le T$, denotes the continuous modification $\{J_t\}_{0 \le t \le t}$ in Theorem 2.7.
- The processes of the stochastic integrals can be seen as a linear map from $\mathcal{L}_2$ into $\mathcal{M}_2$. Namely, for $\{\phi_t\}$, $\{\psi_t\} \in \mathcal{L}_2$ and $\alpha, \beta \in \mathbb{R}$ we have $I_t(\alpha\phi + \beta\psi) = \alpha I_t(\phi) + \beta I_t(\psi)$.
- We define, for $s \le t$,

$$\int_s^t \phi_u dW_u = \int_0^t \phi_u dW_u - \int_0^s \phi_u dW_u.$$

  Then it follows that for $A \in \mathcal{F}_s$

$$\int_s^t 1_A \phi_u 1_{\{s<u\}} dW_u = 1_A \int_s^t \phi_u dW_u, \tag{2.1.5}$$

  which can be verified by the approximation argument with simple processes.

Next, we consider the stopped process $I_{\cdot\wedge\tau}(\phi)$ defined for an $\mathbb{F}$-stopping time τ (see Chapter 1). The following proposition gives a representation for $I_{t\wedge\tau}(\phi)$:

Proposition 2.8

For any $\{\phi_t\} \in \mathcal{L}_2$ and $\mathbb{F}$-stopping τ,

$$\int_0^{t\wedge\tau} \phi_s dW_s = \int_0^t \phi_s 1_{\{s\le\tau\}} dW_s, \quad 0 \le t \le T, \quad \text{a.s.}$$

Proof. It suffices to show the proposition in the case that τ is $[0,t]$-valued for some fixed $t \in [0,T]$.

First assume that τ is represented as $\tau = \sum_{i=1}^n t_i 1_{A_i}$, where $0 < t_1 < \cdots < t_n = t$ and $A_i \in \mathcal{F}_{t_i}$ such that $\{A_i\}$ is disjoint. From $\{s > \tau\} = \cup_{i=1}^n \{s > t_i\} \cap A_i$, the fact that $s \mapsto 1_{A_i} 1_{\{s>t_i\}} \phi_s$ is adapted and the linearity of the stochastic integrals we obtain

$$\int_0^t 1_{\{s>\tau\}} \phi_s dW_s = \sum_{i=1}^n \int_0^t 1_{A_i} 1_{\{s>t_i\}} \phi_s dW_s.$$

Applying (2.1.5) to the right-hand side in the equality just above, we find

$$\int_0^t 1_{\{s>\tau\}} \phi_s dW_s = \sum_{i=1}^n 1_{A_i} \int_{t_i}^t \phi_s dW_s = \int_\tau^t \phi_s dW_s.$$

For a general $[0,t]$-valued stopping time τ, we consider an approximation of τ with

$$\tau_n = \sum_{i=0}^{2^n} (i+1) 2^{-n} t 1_{\{i2^{-n}t \le \tau < (i+1)2^{-n}t\}}.$$

Since $\tau_n \to \tau$ a.s. and $s \mapsto \int_0^s \phi_u dW_u$ is continuous almost surely, the sequence of the random variables $\int_0^{\tau_n} \phi_s dW_s$ converges to $\int_0^{\tau} \phi_s dW_s$ almost surely.

On the other hand, by the dominated convergence theorem, as $n \to \infty$,

$$\mathbb{E}\left|\int_0^t 1_{\{s\le\tau\}}\phi_s dW_s - \int_0^t 1_{\{s\le\tau_n\}}\phi_s dW_s\right|^2 = \mathbb{E}\int_0^t 1_{\{\tau<s\le\tau_n\}}\phi_s^2 ds \to 0.$$

Therefore, $\int_0^t 1_{\{s\le\tau_{n_k}\}}\phi_s dW_s \to \int_0^t 1_{\{s\le\tau\}}\phi_s dW_s$ a.s. for some subsequence $n_k \nearrow \infty$. Thus the proposition follows. □

Itô integrals for general integrands

We shall define the stochastic integrals for the class

$$\mathcal{L}_{2,\mathrm{loc}} := \left\{\{\phi_t\}_{0\le t\le T} : \mathbb{F}\text{-adapted}, \int_0^T \phi_t^2 dt < \infty \text{ a.s.}\right\}$$

that is larger than $\mathcal{L}_2$. To this end, we introduce local martingales.

Definition 2.9. We say that $\{M_t\}_{t\ge0}$ is an $\mathbb{F}$-*local martingale* if there exists an increasing sequence $\{\tau_n\}_{n\ge1}$ of stopping times such that $\lim_{n\to\infty}\tau_n = \infty$ and that $\{M_t^{\tau_n}\}_{t\ge0}$ is an $\mathbb{F}$-martingale.

Denote by $\mathcal{M}_{\mathrm{loc}}$ the collection of all $\mathbb{F}$-local martingales $M = \{M_t\}_{0\le t\le T}$ with almost surely continuous paths and $M_0 = 0$. For $\{\phi_t\} \in \mathcal{L}_{2,\mathrm{loc}}$, we consider the random variable

$$\tau_n = \inf\left\{s \in [0,T] : \int_0^s \phi_u^2 du \ge n\right\}.$$

Here $\inf\emptyset = \infty$ by convention. Then, since $\{\tau_n \le t\} = \{\int_0^t \phi_s^2 ds \ge n\}$ and $\int_0^t \phi_s^2 ds$ is $\mathcal{F}_t$-measurable by Proposition 1.21, each τ_n is a stopping time.

Now, define the process $\{\phi_t^{(n)}\}$ by

$$\phi_t^{(n)} = \phi_t 1_{\{t\le\tau_n\}}.$$

Then $\{\phi_t^{(n)}\} \in \mathcal{L}_2$. By definition, we obtain

$$\int_0^t \phi_s^{(n)} dW_s = \int_0^t 1_{\{s\le\tau_n\}}\phi_s^{(n+1)} dW_s.$$

Moreover, by Proposition 2.8,

$$\int_0^t \phi_s^{(n)} dW_s = \int_0^{t\wedge\tau_n} \phi_s^{(n+1)} dW_s.$$

Therefore, on the event $\{t \le \tau_n\} = \{\int_0^t \phi_s^2 ds < n\}$ we have $\int_0^t \phi^{(n)} dW_s = \int_0^t \phi_s^{(n+1)} dW_s$. Also, since

$$\bigcup_{n\ge0}\left\{\int_0^t \phi_s^2 ds < n\right\} = \left\{\int_0^t \phi_u^2 du < +\infty\right\},$$

we can consistently define $\{\tilde{J}(\phi)_t\}$ by

$$\tilde{J}(\phi)_t := \int_0^t \phi_s^{(n)} dW_s, \quad 0 \le t \le \tau_n \wedge T.$$

Then $\{\tilde{J}(\phi)_t\} \in \mathcal{M}_{\mathrm{loc}}$ and $\tilde{J}(\phi)_t = \int_0^t \phi_s dW_s$ for any $\{\phi_t\} \in \mathcal{L}_2$. We write $\tilde{J}(\phi)_t = \int_0^t \phi_s dW_s$, $0 \le t \le T$, and call it the *Itô integral* or *stochastic integral* of $\{\phi_t\} \in \mathcal{L}_{2,loc}$.

Multidimensional cases

We shall define the Itô integrals for multidimensional Brownian motions. Let $W_t = (W_t^1, \dots, W_t^m)$, $t \geq 0$, be an m-dimensional $\mathbb{F}$-Brownian motion.

Definition 2.10. Let $\theta_t = (\theta_t^1, \dots, \theta_t^m)$, $0 \leq t \leq T$, be an $\mathbb{R}^m$-valued process such that $\{\theta_t^i\}_{0\leq t\leq T} \in \mathcal{L}_{2,\mathrm{loc}}$ for each $i = 1, \dots, m$. Then, we define the Itô integral of $\{\theta_t\}$ with respect to $\{W_t\}$ by

$$\int_0^t \theta_s^\top dW_s = \sum_{i=1}^m \int_0^t \theta_s^i dW_s^i.$$

Similarly, for $\mathbb{R}^{d\times m}$-valued process $\sigma_t = (\sigma_t^{ij})$, $0 \leq t \leq T$, such that $\{\sigma_t^{ij}\}_{0\leq t\leq T} \in \mathcal{L}_{2,\mathrm{loc}}$ for each i, j, we define the Itô integral of $\{\sigma_t\}$ with respect to $\{W_t\}$ by

$$\int_0^t \sigma_s dW_s = \left(\sum_{j=1}^m \int_0^t \sigma_s^{1j} dW_s^j, \cdots, \sum_{j=1}^m \int_0^t \sigma_s^{dj} dW_s^j \right)^\top.$$

Pathwise construction

Assume here that $m = 1$, and let $(\phi_t)_{t\geq 0}$ be a continuous adapted process. For each $n \in \mathbb{N}$, we define the sequence $\{\tau_i^n\}_{i=0}^\infty$ of the stopping times by

$$\begin{cases} \tau_0^n = 0, \\ \tau_{i+1}^n = \inf\{t \geq \tau_i^n : |\phi_t - \phi_{\tau_i^n}| \geq 2^{-n}\}, \quad i \in \mathbb{N} \cup \{0\}, \end{cases}$$

Further, for every $n \in \mathbb{N}$, we define the process $(Y_t^n)_{t\geq 0}$ by

$$Y_t^n = \sum_{i=1}^k \phi_{\tau_{i-1}^n}(W_{\tau_i^n} - W_{\tau_{i-1}^n}) + \phi_{\tau_k^n}(W_t - W_{\tau_i^n}), \quad t \in [\tau_k^n, \tau_{k+1}^n), \quad k \in \mathbb{N} \cup \{0\},$$

with convention $\sum_{i=1}^0 = 0$. Then the process (Y_t^n) converges to the corresponding Itô integral *almost surely*.

Theorem 2.11

For $T \in (0, \infty)$, we have

$$\lim_{n\to\infty} \sup_{0\leq t\leq T} \left| Y_t^n - \int_0^t \phi_s dW_s \right| \to 0, \quad \text{a.s.}$$

Proof. Note that Y_t^n can be written as $Y_t^n = \int_0^t \phi_s^n dW_s$ where $\phi_s^n = \phi_{\tau_k^n}$ for $t \in [\tau_k^n, \tau_{k+1}^n)$. Then, by definition, $|\phi_t^n - \phi_t| \leq 2^{-n}$. Thus, using Doob's maximal inequality, we see

$$\mathbb{E} \sup_{0\leq t\leq T} \left| Y_t^n - \int_0^t \phi_s dW_s \right|^2 \leq 4\mathbb{E} \int_0^T |\phi_s^n - \phi_s|^2 ds \leq 4T 2^{-2n}.$$

This together with Cauchy-Schwartz inequality yields

$$\mathbb{E} \sum_{n=1}^\infty \sup_{0\leq t\leq T} \left| Y_t^n - \int_0^t \phi_s dW_s \right| = \sum_{n=1}^\infty \mathbb{E} \sup_{0\leq t\leq T} \left| Y_t^n - \int_0^t \phi_s dW_s \right| \leq \sum_{n=1}^\infty 2\sqrt{T} 2^{-n} < \infty,$$

whence

$$\sum_{n=1}^\infty \sup_{0\leq t\leq T} \left| Y_t^n - \int_0^t \phi_s dW_s \right| < \infty \quad \text{a.s.}$$

Thus the theorem follows. □

2.2 Itô Formula

Recall that if the function $f(t, x(t))$ is smooth, then the *chain rule*

$$\frac{df(t,x(t))}{dt} = \frac{\partial f}{\partial t}(t,x(t)) + \frac{\partial f}{\partial x}(t,x(t))\frac{dx(t)}{dt}$$

holds. By the fundamental theorem in calculus, this can be written in the integral form

$$f(t,x(t)) = f(0,x(0)) + \int_0^t \frac{\partial f}{\partial s}(s,x(s))ds + \int_0^t \frac{\partial f}{\partial x}(s,x(s))dx(s).$$

In this section, we shall derive its stochastic version, i.e., a chain rule for $f(t, X_t)$ when X_t is a stochastic process.

In what follows, we fix an m-dimensional $\mathbb{F}$-Brownian motion $W_t = (W_t^1, \ldots, W_t^m)$, $0 \le t \le T$.

Itô processes

Definition 2.12. A d-dimensional process $X_t = (X_t^1, \ldots, X_t^d)$, $0 \le t \le T$, is called an *Itô process* if each component is written as

$$X_t^i = X_0^i + \int_0^t K_s^i ds + \sum_{j=1}^m \int_0^t H_s^{ij} dW_s^j, \quad 0 \le t \le T, \ \ i = 1, \ldots, d, \tag{2.2.1}$$

where X_0^i is $\mathcal{F}_0$-measurable, $\{K_t^i\}$ and $\{H_t^{ij}\}$ are adapted with $\int_0^T |K_t^i| dt < \infty$, $\int_0^T (H_t^{ij})^2 dt < \infty$, a.s., $i = 1, \ldots, d$, $j = 1, \ldots, m$.

- Propositions 1.17 and 1.21 means that the processes $\int_0^t K_s^i ds$, $i = 1, \ldots, d$, are adapted and so is $\{X_t\}$.

It should be noted that the representation of an Itô process is uniquely determined. To see this, assume $m = d = 1$ for simplicity and that $\{X_t\}$ has representations

$$X_t = X_0 + \int_0^t K_s ds + \int_0^t H_s dW_s = X_0 + \int_0^t K_s' ds + \int_0^t H_s' dW_s.$$

Then,

$$A_t := \int_0^t (K_s - K_s')ds = \int_0^t (H_s - H_s')dW_s, \quad 0 \le t \le T$$

is a local martingale, whence, by Lemma 2.13 below, we necessarily have $A_t = 0$ a.e. This yileds $K_t = K_t'$, $dt \times \mathbb{P}$-a.e., and so $H_t = H_t'$, $dt \times \mathbb{P}$-a.e.

Lemma 2.13

If the Itô process $Y_t = \int_0^t b_s ds$, $0 \le t \le T$, is a local martingale, then $b_t = 0$, $dt \times \mathbb{P}$-a.e.

Proof.* Let $\{\tau_n\}_{n=0}^\infty$ be a sequence of the stopping times such that $\tau_n \nearrow +\infty$ and $\{Y_t^{\tau_n}\}_{0 \le t \le T}$ is a martingale. Then since

$$Y_t^{\tau_n} = Y_{t \wedge \tau_n} = \int_0^t \tilde{b}_u^{(n)} du,$$

where $\tilde{b}_u^{(n)} = b_u 1_{\{u < \tau_n\}}$, the martingale property implies

$$\mathbb{E} \int_s^t \tilde{b}_u^{(n)} du = 0, \quad 0 \le s < t \le T.$$

Therefore,

$$\mathcal{A} = \left\{ A \in \mathcal{B}([0,T]) : \mathbb{E} \int_A \tilde{b}_u^{(n)} du = 0 \right\}$$

forms a σ-algebra that contains $\{(s,t] : 0 \le s \le t \le T\}$. This means that $\mathcal{A} = \sigma((s,t] : 0 \le s \le t \le T) = \mathcal{B}([0,T])$. Consequently, we have

$$\mathbb{E} \int_A \tilde{b}_u^{(n)} du = 0, \quad A \in \mathcal{B}([0,T]),$$

whence $\tilde{b}_t^{(n)} = 0$, $dt \times \mathbb{P}$-a.e. Letting $n \to \infty$ in this equality, we obtain the lemma. □

Chain rule

The following theorem gives a chain rule for Itô processes:

Theorem 2.14: Itô formula

Let $X_t = (X_t^1, \dots, X_t^d)$, $0 \le t \le T$, be an Itô process with representation

$$X_t^i = X_0^i + \int_0^t K_s^i ds + \sum_{j=1}^m \int_0^t H_s^{ij} dW_s^j, \quad i = 1, \dots, d.$$

Suppose that $f \in C^{1,2}([0,T] \times \mathbb{R}^d)$. Then $\{f(t, X_t)\}_{0 \le t \le T}$ is an Itô process and represented as

$$\begin{aligned} f(t, X_t) &= f(0, X_0) + \sum_{i=1}^d \sum_{j=1}^m \int_0^t \partial_{x_i} f(s, X_s) H_s^{ij} dW_s^j \\ &\quad + \int_0^t \left\{ \partial_s f(s, X_s) + \sum_{i=1}^d \partial_{x_i} f(s, X_s) K_s^i + \frac{1}{2} \sum_{i,j=1}^d \sum_{k=1}^m \partial^2_{x_i x_j} f(s, X_s) H_s^{ik} H_s^{jk} \right\} ds. \end{aligned}$$

It is useful to state the Itô formula in the case of $m = d = 1$.

Corollary 2.15

Assume $m = d = 1$. Let $\{X_t\}$ be an Itô process with representation

$$X_t = X_0 + \int_0^t K_s ds + \int_0^t H_s dW_s. \tag{2.2.2}$$

Suppose that $f \in C^{1,2}([0,T] \times \mathbb{R})$. Then we have

$$\begin{aligned} f(t, X_t) &= f(0, X_0) + \int_0^t \partial_x f(s, X_s) H_s dW_s \\ &\quad + \int_0^t \left\{ \partial_s f(s, X_s) + \partial_x f(s, X_s) K_s + \frac{1}{2} \partial^2_{xx} f(s, X_s) H_s^2 \right\} ds. \end{aligned}$$

- The representation (2.2.2) of an Itô process is often written as the *differential form*

  $$dX_t = K_t dt + H_t dW_t.$$

  Notice that this is only a formal expression and a simplified way of representing the integral form (2.2.2). Further, for any adapted process $\{\sigma_t\}$ such that

  $$\int_0^T |\sigma_t| (|K_t| + |H_t|^2) dt < \infty, \quad \text{a.s.},$$

we define

$$\int_0^t \sigma_s dX_s = \int_0^t \sigma_s K_s ds + \int_0^t \sigma_s H_s dW_s.$$

With this definition, we can write

$$\sigma_t dX_t = \sigma_t (K_t dt + H_t dW_t).$$

Multidimensional cases are treated in a similar way.

Writing down the Itô formula in one dimension with the differential form, we have

$$df(t, X_t) = \partial_t f(t, X_t)dt + \partial_x f(t, X_t)dX_t + \frac{1}{2}\partial^2_{xx} f(t, X_t)(H_t)^2 dt. \tag{2.2.3}$$

Now suppose that $f(t, x(t))$ is smooth. Then the Taylor expansion up to 2nd terms gives

$$f(t + \Delta t, x(t + \Delta t)) - f(t, x(t)) = \partial_t f(t, x(t))\Delta t + \partial_x f(t, x(t))x'(t)\Delta t$$
$$+ \frac{1}{2}\partial^2_{tt} f(t, x(t))(\Delta t)^2 + \partial^2_{tx} f(t, x(t))\Delta t x'(t)\Delta t + \frac{1}{2}\partial^2_{xx} f(t, x(t))x''(t)(\Delta t)^2 + o((\Delta t)^2).$$

Formally, this can be written as

$$\begin{aligned} df(t, x(t)) = \partial_t f(t, x(t))dt + \partial_x f(t, x(t))dx(t) + \frac{1}{2}\partial^2_{tt} f(t, x(t))dtdt \\ + \partial^2_{tx} f(t, x(t))dtdx(t) + \frac{1}{2}\partial^2_{xx} f(t, x(t))dx(t)dx(t). \end{aligned}$$

Comparing each term in the equality just above with one in (2.2.3), we obtain

$$\begin{aligned} dtdt &= 0, \\ dtdX_t &= K_t dtdt + H_t dW_t dt = 0, \\ dX_t dX_t &= K_t^2 dtdt + 2K_t H_t dtdW_t + H_t^2 dW_t dW_t = H_t^2 dt, \end{aligned}$$

from which the *Itô's rule*:

$$dtdt = 0, \quad dtdW_t = 0, \quad dW_t dW_t = dt$$

is derived. In multidimensional cases, similarly we have

$$dtdW_t^i = 0, \quad dW_t^i dW_t^j = \delta_{ij} dt$$

where δ_{ij} is the Kronecker delta. Consequently, the chain rule of $f(t, X_t)$ can be derived by expanding it up to 2nd terms as follows:

$$\begin{aligned} df(t, X_t) = \partial_t f(t, X_t)dt + \partial_x f(t, X_t)dX_t + \frac{1}{2}\partial^2_{tt} f(t, X_t)dtdt \\ + \partial^2_{tx} f(t, X_t)dtdX_t + \frac{1}{2}\partial^2_{xx} f(t, X_t)dX_t dX_t \end{aligned}$$

and then by applying the Itô's rule to the expansion.

Proof of Theorem 2.14. We will show the claim in the case where

$$m = d = 1, \;\; f \text{ do not depend on } t, \;\; f'(x) \text{ and } f''(x) \text{ are bounded}, \;\; \{H_t\} = \{H_t^{ij}\} \in \mathcal{L}_2.$$

For the general case we refer to the references given in the last part of these notes.

First, assume that $\{K_t\} = \{K_t^i\}$ and $\{H_t\}$ are simple processes. Taylor's theorem gives

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + (x - x_0)^2 r(x, x_0), \tag{2.2.4}$$

where $r(x, x_0)$ is a bounded function such that $\lim_{x \to x_0} r(x, x_0) = 0$. We may assume that K_t and H_t have a common partition $0 = t_0 < t_1 < \cdots < t_N < t_{N+1} = t$ without loss of generality. Then we use the representation

$$f(X_t) = f(X_0) + \sum_{k=0}^{N} \Delta f_k$$

with $\Delta f_k := f(X_{t_{k+1}}) - f(X_{t_k})$. Furthermore, we divide Δf_k as follows:

$$\Delta f_k = \sum_{j=1}^{2^m} (f(X_{s_j^m}) - f(X_{s_{j-1}^m})),$$

where $s_j^m = t_k + j2^{-m}(t_{k+1} - t_k)$. Since K_t and H_t are constant on $[t_k, t_{k+1})$, we have

$$X_{s_j^m} - X_{s_{j-1}^m} = K_{t_k} 2^{-m}(t_{k+1} - t_k) + H_{t_k}(W_{s_j^m} - W_{s_{j-1}^m}).$$

Applying (2.2.4) to $f(X_{s_j^m}) - f(X_{s_{j-1}^m})$, we obtain

$$\begin{aligned}
\Delta f_k = & \sum_{j=1}^{2^m} f'(X_{s_{j-1}^m})(K_{t_k} 2^{-m}(t_{k+1} - t_k) + H_{t_k}(W_{s_j^m} - W_{s_{j-1}^m})) \\
& + \sum_{j=1}^{2^m} \frac{1}{2} f''(X_{s_{j-1}^m})(K_{t_k} 2^{-m}(t_{k+1} - t_k) + H_{t_k}(W_{s_j^m} - W_{s_{j-1}^m}))^2 \\
& + \sum_{j=1}^{2^m} (K_{t_k} 2^{-m}(t_{k+1} - t_k) + H_{t_k}(W_{s_j^m} - W_{s_{j-1}^m}))^2 r(X_{s_j^m}, X_{s_{j-1}^m}).
\end{aligned} \tag{2.2.5}$$

By the boundedness of $f'(x)$, the first term of the right-hand side in (2.2.5) converges to

$$K_{t_k} \int_{t_k}^{t_{k+1}} f'(X_s) ds + H_{t_k} \int_{t_k}^{t_{k+1}} f'(X_s) dW_s$$

in L^2 as $m \to \infty$.

Next, the second term of the right-hand side in (2.2.5) is written as $I_1 + I_2 + I_3$ with

$$\begin{aligned}
I_1 &= \frac{1}{2} \cdot 2^{-m}(t_{k+1} - t_k)^2 K_{t_k}^2 \sum_{j=1}^{2^m} f''(X_{s_{j-1}^m}) 2^{-m}, \\
I_2 &= 2^{-m}(t_{k+1} - t_k) K_{t_k} H_{t_k} \sum_{j=1}^{2^m} f''(X_{s_{j-1}^m})(W_{s_j^m} - W_{s_{j-1}^m}), \\
I_3 &= \frac{1}{2} H_{t_k}^2 \sum_{j=1}^{2^m} f''(X_{s_{j-1}^m})(W_{s_j^m} - W_{s_{j-1}^m})^2.
\end{aligned}$$

Since $f''(x)$ is bounded, as $m \to \infty$, the random variable $\sum_{j=1}^{2^m} f''(X_{s_{j-1}^m}) 2^{-m}$ converges to $\int_{t_k}^{t_{k+1}} f''(X_s) ds$ almost surely, and $\sum_{j=1}^{2^m} f''(X_{s_{j-1}^m})(W_{s_j^m} - W_{s_{j-1}^m})$ converges to $\int_{t_k}^{t_{k+1}} f''(X_s) dW_s$

in L^2, from which $I_1 + I_2$ converges to 0 in L^2. To see a limiting behavior of I_3, observe

$$\mathbb{E}\left[\left(\sum_{j=1}^{2^m} f''(X_{s_{j-1}^m})(W_{s_j} - W_{s_{j-1}^m})^2 - \int_{t_k}^{t_{k+1}} f''(X_s)ds\right)^2\right]$$

$$\leq 2\mathbb{E}\left[\left\{\sum_{j=1}^{2^m} f''(X_{s_{j-1}^m})((W_{s_j^m} - W_{s_{j-1}^m})^2 - 2^{-m})\right\}^2\right]$$

$$+ 2\mathbb{E}\left[\left(\sum_{j=1}^{2^m} f''(X_{s_{j-1}^m})2^{-m} - \int_{t_k}^{t_{k+1}} f''(X_s)ds\right)^2\right].$$

By the boundedness of $f''(x)$ and the argument in the proof of Theorem 1.42, there exist positive constants C_1 and C_2 such that the right-hand side in the inequality just above is at most

$$C_1\mathbb{E}\left[\sum_{j=1}^{2^m}((W_{s_j^m} - W_{s_{j-1}^m})^2 - 2^{-m})^2\right] \leq C_2 2^{-m}$$

Therefore, I_3 converges to $(1/2)H_{t_k}^2 \int_{t_k}^{t_{k+1}} f''(X_s)ds$ in L^2 as $m \to \infty$.

Moreover, the 3rd term of the right-hand side in (2.2.5) is at most

$$2\sup_j |r(X_{s_j^m}, X_{s_{j-1}^m})| \left\{K_{t_k}^2(t_{k+1} - t_k)^2 + H_{t_k}^2 \sum_{j=1}^{2^m}(W_{s_j^m} - W_{s_{j-1}^m})^2\right\}.$$

The term $\sum_{j=1}^{2^m}(W_{s_j^m} - W_{s_{j-1}^m})^2$ converges to $t_{k+1} - t_k$ in L^2, and $\sup_j |r(X_{s_j^m}, X_{s_{j-1}^m})|$ is a bounded random variable that converges to 0 almost surely. Hence the 3rd term of the right-hand side in (2.2.5) converges to 0 almost surely along with some subsequence.

Consequently, taking an a.s. convergent subsequence, we obtain

$$\Delta f_k = \int_{t_k}^{t_{k+1}} f'(X_s)K_s ds + \int_{t_k}^{t_{k+1}} f'(X_s)H_s dW_s + \frac{1}{2}\int_{t_k}^{t_{k+1}} f''(X_s)H_s^2 ds,$$

from which the Itô formula follows by summing up the both side in the equality just above from $k = 0$ to n.

In general cases where $\{K_t\}$ and $\{H_t\}$ are not necessarily simple, choose approximating simple processes $\{K_t^{(n)}\}$ and $\{H_t^{(n)}\}$ such that

$$\int_0^T |K_s - K_s^{(n)}|ds \to 0, \text{ a.s.}, \quad \mathbb{E}\int_0^T |H_s - H_s^{(n)}|^2 ds \to 0,$$

apply the derived Itô formula for simple process, and take limits. We are done. □

Example 2.16. Let $m = 1$. Recall that in Example 2.5 we compute $\int_0^T W_t dW_t$ directly from the definition of the Itô integrals. Here we shall compute it using Itô formula. Applying Corollary 2.15 with $f(x) = x^2/2$, we have

$$df(W_t) = f'(W_t)dW_t + \frac{1}{2}f''(W_t)dt = W_t dW_t + \frac{1}{2}dt,$$

whence

$$\int_0^T W_t dW_t = \frac{1}{2}W_T^2 - \frac{T}{2}.$$

We obtain the following the *product formula* by Theorem 2.14 with $f(x, y) = xy$.

Proposition 2.17: Product formula

For one dimensional Itô processes $\{X_t\}$ and $\{Y_t\}$, we have

$$d(X_tY_t) = X_tdY_t + Y_tdX_t + dX_tdY_t.$$

Example 2.18. Let us compute $\int_0^t sdW_s$. Use the product formula with $X_t = t$ and $Y_t = W_t$ and $dtdW_t = 0$ to see

$$d(tW_t) = tdW_t + W_tdt.$$

Thus,

$$\int_0^t sdW_s = tW_t - \int_0^t W_sds.$$

Example 2.19. Let $\{W_t\}$ is a scalar Brownian motion. Suppose that an $\mathbb{R}$-valued process $\{X_t\}$ satisfies the *stochastic differential equation*

$$dX_t = bX_tdt + \sigma dW_t, \tag{2.2.6}$$

where $b \in \mathbb{R}$ and $\sigma > 0$.

Applying the product formula for e^{-bt} and X_t, we observe

$$d(e^{-bt}X_t) = -be^{-bt}X_tdt + e^{-bt}(bX_tdt + \sigma dW_t) = \sigma e^{-bt}dW_t.$$

Hence, the solution of (2.2.6) is given by

$$X_t = e^{bt}X_0 + \sigma \int_0^t e^{b(t-s)}dW_s,$$

which is called an *Ornstein-Uhlenbeck process*.

2.3 Girsanov–Maruyama Theorem

In this section, we will see that a Brownian motion with drift $bt+W_t$ turns out to be a Brownian motion under a probability measure different from $\mathbb{P}$.

We start with two examples of changing drifts.

Example 2.20. Let X be a standard Gaussian random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, i.e.,

$$\mu_X(A) = \mathbb{P}(X \in A) = \int_A \frac{e^{-x^2/2}}{\sqrt{2\pi}}dx, \quad A \in \mathcal{B}(\mathbb{R}).$$

Then, for any $a \in \mathbb{R}$, the random variable $Y := X + a$ of course follows a normal distribution with mean a and variance 1 under $\mathbb{P}$. Namely,

$$\mu_Y(A) = \mathbb{P}(Y \in A) = \int_A \frac{e^{-(x-a)^2/2}}{\sqrt{2\pi}}dx, \quad A \in \mathcal{B}(\mathbb{R}).$$

Since the probability measures μ_X and μ_Y are equivalent and

$$\mu_X(A) = \int_A \frac{e^{-x^2/2+(x-a)^2/2-(x-a)^2/2}}{\sqrt{2\pi}}dx = \int_A e^{-x^2/2+(x-a)^2/2}d\mu_Y(x),$$

we have

$$\frac{d\mu_X}{d\mu_Y}(x) = e^{-x^2/2+(x-a)^2/2} = e^{-ax+a^2/2}.$$

Therefore, the probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ defined by

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{-aY+a^2/2} = e^{-aX-a^2/2}$$

satisfies, for $A \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}\mathbb{Q}(Y \in A) = \mathbb{E}[1_{\{Y\in A\}}e^{-aY+a^2/2}] &= \int_A e^{-ax+a^2/2} d\mu_Y(x) \\ &= \int_A \frac{d\mu_X}{d\mu_Y}(x) d\mu_Y(x) = \mu_X(A) = \int_A \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.\end{aligned}$$

Thus, $Y \sim N(0,1)$ under $\mathbb{Q}$.

Example 2.21. Consider the symmetric random walk $S_n = \sum_{i=0}^n X_i$ starting from 0. Then $\{S_n\}$ is a martingale with respect to the filtration $\{\mathcal{G}_n\}$ given by $\mathcal{G}_n = \sigma(X_i : i \le n)$. Let $\{\theta_n\}$ be a process such that θ_n is $\mathcal{G}_{n-1}$-measurable and satisfies $|\theta_n| < 1$ for each n. Then

$$L_n := \prod_{i=1}^n (1 + \theta_i X_i), \quad L_0 := 1$$

is a positive martingale.

Define the probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{G}_N)$ by $d\mathbb{Q}/d\mathbb{P} = L_N$, and consider the process

$$\tilde{S}_n = S_n - \sum_{i=1}^n \theta_i, \quad \tilde{S}_0 = 0.$$

Then the *Bayes formula*

$$\mathbb{E}_{\mathbb{Q}}[\tilde{S}_{n+1}|\mathcal{G}_n] = L_n^{-1}\mathbb{E}[L_{n+1}\tilde{S}_{n+1}|\mathcal{G}_n]$$

and $\mathbb{E}[(1+\theta_{n+1}X_{n+1})(X_{n+1} - \theta_{n+1})|\mathcal{G}_n] = 0$ lead to $\mathbb{E}_{\mathbb{Q}}[\tilde{S}_{n+1}|\mathcal{G}_n] = \tilde{S}_n$, whence $\{\tilde{S}_n\}_{n=0}^N$ is a $\mathbb{Q}$-martingale.

Now we consider the change of drifts of Brownian motions. To this end, we show some preliminary results.

Lemma 2.22

Let $\{M_t\}_{0\le t\le T}$ be a nonnegative local martingale. Then $\{M_t\}$ is a supermartingale. Moreover if $\mathbb{E}[M_T] = \mathbb{E}[M_0]$ then $\{M_t\}$ is a martingale.

Proof. Let $\{\tau_n\}_{n=1}^\infty$ be a sequence of stopping times such that $\tau_n \nearrow \infty$ and M_t^τ is a martingale. By Fatou's lemma, we have

$$\mathbb{E}[M_t] = \mathbb{E}[\lim_{n\to\infty} M_{t\wedge\tau_n}] \le \liminf_{n\to\infty} \mathbb{E}[M_{t\wedge\tau_n}] = \mathbb{E}[M_0] < \infty,$$

whence $M_t \in L^1$ for any t. Then Fatou's lemma for the conditional expectations yields, for $s \le t$,

$$\mathbb{E}[M_t|\mathcal{F}_s] \le \liminf_{n\to\infty} \mathbb{E}[M_{t\wedge\tau}|\mathcal{F}_s] = \liminf_{n\to\infty} M_{s\wedge\tau_n} = M_s,$$

from which $\{M_t\}$ is a supermartingale. In particular, $\mathbb{E}[M_T] \le \mathbb{E}[M_t] \le \mathbb{E}[M_s] \le \mathbb{E}[M_0]$ for $s \le t$. Thus, if $\mathbb{E}[M_T] = \mathbb{E}[M_0]$, then $Z := M_s - \mathbb{E}[M_t|\mathcal{F}_s]$ satisfies $Z \ge 0$ a.s. and $\mathbb{E}[Z] = 0$. This means $Z = 0$ a.s. □

Now, Let $\{W_t\}_{0\le t\le T}$ be a d-dimensional $\mathbb{F}$-Brownian motion, and $\theta_t = (\theta_t^1, \dots, \theta_t^d)$, $0 \le t \le T$, a d-dimensional process such that $\{\theta_t^i\} \in \mathcal{L}_{2,loc}$, $i = 1, \dots, d$. Then consider the process

$$Z_t := \exp\left(-\int_0^t \theta_s^\top dW_s - \frac{1}{2}\int_0^t \theta_s^2 ds \right), \quad 0 \le t \le T, \tag{2.3.1}$$

which is a local martingale (take $\tau_n = \inf\{t \ge 0 : \int_0^t Z_s |\theta_s|^2 ds \ge n\}$ as a localizing sequence). By the previous lemma, $\{Z_t\}$ is a nonnegative supermartingale. Moreover, under the condition $\mathbb{E}[Z_T] = 1$, it is a martingale, and we can define the probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$ by $d\mathbb{Q}/d\mathbb{P} = Z_T$.

Theorem 2.23: Girsanov–Maruyama Theorem

Let $\{Z_t\}_{0\le t\le T}$ be given by (2.3.1). Then the process

$$X_t := W_t + \int_0^t \theta_s ds, \quad 0 \le t \le T,$$

is a d-dimensional $\mathbb{F}$-Brownian motion under $\mathbb{Q}$.

Proof. We will prove the theorem under the boundedness of $\{\theta_t\}$. We refer to [21, Chapter 3] for a proof for general cases.

It is clear that $X_0 = 0$. Thus it suffices to show that for every $s \le t$ and bounded $\mathcal{F}_s$-measurable random variable Y the increments $X_t - X_s$ is independent of Y and follows $N(0, (t-s)I_d)$. To this end, let $\alpha \in \mathbb{R}^d$ and $\beta \in \mathbb{R}$. Then,

$$\begin{aligned}
\mathbb{E}_\mathbb{Q}[e^{i\alpha^\top(X_t - X_s) + i\beta Y}] &= \mathbb{E}_\mathbb{P}[Z_T e^{i\alpha^\top(X_t - X_s) + i\beta Y}] = \mathbb{E}_\mathbb{P}[Z_t e^{i\alpha^\top(X_t - X_s) + i\beta Y}] \\
&= \mathbb{E}_\mathbb{P}\left[Z_s e^{\int_s^t (i\alpha - \theta_s)^\top dW_u + \int_s^t (i\alpha - \theta_u/2)^\top \theta_u du + i\beta Y} \right] \\
&= e^{-|\alpha|^2(t-s)/2} \mathbb{E}_\mathbb{P}\left[Z_s e^{\int_s^t (i\alpha - \theta_u)^\top dW_u - \frac{1}{2}\int_s^t |i\alpha - \theta_u|^2 du + i\beta Y} \right],
\end{aligned}$$

where $i = \sqrt{-1}$ denotes the imaginary unit.

Now, by the Itô formula, the process

$$M_t := \exp\left[\int_0^t (i\alpha - \theta_u)^\top dW_u - \frac{1}{2}\int_0^t |i\alpha - \theta_u|^2 du \right], \quad 0 \le t \le T,$$

satisfies

$$M_t = 1 + \int_0^t M_u (i\alpha - \theta_u)^\top dW_u,$$

and so is a local martingale under $\mathbb{P}$. This and the boundedness of $\{\theta_t\}$ mean that it is indeed a martingale under $\mathbb{P}$. Thus, $\mathbb{E}_\mathbb{P}[e^{\int_s^t (i\alpha - \theta_u)^\top dW_u - \frac{1}{2}\int_s^t |i\alpha - \theta_u|^2 du} | \mathcal{F}_s] = 1$. Consequently,

$$\mathbb{E}_\mathbb{Q}[e^{i\alpha^\top(X_t - X_s) + i\beta Z}] = e^{-|\alpha|^2(t-s)/2} \mathbb{E}_\mathbb{P}\left[Z_s e^{i\beta Z} \right] = e^{-|\alpha|^2(t-s)/2} \mathbb{E}_\mathbb{Q}[e^{i\beta Z}],$$

from which the theorem follows. $\square$

We give a sufficient condition for which $\{Z_t\}$ in (2.3.1) satisfies $\mathbb{E}[Z_T] = 1$, without a proof, which is known as the *Novikov's condition*.

Theorem 2.24: Novikov

Let $\theta_t = (\theta_t^1, \ldots, \theta_t^d)$, $0 \le t \le T$, be a d-dimensional process such that each component belongs to $\mathcal{L}_{2,\mathrm{loc}}$. Suppose that

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T |\theta_t|^2 dt\right)\right] < \infty.$$

Then $\{Z_t\}_{0\le t\le T}$ given by (2.3.1) is a martingale.

2.4 Martingale Representation Theorem

As seen in Section 2.1, for $\{X_t\} \in \mathcal{L}_2$ the process $\{I(X)_t\}$ of Itô integrals is L^2-martingale. In this section, conversely, we will show that any L^2-martingale is represented as a process of Itô integrals. In doing so, we will see that any random variable in L^2 is represented as an Itô integral.

Let $\{W_t\}_{0\le t\le T}$ be a a d-dimensional Brownian motion. Recall that for any C^1-function f the fundamental theorem of calculus tells us that $f(t) = f(0) + \int_0^t f'(s)ds$. In stochastic analysis, however, Itô formula tells us that the analogous result $f(W_t^1) = f(0) + \int_0^t f'(W_s^1)dW_s^1$ does not hold in general.

Throughout this section, we assume that $\mathbb{F} = \{\mathcal{F}_t\}_{0\le t\le T}$ is given by the augmented natural filtration generated by $\{W_t\}$, i.e., assume that

$$\mathcal{F}_t = \sigma(\mathcal{F}_t^W \cup \mathcal{N}), \quad 0 \le t \le T.$$

The following is the martingale representation theorem:

Theorem 2.25: Martingale representation theorem

Let $\{M_t\}_{0\le t\le T}$ be an $\mathbb{F}$-martingale with $M_T \in L^2$. Then there exists a unique $\mathbb{R}^d$-valued process $\{\phi_t\}_{0\le t\le T}$ with each component belonging to $\mathcal{L}_2$ such that

$$M_t = M_0 + \int_0^t \phi_s^\mathsf{T} dW_s, \ \text{ a.s.}, \ \ 0 \le t \le T.$$

- The uniqueness here means that two processes coincides with each other up to null sets with respect to the measure $dt \times \mathbb{P}$. Namely, if

$$M_t = M_0 + \int_0^t \phi_s^\mathsf{T} dW_s = M_0 + \int_0^t \psi_s^\mathsf{T} dW_s, \ \text{ a.s.}, \ \ 0 \le t \le T,$$

  for $\{\phi_t^i\}, \{\psi_t^i\} \in \mathcal{L}_2$, $i = 1, \ldots, d$, then $\phi_t^i(\omega) = \psi_t^i(\omega)$ holds for almost all $(t, \omega) \in [0, T] \times \Omega$ for any i.

Theorem 2.25 is a corollary of the following result:

Theorem 2.26: Itô representation theorem

Let X be an $\mathcal{F}_T$-measurable random variable in L^2. Then, there exists a unique $\mathbb{R}^d$-valued process $\{\phi_t\}$ with each component belonging to $\mathcal{L}_2$ such that

$$X = \mathbb{E}[X] + \int_0^T \phi_t^\mathsf{T} dW_t, \ \text{ a.s.} \tag{2.4.1}$$

Here, the uniqueness is understood as in above.

Proof. The uniqueness follows from the Itô isometry. We prove the existence. First we prove that it suffices to show the representation (2.4.1) holds for $X = f(W_{t_1}, \dots, W_{t_n})$ with bounded Borel functions f on $(\mathbb{R}^d)^n$ and $0 \le t_1 < \cdots < t_n \le T$. To this end, consider

$$\mathcal{X} = \left\{ X \in L^2(\mathcal{F}_T) : \text{the representation (2.4.1) holds for some } \{\phi_t\} \in \mathcal{L}_2 \right\}.$$

Notice that $\mathcal{X}$ is a closed subspace in $L^2(\mathcal{F}_T)$. Suppose that $\mathcal{X}$ contains all random variables of the form $X = 1_A(W_{t_1}, \dots, W_{t_n})$ where $A \in \mathcal{B}((\mathbb{R}^d)^n)$ and $0 \le t_1 < \cdots < t_n \le T$. Then, for $Y \in \mathcal{X}^\perp$, A and t_i's as above,

$$\mathbb{E}[Y 1_A(W_{t_1}, \dots, W_{t_n})] = 0$$

or

$$\mathbb{E}[Y^+ 1_A(W_{t_1}, \dots, W_{t_n})] = \mathbb{E}[Y^- 1_A(W_{t_1}, \dots, W_{t_n})].$$

This means that two probability measures defined by Y^+ and Y^- as their Radon-Nikodym derivatives coincide with each other on the π-system $\mathcal{C} := \{(W_{t_1}, \dots, W_{t_n}) \in A : 0 = t_0 \le t_1 < \cdots < t_n = T, \quad A \in (\mathbb{R}^d)^n, \quad n \ge 1\}$. This together with $\sigma(\mathcal{C}) = \mathcal{F}_T$ and Lemma A.44 yields $Y^+ = Y^-$ a.s., whence $\mathcal{X}^\perp = \{0\}$.

Next we show that the martingale representation holds for $X = f(W_{t_1}, \dots, W_{t_n})$ with f and t_i's as above. Define the function $v_k : [t_{k-1}, t_k] \times (\mathbb{R}^d)^k \to \mathbb{R}$, $k = 1, \dots, n$, inductively by

$$v_n(t, x_1, \dots, x_n) = \mathbb{E}[f(x_1, \dots, x_{n-1}, x_{n-1} + W_{t_n - t})], \quad t_{n-1} \le t \le t_n,$$

and for $k = n-1, n-2, \dots, 1$,

$$v_k(t, x_1, \dots, x_k) = \mathbb{E}[v_{k+1}(t_k, x_1, \dots, x_k, x_k + W_{t_k - t})], \quad t_{k-1} \le t \le t_k.$$

Then by Chapter 2, the function $(t_{k-1}, t_k) \times \mathbb{R}^d \ni (t, x_k) \mapsto v_k(t, x_1, \dots, x_k)$ is C^∞ and satisfies

$$\partial_t v_k + \frac{1}{2} \Delta_{x_k} v_k = 0,$$

where Δ_{x_k} is the Laplacian with respect to the variable x_k. Thus Itô formula yields

$$\begin{aligned} v_k(t, W_{t_1}, \dots, W_{t_k}) = {} & v_k(t_{k-1}, W_{t_1}, \dots, W_{t_{k-1}}, W_{t_{k-1}}) \\ & + \int_{t_{k-1}}^t D_{x_k} v_k(t_{k-1}, W_{t_{k-1}}, \dots, W_{t_{k-1}}, W_s)^\mathsf{T} dW_s, \quad t_{k-1} < t < t_k, \end{aligned}$$

where D_{x_k} is the gradient with respect to the variable x_k, from which we obtain

$$v_k(t_k, W_{t_1}, \dots, W_{t_k}) = v_{k-1}(t_{k-1}, W_{t_1}, \dots, W_{t_{k-1}}) + \int_{t_{k-1}}^{t_k} \phi_s^\mathsf{T} dW_s$$

with $\phi_s = D_{x_k} v_k(t_{k-1}, W_{t_{k-1}}, \dots, W_{t_{k-1}}, W_s)$, $s \in [t_{k-1}, t_k]$. Notice that $\phi \in \mathcal{L}_2$ since f is bounded. Consequently,

$$\begin{aligned} f(W_{t_1}, \dots, W_{t_n}) &= v_n(t_n, W_{t_1}, \dots, W_{t_n}) \\ &= v_{n-1}(t_{n-1}, W_{t_1}, \dots, W_{t_{n-1}}) + \int_{t_{n-1}}^{t_n} \phi_s^\mathsf{T} dW_s \\ &= v_{n-2}(t_{n-2}, W_{t_1}, \dots, W_{t_{n-2}}) + \int_{t_{n-2}}^{t_n} \phi_s^\mathsf{T} dW_s. \end{aligned}$$

Repeating this argument, we deduce

$$f(W_{t_1}, \dots, W_{t_n}) = v_1(0, 0) + \int_0^{t_n} \phi_s^\mathsf{T} dW_s,$$

as required. $\square$

We state a more general martingale representation theorem. For a proof we refer to the references on stochastic analysis.

Theorem 2.27

For every local $\mathbb{F}$-martingale $\{M_t\}$, there exists a unique $\mathbb{R}^d$-valued process $\{\phi_t\}$ with each component belonging to $\mathcal{L}_{2,loc}$ such that

$$M_t = M_0 + \int_0^t \phi_s^\top dW_s, \quad \text{a.s.}, \quad 0 \le t \le T.$$

2.5 Stochastic Integrals for Continuous Local Martingales

This section is devoted to a brief introduction to stochastic integration theory for general continuous local martingales. As an application, we will prove *Lévy's theorem*, providing a sufficient condition for which a given continuous local martingale is a Brownian motion. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a (complete) probability space equipped with a filtration $\mathbb{F} = \{\mathcal{F}_t\}_{0\le t\le T}$ satisfying the usual conditions. Recall from §2.1 that $\mathcal{M}_{loc}$ is the set of all continuous $\mathbb{F}$-local martingale starting from zero.

We shall start with the following theorem:

Theorem 2.28

Let $M = \{M_t\}_{0\le t\le T} \in \mathcal{M}_{loc}$. Then there exists a unique continuous, adapted, and monotonically nondecreasing process $\langle M\rangle = \{\langle M\rangle_t\}_{0\le t\le T}$ such that $\langle M\rangle_0 = 0$, $M^2 - \langle M\rangle \in \mathcal{M}_{loc}$, and

$$\sup_{0\le t\le T} \left| \langle M\rangle_t - \sum_{i=0}^{n-1} (M_{t_{i+1}\wedge t} - M_{t_i\wedge t})^2 \right| \to 0, \tag{2.5.1}$$

in probability as $\max_i(t_{i+1} - t_i) \to 0$, where $0 = t_0 < t_1 < \cdots < t_n = T$.

Proof. We will prove the existence of $\langle M\rangle$ as in the statement in the case where M is given by $M_t = \int_0^t \phi_s dW_s$ for some one dimensional $\mathbb{F}$-Brownian motion W and bounded adapted process $\{\phi_t\}_{0\le t\le T}$. For a proof of the general existence we refer to e.g. [49] and [21, Chapter 1].

Define $\langle M\rangle$ by

$$\langle M\rangle_t = \int_0^t \phi_s^2 ds, \quad 0 \le t \le T.$$

Then $\langle M\rangle$ is continuous, adapted, and monotonically nondecreasing with $\langle M\rangle_0 = 0$. By Itô formula, we have

$$dM_t^2 = 2M_t\phi_t dW_t + \phi_t^2 dt,$$

whence $M^2 - \langle M\rangle \in \mathcal{M}_{loc}$. Further, for any $\{t_i\}_{i=0}^n$ with $0 = t_0 < t_1 < \cdots < t_n = T$,

$$\begin{aligned}\sum_{i=0}^{n-1}(M_{t_{i+1}\wedge t} - M_{t_i\wedge t})^2 &= \sum_{i=0}^{n-1}\left[2\int_{t_i\wedge t}^{t_{i+1}\wedge t}(M_s - M_{t_i\wedge t})\phi_s dW_s + \int_{t_i\wedge t}^{t_{i+1}\wedge t}\phi_s^2 ds\right] \\ &= 2\int_0^t K_s^{(n)}\phi_s dW_s + \langle M\rangle_t,\end{aligned}$$

where

$$K_s^{(n)} = \sum_{i=0}^{n-1}(M_s - M_{t_i\wedge t})1_{(t_i\wedge t, t_{i+1}\wedge t]}(s).$$

The continuity of M yields that $K_s^{(n)} \to 0$, a.s., as $\Delta := \max_i(t_{i+1} - t_i) \to 0$, for any s. Further, $|K_s^{(n)}| \le 2\max_{0\le t\le T}|M_t|$ and by Doob's maximal inequality (Theorem 1.34)

$$\mathbb{E}\max_{0\le t\le T}|M_t|^2 \le 4\mathbb{E}M_T^2 = 4\mathbb{E}\int_0^t |\phi_t|^2 dt.$$

Thus the dominated convergence theorem and the boundedness of ϕ lead to

$$\mathbb{E}\int_0^T |K_s^{(n)}\phi_s|^2 ds \le C\mathbb{E}\int_0^T |K_s^{(n)}|^2 ds \to 0, \quad \text{as } \Delta \to 0.$$

Again by Doob's maximal inequality, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{0\le t\le T}\left|\int_0^t K_s^{(n)}\phi_s dW_s\right| > \varepsilon\right) \le \frac{1}{\varepsilon^2}\mathbb{E}\left|\int_0^T K_s^{(n)}\phi_s dW_s\right|^2 = \frac{1}{\varepsilon^2}\mathbb{E}\int_0^T \left|K_s^{(n)}\phi_s\right|^2 ds \to 0,$$

as $\Delta \to 0$, whence $\langle M\rangle$ satisfies (2.5.1).

Next we will prove the uniqueness. Let $A = \{A_t\}_{0\le t\le T}$ satisfy (2.5.1). Then, for any $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}\left(\sup_{0\le t\le T}|\langle M\rangle_t - A_t| > \varepsilon\right) &\le \mathbb{P}\left(\sup_{0\le t\le T}\left|\langle M\rangle_t - \sum_{i=0}^{n-1}(M_{t_{i+1}\wedge t} - M_{t_i\wedge t})^2\right| > \frac{\varepsilon}{2}\right)\\ &\quad + \mathbb{P}\left(\sup_{0\le t\le T}\left|A_t - \sum_{i=0}^{n-1}(M_{t_{i+1}\wedge t} - M_{t_i\wedge t})^2\right| > \frac{\varepsilon}{2}\right)\\ &\to 0,\end{aligned}$$

as $\Delta \to 0$. This means $\sup_{0\le t\le T}|\langle M\rangle_t - A_t| = 0$, a.s. ☐

- $\langle M\rangle$ is called the *quadratic variation* of M. Recall that in the case where M is a Brownian motion, the fact that $\langle M\rangle_t = t$ is derived in §1.3.

Theorem 2.29

Let $M, N \in \mathcal{M}_{loc}$. Then there exists a unique continuous adapted process $\langle M, N\rangle = \{\langle M, N\rangle_t\}_{0\le t\le T}$ with finite total variation such that $\langle M, N\rangle_0 = 0$, $MN - \langle M, N\rangle \in \mathcal{M}_{loc}$, and

$$\sup_{0\le t\le T}\left|\langle M, N\rangle_t - \sum_{i=0}^{n-1}(M_{t_{i+1}\wedge t} - M_{t_i\wedge t})(N_{t_{i+1}\wedge t} - N_{t_i\wedge t})\right| \to 0,$$

in probability as $\max_i(t_{i+1} - t_i) \to 0$, where $0 = t_0 < t_1 < \cdots < t_n = T$.

Proof. The process

$$\langle M, N\rangle := \frac{1}{4}\left(\langle M + N\rangle - \langle M - N\rangle\right)$$

satisfies the required properties. Indeed,

$$MN - \langle M, N\rangle = \frac{1}{4}\left((M+N)^2 - \langle M+N\rangle\right) - \frac{1}{4}\left((M-N)^2 - \langle M-N\rangle\right) \in \mathcal{M}_{loc}.$$

Proofs of the other statements are left to the reader. ☐

- The process $\langle M, N\rangle$ is called the *quadratic covariation* or *quadratic cross variation* of M and N.

- Let M and N be given by $M_t = \int_0^t f_s dW_s$ and $N_t = \int_0^t g_s dW_s$ for some $f, g \in \mathcal{L}_{2,loc}$, where $\{W_t\}$ is a one-dimensional $\mathbb{F}$-Brownian motion. Then, the product Itô formula yields

$$\langle M, N\rangle_t = \int_0^t f_s g_s ds.$$

Thus we can write

$$dM_tN_t = M_t dN_t + N_t dM_t + d\langle M, N\rangle_t.$$

Now introduce the class

$$\mathcal{M}_2^0 := \left\{ M \in \mathcal{M}_2 : \langle M\rangle_t = \int_0^t a_s ds \text{ for some nonnegative and adapted process } \{a_t\} \right\}.$$

Let $M \in \mathcal{M}_2^0$. We shall first define the Itô integral with respect to M. To this end, define the class $\mathcal{L}_2^M$ of the integrands by

$$\mathcal{L}_2^M = \left\{ \phi = \{\phi_t\}_{0\le t\le T} : \text{adapted, } \mathbb{E}\int_0^T |\phi_t|^2 dt\langle M\rangle_t < \infty \right\}.$$

As in the case of Brownian motions, for any simple process ϕ of the form

$$\phi_t = \psi_0 1_{\{0\}}(t) + \sum_{i=0}^n \varphi_i 1_{(t_i, t_{i+1}]}(t),$$

the Itô integral $\int_0^T \phi_t dM_t$ is defined by

$$\int_0^T \phi_t dM_t = \sum_{i=0}^n \varphi_i (M_{t_{i+1}} - M_{t_i}).$$

The following result is an analog to Proposition 2.2:

Proposition 2.30

Let $M \in \mathcal{M}_2^0$. For every simple process ϕ, we have

$$\mathbb{E}\left|\int_0^T \phi_t dM_t\right|^2 = \mathbb{E}\int_0^T |\phi_t|^2 d\langle M\rangle_t.$$

A proof of this proposition is left to the reader.

The class of simple processes is also dense in $\mathcal{L}_2^M$ in the following sense:

Lemma 2.31

Let $M \in \mathcal{M}_2^0$. For any $\phi \in \mathcal{L}_2^M$, there exists a sequence $(\phi^{(n)})_{n=1}^\infty$ of simple processes such that

$$\lim_{n\to\infty} \mathbb{E}\int_0^T |\phi_t - \phi_t^{(n)}|^2 d\langle M\rangle_t = 0.$$

Proof.* Since $M \in \mathcal{M}_2^0$, there exists a nonnegative and adapted process $\{a_t\}$ such that $\langle M\rangle_t = \int_0^t a_s ds$ and $\mathbb{E}\int_0^T a_t dt < \infty$.

First we will prove the lemma in the case where $|\phi_t| \le K$ on $[0, T] \times \Omega$ for some constant $K > 0$. Since $\phi \in \mathcal{L}_2$, by Lemma 2.3, there exists a sequence $(\phi^{(n)})$ of simple processes such that

$$\lim_{n\to\infty} \mathbb{E}\int_0^T |\phi_t - \phi_t^{(n)}|^2 dt = 0.$$

This means that there exists a subsequence $(\phi^{(n_j)})$ such that

$$\lim_{j\to\infty} \phi_t^{(n_j)} = \phi_t, \quad dt \times d\mathbb{P}\text{-a.e.}$$

We can take $\phi_t^{(n_j)}$ so that $|\phi_t^{(n_j)}| \leq K$ as in the proof of Lemma 2.3. Thus, by the dominated convergence theorem,

$$\lim_{j\to\infty} \mathbb{E} \int_0^T |\phi_t - \phi_t^{(n_j)}|^2 a_t dt = 0.$$

To prove the lemma for general $\phi \in \mathcal{L}_2^M$, consider the truncated process $\phi_t^{(N)} := \phi_t 1_{\{|\phi_t| \leq N\}}$ and follow the arguments as in the proof of Lemma 2.3. □

Definition 2.32. Let $\{\phi_t\} \in \mathcal{L}_2^M$, and let $(\phi^{(n)})$ be a sequence of simple processes as in Lemma 2.31. Then we define the Itô integral $\int_0^T \phi_t dM_t$ of $\{\phi_t\}$ by

$$\int_0^T \phi_t dM_t = \lim_{n\to\infty} \int_0^T \phi_t^{(n)} dM_t \quad \text{in } L^2.$$

As in Theorem 2.7, there exists a continuous modification J_t of $\int_0^t \phi_s dM_s := \int_0^T \phi_s 1_{[0,t]}(s) dM_s$. Thus we shall call $\{J_t\}$ as the process of stochastic integral of ϕ_t and write $J_t = \int_0^t \phi_s dM_s$ by abuse of notation.

Definition 2.33. Let $\phi \in \mathcal{L}_2^M$. The process $\int_0^t \phi_s dM_s$, $0 \leq t \leq T$, is a unique element in $\mathcal{M}_2$ such that for any sequence $(\phi^{(n)})_{n=1}^\infty$ of simple processes satisfying

$$\lim_{n\to\infty} \mathbb{E} \int_0^T |\phi_s - \phi_s^{(n)}|^2 d\langle M\rangle_s = 0,$$

we have

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{0\leq t\leq T} \left|\int_0^t \phi_s dM_s - \int_0^t \phi_s^{(n)} dM_s\right|^2\right] = 0.$$

Next, we shall extend the definition of Itô integral to the case where M belongs to the class

$$\mathcal{M}_{loc}^0 := \left\{ M \in \mathcal{M}_{loc} : \langle M\rangle_t = \int_0^t a_s ds \text{ for some nonnegative and adapted process } \{a_s\} \right\}.$$

The procedure for doing this is completely parallel to the Brownian case. Let $M \in \mathcal{M}_{loc}^0$. Consider the space $\mathcal{L}_{2,loc}^M$ of the integrands defined by

$$\mathcal{L}_{2,loc}^M = \left\{ \phi = \{\phi_t\}_{0\leq t\leq T} : \text{adapted}, \ \int_0^T \phi_t^2 d\langle M\rangle_t < \infty, \ \text{a.s.} \right\}.$$

Then, for any $\phi \in \mathcal{L}_{2,loc}^M$, the process

$$\phi_t^{(n)} := \phi_t 1_{\{\tau_n \leq t\}}, \quad 0 \leq t \leq T,$$

is in $\mathcal{L}_2^M$, where

$$\tau_n = \inf\left\{ 0 \leq t \leq T : \int_0^t \phi_s^2 d\langle M\rangle_s \geq n \right\}.$$

The Itô integral $\int_0^t \phi_s dM_s$ is now defined as

$$\int_0^t \phi_s dM_s = \int_0^t \phi_s^{(n)} dM_s, \quad 0 \leq t \leq \tau_n \wedge T.$$

Problem 2.34. Let $M \in \mathcal{M}_{loc}^0$ and ϕ be a left-continuous and adapted process. Show that the Itô integral $\int_0^t \phi_s dM_s$ is well-defined and satisfies

$$\int_0^t \phi_s dM_s = \lim_{\Delta \to 0} \sum_{i=0}^{n-1} \phi_{t_i}(M_{t_{i+1}} - M_{t_i})$$

in probability, where $\{t_i\}$ is any sequence such that $0 = t_0 < t_1 < \cdots < t_n = t$ and $\Delta = \max_i(t_{i+1} - t_i)$.

- In the case where M is in $\mathcal{M}_{loc}$ but not in $\mathcal{M}_{loc}^0$, the Itô integral $\int_0^t \phi_s dM_s$ is well-defined and is in $\mathcal{M}_{loc}$ if ϕ is $\mathbb{F}$-progressively measurable with $\int_0^T \phi_t^2 d\langle M \rangle_t < \infty$, a.s. We refer to e.g. [21], [18], [49] for details.
- In particular, if ϕ is continuous and adapted, then by Proposition 1.19, it is progressively measurable and satisfies

$$\int_0^t \phi_s dM_s = \lim_{\Delta \to 0} \sum_{i=0}^{n-1} \phi_{t_i}(M_{t_{i+1}} - M_{t_i})$$

  in probability, where $\{t_i\}$ is any sequence such that $0 = t_0 < t_1 < \cdots < t_n = t$ and $\Delta = \max_i(t_{i+1} - t_i)$.

We say that a one-dimensional process $X = \{X_t\}_{0 \le t \le T}$ is a *semimartingale* if it is represented as

$$X_t = X_0 + \int_0^t b_s ds + M_t, \quad 0 \le t \le T, \tag{2.5.2}$$

where

- X_0 is $\mathcal{F}_0$-measurable.
- $M \in \mathcal{M}_{loc}$.
- $\{b_t\}$ is an adapted process with $\int_0^T |b_t| dt < \infty$, a.s.

Let X be a semimartingale represented as in (2.5.2). For every continuous and adapted process ϕ, the *Itô integral* $\int_0^t \phi_s dX_s$ with respect to X is defined by

$$\int_0^t \phi_s dX_s = \int_0^t \phi_s b_s ds + \int_0^t \phi_s dM_s, \quad 0 \le t \le T.$$

Definition 2.35. Let X and Y be semimartingales. The *quadratic variation* $\langle X \rangle = \{\langle X \rangle_t\}_{0 \le t \le T}$ of X is defined by

$$\langle X \rangle_t = X_t^2 - 2 \int_0^t X_s dX_s, \quad 0 \le t \le T.$$

The *quadratic covariation* $\langle X, Y \rangle = \{\langle X, Y \rangle_t\}_{0 \le t \le T}$ of X and Y is defined by

$$\langle X, Y \rangle_t = X_t Y_t - \int_0^t X_s dY_s - \int_0^t Y_s dX_s, \quad 0 \le t \le T.$$

Theorem 2.36

Let X and Y be semimartingales. Then,

$$\langle X, Y \rangle_t = X_0 Y_0 + \lim_{\Delta \to 0} \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i}), \quad \text{in prob.},$$

where the limit is taken with respect to any sequence $\{t_i\}_{i=0}^n$ with $0 = t_0 < t_1 < \cdots < t_n = t$ and $\Delta = \max_i(t_{i+1} - t_i)$.

Proof. Using $X_tY_t = X_0Y_0 + \sum_{i=0}^{n-1}(X_{t_{i+1}}Y_{t_{i+1}} - X_{t_i}Y_{t_i})$ we observe

$$\begin{aligned}
&X_tY_t - \sum_{i=0}^{n-1} X_{t_i}(Y_{t_{i+1}} - Y_{t_i}) - \sum_{i=0}^{n-1} Y_{t_i}(X_{t_{i+1}} - X_{t_i}) \\
&= X_0Y_0 + \sum_{i=0}^{n-1} \big(X_{t_{i+1}}Y_{t_{i+1}} - X_{t_i}Y_{t_i} - X_{t_i}(Y_{t_{i+1}} - Y_{t_i}) - Y_{t_i}(X_{t_{i+1}} - X_{t_i})\big) \\
&= X_0Y_0 + \sum_{i=0}^{n-1}(X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i}).
\end{aligned}$$

From this and Problem 2.34 the theorem follows. □

By this theorem, we can confirm that the quadratic covariation $\langle M, N\rangle$ of $M, N \in M_{loc}$ coincides with the one in Definition 2.35.

Here is a generalized Itô formula for semimartingales.

Theorem 2.37

Let $\{X_t\}$ be a semimartingale of the form

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dM_s.$$

Let $f \in C^{1,2}([0,T] \times \mathbb{R}^d)$. Then

$$df(t, X_t) = \partial_t f(t, X_t)dt + \partial_x f(t, X_t)dX_t + \frac{1}{2}\partial_{xx}^2 f(t, X_t)\sigma_t^2 d\langle M\rangle_t.$$

- A generalized Itô rule is given by

$$(dM_t)^2 = d\langle M\rangle_t, \quad dtdM_t = 0.$$

- Applying Theorem 2.37 with $f(x) = x^2$, we find

$$\langle X\rangle_t = X_0^2 + \int_0^t \sigma_s^2 d\langle M\rangle_s.$$

  Thus formally

$$(dX_t)^2 = d\langle X\rangle_t = \sigma_t^2 d\langle M\rangle_t,$$

  and $df(t, X_t)$ is described by

$$df(t, X_t) = \partial_t f(t, X_t)dt + \partial_x f(t, X_t)dX_t + \frac{1}{2}\partial_{xx}^2 f(t, X_t)d\langle X\rangle_t.$$

The multidimensional Itô formula is as follows:

Theorem 2.38

Let $X = (X^1, \ldots, X^d)$ a d-dimensional process such that X^i is a semimartingale for each i. Let $f \in C^{1.2}$. Then,

$$df(t, X_t) = \partial_t f(t, X_t)dt + \sum_{i=1}^d \partial_{x_i} f(t, X_t)dX_t^i + \frac{1}{2}\sum_{i,j=1}^d \partial_{x_ix_j}^2 f(t, X_t)d\langle X^i, X^j\rangle_t.$$

- We call d-dimensional process with each component being semimartingale a *d-dimensional semimartingale*.
- When X is represented as $dX_t^i = b_t^i dt + \sum_{k=1}^m \sigma_t^{ik} dM_t^k$, then formally

$$dX_t^i dX_t^j = \sum_{k,\ell=1}^m \sigma_t^{ik}\sigma_t^{i\ell} dM_t^i dM_t^j = \sum_{k,\ell=1}^m \sigma_t^{ik}\sigma_t^{i\ell} d\langle M^i, M^j\rangle_t.$$

  Moreover, we can prove that

$$\langle X^i, X^j\rangle_t = X_0^i X_0^j + \sum_{k,\ell=1}^m \int_0^t \sigma_s^{ik}\sigma_s^{i\ell} d\langle M^i, M^j\rangle_s.$$

To emphasize the remarks above, we shall state the product Itô formula as a corollary of Theorem 2.37.

Corollary 2.39

Let $M \in \mathcal{M}_{loc}^0$. Let X and Y be semimartingales with representation

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dM_s,$$
$$Y_t = Y_0 + \int_0^t f_s ds + \int_0^t g_s dM_s.$$

Then,

$$\langle X, Y\rangle_t = X_0 Y_0 + \int_0^t \sigma_s g_s d\langle M\rangle_s.$$

Recall that by definition of the quadratic covariation, for semimartingales X and Y,

$$X_t Y_t = X_0 Y_0 + \int_0^t Y_s dX_s + \int_0^t X_s dY_s + \langle X, Y\rangle_t.$$

The existence of the *correction term* $\langle X, Y\rangle$ makes the Itô calculus different from the ordinary calculus. The Stratonovich integral provides a useful means of developing stochastic analysis as in ordinary calculus.

Definition 2.40. Let X and Y be one-dimensional semimartingales. The Stratonovich integral $\int_0^t Y_s \circ dX_s$ of Y with respect to X is defined as

$$\int_0^t Y_s \circ dX_s := \int_0^t Y_s dX_s + \frac{1}{2}\langle X, Y\rangle_t.$$

- By definition, for semimartingales X and Y,

$$X_t Y_t = X_0 Y_0 + \int_0^t Y_s \circ dX_s + \int_0^t X_s \circ dY_s.$$

- If X_t and Y_t are represented as

$$dX_t = b_t dt + \sigma_t dM_t,$$
$$dY_t = f_t dt + g_t dM_t$$

  for some $M \in \mathcal{M}_{loc}^0$, respectively, then

$$\int_0^t Y_s \circ dX_s = \int_0^t Y_s(b_s ds + \sigma_s dM_s) + \frac{1}{2}\int_0^t \sigma_s g_s \langle M\rangle_s.$$

The following chain rule holds for the Stratonovich integral under more smoothness condition than that in the Itô formula:

Proposition 2.41

Let $X = (X^1, \ldots, X^d)$ be a d-dimensional semimartingale, and $f \in C^3(\mathbb{R}^d)$. Then

$$df(X_t) = \sum_{i=1}^{d} \partial_{x_i} f(X_t) \circ dX_t^i.$$

Proof. For simplicity we shall assume that $d = 1$ and X_t is represented as $dX_t = b_t dt + \sigma dM_t$ with $\langle M \rangle_t = a_t dt$. Theorem 2.37 yields

$$\begin{aligned} df(X_t) &= f'(X_t)dX_t + \frac{1}{2}f''(X_t)d\langle X \rangle_t \\ &= f'(X_t)dX_t + \frac{1}{2}f''(X_t)\sigma_t^2 a_t dt, \end{aligned}$$

as well as

$$\begin{aligned} df'(X_t) &= f''(X_t)dX_t + \frac{1}{2}f'''(X_t)d\langle X \rangle_t \\ &= f''(X_t)\left(b_t + \frac{1}{2}f'''(X_t)\sigma_t^2 a_t\right) dt + f''(X_t)\sigma_t dM_t. \end{aligned}$$

Thus, by definition of the Stratonovich integral, we have

$$\begin{aligned} f'(X_t) \circ dX_t &= f'(X_t)dX_t + \frac{1}{2}df'(X_t)dX_t = f'(X_t)dX_t + \frac{1}{2}f''(X_t)\sigma_t^2 a_t dt \\ &= df(X_t), \end{aligned}$$

as required. □

Proposition 2.42

Let X and Y are one-dimensional semimartingales. Then,

$$\int_0^t Y_s \circ dX_s = \lim_{\Delta \to 0} \sum_{i=0}^{n-1} \left(\frac{1}{2}Y_{t_i} + \frac{1}{2}Y_{t_{i+1}}\right)(X_{t_{i+1}} - X_{t_i}), \quad \text{in prob.},$$

where the limit is taken with respect to any sequence $\{t_i\}_{i=0}^n$ with $0 = t_0 < t_1 < \cdots < t_n = t$ and $\Delta = \max_i(t_{i+1} - t_i)$.

Proof. From this and $(Y_{t_{i+1}} + Y_{t_i})/2 = (Y_{t_{i+1}} - Y_{t_i})/2 + Y_{t_i}$ we find

$$\begin{aligned} \sum_{i=0}^{n-1} \left(\frac{1}{2}Y_{t_{i+1}} + \frac{1}{2}Y_{t_i}\right) &= \sum_{i=0}^{n-1} Y_{t_i}(X_{t_{i+1}} - X_{t_i}) + \frac{1}{2}\sum_{i=1}^{n-1}(Y_{t_{i+1}} - Y_{t_i})(X_{t_{i+1}} - X_{t_i}) \\ &\to \int_0^t Y_s dX_s + \frac{1}{2}\langle X, Y \rangle_t \end{aligned}$$

in probability as $\Delta \to 0$. Thus the proposition follows. □

The notion of the *backward Itô integral* will be used in the analysis of the *reverse-time diffusions* (see Section 3.8).

Definition 2.43. Let $X = \{X_t\}_{0\le t\le T}$ and $Y = \{Y_t\}_{0\le t\le T}$ be continuous processes. If there exists a limit of

$$\sum_{i=0}^{n-1} Y_{t_{i+1}}(X_{t_{i+1}} - X_{t_i})$$

in probability as $\Delta \to 0$, where the limit is taken with respect to any sequence $\{t_i\}_{i=0}^n$ with $0 = t_0 < t_1 < \cdots < t_n = t$ and $\Delta = \max_i(t_{i+1} - t_i)$, then we write $\int_0^t Y_s \overleftarrow{d} X_t$ for this limit and call it the *backward Itô integral.*

Proposition 2.44

Let X and Y be one-dimensional semimartingales. Then,

$$\int_0^t Y_s \overleftarrow{d} X_t = \int_0^t Y_s dX_s + \langle Y, X\rangle_t.$$

Proof. Use the relation $Y_{t_{i+1}}(X_{t_{i+1}} - X_{t_i}) = Y_{t_i}(X_{t_{i+1}} - X_{t_i}) + (Y_{t_{i+1}} - Y_{t_i})(X_{t_{i+1}} - X_{t_i})$ and apply the results from Problem 2.34 and Theorem 2.36. □

We close this section by showing Lévy's theorem as announced in the beginning.

Theorem 2.45: Lévy's characterization of Brownian motions

Let $M = (M^1, \cdots, M^d)$ be such that $M^i \in \mathcal{M}_{loc}$ for each i, and $\langle M^i, M^j\rangle_t = \delta_{ij}t$, where δ_{ij} is the Kronecker's delta. Then, M is a d-dimensional $\mathbb{F}$-Brownian motion.

Proof. We will show this theorem in the case where $d = 1$ and $M = M^1 \in \mathcal{M}^0_{loc}$. For a proof of general cases we refer to, e.g., [49] and [21]. Let $\{\tau_n\}_{n=1}^\infty$ be a sequence of stopping times such that $\{M_{t\wedge\tau_n}\}_{0\le t\le T}$ is a martingale. Since $M^2 - \langle M\rangle \in \mathcal{M}_{loc}$, we find

$$\mathbb{E}[M^2_{t\wedge\tau_n}] = \mathbb{E}[\langle M\rangle_{t\wedge\tau_n}] = \mathbb{E}[t \wedge \tau_n] \le t.$$

Applying Fatou's lemma, we have

$$\mathbb{E}[M_t^2] = \mathbb{E}\left[\lim_{n\to\infty} M^2_{t\wedge\tau_n}\right] \le \liminf_{n\to\infty} \mathbb{E}[M^2_{t\wedge\tau_n}] \le t,$$

whence $M \in \mathcal{M}^0_2$.

By Itô formula and $\langle M\rangle_t = t$,

$$de^{i\xi M_t} = i\xi e^{i\xi M_t} dM_t - \frac{\xi^2}{2} e^{i\xi M_t} dt$$

for $\xi \in \mathbb{R}^d$, where $i = \sqrt{-1}$ is the imaginary unit. Since $|e^{i\xi M_t}| = 1$, the complex valued process $\int_0^t e^{i\xi M_s} dM_s$ is a martingale. Thus, for any $A \in \mathcal{F}_s$ and $t \ge s$,

$$\mathbb{E}\left[e^{i\xi(M_t - M_s)} 1_A\right] = \mathbb{P}(A) - \frac{\xi^2}{2}\int_s^t \mathbb{E}\left[e^{i\xi(M_r - M_s)} 1_A\right] dr.$$

Solving this ODE, we obtain

$$\mathbb{E}\left[e^{i\xi(M_t - M_s)} 1_A\right] = \mathbb{P}(A) e^{-\frac{\xi^2}{2}(t-s)} = \mathbb{E}\left[e^{-\frac{\xi^2}{2}(t-s)} 1_A\right].$$

Since $A \in \mathcal{F}_s$ is arbitrary, we deduce

$$\mathbb{E}\left[e^{i\xi(M_t - M_s)} \middle| \mathcal{F}_s\right] = e^{-\frac{\xi^2}{2}(t-s)}, \quad t > s,$$

from which the theorem follows. □

CHAPTER 3

Stochastic Differential Equations

3.1 Introduction

Ordinary differential equations with white noise

We are concerned with ordinary differential equations (ODEs) with random noises. For example, such ODEs can be of the form

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t)\xi_t, \tag{3.1.1}$$

where $\{\xi_t\}$ is a stochastic process providing random disturbance to the system process $\{X_t\}$. In science and engineering, a natural candidate for the disturbance processes is a Gaussian white noise, i.e., it is natural to assume that ξ_t is a Gaussian process with mean zero and covariance $\mathbb{E}[\xi_t\xi_s] = \delta(t-s)$, $t, s \in \mathbb{R}$, where $\delta(\cdot)$ is the delta function. Unfortunately, this natural formulation for nonlinear ODEs (3.1.1) comes up against an obstacle since the delta function is not a usual function but a *distribution* rigorously. Indeed, $\{\xi_t\}_{t\in\mathbb{R}}$ is not a stochastic process in the usual sense but a *random distribution* (see Itô [20]).

Changing the approach to (3.1.1), we use the fact that ξ_t is given by the time derivative, in the sense of the distribution, of a one-dimensional Brownian motion W_t (see again [20]). Then, replacing ξ_t with dW_t/dt in (3.1.1), we get

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t)\frac{dW_t}{dt},$$

whence, by a formal integration,

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s. \tag{3.1.2}$$

The integral equation (3.1.2) is equivalent to (3.1.1) formally, as well as can be defined rigorously since the term $\int_0^t \sigma(s, X_s)dW_s$ is understood as the Itô integral. Then, we write

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \tag{3.1.3}$$

for (3.1.2). This is a modern approach to *stochastic differential equations* (SDEs), which is originated by Itô [42] (see also Itô [19]) and has achieved remarkable successes.

In this chapter, we present some basic results on SDEs. We refer to [31], [49], [45], [21] for more detailed accounts. Before presenting examples of SDEs, we give a formal characterization of the coefficients b and σ in (3.1.3). By (3.1.2), we have

$$X_{t+\Delta t} = X_t + \int_t^{t+\Delta t} b(s, X_s)ds + \int_t^{t+\Delta t} \sigma(s, X_s)dW_s.$$

Under the assumption that $\{\sigma(t, X_t)\} \in \mathcal{L}_2$ (recall from Chapter 2), it follows that at least formally,

$$\begin{aligned} b(t,x) &= \lim_{\Delta t \searrow 0} \frac{1}{\Delta t}\mathbb{E}[X_{t+\Delta t} - X_t | X_t = x], \\ \sigma(t,x)^2 &= \lim_{\Delta t \searrow 0} \frac{1}{\Delta t}\mathbb{V}[X_{t+\Delta t} - X_t | X_t = x]. \end{aligned} \tag{3.1.4}$$

The functions b and σ are called the *drift* and *diffusion* coefficients, respectively.

Black–Scholes model for stock prices

Let us consider a stock with price S_t at time $t \geq 0$. Then the *return rate* $R_{t,t+\Delta t}$ of this stock between t and $t + \Delta t$ is given by $R_{t,t+\Delta t} = (S_{t+\Delta t} - S_t)/S_t$. Using the normalization $I_{t,t+\Delta t}$ of $R_{t,t+\Delta t}$, i.e., $I_{t,t+\Delta t} = (R_{t,t+\Delta t} - \mathbb{E}[R_{t,t+\Delta t}])/\sqrt{\mathbb{V}(R_{t,t+\Delta t})}$, we have

$$R_{t+\Delta t} = \mathbb{E}[R_{t+\Delta t}] + \sqrt{\mathbb{V}(R_{t,t+\Delta t})}I_{t,t+\Delta t}.$$

Now, assume that the expected return rate $b = \mathbb{E}[R_{t+\Delta t}]/\Delta t$ per time and the variance $\sigma^2 = \mathbb{V}(R_{t,t+\Delta t})/\Delta t$ of the return rate per time are constant with respect to t. Then,

$$\frac{S_{t+\Delta t} - S_t}{S_t} = b\Delta t + \sigma\sqrt{\Delta t}I_{t,t+\Delta t}.$$

Thus,

$$\begin{aligned} \lim_{\Delta t \searrow 0} \frac{1}{\Delta t}\mathbb{E}[S_{t+\Delta t} - S_t | S_t = s] &= bs, \\ \lim_{\Delta t \searrow 0} \frac{1}{\Delta t}\mathbb{V}[S_{t+\Delta t} - S_t | S_t = s] &= \sigma^2 s^2. \end{aligned}$$

So, assuming that $\{S_t\}$ is described by an SDE and then using (3.1.4), we obtain

$$dS_t = S_t(bdt + \sigma dW_t). \tag{3.1.5}$$

This SDE is called the *Black–Scholes model.* As remarked in the above, this equation should be interpreted as the following integral form:

$$S_t = S_0 + b\int_0^t S_r dr + \sigma\int_0^t S_r dW_r.$$

Now suppose temporarily that there exists a solution S_t to the equation (3.1.5). Then, applying Itô formula for $\log(S_t)$, formally we have

$$d(\log S_t) = \frac{dS_t}{S_t} - \frac{1}{2S_t^2} \cdot S_t^2\sigma^2 dt = bdt + \sigma dW_t - \frac{1}{2}\sigma^2 dt.$$

Thus the solution S_t of the Black–Scholes model is explicitly given by

$$S_t = S_0 \exp((b - \sigma^2/2)t + \sigma W_t).$$

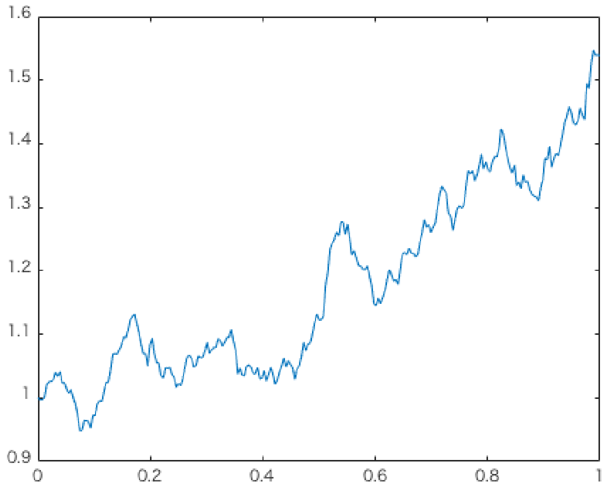

Figure 3.1.1: A sample path of Black–Scholes model in the case of $b = 0.5$ and $\sigma = 0.2$.

Predator-prey model

Consider a biological system consisting of two species where one is a predator and the other is a prey, whose populations at time t are denoted by X_t^1 and X_t^2, respectively. We assume that in a small time interval $[t, t + \Delta t]$, the probability of the predator being given a single birth without death and the population of the prey remaining unchanged is

$$\mathbb{P}(\Delta X_t^1 = 1, \Delta X_t^2 = 0 \,|\, X_t^1 = x_1, X_t^2 = x_2) = b_1 x_1 \Delta t + o(\Delta t).$$

Similarly, we assume

$$\begin{aligned}
\mathbb{P}(\Delta X_t^1 = 0, \Delta X_t^2 = 1 \,|\, X_t^1 = x_1, X_t^2 = x_2) &= b_2 x_2 \Delta t + o(\Delta t), \\
\mathbb{P}(\Delta X_t^1 = -1, \Delta X_t^2 = 0 \,|\, X_t^1 = x_1, X_t^2 = x_2) &= d_1 x_1 \Delta t + o(\Delta t), \\
\mathbb{P}(\Delta X_t^1 = 0, \Delta X_t^2 = -1 \,|\, X_t^1 = x_1, X_t^2 = x_2) &= d_2 x_2 \Delta t + o(\Delta t).
\end{aligned}$$

In view of the predator-prey relation, we further assume that b_2, d_1 are positive constants and that

$$b_1 = c_1 x_2, \quad d_2 = c_2 x_1,$$

with some positive constants c_1, c_2. Moreover, the probabilities of multiple births or deaths are assumed to be $o(\Delta t)$. Then, it is straightforward to see

$$\begin{aligned}
&\lim_{\Delta t \searrow 0} \frac{1}{\Delta t} \mathbb{E}[\Delta X_t^1 | X_t^1 = x_1, X_t^2 = x_2] = (c_1 x_2 - d_1) x_1, \\
&\lim_{\Delta t \searrow 0} \frac{1}{\Delta t} \mathbb{E}[\Delta X_t^1 | X_t^1 = x_1, X_t^2 = x_2] = (b_2 - c_2 x_1) x_2, \\
&\lim_{\Delta t \searrow 0} \frac{1}{\Delta t} \mathbb{V}[\Delta X_t^1 | X_t^1 = x_1, X_t^2 = x_2] = (c_1 x_2 + d_1) x_1, \\
&\lim_{\Delta t \searrow 0} \frac{1}{\Delta t} \mathbb{V}[\Delta X_t^2 | X_t^1 = x_1, X_t^2 = x_2] = (b_2 + c_2 x_1) x_2, \\
&\lim_{\Delta t \searrow 0} \frac{1}{\Delta t} \mathrm{Cov}[\Delta X_t^1, \Delta X_t^2 | X_t^1 = x_1, X_t^2 = x_2] = 0.
\end{aligned}$$

By a multidimensional analog of (3.1.4), we derive the SDE

$$
\begin{aligned}
dX_t^1 &= (c_1X_t^2 - d_1)X_t^1dt + \sqrt{(c_1X_t^2 + d_1)X_t^1}dW_t^1, \\
dX_t^2 &= (b_2 - c_2X_t^1)X_t^2dt + \sqrt{(b_2 + c_2X_t^1)X_t^2}dW_t^2
\end{aligned}
$$

for the predator-prey system, where (W_t^1, W_t^2) is a 2-dimensional Brownian motion.

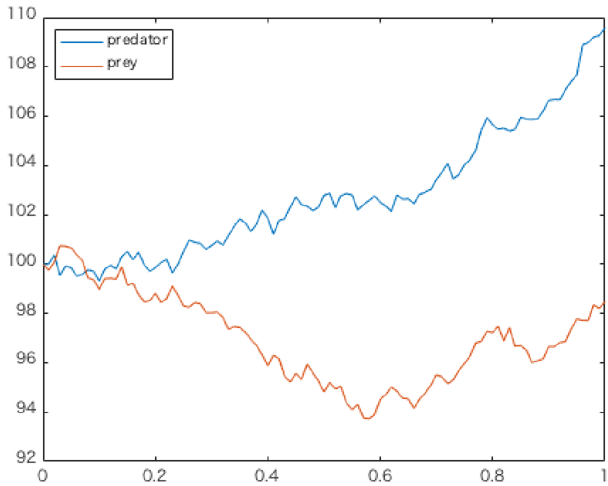

Figure 3.1.2: A sample path of the predator-prey model in the case of $d_1 = 0.01$, $b_2 = 0.05$, $c_1 = c_2 = 0.005$, and $X_0^1 = X_0^2 = 100$. Generated by the Euler-Maruyama method (see Section 3.4).

3.2 Existence and Uniqueness

In what follows, $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space equipped with filtration $\mathbb{F}$ satisfying the usual conditions, and $\{W_t\}$ is an m-dimensional $\mathbb{F}$-Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$. We fix a time horizon $T \in (0, \infty)$.

Definition 3.1. Let $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$ be Borel measurable, and let ξ be an $\mathcal{F}_0$-measurable random variable. We say that an $\mathbb{R}^d$-valued process $\{X_t\}_{0\le t\le T}$ is a solution of the *stochastic differential equation* (SDE)

$$
dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t
$$

with initial condition $X_0 = \xi$ if the following conditions are satisfied:

(i) $\{X_t\}$ is a.s. continuous and $\mathbb{F}$-adapted.

(ii) $\int_0^T |b(s, X_s)|ds + \int_0^T |\sigma(s, X_s)|^2 ds < \infty$, a.s.

(iii) $\{X_t\}$ is represented as

$$
X_t = \xi + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \quad \text{a.s.}, \quad 0 \le t \le T.
$$

The following is the fundamental existence and uniqueness result for SDEs:

Theorem 3.2

Suppose that the functions b, σ and the random variable ξ in Definition 3.1 satisfy

(i) Lipschitz continuity: there exists $K_0 > 0$ such that

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \le K_0|x-y|, \quad (t,x),(t,y) \in [0,T] \times \mathbb{R}^d,$$

(ii) Linearly growth condition: there exists $K_1 > 0$ such that

$$|b(t,x)| + |\sigma(t,x)| \le K_1(1+|x|), \quad (t,x) \in [0,T] \times \mathbb{R}^d.$$

(iii) $\xi \in L^2$.

Then, the stochastic differential equation

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dW_t \tag{3.2.1}$$

with initial condition $X_0 = \xi$ has a solution $\{X_t\}_{t\in[0,T]}$ satisfying $\mathbb{E}\left[\sup_{0\le t\le T}|X_t|^2\right] < \infty$. Moreover, the existence of the solution is unique in the sense of the indistinguishability, i.e., for any other solution $\{Y_t\}$ we have $X_t = Y_t$, $0 \le t \le T$, a.s.

We prove Theorem 3.2 with arguments similar to those in the existence proof for ordinary differential equations. Recall that Gronwall lemma plays an important role in that case.

Lemma 3.3: Gronwall lemma

Suppose that a nonnegative, bounded and Borel function $v : [0,T] \to \mathbb{R}$ satisfies

$$v(t) \le C + A\int_0^t v(s)ds, \quad 0 \le t \le T$$

for some positive constants C, A. Then,

$$v(t) \le Ce^{At}, \quad 0 \le t \le T.$$

Proof. By an iterative application of the condition on v, we obtain

$$\begin{aligned} v(t) &\le C + CAt + A^2\int_0^t\int_0^s v(r)dr \\ &\le C + CAt + \frac{CA^2t^2}{2} + \cdots + \frac{CA^nt^n}{n!} + A^{n+1}\int_0^t\int_0^{s_1}\cdots\int_0^{s_n} v(s_{n+1})ds_{n+1}ds_n\cdots ds_1 \end{aligned}$$

for $n \ge 1$. The last term is at most $\sup_{0\le t\le T} v(t)(At)^{n+1}/(n+1)!$ and goes to zero as $n \to \infty$. Thus the lemma follows. □

Proof of Theorem 3.2. First we show the uniqueness. Let $\{X_t\}$ and $\{Y_t\}$ be two solution, and put $a_t = b(t,X_t) - b(t,Y_t)$, $\gamma_t = \sigma(t,X_t) - \sigma(t,Y_t)$. Then, from $\mathbb{E}[\max_{0\le t\le T}|X_t - Y_t|^2] < \infty$ and the Lipschitz continuity, we have $\{\gamma_t\} \in \mathcal{L}_2$. This together with the inequality $|x+y|^2 \le 2(|x|^2+|y|^2)$

yields

$$\begin{aligned}\mathbb{E}|X_t - Y_t|^2 &= \mathbb{E}\left|\int_0^t a_s ds + \int_0^t \gamma_s dW_s\right|^2 \le 2\mathbb{E}\left|\int_0^t a_s ds\right|^2 + 2\mathbb{E}\left|\int_0^t \gamma_s dW_s\right|^2 \\ &\le 2t\mathbb{E}\int_0^t |a_s|^2 ds + 2\mathbb{E}\int_0^t |\gamma_s|^2 ds \\ &\le 2(1+t)K_0^2 \int_0^t \mathbb{E}|X_s - Y_s|^2 ds.\end{aligned}$$

Hence the function $v(t) := \mathbb{E}|X_t - Y_t|^2$ satisfies $v(t) \le 2(1+T)K_0^2 \int_0^t v(s)ds$. Gronwall lemma now implies that $v(t) = 0$, which means that X_t and Y_t are modifications of each other. Moreover, since these two are continuous, by Proposition 1.19, X_t and Y_t are indistinguishable.

Next we prove the existence. Put $Y_t^{(0)} = X_0$, and then define $Y_t^{(k)}$, $k = 1, 2, \ldots$, recursively by

$$Y_t^{(k+1)} = X_0 + \int_0^t b(s, Y_s^{(k)})ds + \int_0^t \sigma(s, Y_s^{(k)})dW_s. \tag{3.2.2}$$

Then by $X_0 \in L^2$ and the linearly growth condition for σ, we find $\{\sigma(s, Y_s^{(0)})\} \in \mathcal{L}_2$. From this and Doob's maximal inequality it follows that $\mathbb{E}[\max_{0\le t\le T} |Y_t^{(1)}|^2] < \infty$. Applying this argument recursively, we deduce that $\mathbb{E}[\max_{0\le t\le T} |Y_t^{(k)}|^2] < \infty$ for every $k \ge 0$. Then, as in the case of the uniqueness proof, for $k \ge 1$,

$$\mathbb{E}\max_{0\le s\le t} |Y_s^{(k+1)} - Y_s^{(k)}|^2 \le (2+8T)K_0^2 \mathbb{E}\int_0^t |Y_s^{(k)} - Y_s^{(k-1)}|^2 ds. \tag{3.2.3}$$

Here, we can use Doob's maximal inequality to estimate $\mathbb{E}\max_{0\le s\le t}\left|\int_0^s \gamma_u dW_s\right|^2$. Hence, by repeating the estimation (3.2.3) recursively, we obtain

$$\mathbb{E}\max_{0\le t\le T} |Y_t^{(k+1)} - Y_t^{(k)}|^2 \le K_2 \frac{K_3^k T^k}{k!}, \quad k \ge 0,$$

where

$$K_2 = \mathbb{E}\max_{0\le t\le T} |Y_t^{(1)} - Y_t^{(0)}|^2 < \infty$$

and $K_3 = (2+8T)K_0^2$. Chebyshev's inequality then leads to

$$\mathbb{P}\left(\max_{0\le t\le T} |Y_t^{(k+1)} - Y_t^{(k)}| > 2^{-k}\right) \le K_2 \frac{(4K_3T)^k}{k!}.$$

The series for the sequence in the right-hand side of the inequality just above converges, whence by Borel-Cantelli lemma, there exists $\Omega_0 \in \mathcal{F}_T$ with $\mathbb{P}(\Omega_0) = 1$ such that

$$\max_{0\le t\le T} |Y_t^{(k+1)}(\omega) - Y_t^{(k)}(\omega)| \le 2^{-k}, \quad k \ge n_0(\omega), \quad \omega \in \Omega_0,$$

for some $n_0(\omega)$ defined for each $\omega \in \Omega_0$. From this $\sum_{k=n_0(\omega)}^{\infty} \max_{0\le t\le T} |Y_t^{(k+1)}(\omega) - Y_t^{(k)}(\omega)| < \infty$ and so $Y_t^{(k)}(\omega)$ converges uniformly on $[0,T]$. Therefore, there exists a limiting function $X_t(\omega)$ such that $\sup_{0\le t\le T} |Y_t^{(k)}(\omega) - X_t(\omega)| \to 0$ (see, e.g., [47, 定理 13.4]). Since a uniformly converging limit of continuous functions is also continuous, we deduce that $\{X_t\}_{0\le t\le T}$ is adapted and a.s. continuous. Further, by Fatou's lemma, $\mathbb{E}[\max_{0\le t\le T} |X_t|^2] < \infty$. Hence in particular, $\{X_t\}$ satisfies the conditions (i) and (ii) in Definition 3.1. Moreover, since $\int_0^T |\sigma(t, Y_t^{(k)}) - \sigma(t, X_t)|^2 dt \to 0$, a.s. and there exists some subsequence $k_n \nearrow \infty$ such that $\int_0^t \sigma(s, Y_s^{k_n})dW_s \to \int_0^t \sigma(s, X_s)dW_s$ a.s. On the other hand, we have $Y_t^{k_n+1} \to X_t$, a.s. and $\int_0^t b(s, Y_s^{k_n})ds \to \int_0^t b(s, X_s)ds$, a.s. Thus, letting $k = k_n$, $n \to \infty$ in (3.2.2), we deduce that $\{X_t\}$ satisfies the condition (iii) in Definition 3.1. □

3.3 Explicit Solutions

We describe classes of SDEs having explicit solutions.

Linear cases

First assume that $m = 1$, i.e., consider the case of a scalar Brownian motion. It follows from Example 2.19 and Theorem 3.2 that the unique solution of the SDE

$$dX_t = bX_t dt + \sigma dW_t$$

is given by

$$X_t = e^{bt} X_0 + \int_0^t e^{b(t-s)} dW_s.$$

Then let us consider the more general SDE

$$dX_t = [a(t) + b(t)X_t]dt + \sigma(t)dW_t, \tag{3.3.1}$$

where $a, b, \sigma : [0, T] \to \mathbb{R}$ are bounded and Borel measurable. As in Example 2.19, using the product Itô formula, we observe

$$d\left(e^{-\int_0^t b(s)ds} X_t\right) = e^{-\int_0^t b(s)ds}(a(t)dt + \sigma(t)dW_t).$$

Thus, the unique solution of (3.3.1) is given by

$$X_t = e^{\int_0^t b(s)ds} X_0 + \int_0^t e^{\int_s^t b(r)dr}(a(s)ds + \sigma(s)dW_s).$$

Problem 3.4. Here consider general cases $m \geq 1$ and the scalar SDE

$$dX_t = [a(t) + b(t)X_t]dt + [X_t \gamma(t) + \sigma(t)]^\mathsf{T} dW_t, \tag{3.3.2}$$

where $a, b : [0, T] \to \mathbb{R}$ and $\gamma, \sigma : [0, T] \to \mathbb{R}^m$ are bounded and Borel measurable. Show that the unique solution of (3.3.2) is

$$X_t = Z_t \left[X_0 + \int_0^t Z_s^{-1}(a(s) - \gamma(s)^\mathsf{T}\sigma(s))ds + \int_0^t Z_s^{-1}\sigma(s)^\mathsf{T} dW_s\right],$$

where

$$Z_t = \exp\left[\int_0^t \left(b(s) - \frac{1}{2}|\gamma(s)|^2\right) ds + \int_0^t \gamma(s)^\mathsf{T} dW_s\right].$$

Problem 3.5. Consider the d-dimensional SDE

$$dX_t = (a(t) + b(t)X_t)dt + \sigma(t)dW_t, \tag{3.3.3}$$

where $a : [0, T] \to \mathbb{R}$, $b : [0, T] \to \mathbb{R}^{d\times d}$, and $\sigma : [0, T] \to \mathbb{R}^{d\times m}$, are bounded and Borel measurable. Assume that X_0 has a d-variate normal distribution with mean vector μ and covariance matrix ρ. Then, show that $\{X_t\}_{t\geq 0}$ is a Gaussian process with the representation

$$X_t = \Phi(t)\left(X_0 + \int_0^t \Phi^{-1}(s)a(s)ds + \int_0^t \Phi^{-1}(s)\sigma(s)dW_s\right),$$

and that the mean vector $\mu(t) = \mathbb{E}[X_t]$ and the covariance matrix $\rho(s, t) = \mathbb{E}[(X_s - m(s))(X_t - m(t))^\mathsf{T}]$, $s, t \geq 0$, are given respectively by

$$\mu(t) = \Phi(t)\left[\mu + \int_0^t \Phi^{-1}(s)a(s)ds\right],$$

$$\rho(s, t) = \Phi(s)\left[\rho + \int_0^{s\wedge t} \Phi^{-1}(r)\sigma(r)(\Phi^{-1}(r)\sigma(r))^\mathsf{T} dr\right]\Phi(t)^\mathsf{T}.$$

Here, a process is said to be Gaussian if any finite dimensional distribution is jointly normal, and $\Phi(t)$ is the unique solution of the matrix ODE

$$d\Phi(t) = b(t)\Phi(t)dt, \quad \Phi(0) = I_d.$$

Problem 3.6. Solve 2-dimensional SDE

$$dX_t = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} X_t dt + \begin{pmatrix} 0 \\ \sigma^2 \end{pmatrix} dW_t,$$

where $\{W_t\}_{t\geq 0}$ is one-dimensional.

Reducible cases

Here assume $m = 1$. Consider the one-dimensional SDE

$$dX_t = \frac{1}{2}\sigma(X_t)\sigma'(X_t)dt + \sigma(X_t)dW_t, \tag{3.3.4}$$

where $\sigma(\cdot) > 0$. To obtain the solution, we use the function

$$g(x) = \int_0^x \frac{1}{\sigma(\xi)}d\xi, \tag{3.3.5}$$

defined for x in a possible state space of $\{X_t\}$. Then, since

$$(g^{-1})'(x) = \sigma(g^{-1}(x)), \quad (g^{-1})''(x) = \sigma(g^{-1}(x))\sigma'(g^{-1}(x)),$$

the process $X_t := g^{-1}(W_t + g(X_0))$ satisfies $(g^{-1})'(W_t + g(X_0)) = \sigma(X_t)$ and $(g^{-1})''(W_t + g(X_0)) = \sigma(X_t)\sigma'(X_t)$. Thus, by Itô formula, we find that X_t is a solution to (3.3.4).

Problem 3.7. Solve

$$dX_t = \frac{1}{2}a^2 X_t dt + a\sqrt{1 + X_t^2}dW_t.$$

Problem 3.8. Solve

$$dX_t = \frac{1}{2}a(a-1)X_t^{1-2/a}dt + aX_t^{1-1/a}dW_t.$$

Next consider the SDE of the form

$$dX_t = \left(\alpha\sigma(X_t) + \frac{1}{2}\sigma(X_t)\sigma'(X_t)\right)dt + \sigma(X_t)dW_t, \tag{3.3.6}$$

As in the previous case, we observe the process $X_t := g^{-1}(\alpha t + W_t + g(X_0))$ satisfies (3.3.6), where g is given by (3.3.5).

Problem 3.9. Solve

$$dX_t = \left(\frac{1}{2}X_t + \sqrt{1 + X_t^2}\right)dt + \sqrt{1 + X_t^2}dW_t.$$

Problem 3.10. Solve

$$dX_t = -(\alpha + \beta^2 X_t)(1 - X_t^2)dt + \beta(1 - X_t^2)dW_t.$$

Generalizing these results, we have the following. The proof is left to the reader.

Proposition 3.11

Suppose that b is Lipschitz continuous on $\mathbb{R}$ and σ is of class $C^2(\mathbb{R})$ with bounded first and second derivatives. Then the unique solution $\{X_t\}_{t\geq 0}$ of the one-dimensional SDE

$$dX_t = \left[b(X_t) + \frac{1}{2}\sigma(X_t)\sigma'(X_t)\right]dt + \sigma(X_t)dW_t \tag{3.3.7}$$

is represented as $X_t = u(W_t, Y_t)$, where $u : \mathbb{R}^2 \to \mathbb{R}$ is the solution of the ODE

$$\partial_x u(x,y) = \sigma(u(x,y)), \quad u(0,y) = y,$$

and the process $\{Y_t\}_{t\geq 0}$ is the solution of the ODE

$$dY_t = f(W_t, Y_t)dt, \quad Y_0 = X_0$$

with

$$f(x,y) = \exp\left(-\int_0^x \sigma'(u(z,y))dz\right) b(u(x,y)).$$

3.4 Numerical Solutions

When explicit solutions of SDEs are unavailable, we need to approximate the equations to generate the sample paths in computer simulations or to compute the expectation of quantities involving the solutions. Here we present the Euler–Maruyama method, which is a most popular one for the time discretization, and can be seen as a stochastic version of the Euler method in ODEs.

Consider the SDE (3.2.1) with the drift coefficient b and the diffusion coefficient σ. We impose the following conditions on b and σ:

Assumption 3.12

There exists a positive constant C_0 such that

$$|b(t,x) - b(s,y)| + |\sigma(t,x) - \sigma(s,y)| \leq C_0(|t-s|^{1/2} + |x-y|), \quad t,s \in [0,T], \ \ x,y \in \mathbb{R}^d.$$

Assumption 3.12 means the conditions in Theorem 3.2. Thus, under Assumption 3.12, there exists a unique solution $\{X_t\}$ of (3.2.1).

First, set $t_k = kT/n$, $k = 0, \ldots, n$. We start with the representation

$$X_{t_k} = X_{t_{k-1}} + \int_{t_{k-1}}^{t_k} b(s, X_s)ds + \int_{t_{k-1}}^{t_k} \sigma(s, X_s)dW_s.$$

Since $\{X_t\}$ has continuous sample paths, the approximation $X_s \approx X_{t_{k-1}}$, $s \in [t_{k-1}, t_k]$, is reasonable for sufficiently large n. Applying this approximation, we have

$$X_{t_k} \approx X_{t_{k-1}} + \int_{t_{k-1}}^{t_k} b(t_{k-1}, X_{t_{k-1}})ds + \int_{t_{k-1}}^{t_k} \sigma(t_{k-1}, X_{t_{k-1}})dW_s,$$

which is equivalent to

$$X_{t_k} \approx X_{t_{k-1}} + b(t_{k-1}, X_{t_{k-1}})(t_k - t_{k-1}) + \sigma(t_{k-1}, X_{t_{k-1}})(W_{t_k} - W_{t_{k-1}}).$$

The random variable $W_{t_k} - W_{t_{k-1}}$ follows the normal distribution with mean vector 0 and covariance matrix $(T/N)I_d$, which can be generated by pseudo random numbers. Therefore, the sequence $\{Y_k\}_{k=0}^n$ defined by

$$Y_{k+1} = Y_k + b(t_k, Y_k)(t_{k+1} - t_k) + \sigma(t_k, Y_k)(W_{t_{k+1}} - W_{t_k}) \tag{3.4.1}$$

with $Y_0 = X_0$ is a candidate of an implementable numerical solution for (3.2.1).

Hereafter, we discuss a rate of convergence of $\{Y_k\}$ to $\{X_t\}$.

Lemma 3.13

Suppose that Assumption 3.12 hold. Let $\{X_t\}_{0\le t\le T}$ be as above. Then, there exists a positive constant C such that

$$\mathbb{E}|X_t - X_s|^2 \le C(t-s), \quad 0 \le s \le t \le T.$$

Proof. Using the inequality $(a+b)^2 \le 2(a^2+b^2)$, we see

$$\mathbb{E}|X_t - X_s|^2 \le 2\mathbb{E}\left[\left|\int_s^t b(r, X_r)dr\right|^2\right] + 2\mathbb{E}\left[\left|\int_s^t \sigma(r, X_r)dW_r\right|^2\right]. \tag{3.4.2}$$

By the linear growth condition, the 1st term of the right-hand side in (3.4.2) is at most

$$2\mathbb{E}\left|\int_s^t b(r, X_r)dr\right|^2 \le 2(t-s)\int_s^t \mathbb{E}|b(r, X_r)|^2 dr \le C'\left(1 + \mathbb{E}\left[\sup_{0\le r\le T}|X_r|^2\right]\right)(t-s),$$

where C' is a positive constant. A similar estimation works for the 2nd term of the right-hand side in (3.4.2). □

Roughly speaking, the approximation error for the Euler-Maruyama methods is $O(n^{-1/2})$.

Theorem 3.14

Suppose that Assumption 3.12 hold. Let $\{X_t\}_{0\le t\le T}$ be as above and let $\{Y_k\}_{k=0}^n$, $n \in \mathbb{N}$, be the sequences defined by (3.4.1). Then, there exists a positive constant C_1 such that

$$\max_{k=0,1,\dots,n} \mathbb{E}|X_{t_k} - Y_k|^2 \le \frac{C_1}{n}.$$

Proof. By C we denote positive constants that do not depend on n and $k = 0, 1, \dots, n$ and that may vary from line to line.

First notice that Y_k is $\mathcal{F}_{t_k}$-measurable and in L^2 for each $k = 0, 1, \dots, n$. To confirm the latter property, assume that $Y_k \in L^2$ for some k and observe

$$|Y_{k+1}|^2 \le 3|Y_k|^2 + 3|b(t_k, Y_k)|^2(\Delta t)^2 + 3|\sigma(t_k, Y_k)\Delta W_{k+1}|^2, \tag{3.4.3}$$

where $\Delta t = T/n$ and $\Delta W_{k+1} = W_{t_{k+1}} - W_{t_k}$. From (3.4.3), the linearly growth conditions on b, σ, and $\mathbb{E}|Y_k|^2|\Delta W_{k+1}|^2 = \mathbb{E}|Y_k|^2\mathbb{E}|\Delta W_{k+1}|^2$ it follows that $\mathbb{E}|Y_{k+1}|^2 \le C\mathbb{E}|Y_k|^2 < \infty$.

Next, observe

$$X_{t_{k+1}} - Y_{k+1} = X_{t_k} - Y_k + \int_{t_k}^{t_{k+1}} \Delta b_s ds + \int_{t_k}^{t_{k+1}} \Delta\sigma_s dW_s,$$

where

$$\Delta b_s = b(s, X_s) - b(t_k, Y_k), \quad \Delta\sigma_s = \sigma(s, X_s) - \sigma(t_k, Y_k), \quad s \in [t_k, t_{k+1}).$$

Furthermore we have

$$
\begin{aligned}
&|X_{t_{k+1}} - Y_{k+1}|^2 \\
&= |X_{t_k} - Y_k|^2 + \left| \int_{t_k}^{t_{k+1}} \Delta b_s ds \right|^2 + \left| \int_{t_k}^{t_{k+1}} \Delta \sigma_s dW_s \right|^2 + 2(X_{t_k} - Y_k)^\mathsf{T} \int_{t_k}^{t_{k+1}} \Delta b_s ds \\
&\quad + 2(X_{t_k} - Y_k)^\mathsf{T} \int_{t_k}^{t_{k+1}} \Delta \sigma_s dW_s + 2 \left(\int_{t_k}^{t_{k+1}} \Delta b_s ds \right)^\mathsf{T} \int_{t_k}^{t_{k+1}} \Delta \sigma_s dW_s.
\end{aligned}
$$

By Cauchy-Schwartz inequality, the Lipschitz continuity of b, and Lemma 3.13,

$$
\begin{aligned}
\mathbb{E} \left| \int_{t_k}^{t_{k+1}} \Delta b_s ds \right|^2 &\leq \Delta t \int_{t_k}^{t_{k+1}} \mathbb{E} |\Delta b_s|^2 ds \\
&\leq C \Delta t \int_{t_k}^{t_{k+1}} \mathbb{E}[s - t_k + |X_s - X_{t_k}|^2 + |X_{t_k} - Y_k|^2] ds \\
&\leq C(\Delta t)^3 + C(\Delta t)^2 \mathbb{E}|X_{t_k} - Y_k|^2.
\end{aligned}
$$

Using Itô isometry, similarly we have

$$
\mathbb{E} \left| \int_{t_k}^{t_{k+1}} \Delta \sigma_s dW_s \right|^2 \leq C(\Delta t)^2 + C \Delta t \mathbb{E} |X_{t_k} - Y_k|^2,
$$

whence

$$
\begin{aligned}
2\mathbb{E} \left(\int_{t_k}^{t_{k+1}} \Delta b_s ds \right)^\mathsf{T} \int_{t_k}^{t_{k+1}} \Delta \sigma_s dW_s &\leq \mathbb{E} \left| \int_{t_k}^{t_{k+1}} \Delta b_s ds \right|^2 + \mathbb{E} \left| \int_{t_k}^{t_{k+1}} \Delta \sigma_s dW_s \right|^2 \\
&\leq C(\Delta t)^2 + C \Delta t \mathbb{E}|X_{t_k} - Y_k|^2.
\end{aligned}
$$

Using Young's inequality $ab \leq ca^2/2 + b^2/(2c)$ for $a, b \in \mathbb{R}$ and $c > 0$, we find

$$
\begin{aligned}
2\mathbb{E}(X_{t_k} - Y_k)^\mathsf{T} \int_{t_k}^{t_{k+1}} \Delta b_s ds &\leq \Delta t \mathbb{E} |X_{t_k} - Y_k|^2 + \frac{1}{\Delta t} \mathbb{E} \left| \int_{t_k}^{t_{k+1}} \Delta b_s ds \right|^2 \\
&\leq C(\Delta t)^2 + C \Delta t \mathbb{E}|X_{t_k} - Y_k|^2.
\end{aligned}
$$

As for the remaining term, we have

$$
\mathbb{E}(X_{t_k} - Y_k)^\mathsf{T} \int_{t_k}^{t_{k+1}} \Delta \sigma_s dW_s = \mathbb{E} \left[(X_{t_k} - Y_k)^\mathsf{T} \mathbb{E} \left[\int_{t_k}^{t_{k+1}} \Delta \sigma_s dW_s \middle| \mathcal{F}_{t_k} \right] \right] = 0.
$$

Collecting the estimates above, we deduce

$$
\mathbb{E}|X_{t_{k+1}} - Y_{k+1}|^2 \leq (1 + C\Delta t) \mathbb{E} |X_{t_k} - Y_k|^2 + C(\Delta t)^2, \quad k = 0, \ldots, n-1.
$$

From this the theorem easily follows. □

Example 3.15. Let us examine the Euler-Maruyama approximation for the SDE

$$
dX_t = X_t(0.5\, dt + 0.2\, dW_t), \quad 0 \leq t \leq 1,
$$

with $X_0 = 1$. The time grids are set to be $t_i = i/n$, $i = 0, 1, \ldots, n$. We execute the simulation $M = 10^6$ times and compute the resulting mean squared error

$$
L^2\text{-error} = \max_{i=1,\ldots,n} \frac{1}{M} \sum_{k=1}^{M} (X_{t_i}^{(k)} - Y_{t_i}^{(k)})^2,
$$

where $\{X_{t_i}^{(k)}\}$ and $\{Y_{t_i}^{(k)}\}$ denotes the k-th sample paths of the true and approximate solutions, respectively. See Figure 3.4.1 below.

Problem 3.16. As in Example 3.15, evaluate the performance of the Euler-Maruyama method for the SDE in Problem 3.7.

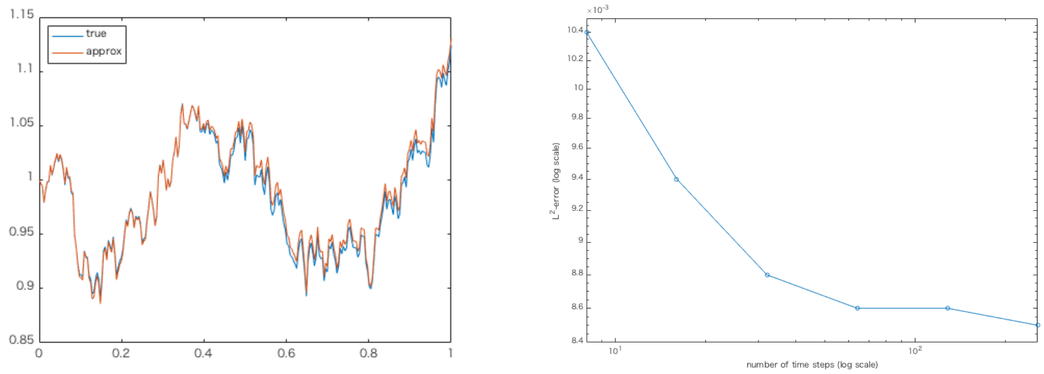

Figure 3.4.1: Sample paths of the true and approximate solutions in the case of $n = 2^8$ (left) and plotting L^2-errors for $n = 2^3, 2^4, 2^5, 2^6, 2^7, 2^8$ (right).

3.5 Fundamental Properties

We write $\{X_s^{t,x}\}_{t\le s\le T}$ for the solution of the SDE with initial condition $X_t = x$, i.e.,

$$X_s^{t,x} = x + \int_t^s b(r, X_r^{t,x})dr + \int_t^s \sigma(r, X_r^{t,x})dW_r. \tag{3.5.1}$$

Notice that we can ensure the existence and uniqueness of this SDE by considering the SDE on $[0,T]$ with coefficients $\tilde{b}(s,x) = b(s,x)1_{[t,T]}(r)$ and $\tilde{\sigma}(r,x) = \sigma(r,x)1_{[t,T]}(r)$, provided that b and σ satisfy the conditions (i) and (ii) imposed in Theorem 3.2.

In what follows, we often drop the superscripts t,x in $(X_s^{t,x})$ and write $\mathbb{E}^{t,x}[Z]$ for $\mathbb{E}[Z]$ when Z depends on $(X_s^{t,x})$. Using Itô formula, we observe

$$\begin{aligned} b(t,x) &= \lim_{\Delta t\to 0}\frac{1}{\Delta t}\mathbb{E}^{t,x}[X_{t+\Delta t} - X_t], \\ \sigma(t,x)\sigma^{\mathsf{T}}(t,x) &= \lim_{\Delta t\to 0}\frac{1}{\Delta t}\mathbb{E}^{t,x}[(X_{t+\Delta t} - X_t)(X_{t+\Delta t} - X_t)^{\mathsf{T}}]. \end{aligned} \tag{3.5.2}$$

Here, the expectations are taken to be component-wise. In general, the coefficients $b(t,x)$ and $\sigma(t,x)$ of the SDE are called the *drift term* and the *diffusion term*, respectively.

Problem 3.17. Prove (3.5.2).

Markov property

We begin with Markov property.

Theorem 3.18

Suppose that b, σ, and ξ satisfy the assumptions in Theorem 3.2. Then the unique solution $\{X_t\}_{0\le t\le T}$ of the SDE (3.2.1) is an $\mathbb{F}$-Markov process.

Proof. We will give a proof in the case where b and σ satisfy Assumption 3.12. We refer to standard textbooks on the stochastic analysis such as [31] for a proof of the general claim.

Fix $t \in [0,T]$ and $s \in [0, T-t]$. Then $\{X_r\}$ satisfies.

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

Let $t_k = ks/n + t$, $k = 0, \dots, n$, $n \in \mathbb{N}$, and $\{Y_{t_k}\}_{k=0}^n$ the Euler-Maruyama approximation of $\{X_r\}_{t\le r\le t+s}$, i.e.,

$$Y_{t_k} = Y_{t_{k-1}} + b(t_{k-1}, Y_{t_{k-1}})(t_k - t_{k-1}) + \sigma(t_{k-1}, Y_{t_{k-1}})(W_{t_k} - W_{t_{k-1}}), \quad k = 1, \dots, n,$$

with $Y_{t_0} = X_t$. Theorem 3.14 then yields $X_{t+s} = \lim_{n\to\infty} Y_{t_n}$ a.s. possibly along subsequence. Since $W_{t_k} - W_{t_{k-1}} = W_{ks/n+t} - W_t - (W_{(k-1)s/n+t} - W_t)$, by induction, we observe that Y_{t_n} is $\sigma(X_t, W_{r+t} - W_t : 0 \le r \le s)$-measurable, whence so is $\limsup_{n\to\infty} Y_{t_n}$. Therefore, by Theorem 1.9, $X_{t+s} = F_{t+s}(X_t, (W_{r+t} - W_t)_{0\le r\le s})$ a.s. for some Borel function F_{t+s} on $\mathbb{R}^d \times C([0, s]; \mathbb{R}^d)$ for $0 \le s \le T - t$. Since $(W_{r+t} - W_t)_{0\le r\le s}$ is independent of $\mathcal{F}_t$, using Lemma 1.45, we have, for every bounded Borel function f,

$$\begin{aligned}\mathbb{E}[f(X_{t+s})|\mathcal{F}_t] &= \mathbb{E}[f(F_{t+s}(X_t, (W_{t+r} - W_t)_{r\le s})|\mathcal{F}_t] = \mathbb{E}[f(F_{t+s}(x, (W_{t+r} - W_t)_{r\le s})]|_{x=X_t} \\ &= \mathbb{E}[f(X_{t+s})|X_t],\end{aligned}$$

as required. □

By arguments similar to that In the proof of Theorem 3.18, we have the following result:

Corollary 3.19

Let b, σ, and $\{X_t\}$ be as above. Then,

$$\mathbb{E}[f(X_{t+s})|\mathcal{F}_t] = \mathbb{E}[f(X^{t,x}_{t+s})]|_{x=X_t}, \quad \text{a.s.}$$

for any $t, s \in [0, T]$ with $0 \le t + s \le T$ and any bounded Borel measurable function f.

Proof. Let f be a bounded Lipschitz continuous function. As in the proof the previous theorem, there exists a sequence G_n of $\mathcal{B}(\mathbb{R}^d) \times C([0, s]; \mathbb{R}^d)$-measurable functions such that

$$\begin{aligned}&\lim_{n\to\infty} \mathbb{E}|X^{t,x}_{t+s} - G_n(x, (W_{r+t} - W_t)_{0\le r\le s})|^2 = 0, \quad x \in \mathbb{R}^d, \\ &\lim_{n\to\infty} \mathbb{E}|X^{t,X_t}_{t+s} - G_n(X_t, (W_{r+t} - W_t)_{0\le r\le s})|^2 = 0,\end{aligned}$$

and $X^{t,X_t}_{t+s} = X_t$, a.s. This means

$$\lim_{n\to\infty} \mathbb{E}[f(G_n(x, (W_{r+t} - W_t)_{0\le r\le s})] = \mathbb{E}[f(X^{t,x}_{t+s})], \quad x \in \mathbb{R}^d,$$

and

$$\lim_{n\to\infty} \mathbb{E}[f(G_n(X_t, (W_{r+t} - W_t)_{0\le r\le s})|X_t] = \mathbb{E}[f(X^{t,X_t}_{t+s})|X_t]$$

in L^2, whence

$$\lim_{n\to\infty} \mathbb{E}[f(G_n(x, (W_{r+t} - W_t)_{0\le r\le s})]|_{x=X_t} = \mathbb{E}[f(X^{t,x}_{t+s})]|_{x=X_t},$$

and

$$\lim_{k\to\infty} \mathbb{E}[f(G_{n_k}(X_t, (W_{r+t} - W_t)_{0\le r\le s})|X_t] = \mathbb{E}[f(X^{t,X_t}_{t+s})|X_t], \quad \text{a.s.},$$

for some subsequence $\{n_k\}$. Combining these observations and Theorem 3.18, we obtain

$$\mathbb{E}[f(X_{t+s})|\mathcal{F}_t] = \mathbb{E}[f(X^{t,X_t}_{t+s})|X_t] = \mathbb{E}[f(X^{t,x}_{t+s})]|_{x=X_t}. \tag{3.5.3}$$

Using Lemma below, we can show that (3.5.3) holds true for any bounded Borel measurable function f. □

Lemma 3.20

Let X, Y be $\mathbb{R}^d$-valued random variables, and $\mathcal{G}$ a sub σ-algebra. Suppose that

$$\mathbb{E}[f(X)|\mathcal{G}] = \mathbb{E}[f(Y)|\mathcal{G}], \quad \text{a.s.} \tag{3.5.4}$$

for any bounded Lipschitz continuous function f. Then (3.5.4) holds for any bounded Borel function f.

*Proof**. Step (i). Let $A \subset \mathbb{R}^d$ be closed. Then

$$g_n(x) := (1 - n\gamma(x, A))^+, \quad x \in \mathbb{R}^d,$$

where $\gamma(x, A) = \inf_{y \in A} |x-y|$, is bounded and Lipschitz continuous. Indeed, for $\varepsilon > 0$ take $z \in A$ such that $|y - z| \le \gamma(y, A) + \varepsilon$. Then $g_n(x) - g_n(y) \le n(|x - z| - |y - z| + \varepsilon) \le n(|x - y| + \varepsilon)$, from which we find $|g_n(x) - g_n(y)| \le n|x - y|$. Further, $1_A(x) \le g_n(x) \le 1_{A^{1/n}}(x)$, where $A^{1/n} = \{x : \gamma(x, A) < 1/n\}$. Thus,

$$\mathbb{E}[1_A(X)|\mathcal{G}] \le \mathbb{E}[g_n(X)|\mathcal{G}] = \mathbb{E}[g_n(Y)|\mathcal{G}] \le \mathbb{E}[1_{A^{1/n}}(Y)|\mathcal{G}].$$

Letting $n \to \infty$, we obtain $\mathbb{E}[1_A(X)|\mathcal{G}] \le \mathbb{E}[1_A(Y)|\mathcal{G}]$. Changing the role of X and Y, we have the converse inequality, whence the equality.

Step (ii). Let $A \in \mathcal{B}(\mathbb{R}^d)$. We will use the fact that for each n there exists a closed set F_n such that $F_n \subset A$ and $\mathrm{Leb}(A \setminus F_n) \le 1/2^n$, where Leb denotes the Lebesgue measure on $\mathbb{R}^d$ (see, e.g., [43, Theorem 7.6] or [5, Theorem 1.1]). Then, we inductively define the sequence $\{A_n\}$ of closed sets by $A_{n+1} = A_n \cup F_n$. By this construction, 1_{A_n} is monotone nondecreasing and converges to 1_A, Leb-a.e. Applying the monotone convergence theorem, we obtain (3.5.4) for $f = 1_A$.

Step (iii). Any bounded Borel function can be represented as the difference of bounded nonnegative Borel functions, and each part can be approximated by a monotonically nondecreasing sequence of step functions. Thus the lemma follows. □

Next consider the homogeneous case, i.e., the SDE of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x. \tag{3.5.5}$$

Here, $b : \mathbb{R}^d \to \mathbb{R}$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are assumed to be Lipschitz continuous. Then, by Theorem 3.2, the SDE (3.5.5) has a unique solution $\{X_t\}_{t\ge 0}$. Then we have the following strong Markov property for $\{X_t\}$:

Theorem 3.21

Let b, σ, and $\{X_t\}$ be as above. Further, let θ be a stopping time with $\theta < \infty$, a.s. Then, for any bounded Borel measurable function f on $\mathbb{R}^d$, we have

$$\mathbb{E}[f(X_{t+\theta})|\mathcal{F}_\theta] = \mathbb{E}[f(X_{t+\theta})|X_\theta], \quad \text{a.s.}$$

Proof. Fix $t \ge 0$. Let $t_k = \theta + tk/n$, $k = 0, \dots, n$, $n \in \mathbb{N}$. Then consider the Euler-Maruyama approximation $\{Y_k\}_{k=0}^n$ of $\{X_s\}_{\theta \le s \le \theta+t}$, defined by

$$Y_{k+1} = Y_k + b(Y_k)(t_{k+1} - t_k) + \sigma(Y_k)(W_{t_{k+1}} - W_{t_k}), \quad Y_0 = X_\theta.$$

Then, we see that Y_n is $\sigma(X_\theta, (W_{s+\theta} - W_\theta)_{0\le s\le t})$-measurable and $X_{t+\theta} = \lim_{n\to\infty} Y_n$, a.s. possibly along subsequence. Thus, there exists a Borel measurable map F_t from $\mathbb{R}^d \times C([0, s]; \mathbb{R}^d)$

into $\mathbb{R}^d$ such that $X_{t+\theta} = F_t(X_\theta, (W_{s+\theta} - W_\theta)_{0\le s\le t})$ a.s. Since $(W_{s+\theta} - W_\theta)_{0\le s\le t}$ is independent of $\mathcal{F}_\theta$ by Theorem 1.46, we have

$$\begin{aligned}\mathbb{E}[f(X_{t+\theta})|\mathcal{F}_\theta] &= \mathbb{E}[f(F_t(X_\theta, (W_{s+\theta} - W_\theta)_{s\le t})|\mathcal{F}_\theta] = \mathbb{E}[f(F_t(y, (W_{s+\theta} - W_\theta)_{s\le t})]|_{y=X_\theta} \\ &= \mathbb{E}[f(X_{t+\theta})|X_\theta],\end{aligned}$$

whence the claim. □

- In the theory of Markov processes, a strong Markov process with continuous sample paths is called a *diffusion* process.

Feynman-Kac formula

Let $\{X_t\}$ be the unique solution of the SDE (3.2.1) with nonrandom initial condition. With the coefficients b and σ, we consider the differential operator

$$(\mathcal{A}_t f)(x) := \sum_{i=1}^d b_i(t,x)\partial_{x_i} f(x) + \frac{1}{2}\sum_{i,j=1}^d \sum_{k=1}^m \sigma_{ik}(t,x)\sigma_{jk}(t,x)\partial^2_{x_i x_j} f(x), \quad f \in C^2(\mathbb{R}^d).$$

We write $(\mathcal{A}_t f)(t,x) = (\mathcal{A}_t f(t,\cdot))(x)$ when f also depends on the time variable t. Notice that the term $\mathcal{A}_t f$ appears in applying Itô formula to $f(t, X_t)$.

Now, suppose that the partial differential equation (PDE)

$$\begin{aligned}\partial_t u + \mathcal{A}_t u = 0, &\quad \text{on } [0,T) \times \mathbb{R}^d, \\ u(T,\cdot) = g, &\quad \text{on } \mathbb{R}^d\end{aligned} \tag{3.5.6}$$

has a solution $u(t,x)$ of $C^{1,2}$ class. Then by Itô formula,

$$\begin{aligned}g(X_T) &= u(T, X_T) \\ &= u(0,X_0) + \int_0^T (\partial_t u + \mathcal{A}_t u)(t, X_t)dt + \sum_{i=1}^d \sum_{k=1}^m \int_0^T \partial_{x_i} u(t,X_t)\sigma_{ik}(t,X_t)dW_t^k.\end{aligned}$$

Since u satisfies the PDE (3.5.6), the "dt term" turns out to be zero. Moreover, if the term of the stochastic integral is a martingale, which is the case of the integrand belongs to $\mathcal{L}_2$, then by taking the expectation, we get

$$\mathbb{E}[g(X_T)] = u(0, X_0).$$

Let us generalize the argument above. Consider continuous functions $g : \mathbb{R}^d \to \mathbb{R}$, $f : [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $\ell : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ such that for any $t \in [0,T]$ and $x \in \mathbb{R}^d$

$$\begin{aligned}|g(x)| + |f(t,x)| &\le C_0(1 + |x|^2), \\ \ell(t,x) &\ge 0\end{aligned} \tag{3.5.7}$$

for some constant $C_0 > 0$. Further, consider the PDE

$$\begin{aligned}\partial_t u + \mathcal{A}_t u + f - \ell u = 0, &\quad \text{on } [0,T) \times \mathbb{R}^d, \\ u(T,\cdot) = g, &\quad \text{on } \mathbb{R}^d.\end{aligned} \tag{3.5.8}$$

Theorem 3.22: Feynman-Kac

Suppose that b, σ, and ξ satisfies the conditions in Theorem 3.2. Let $\{X_t\}_{0\le t\le T}$ be the unique solution of (3.2.1). Suppose moreover that (3.5.7) holds and the PDE (3.5.8) has a classical solution $u(t,x)$ of $C^{1,2}$-class. Further, assume that there exists a constant $M>0$ such that

$$\max_{0\le t\le T}|u(t,x)|\le M(1+|x|^2),\quad x\in\mathbb{R}^d.$$

Then,

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

- This result and Corollary 3.19 imply

$$\mathbb{E}\left[g(X_T)e^{-\int_t^T \ell(r,X_r)dr}+\int_t^T f(s,X_s)e^{-\int_t^s \ell(r,X_r)dr}ds\,\middle|\,\mathcal{F}_t\right]=u(t,X_t).$$

- The condition (3.5.7) and the growth condition on u can be weakened. We refer to [21, Chpater 5] for details on this point and for a sufficient condition for which the PDE (3.5.8) has a classical solution.

Proof of Theorem 3.22. Consider the stopping times $\tau_n=\inf\{s\ge t:|X_s^{t,x}|\ge n\}$, $n\ge 1$. Applying Itô formula to $e^{-\int_t^s \ell(r,X_r^{t,x})dr}u(s,X_s^{t,x})$, we find

$$\begin{aligned}e^{-\int_t^{T\wedge\tau_n} \ell(r,X_r^{t,x})dr}u(T\wedge\tau_n,X_{T\wedge\tau_n}^{t,x})&=u(t,x)-\int_t^{T\wedge\tau_n}e^{-\int_t^s \ell(r,X_r^{t,x})dr}f(s,X_s^{t,x})ds\\&\quad+\sum_{i=1}^d\sum_{k=1}^m\int_t^{T\wedge\tau_n}\partial_{x_i}u(s,X_s^{t,x})\sigma_{ik}(s,X_s^{t,x})dW_s^k.\end{aligned}$$

Since $|X_s^{t,x}|\le n$ for $s\le T\wedge\tau_n$, the process $\partial_{x_i}u(s,X_s^{t,x})\sigma_{ik}(s,X_s^{t,x})1_{\{s\le\tau_n\}}$, $t\le s\le T$, belongs to $\mathcal{L}_2$. Therefore,

$$u(t,x)=\mathbb{E}\left[e^{-\int_t^{T\wedge\tau_n} \ell(r,X_r^{t,x})dr}u(T\wedge\tau_n,X_{T\wedge\tau_n}^{t,x})+\int_t^{T\wedge\tau_n}e^{-\int_t^s \ell(r,X_r^{t,x})dr}f(s,X_s^{t,x})ds\right].$$

By (3.5.8), the growth condition on u, and $\max_{0\le s\le T}|X_s|^2\in L^2$, we can use the dominated convergence theorem to obtain the required result by letting $n\to\infty$. □

Transition density

Suppose that b, σ, and ξ satisfies the conditions in Theorem 3.2. Let $\{X_t\}_{0\le t\le T}$ be the unique solution of (3.2.1). By definition, the sample paths of X is almost surely continuous. In other words, $\mathbb{P}(X\in\mathbb{W}^d)=1$, where $\mathbb{W}^d=C([0,T];\mathbb{R}^d)$ is a Banach space with sup norm. Thus, X induces the measure

$$\mu_X(B):=\mathbb{P}(X\in B),\quad B\in\mathcal{B}(\mathbb{W}^d),$$

on $(\mathbb{W}^d,\mathcal{B}(\mathbb{W}^d))$, which is the *law* of the solution $\{X_t\}$ of the SDE as $\mathbb{W}^d$-valued random variable. Let $\mathcal{C}$ be the totality of sets of the form

$$B=\{w\in\mathbb{W}^d:(w(t_1),\dots,w(t_n))\in E\},\quad 0\le t_1\le\cdots\le t_n\le T,\;\; E\in\mathcal{B}(\mathbb{R}^{nd}),\;\; n\ge 1.$$

An element of $\mathcal{C}$ is called a *cylinder set*. Since the mapping $\mathbb{W}^d\ni w\mapsto(w(t_1),\dots,w(t_n))$ is continuous, we have $\mathcal{C}\subset\mathcal{B}(\mathbb{W}^d)$. We say the family $\mu_{t_1,\dots,t_n}$, $0\le t_1\le\cdots\le t_n\le T$,

$n \geq 1$, of probability measures defined by $\mu_{t_1,\dots,t_n}(E) = \mathbb{P}((X_{t_1},\dots,X_{t_n}) \in E)$, $E \in \mathbb{R}^{nd}$, the *finite dimensional distributions* of $\{X_t\}$. Thus, the finite dimensional distributions of $\{X_t\}$ are described by the values of μ_X on cylinder sets.

Proposition 3.23

The law of X is uniquely determined by its finite dimensional distributions.

*Proof**. First note that there exists a countable base for the topology of $\mathbb{W}^d$, consisting of sets of the form $\{w : \max_{0\leq t\leq T} |w(t) - w_0(t)| < \delta\}$, $w_0 \in \mathbb{W}^d$, $\delta > 0$. Observe

$$\begin{aligned}\{w : \max_{0\leq t\leq T} |w(t) - w_0(t)| < \delta\} &= \bigcup_{n=1}^{\infty} \left\{w : \max_{0\leq t\leq T} |w(t) - w_0(t)| \leq \delta - \frac{1}{n}\right\} \\ &= \bigcup_{n=1}^{\infty} \bigcap_{q\in\mathbb{Q}\cap[0,T]} \left\{w : |w(q) - w_0(q)| \leq \delta - \frac{1}{n}\right\} \in \sigma(\mathcal{C}),\end{aligned}$$

from which we have $\mathcal{B}(\mathbb{W}^d) \subset \sigma(\mathcal{C})$. Therefore $\mathcal{B}(\mathbb{W}^d) = \sigma(\mathcal{C})$.

Suppose that $\mu_X = \nu$ on $\mathcal{C}$ for some probability measure ν on $(\mathbb{W}^d, \mathcal{B}(\mathbb{W}^d))$. Then, since $\mathcal{C}$ is a π-system, we can apply π-system lemma (see Lemma A.44) to deduce $\mu_X = \nu$ on $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{W}^d)$. $\square$

A nonnegative Borel function $p(t, x; s, y)$, $0 \leq t < s \leq T$, $x, y \in \mathbb{R}^d$, said to be the *transition probability density* of $\{X_s\}$ if it satisfies

$$\mathbb{P}(X_s^{t,x} \in A) = \int_A p(t, x; s, y)dy, \quad A \in \mathcal{B}(\mathbb{R}^d).$$

Now suppose that $\{X_t\}$ has the transition density $p(t, x; s, y)$. Then we will represent the finite dimensional distribution of $\{X_t\}$ by p. To this end, choose $0 < t_1 < t_2 < t_3 \leq T$ and a bounded Borel function f on $\mathbb{R}^d$. Then, by the Markov property (Corollary 3.19,

$$\mathbb{E}[f(X_{t_3})|\mathcal{F}_{t_2}] = g(X_{t_2}),$$

where

$$g(x_2) = \mathbb{E}[f(X_{t_3}^{t_2,x_2})] = \int_{\mathbb{R}^d} f(x_3)p(t_2, x_2; t_3, x_3)dx_3.$$

Hence, by the definition of the conditional expectation,

$$\mathbb{E}[f(X_{t_3})1_{\{X_0\in B_0, X_{t_1}\in B_1, X_{t_2}\in B_2\}}] = \mathbb{E}[g(X_{t_2})1_{\{X_0\in B_0, X_{t_1}\in B_1, X_{t_2}\in B_2\}}],$$

whence

$$\mathbb{E}[g(X_{t_2})1_{\{X_0\in B_0, X_{t_1}\in B_1, X_{t_2}\in B_2\}}] = \mathbb{E}[h(X_{t_1})1_{\{X_0\in B_0, X_{t_1}\in B_1\}}].$$

Here,

$$h(x_1) = \mathbb{E}[g(X_{t_2}^{t_1,x_1})1_{\{X_{t_2}\in B_2\}}] = \int_{B_2} g(x_2)p(t_1, x_1; t_2, x_2)dx_2.$$

Consequently we obtain

$$\begin{aligned}&\mathbb{E}[f(X_{t_3})1_{\{X_0\in B_0, X_{t_1}\in B_1, X_{t_2}\in B_2\}}] = \mathbb{E}[h(X_{t_1})1_{\{X_0\in B_0, X_{t_1}\in B_1\}}] = \int_{B_0}\int_{B_1} h(x_1)p(0, x_0; t_1, x_1)dx_1\mu_0(dx_0) \\ &= \int_{\mathbb{R}^d}\int_{B_2}\int_{B_1}\int_{B_0} f(x_3)p(t_2, x_2; t_3, x_3)p(t_1, x_1; t_2, x_2)p(0, x_0; t_1, x_1)\mu_0(dx_0)dx_1dx_2dx_3,\end{aligned}$$

where μ_0 denotes the distribution of μ_0. Repeating this argument, we find that for $0 < t_1 < \cdots < t_n \le T$ the joint distribution of $(X_0, X_{t_1}, \ldots, X_{t_n})$ is given by

$$\mathbb{P}(X_0 \in B_0, X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n) = \int_{B_n} \cdots \int_{B_1} \int_{B_0} \prod_{i=1}^{n} p(t_{i-1}, x_{i-1}; t_i, x_i) \mu_0(dx_0) dx_1 \cdots dx_n$$

for $B_0, B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R}^d)$.

Remark 3.24. A set of conditions sufficient for which $\{X_t\}$ has a transition density is, in addition to the Lipschitz continuity of b and σ,

(i) the uniform ellipticity: there exists a positive constant c such that

$$|\sigma(t,x)^{\mathsf{T}}\xi|^2 \ge c|\xi|^2, \quad \xi \in \mathbb{R}^d, \;\; (t,x) \in [0,T] \times \mathbb{R}^d;$$

(ii) the boundedness: the functions b and σ are bounded on $[0,T] \times \mathbb{R}^d$.

In general, the transition probability density p of $\{X_t\}$ can be seen as the *fundamental solution* of the corresponding PDE. Indeed, under suitable conditions,

$$u(t,x) = \int_{\mathbb{R}^d} p(t,x,T,y) g(y) dy, \quad t \in [0,T], \;\; x \in \mathbb{R}^d,$$

turns out to be a classical solution of the PDE (3.5.6). We refer to [21, Chpater 5] for details.

3.6 Statistical Inference

In this section, we discuss estimation methods for the drift and diffusion coefficients in SDEs with observed data. We refer to Prakasa Rao [33], Iacus [16] and the references therein for more details.

Maximum Likelihood Estimation

Consider the following *parametrized* SDE:

$$dX_t = b(X_t, \theta)dt + \sigma(X_t, \theta)dW_t, \quad X_0 = x_0, \tag{3.6.1}$$

where $\{W_t\}_{t \ge 0}$ is a one-dimensional Brownian motion and x_0 is a given constant. $\theta \in \mathbb{R}^p$ denotes some parameters of this system, and θ belongs to some parameter space $\Theta \subset \mathbb{R}^p$. We assume that (3.6.1) admits a unique solution and do not impose explicit conditions on the coefficients $b : \mathbb{R} \times \Theta \to \mathbb{R}$ and $\sigma : \mathbb{R} \times \Theta \to (0, \infty)$. Moreover, we assume here that there exists the transition density $p_\theta(t, y; s, x)$ of $\{X_t\}$.

Suppose that sample X_i is observed at time $t_i = i\Delta$, $i = 1, \ldots, n$, where $\Delta \equiv \Delta_{n,T} = T/n$. Denote by θ_0 a true parameter of the system to be estimated. The *maximum likelihood estimation* (MLE) is an estimation method based on the hypothesis "most likely data are observed". Namely, MLE adopts parameters that maximize some *likelihood function*. In general, for the sample $Y_1, \ldots, Y_n$, the likelihood function is defined by the joint density of $Y_1, \ldots, Y_n$ as a function of θ. For example, let Y a random variable with density $p(x, \theta_0)$, and consider the estimation problem of the parameter θ_0 from an IID sample $Y_1, \ldots, Y_n$. Then, by the independence, the joint density is given by the products of p's. More precisely, the likelihood function $L(\theta)$ here is given by

$$L(\theta) = \prod_{i=1}^{n} p(Y_i, \theta).$$

As an estimated parameter, we adopt a local maximizer of the logarithm of the likelihood function, i.e., a solution θ of the equation

$$\frac{\partial}{\partial\theta}\log L(\theta) = 0$$

is adopted as an *estimator*.

In case of SDEs, as seen in Section 3.3, the finite dimensional distribution can be described by the transition density. Thus we adopt it as the likelihood function and a maximizer θ of

$$L(\theta) = \prod_{j=1}^{n} p_\theta(j\Delta, X_j; (j-1)\Delta, X_{j-1})$$

as an estimator of θ_0.

Example 3.25. Consider the following Ornstein-Uhlenbeck process

$$dX_t = -bX_t dt + \sigma dW_t.$$

Recall from Example 2.19 that the unique solution with initial condition $X_t = x$ is given by

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

Since $X_s^{t,x}$ follows a Gaussian distribution with mean $m(s-t,x) := xe^{-b(s-t)}$ and variance $v(s-t,x) := \sigma^2(1-e^{-2b(s-t)})/(2b)$, the transition probability p_θ with $\theta = (b, \sigma)$ is given by

$$p_\theta(t,x;s,y) = \frac{\exp\left(-(y-m(s-t,x))^2/(2v(s-t,x))\right)}{\sqrt{2\pi v(s-t,x)}}.$$

Hence

$$\begin{aligned}
\log L(\theta) &= \sum_{j=1}^{n} \log p_\theta(t_j, X_j; t_{j-1}, X_{j-1}) \\
&= \sum_{j=1}^{n} \left[-\frac{(X_j - m(\Delta, X_{j-1}))^2}{2v(\Delta, X_{j-1})} - \frac{1}{2}\log(2\pi v(\Delta, X_{j-1})) \right].
\end{aligned}$$

Therefore, the maximum likelihood estimator $\hat{b}$ for b is approximately given by

$$\hat{b} \approx -\frac{1}{\Delta}\log\left(\frac{\sum_{j=1}^{n} X_{j-1}X_j}{\sum_{j=1}^{n} X_{j-1}^2}\right).$$

Note that this quantity can be defined only when $\sum_{j=1}^{n} X_{j-1}X_j > 0$. Under this condition, it is straightforward to see that the maximum likelihood estimator $\hat{\sigma}$ for σ is given by

$$\hat{\sigma} = \sqrt{\frac{2\hat{b}}{n(1-e^{-2\hat{b}\Delta})}\sum_{j=1}^{n}(X_j - X_{j-1}e^{-\hat{b}\Delta})^2}.$$

Example 3.26. Consider the geometric Brownian motion

$$dX_t = bX_t dt + \sigma X_t dW_t,$$

where $b \in \mathbb{R}$ and $\sigma \geq 0$. As seen in Section 3.1 in Chapter 3, we have

$$X_s^{t,x} = x\exp\left((b-\sigma^2/2)(s-t) + \sigma(W_s - W_t)\right), \quad s \geq t, \quad x > 0,$$

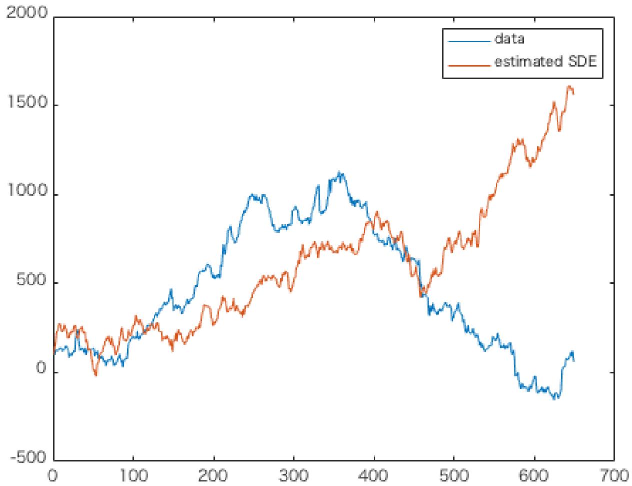

Figure 3.6.1: The difference of the stock prices of Tokyu Corp. and Keikyu Corp. from 2016/1/4 to 2018/9/4 (blue line), and a sample path of the Ornstein-Uhlenbeck process with estimated parameter $\hat{b} = 0.2111$ and $\hat{\sigma} = 372.6866$ (red line).

whence

$$\mathbb{P}(X_{t+\Delta}^{t,x} \leq y) = \mathbb{P}((b - \sigma^2/2)\Delta + \sigma W_\Delta \leq \log y - \log x).$$

Thus the transition density p_θ is given by

$$p_\theta(t+\Delta, y; t, x) = \frac{1}{\sigma y \sqrt{2\pi\Delta}} \exp\left(-\frac{1}{2} \left(\frac{\log y - \log x - (b - \sigma^2/2)\Delta}{\sigma\sqrt{\Delta}} \right)^2 \right).$$

Hence,

$$\log L(\theta) = -\sum_{j=1}^{n} \left\{ \frac{1}{2} \left(\frac{\log X_j - \log X_{j-1} - (b - \sigma^2/2)\Delta}{\sigma\sqrt{\Delta}} \right)^2 - \log(\sigma X_j \sqrt{2\pi\Delta}) \right\}.$$

Unfortunately, the transition probability density for diffusion processes are rarely available. One of approximation methods for the likelihood functions is to apply the Euler-Maruyama approximation

$$X_{t+\Delta} - X_t = b(X_t, \theta)\Delta + \sigma(X_t, \theta)(W_{t+\Delta} - W_t)$$

to (3.6.1). The right-hand side in the equation just above follows a (conditional) Gaussian distribution with mean $b(X_t, \theta)\Delta$ and $\sigma(X_t, \theta)^2$. Thus, the transition density p_θ is approximated with

$$\tilde{p}_\theta(t+\Delta, y; t, x) := \frac{1}{\sqrt{2\pi\Delta\sigma^2(x, \theta)}} \exp\left\{ -\frac{1}{2} \frac{(y - x - b(x, \theta)\Delta)^2}{\Delta\sigma^2(x, \theta)} \right\}.$$

Now, we will present a consistency result for the pseudo-likelihood methods. To this end, we restrict ourselves to the case where the SDEs are described by

$$dX_t = b(X_t, \theta)dt + \sigma dW_t, \tag{3.6.2}$$

where $\theta \in \Theta$ is as in above and $\sigma > 0$ is also a unknown parameter independent of θ. Then, the maximization of $L(\theta)$ is equivalent to the least-squares problem

$$L_1(\theta) = \sum_{j=1}^{n} (X_j - X_{j-1} - b(X_{j-1}, \theta)\Delta)^2.$$

We denote by $\hat{\theta}$ its estimator, i.e.,

$$\hat{\theta} = \arg\min_{\theta \in \Theta} L_1(\theta).$$

With this $\hat{\theta}$, we adopt

$$\hat{\sigma}^2 := \frac{1}{n\Delta} \sum_{j=1}^{n} (X_j - X_{j-1} - b(X_{j-1}, \hat{\theta})\Delta)^2$$

as an estimator for σ^2.

To prove the consistency of the estimators above, we assume that

$$\begin{aligned} &\int_0^x \exp\left\{-\frac{2}{\sigma^2}\int_0^y b(z)dz\right\} dy \to \pm\infty, \quad \text{as } x \to \pm\infty, \\ &c := \int_{-\infty}^{\infty} \exp\left\{\frac{2}{\sigma^2}\int_0^x b(z)dz\right\} dx < \infty. \end{aligned} \tag{3.6.3}$$

Then, it is known that $\{X_t\}_{t\geq 0}$ is *ergodic* with *invariant measure* ν defined by

$$\frac{d\nu}{dx} = \frac{1}{c}\exp\left\{\frac{2}{\sigma^2}\int_0^x b(z)dz\right\}$$

for $\theta = \theta_0$, i.e., for any Borel measurable function h on $\mathbb{R}$ that is integrable with respect to ν,

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T h(X_t)dt = \int_{-\infty}^{\infty} h(x)\nu(dx), \quad \text{a.s.}$$

Moreover, we assume that the following conditions are satisfied:

Assumption 3.27

(i) There exists a unique solution $\{X_t\}_{t\geq 0}$ of (3.6.2) satisfying $\sup_{t\geq 0} \mathbb{E}|X_t|^p < \infty$ for every $p \geq 1$.

(ii) There exist a positive constant C_0 and q such that for any $x \in \mathbb{R}$ and $\theta \in \Theta$,

$$\begin{aligned} |b(x,\theta)| &\leq C_0(1+|x|^q), \\ |b(x,\theta) - b(y,\theta)| &\leq C_0|x-y|. \end{aligned}$$

(iii) The function $b(x,\cdot) \in C^2(\Theta)$ for any $x \in \mathbb{R}$ and

$$|\partial_{\theta_i} b(x,\theta)| + |\partial^2_{\theta_i\theta_j} b(x,\theta)| \leq C_1(1+|x|^{q_1}), \quad x \in \mathbb{R},$$

for some constants $C_1, q_1 > 0$.

(iv) The function

$$\int_{\mathbb{R}} b(\theta,x)\left\{b(\theta_0,x) - \frac{1}{2}b(\theta,x)\right\}\nu(dx)$$

has a unique maximum at $\theta = \theta_0$ in Θ.

(v) The functions b and $\partial_{\theta_i} b$, $i = 1,\ldots,p$, are smooth in x and their derivatives are of polynomial growth in x uniformly in $\theta \in \Theta$.

(vi) The matrix

$$\Phi = \int_{\mathbb{R}} D_\theta b(\theta_0,x)^\mathsf{T} D_\theta b(\theta_0,x)\nu(dx)$$

is positive definite.

Under the complicated conditions in Assumption 3.27, we can show the consistency of $\hat{\sigma}$ and $\hat{\theta}$. More precisely, we have the following result:

Theorem 3.28

Suppose that (3.6.3) and Assumption 3.27 holds. Then,

$$(\sqrt{n}(\hat{\sigma}-\sigma_0), \sqrt{T}(\hat{\theta}-\theta_0)) \longrightarrow N(0,H)$$

in distribution, provided that $n, T \to \infty$, $\Delta_{n,T} \to 0$, and $(\Delta_{n,T})^3 n = o(1)$, where

$$H = \begin{pmatrix} \sigma_0^2 & 0 \\ 0 & 1 \end{pmatrix}.$$

For a proof we refer to Yoshida [41] (see also Section 3.4 in [33]).

Nonparametric estimation

Let D be a domain in $\mathbb{R}^d$. Here we consider a nonparametric estimation for the D-valued SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t. \tag{3.6.4}$$

Namely, we consider the problem of estimating the functions b and σ from observed data. Accordingly, we assume that b and σ are Lipschitz continuous so that (3.6.4) has a unique solution $\{X_t\}_{t\geq 0}$. Moreover, assume that we observe X_i at time $t_i = i\Delta$, $i = 1, \dots, n$, where $\Delta \equiv \Delta_{n,T} = T/n$.

Put $a(x) = \sigma(x)\sigma(x)^{\mathsf{T}}$, $x \in D$. By (3.5.2), the functions b and a can be represented as

$$\begin{aligned} \mathbb{E}[X^{t,x}_{t+\Delta t} - x] &= \Delta t b(x) + o(\Delta t), \\ \mathbb{E}[(X^{t,x}_{t+\Delta t} - x)(X^{t,x}_{t+\Delta t} - x)^{\mathsf{T}}] &= \Delta t a(t) + o(\Delta t). \end{aligned} \tag{3.6.5}$$

By (3.6.5), formally we have

$$\begin{aligned} b(x) &\simeq \frac{1}{\Delta}\mathbb{E}[X_{t+\Delta} - x | X_t = x], \\ a(x) &\simeq \frac{1}{\Delta}\mathbb{E}[(X_{t+\Delta} - x)(X_{t+\Delta} - x)^{\mathsf{T}} | X_t = x]. \end{aligned}$$

Thus, by kernel regression, the functions

$$\begin{aligned} \hat{b}(x) &= \frac{\sum_{i=1}^{n-1} K((X_i - x)/h)(X_{i+1} - X_i)}{\Delta \sum_{i=1}^{n} K((X_i - x)/h)}, \\ \hat{a}(x) &= \frac{\sum_{i=1}^{n-1} K((X_i - x)/h)(X_{i+1} - X_i)(X_{i+1} - X_i)^{\mathsf{T}}}{\Delta \sum_{i=1}^{n} K((X_i - x)/h)} \end{aligned}$$

are adopted as estimators for $b(x)$ and $a(x)$, respectively. Here, K is a nonnegative function on $\mathbb{R}^d$, called a *kernel*, and a parameter $h \equiv h_{n,T} > 0$ determines the smoothness of the estimators. For examples, the function K can be

- the naive kernel: $K(x) = 1_{\{|x|\leq 1\}}$;
- the quadratic kernel: $K(x) = (1 - |x|^2)_+$;
- the Gaussian kernel: $K(x) = e^{-|x|^2}$.

We refer to, e.g., Györfi et.al [13] for the theory of nonparametric estimation of the conditional expectations.

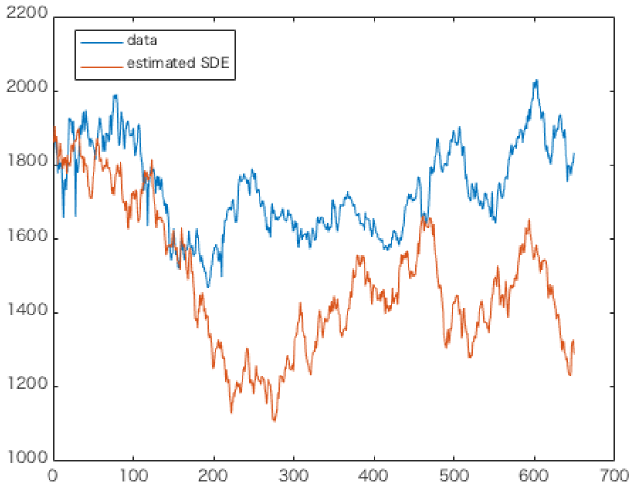

Figure 3.6.2: The stock prices of Tokyu Corp. from 2016/1/4 to 2018/9/4 (blue line), and a sample path of the SDE estimated by the kernel regression (red line). The quadratic kernel $K(x) = (1 - |x|^2)_+$ with $h = 0.8$ is used.

Problem 3.29. Perform the kernel-based estimation above using simulated paths from a geometric Brownian motion as the sample data. Observe how different the original model and the estimated one are.

Now let us see the theoretical side. To guarantee the consistency of the estimators, we impose the following conditions on the coefficients of the SDE to be estimated:

Assumption 3.30

(i) There exists a positive constant C_0 such that

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \le C_0|x - y|, \quad x, y \in D.$$

(ii) For every open and bounded set $A \subset D$,

$$\min_{x \in \overline{A}} a_{ii}(x) > 0$$

for some $i \in \{1, \dots, d\}$, where $\overline{A}$ is the closure of A.

(iii) There exists a function $\varphi : \mathbb{R}^d \setminus \{0\} \to \mathbb{R}$ of the class C^2 such that

$$b(x)^\mathsf{T} D\varphi(x) + \frac{1}{2}\mathrm{tr}(a(x)D^2\varphi(x)) \le 0, \quad x \in \mathbb{R}^d \setminus \{0\},$$

and that the function $r \mapsto \min_{|x|=r} \varphi(x)$ is strictly increasing and diverges to infinity as $r \to \infty$.

It is known that, under Assumption 3.30, there exists a σ-finite measure ν on $(D, \mathcal{B}(D))$ such that

$$\nu(A) = \int_D \mathbb{P}(X_t^{0,x} \in A)\nu(dx), \quad A \in \mathcal{B}(D). \tag{3.6.6}$$

We restrict ourselves to the case where the kernel K is of the form $K(x) = \prod_{i=1}^d \rho(x_i)$ for $x = (x_1, \dots, x_d)^\mathsf{T} \in \mathbb{R}^d$. Moreover, we make the following conditions on ρ:

Assumption 3.31

(i) The function ρ is nonnegative, bounded, continuous, symmetric function on $\mathbb{R}$ with $\int_{\mathbb{R}} \rho(s)ds = 1$, $\int_{\mathbb{R}} \rho^2(s)ds < \infty$, and $\int_{\mathbb{R}} s^2\rho(s)ds < \infty$.

(ii) There exists a nonnegative function H on $\mathbb{R}^d \times (0, \infty)$ such that

$$|K(x) - K(\xi)| \le H(\xi, \varepsilon)|x - \xi|$$

for $x, \xi \in \mathbb{R}^d$ satisfying $|x - \xi| < \varepsilon$ and that

$$\lim_{\varepsilon \to 0} \int_{\mathbb{R}^d} H(\xi, \varepsilon)d\xi < \infty, \qquad \int_D H(\xi, \varepsilon)\nu(d\xi) < \infty$$

for any $\varepsilon > 0$.

Further, we introduce the quantity

$$\hat{L}_{n,T}(T, x) = \Delta \sum_{i=1}^{n} K_h(X_{i\Delta} - x), \quad x \in D,$$

and impose the following conditions on this and the other parameters:

Assumption 3.32

When $n, T \to \infty$, we have $\Delta_{n,T} \to 0$, $h_{n,T} \to 0$, and

$$\hat{L}_{n,T}(T, x) \to 0, \quad (\Delta_{n,T} \log(1/\Delta_{n,T}))^{1/2}\, h_{n,T}^{-d} \to 0, \quad \text{a.s.},$$

for any $x \in D$.

Under the assumptions above, we have the following consistency results:

Theorem 3.33

Suppose that Assumptions 3.30–3.32 hold. Then, for any $x \in D$, we have

$$\hat{b}_{n,T}(x) \to b(x), \quad \hat{a}_{n,T}(x) \to a(x), \quad \text{a.s.},$$

as $n, T \to \infty$.

For a proof of this theorem we refer to Bandi and Moloche [1], where the asymptotic normality of the estimators are also obtained under additional conditions.

3.7 Weak Solutions

Here we introduce the notion of *weak solutions* of SDEs, which differ from solutions of SDEs appeared in previous sections in that a filtered probability space and a Brownian motion are parts of the solution.

Let $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d \times m}$ be Borel measurable.

Definition 3.34. A 6-tuple $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, W, X)$ is said to be a *weak solution* of (3.1.3) if

(i) $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space with filtration $\mathbb{F} = \{\mathcal{F}_t\}_{0 \le t \le T}$ satisfying the usual conditions;

(ii) $W = \{W_t\}_{0 \le t \le T}$ is an m-dimensional $\mathbb{F}$-Brownian motion and $X = \{X_t\}_{0 \le t \le T}$ is a d-dimensional process defined on $(\Omega, \mathcal{F}, \mathbb{P})$;

(iii) X is a solution of (3.1.3) in the sense of Definition 3.1, where W is the given Brownian motion.

Solution of SDEs where a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F})$ and a Brownian motion W are fixed a priori, i.e., solutions introduced in Section 3.2, are actually called *strong solutions* for SDEs. By definition, a strong solution of (3.1.3) is a weak solution of (3.1.3). The notion of weak solutions is often natural in application since in many cases of modeling we cannot specify a probability space and Brownian motion a priori, and is even useful in theory since we can show the existence of solutions under weaker conditions on the drift term $b(t, x)$.

We say that the weak solution of (3.1.3) is *unique in the sense of probability law* if any two weak solutions $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, W, X)$ and $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{F}}, \tilde{\mathbb{P}}, \tilde{W}, \tilde{X})$ of (3.1.3) with

$$\mathbb{P}(X_0 \in A) = \tilde{\mathbb{P}}(X_0 \in A), \quad A \in \mathcal{B}(\mathbb{R}^d),$$

we have

$$\mathbb{P}(X \in \Gamma) = \tilde{\mathbb{P}}(X \in \Gamma), \quad \Gamma \in \mathcal{B}(\mathbb{W}^d).$$

Thus, by Proposition 3.23, the uniqueness in this sense holds if two solutions have the same finite dimensional distributions.

Example 3.35. Consider the one-dimensional SDE

$$dX_t = \mathrm{sgn}(X_t) dW_t, \quad X_0 = 0, \tag{3.7.1}$$

where $\mathrm{sgn}(x) = 1$ for $x > 0$ and $= -1$ for $x \le 0$. Let us see that this SDE has a weak solution but does not admit a strong solution. Let $\{X_t\}$ be a one dimensional Brownian motion on a given $(\Omega, \mathcal{F}, \mathbb{P})$. Then,

$$W_t := \int_0^t \mathrm{sgn}(X_s) dX_s$$

is a martingale with respect to $\mathbb{F} = \{\mathcal{F}_t\}$, the augmented natural filtration generated by $\{X_t\}$. Since $\{W_t\}$ is an Itô process,

$$\int_0^t \mathrm{sgn}(X_s) dW_s = \int_0^t \mathrm{sgn}(X_s)^2 dX_s = X_t.$$

Thus $\{X_t\}$ and $\{W_t\}$ satisfy (3.7.1). Observe $dW_t dW_t = dt$ and so by Itô's formula,

$$\mathbb{E}[e^{i\xi(W_t - W_s)} | \mathcal{F}_s] = 1 - \frac{\xi^2}{2} \int_s^t \mathbb{E}[e^{i\xi(W_u - W_s)} | \mathcal{F}_s] du,$$

for $0 \le s \le t \le T$ and $\xi \in \mathbb{R}$, where i denotes the imaginary unit. Solving this equation, we obtain

$$\mathbb{E}[e^{i\xi(W_t - W_s)} | \mathcal{F}_s] = e^{-\xi^2(t-s)/2},$$

whence $\{W_t\}$ is a $\{\mathcal{F}_t\}$-Brownian motion. Therefore $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, W, X)$ is a weak solution of (3.7.1).

On the other hand, suppose that $\{X_t\}$ satisfies (3.7.1) on a given filtered space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ and for a given $\mathbb{F}$-Brownian motion $\{W_t\}$. This means in particular that

$$\sigma(X_s : 0 \le s \le t) \subset \sigma(\sigma(W_s : 0 \le s \le t) \cup \mathcal{N}) \tag{3.7.2}$$

where $\mathcal{N}$ denotes the $\mathbb{P}$-null sets from $\mathcal{F}$. Then, the above arguments shows that $\{X_t\}$ is necessarily a Brownian motion and

$$W_t = \int_0^t \mathrm{sgn}(X_s) dX_s.$$

Applying Tanaka's formula (see, e.g., [31, Chapter 4]) for the right-hand side, we have

$$W_t = |X_t| - \lim_{\varepsilon \searrow 0} \frac{1}{2\varepsilon} \mathrm{Leb}(0 \le s \le t : |X_s| \le \varepsilon), \quad 0 \le t \le T, \ \mathbb{P}\text{-a,s.},$$

where Leb denotes the Lebesgue measure on $[0, T]$. This leads to

$$\sigma(W_s : 0 \le s \le t) \subset \sigma(\sigma(|X_s| : 0 \le s \le t) \cup \mathcal{N}) \subsetneq \sigma(\sigma(X_s : 0 \le s \le t) \cup \mathcal{N}),$$

contradicting to (3.7.2).

Using Girsanov's theorem (see Section 2.3), we can obtain the existence and uniqueness of weak solutions for SDEs with *measurable* drift. Namely, we can remove the continuity condition for the drift coefficients in the framework of weak solutions.

Suppose that σ is an $\mathbb{R}^{d\times d}$-valued function on $[0, T] \times \mathbb{R}^d$ satisfying the following condition: the inverse $\sigma^{-1}(t, x)$ exists for all $(t, x) \in [0, T] \times \mathbb{R}^d$; σ^{-1} is bounded on $[0, T] \times \mathbb{R}^d$; the Lipschitz continuity condition imposed in Theorem 3.2 holds. Further, let $\{W_t\}$ be a d-dimensional $\mathbb{F} = \{\mathcal{F}_t\}$-Brownian motion on a given $(\Omega, \mathcal{F}, \mathbb{P})$. Then there exists a unique strong solution $\{X_t\}_{0\le t\le T}$ of the SDE

$$dX_t = \sigma(t, X_t) dW_t, \quad X_0 = \xi$$

for a given $\xi \in L^2(\Omega, \mathcal{F}_0, \mathbb{P})$. Let $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ be bounded and Borel mesurable. Then, by Girsanov's theorem (Theorem 2.23),

$$B_t := W_t - \int_0^t (\sigma^{-1} b)(s, X_s) ds, \quad 0 \le t \le T,$$

is a d-dimensional Brownian motion under the probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ defined by

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left[\int_0^T (\sigma^{-1} b)(t, X_t) dW_t - \frac{1}{2} \int_0^T |(\sigma^{-1} b)(t, X_t)|^2 dt\right].$$

Since $\{W_t\}$ and $\{B_t\}$ are Itô-processes under both $\mathbb{P}$ and $\mathbb{Q}$, we have

$$\int_0^t \sigma(X_s) dB_s = \int_0^t \sigma(X_s) dW_s - \int_0^t b(s, X_s) ds, \quad 0 \le t \le T, \ \mathbb{P} \text{ and } \mathbb{Q}\text{-a.s.}$$

Thus, $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{Q}, \{B_t\}, \{X_t\})$ is a weak solution of (3.1.3).

Theorem 3.36

Let $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ be bounded and Borel measurable, and $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$ satisfies the following condition: the inverse $\sigma^{-1}(t, x)$ exists for all $(t, x) \in [0, T] \times \mathbb{R}^d$; σ^{-1} is bounded on $[0, T] \times \mathbb{R}^d$; the Lipschitz continuity condition imposed in Theorem 3.2 holds. Then (3.1.3) admits a weak solution that is unique in the sense of probability law.

Proof.* The existence is proved by the argument above. To show the uniqueness, let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, W, X)$ be a weak solution (3.1.3) with initial distribution

$$\mu_0(A) := \mathbb{P}(X_0 \in A), \quad A \in \mathcal{B}(\mathbb{R}^d),$$

Define the probability measures $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ by

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left[-\int_0^T (\sigma^{-1} b)(t, X_s) dW_s - \frac{1}{2} \int_0^T |(\sigma^{-1} b)(t, X_t)|^2 dt\right].$$

Then,

$$B_t := W_t + \int_0^t (\sigma^{-1}b)(s, X_s)ds, \quad 0 \le t \le T,$$

is an $\mathbb{F}$-Brownian motion under $\mathbb{Q}$. The process $\{X_t\}$ is a unique strong solution of

$$dX_t = \sigma(t, X_t)dB_t$$

under $\mathbb{Q}$, whence $\sigma(X_s : 0 \le s \le t) \subset \mathcal{G}_t := \sigma(X_0, \{B_s\}_{s \le t}, \mathcal{N})$, where $\mathcal{N}$ denotes the collections of $\mathbb{Q}$-null sets. Thus, $\{(\sigma^{-1}b)(t, X_t)\}$ is a $\{\mathcal{G}_t\}$-progressively measurable and so is $\{W_t\}$. Since the integral $\int_0^t (\sigma^{-1}b)(s, X_s)dW_s$ is an L^2-limit of some $\mathcal{G}_t$-measurable random variables. Therefore, $X = F(X_0, \{B_t\}_{0 \le t \le T})$ and $d\mathbb{Q}/d\mathbb{P} = G(X_0, \{B_s\}_{0 \le s \le T})$ a.s. for some measurable function F and G on $(\mathbb{R}^d \times C([0,T]; \mathbb{R}^d), \mathcal{B}(\mathbb{R}^d) \times \mathcal{B}(\mathbb{W}^d))$, respectively. This means that for $\Gamma \in \mathcal{B}(\mathbb{W}^d)$

$$\mathbb{P}(X \in \Gamma) = \mathbb{E}_{\mathbb{Q}}\left[1_{\{F(X_0, \{B_t\}_{0 \le t \le T}) \in \Gamma\}} G(X_0, \{B_s\}_{0 \le s \le T})\right] = \int_{\{F \in \Gamma\}} G(x, \xi)\mu_0(dx)\mu_W(d\xi),$$

where μ_W denotes the Wiener measure. It is clear that F and G do not depend on a particular choice of weak solution, whence the uniqueness in the sense of probability law follows. □

3.8 Time Reversal

Consider a solution $\{X_t\}$ of the SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \tag{3.8.1}$$

and the process $\overline{X}_t := X_{T-t}$. Our aim here is to find a *reverse-time SDE*

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

for $\overline{X}_t$. More precisely, we aim to prove that $\{\overline{X}_t\}$ is a weak solution of the SDE above for appropriate $\overline{b}$ and $\overline{\sigma}$. Of course we want to give explicit representations for these functions.

In this section, we assume that $\{W_t\}_{0 \le t \le T}$ is a d-dimensional $\mathbb{F}$-Brownian motion. Let ξ be an $\mathbb{R}^d$-valued $\mathcal{F}_0$-measurable random variable such that

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

Let $\overline{\mathbb{F}} = \{\overline{\mathcal{F}}_t\}_{0 \le t \le T}$ be the augmented natural filtration generated by $\overline{X}$, i.e., $\overline{\mathcal{F}}_t = \sigma(\sigma(\overline{X}_s; s \le t) \cup \mathcal{N})$, $0 \le t \le T$.

Assumption 3.37

The functions b and σ satisfy the following:

(i) The inverse matrix $\sigma^{-1}(t, x)$ exists for any $t \in [0, T]$ and $x \in \mathbb{R}^d$.

(ii) There exists a positive constant C such that for $\varphi = b^i, \sigma^{ij}, (\sigma^{-1})^{ij}$, $i, j = 1, \ldots, d$, $t \in [0, T]$, $x, y \in \mathbb{R}^d$,

$$|\varphi(t, x) - \varphi(t, y)| \le C|x - y|,$$
$$|b^i(t, x)| \le C(1 + |x|).$$

(iii) $\sigma \in C^{1,2}([0, T] \times \mathbb{R}^d)$.

Under Assumption 3.37, by Theorem 3.2 there exists a unique strong solution $X = \{X_t\}_{0 \le t \le T}$ of (3.8.1) with $X_0 = \xi$.

Assumption 3.38

The distribution of ξ has a continuous density ρ_0. Moreover, There exists a transition density $p(t,x,s,y)$ of X that is everywhere positive such that the following hold:

(i) for any $y \in \mathbb{R}^d$ the function $(t,x) \mapsto p(t,x,T,y)$ is in $C^{1,2}([0,T) \times \mathbb{R}^d)$;

(ii) for any $t < T$ and $x \in \mathbb{R}^d$, the functions $\partial_t p(t,x,T,y)$, $\partial_{x_i} p(t,x,T,y)$, $\partial^2_{x_i x_j} p(t,x,T,y)$, $i,j = 1,\dots,d$, are all continuous in y on $\mathbb{R}^d$;

Under Assumption 3.38, the density $p(t,x)$ of X_t exists and is given by

$$p(t,x) = \int_{\mathbb{R}^d} p(0,y,t,x)\rho_0(y)dy, \quad 0 \le t \le T, \;\; x \in \mathbb{R}^d.$$

Put $a = (a^{ij})_{1 \le i,j \le d} = \sigma\sigma^{\mathsf{T}}$. We further make the following assumption:

Assumption 3.39

For any $t \le T$ the function $p(t,\cdot)$ is $C^1(\mathbb{R}^d)$ and satisfies

$$\int_0^T \int_{\mathbb{R}^d} |\partial_{y_i}(a^{ij}(t,y)p(t,y))| dy dt < \infty, \quad i,j = 1,\dots,d.$$

Introduce another drift function $\tilde{b} = (\tilde{b}^1, \cdots, \tilde{b}^d)$ defined by

$$\tilde{b}^i(t,x) = b^i(t,x) - \frac{1}{p(t,x)} \sum_{j=1}^d \partial_{x_j}\left(a^{ij}(t,x)p(t,x)\right).$$

Further, define $\overline{b}(t,x) = (\overline{b}^1(t,x), \cdots, \overline{b}^d(t,x))$ and $\overline{\sigma}(t,x) = (\overline{\sigma}^{ij}(t,x))_{1 \le i,j \le d}$ by

$$\begin{aligned} \overline{b}^i(t,x) &= -b^i(T-t,x) + \frac{1}{p(T-t,x)} \sum_{j=1}^d \partial_{x_j}\left(a^{ij}(T-t,x)p(T-t,x)\right), \\ \overline{\sigma}(t,x) &= \sigma(T-t,x) \end{aligned}$$

for $i = 1,\dots,d$ and $(t,x) \in [0,T] \times \mathbb{R}^d$. Notice that by Assumption 3.39,

$$\int_0^T \mathbb{E}\left|(p^{-1}\partial_{x_i}(a^{ij}p))(r,X_r)\right| dr = \int_0^T \int_{\mathbb{R}^d} \left|\partial_{x_i}(a^{ij}p)(r,x)\right| dx dr < \infty.$$

This means

$$\int_0^T |\overline{b}(r,X_r)| dr < \infty, \quad \text{a.s.}$$

Here is a main result in this section.

Theorem 3.40

Suppose that Assumptions 3.37–3.39 hold. Then there exists a d-dimensional $\overline{\mathbb{F}}$-Brownian motion $\overline{W} = \{\overline{W}_t\}_{0\le t\le T}$ such that

$$\overline{X}_t = \overline{X}_0 + \int_0^t \overline{b}(s, \overline{X}_s)ds + \int_0^t \overline{\sigma}(s, \overline{X}_s)d\overline{W}_s.$$

Moreover, the process $\{X_t\}$ is represented as

$$X_T = X_t + \int_t^T \tilde{b}(s, X_s)ds + \int_t^T \sigma(s, X_s)\overleftarrow{d}B_s,$$

where $B_t = (B_t^1, \dots, B_t^d)$ is given by

$$B_t^i = W_t^i + \int_0^t \frac{1}{p(s, X_s)} \sum_{j=1}^d \partial_{x_j}\left(\sigma^{ij}(s, X_s)p(s, X_s)\right) ds, \quad i = 1, \dots, d$$

and is a d-dimensional Brownian motion such that $B_t - B_s$ is independent of $\sigma(X_u; u \ge t)$ for any $t > s$.

Remark 3.41. Actually, the assumptions imposed in Theorem 3.40 can be slightly weakened. More general analysis can be found in Haussmann and Pardoux [15].

The rest of this section is devoted to a proof of Theorem 3.40. From now on, we suppose Assumptions 3.37–3.39 always hold. For simplicity we shall assume that $T = 1$ and $d = 1$. Thus we suppress the superscript that used for d-dimensional vectors and $d \times d$-matrices. E.g., we write $b(t, x)$ for $b^1(t, x)$.

We start by proving the Markovian property of $\overline{X}$.

Proposition 3.42

The process $\{\overline{X}_t\}_{0\le t\le 1}$ is Markov.

Proof. Let $t > s$. As a generalization of the Markov property of $\{X_t\}$, we actually have

$$\mathbb{E}[G(X_{\cdot\vee(1-s)})|\mathcal{F}_{1-s}] = \mathbb{E}[G(X_{\cdot\vee(1-s)}|X_{1-s}] =: g(X_{1-s})$$

for any bounded Borel measurable function G on $\mathbb{W}^d$. See, e.g. [31, Chapter 7]. Thus, for any bounded Borel function f on $\mathbb{R}^d$,

$$\begin{aligned}\mathbb{E}[f(X_{1-t})G(X_{\cdot\vee(1-s)})] &= \mathbb{E}[f(X_{1-t})g(X_{1-s})] = \mathbb{E}\left[\mathbb{E}[f(X_{1-t})|X_{1-s}]g(X_{1-s})\right] \\ &= \mathbb{E}\left[\mathbb{E}[f(X_{1-t})|X_{1-s}]G(X_{\cdot\vee(1-s)})\right].\end{aligned}$$

This leads to

$$\mathbb{E}[f(\overline{X}_t)1_A] = \mathbb{E}[\mathbb{E}[f(\overline{X}_t)|\overline{X}_s]1_A], \quad A \in \sigma(X_u; u \ge 1-s),$$

whence the proposition follows. ☐

The following is a key result:

Lemma 3.43

Let $f \in C_c^\infty(\mathbb{R})$. For $t > s$ we have

$$\mathbb{E}[f(\overline{X}_t) - f(\overline{X}_s)|\overline{X}_s] = \int_s^t \mathbb{E}[\overline{\mathcal{L}}_r f(\overline{X}_r)|\overline{X}_s]dr,$$

where

$$\overline{\mathcal{L}}_r = \bar{b}(r,x)f'(x) + \frac{1}{2}\overline{\sigma}^2(r,x)f''(x).$$

Proof. Let $t > s$ be fixed. Observe

$$\begin{aligned} f(\overline{X}_t) - f(\overline{X}_s) &= -(f(X_{1-s}) - f(X_{1-t})) \\ &= -\left[\int_{1-t}^{1-s} \mathcal{L}_r f(X_r)dr + \int_{1-t}^{1-s} (\sigma f')(r, X_r)dW_r\right]. \end{aligned}$$

where $\mathcal{L}_r f(x) = b(r,x)f'(x) + (1/2)\sigma^2(r,x)f''(x)$. Let $\phi \in C_c^\infty(\mathbb{R})$ be arbitrary. Consider the function $V(r,y) := \mathbb{E}[\phi(X_{1-s}^{r,y})]$ defined for $0 \le r \le 1-s$ and $y \in \mathbb{R}$. By Assumption 3.38, the function V is represented as

$$V(r,y) = \int_{\mathbb{R}^d} \phi(z)p(r,y,1-s,z)dz$$

and in $C^{1,2}([0,1-s) \times \mathbb{R})$. Since X is $\mathbb{F}$-Markov,

$$V_r := V(r, X_r) = \mathbb{E}[\phi(X_{1-s})|\mathcal{F}_r]$$

and so $\{V_r\}_{0\le r\le 1-s}$ is an $\mathbb{F}$-martingale.

Choose $\varepsilon > 0$ so that $1-t \le 1-s-\varepsilon$. Itô formula yields

$$V_{1-s-\varepsilon} = V_{1-t} + \int_{1-t}^{1-s-\varepsilon} (\partial_r + \mathcal{L}_r)V(r,X_r)dr + \int_{1-t}^{1-s-\varepsilon} (\partial_x V\sigma)(r,X_r)dW_r.$$

Applying Lemma 2.13, we get $(\partial_r + \mathcal{L}_r)V(r,X_r) = 0$, a.e., whence

$$V_r = V_{1-t} + \int_{1-t}^r (\partial_x V\sigma)(u,X_u)dW_u, \quad 0 \le r \le 1-s-\varepsilon.$$

The product Itô formula gives

$$df(X_r)V_r = V_r(\mathcal{L}_r f(X_r)dr + (\sigma f')(r,X_r)dW_r) + f(X_r)dV_r + (\partial_x V\sigma^2 f')(r,X_r)dr.$$

Since $f \in C_c^\infty(\mathbb{R})$, the Itô integral in the equality just above is a martingale. Thus,

$$\mathbb{E}[f(X_{1-s-\varepsilon})V_{1-s-\varepsilon}] = \mathbb{E}[f(X_{1-t})V_{1-t}] + \int_{1-t}^{1-s-\varepsilon} \mathbb{E}\left[V_r\mathcal{L}_r f(X_r) + (\partial_x V\sigma^2 f')(r,X_r)\right]dr.$$

The integration-by-parts formula yields

$$\begin{aligned} \mathbb{E}\left[(\partial_x V\sigma^2 f')(r,X_r)\right] &= -\int_{-\infty}^{\infty} V(r,x)\partial_x((\sigma^2 pf')(r,x))dx \\ &= -\int_{-\infty}^{\infty} V(r,x)\left(\partial_x(\sigma^2 p)f' + \sigma^2 pf''\right)(r,x)dx \\ &= -\mathbb{E}[(Vp^{-1}\partial_x(\sigma^2 p)f')(r,X_r)] - \mathbb{E}[(V\sigma^2 f'')(r,X_r)], \end{aligned}$$

whence

$$\mathbb{E}[f(X_{1-s-\varepsilon})V_{1-s-\varepsilon}] = \mathbb{E}[f(X_{1-t})V_{1-t}] - \int_{1-t}^{1-s-\varepsilon} \mathbb{E}\left[V_r\overline{\mathcal{L}}_{1-r}f(X_r)\right] dr.$$

Letting $\varepsilon \to 0$ and using the martingale property of V, we obtain

$$\mathbb{E}[f(X_{1-s})\phi(X_{1-s})] = \mathbb{E}\left[\left(f(X_{1-t}) - \int_{1-t}^{1-s} \overline{\mathcal{L}}_{1-r}f(X_r)dr\right)\phi(X_{1-s})\right],$$

Since ϕ is arbitrary, the lemma follows. □

Proof of Theorem 3.40. Step (i). Define the process $\{M_t\}$ by

$$M_t = \overline{X}_t - \overline{X}_0 - \int_0^t \overline{b}(r, \overline{X}_r)dr, \quad 0 \le t \le 1.$$

For $N \ge 1$, take $f_N, g_N \in C_c^\infty(\mathbb{R})$ such that $f_N(x) = x$ and $g_N(x) = x^2$ for $|x| \le N$. Consider the stopping time $\tau_N = \inf\{t \ge 0; |\overline{X}_t| > N\}$. By Proposition 3.42 and Lemma 3.43, for $t > s$ and $\varphi = f_N, g_N$,

$$\begin{aligned}
&\mathbb{E}\left[\varphi(\overline{X}_t) - \varphi(\overline{X}_s)\middle|\overline{\mathcal{F}}_s\right] - \int_s^t \mathbb{E}\left[\overline{\mathcal{L}}_r\varphi(\overline{X}_r)\middle|\overline{\mathcal{F}}_s\right] dr \\
&= \mathbb{E}\left[\overline{X}_t - \overline{X}_s\middle|\overline{X}_s\right] - \int_s^t \mathbb{E}\left[\overline{\mathcal{L}}_r\varphi(\overline{X}_r)\middle|\overline{X}_s\right] dr = 0.
\end{aligned}$$

whence $\{\varphi(\overline{X}_t)\}$ is a continuous $\overline{\mathbb{F}}$-martingale. Since $f_N(\overline{X}_{t\wedge\tau_N}) = \overline{X}_{t\wedge\tau_N}$ and $\overline{\mathcal{L}}_rf_N(\overline{X}_r) = \overline{b}(r, \overline{X}_r)$ for $r \le \tau_N$, the process $\{M_t\}$ is an $\overline{\mathbb{F}}$-local martingale.

Similarly, with the function g_N, by Proposition 3.42 and Lemma 3.43, we see that

$$\tilde{M}_t := \overline{X}_t^2 - \overline{X}_0^2 - \int_0^t \left(2\overline{b}(s, \overline{X}_s)\overline{X}_s + \overline{\sigma}^2(s, \overline{X}_s)\right) ds$$

is an $\overline{\mathbb{F}}$-local martingale.

By the definition of the quadratic variation (Definition 2.35),

$$\begin{aligned}
\tilde{M}_t + 2\int_0^t \overline{b}(s, \overline{X}_s)\overline{X}_sds &= \overline{X}_t^2 - \overline{X}_0^2 - \int_0^t \overline{\sigma}^2(s, \overline{X}_s)ds \\
&= 2\int_0^t \overline{X}_sd\overline{X}_s + \langle M\rangle_t - \int_0^t \overline{\sigma}^2(s, X_s)ds \\
&= 2\int_0^t \overline{b}(s, \overline{X}_s)\overline{X}_sds + 2\int_0^t \overline{X}_sdM_s + \langle M\rangle_t - \int_0^t \overline{\sigma}^2(s, X_s)ds.
\end{aligned}$$

From this,

$$\langle M\rangle_t - \int_0^t \overline{\sigma}^2(s, \overline{X}_s)ds = \tilde{M}_t - 2\int_0^t \overline{X}_sdM_s.$$

The right-hand side in the equality just above is a local martingale. So by the uniqueness of $\langle M\rangle$ (Theorem 2.28),

$$\langle M\rangle_t = \int_0^t \overline{\sigma}^2(s, \overline{X}_s)ds.$$

By Assumption 3.37, the process

$$\overline{W}_t := \int_0^t \overline{\sigma}^{-1}(s, \overline{X}_s)dM_s, \quad 0 \le t \le 1,$$

is in $\mathcal{M}_{loc}$ and satisfies

$$\langle W \rangle_t = \int_0^t \left(\overline{\sigma}^{-1}(s, \overline{X}_s)\right)^2 d\langle M \rangle_s = \int_0^t \left(\overline{\sigma}^{-1}(s, \overline{X}_s)\right)^2 \overline{\sigma}(s, \overline{X}_s)^2 ds = t.$$

So we can apply Lévy's theorem (Theorem 2.45) to deduce that $\overline{W}$ is an $\overline{\mathbb{F}}$-Brownian motion. With this $\overline{W}$, we have the representation

$$\overline{X}_t - \overline{X}_0 - \int_0^t \overline{b}(s, \overline{X}_s)ds = M_t = \int_0^t \overline{\sigma}(s, \overline{X}_s)d\overline{W}_s.$$

Step (ii). Define the process $B = \{B_t\}$ by $B_t = \overline{W}_{1-t} - \overline{W}_1$. It is straightforward to see that B is a Brownian motion. Let $t > s$. It follows from $B_t - B_s = -(\overline{W}_{1-s} - \overline{W}_{1-t})$ and $\overline{\mathcal{F}}_{1-t} \supset \sigma(X_r; r \geq t)$ that $B_t - B_s$ is independent of $\sigma(X_r; r \geq t)$.

For a fixed $t \geq$, take an arbitrary partition $\{t_i\}_{i=0}^n$ of $[1-t, 1]$ such that $1 - t = t_0 < \cdots < t_n = 1$. Using the result from Problem 2.34, we find

$$\begin{aligned}\int_{1-t}^1 \overline{\sigma}^{-1}(r, \overline{X}_r)d\overline{X}_r &= \lim_{\Delta \to 0} \sum_{i=0}^{n-1} \overline{\sigma}^{-1}(t_i, \overline{X}_{t_i})(\overline{X}_{t_{i+1}} - \overline{X}_{t_i}) \\ &= -\lim_{\Delta \to 0} \sum_{i=0}^{n-1} \sigma^{-1}(1 - t_i, X_{1-t_i})(X_{1-t_i} - X_{1-t_{i+1}})\end{aligned}$$

in probability, where $\Delta = \max_i(t_{i+1} - t_i)$. Put $s_i = 1 - t_{n-i}$. Then $0 = s_0 < \cdots < s_n = 1 - t$ and $\Delta = \max_i(s_{i+1} - s_i)$. The observation just above means that the limit

$$\lim_{\Delta \to 0} \sum_{i=1}^{n} \sigma^{-1}(s_i, X_{s_i})(X_{s_i} - X_{s_{i-1}})$$

exists in probability, whence by Definition 2.43

$$\int_{1-t}^1 \overline{\sigma}^{-1}(r, \overline{X}_r)d\overline{X}_r = -\int_0^t \sigma^{-1}(r, X_r)\overleftarrow{d}X_r.$$

Since we have assumed that σ is positive and a $C^{1,2}$-function, we see

$$d\sigma^{-1}(t, X_t) = \left(\partial_t \sigma^{-1} + b\partial_x \sigma^{-1} + \frac{1}{2}\sigma^2 \partial_{xx}^2 \sigma^{-1}\right)(t, X_t)dt + (\partial_x \sigma^{-1}(t, X_t))\sigma(t, X_t)dW_t$$

and so

$$d\langle \sigma^{-1}(\cdot, X_\cdot), X \rangle_t = (\partial_x \sigma^{-1}(t, X_t))\sigma^2(t, X_t)dt = -\partial_x \sigma(t, X_t)dt.$$

Therefore by Proposition 2.44,

$$\int_0^t \sigma^{-1}(r, X_r)\overleftarrow{d}X_r = \int_0^t \sigma^{-1}(r, X_r)dX_r + \langle \sigma^{-1}(\cdot, X_\cdot), X \rangle_t = W_t + \int_0^t (\sigma^{-1}b - \partial_x \sigma)(r, X_r)dr.$$

So,

$$\begin{aligned}B_t = -\int_{1-t}^1 \overline{\sigma}^{-1}(r, \overline{X}_r)(d\overline{X}_r - \overline{b}(r, \overline{X}_r)dr) &= \int_0^t \sigma^{-1}(r, X_r)\overleftarrow{d}X_r - \int_0^t (\sigma^{-1}\tilde{b})(r, X_r)dr \\ = W_t + \int_0^t (p^{-1}\partial_x(\sigma p))(r, X_r)dr.\end{aligned}$$

In particular, B is an $\mathbb{F}$-semimartingale. Again by Proposition 2.44,

$$\int_t^1 \sigma(r, X_r)\overleftarrow{d}B_r = \int_t^1 \sigma(r, X_r)dB_r + \langle \sigma(\cdot, X_\cdot), B \rangle_t = \int_t^1 \sigma(r, X_r)dW_r + \int_t^1 (p^{-1}\partial_x(\sigma^2 p))(r, X_r)dr.$$

From this we obtain

$$X_1 = X_t + \int_t^1 b(r, X_r)dr + \int_t^1 \sigma(r, X_r)dW_r = X_t + \int_t^1 (b - p^{-1}\partial_x(\sigma^2 p))(r, X_r)dr + \int_t^1 \sigma(r, X_r)\overleftarrow{d}B_r,$$

as wanted. $\square$

CHAPTER 4

Stochastic Controls

The term *stochastic controls* generally refers to the optimization problems defined for stochastic dynamical systems with *control inputs*. Here we present a basic approach to stochastic controls in the framework of SDEs. We refer to Øksendal [31], Fleming and Rishel [10], Bensoussan [4], Fleming and Soner [11], Pham [32], Yong and Zhou [40], and to the lecture notes Touzi [36] and van Handel [37] for more quick overviews and for more detailed accounts.

Throughout this chapter, $T \in (0, \infty)$ is a fixed constant representing a time maturity, and we assume that $\{W_t\}_{0 \le t \le T}$ is an m-dimensional Brownian motion unless stated otherwise.

4.1 Optimization Problems

We consider the stochastic dynamical systems with control input through the SDEs with exogenous variables. Namely, we consider the *controlled stochastic differential equations*, described in the form

$$dX_s = b(s, X_s, \alpha_s)ds + \sigma(s, X_s, \alpha_s)dW_s. \tag{4.1.1}$$

We call $\{\alpha_t\}$ a *control process*. Suppose that our objective is to optimize a performance of the controlled SDEs with suitable criterion over control processes. This leads to the following optimization problem:

$$\min_{\{\alpha_t\}_{0 \le t \le T}} \mathbb{E}\left[\int_0^T f(s, X_s, \alpha_s)ds + g(X_T)\right]. \tag{4.1.2}$$

The function g evaluates the terminal value of the SDE and f indicates a running cost. The problem (4.1.2) is generally called a *stochastic control problem*.

Before discussing the stochastic control problems rigorously, we shall present a few examples.

Example 4.1 (Merton Problem [26], [27])*.* Let S_t be the price of a stock at time t, and B_t the price of a riskless bond at time t. Suppose that we are in a position to invest our wealth into these two assets by dynamically changing the *fraction* of the wealth to the stock. Denote by X_t our wealth at time t. If we have ϕ_t shares of the stock at time t, then the resulting fraction α_t to the stock is

$$\alpha_t = \frac{\phi_t S_t}{X_t},$$

whence $\phi_t = \alpha_t X_t / S_t$. The remaining fraction $1 - \alpha_t$ is invested into the riskless bond, and so the number of shares invested into the riskless bond at time t is $(1 - \alpha_t)X_t/B_t$. Thus, assuming there is neither income nor consumption in the period $[t, t + \Delta t]$, we obtain

$$X_{t+\Delta t} - X_t = \frac{\alpha_t X_t}{S_t}(S_{t+\Delta t} - S_t) + \frac{(1-\alpha_t)X_t}{B_t}(B_{t+\Delta t} - B_t).$$

This leads to the SDE

$$\frac{dX_t}{X_t} = \alpha_t \frac{dS_t}{S_t} + (1 - \alpha_t) \frac{dB_t}{B_t} \tag{4.1.3}$$

for the wealth process. In the simplest case, the price dynamics of the two assets are assumed to be described respectively by

$$\begin{aligned} dS_t &= S_t(bdt + \sigma dW_t), \\ dB_t &= rB_t dt, \end{aligned}$$

where $m = 1$, and b, σ, r are constants with $\sigma > 0$ and $r \geq 0$. Then (4.1.3) turns out to be

$$dX_t = X_t[r + (b - r)\alpha_t]dt + X_t \alpha_t dW_t. \tag{4.1.4}$$

The investor's problem here is to maximize the expected utility of the wealth

$$\mathbb{E}[U(X_T)] \tag{4.1.5}$$

over all portfolio proportion processes $\{\alpha_t\}$. Here $U : (0, \infty) \to \mathbb{R}$ satisfies $U' > 0$ and $U'' < 0$, which is called a *utility function*.

Example 4.2 (Aircraft trajectory planning [25]). Consider an aircraft's motion in the 2-dimensional horizontal plane. We assume that the local navigation frame is described by the 2-dimensional Euclidean plane where x-axis points the east and y-axis points the north. Then, the state X_t of the aircraft is described by a vector in $\mathbb{R}^2$. We further assume that the current heading of the aircraft is determined by the control variable $\alpha_t \in A = [0, 2\pi)$. With these assumptions, the dynamic of X_t can be described by

$$dX_t = \begin{pmatrix} \cos(\alpha_t) \\ \sin(\alpha_t) \end{pmatrix} v_c dt + dY_t.$$

where v_c is the aircraft's cruise speed, assumed to be constant, and $Y_t = Y_t(x)$ describes the wind disturbance at the position x. A simple model for the wind disturbance is

$$dY_t(x) = y(t, x)dt + \sigma(t, x)dW_t.$$

Here $y(t, x)$ describes a mean behavior of the wind, which is a deterministic vector field, and $\sigma(t, x)$ is a magnitude of random fluctuations at (t, x), both of which are estimated by weather charts. Further, W_t is a 2-dimensional Brownian motion. Thus, the controlled process X_t is given by

$$dX_t = \left[\begin{pmatrix} \cos(\alpha_t) \\ \sin(\alpha_t) \end{pmatrix} v_c + y(t, X_t) \right] dt + \sigma(t, X_t)dW_t.$$

The objective of the trajectory planning here is to control the movement of the airplane so as to enter a given area S_0 at the terminal time T while avoiding a forbidden area S_1. Then the problem is

$$\min_{\{\alpha_t\}} \mathbb{E}\left[d(X_T, S_0) + \lambda \int_0^T e^{-\gamma d(X_t, S_1)} dt \right],$$

where $\lambda, \gamma > 0$ and $d(x, S_1)$ denotes a distance between a point $x \in \mathbb{R}^d$ and a set $S \subset \mathbb{R}^2$.

We turn to the rigorous formulation. In what follows, we fix an $\mathcal{F}_0$-measurable random variable $X_0 \in L^2$ and a closed subset A of $\mathbb{R}^{d_1}$. We assume that the evaluation functions g on $\mathbb{R}^d$ and f on $[0, T] \times \mathbb{R}^d \times A$ are Borel measurable. Denote by $\mathcal{A}$ the collection of all processes $\alpha = \{\alpha_t\}_{0 \leq t \leq T}$ such that

(i) α is A-valued and $\mathbb{F}$-adapted;

(ii) the SDE (4.1.1) has a unique solution $\{X_t^\alpha\}_{0\le t\le T}$ with initial condition $X_0^\alpha = X_0$;

(iii) The criterion is finite, i.e.,

$$\mathbb{E}\left[\int_0^T f(s, X_s^\alpha, \alpha_s)ds + g(X_T^\alpha)\right]$$

is finite.

We call elements in $\mathcal{A}$ *control processes*. Then, given a subset $\tilde{\mathcal{A}} \subset \mathcal{A}$, our stochastic control problem is describe by

$$\inf_{\alpha\in\tilde{\mathcal{A}}} \mathbb{E}\left[\int_0^T f(s, X_s^\alpha, \alpha_s)ds + g(X_T^\alpha)\right]. \tag{4.1.6}$$

- We say that (4.1.6) is a *finite time horizon problem*.
- The stochastic control problem

$$\inf_{\alpha\in\tilde{\mathcal{A}}} \mathbb{E}\left[\int_0^{\tau^\alpha} f(s, X_s^\alpha, \alpha_s)ds + g(X_{\tau^\alpha}^\alpha)\right],$$

  where τ^α is the first exit time of $\{X_s^{0,x,\alpha}\}$ from a given set $S \subset \mathbb{R}^d$, is called an *indefinite time horizon problem*, and the one

$$\inf_{\alpha\in\tilde{\mathcal{A}}} \mathbb{E}\int_0^\infty e^{-\lambda s} f(s, X_s^\alpha, \alpha_s)ds,$$

  where $\lambda \ge 0$, is called an *infinite time horizon problem*. The both problems have many important applications. However, we omit to deal with them for simplicity of the presentation.
- Suppose that $\{X_t^*\}_{0\le t\le T}$ is a unique solution of

$$dX_t^* = b(t, X_t^*, a(t, X_t^*))dt + \sigma(t, X_t^*, a(t, X_t^*))dW_t$$

  for some Borel function a and that $\alpha_t^* := a(t, X_t^*)$, $0 \le t \le T$, is in $\mathcal{A}$. Then, by the uniqueness, $X_t^{\alpha^*} = X_t^*$. We call such α^* a *Markov control*.
- Of course $\alpha_t := a(t, \max_{0\le s\le t} X_s^*)$, $0 \le t \le T$, is not a Markov control. Thus, in general, the controlled SDEs (4.1.1) differ from those considered in Chapter 3 in that the former depends on possibly non-Markovian processes.

To discuss the existence and uniqueness of (4.1.1), we assume here that $b : [0, T] \times \mathbb{R}^d \times A \to \mathbb{R}^d$ and $\sigma : [0, T] \times \mathbb{R}^d \times A \to \mathbb{R}^{d\times m}$ continuous functions and that there exists a positive constant C_0 such that for $(t, x, y, a) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times A$,

$$|b(t, x, a) - b(t, y, a)| + |\sigma(t, x, a) - \sigma(t, y, a)| \le C_0|x - y|, \tag{4.1.7}$$

and that

$$\mathbb{E}\int_0^T \left(|b(t, 0, \alpha_t)|^2 + \sigma(t, 0, \alpha_t)|^2\right) dt < \infty \tag{4.1.8}$$

for a given A-valued and adapted process α.

We can apply the same argument as in the proof of Theorem 3.2 to obtain the following:

Theorem 4.3

Suppose that the conditions (4.1.7) and (4.1.8) hold. Then, there exists a unique solution $\{X_t^\alpha\}_{0\le t\le T}$ of (4.1.1) with initial condition $X_0^\alpha = X_0$.

Actually, Theorem 4.3 is a corollary of the following result:

Theorem 4.4

Let $t \in [0, T]$. Consider the equation

$$X_s = \xi + \int_t^s \bar{b}(s, X_s)ds + \int_t^s \bar{\sigma}(s, X_s)dW_s, \quad t \le s \le T, \tag{4.1.9}$$

where ξ, $\bar{b} : \Omega \times [t, T] \times \mathbb{R}^d$ and $\bar{\sigma} : \Omega \times [t, T] \times \mathbb{R}^{d \times m}$ satisfy

(i) ξ is an $\mathbb{R}^d$-valued and $\mathcal{F}_t$-measurable random variable with $\mathbb{E}|\xi|^2 < \infty$.

(ii) $\bar{b}(s, x)$ and $\bar{\sigma}(s, x)$ are adapted for each $(s, x) \in [t, T] \times \mathbb{R}^d$.

(iii) There exists a positive constant C_1 such that

$$|\bar{b}(s, x) - \bar{b}(s, y)| \le C_1 |x - y|, \quad s \in [t, T], \ \ x, y \in \mathbb{R}^d.$$

(iv) The processes $\{\bar{b}(s, 0)\}$ and $\{\bar{\sigma}(s, 0)\}$ are in $\mathcal{L}^2$, i.e.,

$$\mathbb{E} \int_t^T \left(|\bar{b}(s, 0)|^2 + |\bar{\sigma}(s, 0)|^2 \right) dt < \infty.$$

Then, there exists a unique solution $\{X_s\}_{t \le s \le T}$ of (4.1.9) satisfying $\mathbb{E} \sup_{t \le s \le T} |X_s|^2 < \infty$.

- As in Chapter 3, we write $\{X_s^{t,\xi,\alpha}\}_{t \le s \le T}$ for the unique solution of (4.1.1) with initial condition $X_t^{t,\xi,\alpha} = \xi$.

4.2 Verification Theorem

Consider the following nonlinear second order PDE, called the *Hamilton-Jacobi-Bellman* (HJB) *equation*:

$$\begin{aligned} \partial_t V(t, x) + \inf_{a \in A} H^a(t, x, DV(t, x), D^2 V(t, x)) &= 0, \quad (t, x) \in [0, T) \times \mathbb{R}^d, \\ V(T, x) &= g(x), \quad x \in \mathbb{R}^d. \end{aligned} \tag{4.2.1}$$

A central tool for solving stochastic control problems is the HJB equation, which characterizes the value function of the stochastic control problems. Once a candidate value function is identified, one needs to verify that it indeed represents the optimal cost. The purpose of this section is to provide a rigorous formulation of this verification step. The verification theorem establishes sufficient conditions under which a suitably smooth solution of the HJB equation coincides with the value function and yields an optimal control.

Theorem 4.5: Verification theorem

Suppose that there exists a $C^{1,2}$-function V on $[0,T]\times\mathbb{R}^d$ that is a solution of (4.2.1). Suppose moreover that the following are satisfied:

(i) For every $\alpha\in\tilde{\mathcal{A}}$,

$$\mathbb{E}\left[\sup_{0\le t\le T}|V(t,X_t^\alpha)|+\int_0^T|f(t,X_t^\alpha,\alpha_t)|dt\right]<\infty.$$

(ii) There exists a Borel function a^* on $[0,T]\times\mathbb{R}^d$ such that

$$\inf_{a\in A}H^a(t,x,DV(t,x),D^2V(t,x))=H^{a^*(t,x)}(t,x,DV(t,x),D^2V(t,x)),\quad (t,x)\in[0,T]\times\mathbb{R}^d.$$

(iii) There exists a unique solution $\{X_t^*\}_{0\le t\le T}$ of the SDE

$$dX_t^*=b(t,X_t^*,a^*(t,X_t^*))dt+\sigma(t,X_t^*,a^*(t,X_t^*))dW_t,\quad X_0^*=X_0.$$

(iv) The process $\alpha_t^*:=a^*(t,X_t^*)$, $0\le t\le T$, belongs to $\tilde{\mathcal{A}}$.

Then α^* is optimal for the problem (4.1.6).

Proof. For $\alpha\in\tilde{\mathcal{A}}$ and $n\in\mathbb{N}$ define the stopping time τ_n^α by

$$\tau_n^\alpha=\inf\{t\in[0,T]:|X_t^\alpha|>n\}\wedge T.$$

Then, using Itô formula and (4.2.1), we have

$$\mathbb{E}\left[\int_0^{\tau_n^\alpha}f(s,X_s^\alpha,\alpha_s)ds+V(\tau_n^\alpha,X_{\tau_n^\alpha}^\alpha)\right]\ge\mathbb{E}[V(0,X_0)].$$

Then, by the dominated convergence theorem,

$$\mathbb{E}\left[\int_0^T f(s,X_s^\alpha,\alpha_s)ds+g(X_T^\alpha)\right]\ge\mathbb{E}[V(0,X_0)],$$

whence

$$\inf_{\alpha\in\tilde{\mathcal{A}}}\mathbb{E}\left[\int_0^T f(s,X_s^\alpha,\alpha_s)ds+g(X_T^\alpha)\right]\ge\mathbb{E}[V(0,X_0)].$$

On the other hand, by the uniqueness, $X_t^{\alpha^*}=X_t^*$, $0\le t\le T$, a.s. Thus, using the conditions in Theorem 4.5 and the localizing argument as in above,

$$\begin{aligned}\mathbb{E}\left[\int_0^T f(s,X_s^{\alpha^*},\alpha_s^*)ds+g(X_T^{\alpha^*})\right]&=\mathbb{E}[V(0,X_0)]\\&\le\inf_{\alpha\in\tilde{\mathcal{A}}}\mathbb{E}\left[\int_0^T f(s,X_s^\alpha,\alpha_s)ds+g(X_T^\alpha)\right].\end{aligned}$$

Since $\alpha^*\in\tilde{\mathcal{A}}$, we deduce that α^* is optimal. □

- It is straightforward to see that V coincides with the *value function*

$$v(t,x):=\inf_{\alpha\in\tilde{A}}\mathbb{E}\left[\int_t^T f(s,X_s^{t,x,\alpha},\alpha_s)ds+g(X_T^{t,x,\alpha})\right],\quad (t,x)\in[0,T]\times\mathbb{R}^d,$$

where $\{X_s^{t,x,\alpha}\}_{t\le s\le T}$ is a unique solution of

$$dX_s^{t,x,\alpha} = b(s, X_s^{t,x,\alpha}, \alpha_s)ds + \sigma(s, X_s^{t,x,\alpha}, \alpha_s)dW_s,$$

with initial condition $X_t^{t,x,\alpha} = x$. That is, we have

$$v(t,x) = V(t,x), \quad (t,x) \in [0,T] \times \mathbb{R}^d,$$

provided that v is well-defined and the conditions in Theorem 4.5 hold.

Example 4.6 (Linear regulator problem). Consider the controlled SDE

$$dX_t^\alpha = (b(t)X_t^\alpha + c(t)\alpha_t)\,dt + \sigma(t)dW_t, \tag{4.2.2}$$

where $b : [0,T] \to \mathbb{R}^{d\times d}$, $c : [0,T] \to \mathbb{R}^{d\times d_1}$, and $\sigma : [0,T] \to \mathbb{R}^{d\times m}$, all of which are continuous. The problem is to minimize

$$\mathbb{E}\left[(X_T^\alpha)^\mathsf{T} R X_T^\alpha + \int_0^T \left\{(X_t^\alpha)^\mathsf{T} P(t) X_t^\alpha + (\alpha_t)^\mathsf{T} Q(t)\alpha_t\right\} dt\right]$$

over all $\mathbb{R}^{d_1}$-valued process $\alpha = \{\alpha_t\}_{0\le t\le T}$ with each component belonging to $\mathcal{L}_2$. Here, $R \in \mathbb{S}^d$ and the functions $P : [0,T] \to \mathbb{S}^d$, $Q : [0,T] \to \mathbb{S}^{d_1}$ are assumed to be continuous and nonnegative definite. Further, $Q(t)$ is assumed to be positive definite for any $t \in [0,T]$. By Theorem 4.3 (and Theorem 4.4 or by a direct esitmation), there exists a unique solution $\{X_t^\alpha\}_{0\le t\le T}$ of (4.2.2) for any $\alpha = \{\alpha_t\}_{0\le t\le T}$ as in above and initial condition $X_0 \in L^2$ such that $\mathbb{E}\sup_{0\le t\le T}|X_t^\alpha|^2 < \infty$. Thus the criterion is always finite. So we take $\tilde{\mathcal{A}}$ to be the set of all $\mathbb{R}^{d_1}$-valued processes α such that each component is in $\mathcal{L}_2$. Then $\tilde{\mathcal{A}} \subset \mathcal{A}$.

Theorem 4.5 suggests that if the HJB equation has an explicit solution then the solution gives a candidate of an optimal solution. In our case,

$$\begin{aligned}
&H^a(t, x, Dv(t,x), D^2v(t,x)) \\
&= (b(t)x + c(t)a)^\mathsf{T} DV(t,x) + \frac{1}{2}\mathrm{tr}(\sigma(t)\sigma(t)^\mathsf{T} D^2V(t,x)) + x^\mathsf{T} P(t)x + a^\mathsf{T} Q(t)a.
\end{aligned}$$

Therefore, the infimum of H^a's is attained by

$$a^*(t,x) = -\frac{1}{2}Q(t)^{-1}DV(t,x)^\mathsf{T} c(t).$$

In view of the linear-quadratic structure of the problem, we look for a solution V of the HJB equation by assuming $V(t,x) = x^\mathsf{T} F(t)x + G(t)$ for some deterministic functions $F : [0,T] \to \mathbb{S}^d$ and $G : [0,T] \to \mathbb{R}$. Substituting this form into the HJB equation, we see

$$\begin{aligned}
&x^\mathsf{T}\left[F'(t) - F(t)c(t)Q(t)^{-1}c(t)^\mathsf{T} F(t) + P(t) + b(t)^\mathsf{T} F(t) + F(t)b(t)\right] x \\
&+ G'(t) + \mathrm{tr}(\sigma(t)\sigma(t)^\mathsf{T} F(t)) = 0
\end{aligned}$$

for $(t,x) \in [0,T) \times \mathbb{R}^d$, where $\dot{L}(t) = dL(t)/dt$. This leads to the ODEs

$$\begin{gathered}
F'(t) - F(t)c(t)Q(t)^{-1}c(t)^\mathsf{T} F(t) + P(t) + b(t)^\mathsf{T} F(t) + F(t)b(t) = 0, \quad F(T) = R, \\
G'(t) + \mathrm{tr}(\sigma(t)\sigma(t)^\mathsf{T} F(t)) = 0, \quad G(T) = 0.
\end{gathered} \tag{4.2.3}$$

It is known that there exists a solution of the *matrix Riccati differential equation* (4.2.3) (see Theorem 5.2 in [10]). With this F, the function G is explicitly determined and so $V(t,x) = x^\mathsf{T} F(t)x + G(t)$ is a solution of the HJB equation. Consequently, $a^*(t,x) = -Q(t)^{-1}c(t)^\mathsf{T} F(t)x$.

Problem 4.7. In Example 4.6, complete the remaining arguments to be done and obtain an optimal control using Theorem 4.5.

Problem 4.8. Try to find an optimal control for a more general problem than that in Example 4.6.

Before turning to next example, we observe that the following theorem holds:

Theorem 4.9

Let $\{b_t\}_{0\le t\le T}$ and $\{\sigma_t\}_{0\le t\le T}$ be $\mathbb{R}$-valued and $\mathbb{R}^m$-valued adapted processes such that

$$\int_0^T |b_t|dt + \int_0^T |\sigma_t|^2 dt < \infty, \quad \text{a.s.},$$

respectively. Then there exists a unique solution $\{Z_t\}_{0\le t\le T}$ of the SDE

$$dZ_t = Z_t(b_t dt + \sigma_t^{\mathsf{T}} dW_t), \quad Z_0 = 1. \tag{4.2.4}$$

Proof. Put

$$Y_t = \int_0^t \left(b_s - \frac{1}{2}|\sigma_s|^2\right) ds + \int_0^t \sigma_s^{\mathsf{T}} dW_s, \quad 0 \le t \le T.$$

Then, with Itô formula, it is straightforward to see that $Z_t := e^{Y_t}$, $0 \le t \le T$, is a solution of (4.2.4). Let Z_t', $0 \le t \le T$, be another solution. Then, Itô formula yields $dZ_t' e^{-Y_t} = 0$. Thus $Z_t = Z_t'$, $0 \le t \le T$. □

Example 4.10 (Merton problem). Recall the investment problem in Example 4.1. By Theorem 4.9 there exists a unique solution $\{X_t^\alpha\}_{0\le t\le T}$ of (4.1.4) for any $\mathbb{R}$-valued adapted process $\alpha \in \mathcal{L}_{2,loc}$, given by

$$X_t = X_0 \exp\left[\int_0^t \left(r + (b-r)\alpha_s - \frac{1}{2}\sigma^2\alpha_s^2\right) ds + \sigma\int_0^t \alpha_s dW_s\right], \quad 0 \le t \le T.$$

Here we take $U(x) = x^q$, $x > 0$, for some $q \in (0,1)$, and then define $\tilde{\mathcal{A}}$ by the set of all $\mathbb{R}$-valued processes $\alpha \in \mathcal{L}_{2,loc}$ such that $\mathbb{E}\sup_{0\le t\le T} U(X_t^\alpha) < \infty$. Moreover, we assume that X_0 is a positive constant.

To solve the control problem, we consider

$$Y_t^\alpha := q\int_0^t \left(r + (b-r)\alpha_s - \frac{1}{2}\sigma^2\alpha_s^2\right) ds + q\sigma\int_0^t \alpha_s dW_s, \quad 0 \le t \le T.$$

as a state variable. Then the corresponding HJB equation is

$$\begin{aligned} \partial_t v(t,y) + \sup_{a\in\mathbb{R}} H^a(y, Dv(t,y), D^2 v(t,y)) &= 0, \quad (t,y) \in [0,T) \times \mathbb{R}, \\ v(T,y) &= e^y, \quad y \in \mathbb{R}. \end{aligned} \tag{4.2.5}$$

where

$$H^a(x,p,\gamma) = q(r + (b-r)a - \frac{1}{2}\sigma^2 a^2)p + \frac{1}{2}q^2\sigma^2 a^2\gamma.$$

We look for a solution of (4.2.5) of the form $v(t,y) = w(t)e^y$, where w is a positive deterministic function. Substituting this form into (4.2.5), we observe

$$\begin{aligned} 0 &= e^y\left\{w'(t) + qw(t)\sup_{a\in\mathbb{R}}\left[r + (b-r)a - \frac{1}{2}(1-q)\sigma^2 a^2\right]\right\} \\ &= e^y\left\{w'(t) + qw(t)\left[r + \frac{q(b-r)^2}{2\sigma^2(1-q)}\right]\right\}, \end{aligned}$$

where the supremum is attained by $a^* := (b-r)/(\sigma^2(1-q))$. Thus, $v(t,y) = \exp(q\theta(T-t)+y)$ with $\theta = r + (b-r)^2/(2\sigma^2(1-q))$ is a solution of (4.2.5), and the constant control $\alpha_t^* := a^*$ is a candidate of an optimal portfolio proportion. By the verification theorem, we can show that α^* is indeed optimal.

Problem 4.11. In Example 4.10, check the conditions in Theorem 4.5 hold to confirm the optimality of α^*. Doob's maximal inequality will help you.

- As we have seen so far, the verification theorem gives a way of constructing an optimal control. In particular, Theorem 4.5 gives sufficient conditions for which optimal control *exists*.
- To apply Theorem 4.5 for applications, we need to obtain an explicit solution of the HJB equation, which is rarely available, however. Even more, a classical solution may not exist.

As for the existence of optimal Markovian controls, we have the following result:

Theorem 4.12

Suppose that A is compact, b, σ, g are all bounded continuous functions, and $f = 0$. Suppose moreover that the set

$$\{(\sigma(t,x,a)\sigma(t,x,a)^\mathsf{T}, b(t,x,a)) : a \in A\}$$

is convex for all $(t,x) \in [0,T] \times \mathbb{R}^d$. Then, there exist a filtered probability space $(\Omega^*, \mathcal{F}^*, \mathbb{F}^*, \mathbb{P}^*)$, a process $\alpha^* \in \mathcal{A}$, and a Borel function a^* on $[0,T] \times \mathbb{R}^d$ such that α^* is optimal for the stochastic control problem (4.1.6) defined on this filtered probability space, where $\mathcal{A} = \tilde{\mathcal{A}}$ is defined by the set of all A-valued adapted processes, and that

$$\alpha_t^* = a^*(t, X_t^{\alpha^*}), \quad \text{a.s.}, \quad 0 \le t \le T.$$

For a proof of this theorem, we refer to Haussmann [14].

We close this section by giving an example of HJB equations having no classical solutions.

Example 4.13. Consider the case where the controlled SDE $\{X_t^\alpha\}$ is given by

$$dX_t^\alpha = \alpha_t dW_t,$$

with a nonrandom initial condition, and then the optimal control problem

$$\sup_{\alpha \in \tilde{\mathcal{A}}} \mathbb{E}[g(X_T^\alpha)]$$

where $\tilde{\mathcal{A}}$ is the set of all $\mathbb{R}$-valued processes in $\mathcal{L}_2$, and

$$g(x) = \begin{cases} \sin x & (x \ge 0), \\ x & (x < 0). \end{cases}$$

Suppose that there exists a $C^{1,2}([0,T] \times \mathbb{R})$-solution V of the corresponding HJB equation

$$\partial_t V(t,x) + \frac{1}{2} \sup_{a \in \mathbb{R}} [a^2 D^2 V(t,x)] = 0, \quad (t,x) \in [0,T) \times \mathbb{R},$$
$$V(T,x) = g(x), \quad x \in \mathbb{R}.$$

Then, $D^2 V(t,x) \le -2\partial_t V(t,x)/a^2$ for $a \ne 0$, and so letting $a \to \infty$ we have $D^2 V(t,x) \le 0$ for every $(t,x) \in [0,T) \times \mathbb{R}$. Hence, $V(t,\cdot)$ is concave on $\mathbb{R}$ and $V(t,\cdot) = g$, which is a contradiction.

4.3 Problems with Terminal Time Constraints

This section is devoted to a special class of control problems described by the following:

Problem $(\mathcal{S})$**.** Given two Borel probability measures μ_0, μ_1 on $\mathbb{R}^d$ and a positive constant σ, minimize

$$\mathbb{E}\int_0^1 |u_t|^2 dt$$

over all $\mathbb{R}^d$-valued control processes $\{u_t\}$ such that the corresponding controlled diffusions

$$X_t = X_0 + \int_0^t u_s ds + \sigma W_t$$

satisfy $\mathbb{P}(X_0 \in dx) = \mu_0(dx)$ and $\mathbb{P}(X_1 \in dy) = \mu_1(dy)$, where $\{W_t\}$ is a d-dimensional Brownian motion.

We shall adopt a weak formulation of the control problems above, i.e., the minimization are taken over all possible probability measures, Brownian motions and control processes. We will present a rigorous formulation below.

When there is no terminal time constraint $\mathbb{P}(X_1 \in dy) = \mu_1(dy)$ and μ_0 is a Gaussian distribution, then the problem $(\mathcal{S})$ is a special case of the linear regulator problem (see Example 4.6). Our problem here is to find a controlled diffusion $dX_t = u_t dt + \sigma dW_t$ that starts from the initial distribution μ_0 and arrives at a predetermined final distribution μ_1, with minimum "energy" $\mathbb{E}\int_0^1 |u_t|^2 dt$.

Background

In the two papers [34] and [35], Erwin Schrödinger considered the following thought experiment: for N-independent Brownian particles $X^{(1)}, \ldots, X^{(N)}$, suppose that at time $t = 0$, this cloud approximately follows $\mu_0(dx)$ and at time $t = 1$ the observed distribution of the could follows $\mu_1(dy)$. Then, what is a could evolution that most likely occurs?

The law of large numbers tells us that the above transition is a *rare* event. To be precise, if the initial distribution of each particle $X^{(j)}$ follows μ_0 then by the strong law of large number, the empirical measure $(1/N)\sum_{j=1}\delta_{X_1^{(j)}}$ of this could at terminal time converges to

$$\int p(0, x, 1, y)\mu_0(dx)dy \neq \mu_1(dy),$$

almost surely as $N \to \infty$, where p is the transition density of a Brownian motion. To determine a reasonable cloud transition probability among these unlikely possibilities, Schrödinger used a particle migration model with space discretization, exactly computed the distribution of the random variable of the particle migrations under the initial and terminal time constrains, and then adopted the *maximum entropy principle*. Then, after taking the continuous limit, he derived a system of partial differential equations for the optimal transition probability, the so called *Schrödinger system* or *Schrödinger's functional equation* (see (4.3.5) below). We refer to an english translation [7] of [34] for an exposition of the Schrödinger's original approach.

Föllmer [12] discovers the Schrödinger's problem is nothing but the one of *large deviation*. By Sanov's theorem (see, e.g., [9]) for the large deviation principles on empirical measures, the problem of computing

$$\frac{1}{N}\log\mathbb{P}\left(\frac{1}{N}\sum_{j=1}^{N}\delta_{X_t^{(j)}} \text{ follows } \mu_t, \quad t = 0, 1\right)$$

is nearly equivalent to the minimization problem of the relative entropy

$$H(Q|P) := \begin{cases} \mathbb{E}_Q\left[\log \frac{dQ}{dP}\right], & (Q \ll P), \\ +\infty, & (\text{otherwise}). \end{cases}$$

over all Borel probability measures Q on $\mathbb{W}^d$ such that the initial and terminal marginals are given by μ_0 and μ_1, respectively, where P is the a priori law of Brownian particles. An optimal measure Q that solves this minimization problem is called the *Schrödinger's bridge* between μ_0 and μ_1.

Assume here that $\Omega = \mathbb{W}^d$ and X is the coordinate process: $X_t(\omega) = \omega(t)$. Under P, the process X is represented as $X_t = X_0 + W_t$, where X_0 follows μ_0 and W is a Brownian motion that is independent of X_0. Roughly speaking, by the martingale representation theorem, every Q above can be represented as

$$\frac{dQ}{dP} = \exp\left[\int_0^1 u_t^\top dW_t - \frac{1}{2}\int_0^1 |u_t|^2 dt\right]$$

for some process $\{u_t\}$ adapted to the canonical filtration. By Girsanov's theorem, the process $\tilde{W}_t := W_t - \int_0^t u_s ds$ is a Brownian motion under Q, whence X is represented as a controlled diffusion given by

$$X_t = X_0 + \int_0^t u_s ds + \tilde{W}_t$$

under Q. Assuming the square integrability for $\{u_t\}$ formally, we obtain

$$H(Q|P) = \frac{1}{2}\mathbb{E}_Q \int_0^1 |u_t|^2 dt.$$

Consequently, finding a Schrödinger's bridge is roughly equivalent to solving the problem $(\mathcal{S})$ with $\sigma = 1$. For this reason, we will call $(\mathcal{S})$ the *Schrödinger's bridge problem.* We refer to, e.g., [6] and [24] for surveys of Schrödinger's bridges.

Connections with optimal transport

Let Q_{01} be the law of (X_0, X_1) under Q, where Q is a Borel probability measure on $\Omega = \mathbb{W}^d$. Then, one can prove that the problem

$$\inf_{Q \in \mathcal{P}(\mu_0, \mu_1)} H(Q|P),$$

where $\mathcal{P}(\mu_0, \mu_1)$ is the set of all Borel probability measure Q on Ω such that $QX_0^{-1} = \mu_0$ and $QX^{-1} = \mu_1$, is equivalent to the *static* problem

$$\inf_{R \in \mathcal{P}_s(\mu_0, \mu_1)} H(R|P_{01}), \tag{4.3.1}$$

where $\mathcal{P}_s(\mu_0, \mu_1)$ is the totality of all Borel probability measures R on $\mathbb{R}^d \times \mathbb{R}^d$ satisfying

$$R(A \times \mathbb{R}^d) = \mu_0(A), \quad R(\mathbb{R}^d \times A) = \mu_1(A), \quad A \in \mathcal{B}(\mathbb{R}^d)$$

(see Proposition 4.19 below).

Let us slightly generalize the situation to the case $\sigma > 0$, i.e., $X_t = X_0 + \sigma W_t$, $0 \le t \le 1$. It is straightforward to show the same results as above, where P_{01} is now given by

$$P_{01}(A \times B) = \int_A \int_B p(0, x, 1, y)\mu_0(dx)dy$$

with the transition density

$$p(t,x,s,y) = \frac{e^{-|y-x|^2/(2\sigma^2(s-t))}}{(2\pi\sigma^2(s-t))^{d/2}}, \quad 0 \le t < s \le 1, \;\; x, y \in \mathbb{R}^d. \tag{4.3.2}$$

Assume here that μ_0 has a positive density ρ_0. In this case, for $R(dxdy) = \rho(x,y)dxdy$,

$$\log \frac{dR}{dP_{01}}(x,y) = \frac{1}{2\sigma^2}|y-x|^2 + \frac{d}{2}\log(2\pi\sigma^2) + \log\frac{\rho(x,y)}{\rho_0(x)}, \quad \text{a.e.},$$

whence the static problem (4.3.1) is roughly equivalent to the minimization problem

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^2 R(dxdy) + 2\sigma^2 \int_{\mathbb{R}^d\times\mathbb{R}^d} \left(\log\frac{\rho(x,y)}{\rho_0(x)}\right)\rho(x,y)dxdy + C_\sigma$$

over such R's, where C_σ is a constant independent of R, satisfying $C_\sigma \to 0$ as $\sigma \to 0$. This means that the problem (4.3.1) can be seen as an entropic regularization of the so-called Monge-Kantorovich optimal mass transport problem

$$\inf_{R\in\mathcal{P}_s(\mu_0,\mu_1)} \int_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^2 R(dxdy).$$

It is known that under some regularity conditions there exists a measurable map $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $R^* := (Id \times T)_\sharp\mu_0$ is optimal to the Monge-Kantorovich problem, where Id is the identity map and $f_\sharp\mu_0$ denotes the pushforward of μ_0 with f.

Consequently, solving the Schrödinger's bridge problem with small σ is amount to giving an approximation of the *displacement interpolation* $\mu_t = (T_t)_\sharp\mu_0$ where $T_t(x) = (1-t)x + tT(x)$, $0 \le t \le 1$, by the marginal distributions of a controlled diffusion $\{X_t\}_{0\le t\le 1}$. We refet to [29] for a rigorous convergence analysis in the zero-noise limit of Schrödinger's bridges.

Rigorous formulation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space. Let $\{W_t\}_{0\le t\le 1}$ be a d-dimensional Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$ with respect to a given filtration $\mathbb{F}$ satisfying the usual conditions. Then consider the process

$$X_t = X_0 + \sigma W_t, \quad 0 \le t \le 1.$$

Assume that X_0 follows μ_0 under $\mathbb{P}$, i.e., that $\mathbb{P}(X_0 \in A) = \mu_0(A)$ for any $A \in \mathcal{B}(\mathbb{R}^d)$.

Definition 4.14. We say that a triple $\pi = (\mathbb{Q}, B, u)$ is an *admissible control system* if

(i) $\mathbb{Q}$ is a probability measure on $(\Omega, \mathcal{F})$ such that $\mathbb{Q} \sim \mathbb{P}$;

(ii) $B = \{B_t\}_{0\le t\le 1}$ is a d-dimensional $\mathbb{F}$-Brownian motion on $(\Omega, \mathcal{F}, \mathbb{Q})$;

(iii) $u = \{u_t\}_{0\le t\le 1}$ is an $\mathbb{R}^d$-valued $\mathbb{F}$-adapted process such that

$$\mathbb{E}_{\mathbb{Q}}\left[\int_0^1 |u_s|^2 ds\right] < \infty,$$

$$\mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_0^1 u_s^\top dB_s - \frac{1}{2}\int_0^1 |u_s|^2 ds\right)\right] = 1,$$

and the controlled process

$$X_t^u := X_0 + \int_0^t u_s ds + \sigma B_t, \quad 0 \le t \le 1,$$

satisfies $\mathbb{Q}(X_0^u \in A) = \mu_0(A)$ and $\mathbb{Q}(X_1^u \in A) = \mu_1(A)$ for all $A \in \mathcal{B}(\mathbb{R}^d)$.

We write Π for the set of all admissible control systems.

Note that $X_t^u = X_t$ for the admissible control system $(\mathbb{P}, W, u)$ where $u \equiv 0$. For any admissible control system $\pi = (\mathbb{Q}, W, u) \in \Pi$, define the control criterion by

$$J(\pi) := \mathbb{E}_{\mathbb{Q}} \left[\int_0^1 |u_s|^2 ds \right].$$

Then consider the optimal control problem

$$J^* := \inf_{\pi \in \Pi} J(\pi). \tag{4.3.3}$$

We shall reformulate the problem $(\mathcal{S})$ by (4.3.3). Namely, we redefine the problem $(\mathcal{S})$ by the one of finding an admissible control system that attains the infimum in (4.3.3).

Solutions

As a first step, we consider the static Schrödinger's problem (4.3.1), and derive a first order optimality condition. To this end, first assume that μ_0 and μ_1 have the densities ρ_0 and ρ_1, respectively. Then, the measure $P_{01}(dxdy) = \mathbb{P}((X_0, X_1) \in dxdy)$ has the density $\rho_0(x)p(0, x, 1, y)$, where p is the transition density of $\{X_t\}$ under $\mathbb{P}$, defined by (4.3.2). Further, for any $R \in \mathcal{P}_s(\mu_0, \mu_1)$ such that its density $q\rho_0$ exists, we have

$$H(R|P_{01}) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(\log \frac{q(x,y)}{p(0,x,1,y)} \right) q(x,y)\rho_0(x) dxdy.$$

Introduce Lagrange multipliers $\lambda(x)$ and $\eta(y)$, and then consider the functional

$$L(q, \lambda, \eta) := \int_{\mathbb{R}^d \times \mathbb{R}^d} \left[q(x,y)\rho_0(x) \log \frac{q(x,y)}{p(0,x,1,y)} + (\lambda(x) + \eta(y))(\rho_1(y) - q(x,y))\rho_0(x) \right] dxdy.$$

Assume here that $R(dxdy) = q(x,y)\rho_0(x)dxdy$ is optimal. Let q' be an arbitrary such that $\int_{\mathbb{R}^d} q'(x,y)dy = 0$. By an elementary computation, the first order condition

$$\left. \frac{d}{d\varepsilon} L(q + \varepsilon q')) \right|_{\varepsilon = 0} = 0$$

leads to

$$q(x,y) = p(0,x,1,y)e^{\lambda(x) + \eta(y)}.$$

Now put $\varphi(x) = \rho_0(x)e^{\lambda(x)}$ and $\hat{\varphi}(y) = e^{\eta(y)}$. The constraint $R \in \mathcal{R}_s(\mu_0, \mu_1)$ can be written as

$$\begin{aligned} \varphi(x) \int_{\mathbb{R}^d} p(0,x,1,y)\hat{\varphi}(y)dy &= \rho_0(x), \quad x \in \mathbb{R}^d, \\ \hat{\varphi}(y) \int_{\mathbb{R}^d} \varphi(x)p(0,x,1,y)dx &= \rho_1(y), \quad y \in \mathbb{R}^d. \end{aligned} \tag{4.3.4}$$

Then, consider a generalized version

$$\begin{aligned} \mu_0^*(dx) \int_{\mathbb{R}^d} p(0,x,1,y)\mu_1^*(dy) &= \mu_0(dx), \\ \mu_1^*(dy) \int_{\mathbb{R}^d} p(0,x,1,y)\mu_0^*(dx) &= \mu_1(dy), \end{aligned} \tag{4.3.5}$$

of (4.3.4). The system (4.3.5) of equations is called the *Schrödinger's system* or *Schrödinger's functional equation*. A solution of the Schrödinger's system is thus a pair (μ_0^*, μ_1^*) of σ-finite measures on $\mathbb{R}^d$ satisfying (4.3.5). Note that if a solution (μ_0^*, μ_1^*) of the Schrödinger's system exists then it satisfies $\mu_0 \ll \mu_0^*$ and $\mu_1 \ll \mu_1^*$, and $(\lambda\mu_0^*, \lambda^{-1}\mu_1^*)$ is also a solution for any $\lambda > 0$.

Actually the Schrödinger's system admits a solution. A proof of the following result can be found in [23].

Theorem 4.15

There exists a solution (μ_0^*, μ_1^*) of the Schrödinger's system (4.3.5) such that $\mu_0^* \sim \mu_0$ and $\mu_1^* \sim \mu_1$. Moreover, (μ_0^*, μ_1^*) is uniquely determined up to positive transformation, i.e., if $(\bar{\mu}_0^*, \bar{\mu}_1^*)$ is another solution, then $\mu_0^* = \lambda\bar{\mu}_0^*$ and $\mu_1^* = \lambda^{-1}\bar{\mu}_1^*$ for some $\lambda > 0$.

Hereafter, denote by (μ_0^*, μ_1^*) the unique solution of (4.3.5) in sense of the theorem above. Further, assume that μ_1 is equivalent to the Lebesgue measure. Then, since $\mu_1^* \sim \mu_1$, the Radon-Nikodym derivative of μ_1^* with respect to the Lebesgue measure exists. So define the functions φ_1 on $\mathbb{R}^d$ by

$$\varphi_1(y) = \frac{d\mu_1^*}{dy}(y), \quad y \in \mathbb{R}^d.$$

We further assume that φ_1 is continuous on $\mathbb{R}^d$. Then,

$$h(t,x) = \mathbb{E}_{\mathbb{P}}[\varphi_1(X_1^{t,x})], \quad (t,x) \in [0,1] \times \mathbb{R}^d$$

solves the Cauchy problem

$$\begin{aligned} \partial_t h(t,x) + \frac{1}{2}\sigma^2 \Delta h(t,x) = 0, \quad (t,x) \in [0,1) \times \mathbb{R}^d, \\ h(1,x) = \varphi_1(x), \quad x \in \mathbb{R}^d \end{aligned} \tag{4.3.6}$$

(see Section 1.3). By (4.3.5), we have

$$h(0,x) = \int_{\mathbb{R}^d} p(0,x,1,y)\varphi_1(y)dy = \int_{\mathbb{R}^d} p(0,x,1,y)\mu_1^*(dy) = \frac{d\mu_0}{d\mu_0^*}(x) > 0, \tag{4.3.7}$$

whence

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}\left[\frac{h(1,X_1)}{h(0,X_0)}\right] &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{\varphi_1(y)}{h(0,x)} p(0,x,1,y)dy\mu_0(dx) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} p(0,x,1,y)\mu_1^*(dy)\mu_0^*(dx) \\ &= \mu_0(\mathbb{R}^d) = 1. \end{aligned} \tag{4.3.8}$$

Thus, we define the probability measure $\mathbb{P}^*$ on $(\Omega, \mathcal{F})$ by

$$\frac{d\mathbb{P}^*}{d\mathbb{P}} = \frac{h(1,X_1)}{h(0,X_0)}.$$

Let $u^* = \{u_t^*\}_{0\le t\le 1}$ and $W^* = \{W_t^*\}_{0\le t\le 1}$ be the processes defined respectively by

$$u_t^* = \sigma^2 D \log h(t,X_t), \quad W_t^* = W_t - \frac{1}{\sigma}\int_0^t u_s^* ds.$$

Theorem 4.16

Suppose that μ_1 is equivalent to the Lebesgue measure and φ_1 is a bounded continuous function on $\mathbb{R}^d$. Moreover, suppose that

$$H(R^*|P_{01}) < \infty,$$

where $R^*(dxdy) = p(0,x,1,y)\mu_0^*(dx)\mu_1^*(dy)$. Then, $\pi^* := (\mathbb{P}^*, W^*, u^*) \in \Pi$ and is optimal to the problem (S).

Proof. Step (i). By (4.3.6) and Itô formula,

$$dh(t, X_t) = \sigma Dh(t, X_t)^\top dW_t = \sigma h(t, X_t) D \log h(t, X_t)^\top dW_t = \frac{1}{\sigma} h(t, X_t)(u_t^*)^\top dW_t.$$

Thus,

$$\frac{d\mathbb{P}^*}{d\mathbb{P}} = \exp\left[\frac{1}{\sigma}\int_0^1 (u_t^*)^\top dW_t - \frac{1}{2\sigma^2}\int_0^1 |u_t^*|^2 dt\right].$$

This together with Girsanov's theorem yields that $\{W_t^*\}$ is a d-dimensional $\mathbb{F}$-Brownian motion under $\mathbb{P}^*$. Further, the controlled process $X_t^{u^*}$, $0 \le t \le 1$, is given by

$$X_t^{u^*} = X_0 + \int_0^t u_s^* ds + \sigma W_t^* = X_t.$$

Hence, the underlying process $\{X_t\}$ is seen as the controlled diffusion with input u^* under $\mathbb{P}^*$. By (4.3.5) and (4.3.6), for $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned}\mathbb{P}^*(X_0 \in A) &= \mathbb{E}_{\mathbb{P}}\left[1_{\{X_0 \in A\}} \frac{h(1, X_1)}{h(0, X_0)}\right] = \int_A \frac{1}{h(0,x)} \left(\int_{\mathbb{R}^d} \varphi_1(y) p(0, x, 1, y) dy\right) \mu_0(dx) \\ &= \mu_0(A),\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}^*(X_1 \in A) &= \mathbb{E}_{\mathbb{P}}\left[1_{\{X_1 \in A\}} \frac{h(1, X_1)}{h(0, X_0)}\right] = \int_A \int_{\mathbb{R}^d} \frac{d\mu_0^*}{d\mu_0}(x) p(0, x, 1, y) \mu_0(dx) \mu_1^*(dy) \\ &= \mu_1(A).\end{aligned}$$

Therefore the controlled process $\{X_t^{u^*}\}$ satisfies the distribution constraints.

Next we will prove $\mathbb{E}_{\mathbb{P}^*} \int_0^1 |u_t^*|^2 dt < \infty$, which leads to $\pi^* \in \Pi$. To do so, consider the stopping times $\tau_n := \inf\{t > 0; |u_t^*| > n\}$, $n \in \mathbb{N}$. Then, for each n define the probability measure $\mathbb{P}_n$ by

$$\begin{aligned}\frac{d\mathbb{P}_n}{d\mathbb{P}} :&= \frac{h(1 \wedge \tau_n, X_{1\wedge\tau_n})}{h(0, X_0)} = \exp\left[\int_0^1 (\psi_t^{(n)})^\top dW_t - \frac{1}{2}\int_0^1 |\psi_t^{(n)}|^2 dt\right] \\ &= \exp\left[\int_0^1 (\psi_t^{(n)})^\top dW_t^* + \frac{1}{2}\int_0^1 |\psi_t^{(n)}|^2 dt\right],\end{aligned}$$

where $\psi_t^{(n)} = (1/\sigma) u_t^* 1_{\{t \le \tau_n\}}$. By the monotone convergence theorem,

$$\mathbb{E}_{\mathbb{P}^*} \int_0^1 |u_t^*|^2 dt = \sigma^2 \lim_{n\to\infty} \mathbb{E}_{\mathbb{P}^*}\left[\int_0^1 |\psi_t^{(n)}|^2 dt\right] = \sigma^2 \lim_{n\to\infty} \mathbb{E}_{\mathbb{P}^*}\left[\log \frac{d\mathbb{P}_n}{d\mathbb{P}}\right]. \tag{4.3.9}$$

On the other hand, since the relative entropy is nonnegative, we obtain

$$\begin{aligned}\mathbb{E}_{\mathbb{P}^*}\left[\log \frac{d\mathbb{P}_n}{d\mathbb{P}}\right] &\le \mathbb{E}_{\mathbb{P}^*}\left[\log \frac{d\mathbb{P}_n}{d\mathbb{P}}\right] + \mathbb{E}_{\mathbb{P}^*}\left[\log \frac{d\mathbb{P}^*}{d\mathbb{P}_n}\right] = \mathbb{E}_{\mathbb{P}^*}\left[\log \frac{d\mathbb{P}^*}{d\mathbb{P}}\right] \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{d\mu_0^*}{d\mu_0}(x) \varphi_1(y) \log\left(\frac{d\mu_0^*}{d\mu_0}(x)\varphi_1(y)\right) p(0, x, 1, y) dy \mu_0(dx) \\ &= H(R^* | P_{01}) < \infty.\end{aligned}$$

From this and (4.3.9) the announced result follows.

Step (ii). We will prove the optimality of π^*. Let $\pi = (\mathbb{Q}, B, u) \in \Pi$ be arbitrary. Then the process

$$\hat{W}_t := B_t + \frac{1}{\sigma}\int_0^t u_s ds, \quad 0 \le t \le 1,$$

is a d-dimensional $\mathbb{F}$-Brownian motion under $\hat{\mathbb{P}}$ defined by

$$\frac{d\hat{\mathbb{P}}}{d\mathbb{Q}} = \exp\left[-\frac{1}{\sigma}\int_0^1 u_t^\mathsf{T} dB_t - \frac{1}{2\sigma^2}\int_0^1 |u_t|^2 dt\right].$$

Since the controlled process X_t^u satisfies $X_t^u = X_0 + \sigma\hat{W}_t$, the distribution of X^u under $\hat{P}$ is the same as that of X under $\mathbb{P}$. Hence,

$$\begin{aligned}1 &= \mathbb{E}_\mathbb{P}\left[\frac{h(1,X_1)}{h(0,X_0)}\right] = \mathbb{E}_{\hat{\mathbb{P}}}\left[\frac{h(1,X_1^u)}{h(0,X_0^u)}\right] = \mathbb{E}_\mathbb{Q}\left[\frac{h(1,X_1^u)}{h(0,X_0^u)}\exp\left[-\frac{1}{\sigma}\int_0^1 u_t^\mathsf{T} dB_t - \frac{1}{2\sigma^2}\int_0^1 |u_t|^2 dt\right]\right] \\ &\geq \exp\left\{\mathbb{E}_\mathbb{Q}\left[\log\frac{h(1,X_1^u)}{h(0,X_0^u)} - \frac{1}{\sigma}\int_0^1 u_t^\mathsf{T} dB_t - \frac{1}{2\sigma^2}\int_0^1 |u_t|^2 dt\right]\right\},\end{aligned}$$

where we have used Jensen's inequality in the last inequality. Therefore,

$$\begin{aligned}&\frac{1}{2\sigma^2}\mathbb{E}_\mathbb{Q}\left[\int_0^1 |u_t|^2 dt\right] \\ &\geq \mathbb{E}_\mathbb{Q}\left[\log\frac{h(1,X_1^u)}{h(0,X_0^u)}\right] = \mathbb{E}_\mathbb{Q}\left[\log\frac{d\mu_0}{d\mu_0^*}(X_0) + \log\varphi_1(X_1^u)\right] \\ &= \int_{\mathbb{R}^d}\left(\log\frac{d\mu_0}{d\mu_0^*}(x)\right)\mu_0^*(dx) + \int_{\mathbb{R}^d}(\log\varphi_1(y))\,\mu_1(dy) \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left(\log\frac{d\mu_0}{d\mu_0^*}(x)\right)p(0,x,1,y)\mu_0^*(dx)\mu_1^*(dy) + \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}(\log\varphi_1(y))\,p(0,x,1,y)\mu_0^*(dx)\mu_1^*(dy) \\ &= H(R^*|P_{01}) = \frac{1}{2\sigma^2}\mathbb{E}_{\mathbb{P}^*}\left[\int_0^1 |u_t^*|^2 dt\right].\end{aligned}$$

Consequently, we deduce that π^* is optimal to (S). ☐

Proposition 4.17

Suppose that $\mu_0(dx) = \delta_{x_0}(dx)$ for some $x_0 \in \mathbb{R}^d$ and $\mu_1(dy) = \rho(y)dy$ for some positive, bounded, and continuous function ρ satisfying

$$\int_{\mathbb{R}^d}\rho(y)\left\{|y|^2 + \log\rho(y)\right\}dy < \infty. \tag{4.3.10}$$

Then, the pair (μ_0^*, μ_1^*) of σ-finite measures defined respectively by

$$\mu_0^*(dx) = \mu_0(dx), \quad \mu_1^*(dy) = \frac{\rho(y)}{p(0,x_0,1,y)}dy$$

is a solution of the Schrödinger's system (4.3.5). Moreover,

$$J^* = H(R^*|P_{01}) = \int_{\mathbb{R}^d}\rho(y)\log\left(\frac{\rho(y)}{p(0,x_0,1,y)}\right)dy < \infty.$$

Proof. By definition, we have

$$\frac{d\mu_0^*}{d\mu_0}(x) = 1, \quad \frac{d\mu_1^*}{d\mu_1}(y) = \frac{1}{p(0,x_0,1,y)}.$$

Thus, for any $A \in \mathcal{B}(\mathbb{R}^d)$ with $A \ni x_0$,

$$\int_A\int_{\mathbb{R}^d}p(0,x,1,y)\mu_1^*(dy)\mu_0^*(dx) = \int_{\mathbb{R}^d}p(0,x_0,1,y)\frac{\mu_1(dy}{p(0,x_0,1,y)} = 1 = \mu_0(A).$$

Further, for any $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\int_B \int_{\mathbb{R}^d} p(0,x,1,y)\mu_0^*(dx)\mu_1^*(dy) = \int_B p(0,x_0,1,y)\frac{\mu_1(dy)}{p(0,x_0,1,y)} = \mu_1(B).$$

Therefore (μ_0^*, μ_1^*) satisfies (4.3.5).

As confirmed in the proof of the theorem above, $J^* = H(R|P_{01})$. Under the present assumptions,

$$\begin{aligned} H(R^*|P_{01}) &= \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{\rho(y)}{p(0,x_0,1,y)} \log\left(\frac{\rho(y)}{p(0,x_0,1,y)}\right) p(0,x,1,y)dy\delta_{x_0}(dx) \\ &= \int_{\mathbb{R}^d} \rho(y)\log\rho(y)dy - \int_{\mathbb{R}^d} \rho(y)\log p(0,x_0,1,y)dy < \infty. \end{aligned}$$

Thus the proposition follows. □

Example 4.18. Consider the one dimensional case where $\mu_0(dx) = \delta_0(dx)$ and $\mu_1(dy) = \rho(y)dy$ with

$$\rho(y) = \theta\rho_1(y) + (1-\theta)\rho_2(y), \quad y \in \mathbb{R}.$$

Here, $0 \le \theta \le 1$ and

$$\rho_i(y) = \frac{e^{-|y-m_i|^2/(2v_i)}}{\sqrt{2\pi v_i}}, \quad i = 1, 2,$$

with $m_i \in \mathbb{R}$, $0 < v_i < \sigma^2$. Namely, ρ is the density of a Gaussian mixture distribution. Let us confirm that ρ satisfies the integrability condition (4.3.10). Since the function $F(y) := y\log y - y + 1$ is nonnegative and convex, we find

$$0 \le \int_{\mathbb{R}} \rho(y)\log\rho(y)dy = \int_{\mathbb{R}} F(\rho(y))dy \le \theta\int_{\mathbb{R}} \rho_1(y)\log\rho_1(y)dy + (1-\theta)\int_{\mathbb{R}} \rho_2(y)\log\rho_2(y)dy.$$

It is easy to check that each ρ_i satisfies (4.3.10). Thus ρ satisfies (4.3.10).

Let us give an explicit representation of the drift term u_t^* of the optimal controlled diffusion. Since $\varphi_1(y) = \rho(y)/p(0,0,1,y)$, we have

$$\begin{aligned} h(t,x) &= \mathbb{E}_{\mathbb{P}}[\varphi_1(X_1^{t,x})] \\ &= \theta\int_{\mathbb{R}} \frac{\rho_1(x+\sigma\sqrt{1-t}z)}{p(0,0,1,x+\sigma\sqrt{1-t}z)}\frac{e^{-z^2/2}}{\sqrt{2\pi}}dz + (1-\theta)\int_{\mathbb{R}} \frac{\rho_2(x+\sigma\sqrt{1-t}z)}{p(0,0,1,x+\sigma\sqrt{1-t}z)}\frac{e^{-z^2/2}}{\sqrt{2\pi}}dz. \end{aligned}$$

A tedious computation gives

$$\frac{d}{dx}\log h(t,x) = -\frac{\theta\alpha_1(t)((\frac{1}{v_1}-\frac{1}{\sigma^2})x+\frac{m_1}{v_1})\gamma_1(t)e^{-g_1(t,x)} + (1-\theta)\alpha_2(t)((\frac{1}{v_2}-\frac{1}{\sigma^2})x+\frac{m_2}{v_2})\gamma_2(t)e^{-g_2(t,x)}}{\theta\alpha_1(t)e^{-g_1(x)} + (1-\theta)\alpha_2(t)e^{-g_2(x)}},$$

where

$$\begin{aligned} \gamma_i(t) &= \frac{1}{\left(\frac{1}{v_i}-\frac{1}{\sigma^2}\right)\sigma^2(1-t)+1}, \\ \alpha_i(t) &= \frac{1}{\sqrt{v_i\gamma_i(t)}}, \\ g_i(t,x) &= \frac{1}{2}\left(\frac{1}{v_i}-\frac{1}{\sigma^2}\right)\gamma_i(t)x^2 - \frac{m_i}{v_i}\gamma_i(t)x + \frac{m_i^2}{2v_i} - \frac{m_i^2}{2v_i}\sigma^2(1-t)\gamma_i(t). \end{aligned}$$

Finally, we prove the auxiliary result that claims the equivalence between the dynamic and static Schrödinger's problems.

Proposition 4.19

Assume that P is the law of the process $X_t = X_0 + \sigma W_t$, where $\sigma > 0$. Suppose that $\inf_{R\in\mathcal{P}_s(\mu_0,\mu_1)} H(R|P_{01}) < \infty$. Then,

$$\inf_{Q\in\mathcal{P}(\mu_0,\mu_1)} H(Q|P) = \inf_{R\in\mathcal{P}_s(\mu_0,\mu_1)} H(R|P_{01}). \tag{4.3.11}$$

Proof.* First we will show that for $Q \ll P$,

$$\frac{dQ_{01}}{dP_{01}}(X_0, X_1) = \mathbb{E}_P\left[\frac{dQ}{dP}\middle| X_0, X_1\right], \quad P\text{-a.s.} \tag{4.3.12}$$

To this end, take an arbitrary $A \in \sigma(X_0, X_1)$. Then $A = \{(X_0, X_1) \in B\}$ for some $B \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d)$. We observe

$$\mathbb{E}_P\left[\frac{dQ_{01}}{dP_{01}}(X_0,X_1)1_A\right] = \int_B \frac{dQ_{01}}{dP_{01}}(z)P_{01}(dz) = Q_{01}(B) = \mathbb{E}_Q[1_{\{(X_0,X_1)\in B\}}] = \mathbb{E}_P\left[1_A\frac{dQ}{dP}\right],$$

leading to (4.3.12).

Next, represent the probability measure P by the disintegration formula

$$P(\Gamma) = \int_{\mathbb{R}^d\times\mathbb{R}^d} (P_{01})_z(\Gamma)P_{01}(dz), \quad \Gamma \in \mathcal{B}(\mathbb{W}^d),$$

where $\{(P_{01})_z\}_{z\in\mathbb{R}^d\times\mathbb{R}^d}$ is the family of probability measures on $(\mathbb{W}^d, \mathcal{B}(\mathbb{W}^d))$ as in Theorem 1.10. Then, take $R \in \mathcal{P}_s(\mu_0, \mu_1)$ and consider

$$Q(\Gamma) := \int_{\mathbb{R}^d\times\mathbb{R}^d} (P_{01})_z(\Gamma)R(dz), \quad \Gamma \in \mathcal{B}(\mathbb{W}^d).$$

Note that $Q \in \mathcal{P}(\mu_0, \mu_1)$. For $\Gamma \in \mathcal{B}(\mathbb{W}^d)$, by $R \ll P_{01}$ and Theorem 1.10,

$$Q(\Gamma) = \mathbb{E}_P\left[(P_{01})_z(\Gamma)\frac{dR}{dP_{01}}(z)\bigg|_{z=(X_0,X_1)}\right] = \mathbb{E}_P\left[\mathbb{E}_P[1_\Gamma|X_0,X_1]\frac{dR}{dP_{01}}(X_0,X_1)\right] = \mathbb{E}_P\left[1_\Gamma\frac{dR}{dP_{01}}(X_0,X_1)\right].$$

This means that $Q \ll P$ and

$$\frac{dQ}{dP}(w) = \frac{dQ_{01}}{dP_{01}}(w(0), w(1)), \quad w = (w(t))_{0\le t\le 1} \in \mathbb{W}^d, P\text{-a.e.}$$

Hence $H(Q|P) = H(R|P_{01})$. Therefore,

$$\inf_{Q'\in\mathcal{P}(\mu_0,\mu_1)} H(Q'|P) \le \inf_{R'\in\mathcal{P}_s(\mu_0,\mu_1)} H(R'|P_{01}). \tag{4.3.13}$$

On the other hand, for $Q \in \mathcal{P}(\mu_0, \mu_1)$ with $H(Q|P) < \infty$, applying Jensen's inequality for the conditional expectation for the convex function $f(y) = y\log y - y + 1$, $y \in (0, \infty)$, we have

$$\begin{aligned} H(Q|P) &= \mathbb{E}_Q\left[f\left(\frac{dQ}{dP}\right)\right] = \mathbb{E}_Q\left[\mathbb{E}_Q\left[f\left(\frac{dQ}{dP}\right)\middle| X_0, X_1\right]\right] \\ &\ge \mathbb{E}\left[f\left(\mathbb{E}_Q\left[\frac{dQ}{dP}\middle| X_0, X_1\right]\right)\right] = \mathbb{E}_{Q_{01}}\left[\log\frac{dQ_{01}}{dP_{01}}\right] \ge \inf_{R\in\mathcal{P}_s(\mu_0,\mu_1)} H(R|P_{01}). \end{aligned}$$

The last inequality and (4.3.13) leads to (4.3.11). $\square$

CHAPTER 5

Controlled Diffusions and Viscosity Solutions

As seen at the end of Chapter 4, in general we cannot expect the existence of smooth solutions of HJB equations. The *viscosity solutions* are the most useful and elegant notion for weak solutions of nonlinear elliptic and parabolic partial differential equations (PDEs), as well as open up the possibility of rigorous numerical analysis of HJB equations whose classical solutions might not exist. In this chapter, we describe basic parts in the theory of viscosity solutions. We refer to Crandall et.al [8], [11], [32], and [36] for more detailed accounts.

5.1 Dynamic Programming Principle

The *dynamic programming principle* (DPP) by Bellman [3] gives a recursive method of solving optimal control problems. In discrete-time framework, by the dynamic programming, we can directly obtain optimal control processes at least theoretically. In continuous-time, the situation is slightly different, and the DPP leads to nonlinear partial differential equations for the stochastic control problems, so-called *Hamilton-Jacobi-Bellman* (HJB) *equations*. This section is devoted to the statement and the proof of the DPP under mild assumptions, and in the next section, the connection between the DPP and HJB equations is discussed.

Consider the stochastic control problem (4.1.6). Here we assume that the following is satisfied:

Assumption 5.1

(i) The set A is compact and convex in $\mathbb{R}^{d_1}$.

(ii) For each $\phi = b, \sigma, f$, the function ϕ is continuous on $[0,T] \times \mathbb{R}^d \times A$.

(iii) There exists a positive constant C_0 such that for each $\phi = b, \sigma, f$ and for every $(t, t', x, x', a, a') \in [0,T]^2 \times (\mathbb{R}^d)^2 \times A^2$,

$$|\phi(t,x,a) - \phi(t',x',a')| \leq C_0|t-t'|^{1/2} + C_0|x-x'| + C_0|a-a'|,$$
$$|\phi(t,x,a)| \leq C_0.$$

(iv) The function g is bounded and uniformly continuous on $\mathbb{R}^d$.

- It follows from Assumption 5.1 that (4.1.7) and (4.1.8) holds. Thus, by Theorem 4.3, there exists a unique solution $\{X_s^{t,x,\alpha}\}_{t\leq s\leq T}$ of (4.1.1) with initial condition $X_t^{t,x,\alpha} = x$ for any

$(t,x) \in [0,T] \times \mathbb{R}^d$ and for any A-valued and adapted process α.

- The above fact together with the boundedness of g and f shows that $\mathcal{A}$ is the set of all A-valued and adapted processes.
- We take here $\tilde{\mathcal{A}} = \mathcal{A}$.

The preceding arguments show that the value function

$$v(t,x) := \inf_{\alpha \in \mathcal{A}} \mathbb{E}\left[g(X_T^{t,x,\alpha}) + \int_t^T f(s, X_s^{t,x,\alpha}, \alpha_s)ds\right], \quad (t,x) \in [0,T] \times \mathbb{R}^d, \tag{5.1.1}$$

is real-valued. Moreover, Lemma 5.4 below means that v is bounded and Borel measurable.

In addition to Assumption 5.1, we make the following assumption:

Assumption 5.2

The filtration $\mathbb{F}$ is the augmented one generated by $\{W_t\}_{0 \le t \le T}$.

Now the DPP is stated as follows:

Theorem 5.3

Suppose that Assumptions 5.1 and 5.2 hold. Let v be as in (5.1.1). Then, for any $t, s \in [0,T]$ with $t \le s$ and $x \in \mathbb{R}^d$ we have

$$v(t,x) = \inf_{\alpha \in \mathcal{A}} \mathbb{E}\left[v(s, X_s^{t,x,\alpha}) + \int_t^s f(r, X_r^{t,x,\alpha}, \alpha_r)dr\right].$$

- Assumptions 5.1 and 5.2 can be weakened. See Krylov [22] for the DPP under a more general setting.

The rest of this section is devoted to the proof of Theorem 5.3. There are several variations for the proof of the DPP and all of them are lengthy and technical. Our proof is close to the one in Nisio [30] and can be skipped on a first reading.

To obtain Theorem 5.3, we need several preliminary results. First we show the uniform continuity of the value function.

Lemma 5.4

Under Assumptions 5.1 and 5.2, the value function v is uniformly continuous on $[0,T] \times \mathbb{R}^d$.

Proof. Let $s, t \in [0,T]$ with $s \ge t$, $x, y \in \mathbb{R}^d$ and $\alpha \in \mathcal{A}$. We write C for positive constants that do not depends on particular points in $[0,T] \times \mathbb{R}^d \times \mathcal{A}$ and may vary from line to line. First observe, for $r \ge s$,

$$\begin{aligned} X_r^{t,x,\alpha} - X_r^{s,y,\alpha} = x - y &+ \int_t^s b(u, X_u^{t,x,\alpha}, \alpha_u)du + \int_t^s \sigma(u, X_u^{t,x,\alpha}, \alpha_u)dW_u \\ &+ \int_s^r \left[b(u, X_u^{t,x,\alpha}, \alpha_u) - b(u, X_u^{s,y,\alpha}, \alpha_u)\right] du \\ &+ \int_s^r \left[\sigma(u, X_u^{t,x,\alpha}, \alpha_u) - \sigma(u, X_u^{s,y,\alpha}, \alpha_u)\right] dW_u. \end{aligned}$$

From this and Assumption 5.1, we obtain

$$\mathbb{E}|X_r^{t,x,\alpha} - X_r^{s,y,\alpha}|^2 \le C|x-y|^2 + C(s-t) + C\int_s^r \mathbb{E}\left|X_u^{t,x,\alpha} - X_u^{s,y,\alpha}\right|^2 du.$$

Thus, by Gronwall's lemma,

$$\sup_{s\le r\le T}\mathbb{E}|X_r^{t,x,\alpha}-X_r^{s,y,\alpha}|^2\le C|x-y|^2+C|s-t|. \tag{5.1.2}$$

Now, by Assumption 5.1,

$$\begin{aligned}&|v(t,x)-v(s,y)|\\&\le\sup_{\alpha\in\mathcal{A}}\mathbb{E}\bigg[|g(X_T^{t,x,\alpha})-g(X_T^{s,y,\alpha})|+\int_t^s|f(r,X_r^{t,x,\alpha},\alpha_r)|dr\\&\qquad+\int_s^T|f(r,X_r^{t,x,\alpha},\alpha_r)-f(r,X_r^{s,y,\alpha},\alpha_r)|dr\bigg]\\&\le C\sup_{\alpha\in\mathcal{A}}\bigg[\mathbb{E}|g(X_T^{t,x,\alpha})-g(X_T^{s,y,\alpha})|+(s-t)+\int_s^T\mathbb{E}|X_r^{t,x,\alpha}-X_r^{s,y,\alpha}|dr\bigg].\end{aligned}$$

Since g is uniformly continuous, for $\varepsilon>0$ there exists $\delta_0>0$ such that $|g(z)-g(z')|<\varepsilon$ whenever $z,z'\in\mathbb{R}^d$ satisfy $|z-z'|<\delta_0$. Thus, by (5.1.2),

$$\begin{aligned}\mathbb{E}|g(X_T^{t,x,\alpha})-g(X_T^{s,y,\alpha})|&=\mathbb{E}\left[|g(X_T^{t,x,\alpha})-g(X_T^{s,y,\alpha})|1_{\{|X_T^{t,x,\alpha}-X_T^{s,y,\alpha}|<\delta_0\}}+1_{\{|X_T^{t,x,\alpha}-X_T^{s,y,\alpha}|\ge\delta_0\}}\right]\\&\le\varepsilon+\frac{C}{\delta_0^2}\mathbb{E}|X_T^{t,x,\alpha}-X_T^{s,y,\alpha}|^2\le\varepsilon+\frac{C}{\delta_0^2}(|s-t|+|x-y|^2),\end{aligned}$$

whence

$$|v(t,x)-v(s,y)|\le C\left(\varepsilon+\frac{1}{\delta_0^2}(|s-t|+|x-y|^2)+|s-t|+|x-y|\right)\le C\varepsilon$$

whenever $|x-y|,|s-t|<\delta_1:=\delta_0\sqrt{\varepsilon}\wedge\delta_0^2\varepsilon\wedge\varepsilon$. Thus the lemma follows. □

Lemma 5.5

Suppose that Assumptions 5.1 and 5.2 hold. For any $s,t\in[0,T]$ with $s\ge t$, $\mathcal{F}_t$-measurable random variable $\xi\in L^2$, and $\alpha\in\mathcal{A}$, there exists a Borel measurable map $F_{s,t}$ on $L^2\times\mathcal{L}_2\times C([t,s];\mathbb{R}^d)$ such that

$$X_s^{t,\xi,\alpha}=F_{s,t}(\xi,\alpha,(W_r-W_t)_{t\le r\le s}),\quad \text{a.s.}$$

Proof. Fix $s,t\in[0,T]$ with $s>t$, $\mathcal{F}_t$-measurable random variable $\xi\in L^2$, and $\alpha\in\mathcal{A}$.

Step (i). For any $n\in\mathbb{N}$, put

$$\mathcal{A}_n=\{\beta\in\mathcal{A}:\beta(r)=\alpha(t_{k,n})\text{ for }r\in[t_{k,n},t_{k+1,n}),\ \ k=0,1,\dots,2^n-1\},$$

where $t_{k,n}=t+(s-t)k2^{-n}$, and $\tilde{\mathcal{A}}=\cup_{n=1}^\infty\mathcal{A}_n$. Here we have denoted $\beta_r=\beta(r)$ just for notational convenience. Then, as in the proof of Lemma 2.3 we can show that there exists $\{\alpha^{(n)}\}\subset\tilde{\mathcal{A}}$ such that

$$\lim_{n\to\infty}\mathbb{E}\int_t^s|\alpha_r-\alpha_r^{(n)}|^2dr=0. \tag{5.1.3}$$

To prove (5.1.3), put $\alpha_r=\alpha_0$ for $r\le 0$. Then define the adapted process $\{\beta_r^{(N)}\}_{0\le r\le T}$ with continuous paths by

$$\beta_r^{(N)}=2^N\int_{r-2^{-N}}^r\alpha_u du,\quad N\in\mathbb{N}.$$

Notice that $\beta^{(N)} \in \mathcal{A}$ since A is assumed to be compact and convex. Moreover, since $\beta^{(N)}$ is differentiable a.e., we have $\beta_r^{(N)} \to \alpha_r$, $dt \times \mathbb{P}$-a.e. This together with the boundedness of α yields

$$\mathbb{E}\int_t^T |\alpha_r - \beta_r^{(N)}|^2 \to 0, \quad N \to \infty.$$

Further, put $\beta_r^{(N,\ell)} = \beta^{(N)}(t_{k,\ell})$ for $r \in [t_{k,n}, t_{k+1,n})$ and $\beta_r^{(N,\ell)} = \beta_r^{(N)}$ for $r \in [0,t) \cup [s,T]$, $\ell \in \mathbb{N}$. Then, again $\beta^{(N,\ell)} \in \mathcal{A}$ for each N, ℓ and $\lim_{\ell\to\infty} \beta_r^{N,\ell} = \beta_r^{(N)}$ for any r and N by the continuity of $\beta^{(N)}$. Consequently, we obtain

$$\lim_{N\to\infty}\lim_{\ell\to\infty} \mathbb{E}\int_t^s |\alpha_r - \beta_r^{(N,\ell)}|^2 dr = 0.$$

This means that there exists a sequence $\{(N_n, \ell_n)\}_{n=1}^\infty$ such that $N_n, \ell_n \to \infty$ as $n \to \infty$ and that

$$\lim_{n\to\infty} \mathbb{E}\int_t^s |\alpha_r - \beta_r^{(N_n,\ell_n)}|^2 dr = 0.$$

Thus the process $\alpha_r^{(n)} := \beta_r^{(N_n,\ell_n)}$, $0 \le r \le T$, is the one we aim to construct.

Step (ii). Consider the sequence $\{Y_k^{(n)}\}_{k=0}^{K_n}$ of the random variables defined by

$$Y_{k+1}^{(n)} = Y_k^{(n)} + b(t_k, Y_k^{(n)}, \alpha_{t_k})(t_{k+1} - t_k) + \sigma(t_k, Y_k^{(n)}, \alpha_{t_k})(W_{t_{k+1}} - W_{t_k})$$

for $k = 0, 1, \dots, K_n - 1$ with $Y_0^{(n)} = \xi$. Here we have denoted $K_n = 2^{\ell_n}$ and $t_k = t_{k,2^{\ell_n}}$ for notational simplicity. That is, $\{Y_k^{(n)}\}$ is the Euler-Maruyama approximation of $\{X_r^{t,x,\alpha^{(n)}}\}$. Then, as in the proof of Theorem 3.14,

$$\lim_{n\to\infty} \mathbb{E}[X_s^{t,\xi,\alpha^{(n)}} - Y_{K_n}^{(n)}|^2 = 0.$$

Further, it is now straightforward to see

$$\mathbb{E}|X_s^{t,\xi,\alpha} - X_s^{t,\xi,\alpha^{(n)}}|^2 \le C\mathbb{E}\int_t^s |\alpha_r - \alpha_r^{(n)}|^2 dr$$

for some constant $C > 0$. Therefore, using (5.1.3), we obtain

$$\lim_{n\to\infty} Y_{K_n}^{(n)} = X_s^{t,x,\alpha}, \quad \text{a.s.,} \tag{5.1.4}$$

possibly along a subsequence.

On the other hand, by an inductive argument, $Y_{K_n}^{(n)}$ turns out to be $\sigma(\xi, \alpha, (W_r - W_t)_{t\le r\le s})$-measurable. This and (5.1.4) together with Theorem 1.9 lead to the claim. □

For $(t, x, \alpha) \in [0,T] \times \mathbb{R}^d \times \mathcal{A}$ we write

$$J(t,x,\alpha) = \mathbb{E}\left[g(X_T^{t,x,\alpha}) + \int_t^T f(s, X_s^{t,x,\alpha}, \alpha_s)ds\right].$$

Then of course $v(t,x) = \inf_{\alpha\in\mathcal{A}} J(t,x,\alpha)$, $(t,x) \in [0,T] \times \mathbb{R}^d$. Further, consider the set $\mathcal{A}_t$ of the controls $\alpha \in \mathcal{A}$ such that $\alpha_s = G_s((W_r - W_t)_{t\le r\le s})$ a.s. for some Borel measurable map G_s on $C([t,s];\mathbb{R}^d)$ for each $s \in [t,T]$. Then we have the following:

Lemma 5.6

Under Assumptions 5.1 and 5.2,

$$v(t,x) = \inf_{\alpha\in\mathcal{A}_t} J(t,x,\alpha), \quad (t,x) \in [0,T] \times \mathbb{R}^d.$$

Proof. By Assumption 5.2 and Theorem 1.9, any $\alpha \in \mathcal{A}$ can be represented as $\alpha_s = \tilde{\alpha}_s := G_{\alpha,s}(\{W_r\}_{0\le r\le s})$ a.s., $s \in [0,T]$, for some Borel function $G_{\alpha,s}$ on $C([t,s];\mathbb{R}^d)$. Using the Itô isometry, we find

$$\int_t^s \sigma(r, X_r^{t,x,\alpha}, \alpha_r)dW_r = \int_t^s \sigma(r, X_r^{t,x,\alpha}, \tilde{\alpha}_r)dW_r, \quad t \le s \le T, \text{ a.s.}$$

for any $t \in [0,T]$. This means that $X_s^{t,x,\alpha} = X_s^{t,x,\tilde{\alpha}}$, $t \le s \le T$, a.s. Further, by Lemma 5.5, we find that for any $s \in [t,T]$, there exists a Borel function $F_{s,t}$ on $\mathbb{R}^d \times \mathcal{L}_2 \times C([t,s];\mathbb{R}^d)$ and $\tilde{G}_{\alpha,r}$ on $C([0,t];\mathbb{R}^d) \times C([t,s];\mathbb{R}^d)$, $t \le r \le s$, such that

$$X_s^{t,x,\alpha} = F_{s,t}(x, \tilde{G}_{\alpha,\cdot}(\{W_r\}_{0\le r\le t}, \{W_r - W_t\}_{t\le r\le \cdot}), \{W_r - W_t\}_{t\le r\le s}), \quad \text{a.s.}$$

This together with the tower property of the conditional expectations yields

$$\begin{aligned}
\mathbb{E}[g(X_T^{t,x,\alpha})] &= \mathbb{E}\left[\mathbb{E}\left[g(X_T^{t,x,\tilde{\alpha}})\middle| \mathcal{F}_t\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[g(F_{T,t}(x, \tilde{G}_{\alpha,\cdot}(\{W_r\}_{0\le r\le t}, \{W_r - W_t\}_{t\le r\le \cdot}), \{W_r - W_t\}_{t\le r\le T}))\middle| \mathcal{F}_t\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[g(F_{T,t}(x, \tilde{G}_{\alpha,\cdot}(\phi, \{W_r - W_t\}_{t\le r\le \cdot}), \{W_r - W_t\}_{t\le r\le T}))\right]\Big|_{\phi=\{W_r\}_{0\le r\le t}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[g(X_T^{t,x,\beta(\phi)})\right]\Big|_{\phi=\{W_r\}_{0\le r\le t}}\right],
\end{aligned}$$

where $\beta(\phi) = \tilde{G}_{\alpha,\cdot}(\phi, \{W_r - W_t\}_{t\le r\le \cdot})$. Similarly, we obtain

$$\mathbb{E}[f(s, X_s^{t,x,\alpha}, \alpha_s)] = \mathbb{E}\left[\mathbb{E}[f(s, X_s^{t,x,\beta(\phi)}, \beta(\phi)_s)]\Big|_{\phi=\{W_r\}_{0\le r\le t}}\right].$$

Thus, since $\beta(\phi) \in \mathcal{A}_t$, we deduce

$$J(t,x,\alpha) = \mathbb{E}\left[\mathbb{E}[J(t,x,\beta(\phi))]|_{\phi=\{W_r\}_{0\le r\le t}}\right] \ge \mathbb{E}\left[\inf_{\alpha'\in\mathcal{A}_t} J(t,x,\alpha')\right] = \inf_{\alpha'\in\mathcal{A}_t} J(t,x,\alpha'),$$

whence $v(t,x) \ge \inf_{\alpha'\in\mathcal{A}_t} J(t,x,\alpha')$. The converse inequality is obvious from $\mathcal{A}_t \subset \mathcal{A}$. Thus the lemma follows. □

Proof of Theorem 5.3. Fix $s,t \in [0,T]$ with $s \ge t$, and $x \in \mathbb{R}^d$. By the uniqueness, $X_r^{t,x,\alpha} = X_r^{s,X_s^{t,x,\alpha},\alpha}$ a.s. for $r \in [s,T]$ and for $\alpha \in \mathcal{A}$. As in the proof of Lemma 5.6,

$$\begin{aligned}
\mathbb{E}[g(X_T^{t,x,\alpha})] &= \mathbb{E}\left[\mathbb{E}\left[g(X_T^{s,X_s^{t,x,\alpha},\alpha})\middle| \mathcal{F}_s\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[g(F_{T,s}(\xi, \tilde{G}_{\alpha,\cdot}(\phi, \{W_r - W_s\}_{s\le r\le \cdot}), \{W_r - W_s\}_{s\le r\le T}))\right]\Big|_{\xi=X_\theta^{t,x,\alpha},\phi=\{W_r\}_{t\le r\le s}}\right] \\
&= \mathbb{E}\left[\mathbb{E}[g(X_T^{s,\xi,\beta(\phi)})]\Big|_{\xi=X_s^{t,x,\alpha},\phi=\{W_r\}_{t\le r\le s}}\right].
\end{aligned}$$

where $\beta(\phi) = \tilde{G}_{\alpha,\cdot}(\phi, (W_r - W_s)_{s\le r\le \cdot})$. Similarly,

$$\mathbb{E}\left[f(r, X_r^{t,x,\alpha}, \alpha_r)\right] = \mathbb{E}\left[\mathbb{E}\left[f(r, X_r^{s,\xi,\beta(\phi)}, \beta(\phi)_r)\right]\Big|_{\xi=X_s^{t,x,\alpha},\phi=\{W_r\}_{t\le r\le s}}\right].$$

Hence, for $\alpha \in \mathcal{A}$,

$$\begin{aligned} J(t,x,\alpha) &= \mathbb{E}[g(X_T^{t,x,\alpha})] + \mathbb{E}\int_t^s f(r,X_r^{t,x,\alpha},\alpha_r)dr + \int_s^T \mathbb{E}[f(r,X_r^{t,x,\alpha},\alpha_r)]dr \\ &= \mathbb{E}\left[\mathbb{E}[J(s,X_s^{t,x,\beta(\phi)})]|_{\phi=\{W_r\}_{t\le r\le s}}\right] + \mathbb{E}\int_t^s f(r,X_r^{t,x,\alpha},\alpha_r)dr \\ &\ge \mathbb{E}\left[v(s,X_s^{t,x,\alpha})\right] + \mathbb{E}\int_t^s f(r,X_r^{t,x,\alpha},\alpha_r)dr, \end{aligned}$$

whence

$$v(t,x) \ge \inf_{\alpha\in\mathcal{A}} \mathbb{E}\left[v(s,X_s^{t,x,\alpha}) + \int_t^s f(r,X_r^{t,x,\alpha},\alpha_r)dr\right].$$

To prove the converse inequality, let $\varepsilon > 0$ be arbitrary and take $\delta > 0$ so that

$$|v(s,y) - v(s,y')| \le \varepsilon, \quad \sup_{\alpha\in\mathcal{A}_s} |J(s,y,\alpha) - J(s,y',\alpha)| \le \varepsilon \tag{5.1.5}$$

whenever $y, y' \in \mathbb{R}^d$ satisfy $|y - y'| \le \delta$. This is possible due to Lemma 5.4 and its proof.

Let $\{B_n\}_{n=1}^\infty \subset \mathcal{B}(\mathbb{R}^d)$ be a disjoint partition of $\mathbb{R}^d$ such that $\mathrm{diam}(B_n) \le \delta$. Then, for every n, take $x_n \in B_n$ arbitrary. For this x_n there exists $\alpha_n \in \mathcal{A}_s$ such that

$$v(s,x_n) \ge J(s,x_n,\alpha^n) - \varepsilon.$$

From this and (5.1.5) it follows that, for each n,

$$J(s,y,\alpha^n) \le v(s,y) + 3\varepsilon, \quad y \in B_n. \tag{5.1.6}$$

Now, fix $\alpha \in \mathcal{A}$ and define $\bar\alpha \in \mathcal{A}_s$ by

$$\bar\alpha_r = \alpha_r 1_{\{r\le s\}} + 1_{\{r>s\}} \sum_{n=1}^\infty \alpha_r^n 1_{B_n}(X_s^{t,x,\alpha}), \quad 0 \le r \le T.$$

Since each α^n is independent of $\mathcal{F}_s$, as in the proof of Lemma 5.6,

$$J(t,x,\bar\alpha) = \sum_{n=1}^\infty \mathbb{E}\left[J(s,X_s^{t,x,\alpha},\alpha^n)1_{B_n}(X_s^{t,x,\alpha})\right] + \mathbb{E}\int_t^s f(r,X_r^{t,x,\alpha},\alpha_r)dr.$$

This and (5.1.6) yield

$$v(t,x) \le \mathbb{E}\left[v(s,X_s^{t,x,\alpha}) + \int_t^s f(r,X_r^{t,x,\alpha},\alpha_r)dr\right] + 3\varepsilon,$$

leading to the inequality we wanted. □

5.2 Definition

Let F be a real-valued function on $[0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d$, and consider the PDE

$$F(t,x,v(t,x),\partial_t v(t,x),Dv(t,x),D^2v(t,x)) = 0, \quad (t,x) \in [0,T) \times \mathbb{R}^d. \tag{5.2.1}$$

We are mainly interested with the case where F is of the form

$$F(t,x,u,q,p,M) = -q + \sup_{a\in A}\left[-b(t,x,a)^\mathsf{T} p - \frac{1}{2}\mathrm{tr}(\sigma(t,x,a)\sigma(t,x,a)^\mathsf{T} M) - f(t,x,a)\right], \tag{5.2.2}$$

which is the case of HJB equations.

- The function F is assumed to satisfy the *ellipticity condition*:

$$F(t, x, u, q, p, M_1) \geq F(t, x, u, q, p, M_2), \quad (t, x, u, q, p) \in [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^d, \tag{5.2.3}$$

  for $M_1, M_2 \in \mathbb{S}^d$ with $M_1 \leq M_2$.

- For $A, B \in \mathbb{S}^d$ we write $A \leq B$ if $B - A$ is positive semi-definite.

- The function F is also assumed to satisfy the *parabolicity condition*:

$$F(t, x, u, q_1, p, M) \geq F(t, x, u, q_2, p, M), \quad (t, x, u, p, M) \in [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d, \tag{5.2.4}$$

  for $q_1, q_2 \in \mathbb{R}$ with $q_1 \leq q_2$.

- The nonlinearity F defined by (5.2.2) clearly satisfies (5.2.3) and (5.2.4).

To motivate the notion of viscosity solutions, let us assume that a classical subsolution v of (5.2.1) exists, i.e., (5.2.1) holds with $=$ replaced by $\leq$. Let $\varphi \in C^{1,2}([0, T] \times \mathbb{R}^d)$ and $(t, x) \in [0, T) \times \mathbb{R}^d$ be a global maximum point of $v - \varphi$. By adding a constant if necessary, we can always assume that $(v - \varphi)(t, x) = 0$. Then, we have the three conditions

$$\partial_t(v - \varphi)(t, x) \geq 0, \quad D(v - \varphi)(t, x) = 0, \quad D^2(v - \varphi)(t, x) \leq 0.$$

Note that the first inequality holds with equality if $t > 0$. From these conditions, (5.2.3) and (5.2.4) it follows that

$$F(t, x, \varphi(t, x), \partial_t\varphi(t, x), D\varphi(t, x), D^2\varphi(t, x)) \leq F(t, x, v(t, x), \partial_t v(t, x), Dv(t, x), D^2 v(t, x)) \leq 0.$$

Thus the subsolution property holds at (t, x) for the test function φ.

Similarly, let v be a classical supersolution of (5.2.1), i.e., v satisfy (5.2.1) with $=$ replaced by $\geq$. Then for any $\varphi \in C^{1,2}([0, T] \times \mathbb{R}^d)$ such that $\min_{(s,y) \in [0,T) \times \mathbb{R}^d}(v - \varphi)(s, y) = (v - \varphi)(t, x) = 0$,

$$F(t, x, \varphi(t, x), \partial_t\varphi(t, x), D\varphi(t, x), D^2\varphi(t, x)) \geq 0.$$

Definition 5.7. Let $F : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d \to \mathbb{R}$ satisfy (5.2.3) and (5.2.4), and let $u \in C([0, T] \times \mathbb{R}^d)$.

(i) We say that u is a *viscosity subsolution* of (5.2.1) if

$$F(t, x, \varphi(t, x), \partial_t\varphi(t, x), D\varphi(t, x), D^2\varphi(t, x)) \leq 0$$

for all $\varphi \in C^{1,2}([0, T] \times \mathbb{R}^d)$ and $(t, x) \in [0, T) \times \mathbb{R}^d$ such that $\max_{(s,y) \in [0,T) \times \mathbb{R}^d}(v - \varphi)(s, y) = (v - \varphi)(t, x) = 0$.

(ii) We say that u is a *viscosity supersolution* of (5.2.1) if

$$F(t, x, \varphi(t, x), \partial_t\varphi(t, x), D\varphi(t, x), D^2\varphi(t, x)) \geq 0$$

for all $\varphi \in C^{1,2}([0, T] \times \mathbb{R}^d)$ and $(t, x) \in [0, T) \times \mathbb{R}^d$ such that $\min_{(s,y) \in [0,T) \times \mathbb{R}^d}(v - \varphi)(s, y) = (v - \varphi)(t, x) = 0$.

(iii) We say that v is a *viscosity solution* if it is both a viscosity subsolution and a viscosity supersolution.

5.3 Comparison Principle

The *comparison principle* is a key property for uniqueness of viscosity solutions, and is an important ingredient in numerical analysis of fully nonlinear parabolic PDEs.

An equivalent definition of viscosity solutions

We need an alternative definition of viscosity solutions in terms of superjets and subjets. Observe that for $U \in C([0,T] \times \mathbb{R}^d)$, $\varphi \in C^{1,2}([0,T] \times \mathbb{R}^d)$, and $(t,x) \in [0,T) \times \mathbb{R}^d$ with $\max_{(s,y)\in[0,T)\times\mathbb{R}^d}(U-\varphi)(s,y) = (U-\varphi)(t,x)$, the Taylor expansion up to second order terms yields

$$\begin{aligned} U(s,y) &\le U(t,x) + \varphi(s,y) - \varphi(t,x) \\ &= U(t,x) + \partial_t\varphi(t,x)(s-t) + D\varphi(t,x)^\top(y-x) \\ &\quad + \frac{1}{2}(y-x)^\top D^2\varphi(t,x)(y-x) + o(|s-t| + |y-x|^2). \end{aligned}$$

This leads to the following definition: for $U \in C([0,T] \times \mathbb{R}^d)$ and $(t,x) \in [0,T) \times \mathbb{R}^d$, the set $\mathcal{P}^{2,+}U(t,x)$ is defined by

$$\mathcal{P}^{2,+}U(t,x) = \Bigg\{(q,p,M) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d : \liminf_{(h,y)\to 0} \frac{U(t+h,x+y) - U(t,x) - qh - p^\top y - \frac{1}{2}y^\top My}{|h| + |y|^2} \ge 0\Bigg\}.$$

Similarly, we define the set $\mathcal{P}^{2,-}U(t,x)$ by the

$$\mathcal{P}^{2,-}U(t,x) = \Bigg\{(q,p,M) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d : \limsup_{(h,y)\to 0} \frac{U(t+h,x+y) - U(t,x) - qh - p^\top y - \frac{1}{2}y^\top My}{|h| + |y|^2} \le 0\Bigg\}.$$

- The sets $\mathcal{P}^{2,+}U(t,x)$ and $\mathcal{P}^{2,-}U(t,x)$ are called the *superjet* and *subjet* of U at (t,x), respectively.
- Compare the definitions of the super/sub-jets with that of the subdifferential in convex analysis, if you are familiar with it.
- By definition, for $U \in C([0,T] \times \mathbb{R}^d)$, $\varphi \in C^{1,2}([0,T] \times \mathbb{R}^d)$, and $(t,x) \in [0,T) \times \mathbb{R}^d$ with $\max_{(s,y)\in[0,T)\times\mathbb{R}^d}(U-\varphi)(s,y) = (U-\varphi)(t,x)$,
$$\big(\partial_t\varphi(t,x), D\varphi(t,x), D^2\varphi(t,x)\big) \in \mathcal{P}^{2,+}U(t,x).$$
- The converse implication of the claim just above holds true, i.e., for any $(t,x) \in [0,T) \times \mathbb{R}^d$ and $(q,p,M) \in \mathcal{P}^{2,+}U(t,x)$, there exists $\varphi \in C^{1,2}([0,T]\times\mathbb{R}^d)$ satisfying $\max_{(s,y)\in[0,T)\times\mathbb{R}^d}(U-\varphi)(s,y) = (U-\varphi)(t,x)$ such that
$$(q,p,M) = \big(\partial_t\varphi(t,x), D\varphi(t,x), D^2\varphi(t,x)\big).$$
See [11, Lemma 4.1] for an explicit construction of such φ.
- A similar characterization holds for the subjet. Consequently, for given $(t,x) \in [0,T)\times\mathbb{R}^d$, a point $(q,p,M) \in \mathcal{P}^{2,-}U(t,x)$ if and only if there exists $\varphi \in C^{1,2}([0,T] \times \mathbb{R}^d)$ satisfying $\min_{(s,y)\in[0,T)\times\mathbb{R}^d}(U-\varphi)(s,y) = (U-\varphi)(t,x)$ such that
$$(q,p,M) = \big(\partial_t\varphi(t,x), D\varphi(t,x), D^2\varphi(t,x)\big).$$
- The closures of the subjets and superjets are theoretically useful. We define $\bar{\mathcal{P}}^{2,+}U(t,x)$ by the set of the points $(q,p,M) \in \mathbb{R}\times\mathbb{R}^d\times\mathbb{S}^d$ for which there exists $(t_n,x_n,q_n,p_n,M_n) \in [0,T) \times \mathbb{R}^d \times \mathcal{P}^{2,+}U(t,x)$, $n \in \mathbb{N}$, satisfying $(t_n,x_n,q_n,p_n,M_n) \to (t,x,q,p,M)$. The set $\bar{\mathcal{P}}^{2,-}U(t,x)$ is defined similarly.

With the preliminaries above, we have the following:

Proposition 5.8

Let $F : [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d \to \mathbb{R}$ be continuous and satisfy (5.2.3) and (5.2.4). Then $u \in C([0,T] \times \mathbb{R}^d)$ is a viscosity subsolution (resp. supersolution) of (5.2.1) if and only if for any $(t,x) \in [0,T) \times \mathbb{R}^d$ and $(q,p,M) \in \bar{\mathcal{P}}^{2,+}u(t,x)$ (resp. $(q,p,M) \in \bar{\mathcal{P}}^{2,-}u(t,x)$)

$$F(t,x,u(t,x),q,p,M) \le 0 \ \text{(resp.} \ge 0).$$

Comparison principle

The *Ishii's lemma* is a key to the proof of the comparison principle. Since the proof of this result is lengthy and technical for our introductory notes, we refer to Theorem 8.3 in [8] and [36] for details.

Lemma 5.9: Ishii's Lemma

Assume that $F : [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d \to \mathbb{R}$ is continuous and satisfies (5.2.4), and

$$F(t,x,u,q,p,M) = F(t,x,u,0,p,M) - q$$

for any $(t,x,u,q,p,M) \in [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d$. Let $U, V \in C_b([0,T] \times \mathbb{R}^d)$ be a viscosity subsolution and a viscosity supersolution of (5.2.1), respectively. Let $\phi \in C^{1,1,2,2}([0,T] \times [0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$ and $(\bar{t}, \bar{s}, \bar{x}, \bar{y}) \in [0,T) \times [0,T) \times \mathbb{R}^d \times \mathbb{R}^d$ be a local maximum of $U(t,x) - V(s,y) - \phi(t,s,x,y)$. Then, for every $\eta > 0$, there exist $M_1, M_2 \in \mathbb{S}^d$ such that

$$(\partial_t \phi(\bar{t}, \bar{s}, \bar{x}, \bar{y}), D_x \phi(\bar{t}, \bar{s}, \bar{x}, \bar{y}), M_1) \in \bar{\mathcal{P}}^{2,+}U(\bar{t}, \bar{x}),$$
$$(-\partial_s \phi(\bar{t}, \bar{s}, \bar{x}, \bar{y}), -D_y \phi(\bar{t}, \bar{s}, \bar{x}, \bar{y}), M_2) \in \bar{\mathcal{P}}^{2,-}U(\bar{t}, \bar{x}),$$

and

$$\begin{pmatrix} M_1 & 0 \\ 0 & -M_2 \end{pmatrix} \le D^2_{x,y}\phi(\bar{t}, \bar{s}, \bar{x}, \bar{y}) + \eta \left(D^2_{x,y}\phi(\bar{t}, \bar{s}, \bar{x}, \bar{y})\right)^2.$$

- The space $C^{1,1,2,2}([0,T] \times [0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$ is defined similarly as in the case of $C^{1,2}([0,T] \times \mathbb{R}^d)$.

Hereafter, we assume that the function $F : [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d \to \mathbb{R}$ is represented as

$$F(t,x,u,q,p,M) = -q + \beta u + \sup_{a \in A} \left[-b(t,x,a)^{\mathsf{T}} p - \frac{1}{2}\mathrm{tr}(\sigma(t,x,a)\sigma(t,x,a)^{\mathsf{T}} M) - f(t,x,a) \right] \tag{5.3.1}$$

for $(t,x,u,q,p,M) \in [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d$, where $\beta \in [0,\infty)$, the set A is a subset of $\mathbb{R}^{d_1}$, and each $\phi = b, \sigma, f$ satisfies that there exists a constant $C_0 > 0$ such that

$$|\phi(t,x,a) - \phi(s,y,a)| \le C_0|t-s| + C_0|x-y|$$

for $(t,s,x,y,a) \in [0,T] \times [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times A$.

Now we are ready to prove the comparison principle.

Theorem 5.10: Comparison principle

Suppose that (5.3.1) holds. Let $U, V \in C_b([0,T] \times \mathbb{R}^d)$ be a viscosity subsolution and a viscosity supersolution of (5.2.1), respectively. If $U(T,\cdot) \le V(T,\cdot)$ on $\mathbb{R}^d$, then $U \le V$ on $[0,T] \times \mathbb{R}^d$.

*Proof**. Step (i). Notice that for any $\kappa > 0$, the function $\tilde{U}(t,x) := e^{\kappa t}U(t,x)$, $(t,x) \in [0,T]\times\mathbb{R}^d$, is a viscosity subsolution of (5.2.1) with F replaced by

$$\tilde{F}(t,x,u,q,p,M) = -q+(\beta+\kappa)u+\sup_{a\in A}\left[-b(t,x,a)^\top p - \frac{1}{2}\mathrm{tr}(\sigma(t,x,a)\sigma(t,x,a)^\top M) - e^{\kappa t}f(t,x,a)\right].$$

Indeed, let $\tilde{\varphi} \in C^{1,2}([0,T]\times\mathbb{R}^d)$ and $(t,x) \in [0,T)\times\mathbb{R}^d$ be such that $\max_{(s,y)\in[0,T)\times\mathbb{R}^d}(\tilde{U} - \tilde{\varphi})(s,y) = (\tilde{U}-\tilde{\varphi})(t,x) = 0$, and put $\varphi(s,y) = e^{-\kappa s}\tilde{\varphi}(s,y)$, $(s,y) \in [0,T]\times\mathbb{R}^d$. Then,

$$(U-\varphi)(s,y) = e^{\kappa s}(\tilde{U}-\tilde{\varphi})(s,y) \le e^{\kappa s}(\tilde{U}-\tilde{\varphi})(t,x) = 0 = (U-\varphi)(t,x).$$

Thus, (t,x) is also a global minimum point of $U-\varphi$, whence by the subsolution property,

$$F(t,x,\varphi(t,x),D\varphi(t,x),D^2\varphi(t,x)) \le 0.$$

This together with $\partial_t\varphi(t,x) = e^{-\kappa t}(\partial_t\tilde{\varphi}(t,x) - \kappa\tilde{\varphi}(t,x))$ yields

$$\tilde{F}(t,x,\tilde{\varphi}(t,x),D\tilde{\varphi}(t,x),D^2\tilde{\varphi}(t,x)) \le 0.$$

Hence $\tilde{U}$ is a viscosity subsolution of $\tilde{F} = 0$. A similar relation holds for V, and so we may assume that $\beta > 0$ without loss of generality.

Step (ii). Set $\psi(t,x) = e^{-\lambda t}(1+|x|^2)$, $(t,x) \in [0,T]\times\mathbb{R}^d$, where $\lambda > 0$. Then, it is straightforward to see that for $(t,x) \in [0,T)\times\mathbb{R}^d$,

$$\begin{aligned}&\partial_t\psi(t,x) - \beta\psi(t,x) + \sup_{a\in A}\left[b(t,x,a)^\top D\psi(t,x) + \frac{1}{2}\mathrm{tr}(\sigma(t,x,a)\sigma(t,x,a)^\top D^2\psi(t,x))\right]\\ &\le e^{-\lambda t}(1+|x|^2)(-\lambda-\beta+c_1),\end{aligned} \tag{5.3.2}$$

for some positive constant c_1. Further, for $\delta > 0$ the function $V_\delta := V + \delta\psi$ is a viscosity supersolution of (5.2.1). Indeed, let $\varphi \in C^{1,2}([0,T]\times\mathbb{R}^d)$ and $(t,x) \in [0,T)\times\mathbb{R}^d$ be such that $\min_{[0,T)\times\mathbb{R}^d}(V_\delta - \varphi)(s,y) = (V_\delta-\varphi)(t,x) = 0$. Then $\min_{[0,T)\times\mathbb{R}^d}(V-\varphi_\delta)(s,y) = (V-\varphi_\delta)(t,x) = 0$, where $\varphi_\delta = \varphi - \delta\psi$. The viscosity supersolution property means that $F(t,x,\varphi_\delta(t,x),D\varphi_\delta(t,x),D^2\varphi_\delta(t,x)) \ge 0$. This and (5.3.2) with the choice $\lambda \ge -\beta + c_1$ yield

$$\begin{aligned}0 &\le F(t,x,\varphi_\delta(t,x),D\varphi_\delta(t,x),D^2\varphi_\delta(t,x))\\ &\le F(t,x,\varphi(t,x),D\varphi(t,x),D^2\varphi(t,x) + e^{-\lambda t}(1+|x|^2)(-\lambda-\beta+c_1)\\ &\le F(t,x,\varphi(t,x),D\varphi(t,x),D^2\varphi(t,x)),\end{aligned}$$

whence the claim.

Step (iii). We will show that $U(t,x) \le V_\delta(t,x)$ for all $(t,x) \in [0,T]\times\mathbb{R}^d$ and $\delta > 0$, which leads to the claim of the theorem. To this end, assume that $c := \sup_{(t,x)\in[0,T]\times\mathbb{R}^d}(U-V_\delta)(t,x) > 0$ for some $\delta > 0$. Since

$$\lim_{|x|\to\infty}\sup_{t\in[0,T]}(U-V_\delta)(t,x) = -\infty,$$

there exists a bounded open subset $\mathcal{O}$ of $\mathbb{R}^d$ such that

$$c = \max_{(t,x)\in[0,T)\times\mathcal{O}}(U-V_\delta)(t,x). \tag{5.3.3}$$

Take a sequence $(t_n,s_n,x_n,y_n) \in [0,T]\times[0,T]\times\overline{\mathcal{O}}\times\overline{\mathcal{O}}$, $n\in\mathbb{N}$, that maximizes the function Φ_n on $[0,T]\times[0,T]\times\overline{\mathcal{O}}\times\overline{\mathcal{O}}$ by

$$\Phi_n(t,s,x,y) = U(t,x) - V_\delta(s,y) - \phi_n(t,s,x,y)$$

with

$$\phi_n(t,s,x,y) = \frac{n}{2}\left(|t-s|^2 + |x-y|^2\right)$$

for any $n \in \mathbb{N}$, where $\overline{\mathcal{O}}$ denotes the closure of $\mathcal{O}$. Further, we write c_n for the maximum of Φ_n. Then we have

$$(c_n, \phi_n(t_n, s_n, x_n, y_n)) \to (c, 0), \quad n \to \infty. \tag{5.3.4}$$

To prove this, note that the bounded sequence $\{(t_n, s_n, x_n, y_n)\}_{n\in\mathbb{N}}$ converges to some $(\bar{t}, \bar{s}, \bar{x}, \bar{y}) \in [0,T] \times [0,T] \times \overline{\mathcal{O}} \times \overline{\mathcal{O}}$ possibly along a subsequence. Since $U(t_n, x_n) - V_\delta(s_n, y_n)$, $n \geq 1$, is bounded, it follows from

$$c \leq c_n \leq U(t_n, x_n) - V_\delta(s_n, y_n) - \phi_n(t_n, s_n, x_n, y_n) \leq U(t_n, x_n) - V_\delta(s_n, y_n)$$

that $\phi_n(t_n, s_n, x_n, y_n)$, $n \geq 1$, is also bounded. This means that $\bar{t} = \bar{s}$ and $\bar{x} = \bar{y}$, whence

$$c \leq \lim_{n\to\infty} (U(t_n, x_n) - V_\delta(s_n, y_n)) = U(\bar{t}, \bar{x}) - V_\delta(\bar{t}, \bar{x}) \leq c.$$

From this and (5.3.3) it follows that $c = U(\bar{t}, \bar{x}) - V(\bar{t}, \bar{x})$ and $(\bar{t}, \bar{x}) \in [0,T) \times \mathcal{O}$, which leads to (5.3.4).

Step (iv). Since (t_n, s_n, x_n, y_n) converges to $(\bar{t}, \bar{t}, \bar{x}, \bar{x}) \in [0,T) \times [0,T) \times \mathcal{O} \times \mathcal{O}$ possibly along a subsequence, we may assume that $(t_n, s_n, x_n, y_n) \in [0,T) \times [0,T) \times \mathcal{O} \times \mathcal{O}$ for all n. We apply Lemma 5.9 with these points, ϕ_n's, and $\eta = 1/n$. Direct computation gives $\partial_t \phi_n(t_n, s_n, x_n, y_n) = -\partial_s \phi_n(t_n, s_n, x_n, y_n) = n(t_n - s_n)$ and $D_x \phi_n(t_n, s_n, x_n, y_n) = -D_y \phi_n(t_n, s_n, x_n, y_n) = n(x_n - y_n)$. Thus there exist $M_1, M_2 \in \mathbb{S}^d$ such that $(n(t_n - s_n), n(x_n - y_n), M_1) \in \bar{\mathcal{P}}^{2,+}U(x_n, y_n)$ and $(n(t_n - s_n), n(x_n - y_n), M_2) \in \bar{\mathcal{P}}^{2,-}V_\delta(x_n, y_n)$. Proposition 5.8 now implies that

$$\begin{aligned}
-n(t_n - s_n) + \beta U(t_n, x_n) + F(t_n, x_n, 0, 0, n(x_n - y_n), M_1) \leq 0,\\
-n(t_n - s_n) + \beta V_\delta(s_n, y_n) + F(s_n, y_n, 0, 0, n(x_n - y_n), M_2) \geq 0,
\end{aligned}$$

so that

$$\begin{aligned}
&\beta(U(t_n, x_n) - V_\delta(s_n, y_n))\\
&\leq F(s_n, y_n, 0, 0, n(x_n - y_n), M_2) - F(t_n, x_n, 0, 0, n(x_n - y_n), M_1)\\
&\leq C\phi_n(t_n, s_n, x_n, y_n) + \frac{1}{2}\sup_{a\in A}\left[\mathrm{tr}(\sigma(s_n, y_n, a)\sigma(s_n, y_n, a)^\mathsf{T} M_2) - \mathrm{tr}(\sigma(t_n, x_n, a)\sigma(t_n, x_n, a)^\mathsf{T} M_1)\right]
\end{aligned}$$

for some constant $C > 0$. Here we have used (5.3.1) to derive the last inequality. By the Ishii's lemma and

$$D^2_{x,y}\phi_n(\bar{t}, \bar{s}, \bar{x}, \bar{y}) = n\begin{pmatrix} I_d & -I_d \\ -I_d & I_d \end{pmatrix},$$

we obtain

$$\begin{pmatrix} M_1 & 0 \\ 0 & -M_2 \end{pmatrix} \leq 3n\begin{pmatrix} I_d & -I_d \\ -I_d & I_d \end{pmatrix}.$$

This and (5.3.1) yield

$$\begin{aligned}
&\mathrm{tr}(\sigma(s_n, y_n, a)\sigma(s_n, y_n, a)^\mathsf{T} M_2) - \mathrm{tr}(\sigma(t_n, x_n, a)\sigma(t_n, x_n, a)^\mathsf{T} M_1)\\
&= \mathrm{tr}\left(\Sigma\begin{pmatrix} M_1 & 0 \\ 0 & -M_2 \end{pmatrix}\right) \leq 3n\mathrm{tr}\left(\Sigma\begin{pmatrix} I_d & -I_d \\ -I_d & I_d \end{pmatrix}\right)\\
&= 3n\mathrm{tr}\left((\sigma(s_n, y_n, a) - \sigma(t_n, x_n, a))(\sigma(s_n, y_n, a) - \sigma(t_n, x_n, a))^\mathsf{T}\right)\\
&= 3n|\sigma(s_n, y_n, a) - \sigma(t_n, x_n, a)|^2 \leq C\phi_n(t_n, s_n, x_n, y_n)
\end{aligned}$$

for some constant $C' > 0$ uniformly on A, where

$$\Sigma = \begin{pmatrix} \sigma(s_n, y_n, a)\sigma(s_n, y_n, a)^\top & \sigma(s_n, y_n, a)\sigma(t_n, x_n, a)^\top \\ \sigma(t_n, x_n, a)\sigma(s_n, y_n, a)^\top & \sigma(t_n, x_n, a)\sigma(t_n, x_n, a)^\top \end{pmatrix}.$$

Therefore,

$$\beta(U(t_n, x_n) - V_\delta(s_n, y_n)) \le C''\phi_n(t_n, s_n, x_n, y_n)$$

for some $C'' > 0$, whence by (5.3.4) we have $c \le 0$, a contradiction. □

5.4 HJB Equations in the Viscosity Sense

Recall that the value function v of the stochastic control problem in Section 5.1 is given by

$$v(t, x) = \inf_{\alpha \in \mathcal{A}} \mathbb{E}\left[g(X_T^{t,x,\alpha}) + \int_t^T f(s, X_s^{t,x,\alpha}, \alpha_s)ds\right], \quad (t, x) \in [0, T] \times \mathbb{R}^d.$$

The corresponding HJB equation is

$$-\partial_t V(t, x) + \sup_{a \in A}\left[-b(t, x, a)^\top DV(t, x) - \frac{1}{2}\mathrm{tr}((\sigma\sigma^\top)(t, x)D^2V(t, x)) - f(t, x, a)\right] = 0, \quad (5.4.1)$$

on $[0, T) \times \mathbb{R}^d$ with terminal condition $v(T, x) = g(x)$, $x \in \mathbb{R}^d$.

Theorem 5.11

Suppose that Assumptions 5.1 and 5.2 hold. Let v be defined by (5.1.1). Then v is a unique viscosity solution of (5.4.1) satisfying $v(T, \cdot) = g$ on $\mathbb{R}^d$.

Proof. First note that $v \in C_b([0, T] \times \mathbb{R}^d)$ by Assumption 5.1 and Lemma 5.4. Let $\varphi \in C^{1,2}([0, T] \times \mathbb{R}^d)$ and $(t, x) \in [0, T) \times \mathbb{R}^d$ that is a global maximum of $v - \varphi$ with $v(t, x) = \varphi(t, x)$. For this φ we define the function ϕ on $[0, T] \times \mathbb{R}^d$ by

$$\phi(s, y) = \varphi(s, y)\zeta(s, y) + 2 \sup_{(s', y') \in [0,T] \times \mathbb{R}^d} |v(s', y')|(1 - \zeta(s, y)), \quad (s, y) \in [0, T] \times \mathbb{R}^d,$$

where $\zeta \in C_0^\infty([0, T] \times \mathbb{R}^d)$ is such that $0 \le \zeta \le 1$ on $[0, T] \times \mathbb{R}^d$, $\zeta = 1$ on $B_1(t, x)$, and $\zeta = 0$ on $\mathbb{R}^d \setminus B_2(t, x)$. Then, $\phi \in C_b^{1,2}([0, T] \times \mathbb{R}^d)$ and

$$(v - \phi)(s, y) = (v - \varphi)(s, y)\zeta(s, y) + \left(v - 2 \sup_{[0,T] \times \mathbb{R}^d} |v|\right)(s, y)(1 - \zeta(s, y)) \le 0 = (v - \phi)(t, x).$$

Applying Theorem 5.3 and Itô formula for ϕ, we see

$$\begin{aligned} \phi(t, x) = v(t, x) &\le \mathbb{E}\left[\phi(t + h, X_{t+h}^{t,x,a}) + \int_t^{t+h} f(s, X_s^{t,x,a}, a)ds\right] \\ &= \mathbb{E}\Bigg[\phi(t, x) + \int_t^{t+h} \left[\partial_t\phi(s, X_s^{t,x,a}) + H^a(s, X_s^{t,x,a}, D\phi(s, X_s^{t,x,a}), D^2\phi(s, X_s^{t,x,a}))\right] ds \\ &\qquad + \int_t^{t+h} D\phi(s, X_s^{t,x,a})^\top \sigma(s, X_s^{t,x,a})dW_s\Bigg] \end{aligned}$$

for any $a \in A$. Since σ is bounded and $\phi \in C_b^{1,2}([0, T] \times \mathbb{R}^d)$, the expectation of the Itô integral term in the inequality just above vanishes. Then, dividing the both side by h and letting $h \to 0$, we obtain

$$\partial_t\phi(t, x) + H^a(t, x, D\phi(t, x), D^2\phi(t, x) \le 0, \quad a \in A,$$

whence

$$-\partial_t\varphi(t,x) - \inf_{a\in A} H^a(t,x,D\varphi(t,x),D^2\varphi(t,x)) \le 0.$$

Thus v is a viscosity subsolution.

Let $\varphi \in C^{1,2}([0,T]\times\mathbb{R}^d)$ and $(t,x)\in[0,T)\times\mathbb{R}^d$ that is a global minimum of $v-\varphi$ with $v(t,x)=\varphi(t,x)$. As in above, we can modify φ to be in $C_b^{1,2}([0,T]\times\mathbb{R}^d)$. By Theorem 5.3, for any $\varepsilon>0$ there exists $\alpha^\varepsilon\in\mathcal{A}$ such that

$$v(t,x)+h\varepsilon \ge \mathbb{E}\left[v(t+h, X_{t+h}^{t,x,\alpha^\varepsilon}) + \int_t^{t+h} f(s, X_s^{t,x,\alpha^\varepsilon}, \alpha_s^\varepsilon)ds\right].$$

The condition on φ and the Itô formula yield

$$\begin{aligned}\varepsilon &\ge \frac{1}{h}\mathbb{E}\int_t^{t+h}\left[\partial_t\varphi(s,X_s^{t,x,\alpha^\varepsilon}) + H^{\alpha^\varepsilon}(s,X_s^{t,x,\alpha^\varepsilon}, D\varphi(s,X_s^{t,x,\alpha^\varepsilon}), D^2\varphi(s,X_s^{t,x,\alpha^\varepsilon}))\right]ds\\ &\ge \frac{1}{h}\mathbb{E}\int_t^{t+h}\left[\partial_t\varphi(s,X_s^{t,x,\alpha^\varepsilon}) + \inf_{a\in A} H^{a}(s,X_s^{t,x,\alpha^\varepsilon}, D\varphi(s,X_s^{t,x,\alpha^\varepsilon}), D^2\varphi(s,X_s^{t,x,\alpha^\varepsilon}))\right]ds.\end{aligned}$$

Since $D^2\varphi$ is uniformly continuous by the modification as in above, the function

$$s\mapsto \mathbb{E}\inf_{a\in A} H^a(s,X_s^{t,x,\alpha^\varepsilon}, D\varphi(s,X_s^{t,x,\alpha^\varepsilon}), D^2\varphi(s,X_s^{t,x,\alpha^\varepsilon}))$$

is continuous on $[t,t+h]$. Indeed, by Assumption 5.1, $\varphi\in C_b^{1,2}([0,T]\times\mathbb{R}^d)$ and the uniform continuity of $D^2\varphi$, for $\varepsilon_1>0$ there exits $\delta>0$ such that

$$\left|\inf_{a\in A} H^a(s,y,D\varphi(s,y),D^2\varphi(s,y)) - \inf_{a\in A} H^a(s',y',D\varphi(s',y'),D^2\varphi(s',y'))\right| < \varepsilon_1$$

whenever $|s-s'|+|y-y'|<\delta$. From this and the arguments as in the proof of Lemma 5.4 we find

$$\begin{aligned}&\left|\mathbb{E}\inf_{a\in A} H^a(s,X_s^{t,x,a},D\varphi(s,X_s^{t,x,a}),D^2\varphi(s,X_s^{t,x,a}))\right.\\ &\left.-\,\mathbb{E}\inf_{a\in A} H^a(s',X_{s'}^{t,x,a},D\varphi(s',X_{s'}^{t,x,a}),D^2\varphi(s',X_{s'}^{t,x,a}))\right|\\ &\le \varepsilon_1 + C\frac{1}{\delta^2}\sup_{a\in A}\mathbb{E}|X_s^{t,x,a}-X_{s'}^{t,x,a}|^2 \le \varepsilon_1 + \frac{C'}{\delta^2}|s-s'| \le (1+C')\varepsilon_1\end{aligned}$$

whenever $|s-s'|<\delta_1 := \delta^2\varepsilon\wedge\delta$, where C and C' are some positive constants. Thus the required continuity follows.

Then using the mean-value theorem and letting $h\to 0$, we have

$$\varepsilon \ge \partial_t\varphi(t,x) + \inf_{a\in A} H^a(t,x,D\varphi(t,x),D^2\varphi(t,x)),$$

whence letting $\varepsilon\to 0$,

$$-\partial_t\varphi(t,x) - \inf_{a\in A} H^a(t,x,D\varphi(t,x),D^2\varphi(t,x)) \ge 0.$$

Thus v is a viscosity supersolution.

The uniqueness immediately follows from the comparison principle and the boundary condition. ☐

Theorem 5.11 and the definition of viscosity solutions lead to the following corollary:

Corollary 5.12

Suppose that Assumptions 5.1 and 5.2 hold. If the function v defined by (5.1.1) is in $C^{1,2}([0,T]\times\mathbb{R}^d)$, then v is a unique classical solution of the HJB equation (4.2.1).

5.5 Approximation of Viscosity Solutions

Suppose that we want to prove that a given family $\{v_n\}$ of functions converges to a solution v of the nonlinear PDE (5.2.1). In that case, of course we cannot execute a routine error analysis by assuming a smoothness of v. Thus we are led to work in the framework of viscosity solutions. Then, it is often difficult to know a priori that the limit $\lim_{n\to\infty} v_n$ indeed exists and is continuous if it exists. The notion of *discontinuous viscosity solution* is useful in handling these technical problems.

Discontinuous Viscosity Solutions

Let u be bounded function on $[0,T]\times\mathbb{R}^d$. We define the *upper semi-continuous envelope* u^* of u by

$$u^*(t,x) = \limsup_{\substack{(s,y)\to(t,x)\\(s,y)\in[0,T]\times\mathbb{R}^d}} u(s,y), \quad (t,x)\in[0,T]\times\mathbb{R}^d,$$

and the *lower semi-continuous envelope* u_* of u by

$$u_*(t,x) = \liminf_{\substack{(s,y)\to(t,x)\\(s,y)\in[0,T]\times\mathbb{R}^d}} u(s,y), \quad (t,x)\in[0,T]\times\mathbb{R}^d.$$

- u^* is the smallest upper semi-continuous (u.s.c.) function that is greater than or equal to u.
- u_* is the largest lower semi-continuous (l.s.c.) function that is smaller than or equal to u.

Definition 5.13. Let $F:[0,T]\times\mathbb{R}^d\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{S}^d\to\mathbb{R}$ satisfy (5.2.3) and (5.2.4), and let $u:[0,T]\times\mathbb{R}^d\to\mathbb{R}$ be bounded.

(i) We say that u is a *discontinuous viscosity subsolution* of (5.2.1) if

$$F(t,x,\varphi(t,x),\partial_t\varphi(t,x),D\varphi(t,x),D^2\varphi(t,x)) \le 0$$

for all $\varphi\in C^{1,2}([0,T]\times\mathbb{R}^d)$ and $(t,x)\in[0,T)\times\mathbb{R}^d$ such that $\max_{(s,y)\in[0,T)\times\mathbb{R}^d}(v^*-\varphi)(s,y) = (v^*-\varphi)(t,x)=0$.

(ii) We say that u is a *discontinuous viscosity supersolution* of (5.2.1) if

$$F(t,x,\varphi(t,x),\partial_t\varphi(t,x),D\varphi(t,x),D^2\varphi(t,x)) \ge 0$$

for all $\varphi\in C^{1,2}([0,T]\times\mathbb{R}^d)$ and $(t,x)\in[0,T)\times\mathbb{R}^d$ such that $\min_{(s,y)\in[0,T)\times\mathbb{R}^d}(v_*-\varphi)(s,y) = (v_*-\varphi)(t,x)=0$.

(iii) We say that v is a *discontinuous viscosity solution* if it is both a discontinuous viscosity subsolution and a discontinuous viscosity supersolution.

Under the framework of the discontinuous viscosity solutions, we still have the comparison principle.

Theorem 5.14: Comparison principle

Suppose that (5.3.1) holds. Let $U,V:[0,T]\times\mathbb{R}^d\to\mathbb{R}$ be a bounded discontinuous viscosity subsolution and a bounded discontinuous viscosity supersolution of (5.2.1), respectively. If $U(T,\cdot)\le V(T,\cdot)$ on $\mathbb{R}^d$, then $U\le V$ on $[0,T]\times\mathbb{R}^d$.

Suppose that (5.3.1) holds. Suppose moreover that for a given bounded function u the upper semi-continuous envelope u^* is discontinuous viscosity subsolutions of (5.3.1) satisfying

$u^*(T,\cdot) \le g$ on $\mathbb{R}^d$ and the lower semi-continuous envelope u_* is discontinuous viscosity supersolutions of (5.3.1) satisfying $u_*(T,\cdot) \ge g$ on $\mathbb{R}^d$. Then by the comparison theorem, $u^* \le u_*$ on $[0,T]\times\mathbb{R}^d$. However, by definition, $u_* \le u^*$, and so $u^* = u_*$. This means that $u := u^* = u_*$ is a continuous viscosity solution of (5.2.1). Further by the comparison theorem for continuous viscosity solutions (Theorem 5.10), the uniqueness follows. Consequently, u is a unique continuous viscosity solution.

Barles–Souganidis Method

The abstract method given in Barles and Souganidis [2] is a powerful tool for checking the convergence of a given family of functions to a unique viscosity solution. Let $F : [0,T]\times\mathbb{R}^d\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{S}^d \to \mathbb{R}$. Further, let $\mathcal{C}$ be a class of bounded functions such that $C_b^2(\mathbb{R}^d) \subset \mathcal{C}$, and $\{\Phi^h\}_{h\in(0,1]}$ a family of operators such that $\Phi^h : \mathcal{C} \to \mathcal{C}$, $h \in (0,1]$.

Assume that F satisfies (5.2.3) and (5.2.4), and that the comparison principle holds.

Assumption 5.15

Let $U, V : [0,T]\times\mathbb{R}^d \to \mathbb{R}$ be a bounded discontinuous viscosity subsolution and a bounded discontinuous viscosity supersolution of (5.2.1), respectively. If $U(T,\cdot) \le V(T,\cdot)$ on $\mathbb{R}^d$, then $U \le V$ on $[0,T]\times\mathbb{R}^d$.

Now consider the terminal value problem (5.2.1) with $v(T,\cdot) = g$ on $\mathbb{R}^d$, where $g \in \mathcal{C}$. Suppose that we construct the family $\{v^h(t_k,\cdot)\}_{h\in(0,1]} \subset \mathcal{C}$, $k = 0,\dots,n$, such that

$$\begin{aligned} &v^h(t_k,x) = \Phi^h[v^h(t_{k+1},\cdot)](x), \quad k = 0,\dots,n-1, \ \ x \in \mathbb{R}^d, \\ &v^h(t_n,x) = g(x), \quad x \in \mathbb{R}^d. \end{aligned} \tag{5.5.1}$$

Here, $t_k = kT/n$ for $k = 0,\dots,n$. We assume that $\{t_k\}_{k=0}^n$ is described by the parameter h and that $\Delta t := T/n \to 0$, as $h \to 0$.

Then we make the following conditions on our scheme:

Assumption 5.16

(i) *Monotonicity*:
$$\Phi^h[\phi](x) \le \Phi^h[\psi](x), \quad x \in \mathbb{R}^d$$
for any $\phi, \psi \in \mathcal{C}$ with $\phi \le \psi$ on $\mathbb{R}^d$.

(ii) *Stability*:
$$\sup_{h\in(0,1]} \sup_{x\in\mathbb{R}^d} |v^h(t_k,x)| < \infty, \quad k = 0,\dots,n.$$

(iii) *Consistency I*: for any $(t,x) \in [0,T)\times\mathbb{R}^d$ and $\phi \in C_b^{1,2}([0,T]\times\mathbb{R}^d)$, we have
$$\begin{aligned} &\lim_{\substack{(s,y)\to(t,x) \\ h\to 0,\ c\to 0}} \frac{1}{\Delta t}\Big(\phi(s,y) + c - \Phi^h[\phi(s+\Delta t,\cdot) + c](y)\Big) \\ &= F(t,x,\partial_t\phi(t,x),\phi(t,x),D\phi(t,x),D^2\phi(t,x)) = 0. \end{aligned}$$

(iv) *Consistency II*: for $x \in \mathbb{R}^d$,
$$\lim_{\substack{(t_k,y)\to(T,x) \\ h\to 0}} v^h(t_k,y) = g(x).$$

Here is our main result in this section.

Theorem 5.17

Suppose that Assumptions 5.16 and 5.15 hold. Let v^h, $h \in (0,1]$, be as in (5.5.1). Then, there exists a unique continuous viscosity solution v of (5.2.1), and for any $t \in [0,T]$,

$$\lim_{h\to 0,\, t_k \to t} v^h(t_k, x) = v(t,x),$$

uniformly on any compact subset of $\mathbb{R}^d$.

Proof. We consider

$$\overline{v}(t,x) = \limsup_{\substack{(t_k,y)\to(t,x)\\ h\searrow 0}} v^h(t_k,y), \quad (t,x) \in [0,T]\times\mathbb{R}^d$$

and show that $\overline{v}$ is a discontinuous viscosity subsolution of (5.2.1). Let $\varphi \in C^{1,2}([0,T]\times\mathbb{R}^d)$ such that $\overline{v}-\varphi$ has a global maximum at $(t,x)\in[0,T)\times\mathbb{R}^m$ with $\overline{v}(t,x)=\varphi(t,x)$. As in the proof of Theorem 5.11, we may assume that $\varphi \in C_b^{1,2}([0,T]\times\mathbb{R}^d)$. Then, take $r>0$ such that

$$(\overline{v}-\varphi)(s,y) \le (\overline{v}-\varphi)(t,x), \quad (s,y)\in B_r(t,x)\subset[0,T)\times\mathbb{R}^d.$$

where $B_r(t,x)$ denote the closed ball at (t,x) with radius r. For $(s,y)\in B_r(t,x)$ set

$$\tilde{\varphi}(s,y) = \varphi(s,y) + |s-t|^2 + |y-x|^2.$$

It follows that (t,x) is a strict maximum of $\overline{v}-\tilde{\varphi}$ on $B_r(t,x)$. Also, for (s,y) outside the ball, we choose $\tilde{\varphi}$ so that $\tilde{\varphi}(s,y)\ge 2\sup_{h\in(0,1]}|v^h(s,y)|$ and that $\tilde{\varphi}$ is still in $C_b^{1,2}([0,T]\times\mathbb{R}^d)$. Thus (t,x) is a global strict maximum of $\overline{v}-\tilde{\varphi}$. By abuse of notation, we write φ for $\tilde{\varphi}$.

By definition of $\overline{v}$, there exist h_m, $\tilde{k}_m$, $\tilde{y}_m$, $m\ge 1$, such that $(t_{\tilde{k}_m},\tilde{y}_m)\in B_r(t,x)$ and as $m\to\infty$,

$$h_m \to 0,\ \ (t_{\tilde{k}_m},\tilde{y}_m)\to(t,x),\ \ v^{h_m}(t_{\tilde{k}_m},\tilde{y}_m)\to\overline{v}(t,x).$$

Take k_m and y_m so that

$$(v^{h_m}-\varphi)(t_{k_m},y_m) \ge \max_{k=0,1,\dots,n}\sup_{y\in\mathbb{R}^d}(v^{h_m}-\varphi)(t_k,y) - (\Delta t)_m^2, \tag{5.5.2}$$

where $(\Delta t)_m = \Delta t$ for $h=h_m$. The sequence (t_{k_m},y_m), $m\ge 1$, can be taken from the bounded set $B_r(t,x)$, so there exists a limit point $(\tilde{t},\tilde{x})\in B_r(t,x)$ possibly along a subsequence. Thus, denoting $c_m = (v^{h_m}-\varphi)(t_{k_m},y_m)$, we have

$$0 = (\overline{v}-\varphi)(t,x) = \lim_{m\to\infty}(v^{h_m}-\varphi)(t_{\tilde{k}_m},\tilde{y}_m) \le \liminf_{m\to\infty} c_m \le \limsup_{m\to\infty} c_m \le (\bar{v}-\varphi)(\tilde{t},\tilde{x}).$$

Since (t,x) is a strict maximum, we deduce that $(\tilde{t},\tilde{x}) = (t,x)$. Therefore, it follows that $(t_{k_m},y_m)\to(t,x)$ and $c_m\to 0$.

By (5.5.2), for any $y\in\mathbb{R}^d$,

$$\varphi(t_{k_m+1},y) + c_m + (\Delta t)_m^2 \ge v^{h_m}(t_{k_m+1},y).$$

Thus, using the monotonicity property in Assumption 5.16,

$$\begin{aligned}&\frac{1}{\Delta t}\Phi^{h_m}[\varphi(t_{k_m}+\Delta t,\cdot)+c_m+(\Delta t)_m^2](t_{k_m},y_m)\\ &\ge \frac{1}{\Delta t}v^{h_m}(t_{k_m},y_m) \ge \frac{1}{\Delta t}(\varphi(t_{k_m},y_m)+c_m+(\Delta t)_m^2) - 2(\Delta t)_m.\end{aligned}$$

Combining this with the consistency property in Assumption 5.16, we find that

$$F(t,x,\partial_t\varphi(t,x),\varphi(t,x),D\varphi(t,x),D^2\varphi(t,x)) \geq 0.$$

Thus $\overline{v}$ is a discontinuous viscosity subsolution of (5.2.1).

By a similar argument, we can show that

$$\underline{v}(t,x) = \liminf_{\substack{(t_k,y)\to(t,x)\\ h\searrow 0}} v^h(t_k,y), \quad (t,x)\in[0,T]\times\mathbb{R}^d$$

is a discontinuous viscosity supersolution of (5.2.1). Since $\overline{v}(T,\cdot) = \underline{v}(T,\cdot) = g$, Assumption 5.15 now implies that $\overline{v} \leq \underline{v}$. However, by definition, $\overline{v} \geq \underline{v}$. Hence we obtain $\overline{v} = \underline{v}$. This means that $v := \overline{v} = \underline{v}$ is a discontinuous viscosity solution of (5.2.1). From $v(t,x) = \lim_{(t_k,y)\to(t,x)} \lim_{h\to 0} v^h(t_k,y)$, the continuity of v follows. Hence v is a continuous viscosity solution of (5.2.1).

Now take an arbitrary compact set $K \subset \mathbb{R}^d$. Further fix $t\in[0,T]$ and $\varepsilon > 0$. Then, by the uniform continuity of $v(t,\cdot)$ on K, there exists $\delta_0 > 0$ such that $|v(t,y) - v(t,z)| < \varepsilon$ whenever $|y-z| < \delta_0$. Moreover, for any $x\in K$ there exist $\delta(x) > 0$ and $h(x)\in(0,1]$ such that

$$|v^h(t_k,y) - v(t,x)| < \varepsilon, \quad y\in B_{\delta(x)}(x), \;\; h\leq h(x),$$

where $t_k \to t$ as $h\to 0$. We may assume $\delta(x)\leq\delta_0$ for all $x\in K$. Since $\{B_{\delta(x)}(x)\}_{x\in K}$ is an open coverage of K, there exist $x_1,\ldots,x_k\in K$ such that $K\subset \cup_{i=1}^k B_{\delta(x_i)}(x_i)$. Thus for any $x\in K$ we have $|v^h(t_k,x) - v^h(t,x_i)| < \varepsilon$ for some $i = 1,\ldots,k$ whenever $h\leq h_0 := \min\{h(x_1),\ldots,h(x_k)\}$. This means that $|v^h(t_k,x)-v(t,x)| \leq |v^h(t_k,x)-v(t,x_i)|+|v(t,x_i)-v(t,x)| < 2\varepsilon$. Consequently,

$$\sup_{x\in K} |v^h(t_k,x) - v(t,x)| \leq 2\varepsilon, \quad h\leq h_0.$$

Thus the required uniform convergence follows. □

CHAPTER 6

Numerical Methods for Nonlinear PDEs

6.1 Introduction

The objective of this chapter is to discuss numerical methods for the terminal value problems of the parabolic PDEs:

$$\begin{aligned} -\partial_t v(t,x) + F(t,x,v(t,x),Dv(t,x),D^2v(t,x)) = 0, \quad (t,x) \in [0,T) \times \mathbb{R}^d, \\ v(T,x) = f(x), \quad x \in \mathbb{R}^d, \end{aligned} \tag{6.1.1}$$

where $F : [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d \to \mathbb{R}$. As seen in the previous chapter, under suitable conditions including the ellipticity condition on F, the terminal value problem (6.1.1) has a unique viscosity solution v.

Most popular numerical method is the *finite difference method*. This is powerful and mathematically harmless in the case of $d = 1$. However, its time complexity is growing exponentially as d becomes large, and strong conditions need to ensure the rigorous convergence for $d \geq 2$. We refer to [11] and Ieda [17] for the analysis of the finite difference method.

As an alternative, we present *kernel-based collocation methods*. To explain a basic idea, let $\mathcal{O} \subset \mathbb{R}^d$ be a set on which functions to be approximated, $\Gamma = \{x^{(1)}, \cdots, x^{(N)}\}$ be a finite subset of $\mathcal{O}$, and $\Phi : \mathcal{O} \times \mathcal{O} \to \mathbb{R}$. Suppose that the matrix $A := \{\Phi(x^{(i)}, x^{(j)})\}_{i,j=1,\dots,N}$ is invertible. Then for any $f : \mathcal{O} \to \mathbb{R}$, the linear equation

$$A\alpha = f|_\Gamma$$

has a unique solution $\alpha = (\alpha_1, \dots, \alpha_N)^\mathsf{T} \in \mathbb{R}^N$, where $f|_\Gamma = (f(x^{(1)}), \dots, f(x^{(N)})^\mathsf{T} \in \mathbb{R}^N$. Namely, for $f : \mathcal{O} \to \mathbb{R}$, the function

$$I(f)(x) = \sum_{j=1}^{N} (A^{-1} f|_\Gamma)_j \Phi(x, x^{(j)}), \quad x \in \mathcal{O},$$

interpolates f on Γ, where $(\xi)_j$ denotes the j-th component of $\xi \in \mathbb{R}^N$. This suggests

$$f(x) \approx I(f)(x), \quad x \in \mathcal{O}.$$

Now, by a time-discretization of (6.1.1),

$$v(t_k, x) \simeq v(t_{k+1}, x) - \Delta t F(t_{k+1}, x, v(t_{k+1}, x), Dv(t_{k+1}, x), D^2 v(t_{k+1}, x)), \quad x \in \mathcal{O},$$

where $t_k = kT/n$, $k = 0, 1, \ldots, n$, and $\Delta t = T/n$. Then by replacing the derivatives of $v(t_{k+1}, \cdot)$ with those of $I(v(t_{k+1}, \cdot))$, we obtain

$$v(t_k, x) \simeq v(t_{k+1}, x) - \Delta t F(t_{k+1}, x, v(t_{k+1}, x), DI(v(t_{k+1}, \cdot)(x), D^2 I(v(t_{k+1}, \cdot))(x)), \quad x \in \mathcal{O}.$$

This leads to a recursive equation backward in time that is determined by the *collocation points* $\{t_0, \ldots, t_n\} \times \Gamma$. We analyze this method in details in Section 6.3.

As preliminaries, the next section is devoted to the review of the theory of the function approximations above. We refer to [38] for a complete account.

6.2 Function Approximations with Reproducing Kernels

Let $\mathcal{O} = \{x \in \mathbb{R}^d : |x - \tilde{x}|_0 < R\}$, an open ball centered at some $\tilde{x} \in \mathbb{R}^d$ with a radius $R > 0$ defined by some Euclidean norm $|\cdot|_0$ in $\mathbb{R}^d$.

Definition 6.1. We say that $\Phi : \mathbb{R}^d \to \mathbb{R}$ is a positive definite function if for every $\ell \in \mathbb{N}$, for all pairwise distinct $y_1, \ldots, y_\ell \in \mathbb{R}^d$ and for all $\alpha = (\alpha_i) \in \mathbb{R}^\ell \setminus \{0\}$, we have

$$\sum_{i,j=1}^{\ell} \alpha_i \alpha_j \Phi(y_i - y_j) > 0.$$

Moreover, Φ is said to be a radial function if $\Phi(\cdot) = \phi(|\cdot|)$ for some $\phi : [0, \infty) \to \mathbb{R}$.

For $f \in L^1(\mathbb{R}^d)$ the Fourier transform of f is defined by

$$\widehat{f}(\xi) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) e^{-\sqrt{-1}x^\mathsf{T}\xi} dx, \quad \xi \in \mathbb{R}^d.$$

Theorem 6.2

Suppose that $\Phi \in C(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$. If $\widehat{\Phi}(\xi) > 0$ for any $\xi \in \mathbb{R}^d$, then Φ is positive definite.

Proof. Since $\Phi \in C(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$, we can apply the Fourier inversion formula (see, e.g., [43]) to obtain

$$\Phi(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{\Phi}(\xi) e^{\sqrt{-1}x^\mathsf{T}\xi} d\xi, \quad x \in \mathbb{R}^d.$$

Thus, for every $\ell \in \mathbb{N}$, for all pairwise distinct $y_1, \ldots, y_\ell \in \mathcal{O}$ and for all $\alpha = (\alpha_i) \in \mathbb{R}^\ell$, we have

$$\begin{aligned}\sum_{i,j=1}^{\ell} \alpha_i \alpha_j \Phi(y_i - y_j) &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \sum_{i,j=1}^{\ell} \alpha_i \alpha_j e^{\sqrt{-1}(y_i - y_j)^\mathsf{T}\xi} \widehat{\Phi}(\xi) d\xi \\ &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \left| \sum_{i=1}^{\ell} \alpha_i e^{\sqrt{-1}y_i^\mathsf{T}\xi} \right|^2 \widehat{\Phi}(\xi) d\xi.\end{aligned}$$

Now suppose that $\sum_{i,j=1}^{\ell} \alpha_i \alpha_j \Phi(y_i - y_j) = 0$. Then, since $\widehat{\Phi} > 0$, we have $\sum_{i=1}^{\ell} \alpha_i e^{\sqrt{-1}y_i^\mathsf{T}\xi} = 0$, $d\xi$-a.e. Hence, by continuity, $\sum_{i=1}^{\ell} \alpha_i e^{\sqrt{-1}y_i^\mathsf{T}\xi} = 0$ for any $\xi \in \mathbb{R}^d$. Fix an arbitrary $i \in \{1, \ldots, \ell\}$ and consider $f \in C_0^\infty(\mathbb{R}^d)$ satisfying $f = 1$ on $\{x : |x - y_i| < \varepsilon/2\}$ and $f = 0$ on $\{x : |x - y_i| > \varepsilon\}$, where $\varepsilon > 0$ is sufficiently small such that $f(y_j) = 0$ for every $j \neq i$. Then by the Fourier inversion,

$$\alpha_i = \sum_{j=1}^{\ell} \alpha_j f(y_j) = 0.$$

Thus the theorem follows. □

Example 6.3 (Gaussian kernel). Consider the case where $\Phi(x) = e^{-\alpha|x|^2}$, $x \in \mathbb{R}^d$, $\alpha > 0$. It is straightforward to see that $G(x) := e^{-|x|^2/2}$, $x \in \mathbb{R}^d$, satisfies $\widehat{G} = G$ on $\mathbb{R}^d$. From this it follows that $\widehat{\Phi}(\xi) = \widehat{G}(1/\sqrt{2\alpha})(2\alpha)^{-d/2} > 0$. Hence Φ is positive definite on $\mathbb{R}^d$.

Example 6.4 (Inverse multiquadric kernel). Consider the case where $\Phi(x) = (c^2+|x|^2)^{-\beta}$, $x \in \mathbb{R}^d$, $c > 0$, $\beta > d/2$. Then we confirm by an elementary analysis that

$$\widehat{\Phi}(\xi) = \gamma \left(\frac{|\xi|}{c}\right)^{\beta-d/2} K_{d/2-\beta}(c|\xi|) > 0, \quad x \in \mathbb{R}^d,$$

where

$$1/\gamma = 2^{\beta-1} \int_0^\infty t^{\beta-1} e^{-t} dt,$$

and $K_\nu(z)$, $z > 0$, is the modified Bessel function of 3rd (2nd) kind given by

$$K_\nu(z) = \int_0^\infty e^{-z\cosh(t)} \cosh(\nu t) dt.$$

Hence Φ positive definite on $\mathbb{R}^d$.

Example 6.5 (Wendland kernel). Consider the case where $\Phi(x) = \phi_{d,\tau}(|x|)$. Here,

$$\phi_{d,\tau}(r) = \begin{cases} \displaystyle\int_r^1 s(1-s)^\ell (s^2 - r^2)^{\tau-1} ds, & 0 \le r \le 1, \\ 0, & r > 1, \end{cases}$$

where $\ell = \max\{k \in \mathbb{Z} : k \le d/2\} + \tau + 1$. It is known that Φ is positive definite on $\mathbb{R}^d$ and in $C^{2\tau}(\mathbb{R}^d)$. See [38]. For example,

$$\begin{aligned}
\phi_{1,2}(r) &\doteq \max\{1-r, 0\}^5 (8r^2 + 5r + 1), \\
\phi_{1,3}(r) &\doteq \max\{1-r, 0\}^7 (21r^3 + 19r^2 + 7r + 1), \\
\phi_{1,4}(r) &\doteq \max\{1-r, 0\}^9 (384r^4 + 453r^3 + 237r^2 + 63r + 7), \\
\phi_{2,4}(r) &\doteq \max\{1-r, 0\}^{10} (429r^4 + 450r^3 + 210r^2 + 50r + 5), \\
\phi_{2,5}(r) &\doteq \max\{1-r, 0\}^{12} (2048r^5 + 2697r^4 + 1644r^3 + 566r^2 + 108r + 9),
\end{aligned}$$

where $\doteq$ denotes equality up to a positive constant factor.

- One of advantages in using Wendland kernel, which is complicatedly constructed and has a limited smoothness, is that the corresponding interpolation matrix A is *sparse*.
- Another advantage is that a function space where the approximation works is relatively easy to handle.

In what follows, let $\Phi : \mathbb{R}^d \to \mathbb{R}$ be a fixed positive definite function, and we provide a theoretical validation of the approximation $I(f) \simeq f$.

Theorem 6.6

There exists a unique Hilbert space $\mathcal{N}_\Phi(\mathcal{O}) \subset C(\mathcal{O})$ with inner product $\langle \cdot, \cdot \rangle_{\mathcal{N}_\Phi(\mathcal{O})}$ such that

(i) $\Phi(\cdot - y) \in \mathcal{N}_\Phi(\mathcal{O})$ for all $y \in \mathcal{O}$.

(ii) $f(y) = \langle f, \Phi(\cdot - y) \rangle_{\mathcal{N}_\Phi(\mathcal{O})}$ for all $f \in \mathcal{N}_\Phi(\mathcal{O})$ and $y \in \mathcal{O}$.

- We call $\mathcal{N}_\Phi(\mathcal{O})$ the *native space*.
- Φ is said to be a *reproducing kernel* for $\mathcal{N}_\Phi(\mathcal{O})$.

Example 6.7 (Gaussian kernel). In the case where Φ is given by the Gaussian kernel,

$$\mathcal{N}_\Phi(\mathbb{R}^d) = \left\{ f : \mathbb{R}^d \to \mathbb{R} \,\middle|\, \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 e^{|\xi|^2/(4\alpha)} d\xi < \infty \right\}$$

and there exist $c_1, c_2 > 0$ such that for $f \in \mathcal{N}_\Phi(\mathbb{R}^d)$,

$$c_1 \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 e^{|\xi|^2/(4\alpha)} d\xi \le \|f\|^2_{\mathcal{N}_\Phi(\mathbb{R}^d)} \le c_2 \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 e^{|\xi|^2/(4\alpha)} d\xi.$$

Here, for $f \in L^1(\mathbb{R}^d)$, the function $\widehat{f}$ is the Fourier transform of f, defined as usual by

$$\widehat{f}(\xi) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) e^{-\sqrt{-1}x^\mathsf{T}\xi} dx, \quad \xi \in \mathbb{R}^d.$$

Example 6.8 (Inverse multiquadric kernel). In the case where Φ is given by the inverse multiquadric kernel,

$$\mathcal{N}_\Phi(\mathbb{R}^d) = \left\{ f : \mathbb{R}^d \to \mathbb{R} \,\middle|\, \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 |\xi|^{d-2\beta} / K_{d/2-\beta}(c|\xi|^2) d\xi < \infty \right\}$$

and there exist $c_1, c_2 > 0$ such that for $f \in \mathcal{N}_\Phi(\mathbb{R}^d)$,

$$c_1 \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 |\xi|^{d-2\beta} / K_{d/2-\beta}(c|\xi|^2) d\xi \le \|f\|^2_{\mathcal{N}_\Phi(\mathbb{R}^d)} \le c_2 \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 |\xi|^{d-2\beta} / K_{d/2-\beta}(c|\xi|^2) d\xi.$$

Here, K_ν is the modified Bessel function of the third kind of the order ν.

Example 6.9 (Wendland kernel). In the case where Φ is given by the Wendland kernel,

$$\mathcal{N}_\Phi(\mathbb{R}^d) = \left\{ f : \mathbb{R}^d \to \mathbb{R} \,\middle|\, \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 (1 + |\xi|^2)^{\tau + (d+1)/2} d\xi < \infty \right\}$$

and there exist $c_1, c_2 > 0$ such that for $f \in \mathcal{N}_\Phi(\mathbb{R}^d)$,

$$c_1 \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 (1 + |\xi|^2)^{\tau + (d+1)/2} d\xi \le \|f\|^2_{\mathcal{N}_\Phi(\mathbb{R}^d)} \le c_2 \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 (1 + |\xi|^2)^{\tau + (d+1)/2} d\xi.$$

That is, the native space is given by the L^2-Sobolev space of the order τ with equivalent norm. Moreover, if $\tau + (d+1)/2$ is a positive integer, then

$$\mathcal{N}_\Phi(\mathbb{R}^d) = \left\{ f : \mathbb{R}^d \to \mathbb{R} \,\middle|\, \sum_{|\alpha| \le \tau + (d+1)/2} \int_{\mathbb{R}^d} |D^\alpha f(x)|^2 dx < \infty \right\}.$$

We will show that the approximation $I(f) \simeq f$ works on the native space and the error can be described in terms of $\|f\|_{\mathcal{N}_\Phi(\mathcal{O})} := \langle f, f \rangle^{1/2}_{\mathcal{N}_\Phi(\mathcal{O})}$ and

$$\Delta x := \sup_{x \in \mathcal{O}} \min_{j=1,\ldots,N} |x - x^{(j)}|.$$

That is, Δx is the Hausdorff distance between Γ and $\mathcal{O}$.

Theorem 6.10

Suppose that $\Phi \in C^2(\mathbb{R}^d)$. Then there exists a positive constant $C_{\Phi,\mathcal{O}}$, only depending on Φ and $\mathcal{O}$, such that for any $f \in \mathcal{N}_\Phi(\mathcal{O})$,

$$|f(x) - I(f)(x)| \le C_{\Phi,\mathcal{O}} \Delta x \|f\|_{\mathcal{N}_\Phi(\mathcal{O})}, \quad x \in \mathcal{O}.$$

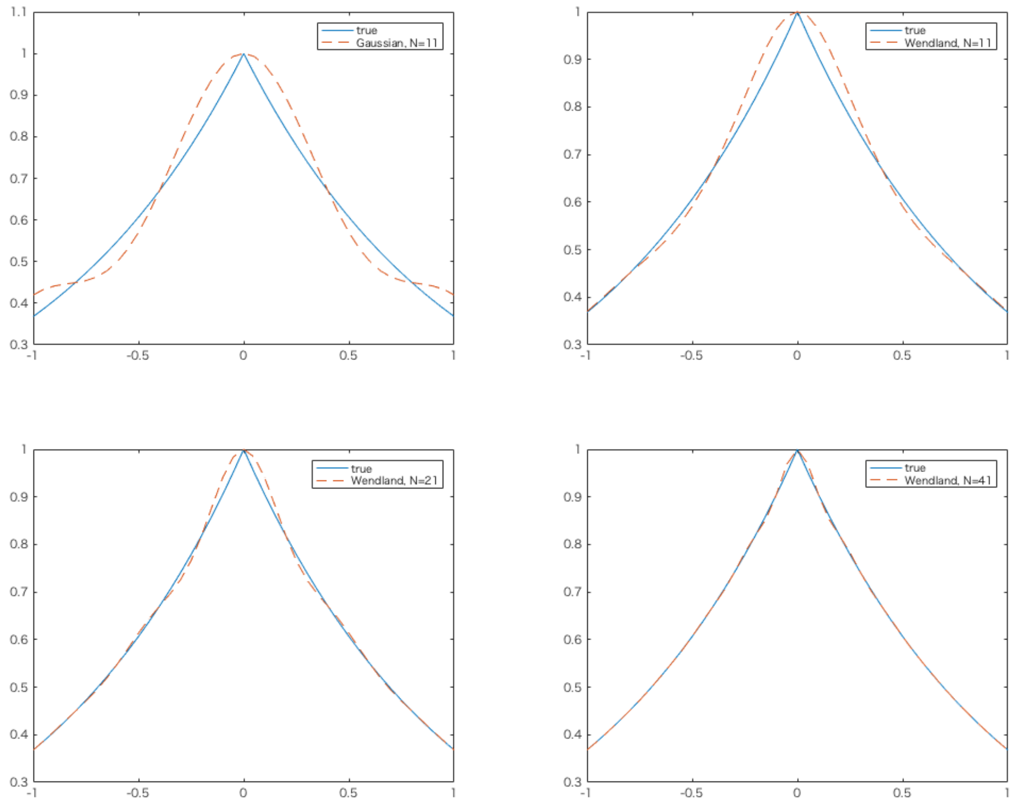

Figure 6.2.1: Approximation of $e^{-|x_1|}$ $(d = 1)$. Gaussian kernel with $\alpha = 1$, $N = 11$ and Wendland kernel $\phi_{1,3}$ for $N = 11, 21, 41$. Γ is set to be the uniform grid on $[-2, 2]$ including the boundary.

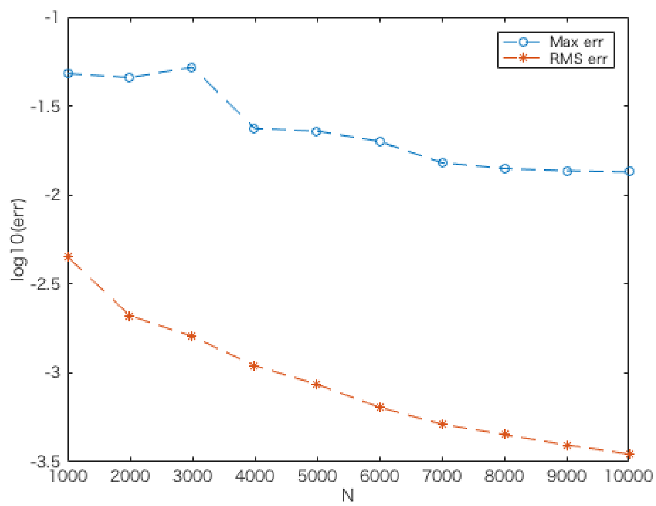

Figure 6.2.2: Approximation errors of $e^{-|x_1|-|x_2|}$ $(d = 2)$. Wendland kernel $\phi_{2,4}$ for $N = 1000, 2000, \ldots, 10000$. Γ is generated by the quasi random number of Halton type on $[-2, 2]^2$. The evaluations are done at 441 uniform grid points on $[-1, 1]^2$ including the boundary.

Outline of the proof.
Step (i). Observe

$$\sup_{x\in\mathcal{O}}|f(x)| \le \max_{j=1,\dots,N}|f(x^{(j)})| + K_f\Delta x$$

for any Lipschitz continuous function f on $\mathcal{O}$ where

$$K_f := \sup_{\substack{x,y\in\mathcal{O}\\ x\neq y}} \frac{|f(x)-f(y)|}{|x-y|}.$$

Step (ii). We will see that for any $f \in \mathcal{N}_\Phi(\mathcal{O})$ we have $\|f - I(f)\|_{\mathcal{N}_\Phi(\mathcal{O})} \le \|f\|_{\mathcal{N}_\Phi(\mathcal{O})}$ and there exists a constant $C > 0$ such that

$$K_f \le C\|f\|_{\mathcal{N}_\Phi(\mathcal{O})}.$$

□

If Φ is of Gaussian or inverse multiquadric types, then we can obtain an arbitrary order of convergence.

Theorem 6.11

Suppose that Φ is one of the Gaussians or the inverse multiquadrics. Let $\ell \in \mathbb{N}$. Then there exist a positive constants δ_0 and C such that for any $f \in \mathcal{N}_\Phi(\mathcal{O})$, $x \in \mathcal{O}$, and $\Delta x \le \delta_0$,

$$|f(x) - I(f)(x)| \le C(\Delta x)^\ell \|f\|_{\mathcal{N}_\Phi(\mathcal{O})}.$$

In the case of Wendland kernels, we have the following:

Theorem 6.12

Suppose that $\Phi = \phi_{d,\tau}(|\cdot|)$ is the Wendland kernel. Then there exist positive constant δ_0 and C such that for any $f \in \mathcal{N}_\Phi(\mathcal{O})$, $x \in \mathcal{O}$, and $\Delta x \le \delta_0$,

$$|f(x) - I(f)(x)| \le C(\Delta x)^{\tau+1/2}\|f\|_{\mathcal{N}_\Phi(\mathcal{O})}.$$

6.3 Kernel-Based Collocation Methods

Construction

In this section, the function Φ is assumed to be the Wendland kernel $\Phi_{d,\tau}$ divided by some positive constant with fixed $\tau \ge 2$. Let $h > 0$ be a parameter that describes approximate solutions, $\Gamma = \{x^{(1)}, \dots, x^{(N)}\} \subset (-R, R)^d$ with $R > 1$, and $\{t_0, \dots, t_n\}$ the set of time grid points such that $t_k = kT/n$, $k = 0, \dots, n$. Then think of the interpolant

$$v^h(t_k, x) = \sum_{j=1}^{N} (A^{-1}v_k^h)_j \Phi(x - x^{(j)}), \quad x \in \mathbb{R}^d, \tag{6.3.1}$$

of $v_k^h = (v_{k,1}^h, \dots, v_{k,N}^h)^\top \in \mathbb{R}^N$ to be specified below. Substituting this into the time discretized equation

$$\frac{v(t_{k+1}, x) - v(t_k, x)}{t_{k+1} - t_k} \simeq F(t_{k+1}, x; v(t_{k+1}, \cdot)),$$

we derive the following equation for $\{v_k^h\}$:

$$v_{k+1,j}^h - v_{k,j}^h = (t_{k+1} - t_k)F_{k+1,j}(v_{k+1}^h), \quad k = 0, \dots, n-1, \;\; j = 1, \dots, N.$$

Here, for any C^2-function φ on $\mathbb{R}^d$,

$$F(t,x;\varphi) = F(t,x,\varphi(x),D\varphi(x),D^2\varphi(x)), \quad x \in \mathbb{R}^d,$$

and $F_{k,j}(v_k^h) = F(t_k, x^{(j)}; v^h(t_k,\cdot))$. The terminal condition leads to $v_{n,j}^h = f(x^{(j)})$, $j = 1,\dots,N$. Thus, denoting $F_k(v_k^h) = (F_{k,1}(v_k^h),\dots,F_{k,N}(v_k^h))^\mathsf{T}$, we get

$$\begin{cases} v_k^h = v_{k+1}^h - (t_{k+1}-t_k)F_{k+1}(v_{k+1}^h), \quad k = 0,\dots,n-1, \\ v_n^h = f|_\Gamma. \end{cases} \tag{6.3.2}$$

Consequently, we define the function $v^h(t_k,x)$, a candidate of an approximate solution of (6.1.1), by (6.3.1) with $\{v_k^h\}$ determined by the equation (6.3.2).

Remark 6.13. The linearity of the interpolant yields, for $x \in \mathbb{R}^d$,

$$v^h(t_k,x) = v^h(t_{k+1},x) - (t_{k+1}-t_k)I(F_{k+1}(v_{k+1}^h))(x),$$

where by abuse of notation we denote $I(\xi)(x) = \sum_{j=1}^N (A^{-1}\xi)_j \Phi(x - x^{(j)})$ for $\xi \in \mathbb{R}^N$.

Let us describe our collocation methods in a matrix form. To this end, we assume here that the nonlinearity F can be written as

$$F(t,x;\varphi) = \sup_{\pi \in K} H(t,x,\varphi(x), b(x,\pi)^\mathsf{T} D\varphi(x), \mathrm{tr}(a(x,\pi)D^2\varphi(x))),$$

where K is a set, $b : \mathbb{R}^d \times K \to \mathbb{R}^d$, $a : \mathbb{R}^d \times K \to \mathbb{S}^d$, and $H : [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. It should be noted that the nonlinearities corresponding to Hamilton-Jacobi-Bellman equations are represented in this form. Then, consider the function $\phi_{d,\tau}^{(1)}(r) := \phi'_{d,\tau}(r)/r$, $r \geq 0$. By definition of $\phi_{d,\tau}$, the function $\phi_{d,\tau}^{(1)}$ is continuous on $[0,\infty)$ and supported in $[0,1]$. With this function, we have

$$\partial_{x_m}\Phi(x) = \phi^{(1)}(|x|)x_m, \quad x = (x_1,\dots,x_d) \in \mathbb{R}^d.$$

Thus,

$$B_\ell(\pi) := \Big(b_\ell(x^{(i)},\pi)\partial_{x_\ell}\Phi(x^{(i)} - x^{(j)})\Big)_{1 \le i,j \le N} = Q_\ell(\pi)(G_\ell A_1 - A_1 G_\ell),$$

where $Q_\ell(\pi) = \mathrm{diag}(b_\ell(x^{(1)},\pi),\dots,b_\ell(x^{(N)},\pi))$, $A_1 = \{\phi_{d,\tau}^{(1)}(|x^{(i)} - x^{(j)}|)\}_{1\le i,j\le N}$ and $G_\ell = \mathrm{diag}(x_\ell^{(1)},\dots,x_\ell^{(N)})$. Hence,

$$\mathbb{R}^N \ni (b_\ell(x^{(i)},\pi)(\partial/\partial_{x_\ell})I(\xi)(x^{(i)}))_{1\le i\le N} = B_\ell(\pi)A^{-1}\xi.$$

Similarly,

$$\partial^2_{x_m x_\ell}\Phi(x) = \begin{cases} \phi_{d,\tau}^{(1)}(|x|) + \phi_{d,\tau}^{(2)}(|x|)x_m^2, & (\ell = m), \\ \phi_{d,\tau}^{(2)}(|x|)x_m x_\ell, & (\ell \neq m), \end{cases}$$

where

$$\phi_{d,\tau}^{(2)}(r) = \frac{1}{r}\frac{d\phi_{d,\tau}^{(1)}}{dr}(r), \quad r \geq 0.$$

Notice that $\phi_{d,\tau}^{(2)}$ is also continuous on $[0,\infty)$ and supported in $[0,1]$. Thus,

$$B_{m\ell}(\pi) := \Big\{a_{m\ell}(x^{(i)},\pi)\partial^2_{x_m x_\ell}\Phi(x^{(i)} - x^{(j)})\Big\}_{1\le i,j\le N}$$

is given by

$$B_{mm}(\pi) = Q_{mm}(\pi)(A_1 + G_m^2 A_2 - 2G_m A_2 G_m + A_2 G_m^2)$$

and for $m \neq \ell$,

$$B_{m\ell} = Q_{m\ell}(\pi)(G_m G_\ell A_2 - G_m A_2 G_\ell - G_\ell A_2 G_m + A_2 G_m G_\ell)$$

with $A_2 = \{\phi_{d,\tau}^{(2)}(|x^{(i)} - x^{(j)}|)\}_{1\le i,j\le N}$ and $Q_{m\ell}(\pi) = \mathrm{diag}(a_{m\ell}(x^{(1)},\pi),\dots,a_{m\ell}(x^{(N)},\pi))$. Consequently, we obtain

$$F_{k,j}(v_k^h) = \sup_{\pi\in K} H\left(t_k, x^{(j)}, \left(\sum_{m=1}^d B_m(\pi)A^{-1}v_k^h\right)_j, \left(\sum_{m,\ell=1}^d B_{m\ell}(\pi)A^{-1}v_k^h\right)_j\right).$$

Numerical examples

Here we consider the following equation for our numerical experiments:

$$\begin{cases} -\partial_t v - \dfrac{1}{2}\displaystyle\sup_{0\le\sigma\le 1/5}\mathrm{tr}(\sigma^2 D^2 v) + G(v, Dv) = 0, \quad (t,x)\in[0,1)\times\mathbb{R}^d, \\ v(1,x) = \sin\left(1 + \displaystyle\sum_{i=1}^d x_i\right), \quad x = (x_1,\dots,x_d)^\mathsf{T}\in\mathbb{R}^d, \end{cases}$$

where $G(z,p) = (1/d)\sum_{i=1}^d p_i - (d/2)\inf_{0\le\sigma\le 1/5}(\sigma^2 z)$ for $z\in\mathbb{R}$, $p = (p_1,\dots,p_d)^\mathsf{T}\in\mathbb{R}^d$. It is straightforward to see that the unique solution is given by $v(t,x) = \sin(t + \sum_{i=1}^d x_i)$.

We apply our method to this equation in the cases of $d = 1$ and $d = 2$. As mentioned in Section 6.1, we use the interpolation method as a practical alternative to the regression one and then show its usefulness through the numerical experiments below.

For each $d = 1, 2$, we choose the parameter $\tau = \tau_d$ of the Wendland kernel as $\tau_1 = 4$ and $\tau_2 = 15$. We construct the set $\Gamma = \Gamma_d$ of collocation points as the equi-spaced points on $[-R_d, R_d]^d$, where

$$R_d = \gamma_d N^{1/d - 1/(d+2\tau_d-3)}.$$

Here, $\gamma_1 = 1/4$ and $\gamma_2 = 1/5$. These choices come from the fact that $\Delta x \sim R_d N^{-1/d}$ and the interpolation error up to the second derivatives is $O((\Delta x)^{\tau_d - 3/2})$ (see Corollary 11.33 in [38]).

To implement the collocation method, we use the matrix representation, by noting $\inf_{0\le\sigma\le 1/5}(\sigma^2 y) = -(1/5)^2\max(-y,0)$, with the uniform time grid. We examine the cases of $n = 2^8$ and $n = 2^{12}$. Figures 6.3.1 and 6.3.2 show the resulting root mean square errors and the maximum errors, defined by

$$\text{Max error} = \max_{\xi\in\Gamma_0,\, i=0,\dots,n}\left|v^h(t_i,\xi) - v(t_i,\xi)\right|,$$

$$\text{RMS error} = \sqrt{\frac{1}{10^d(n+1)}\sum_{\xi\in\Gamma_0}\sum_{i=0}^n |v^h(t_i,\xi) - v(t_i,\xi)|^2},$$

respectively, where Γ_0 is the set of 10^d-evaluation points constructed by a Sobol' sequence on $[-1,1]^d$ for each $d = 1, 2$.

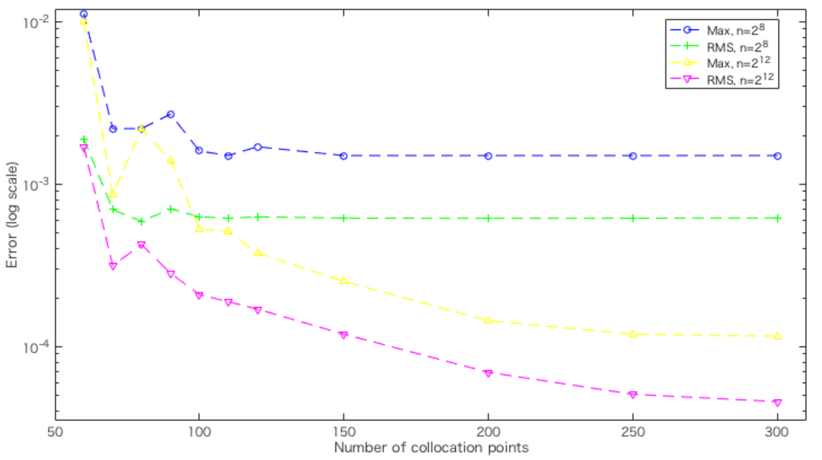

Figure 6.3.1: Max and RSM errors for $d = 1$ with $n = 2^8, 2^{12}$.

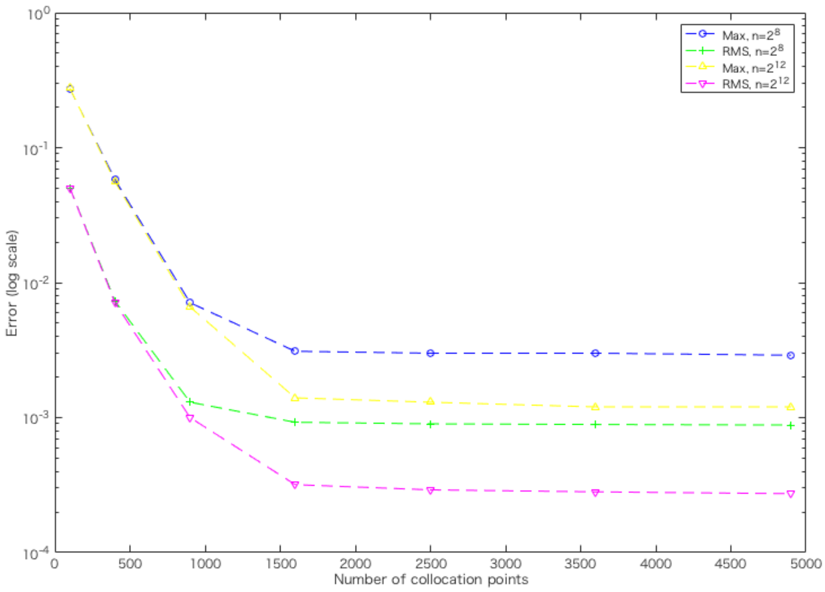

Figure 6.3.2: Max and RSM errors for $d = 2$ with $n = 2^8, 2^{12}$.

APPENDIX A

Review on Probability Theory

This chapter reviews basic facts about measure theoretic probability. We refer to, e.g., [39], [48], [44], and [46] for details.

Probability spaces

Definition A.1. Let Ω be an arbitrary set. A family $\mathcal{F}$ of subsets of Ω is said to be *σ-algebra* or *σ-field* if the following are satisfied:

(i) $\emptyset \in \mathcal{F}$.

(ii) If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$. Here, $A^c = \Omega \setminus A$.

(iii) If $A_1, A_2, \dots \in \mathcal{F}$ then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

- We say that a set $A \in \mathcal{F}$ is *$\mathcal{F}$-measurable* or simply *measurable*. Further, we call $A \in \mathcal{F}$ an *event*.
- The pair $(\Omega, \mathcal{F})$ is called a *measurable space*.

Example A.2. For any set Ω, the set $\mathcal{F}$ of all subsets of Ω, i.e., $\mathcal{F} = 2^{\Omega} := \{A : A \subset \Omega\}$, is a σ-field.

Proposition A.3

Let $(\Omega, \mathcal{F})$ be a measurable space, and let $A_i \in \mathcal{F}$, $i = 1, 2, \dots$. Then, the following sets are all $\mathcal{F}$-measurable:

$$\bigcup_{i=1}^{n} A_i, \quad \bigcap_{i=1}^{n} A_i, \quad \bigcap_{i=1}^{\infty} A_i, \quad \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i, \quad \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i.$$

Remark A.4. Basically, in probability theory, a subset of Ω is interpreted as randomly occurred phenomenon and is a mathematical object for measuring how probable is its occurrence. Then the σ-algebra $\mathcal{F}$ is a class of "well-defined" random phenomenons. For example, suppose that for well-defined phenomenons A and B we are in a position to study the phenomenon that both occurs and the one that A occurs but B does not. Then it is natural to require these phenomenons are also well-defined. Namely, it is convenient for us to have $A \cap B, A \cap B^c \in \mathcal{F}$ whenever $A, B \in \mathcal{F}$. For this purpose, we require a collection of random phenomenons to be a σ-algebra. In other words, since σ-algebras are closed under various set manipulations, complicated events can be well-defined objects to be studied. On the other hand, recall from Example A.2 that the totality of all subsets of Ω is always a σ-algebra. Thus one may naturally ask: is it sufficient to always adopt 2^Ω as the underlying σ-algebra? Are there needs to consider possibly different σ-algebras? We refer to, e.g.,[39] for a complete answer to this question. Here we only mention that there exists a subset of $[0, 1]$ such that the Lebesgue measure (see below) of the set cannot be defined. In general, we need to choose appropriate σ-algebras depending on problems. However, the choices of actually used σ-algebras are limited, so application-oriented reader may not be discouraged with such technicality in measure theory.

For a family $\mathcal{G}$ of subsets of Ω, we set

$$\sigma(\mathcal{G}) := \bigcap\{\mathcal{H} : \sigma\text{-algebra on } \Omega \text{ s.t. } \mathcal{G} \subset \mathcal{H}\}.$$

This is the minimum σ-filed containing $\mathcal{G}$.

Example A.5. Let $A \in \mathcal{F}$. In the case of $\mathcal{G} = \{A\}$, we have $\sigma(\mathcal{G}) = \{\emptyset, A, A^c, \Omega\}$. We usually write $\sigma(A)$ for $\sigma(\{A\})$.

Let Ω be a topological space, and let $\mathcal{G}$ be the set of all open sets in Ω. Then, we call $\sigma(\mathcal{G})$ a *Borel σ-algebras on* Ω, and write $\mathcal{B}(\Omega) = \sigma(\mathcal{G})$. We may take $\Omega = \mathbb{R}^n$, $[a, b]$ for examples. The notation $\mathcal{B}([a, b])$ is often abbreviated as $\mathcal{B}[a, b]$.

Definition A.6. A set function $\mathbb{P} : \mathcal{F} \to [0, 1]$ is said to be a *probability measure on* $(\Omega, \mathcal{F})$ if the following conditions are satisfied:

(i) $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$.

(ii) For $A_1, A_2, \cdots \in \mathcal{F}$ with $A_i \cap A_j = \emptyset$ $(i \neq j)$, we have

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

- We call the triple $(\Omega, \mathcal{F}, \mathbb{P})$ a *probability space.*
- $\mathbb{P}(A) =$ "the probability that the event A occurs".
- When $\mathbb{P}(A) = 1$, we say that "the event A occurs with probability one" or "the event A occurs with almost surely (a.s.)".
- We say that a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is *complete* if all subsets of an arbitrary set in $\mathcal{F}$ with probability zero belong to $\mathcal{F}$, i.e., if

$$B \in \mathcal{F},\ A \subset B,\ \mathbb{P}(B) = 0 \implies A \in \mathcal{F}.$$

Theorem A.7

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Put

$$\overline{\mathcal{F}} = \left\{ A \subset \Omega : \begin{array}{l} A_* \subset A \subset A^*,\ \ \mathbb{P}(A^* \setminus A_*) = 0 \\ \text{for some } A_*, A^* \in \mathcal{F} \end{array} \right\}$$

and set $\overline{\mathbb{P}}(A) = \mathbb{P}(A_*)$, $A \in \overline{\mathcal{F}}$, where A_* is as above. Then $(\Omega, \overline{\mathcal{F}}, \overline{\mathbb{P}})$ is a complete probability space.

- The probability space $(\Omega, \overline{\mathcal{F}}, \overline{\mathbb{P}})$ is said to be a *completion* of $(\Omega, \mathcal{F}, \mathbb{P})$.

Example A.8. Let $(\Omega, \mathcal{F})$ be a measurable space. For a fixed $\omega_0 \in \mathcal{F}$, we define $\mathbb{P} : \mathcal{F} \to [0,1]$ by

$$\mathbb{P}(A) = \begin{cases} 1, & \text{if } \omega_0 \in A, \\ 0, & \text{if } \omega_0 \notin A. \end{cases}$$

Then $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$. This $\mathbb{P}$ is called the Dirac measure at ω_0.

Example A.9. Let Ω be a finite set (i.e., $\#\Omega < \infty$), and let $\mathcal{F}$ be the set of all subsets of Ω. Then we define $\mathbb{P} : \mathcal{F} \to [0,1]$ by $\mathbb{P}(A) = \sum_{\omega \in A} p_\omega$, $A \in \mathcal{F}$, where $\{p_\omega\}_{\omega \in \Omega}$ satisfies $p_\omega \in [0,1]$ for each $\omega \in \Omega$ and $\sum_{\omega \in \Omega} p_\omega = 1$. By this procedure, we can construct any probability measure on $(\Omega, \mathcal{F})$.

Example A.10 (Lebesgue measure). There exists a probability measure μ on $((0,1], \mathcal{B}((0,1]))$ such that

$$\mu((a,b]) = b - a, \quad 0 \le a \le b \le 1.$$

See, e.g., [44], [43], and [39]. That is, μ measures the length of intervals in $[0,1]$. This is called the *Lebesgue measure on* $((0,1], \mathcal{B}(0,1])$. By Definition A.6, we can show that $\mu(\{0\}) = 0$. So it can be seen as a probability measure on $([0,1], \mathcal{B}[0,1])$.

Further, there exists a nonnegative measure ν on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (i.e., a nonnegative set function ν satisfying Definition A.6 (ii)) such that

$$\nu((a,b]) = b - a, \quad -\infty \le a \le b \le +\infty.$$

This is called the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Moreover, since ν defines a measure on $[\alpha, \beta] \subset \mathbb{R}$, the restricted measure is called the Lebesgue measure on $([\alpha, \beta], \mathcal{B}[\alpha, \beta])$.

Proposition A.11

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then we have the following:

(i) $A \in \mathcal{F} \Longrightarrow \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(ii) $A, B \in \mathcal{F}$, $A \subset B \Longrightarrow \mathbb{P}(A) \le \mathbb{P}(B)$.

(iii) $A_n \in \mathcal{F}$, $n = 1, 2, \ldots \Longrightarrow \mathbb{P}(\bigcup_n A_n) \le \sum_n \mathbb{P}(A_n)$.

(iv) $A_n \in \mathcal{F}$, $n = 1, 2, \ldots$, $A_1 \subset A_2 \subset \cdots \Longrightarrow \lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\bigcup_n A_n)$.

(v) $A_n \in \mathcal{F}$, $n = 1, 2, \ldots$, $A_1 \supset A_2 \supset \cdots \Longrightarrow \lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\bigcap_n A_n)$.

The following fact is frequently used:

Lemma A.12: Borel-Cantelli lemma

Suppose that a sequence $\{A_n\} \subset \mathcal{F}$ satisfies $\sum_{n=1}^\infty \mathbb{P}(A_n) < \infty$. Then

$$\mathbb{P}\left(\bigcap_{n \ge 1} \bigcup_{k \ge n} A_k \right) = 0.$$

Proof. It follows from Proposition A.11 that

$$\mathbb{P}\left(\bigcap_{n \ge 1} \bigcup_{k \ge n} A_k \right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k \ge n} A_k \right) \le \lim_{n\to\infty} \sum_{k \ge n} \mathbb{P}(A_k) = 0.$$

□

Random variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *random variable* describes realized values for all source $\omega \in \Omega$ of randomness.

Definition A.13. We say that $X : \Omega \to \mathbb{R} \cup \{\pm\infty\}$ is an *$\mathcal{F}$-measurable random variable* if

$$\{\omega \in \Omega : X(\omega) > a\} \in \mathcal{F}, \quad a \in \mathbb{R}.$$

For $\mathbb{R}^d$-valued random function, we usually adopt the following definition:

Definition A.14. We say that $X : \Omega \to \mathbb{R}^n$ is an *$\mathcal{F}$-measurable random variable* if

$$X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}, \quad B \in \mathcal{B}(\mathbb{R}^n).$$

- Definition A.14 requires that for an arbitrary $B \in \mathcal{B}(\mathbb{R}^n)$, the event that $X(\omega) \in B$ belongs to the "well-defined" class $\mathcal{F}$ of random phenomenons.
- When $\mathcal{F}$ is referred to as an underlying σ-algebra, i.e., the σ-algebra $\mathcal{F}$ is the largest among those appeared in a specified problem, we simply say that X is a random variable.
- The event $\{\omega \in \Omega : X(\omega) \in B\}$ is often written as $\{X \in B\}$.

Sometimes it is convenient to consider a stochastic process as a random variable taking values in a function space. To this end, we describe a generalized version of Definition A.14.

Definition A.15. Let $(S, \mathcal{S})$ and $(U, \mathcal{U})$ be measurable spaces. A mapping $f : S \to U$ is said to be a *measurable mapping from* $(S, \mathcal{S})$ *into* $(U, \mathcal{U})$ if

$$f^{-1}(B) = \{f \in B\} \in \mathcal{S}, \quad \forall B \in \mathcal{U}.$$

In particular, when we work in a probability space $(S, \mathcal{S}, \mathbb{Q})$, the mapping f is said to be a U-valued random variable on $(S, \mathcal{S}, \mathbb{Q})$.

- In the case that U is a topological space, we say that a $\mathcal{B}(U)$-measurable mapping is *Borel measurable.*

Functions and limits of random variables are again random variables.

Proposition A.16

Let $(S, \mathcal{S})$ and $(U, \mathcal{U})$ be measurable spaces. Then we have the following:

(i) If $X : \Omega \to S$ and $f : S \to U$ are measurable, so is $f(X)$.

(ii) Let $\{X_n\}$ be a sequence of random variables $X_n : \Omega \to S$, then $\inf_n X_n$, $\sup_n X_n$, $\liminf_n X_n$, and $\limsup_n X_n$ are all random variables.

(iii) Suppose that Ω is a topological space and $\mathcal{F} = \mathcal{B}(\Omega)$. Then any continuous map $h : \Omega \to \mathbb{R}^n$ is measurable.

- Let $(S, \mathcal{S})$ be a measurable space. For $X : \Omega \to S$, the family

$$\sigma(X) := \{X^{-1}(B) : B \in \mathcal{S}\}$$

of subsets of Ω is the minimum σ-field such that X is measurable.

- One may adopt $\sigma(X_\lambda, \lambda \in \Lambda)$ as an underlying σ-filed when mappings X_λ on Ω, $\lambda \in \Lambda$, are the only random objects to be studied. That is, in that case, it is sufficient for us to set $\mathcal{F} = \sigma(X_\lambda, \lambda \in \Lambda)$.

That a random variable Y is measurable w.r.t. a σ-field $\mathcal{G}$ means that Y can be constructed by the information of $\mathcal{G}$. Precisely speaking, we have the following:

Theorem A.17

Let $(E, \mathcal{E})$ be a measurable space, and $X : \Omega \to \mathbb{R}$, and $Y : \Omega \to E$. Then a necessary and sufficient condition for which X is $\sigma(Y)$-measurable is that there exists an $\mathcal{E}$-measurable function $f : E \to \mathbb{R}$ such that $X = f(Y)$.

The well-known concept of the *distributions* is rigorously formulated in the measure theoretic probability.

Definition A.18. Let $(S, \mathcal{S})$ be a measurable space. Then for S-valued random variable X,

$$\mu_X(B) := P(X^{-1}(B)), \quad B \in \mathcal{S}$$

is a probability measure on $(S, \mathcal{S})$. We call this μ_X as the *distribution* of X.

- When X is real-valued, the nondecreasing and right-continuous function

$$F_X(x) := \mathbb{P}(X \in (-\infty, x]) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R},$$

  is said to be the *distribution function* of X.

- We say that a nonnegative Borel function f on $\mathbb{R}^d$ is a *probability density function* if

$$\int_{\mathbb{R}^d} f(x)dx = 1.$$

  For an $\mathbb{R}^d$-valued random variable X, when there exists a probability density function f such that

$$\mathbb{P}(X \in B) = \int_B f(x)dx, \quad B \in \mathcal{B}(\mathbb{R}^d),$$

  we say that the distribution of X has a density f.

Example A.19. Let $p \in [0, 1]$. Assume that the distribution μ of a $\{0, 1, \ldots, n\}$- valued random variable S_n is given by

$$\mu(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Then we say that S_n follows the *binomial distribution* with parameter (n, p), and write $S_n \sim B(n, p)$.

Example A.20. Let X be an $\mathbb{R}^d$-valued random variable, $m \in \mathbb{R}^d$, and $V \in \mathbb{R}^{d \times d}$ positive definite. We say that X follows a d-dimensional Gaussian distribution if the distribution μ of X satisfies

$$\mu(B) = \frac{1}{\sqrt{(2\pi)^d \det V}} \int_B \exp(-x^*(V^{-1})x/2)dx, \quad B \in \mathcal{B}(\mathbb{R}^d),$$

where $\det(V)$ is the determinant of V. Then we write $X \sim N(m, V)$.

Expectation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. In this section, we assume that all random variables are $\mathbb{R} \cup \{\pm\infty\}$-valued unless otherwise stated.

We define the indicator function 1_A of a set $A \subset \Omega$ by

$$1_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

- 1_A is measurable $\iff A \in \mathcal{F}$.

$X : \Omega \to \mathbb{R}$ is said to be a *simple function* if there exist $A_1, \ldots, A_n \in \mathcal{F}$ and $x_1, \ldots, x_n \in \mathbb{R}$ with

$$A_i \cap A_j = \emptyset \ (i \neq j), \ \ i, j = 1, \ldots, n, \ \ \Omega = \bigcup_{i=1}^{n} A_i$$

such that

$$X(\omega) = \sum_{i=1}^{n} x_i 1_{A_i}(\omega), \quad \omega \in \Omega. \tag{A.1}$$

- If X is a simple function of the form (A.1), then $X(\Omega) = \{x_1, \ldots, x_n\}$ and $\{X = x_i\} \cap \{X = x_j\} = \emptyset \ (i \neq j)$.

Suppose that $X : \Omega \to \mathbb{R}$ is a simple function having representation (A.1). Then we define the *expectation* $\mathbb{E}[X]$ of X by

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i \mathbb{P}(A_i).$$

- It should be emphasized that this definition is *well-defined*, i.e., $\mathbb{E}[X]$ is determined independently of the representations of X as a simple function.
- Notice that for simple functions X, Y with $X(\omega) \leq Y(\omega)$, $\omega \in \Omega$ (in many cases, this is simply written as $X \leq Y$), we have $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

We define the expectations of general random variables by some approximations with those of simple functions. To this end, we need the following lemma:

Lemma A.21

Let $X : \Omega \to \mathbb{R} \cup \{\pm\infty\}$. Then X is a random variable (i.e., $\mathcal{F}$-measurable) if and only if there exists a sequence $\{X_n\}_{n=1}^{\infty}$ of nonnegative simple functions such that for all $\omega \in \Omega$

$$\begin{aligned} 0 \leq X_1(\omega) \leq X_2(\omega) \leq \cdots \leq X(\omega), \\ \lim_{n\to\infty} X_n(\omega) = X(\omega). \end{aligned} \tag{A.2}$$

For any random variable X we define

$$X^+(\omega) := \max\{X(\omega), 0\}, \quad X^-(\omega) := -\min\{X(\omega), 0\}, \quad \omega \in \Omega.$$

The random variables X^+ and X^- are both nonnegative. It follows from Lemma A.21 that there exists a sequence $\{X_n^+\}$ (resp. $\{X_n^-\}$) of simple functions satisfying (A.2) for X^+ (resp. X^-). As remarked above, it follows that $\mathbb{E}[X_n^+] \leq \mathbb{E}[X_{n+1}^+]$, whence $\{\mathbb{E}[X_n^+]\}$ is nonnegative and nondecreasing. Hence the limit $\lim_{n\to\infty} \mathbb{E}[X_n^+] \in [0, \infty]$ exists. Similarly, the limit $\lim_{n\to\infty} \mathbb{E}[X_n^-] \in [0, \infty]$ exists. We define the expectation $\mathbb{E}[X]$ of X by

$$\mathbb{E}[X] = \lim_{n\to\infty} \mathbb{E}[X_n^+] - \lim_{n\to\infty} \mathbb{E}[X_n^-]$$

provided that at least one or both of the two limits are finite.

- This definition is also well-defined.
- Since $|X| = X^+ + X^-$, that $\mathbb{E}[X]$ is finite is equivalent to $\mathbb{E}[|X|] < \infty$.
- The expectation is nothing but the Lebesgue integral with respect to the measure $\mathbb{P}$ and so it can be written as
$$\mathbb{E}[X] = \int X(\omega)\mathbb{P}(d\omega) = \int X d\mathbb{P}.$$
Also, we often write $\mathbb{E}_{\mathbb{P}}[X]$ for the expectation of X to emphasize that it is defined under the probability measure $\mathbb{P}$.

Let X, Y be (real-valued) random variables and denote by $i = \sqrt{-1}$ the imaginary unit. Then $Z := X + iY$ is a complex-valued random variable, and we define its expectation by

$$\mathbb{E}[Z] = \mathbb{E}[X] + i\mathbb{E}[Y].$$

In particular, for a real-valued random variable X and $t \in \mathbb{R}$,

$$\mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)].$$

We list several basic properties of $\mathbb{E}[\cdot]$.

Proposition A.22

Let X and Y be random variables. Assume that the both $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ are defined. Then for $a, b \in \mathbb{R}$ we have the following:

(i) $X = Y$ a.s. $\Longrightarrow \mathbb{E}[X] = \mathbb{E}[Y]$.

(ii) $X \le Y$ a.s. $\Longrightarrow \mathbb{E}[X] \le \mathbb{E}[Y]$.

(iii) $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ (unless the right-hand side is $\infty - \infty$).

(iv) $|\mathbb{E}[X]| \le \mathbb{E}[|X|]$.

(v) $\mathbb{E}[|X|] < \infty \Longrightarrow |X| < \infty$ a.s.

(vi) $X \ge 0$ a.s., $\mathbb{E}[X] = 0 \Longrightarrow X = 0$ a.s.

(vii) $X \ge Y$ a.s., $\mathbb{E}[X] = \mathbb{E}[Y] \Longrightarrow X = Y$ a.s.

The expectation of a random variable can be given by the Lebesgue integral on the set which the variable takes values in.

Proposition A.23

Let $(S, \mathcal{S})$ be a measurable space, X an S-valued random variable, μ_X its distribution, and f a Borel measurable function on S. Then,

$$\mathbb{E}[f(X)] = \int_S f(x) d\mu_X(x).$$

Here, the equality means that if the right-hand side is finite then the other one is also finite and has the same value, and vice versa.

In general, for $\mathbb{R}^d$-valued random variable $X = (X_1, \ldots, X_d)$, we say that

$$\varphi_X(t) = \mathbb{E}\left[e^{i\sum_{k=1}^d t_k X_k}\right], \quad t = (t_1, \ldots, t_d) \in \mathbb{R}^d$$

is the *characteristic function* of X. The distribution of any random variable is completely determined by its characteristic function.

Proposition A.24

Let X and Y be $\mathbb{R}^d$-valued random variables. If $\varphi_X(t) = \varphi_Y(t)$ holds for any $t \in \mathbb{R}^d$, then $\mu_X = \mu_Y$.

Let $p \in [1, \infty]$. For real-valued random variable X, we set

$$\|X\|_p := \begin{cases} (\mathbb{E}[|X|^p])^{\frac{1}{p}} & (p \in [1, \infty)), \\ \inf\{a \geq 0 : |X| \leq a \text{ a.s.}\} & (p = \infty). \end{cases}$$

Denote by $L^p(\Omega, \mathcal{F}, \mathbb{P})$ by the totality of random variables such that $\|X\|_p < \infty$.

- Since $X = 0$ a.s. $\iff \|X\|_p = 0$, if we identify X with Y in the case of $X = Y$ a.s., then $\|\cdot\|_p$ defines a norm. By this identification, $L^p = L^p(\Omega, \mathcal{F}, \mathbb{P})$ becomes a Banach space (i.e., a complete normed space).
- Notice that for $1 \leq p \leq q$ and $X \in L^q$ we have $\|X\|_p \leq \|X\|_q$. Thus $X \in L^p$.
- L^2 is a real Hilbert space with the inner product
$$\langle X, Y \rangle = \mathbb{E}[XY].$$
- A random variable X is said to be *integrable* if $X \in L^1$, i.e., $\mathbb{E}[|X|] < \infty$.

The following several inequalities are frequently used.

Proposition A.25: Chebyshev's inequality

Let X be a nonnegative random variable. Then, for any nondecreasing function $f : [0, \infty) \to [0, \infty)$ and $x > 0$,

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}[f(X)]}{f(x)}.$$

Applying Proposition A.25 for $|X|$ and $f(x) = x$, we obtain the following:

Corollary A.26: Markov's inequality

For any $\mathbb{R}$-valued random variable X and any $x > 0$,

$$\mathbb{P}(|X| \geq x) \leq \frac{\mathbb{E}[|X|]}{x}.$$

Markov's inequality implies that if X is integrable then the tail probability $\mathbb{P}(|X| > x)$ decreases to zero faster than $O(1/x)$. If X has higher moments, then Chebyshev's inequality means that the tail more rapidly decreases to zero.

Proposition A.27: Jensen's inequality

Let X be an integrable random variable, and let $g : \mathbb{R} \to \mathbb{R}$ be convex. Then,

$$g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)].$$

Proposition A.28

Let $p, q \in (1, \infty)$ be such that $(1/p) + (1/q) = 1$. For $X, Z \in L^p$ and $Y \in L^q$ we have

(i) (Hölder's inequality)
$$|\mathbb{E}[XY]| \leq \|X\|_p \|Y\|_q;$$

(ii) (Minkowski's inequality)
$$\|X + Z\|_p \leq \|X\|_p + \|Z\|_p.$$

- Hölder's inequality with $p = 2$ is generally called the *Cauchy-Schwartz inequality.*

Convergence of random variables

Definition A.29. Let $X, X_1, X_2, \ldots$ be random variables.

(i) $\{X_n\}_{n=1}^{\infty}$ converges to X almost surely (we write $X_n \to X$ a.s.) $\stackrel{\text{def}}{\Longleftrightarrow}$ $\mathbb{P}(\{\omega : X_n(\omega) \to X(\omega)\}) = 1$.

(ii) $\{X_n\}_{n=1}^{\infty}$ converges to X in probability $\stackrel{\text{def}}{\Longleftrightarrow}$ $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$ $(n \to \infty)$ for any $\varepsilon > 0$.

(iii) $\{X_n\}_{n=1}^{\infty}$ converges to X in L^p $\stackrel{\text{def}}{\Longleftrightarrow}$ $\lim_{n\to\infty} \|X_n - X\|_p = 0$.

(iv) Assume that $X, X_1, X_2, \ldots$ are all $\mathbb{R}^d$-valued. Then $\{X_n\}_{n=1}^{\infty}$ converges to X in law (or in distribution) $\stackrel{\text{def}}{\Longleftrightarrow}$ $\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$ for any bounded continuous function f.

For $\mathbb{R}$-valued random variables, we have the following relations for the definitions of the convergences:

- $X_n \to X$ a.s. $\Longrightarrow$ $X_n \to X$ in probability.
- $X_n \to X$ in L^p $\Longrightarrow$ $X_n \to X$ in probability.
- $X_n \to X$ in probability $\Longrightarrow$ $X_n \to X$ in law.
- $X_n \to X$ in probability $\Longrightarrow$ $\lim_{k\to\infty} X_{n_k} = X$ a.s. for some subsequence $\{n_k\}_{k=1}^{\infty}$ with $\lim_{k\to\infty} n_k = \infty$.

The following three claims state the interchangeablity between the expectation and the limit of random variables.

Theorem A.30: Monotone convergence theorem

Let $\{X_n\}$ be a sequence of random variables such that $0 \leq X_1 \leq X_2 \leq \cdots$ a.s. Then

$$\mathbb{E}[X_n] \nearrow \mathbb{E}[X] \quad (n \to \infty).$$

Lemma A.31: Fatou lemma

Let $\{X_n\}$ be a sequence of almost surely nonnegative random variables. Then,

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n].$$

Theorem A.32: Dominated convergence theorem

Suppose that random variables X, X_n, $n \in \mathbb{N}$, satisfy the following:

(i) $X_n \to X$ a.s.

(ii) There exists a random variable $Y \in L^1$ such that $|X_n| \leq Y$ a.s. for all $n \in \mathbb{N}$.

Then,

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Independence and product spaces

Definition A.33. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

(i) $A, B \in \mathcal{F}$ are said to be *independent* of each other if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

(ii) A family $\{\mathcal{B}_i\}$, $i \in I$, of subsets of $\mathcal{F}$ is said to be independent if for distinct $i_1, \ldots, i_k \subset I$ we have

$$\mathbb{P}(B_{i_1} \cap \cdots \cap B_{i_k}) = \mathbb{P}(B_{i_1}) \cdots \mathbb{P}(B_{i_k}), \quad B_{i_j} \in \mathcal{B}_{i_j}, \ \ j = 1, \ldots, k.$$

(iii) Let $\{X_i\}_{i\in I}$ be a family of random variables. We say that X_i, $i \in I$, is independent if $\sigma(X_i)$, $i \in I$, is independent.

For given measurable spaces $(\Omega_k, \mathcal{F}_k)$, $k = 1, \ldots, n$, we call the σ-field

$$\prod_{k=1}^{n} \mathcal{F}_k = \sigma\left(\left\{\prod_{k=1}^{n} A_k : A_k \in \mathcal{F}_k, \ k = 1, \ldots, n\right\}\right)$$

as the *product σ-field* on $\prod_{k=1}^{n} \Omega_k$, and $(\prod_{k=1}^{n} \Omega_k, \prod_{k=1}^{n} \mathcal{F}_k)$ as the *product measurable space*.

Proposition A.34

We have $\mathcal{B}(\mathbb{R}^d) = \prod_{k=1}^{d} \mathcal{B}(\mathbb{R})$.

It is known that for probability spaces $(\Omega_k, \mathcal{F}_k, \mathbb{P}_k)$, $k = 1, \ldots, n$, there exists a unique probability measure $\prod_{k=1}^{n} \mathbb{P}_k$ on the product measurable space $(\prod_{k=1}^{n} \Omega_k, \prod_{k=1}^{n} \mathcal{F}_k)$ such that

$$(\textstyle\prod_{k=1}^{n} \mathbb{P}_k)(\prod_{k=1}^{n} A_k) = \prod_{k=1}^{n} \mathbb{P}_k(A_k), \quad A_k \in \mathcal{F}_k, \ \ k = 1, \ldots, n.$$

We call $\prod_{k=1}^{n} \mathbb{P}_k$ as the *product probability measure*, and $(\prod_{k=1}^{n} \Omega_k, \prod_{k=1}^{n} \mathcal{F}_k, \prod_{k=1}^{n} \mathbb{P}_k)$ as the *product probability space*.

Now, let $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ be given probability spaces. Here we will justfy the interchange of the order of integrations for functions on $\Omega_1 \times \Omega_2$. To this end, we need to confirm the measurability of the functions appeared in the iterated integrals. As for this point, it is straightforward to see that for any nonnegative and $\mathcal{F}_1 \times \mathcal{F}_2$-measurable function $X : \Omega_1 \times \Omega_2 \to \mathbb{R}$ the following four claims hold true:

- For $\omega_1 \in \Omega_1$ the function $X(\omega_1, \cdot) : \Omega_2 \to \mathbb{R}$ is $\mathcal{F}_1$-measurable.
- For $\omega_2 \in \Omega_2$ the function $X(\cdot, \omega_2) : \Omega_1 \to \mathbb{R}$ is $\mathcal{F}_2$-measurable.
- The function $\int_{\Omega_2} X(\cdot, \omega_2)\mathbb{P}_2(d\omega_2)$ on Ω_1 is a random variable on the probability space $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$.

- The function $\int_{\Omega_1} X(\omega_1, \cdot)\mathbb{P}_1(d\omega_1)$ on Ω_2 is a random variable on the probability space $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$.

Moreover, if X is not necessarily nonnegative but integrable on $\Omega_1 \times \Omega_2$ then we have the following two propositions:

- For $\mathbb{P}_1$-almost every (a.e.) $\omega_1 \in \Omega_1$, the function $X(\omega_1, \cdot) : \Omega_2 \to \mathbb{R}$ is $\mathcal{F}_1$-measurable, and the function $\int_{\Omega_2} X(\cdot, \omega_2)\mathbb{P}_2(d\omega_2)$ is a random variable on $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$.
- For $\mathbb{P}_2$-a.e. $\omega_2 \in \Omega_2$, the function $X(\cdot, \omega_2) : \Omega_1 \to \mathbb{R}$ is $\mathcal{F}_2$-measurable, and the function $\int_{\Omega_1} X(\omega_1, \cdot)\mathbb{P}_1(d\omega_1)$ is a random variable on $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$.

Basically, the expectation of a random variable on a product probability space is given by the iterated expectation.

Theorem A.35

Let $X(\omega_1, \omega_2)$ be a random variable on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, \mathbb{P}_1 \times \mathbb{P}_2)$.

(i) (Tonelli's theorem) If X is nonnegative, then

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_1\times\mathbb{P}_2}[X] &= \int_{\Omega_2}\left[\int_{\Omega_1} X(\omega_1,\omega_2)\mathbb{P}_1(d\omega_1)\right]\mathbb{P}(d\omega_2)\\ &= \int_{\Omega_1}\left[\int_{\Omega_2} X(\omega_1,\omega_2)\mathbb{P}_2(d\omega_2)\right]\mathbb{P}(d\omega_1).\end{aligned}$$

(ii) (Fubini's theorem) If X is integrable on $\Omega_1 \times \Omega_2$, then the equalities above also hold.

- To check the integrability of X, one may apply Tonelli's theorem for $|X|$ to try one of three integrals above that is easy to compute.

It is also known that Fubini-Tonelli theorem holds for product spaces involving the Lebesgue measure on $([0,\infty), \mathcal{B}[0,\infty))$. For example, if $X_\cdot : [0,\infty)\times\Omega \to \mathbb{R}$ is nonnegative and $\mathcal{B}[0,\infty)\times\mathcal{F}$-measurable, then $\int_0^\infty X_t(\omega)dt$ is an $\mathcal{F}$-measurable random variable and we have

$$\mathbb{E}\left[\int_0^\infty X_t dt\right] = \int_0^\infty \mathbb{E}[X_t]dt.$$

Next we summarize the relation between the independence and the product probability space.

Theorem A.36

Let $X_1, \dots, X_n$ be random variables, μ_i the distribution of X_i for $i = 1, \dots, n$, and μ the distribution of n-dimensional random variable $(X_1, \dots, X_n)$. Then $\{X_i\}_{i=1}^n$ is independent if and only if $\mu = \mu_1 \times \cdots \times \mu_n$.

This theorem leads to the following properties:

- Suppose that $X_1, \dots, X_n$ are independent and $f_1, \dots, f_n$ are Borel functions on $\mathbb{R}$. Then $f_1(X_1), \dots, f_n(X_n)$ are also independent.
- Suppose that $X_1, \dots, X_n$ are independent and integrable. Then
$$\mathbb{E}[X_1 \cdots X_n] = \mathbb{E}[X_1]\cdots\mathbb{E}[X_n].$$
- A necessary and sufficient condition for which random variables $X_1, \dots, X_n$ are independent is
$$\mathbb{E}\left[e^{i\sum_{k=1}^n t_k X_k}\right] = \prod_{k=1}^n \mathbb{E}[e^{it_k X_k}], \quad t_k \in \mathbb{R}, \;\; k = 1, \dots, n,$$
where $i = \sqrt{-1}$.

Change of probability measures

Let $(\Omega, \mathcal{F})$ be a measurable space.

Definition A.37. Let $\mathbb{Q}, \mathbb{P}$ be probability measures on $(\Omega, \mathcal{F})$. We say that $\mathbb{Q}$ is *absolutely continuous* with respect to $\mathbb{P}$ and write $\mathbb{Q} \ll \mathbb{P}$ if we have

$$\mathbb{P}(A) = 0,\ A \in \mathcal{F} \implies \mathbb{Q}(A) = 0.$$

- Suppose that $\mathbb{Q} \ll \mathbb{P}$. Then we have

  $$\mathbb{P}(A) = 1 \implies \mathbb{Q}(A) = 1.$$

  This means that an event almost surely occurs w.r.t. $\mathbb{P}$ also does w.r.t. $\mathbb{Q}$.

- If $\mathbb{Q} \ll \mathbb{P}$ and $\mathbb{P} \ll \mathbb{Q}$, then we say that $\mathbb{Q}$ and $\mathbb{P}$ are *equivalent* and write $\mathbb{Q} \sim \mathbb{P}$.

Theorem A.38: Radon-Nikodym theorem

Let $\mathbb{Q}, \mathbb{P}$ be probability measures on $(\Omega, \mathcal{F})$ such that $\mathbb{Q} \ll \mathbb{P}$. Then there exists an almost surely unique nonnegative random variable Y such that $\mathbb{E}[Y] = 1$ and

$$\mathbb{Q}(A) = \mathbb{E}[Y 1_A], \quad A \in \mathcal{F}.$$

- We say that the random variable Y as in Theorem A.38 is *Radon-Nikodym derivative* of $\mathbb{Q}$ with respect to $\mathbb{P}$, and write $\frac{d\mathbb{Q}}{d\mathbb{P}}$ for Y.

Limit theorems

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

Theorem A.39: Strong law of large number

Let $\{X_n\}$ be a sequence of independent random variables such that $\mathbb{E}[|X_1|] < \infty$. Then

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mathbb{E}[X_1] \ \text{ a.s.}$$

Theorem A.40: Central limit theorem

Let $\{X_n\}$ be an IID sequence with $X_1 \in L^2$ and $N \sim N(0,1)$. Then

$$\frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_1])}{\sqrt{n\mathbb{V}(X)}} \to N \ \text{ in law}, \quad n \to \infty.$$

Since any interval is a continuous set w.r.t. Gaussian measure, we have

$$\lim_{n\to\infty} \mathbb{P}(a < \textstyle\sum_{i=1}^n (X_i - \mu)/\sigma\sqrt{n} \le b) = \displaystyle\int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx, \quad -\infty \le a \le b < \infty,$$

provided that the central limit theorem holds.

Convergence of probability measures

Let (S,d) be a metric space. A sequence $\{\mu_n\}_{n=1}^{\infty}$ of probability measures on $(S,\mathcal{B}(S))$ is said to weakly converge to a probability measure μ on $(S,\mathcal{B}(S))$ if

$$\lim_{n\to\infty}\int_S f(x)\mu_n(dx)=\int_S f(x)\mu(dx)$$

for any bounded continuous function f on S.

Denote by $\overline{A}$ and $\mathring{A}$ the closure and interior of $A\in\mathcal{B}(S)$ respectively. We say that $\partial A:=\overline{A}\backslash\mathring{A}$ is the *boundary set* of A. Moreover, we say that $A\in\mathcal{B}(S)$ is a *μ-continuous set* if $\mu(\partial A)=0$.

Theorem A.41

Let $\{\mu_n\}$ be a sequence of probability measures on $(S,\mathcal{B}(S))$, μ a probability measure on $(S,\mathcal{B}(S))$. Then the following two claims are equivalent:

(i) $\{\mu_n\}$ weakly converges to μ.

(ii) For any μ-continuous set $A\in\mathcal{B}(S)$,

$$\lim_{n\to\infty}\mu_n(A)=\mu(A).$$

We often encounter the case of $S=C[0,\infty)$, the space of continuous functions on $[0,\infty)$. With the metric

$$d(\omega_1,\omega_2):=\sum_{n=1}^{\infty}\frac{1}{2^n}\max_{0\le t\le n}(|\omega_1(t)-\omega_2(t)|\wedge 1),$$

the space $C[0,\infty)$ is complete and separable, and the set $\mathcal{B}(C[0,\infty))$ of all Borel subsets of $C[0,\infty)$ is defined. To discuss the weak convergence in this space, we introduce the *modulus of continuity* of

$$m^T(\omega,\delta):=\max\{|\omega(t)-\omega(s)|:|s-t|\le\delta,\ 0\le s,t\le T\}$$

of $\omega\in C[0,\infty)$ on $[0,T]$ for each $\delta>0$ and $T>0$.

Theorem A.42

Suppose that a sequence $\{\mu_n\}_{n=1}^{\infty}$ of probability measures on $(C[0,\infty),\mathcal{B}(C[0,\infty)))$ satisfies the following two conditions:

(i) For each $\eta>0$ there exist $a\ge 0$ and $n_0\in\mathbb{N}$ such that

$$\mu_n(\omega:|\omega(0)|\ge a)\le\eta,\quad n\ge n_0.$$

(ii) For each $\varepsilon>0$, $T>0$, and $\eta>0$ there exist $\delta\in(0,1)$ and $n_0\in\mathbb{N}$ such that

$$\mu_n(\omega:m^T(\omega,\delta)>\varepsilon)\le\eta,\quad n\ge n_0.$$

There exists a subsequence $\{\mu_{n_k}\}_{k=1}^{\infty}$ that weakly converges to some probability measure on $(C[0,\infty),\mathcal{B}(C[0,\infty)))$.

Lemma on π-systems

Lemma A.44 is a tool for proving some propositions related to σ-algebras. For example, it will be useful when we aim to show that for two probability measures $\mathbb{P}$ and $\mathbb{Q}$ coincides with each other if $\mathbb{P}=\mathbb{Q}$ on a sub σ-algebras.

Definition A.43. Let Ω be a set. A family $\mathcal{C}$ of subsets of Ω is said to be π-*system* if $A \cap B \in \mathcal{C}$ for $A, B \in \mathcal{C}$.

Lemma A.44

Let $(\Omega, \mathcal{F})$ be a measurable space, $\mathcal{C}$ a π-system with $\sigma(\mathcal{C}) = \mathcal{F}$. If two probability measures $\mathbb{P}$ and $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ coincide with each other on $\mathcal{C}$, i.e., $\mathbb{P}(A) = \mathbb{Q}(A)$ for any $A \in \mathcal{C}$, then $\mathbb{P} = \mathbb{Q}$ on $\mathcal{F}$.

Bibliography

[1] F. M. Bandi and G. Moloche. On the functional estimation of multivariate diffusion processes. *Econometric Theory*, 34:896–946, 2018.

[2] G. Barles and P. E. Souganidis. Convergence of approximation schemes for fully nonlinear second order equations. *Asymptot. Anal.*, 4:271–283, 1991.

[3] R. Bellman. *Dynamic programming.* Princeton University Press, Princeton, 1957.

[4] A. Bensoussan. *Stochastic control of partially observable systems.* Cambridge University Press, Cambridge, 1992.

[5] P. Billingsley. *Convergence of Probability Measures.* John Wiley & Sons, New York, 2nd edition, 1999.

[6] Y. Chen, T. T. Georgiou, and M. Pavon. Stochastic control liaisons: Richard Sinkhorn meets Gaspard Monge on a Schrödinger bridge. *SIAM Rev.*, 63:249–313, 2021.

[7] R. Chetrite, P. Muratore-Ginanneschi, and K. Schwieger. E. Schrödinger's 1931 paper "On the Reversal of the Laws of Nature" ["Über die Umkehrung der Naturgesetze", Sitzungsberichte der preussischen Akademie der Wissenschaften, physikalisch-mathematische Klasse, 8 N9 144–153]. *Eur. Phys. J. H*, 46:1–29, 2021.

[8] M. G. Crandall, H. Ishii, and P.-L. Lions. A user's guide to viscosity solutions. *Bull. Amer. Math. Soc. (N.S.)*, 27:1–67, 1992.

[9] A. Dembo and O. Zeitouni. *Large deviations techniques and applications.* Springer, 1998.

[10] W. H. Fleming and R. W. Rishel. *Deterministic and stochastic optimal control.* Springer-Verlag, New York, 1975.

[11] W. H. Fleming and H. M. Soner. *Controlled Markov processes and viscosity solutions.* Springer-Verlag, New York, 2nd edition, 2006.

[12] H. Föllmer. Random fields and diffusion processes. In *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, pages 101–203. Springer, 1988.

[13] L. Györfi, M. Kohler, A. Krzyzak, and H. Walk. *A distribution-free theory of nonparametric regression.* Springer-Verlag, New York, 2002.

[14] U. G. Haussmann. Existence of optimal Markovian controls for degenerate diffusions. In *Lecture Notes in Control and Information Sciences*, volume 78, pages 171–186. Springer-Verlag, Berlin, 1986.

[15] U. G Haussmann and E. Pardoux. Time reversal of diffusions. *Ann. Probab.*, 14:1188–1205, 1986.

[16] S. M. Iacus. *Simulation and Inference for Stochastic Differential Equations.* Springer, Berlin, 2008.

[17] M. Ieda. *Numerical method for combined stochastic control and its application for mathematical finance.* PhD thesis, Tokyo Institute of Technology, Tokyo, Japan, 2018.

[18] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes.* North-Holland/Kodansha, Tokyo, 2nd edition, 2014.

[19] K. Itô. On stochastic differential equations. *Mem. Amer. Math. Soc.*, 4:1–51, 1951.

[20] K. Itô. Stationary random distributions. *Mem. College Sci. Univ. Kyoto Ser. A Math.*, 28:209–223, 1954.

[21] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus.* Springer-Verlag, New York, 1991.

[22] N. V. Krylov. *Controlled diffusion processes.* Springer-Verlag, New York, 1980.

[23] C. Léonard. Revisiting Fortet's proof of existence of a solution to the Schrödinger system. arXiv:1904.13211[Math.PR].

[24] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, 34:1533–1574, 2013.

[25] W. Liu and I. Hwang. Probabilistic aircraft conflict resolution: a stochastic optimal control approach. In *Proceedings of the 2012 American Control Conference*, pages 1714–1719, 2012.

[26] R. C. Merton. Lifetime portfolio selection under uncertainty: the continuous-time case. *Rev. Econom. Statist.*, 51:247–257, 1969.

[27] R. C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *J. Econom. Theory*, 3:373–413, 1971. Erratum: *ibid.*, 3:213–214, 1973.

[28] P. A. Meyer. *Probability and potentials.* Blaisdell Publishing Company, Waltham, Massachusetts, 1966.

[29] T. Mikami. Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probab. Theory Related Fields*, 129:245–260, 2004.

[30] M. Nisio. *Lectures on stochastic control theory.* ISI Lecture Notes, vol. 9. Macmillan India, Delhi, 1981.

[31] B. K. Øksendal. *Stochastic differential equations.* Springer, Berlin, 6th edition, 2013.

[32] H. Pham. *Continuous-time stochastic control and optimization with financial applications.* Springer, Berlin, 2009.

[33] B. L. S. Prakasa Rao. *Statistical inference for diffusion type processes.* John Wiley & Sons, London, 1999.

[34] E. Schrödinger. Über die umkehrung der naturgesetze. *Sitzungsberichte Preuss. Akad. Wiss. Berlin. Phys. Math.*, 144:144–153, 1931.

[35] E. Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *(French) Ann. Inst. H. Poincaré*, 2:269–310, 1932.

[36] N. Touzi. Optimal stochastic control, stochastic target problems, and backward SDE. Lecture Notes, May 2010. Ecole Polytechnique Paris.

[37] R. van Handel. Stochastic calculus and stochastic control. Lecture Notes, May 2010. California Institute of Technology.

[38] H. Wendland. *Scattered data approximation*. Cambridge University Press, Cambridge, 2010.

[39] D. Williams. *Probability with Martingales*. Cambridge University Press, Cambridge, 1991.

[40] J. Yong and X. Y. Zhou. *Stochastic controls: Hamiltonian systems and HJB equations*. Springer-Verlag, New York, 1999.

[41] N. Yoshida. Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.*, 41:220–242, 1992.

[42] 伊藤清. Markoff 過程ヲ定メル微分方程式. 全国紙上数学談話会, 244:1352–1400, 1942.

[43] 伊藤清三. ルベーグ積分入門. 裳華房, 1963.

[44] 佐藤担. 測度から確率へ. 共立出版, 1994.

[45] 舟木直久. 確率微分方程式. 岩波書店, 1997.

[46] 小谷眞一. 測度と確率. 岩波書店, 2005.

[47] 杉浦光夫. 解析入門 I. 東京大学出版会, 1980.

[48] 西尾真喜子. 確率論. 実教出版, 1978.

[49] 長井英生. 確率微分方程式. 共立出版, 1999.