Xu Zhang

Introduction to Stochastic Control

April 6, 2005

Springer
Berlin Heidelberg New York
Barcelona Hong Kong
London Milan Paris
Tokyo

Contents

1

Some Preliminaries in Probability Theory

1.1 Measure and probability, integral and expectation

1.1.1 Basic notations

Fix a nonempty set Ω and a family $\mathcal{F}$ of subsets of Ω. $\mathcal{F}$ is called a σ-field on Ω if, i) $\Omega \in \mathcal{F}$; ii) $\Omega \setminus A \in \mathcal{F}$ whenever $A \in \mathcal{F}$; and iii) $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ whenever each $A_i \in \mathcal{F}$.

If $\mathcal{F}$ is a σ-field on Ω, then $(\Omega, \mathcal{F})$ is called a measurable space. An element $A \in \mathcal{F}$ is called a measurable set on $(\Omega, \mathcal{F})$, or simply a measurable set. A map $\mu : \mathcal{F} \to [0, +\infty]$ is called a measure on $(\Omega, \mathcal{F})$ if $\mu(\emptyset) = 0$ and $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ whenever $A_i \in \mathcal{F}$ and $A_i \bigcap A_j = \emptyset$, $i, j = 1, 2, \cdots, i \neq j$. The triple $(\Omega, \mathcal{F}, \mu)$ is called a measure space. μ is called totally finite (*resp.* σ-finite) if $\mu(\Omega) < \infty$ (*resp.* there exist $A_i \subset \Omega$ so that $\Omega = \bigcup_{i=1}^{\infty} A_i$ and $\mu(A_i) < \infty$ for each i.

A probability space is a totally finite measure space $(\Omega, \mathcal{F}, P)$ for which $P(\Omega) = 1$; the measure P on a probability space is called a probability measure. In this case, a point $\omega \in \Omega$ is called a sample, any $A \in \mathcal{F}$ is called an event and $P(A)$ represents the probability of event A.

A measure space $(\Omega, \mathcal{F}, \mu)$ is said to be complete if for any μ-null set $A \in \mathcal{F}$, i.e. $\mu(A) = 0$, one has $B \in \mathcal{F}$ whenever $B \subset A$. Especially, one may define complete probability space $(\Omega, \mathcal{F}, P)$. If an event $A \in \mathcal{F}$ is such that $P(A) = 1$, then we may alternatively say that A holds, P-a.s., or simply A holds a.s.

If $(X, \mathcal{J})$ is a topological space, then the smallest σ-field containing all open sets $\mathcal{J}$ of X is called the Borel σ-field of X, denoted by $\mathcal{B}(X)$.

Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be two measurable spaces and $f : \Omega \to \Omega'$ be a map. We say f to be $\mathcal{F}/\mathcal{F}'$-measurable or simply measurable if $f^{-1}(\mathcal{F}') \subset \mathcal{F}$. In particular if $(\Omega', \mathcal{F}') = (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, then f is said to be a $\mathcal{F}$-measurable function. In the context of probability theory, f is called a $\mathcal{F}/\mathcal{F}'$-random variable or simply a random variable if no confusion. When $(\Omega', \mathcal{F}') = (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, it is called a $\mathcal{F}$-random variable. Note that measurable map or random variable is defined without measures. Also, it is clear that measurable map and random variable are in fact a same notation.

For a random variable $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$, $X^{-1}(\mathcal{F}')$ is a sub-σ-field of $\mathcal{F}$, which is called the σ-field generated by X, denoted by $\sigma(X)$.

Now, let us introduce an important notations, independence, which distinguishes probability theory from the usual measure theory.

Definition 1.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $A, B \in \mathcal{F}$. We say that A and B are independent if $P(A \cap B) = P(A)P(B)$. Let $\mathcal{J}_1$ and $\mathcal{J}_2$ be two subsets of $\mathcal{F}$. We say that $\mathcal{J}_1$ and $\mathcal{J}_2$ are independent if $P(A \cap B) = P(A)P(B)$ for any $A \in \mathcal{J}_1$ and $B \in \mathcal{J}_2$. Let $X,\ Y : (\Omega, \mathcal{F}) \to (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ be two random variables. We say that X and Y are independent if $\sigma(X)$ and $\sigma(Y)$ are independent.*

The following result is quite useful in probability theory.

Theorem 1.2. (Borel-Cantelli). *Suppose that A_i be a sequence of measurable sets of a measure space $(\Omega, \mathcal{F}, \mu)$.*

i) If $\sum_{i=1}^{\infty} \mu(A_i) < +\infty$, then

$$\mu(\overline{\lim_{i\to\infty}} A_i) = 0.$$

ii) If $(\Omega, \mathcal{F}, \mu)$ is a probability space, and if A_i is a sequence of mutually independent measurable sets, then the condition $\sum_{i=1}^{\infty} \mu(A_i) = +\infty$ implies that $\mu(\overline{\lim_{i\to\infty}} A_i) = 1$.

In the language of probability, i) states for probability contexts a condition that (almost surely) an event occurs only finitely often and ii) states that, in the independence case, if the condition is not satisfied, the event is almost sure to occur infinitely often.

Let $X_i, X_0 : (\Omega, \mathcal{F}, \mu) \to (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, $i = 1, 2, \cdots$, be measurable functions. We say that X_i converges to X a.e. if $\lim_{i\to\infty} |X_i - X_0| = 0$, μ-a.e. We say that X_i converges to X_0 in measure if for any $\varepsilon > 0$, $\lim_{i\to\infty} \mu\{|X_i - X_0| > \varepsilon\} = 0$. In particular, if μ is a probability measure, we say that X_i converges to X_0 in probability.

Theorem 1.3. (Comparison of convergence in measure and convergence almost everywhere). *Suppose that $(\Omega, \mathcal{F}, \mu)$ is a complete measure space, $X_i, X_0 : (\Omega, \mathcal{F}, \mu) \to (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, $i = 1, 2, \cdots$.*

i) If X_i converges to X_0 in measure, then there is a subsequence $\{X_{n_i}\}$ of $\{X_i\}$ such that $X_{n_i} \to X_0$, μ-a.e. as $i \to \infty$.

ii) Suppose in addition that μ is totally finite. Then the μ-a.e. convergence of X_i to X_0 implies that X_i converges to X_0 in measure.

A function f on $(\Omega, \mathcal{F})$ is said to be simple if there is a finite, disjoint class $\{E_1, E_2, \cdots, E_n\}$ of measurable sets and a finite set $\{\alpha_1, \alpha_2, \cdots, \alpha_n\}$ of real numbers such that

$$f(x) = \sum_{i=1}^{n} \alpha_i \chi_{E_i}(x), \quad x \in \Omega. \tag{1.1}$$

It is well-known that every measurable function is the limit of a sequence of simple functions.

A simple function of the form (1.1) on a measure space $(\Omega, \mathcal{F}, \mu)$ is said to be integrable (with respect to measure μ) if $\mu(E_i) < \infty$ for every index i for which $\alpha_i \neq 0$. The integral of f, in symbols

$$\int_\Omega f(x) d\mu(x), \quad \text{or} \quad \int_\Omega f d\mu,$$

is defined by

$$\int_\Omega f(x) d\mu(x) = \sum_{i=1}^{n} \alpha_i \mu(E_i).$$

A measurable function f on a measure space $(\Omega, \mathcal{F}, \mu)$ is said to be integrable (with respect to measure μ) if there exists a sequence $\{f_j\}$ of integrable simple functions such that $\{f_j\}$ converges to f a.e. and $\lim\limits_{j,k\to\infty} \int_\Omega |f_j - f_k| d\mu = 0$. Moreover, the integral of f is defined by

$$\int_\Omega f(x) d\mu(x) = \lim_{j\to\infty} \int_\Omega f_j(x) d\mu(x).$$

We refer to [1] for the properties of integrals defined above.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X : (\Omega, \mathcal{F}, P) \to \mathbb{R}^m$ be a random variable. If X is integrable, then we say that X has a mean, and denoted by

$$EX = \int_\Omega X dP.$$

We also call EX the (mathematical) expectation of X.

We denote $L^p_{\mathcal{F}}(\Omega; \mathbb{R}^m) \triangleq L^p(\Omega, \mathcal{F}, P; \mathbb{R}^m)$ by the set of all random variables X such that $|X|^p$ has means. This is a Banach space with the norm

$$|X|_{L^p_{\mathcal{F}}(\Omega;\mathbb{R}^m)} = \left(\int_\Omega |X|^p dP \right)^{1/p}. \tag{1.2}$$

In particular, $L^2_{\mathcal{F}}(\Omega; \mathbb{R}^m)$ is an Hilbert space.

The following results will play an important role in the sequel.

Theorem 1.4. *If X and Y are independent integral random variables on a probability space $(\Omega, \mathcal{F}, P)$ valued in $\mathbb{R}^m$, then XY^T is integrable, and*

$$E(XY^T) = (EX)(EY)^T. \tag{1.3}$$

1.1.2 Signed measures

We begin with the following definition.

Definition 1.5. *Let $(\Omega, \mathcal{F})$ be a measurable space. A map $\mu : \mathcal{F} \to [-\infty, +\infty]$ is called a signed measure if 1) $\mu(\emptyset) = 0$; 2) $\mu\Big(\bigcup_{j=1}^{\infty} A_j\Big) = \sum_{j=1}^{\infty} \mu(A_j)$ for all $A_j \in \mathcal{F}$ such that $A_j \cap A_k = \emptyset$, $j \neq k$, $j, k = 1, 2, \cdots$; and 3) μ assumes at most one of the values $+\infty$ and $-\infty$.*

Example 1.6. Let f be an integral function in $(\Omega, \mathcal{F}, \mu)$. Then

$$\nu(A) = \int_A f d\mu, \qquad \forall A \in \mathcal{F},$$

defines a signed measure in $(\Omega, \mathcal{F})$. More generally, the above ν is still a signed measure if f is a measurable function in $(\Omega, \mathcal{F})$, and one of f^+ and f^-, the positive and negative part of f, is integrable in $(\Omega, \mathcal{F}, \mu)$.

If μ is a signed measure on a measure space $(\Omega, \mathcal{F})$, we shall call a set $E \subset \Omega$ positive (*resp.* negative) (with respect to μ) if for every $F \in \mathcal{F}$, $E \cap F$ is measurable, and $\mu(E \cap F) \geq 0$ (*resp.* $\mu(E \cap F) \leq 0$).

Theorem 1.7. *If μ is a signed measure on a measure space $(\Omega, \mathcal{F})$, then there exist two disjoint sets A and B, whose union is Ω, such that A is positive and B is negative with respect to μ.*

The sets A and B in Theorem 1.7 are said to form a *Hahn decomposition of Ω with respect to μ*. It is not difficult to construct examples to show that a Hahn decomposition is not unique. If, however,

$$\Omega = A_1 \cup B_1 \quad \text{and} \quad \Omega = A_2 \cup B_2$$

are two Hahn decomposition of Ω, then it is easy to show that, for every measurable set E, it holds

$$\mu(E \cap A_1) = \mu(E \cap A_2) \quad \text{and} \quad \mu(E \cap B_1) = \mu(E \cap B_2).$$

From this fact, we may define unambiguously two set functions μ^+ and μ^- on $(\Omega, \mathcal{F})$ as follows:

$$\mu^+(E) = \mu(E \cap A), \quad \mu^-(E) = -\mu(E \cap B), \qquad \forall\, E \in \mathcal{F}.$$

We call μ^+ and μ^- respectively the upper variation and the lower variation of μ. The set function $|\mu|$, defined for every $E \in \mathcal{F}$ by

$$|\mu|(E) = \mu^+(E) + \mu^-(E)$$

is called the total variation of μ. Obviously

$$\mu(E) = \mu^+(E) - \mu^-(E), \qquad |\mu|(E) \ge |\mu(E)|.$$

Also, it is easy to show that the upper, lower, and the total variations of a signed measure μ are measures.

A signed measure μ on a measurable space $(\Omega, \mathcal{F})$ is said to be totally finite if for any $E \in \mathcal{F}$, one has $|\mu(E)| < \infty$; μ is said to be totally σ-finite if for any $E \in \mathcal{F}$, there exists a sequence $\{E_j\}_{j=1}^{\infty} \subset \mathcal{F}$ such that $|\mu(E_j)| < \infty$ for every j and $E \subset \bigcup_{j=1}^{\infty} E_j$.

Let $(\Omega, \mathcal{F})$ be a measurable space, $f : (\Omega, \mathcal{F}) \to (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ be a measurable function, μ be a signed measure on $(\Omega, \mathcal{F})$. If f is integrable with respect to $|\mu| = \mu^+ + \mu^-$, then f is said to be integrable in Ω with respect to μ. We call

$$\int_\Omega f d\mu^+ - \int_\Omega f d\mu^-$$

the integral of f in Ω with respect to μ, and denote it by $\int_\Omega f d\mu$. Let E be a measurable set, we define

$$\int_E f d\mu = \int_\Omega \chi_E f d\mu.$$

The properties of the integral with respect to a signed measure is similar to the usual integral except that all conditions such as "null measure sets" or "almost everywhere" with respect to μ should be changed to that with respect to $|\mu|$. Note that the μ-null set may be very "large".

Definition 1.8. *If $(\Omega, \mathcal{F})$ is a measurable space and μ and ν are two signed measures on $\mathcal{F}$, we say that ν is absolutely continuous with respect to μ, in symbols $\nu \ll \mu$, if $\nu(E) = 0$ for every $E \in \mathcal{F}$ with $|\mu|(E) = 0$.*

We proceed now to state the fundamental result concerning absolute continuity, which is known as Radon-Nikodym theorem.

Theorem 1.9. *Let $(\Omega, \mathcal{F})$ be a measurable space, μ and ν are two totally σ-finite signed measures and $\nu \ll \mu$. Then there is a finite valued measurable function f on Ω such that*

$$\nu(E) = \int_E f d\mu, \qquad \forall\, E \in \mathcal{F}.$$

The function f is unique in the sense that if also $\nu(E) = \int_E f_1 d\mu$ for some finite valued measurable function f_1 on Ω and for all $E \in \mathcal{F}$, then $f = f_1$, $|\mu| -$ a.e.

The function f in Theorem 1.9 is called Radon-Nikodym derivative, and denoted by

$$f = \frac{d\mu}{d\nu}.$$

By Theorem 1.9, it is easy to conclude that

Theorem 1.10. *Let the assumptions in Theorem 1.9 hold and g is a finite valued measurable function. Then g is integrable with respect to μ if and only if $g\frac{d\mu}{d\nu}$ is integrable with respect to ν. Furthermore,*

$$\int_E g d\mu = \int_E g \frac{d\mu}{d\nu} d\nu, \qquad \forall\, E \in \mathcal{F}.$$

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $(\Omega', \mathcal{F}')$ be a measurable space, and $\Phi : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ be a measurable map. Then Φ induces a measure ν on $(\Omega', \mathcal{F}')$ via

$$\nu(A') \stackrel{\triangle}{=} \mu(\Phi^{-1}(A')), \qquad \forall\, A' \in \mathcal{F}'. \tag{1.4}$$

The following is a change-of-variable formula:

Theorem 1.11. *Let ν be the measure induced by Φ from μ. Then f is integrable in Ω with respect to μ if and only if $f \circ \Phi$ is integrable in Ω' with respect to ν. Furthermore,*

$$\int_\Omega f d\mu = \int_{\Omega'} f \circ \Phi d\nu.$$

1.1.3 Distribution, density, characteristic function and normal distribution

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $(\Omega', \mathcal{F}')$ be a measurable space, and $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ be a random variable. Then X induces a probability measure P_X on $(\Omega', \mathcal{F}')$ via

$$P_X(A') \stackrel{\triangle}{=} P(X^{-1}(A')), \qquad \forall\, A' \in \mathcal{F}'. \tag{1.5}$$

We call P_X the distributions of random variable X. In the case of $(\Omega', \mathcal{F}') = (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, P_X can be uniquely determined by the following function:

$$F(x) \stackrel{\triangle}{=} F(x_1, \cdots, x_m) \stackrel{\triangle}{=} P\{X_i \le x_i,\ 1 \le i \le m\}, \tag{1.6}$$

where $x = (x_1, \cdots, x_m)$ and $(X_1, \cdots, X_m) = X$. We call $F(x)$ the distribution function of X.

If we assume P_X is absolutely continuous with respect to the Lebesgue measure, then by Radon-Nikodym theorem, there exists a (nonnegative) function $f \in L^1(\mathbb{R}^m)$ such that

$$P_X(A) = \int_A f(x) dx, \qquad \forall\, A \in \mathcal{B}(\mathbb{R}^m). \tag{1.7}$$

In particular, one has

$$F(x) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_m} f(\xi_1, \cdots, \xi_m) d\xi_1 \cdots d\xi_m. \tag{1.8}$$

The function $f(x)$ is called the density of the random variable X. As a special case, if $f(x)$ is of the following form:

$$f(x) = [(2\pi)^m \det C]^{-1/2} \exp\left\{-\frac{1}{2}(x-\lambda)C^{-1}(x-\lambda)^T\right\}, \quad x \in \mathbb{R}^m, \quad (1.9)$$

where $\lambda \in \mathbb{R}^m$, $C \in \mathbb{R}^{m\times m}$ with $C^T = C > 0$, then we say that X has a normal distribution with parameter (λ, C) and denoted by $\mathcal{N}(\lambda, C)$.

Now, for any $X, Y \in L^2_{\mathcal{F}}(\Omega; \mathbb{R}^m)$, we define the covariance (matrix) of X and Y by

$$\operatorname{Cov}(X,Y) \stackrel{\triangle}{=} E((X-EX)^T(Y-EY)). \quad (1.10)$$

In particular,

$$\operatorname{Var} X \stackrel{\triangle}{=} \operatorname{Cov}(X,X) \quad (1.11)$$

is called the variance (matrix) of X.

It is not difficult to show that for a normal distribution $\mathcal{N}(\lambda, C)$, λ is the mean and C is the covariance matrix.

Finally, for any $\mathbb{R}^m$-valued random variable X (which is not necessarily in $L^1_{\mathcal{F}}(\Omega; \mathbb{R}^m)$), the following is always well-defined:

$$\varphi_X(x) = Ee^{i\xi\cdot X} = \int_\Omega e^{i\xi\cdot X(\omega)} P(d\omega), \qquad \forall\, \xi \in \mathbb{R}^m. \quad (1.12)$$

We call $\varphi_X(x)$ the characteristic function of X.

Theorem 1.12. *A random variable X has a normal distribution $\mathcal{N}(\lambda, C)$ if and only if*

$$\varphi_X(\xi) = \exp\left\{i\xi\lambda^T - \frac{1}{2}\xi C\xi^T\right\}, \qquad \xi \in \mathbb{R}^m. \quad (1.13)$$

Note that, $\varphi_X(\xi)$ given by (1.13) is still a characteristic function even if $\det C = 0$. In this case, X is called a degenerate normal distribution.

1.1.4 Conditional expectation

We recall the following definition introduced in the elementary probability theory:

Definition 1.13. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $B \in \mathcal{F}$ with $P(B) > 0$. For any event $A \in \mathcal{F}$, put*

$$P(A\,|\,B) = \frac{P(A\cap B)}{P(B)}.$$

Then $P(\cdot\,|\,B)$ is a probability on $(\Omega, \mathcal{F})$, which is called the conditional probability given event B, and denoted by $P_B(\cdot)$. For any given $A \in \mathcal{F}$, $P(A\,|\,B)$ is called the conditional probability of A given B.

Definition 1.14. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and X be a random variable. The conditional expectation of X given an event $B \in \mathcal{F}$ with $P(B) > 0$ is defined by*

$$E(X \mid B) = \int_\Omega X dP_B = \frac{1}{P(B)} \int_B X dP.$$

Clearly, the conditional expectation of X given an event B represents the average value of X on B. On the other hand, it is easy to see that, the conditional probability of A given B can be regarded as a special case of conditional expectation, i.e.,

$$P(A \mid B) = E(\chi_A \mid B), \qquad A \in \mathcal{F}.$$

In many concrete problems, it is not enough to consider the conditional expectation given only one event. Instead, it is quite useful to define the conditional expectation to be a suitable random variable. For example, when consider two conditional expectations $E(X \mid B)$ and $E(X \mid B^c)$ simultaneously, we simply define it as a function $E(X \mid B)\chi_B(\omega) + E(X \mid B^c)\chi_{B^c}(\omega)$ rather than regarding it as two numbers. Generally, we introduce the following definition.

Definition 1.15. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\{B_k\}_{k=1}^\infty \subset \mathcal{F}$ be a partition of Ω (i.e., $\bigcup_{k=1}^\infty B_k = \Omega$ and $B_k \bigcap B_\ell = \emptyset$ whenever $k \neq \ell$), and $P(B_k) > 0$ for all $k = 1, 2, \cdots$. Put $\mathcal{J} = \sigma\{B_1, B_2, \cdots\}$ and assume X to be a random variable with mean EX. Then the following $\mathcal{J}$-measurable function*

$$E(X \mid \mathcal{J})(\omega) \triangleq \sum_{k=1}^\infty E(X \mid B_k)\chi_{B_k}(\omega)$$

is called the conditional expectation of X given σ-field $\mathcal{J}$.

Clearly, when $\mathcal{J} = \sigma\{B_1, B_2, \cdots\}$, the conditional expectation $E(X|\mathcal{J})$ of X takes its average value in every atom B_k of $\mathcal{J}$. Also, it is easy to check that

$$\int_{B_k} E(X \mid \mathcal{J}) dP = \int_{B_k} X dP, \qquad k = 1, 2, \cdots. \tag{1.14}$$

Now, for any given probability space $(\Omega, \mathcal{F}, P)$, let us consider the more general case that $\mathcal{J} \subset \mathcal{F}$ is any given sub-σ-field (hence $\mathcal{J}$ need not to be a partition of Ω). Let X be a random variable with mean EX (which may be infinite). Stimulated by (1.14), we define a set function on $\mathcal{J}$ by

$$\nu(B) \triangleq \int_B X dP, \qquad \forall\, B \in \mathcal{J}.$$

It is easy to see that ν is a signed measure on $(\Omega, \mathcal{J})$. In view of Radon-Nikodym theorem, one may find a unique $\mathcal{J}$-measurable function, denoted by $E(X \mid \mathcal{J})$, such that

$$\int_B E(X \mid \mathcal{J}) dP = \int_B X dP, \qquad \forall\, B \in \mathcal{J}. \tag{1.15}$$

This leads to the following notion:

Definition 1.16. *$E(X \mid \mathcal{J})$, determined by (1.15), is called the conditional expectation of X given σ-field $\mathcal{J}$.*

We collect some basic properties of the conditional expectation as follows.

Theorem 1.17. *Let $\mathcal{J}$ be a sub-σ-field of $\mathcal{F}$. Then the following conclusions hold:*

(1) The map $E(\cdot \mid \mathcal{J}) :\ L^1_{\mathcal{F}}(\Omega, \mathbb{R}^m) \to L^1_{\mathcal{J}}(\Omega, \mathbb{R}^m)$ is linear and continuous.

(2) $E(a \mid \mathcal{J}) = a$, $P|_{\mathcal{J}} -$ a.s., $\forall\, a \in \mathbb{R}$.

(3) If $X, Y \in L^1_{\mathcal{F}}(\Omega, \mathbb{R})$ with $X \geq Y$, then

$$E(X \mid \mathcal{J}) \geq E(Y \mid \mathcal{J}), \qquad P|_{\mathcal{J}} - \text{a.s.}$$

(4) Let $X \in L^1_{\mathcal{J}}(\Omega; \mathbb{R}^m)$ and $Y \in L^1_{\mathcal{F}}(\Omega; \mathbb{R}^k)$ with $XY^T \in L^1_{\mathcal{F}}(\Omega; \mathbb{R}^{m\times k})$. Then

$$E(XY^T \mid \mathcal{J}) = XE(Y \mid \mathcal{J})^T, \qquad P|_{\mathcal{J}} - \text{a.s.}$$

In particular, $E(X \mid \mathcal{J}) = X$, $P|_{\mathcal{J}} -$ a.s. Also, for any $Z \in L^1_{\mathcal{F}}(\Omega; \mathbb{R}^m)$,

$$E(E(Z \mid \mathcal{J})Y^T \mid \mathcal{J}) = E(Z \mid \mathcal{J})E(Y \mid \mathcal{J})^T, \qquad P|_{\mathcal{J}} - \text{a.s.}$$

(5) A random variable X is independent of $\mathcal{J}$ if and only if for any Borel measurable function f such that $Ef(X)$ exists, it holds

$$E(f(X) \mid \mathcal{J}) = Ef(X), \qquad P|_{\mathcal{J}} - \text{a.s.}$$

(6) Let $\mathcal{J}' \subset$ be a sub-σ-field of $\mathcal{J}$. Then

$$E(E(X \mid \mathcal{J}) \mid \mathcal{J}') = E(E(X \mid \mathcal{J}') \mid \mathcal{J}) = E(X \mid \mathcal{J}'), \qquad P|_{\mathcal{J}'} - \text{a.s.}$$

(7) (Jensen's inequality) Let $X \in L^1_{\mathcal{F}}(\Omega; \mathbb{R}^m)$ and $\phi :\ \mathbb{R}^m \to \mathbb{R}$ be a convex function such that $\phi(X) \in \mathbb{R}^m \to \mathbb{R}$. Then

$$\phi(E(X \mid \mathcal{J})) \leq E(\phi(X) \mid \mathcal{J}), \qquad P|_{\mathcal{J}} - \text{a.s.}$$

In particular, for any $p \geq 1$, we have

$$\Big| E(X \mid \mathcal{J}) \Big|^p \leq E(|X|^p \mid \mathcal{J}), \qquad P|_{\mathcal{J}} - \text{a.s.}$$

provided that $E|X|^p$ exists.

Remark 1.18. Given two sub-σ-fields $\mathcal{J}^k$ $(k = 1, 2)$ of $\mathcal{F}$ and a random variable X, generally

$$E(E(X|\mathcal{J}^1)|\mathcal{J}^2) \neq E(E(X|\mathcal{J}^2)|\mathcal{J}^1) \neq E(X|\mathcal{J}^1 \cap \mathcal{J}^2), \qquad P|_{\mathcal{J}^1\cap\mathcal{J}^2} - \text{a.s.}$$

1.2 Stochastic processes

In this section, we recall some elements on stochastic processes. In particular, we will study a special class of stochastic process, Brownian Motion, which will play a fundamental role in the sequel.

1.2.1 General considerations

Let us fix a probability space $(\Omega, \mathcal{F}, P)$. We begin with the following definition:

Definition 1.19. *Let I be a nonempty index set and (U, d) be a metric space with metric d. A family of random variables $\{X(t)\}_{t\in I}$ from $(\Omega, \mathcal{F}) \to (U, \mathcal{B}(U))$ is called a stochastic process. For any $\omega \in \Omega$, the map $t \in X(t, \omega)$ is called a sample path (of X).*

In what follows, we will choose $I = [0, T]$ with $T > 0$, or $I = [0, \infty)$. We will interchangeably use $\{X(t)\}_{t\in I}$, $X(\cdot)$, $X(t)$, or even X to denote a stochastic process. Also, a stochastic process will be simply called a process if no ambiguity.

Usually, we will choose (U, d) as $\mathbb{R}^m$ with the standard topology. In this case, for a given stochastic process $\{X(t)\}_{t\in I}$, we set

$$F_{t_1,\cdots,t_j}(x_1, \cdots, x_j) \overset{\triangle}{=} P\Big\{X(t_1) \le x_1, \cdots, X(t_j) \le x_j\Big\}, \tag{1.16}$$

where $j = 1, 2, \cdots$, $t_i \in I$, $x_i \in \mathbb{R}^m$ and $X(t_i) \le x_i$ stands for componentwise inequalities $(i = 1, \cdots, j)$. Functions defined in (1.16) are called the finite dimensional distributions of process X.

As that done by the distribution function of random variable, the finite dimensional distributions $F_{t_1,\cdots,t_j}(x_1, \cdots, x_j)$ of X include the main probabilities of the process.

Definition 1.20. *Two processes $X(t)$ and $\overline{X}(t)$ are said to be stochastically equivalent if*

$$X(t) = \overline{X}(t), \qquad P-\text{a.s.}, \quad \forall\, t \in I.$$

In this case, one is said to be a modification of the other.

Obviously, when $X(t)$ and $\overline{X}(t)$ are stochastically equivalent, their finite dimensional distributions are the same. However, the P-null set N_t depends on t. Therefore, the sample paths of $X(t)$ and $\overline{X}(t)$ can differ significantly. Here is a simple example.

Example 1.21. Let $\Omega = [0, 1]$, $T \ge 1$, P be the Lebesgue measure, $X(t, \omega) \equiv 0$, and

$$\overline{X}(t, \omega) = \begin{cases} 0, & \omega \neq t, \\ 1, & \omega = t. \end{cases}$$

Then $X(t)$ and $\overline{X}(t)$ are stochastically equivalent. But, each sample path $X(\cdot\,,\omega)$ is continuous and none of the sample paths $\overline{X}(\cdot\,,\omega)$ is continuous. In the present case, we actually have

$$\bigcup_{t\in[0,1]} N_t = [0,1] \equiv \Omega.$$

Definition 1.22. *We say that a process $X(t)$ to be stochastically continuous at $s \in [0,T]$ if for any $\varepsilon > 0$,*

$$\lim_{t\to s} P\{|X(t) - X(s)| > \varepsilon\} = 0.$$

Moreover, $X(t)$ is said to be continuous if there is a $P-$null set $N \in \mathcal{F}$, such that for any $\omega \in \Omega \setminus N$, the sample path $X(\cdot,\omega)$ is continuous.

Similarly, one can define the left and the right (stochastically) continuity of a process. It is obvious that continuity implies stochastic continuity.

Next, for a given measurable space $(\Omega, \mathcal{F})$, we introduce a monotone family of sub-σ-fields $\mathcal{F}_t \subset \mathcal{F}$, $t \in [0,T]$. Here, by monotonicity we mean that

$$\mathcal{F}_{t_1} \subset \mathcal{F}_{t_2}, \qquad \forall\, 0 \le t_1 \le t_2 \le T.$$

Such a family is called a filtration. For any $t \in [0,T]$, we put

$$\mathcal{F}_{t+} \stackrel{\triangle}{=} \bigcap_{s>t} \mathcal{F}_s, \qquad \mathcal{F}_{t-} \stackrel{\triangle}{=} \bigcup_{s<t} \mathcal{F}_s.$$

If $\mathcal{F}_{t+} = \mathcal{F}_t$ (*resp.* $\mathcal{F}_{t-} = \mathcal{F}_t$), then $\{\mathcal{F}_t\}_{t\ge 0}$ is said to be right (*resp.* left) continuous. We call $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\ge0})$ a filtered measurable space and $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\ge0}, P)$ a filtered probability space.

In the sequel, we will say that $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\ge0}, P)$ satisfy the usual condition if $(\Omega, \mathcal{F}, P)$ is complete, $\mathcal{F}_0$ contains all $P-$null sets in $\mathcal{F}$, and $\{\mathcal{F}_t\}_{t\ge0}$ is right continuous.

Definition 1.23. *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\ge0})$ be a filtered measurable space and $X(t)$ be a process which values in a metric space (U, d).*

(1) $X(t)$ is said to be measurable if the map $(t,\omega) \mapsto X(t,\omega)$ is $(\mathcal{B}[0,\infty) \times \mathcal{F})/\mathcal{B}(U)$-measurable;

(2) $X(t)$ is said to be $\{\mathcal{F}_t\}_{t\ge0}$-adapted if for all $t \ge 0$, the map $\omega \mapsto X(t,\omega)$ is $\mathcal{F}_t/\mathcal{B}(U)$-measurable;

(3) $X(t)$ is said to be $\{\mathcal{F}_t\}_{t\ge0}$-progressively measurable if for all $t \ge 0$, the map $(s,\omega) \mapsto X(s,\omega)$ is $(\mathcal{B}[0,t] \times \mathcal{F}_t)/\mathcal{B}(U)$-measurable.

It is clear that if $X(t)$ is $\{\mathcal{F}_t\}_{t\ge0}$-progressively measurable, it must be measurable and $\{\mathcal{F}_t\}_{t\ge0}$-adapted. Conversely, it can be proved that, for any measurable and $\{\mathcal{F}_t\}_{t\ge0}$-adapted process $X(t)$ on a filtered probability space, there is a $\{\mathcal{F}_t\}_{t\geq 0}$-progressively measurable process $\widetilde{X}(t)$ which is stochastically equivalent to $X(t)$. For this reason, unless otherwise indicated, in the sequel, in a filtered probability space, by saying that a process $X(t)$ is $\{\mathcal{F}_t\}_{t\geq 0}$-adapted, we mean that $X(t)$ is $\{\mathcal{F}_t\}_{t\geq 0}$-progressively measurable.

Let H be a Banach space and $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ be a filtered probability space satisfying the usual condition. We denote by $L^p_{\mathcal{F}}(0,T;H)$ the set of all $H-$valued $\{\mathcal{F}_t\}_{t\geq 0}$-adapted process $X(\cdot)$ such that

$$E\int_0^T |X(t)|_H^p dt < \infty;$$

by $L^\infty_{\mathcal{F}}(0,T;H)$ the set of all $H-$valued $\{\mathcal{F}_t\}_{t\geq 0}$-adapted essentially bounded processes; and by $L^p_{\mathcal{F}}(\Omega; C([0,T];H))$ the set of all $H-$valued $\{\mathcal{F}_t\}_{t\geq 0}$-adapted continuous process $X(\cdot)$ such that

$$E(|X(\cdot)|^p_{C([0,T];H)}) < \infty.$$

In the sequel, we simply write $L^p_{\mathcal{F}}(0,T;\mathbb{R})$ as $L^p_{\mathcal{F}}(0,T)$. One has a similar notations for $L^\infty_{\mathcal{F}}(0,T)$ and $L^p_{\mathcal{F}}(\Omega; C[0,T])$.

1.2.2 Brownian motion

We now introduce an extremely important example of stochastic process, called the Brownian motion.

The Brownian motion of pollen particles in a liquid, which owes its name to its discovery by the English botanist R. Brown in 1827, is due to the incessant hitting of pollen by the much small molecules of the liquid. The hits occur a large number of times in any small interval of time, independently of each other and the effect of a particular hit is so small compared to the total effect. The physical theory of this motion was set up by A. Einstein in 1905. It's suggested that this motion is random, and has the following properties:

1) The displacement of a pollen particle over disjoint time intervals are independent (random variables);

2) The displacements are Gaussian random variables;

3) The motion is continuous.

Property 1) means the Brownian motion has independent increments. Property 2) is not surprising in view of the central-limit theorem.

Now, let us give a mathematical definition of the Brownian motion.

Definition 1.24. *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ be a filtered probability. A continuous $\{\mathcal{F}_t\}_{t\geq 0}$-adapted $\mathbb{R}^m$-valued process $W(\cdot)$ is called a m-dimensional $\{\mathcal{F}_t\}_{t\geq 0}$-Brownian motion over $[0,\infty)$, if for all $0 \leq s < t < \infty$, $W(t) - W(s)$ is independent of $\mathcal{F}_s$, and is normally distributed with mean 0 and covariance $(t-s)I$. In addition, if $P(W(0) = 0) = 1$, then $W(\cdot)$ is called a m-dimensional standard $\{\mathcal{F}_t\}_{t\geq 0}$-Brownian motion over $[0,\infty)$.*

One can also define a Brownian motion $W(\cdot)$ naturally over any time interval $[a,b]$ or $[a,b)$ for any $0 \le a < b \le \infty$.

In general, if $W(\cdot)$ is a Brownian motion defined on some filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\ge 0}, P)$, we may define

$$\mathcal{F}_t^W \stackrel{\triangle}{=} \sigma(W(s),\ 0 \le s \le t) \subset \mathcal{F}_t, \qquad \forall\, t \ge 0. \tag{1.17}$$

Generally, filtration $\{\mathcal{F}_t^W\}_{t\ge 0}$ is left-continuous, but not necessarily right-continuous. On the other hand, the augmentation $\{\hat{\mathcal{F}}_t^W\}_{t\ge 0}$ of $\{\mathcal{F}_t^W\}_{t\ge 0}$ by adding all P-null sets is continuous, and $W(t)$ is still a Brownian motion on the (augmented) filtered probability space $(\Omega, \mathcal{F}, \{\hat{\mathcal{F}}_t\}_{t\ge 0}, P)$ (*see* [2, p. 89 and p. 122] for detailed discussion). In the sequel, by saying that $\{\mathcal{F}_t\}_{t\ge 0}$ is the natural filtration generated by the Brownian motion W, we mean that $\{\mathcal{F}_t\}_{t\ge 0}$ is generated as (1.17) with the above augmentation, (and hence $\{\mathcal{F}_t\}_{t\ge 0}$ is continuous).

We collect a few properties of Brownian motion.

Proposition 1.25. *Let W be a standard Brownian motion. Then*

(i) (time-homogeneity). For any $a > 0$, the process $W(t+a) - W(a)$, $t \ge 0$, is a Brownian motion independent of $\sigma(W(u),\ u \le a)$;

(ii) (symmetry). The process $-W(t)$, $t \ge 0$, is a Brownian motion;

(iii) (scaling). For any $\lambda > 0$, the process $\lambda^{-1}W(\lambda^2 t)$, $t \ge 0$, is a Brownian motion;

(iv) (time-inversion). The process X, defined by $X_0 = 0$, $X(t) = tW(t^{-1})$ for $t > 0$, is a Brownian motion;

(v) For any $x \in \mathbb{R}^m$, the process $W^x(t) = x + W(t)$ is called the Brownian motion started at x. Obviously, for any $A \in \mathcal{B}(\mathbb{R}^m)$, $t > 0$,

$$P(W^x(t) \in A) = \frac{1}{(2\pi t)^{m/2}} \int_A \exp\left\{-\frac{|x-y|^2}{2t}\right\} dy.$$

(vi) If $W_1(t), \cdots, W_k(t)$ are k independent copies of a m-dimensional Brownian motion $W_0(t)$, then $W(t) = (W_1(t), \cdots, W_k(t))^T$ is a mk-dimensional Brownian motion. Obviously, if $W(t)$ is a mk-dimensional Brownian motion, then $W_1(t), \cdots, W_k(t)$ are k independent m-dimensional Brownian motion.

(vii) For any $0 \le s \le t$, it holds

$$\begin{cases} E(W(t) - W(s) \mid \mathcal{F}_s) = 0, & P-\text{a.s.} \\ E((W(t) - W(s)(W(t) - W(s))^T \mid \mathcal{F}_s) = (t-s)I, & P-\text{a.s.} \end{cases}$$

We shall define the integral

$$\int_0^T X(t)dW(t) \tag{1.18}$$

of a stochastic process $X(\cdot)$ with respect to a Brownian motion $W(t)$. Such an integral will play an essential role in the sequel. Note that if for $\omega \in \Omega$, the map $t \mapsto W(t,\omega)$ was of bounded variation, then a natural definition of (1.18) would be a Lebesgue-Stieltjes type integral, regarding ω as a parameter. Unfortunately, we will see below that the map $t \mapsto W(t,\omega)$ is not of bounded variation for almost all ω. Thus one needs to define (1.18) in a different way.

Theorem 1.26. *Let $W(t)$ be a (one-dimensional) Brownian motion. Then, for almost all $\omega \in \Omega$, the map $t \mapsto W(t,\omega)$ is nowhere differentiable.*

Proof. It suffices to show that

$$P(\overline{\lim_{s\to 0+}} \frac{|W(t+s)-W(t)|}{s} = +\infty) = 1, \qquad \forall\, t \geq 0.$$

For this purpose, for any $t \geq 0$, $s > 0$ and positive integer n, denote

$$A_{t,n,s} \overset{\triangle}{=} (|W(t+s)-W(t)| < ns).$$

Noting that $W(t+s)-W(t) \sim \mathcal{N}(0,s)$, we get

$$P(A_{t,n,s}) = \frac{1}{\sqrt{2\pi s}} \int_{-ns}^{ns} e^{-\frac{x^2}{2s}}\, dx \leq C\sqrt{s},$$

where $C = C(n) > 0$ is a generic constant. We choose a sequence $\{s_k\}_{k=1}^{\infty}$ so that $\sum_{k=1}^{\infty} \sqrt{s_k} < \infty$. Then

$$\sum_{k=1}^{\infty} P(A_{t,n,s_k}) < \infty.$$

Now, thanks to the Borel-Cantalli's theorem, we conclude that

$$P(\underline{\lim_{k\to\infty}}\, (A_{t,n,s_k})^c) = 1.$$

Denote

$$A_{t,n} = \underline{\lim_{k\to\infty}}\, (A_{t,n,s_k})^c.$$

Then for any $\omega \in A_{t,n}$, there is a $k(\omega)$, such that

$$\left| \frac{W(t+s_k,\omega)-W(t,\omega)}{s_k} \right| \geq n, \qquad \forall\, k \geq k(\omega).$$

Therefore,

$$\overline{\lim_{s\to 0+}} \frac{|W(t+s,\omega)-W(t,\omega)|}{s} \geq n, \qquad \forall\, \omega \in A_{t,n}.$$

Put

$$B = \bigcap_{n\geq 1} A_{t,n}.$$

Then, $P(B) = 1$, and

$$\overline{\lim_{s\to 0+}} \frac{|W(t+s,\omega) - W(t,\omega)|}{s} = +\infty, \qquad \forall\, \omega \in B.$$

This completes the proof. □

1.3 Stopping times

In this section, we shall introduce a special class of random variables, which is one of fundamental notions in the modern probability theory.

Definition 1.27. *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0})$ be a filtered measurable space. A mapping $\tau : \ \Omega \to [0,\infty]$ is called a $\{\mathcal{F}_t\}_{t\geq 0}$-stopping time if*

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

Obviously, when τ is a stopping time, it is a positive random variable taking possibly infinite values. Further, define

$$\mathcal{F}_\tau = \Big\{ A \in \mathcal{F} \;\Big|\; A \cap (\tau \leq t) \in \mathcal{F}_t,\ \forall\, t \geq 0 \Big\}.$$

It is clear that $\mathcal{F}_\tau$ is a sub-σ-field of $\mathcal{F}$. The sets in $\mathcal{F}_\tau$ can be thought of as events which may occur before time τ. The constants, i.e., $\tau(\omega) \equiv s$ for every ω, are stopping times and in that case $\mathcal{F}_\tau = \mathcal{F}_s$ (recall that $\{\mathcal{F}_t\}$ can be considered as describing the history of some phenomenon and $\mathcal{F}_t$ is the σ-fields of events prior to some t). Stopping times thus appear as generalizations of constant times for which one can define a "past" which is consistent with the "pasts" of constant times. A stopping time τ is a random variable such that the event "τ has occurred up to time t" depends only on the history up to time t, and not on any further information about future.

By means of the right-continuity of $\mathcal{F}_t$, one has

Proposition 1.28. *(i) A map $\tau : \ \Omega \to [0,\infty]$ is a stopping time if and only if $(\tau < t) \in \mathcal{F}_t$ for all $t > 0$.*

(ii) Let τ be a stopping time. Then $A \in \mathcal{F}_\tau$ if and only if $A \cap (\tau < t) \in \mathcal{F}_t$ for all $t > 0$.

A stopping time may be thought of as the first time when some physical event occurs.

Example 1.29. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ be a filtered probability space and $X(t)$ be a $\{\mathcal{F}_t\}_{t\geq 0}$-adapted and continuous process with values in $\mathbb{R}^m$. Let $V \subset \mathbb{R}^m$ be an open set. Then the first hitting time of the process $X(t)$ to V, i.e.,

$$\sigma_V(\omega) \stackrel{\triangle}{=} \inf \Big\{ t \geq 0 \;\Big|\; X(t,\omega) \in V \Big\},$$

and the first exit time of the process $X(t)$ from V, i.e.,

$$\tau_V(\omega) \stackrel{\triangle}{=} \inf \Big\{ t \geq 0 \;\Big|\; X(t,\omega) \notin V \Big\}$$

are both $\{\mathcal{F}_t\}_{t\geq 0}$-stopping times. (Here, we agree that $\inf \emptyset \stackrel{\triangle}{=} +\infty$).

The basic properties of stopping times are listed in the following proposition.

Proposition 1.30. *Let σ, τ and σ_i $(i = 1, 2, \cdots)$ are stopping times. Then*

(i) $\sigma + \tau$, $\sup\limits_i \sigma_i$, $\inf\limits_i \sigma_i$, $\overline{\lim\limits_{i\to\infty}}\ \sigma$ and $\underline{\lim\limits_{i\to\infty}}\ \sigma$ are stopping times. Also, events $(\sigma > \tau)$, $(\sigma \geq \tau)$ and $(\sigma = \tau)$ belong to $\mathcal{F}_{\sigma\wedge\tau}$;

(ii) The process $Y(t) \stackrel{\triangle}{=} \tau \wedge t$ is $\{\mathcal{F}_t\}_{t\geq 0}$-progressively measurable;

(iii) For any $A \in \mathcal{F}_\sigma$, it holds

$$A \cap (\sigma \leq \tau) \in \mathcal{F}_\tau.$$

In particular, if $\sigma \leq \tau$, P-a.s., then $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$;

(iv) Let $\hat{\sigma} = \inf\limits_i \sigma_i$. Then

$$\bigcap_i \mathcal{F}_{\sigma_i} = \mathcal{F}_{\hat{\sigma}};$$

In particular, $\mathcal{F}_{\sigma_1} \cap \mathcal{F}_{\sigma_2} = \mathcal{F}_{\sigma_1\wedge\sigma_2}$.

The following result will be sometimes technically useful.

Proposition 1.31. *Every stopping time is the decreasing limit of a sequence of stopping times taking only finitely many values.*

Proof. For a stopping time τ, one sets

$$\tau_k = \sum_{q=1}^{k2^k} \frac{q}{2^k} \chi_{(\frac{q-1}{2^k} \leq \tau < \frac{q}{2^k})} + (+\infty)\chi_{(\tau \geq k)}.$$

It is easy to check that τ_k is a stopping time and that $\{\tau_k\}$ decreases to τ. □

The following result provides a characterization of $\mathcal{F}_\tau$-random variable.

Proposition 1.32. *Let τ be a stopping time and ξ be a random variable with values in $\mathbb{R}^m$. Then ξ is $\mathcal{F}_\tau$-measurable if and only if $\xi\chi_{(\tau\leq t)}$ is $\mathcal{F}_t$-measurable for all $t \geq 0$.*

Proof. It suffices to consider the case $m = 1$. If $\xi \in \mathcal{F}_\tau$, then there is a sequence of $\mathcal{F}_\tau$-measurable simple functions

$$\xi_j \overset{\triangle}{=} \sum_i \xi_j^i \chi_{A_j^i} \to \xi, \qquad \text{as } j \to \infty, \ P-\text{a.s.},$$

where $\xi_j^i \in \mathbb{R}$ and $A_j^i \in \mathcal{F}_\tau$. Obviously,

$$\xi_j \chi_{(\tau \le t)} = \sum_i \xi_j^i \chi_{A_j^i \cap (\tau \le t)}$$

is $\mathcal{F}_t$-measurable. Letting $j \to \infty$, we see that $\xi \chi_{(\tau \le t)}$ is also $\mathcal{F}_t$-measurable for all $t \ge 0$.

Conversely, if for all $t \ge 0$, $\xi \chi_{(\tau \le t)}$ is $\mathcal{F}_t$-measurable, then we have

$$\begin{aligned}
&(\xi \le a) \cap (\tau \le t) = (\xi \chi_{(\tau \le t)} \le a) \in \mathcal{F}_t, \qquad \forall\, a < 0,\\
&(\xi > a) \cap (\tau \le t) = (\xi \chi_{(\tau \le t)} > a) = (\xi \chi_{(\tau \le t)} \le a)^c \in \mathcal{F}_t, \qquad \forall\, a \ge 0.
\end{aligned}$$

Therefore, $(\xi \le a) \in \mathcal{F}_\tau$ for all $a \in \mathbb{R}$. This implies that ξ is $\mathcal{F}_\tau$-measurable. □

Proposition 1.33. *Let σ and τ be stopping times and X be an integrable random variable with values in $\mathbb{R}^m$. Then*

$$\begin{cases}
\chi_{(\sigma > \tau)} E(X \mid \mathcal{F}_\tau) = E(\chi_{(\sigma > \tau)} X \mid \mathcal{F}_\tau) = \chi_{(\sigma > \tau)} E(X \mid \mathcal{F}_{\sigma \wedge \tau}),\\
\chi_{(\sigma \ge \tau)} E(X \mid \mathcal{F}_\tau) = E(\chi_{(\sigma \ge \tau)} X \mid \mathcal{F}_\tau) = \chi_{(\sigma \ge \tau)} E(X \mid \mathcal{F}_{\sigma \wedge \tau}),\\
E(E(X \mid \mathcal{F}_\tau) \mid \mathcal{F}_\sigma) = E(X \mid \mathcal{F}_{\sigma \wedge \tau}).
\end{cases}$$

Proof. The first equalities in the first two assertions are obvious.

To prove the second equality in the first assertion, we note that

$$\chi_{(\sigma > \tau)} E(X \mid \mathcal{F}_\tau) \chi_{(\sigma \wedge \tau \le t)} = E(X \mid \mathcal{F}_\tau) \chi_{(\tau \le t)} \chi_{(\sigma > \tau,\, \sigma \wedge \tau \le t)}.$$

Recall that $E(X \mid \mathcal{F}_\tau)$ is $\mathcal{F}_\tau$-measurable. Thus, by Proposition 1.32, $E(X \mid \mathcal{F}_\tau)\chi_{(\tau \le t)}$ is $\mathcal{F}_t$-measurable. Recall also that $(\sigma > \tau) \in \mathcal{F}_{\sigma \wedge \tau}$. Thus $\chi_{(\sigma > \tau,\, \sigma \wedge \tau \le t)}$ is $\mathcal{F}_t$-measurable. Hence $\chi_{(\sigma > \tau)} E(X \mid \mathcal{F}_\tau) \chi_{(\sigma \wedge \tau \le t)}$ is $\mathcal{F}_t$-measurable for all $t \ge 0$. Hence, by Proposition 1.32 again, $\chi_{(\sigma > \tau)} E(X \mid \mathcal{F}_\tau)$ is $\mathcal{F}_{\sigma \wedge \tau}$-measurable. Then,

$$\begin{aligned}
&\chi_{(\sigma > \tau)} E(X \mid \mathcal{F}_\tau) = E(\chi_{(\sigma > \tau)} E(X \mid \mathcal{F}_\tau) \mid \mathcal{F}_{\sigma \wedge \tau})\\
&= \chi_{(\sigma > \tau)} E(E(X \mid \mathcal{F}_\tau) \mid \mathcal{F}_{\sigma \wedge \tau}) = \chi_{(\sigma > \tau)} E(X \mid \mathcal{F}_{\sigma \wedge \tau}),
\end{aligned}$$

which proves the first assertion. The second one can be proved similarly. Finally,

$$\begin{aligned}
&E(E(X \mid \mathcal{F}_\tau) \mid \mathcal{F}_\sigma)\\
&= E(\chi_{(\sigma > \tau)} E(X \mid \mathcal{F}_\tau) \mid \mathcal{F}_\sigma) + E(\chi_{(\tau \ge \sigma)} E(X \mid \mathcal{F}_\tau) \mid \mathcal{F}_\sigma)\\
&= E(\chi_{(\sigma > \tau)} E(X \mid \mathcal{F}_{\sigma \wedge \tau}) \mid \mathcal{F}_\sigma) + \chi_{(\tau \ge \sigma)} E(E(X \mid \mathcal{F}_\tau) \mid \mathcal{F}_{\tau \wedge \sigma})\\
&= \chi_{(\sigma > \tau)} E(X \mid \mathcal{F}_{\sigma \wedge \tau}) + \chi_{(\tau \ge \sigma)} E(X \mid \mathcal{F}_{\tau \wedge \sigma})\\
&= E(X \mid \mathcal{F}_{\tau \wedge \sigma}),
\end{aligned}$$

which gives the third assertion. □

Finally, we show the following interesting result.

Proposition 1.34. *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ be a filtered probability space and $X(t)$ be a $\{\mathcal{F}_t\}$-progressively measurable process with values in $\mathbb{R}^m$, and τ be a $\{\mathcal{F}_t\}$-stopping time. Then the random variable $X(\tau)$ is $\mathcal{F}_\tau$-measurable and the process $X(\tau \wedge t)$ is $\{\mathcal{F}_t\}$-progressively measurable.*

Proof. We first prove that $X(\tau \wedge t)$ is a $\{\mathcal{F}_t\}$-progressively measurable process. Recall the process $\tau \wedge t$ is $\{\mathcal{F}_t\}$-progressively measurable. Thus for each $t \geq 0$, the map $(s, \omega) \mapsto (\tau(\omega) \wedge s, \omega)$ is measurable from $([0, t] \times \Omega, \mathcal{B}[0, t] \times \mathcal{F}_t)$ into itself. On the other hand, by the progressively measurability of $X(t)$, the map $(s, \omega) \mapsto X(s, \omega)$ is measurable from $([0, t] \times \Omega, \mathcal{B}[0, t] \times \mathcal{F}_t)$ into $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$. Hence, the map $(s, \omega) \mapsto X(\tau(\omega) \wedge s, \omega)$ is measurable from $([0, t] \times \Omega, \mathcal{B}[0, t] \times \mathcal{F}_t)$ into $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, which yields the $\{\mathcal{F}_t\}$-progressive measurability of $X(\tau \wedge t)$. In particular, $X(\tau \wedge t)$ is $\{\mathcal{F}_t\}$-measurable for all $t \geq 0$.

Next, for any $B \in \mathcal{B}(\mathbb{R}^m)$,

$$(X(\tau) \in B) \cap (\tau \leq t) = (X(\tau \wedge t) \in B) \cap (\tau \leq t) \in \mathcal{F}_t, \quad \forall\, t \geq 0.$$

Therefore, $(X(\tau) \in B) \in \mathcal{F}_\tau$. Thus $X(\tau)$ is $\mathcal{F}_\tau$-measurable. □

1.4 Uniform integrability

We fix a probability space $(\Omega, \mathcal{F}, P)$.

Definition 1.35. *Suppose $\mathcal{K} \subset L^1_{\mathcal{F}}(\Omega; \mathbb{R}^m)$. We call $\mathcal{K}$ to be a uniformly integrable subset of $L^1_{\mathcal{F}}(\Omega; \mathbb{R}^m)$ if*

$$\int_{(|X|\geq s)} |X| dP$$

converges to 0 *uniformly for $X \in \mathcal{K}$ as $s \to +\infty$.*

Obviously, if $\mathcal{K}$ is a uniformly integrable subset of $L^1_{\mathcal{F}}(\Omega; \mathbb{R}^m)$, then it is also bounded in $L^1_{\mathcal{F}}(\Omega; \mathbb{R}^m)$.

Proposition 1.36. *Let $\xi \in L^1_{\mathcal{F}}(\Omega; \mathbb{R}^m)$, and $\{\mathcal{F}_\alpha\}_{\alpha \in \Lambda}$ be a family of sub-σ-fields. Then $\{E(\xi \mid \mathcal{F}_\alpha)\}_{\alpha \in \Lambda}$ is uniformly integrable.*

Proof. For any $s > 0$,

$$P(|E(\xi \mid \mathcal{F}_\alpha)| \geq s) \leq s^{-1} E|E(\xi \mid \mathcal{F}_\alpha)| \leq s^{-1} E|\xi|, \qquad \forall\, \alpha \in \Lambda.$$

Hence

$$\int_{(|E(\xi \mid \mathcal{F}_\alpha)|\geq s)} |E(\xi \mid \mathcal{F}_\alpha)|dP \leq \int_{(|E(\xi \mid \mathcal{F}_\alpha)|\geq s)} |\xi|dP$$
$$\leq \sqrt{s}P(|E(\xi \mid \mathcal{F}_\alpha)| \geq s) + \int_{(\xi|\geq\sqrt{s})} |\xi|dP \leq \frac{1}{\sqrt{s}}E|\xi| + + \int_{(\xi|\geq\sqrt{s})} |\xi|dP,$$

which yields the uniform integrability of $\{E(\xi \mid \mathcal{F}_\alpha)\}_{\alpha\in\Lambda}$. □

We have the following characterization of L^1-convergence in terms of the uniform integrability.

Theorem 1.37. *Let $X_0, X_1, X_2, \cdots$ be integrable random variables. Then $X_n \to X_0$ in $L^1_{\mathcal{F}}(\Omega;\mathbb{R}^m)$ as $n \to \infty$ if and only if $\{X_n\}$ converges to X_0 in probability and $\{X_n \mid n \in \mathbb{N}\}$ is uniformly integrable.*

For nonnegative random variable, we have the following simple result:

Theorem 1.38. *Let $X_0, X_1, X_2, \cdots$ be nonnegative integrable random variables. Then $X_n \to X_0$ in $L^1_{\mathcal{F}}(\Omega)$ as $n \to \infty$ if and only if $\{X_n\}$ converges to X_0 in probability and $\lim_{n\to\infty} EX_n = E\xi$.*

Proof. The "only if" part is obvious. We now show the "if" part.

By $0 \leq X_n \wedge \xi \leq \xi \in L^1_{\mathcal{F}}(\Omega)$, we get

$$\lim_{n\to\infty} E(X_n \wedge \xi) = E\xi.$$

On the other hand, by $\lim_{n\to\infty} EX_n = E\xi$, one deduces that

$$\lim_{n\to\infty} E(X_n + \xi) = 2E\xi.$$

Hence, from

$$X_n + \xi = X_n \wedge \xi + X_n \vee \xi, \qquad \forall\, n \in \mathbb{N},$$

we find

$$\lim_{n\to\infty} E(X_n \vee \xi) = E\xi.$$

Hence

$$\lim_{n\to\infty} E|X_n - \xi| = \lim_{n\to\infty} E(X_n \vee \xi) - \lim_{n\to\infty} E(X_n \wedge \xi) = 0.$$

This gives the L^1-convergence. □

1.5 Martingales

Let I be the time parameter set: $I = \{0, 1, 2, \cdots\}$ in the discrete time case or $I = [0, \infty)$ in the continuous time case. Let $\hat{I} = [0, \infty]$. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in I}, P)$ be a filtered probability space. Recall that for the continuous time case, $\{\mathcal{F}_t\}$ is assumed to be right-continuous.

Definition 1.39. *A real $\{\mathcal{F}_t\}$-adapted process $X = \{X(t)\}_{t\in I}$ is called a $\{\mathcal{F}_t\}$-martingale (resp. supermartingale, submartingale) if*

i) $X(t)$ is integrable for each $t \in I$;

ii) $E(X(t) \mid \mathcal{F}_s) = X(s)$ (resp. $\leq X(s)$, $\geq X(s)$) a.s. for every $t, s \in I$ with $s < t$.

Example 1.40. 1) Let X be an integrable random variable. Then $\{E(X \mid \mathcal{F}_t)\}_{t\in I}$ is a $\mathcal{F}_t$-martingale.

2) Let $X(t)$ be a martingale (*resp.* submartingale) and f be a convex function (*resp.* non-decreasing convex function). Then $f(X(t))$ is a submartingale. In particular, $X(t)^+$ and $X(t)\vee a$ ($\forall\, a \in \mathbb{R}$) are submartingales whenever $X(t)$ is a martingale.

Obviously, $X(t)$ is a martingale if and only if it is both sub- and suppermartingale. Martingales are a class of important stochastic processes, which are easily computable and estimable.

First, we consider the discrete time case. In this case, we write $X(n)$ as X_n, and all stopping times are assumed to be valued in $\{0, 1, 2, \cdots, \infty\}$.

The following theorem is the basis to show the martingale inequalities in the sequel.

Theorem 1.41. (Doob Stopping Theorem) *Let $\{X_n\}_{n\in I}$ be a $\{\mathcal{F}_n\}_{n\in I}$-martingale (resp. supermartingale, submartingale), σ and τ be two bounded stopping times with $\sigma \leq \tau$,* a.s. *Then*

$$E(X_\tau \mid \mathcal{F}_\sigma) = X_\sigma \ (resp. \leq,\ \geq) \text{ a.s.}$$

Proof. It suffices to consider the case of supermartingale.

Suppose $\sigma \vee \tau \leq M$, a.s. We need to show that for every $A \in \mathcal{F}_\sigma$, it holds

$$\int_A X_\sigma dP \geq \int_A X_\tau dP.$$

Suppose first that $\sigma \leq \tau \leq \sigma + 1$. Put

$$B_n = A \cap (\sigma = n) \cap (\tau > \sigma) = A \cap (\sigma = n) \cap (\tau > n) \in \mathcal{F}_n.$$

It is clear that

$$A \cap (\tau > \sigma) = \bigcup_{n=0}^{\infty} B_n.$$

Therefore

$$\begin{aligned}\int_A (X_\sigma - X_\tau)dP &= \int_{A\cap(\tau\geq\sigma)} (X_\sigma - X_\tau)dP = \int_{A\cap(\tau>\sigma)} (X_\sigma - X_\tau)dP \\ &= \sum_{n=0}^{\infty} \int_{B_n} (X_\sigma - X_\tau)dP = \sum_{n=0}^{\infty} \int_{B_n} (X_n - X_{n+1})dP \geq 0.\end{aligned}$$

In the general case, write

$$\gamma_k = \tau \wedge (\sigma + k), \qquad k = 0, 1, 2, \cdots, M.$$

Then γ_k are $\{\mathcal{F}_n\}$-stopping times, and

$$0 \leq \gamma_{j+1} - \gamma_j \leq 1, \qquad j = 0, 1, 2, \cdots, M-1.$$

Note that $\gamma_0 = \sigma$ and $\gamma_M = \tau$. Thus from the case discussed above,

$$\int_A X_\sigma dP = \int_A X_{\gamma_0} dP \geq \int_A X_{\gamma_1} dP \geq \cdots \geq \int_A X_{\gamma_M} dP = \int_A X_\tau dP.$$

□

Theorem 1.42. (Doob inequality) *Let $\{X_n\}_{n\in I}$ be a submartingale. Then for every $\lambda > 0$ and $m \in I$,*

$$\lambda P(\max_{0\leq n\leq m} X_n \geq \lambda) \leq \int_{(\max\limits_{0\leq n\leq m} X_n \geq \lambda)} X_m dP \leq E|X_m|,$$

and

$$\lambda P(\min_{0\leq n\leq m} X_n \leq -\lambda) \leq E(|X_0| + |X_m|).$$

Proof. We define a stopping time as follows

$$\sigma = \begin{cases} \min\{n \leq m \mid X_n \geq \lambda\}, \\ m, \qquad \text{if } \{n \leq m \mid X_n \geq \lambda\} = \emptyset. \end{cases}$$

Obviously, $\sigma \leq m$. Therefore

$$\begin{aligned} EX_m \geq EX_\sigma &= \int_{(\max\limits_{0\leq n\leq m} X_n \geq \lambda)} X_\sigma dP + \int_{(\max\limits_{0\leq n\leq m} X_n < \lambda)} X_\sigma dP \\ &\geq \lambda P(\max_{0\leq n\leq m} X_n \geq \lambda) + \int_{(\max\limits_{0\leq n\leq m} X_n < \lambda)} X_m dP, \end{aligned}$$

which yields the first inequality.

The second one is obtained from $EX_0 \leq EX_\tau$, where

$$\tau = \begin{cases} \min\{n \leq m \mid X_n \leq -\lambda\}, \\ m, \qquad \text{if } \{n \leq m \mid X_n \leq -\lambda\} = \emptyset. \end{cases}$$

□

Corollary 1.43. *Let $\{X_n\}_{n\in I}$ be a martingale such that $E|X_n|^p < \infty$ for some $p \geq 1$ and all $n \in I$. Then, for every $m \in I$ and $\lambda > 0$,*

$$P(\max_{0\leq n\leq m} |X_n| \geq \lambda) \leq \lambda^{-p} E|X_m|^p,$$

and if $p > 1$,

$$E(\max_{0\leq n\leq m} |X_n|^p) \leq \left(\frac{p}{p-1}\right)^p E|X_m|^p.$$

Proof. Obviously, $\{|X_n|^p\}_{n\in I}$ is a submartingale and so the first assertion follows from Theorem 1.42.

As for the second one, we set $Y = \max\limits_{0\le n\le m} |X_n|$. Then, by Theorem 1.42, we have

$$\lambda P(Y \ge \lambda) \le \int_{(Y\ge\lambda)} |X_m| dP.$$

Hence,

$$\begin{aligned} EY^p &= p\int_\Omega dP \int_0^Y \lambda^{p-1} d\lambda = p\int_\Omega dP \int_0^\infty \chi_{(Y\ge\lambda)}\lambda^{p-1}d\lambda \\ &= p\int_0^\infty \lambda^{p-1}P(Y\ge\lambda)d\lambda \le p\int_0^\infty \lambda^{p-2}\int_\Omega \chi_{(Y\ge\lambda)}|X_m|dPd\lambda \\ &= p\int_\Omega |X_m|\int_0^Y \lambda^{p-2}d\lambda dP = \frac{p}{p-1}\int_\Omega |X_m|Y^{p-1}dP. \end{aligned}$$

which yields the desired result. □

We now establish Doob's upcrossing inequality, which is used to derive convergence results for martingales. For a real $\{\mathcal{F}_n\}$-adapted process $X = \{X_n\}_{n\in I}$, and an interval $[a, b]$, with $-\infty < a < b < \infty$, we set

$$\left.\begin{array}{l} \tau_0 = 0, \\ \tau_1 = \min\{n \mid X_n \le a\}, \\ \tau_2 = \min\{n \ge \tau_1 \mid X_n \ge b\}, \\ \cdots \\ \tau_{2k+1} = \min\{n \ge \tau_{2k} \mid X_n \le a\}, \\ \tau_{2k+2} = \min\{n \ge \tau_{2k+1} \mid X_n \ge b\}, \\ \cdots. \end{array}\right\} \tag{1.19}$$

($\min\emptyset = +\infty$ unless otherwise stated). It is clear that $\{\tau_n\}$ is an increasing sequence of stopping times. Set

$$U_m^X(a,b)(\omega) \triangleq \max\{k \mid \tau_{2k}(\omega) \le m\}.$$

Obviously, $U_m^X(a, b)$ is the number of upcrossings of $\{X_n\}_{n=0}^m$ for the interval $[a, b]$.

Theorem 1.44. *Let $X = \{X_n\}_{n\in I}$ be a submartingale. Then for every $m \in I$ and $a < b$,*

$$E(U_m^X(a,b)) \le \frac{1}{b-a}\{E[(X_m - a)^+ - (X_0 - a)^+]\}.$$

Proof. Denote $Y_n = (X_n - a)^+$ and $Y = \{Y_n\}_{n\in I}$. Obviously, Y is a submartingale and $U_m^X(a, b) = U_m^Y(0, b - a)$. Let τ_i be defined as in (1.19) with X, a and b replaced by Y, 0 and $b - a$, respectively. Set

$$\tau'_n = \tau_n \wedge m.$$

Then, if $2j > m$,

$$Y_m - Y_0 = \sum_{n=1}^{2j}(Y_{\tau'_n} - Y_{\tau'_{n-1}}) = \sum_{n=1}^{j}(Y_{\tau'_{2n}} - Y_{\tau'_{2n-1}}) + \sum_{n=0}^{j-1}(Y_{\tau'_{2n+1}} - Y_{\tau'_{2n}}).$$

It is easy to see that the first term in the right-hand side is greater than or equal to $(b-a)U_m^Y(0, b-a)$. Also, $EY_{\tau'_{2n+1}} \geq EY_{\tau'_{2n}}$. These two facts yield the desired results. □

Theorem 1.45. *If $X = \{X_n\}_{n\in I}$ be a submartingale such that*

$$\sup_{n\in I} EX_n^+ < \infty,$$

then $X_\infty = \lim_{n\to\infty} X_n$ exists a.s., *and X_∞ is integrable. In order that $\hat{X} = \{X_n\}_{n\in\hat{I}}$ be a submartingale, i.e.,*

$$X_n \leq E(X_m \mid \mathcal{F}_n), \qquad \forall\, 0 \leq n < m \leq \infty,$$

it is necessary and sufficient that $\{X_n^+\}_{n\in I}$ be uniformly integrable.

Proof. Set $U_\infty^X(a,b) = \lim_{m\to\infty} U_m^X(a,b)$. Clearly,

$$(\varliminf_{n\to\infty} X_n < \varlimsup_{n\to\infty} X_n) \subset \bigcup_{r,r'\in Q, r<r'} (U_\infty^X(r,r') = \infty).$$

However, by Theorem 1.44 and noting $(X_m - r)^+ \leq X_m^+ + |r|$, we get

$$\begin{aligned} E(U_\infty^X(r,r')) &\leq \lim_{m\to\infty} E(U_m^X(r,r')) \leq \frac{1}{r'-r}\lim_{m\to\infty} E((X_m - r)^+ - (X_0 - r)^+) \\ &\leq \frac{1}{r'-r}\lim_{m\to\infty} E(X_m^+ + |r| - (X_0 - r)^+), \end{aligned}$$

and this limit is finite. Consequently,

$$P(\varliminf_{n\to\infty} X_n < \varlimsup_{n\to\infty} X_n) = 0,$$

which proves that $\lim_{n\to\infty} X_n$ exists a.s. The integrability of X_∞ follows from Fatou's lemma.

Now, if $X_n \leq E(X_\infty \mid \mathcal{F}_n)$ for $n = 0, 1, 2, \cdots$, then by the property of conditional expectation, we have

$$X_n^+ \leq E(X_\infty^+ \mid \mathcal{F}_n), \quad \text{a.s.}$$

However, we know that $\{E(X_\infty^+ \mid \mathcal{F}_n)\}_{n\in I}$ is uniformly integrable. Therefore, so is $\{X_n^+\}_{n\in I}$.

Conversely, if $\{X_n^+\}_{n\in I}$ is uniformly integrable, then, by

$$|X_n \vee (-a)| \le a + X_n^+, \qquad \forall a > 0,$$

one concludes that $\{X_n \vee (-a)\}_{n\in I}$ is uniformly integrable. Thus, by Theorem 1.37, we see that

$$\lim_{n\to\infty} X_n \vee (-a) = X_\infty \vee (-a) \text{ in } L^1_{\mathcal{F}}(\Omega).$$

Since $\{X_n \vee (-a)\}_{n\in I}$ is a submartingale,

$$E(X_\infty \vee (-a) \mid \mathcal{F}_n) = \lim_{m\to\infty} E(X_m \vee (-a) \mid \mathcal{F}_n) \ge X_n \vee (-a),$$

and so, letting $a \to +\infty$, one gets $E(X_\infty \mid \mathcal{F}_n) \ge X_n$. □

Theorem 1.46. *Let $X = \{X_n\}_{n\in I}$ be a martingale. Then the following conditions are equivalent.*

i) $\lim_{n\to\infty} X_n = X_\infty$ *exists* a.s. *and in* $L^1_{\mathcal{F}}(\Omega)$*;*

ii) $X_n = E(X_\infty \mid \mathcal{F}_n)$ *for all* $n \in I$*;*

iii) $\{X_n\}_{n\in I}$ *is uniformly integrable.*

Furthermore, under one of these conditions, $\hat{X} = \{X_n\}_{n\in \hat{I}}$ *is a martingale.*

Proof. "i)⇒ii)". For any m,

$$E(X_n \mid \mathcal{F}_m) = X_m, \qquad \forall\, n \ge m.$$

Letting $n \to \infty$, one gets

$$X_m = E(X_\infty \mid \mathcal{F}_m), \text{ a.s.}$$

"ii)⇒iii)". Obviously.

"iii)⇒i)". The uniform integrability of $\{X_n\}_{n\in I}$ implies the boundedness of $\{X_n\}_{n\in I}$ in $L^1_{\mathcal{F}}(\Omega)$ and the uniform integrability of $\{X_n^+\}_{n\in I}$. Now, the desired result follows from Theorem 1.45. □

We consider, for a moment, martingale with "reversed" time. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{\mathcal{F}_n\}_{n=-\infty}^0$ be a family of sub-σ-fields of $\mathcal{F}$ such that $\mathcal{F}_0 \supset \mathcal{F}_{-1} \supset \mathcal{F}_{-2} \supset \cdots \supset$. We say $X = \{X_n\}_{n=-\infty}^0$ to be a martingale (*resp.* supermartingale, submartingale) if X_n is $\mathcal{F}_n$-adapted integrable random variable such that

$$E(X_n \mid \mathcal{F}_m) = X_m \quad (\textit{resp.} \ \le,\ \ge)$$

for every $n, m \in \{0, -1, -2, \cdots\}$ with $n > m$.

Theorem 1.47. *Let $X = \{X_n\}_{n=-\infty}^0$ be a submartingale. Then*

i) $X_{-\infty} = \lim_{n\to-\infty} X_n$ *exists* a.s.*;*

ii) X is uniformly integrable and $\lim_{n\to-\infty} X_n = X_{-\infty}$ *in* $L^1_{\mathcal{F}}(\Omega)$ *provided that*

$$\lim_{n\to-\infty} EX_n > -\infty.$$

Proof. The first assertion can be proved similar to that of Theorem 1.45.

For the second one, it suffices to show the uniform integrability of X. For this, we fix any $\varepsilon > 0$ and choose k such that

$$|EX_{k_1} - EX_{k_2}| < \varepsilon, \qquad \forall\, k_1, k_2 \leq k.$$

Then, if $n \leq k$ and $\lambda > 0$,

$$\int_{(|X_n|>\lambda)} |X_n| dP = \int_{(X_n>\lambda)} X_n dP + \int_{(X_n \geq -\lambda)} X_n dP - EX_n$$
$$\leq \int_{(X_n>\lambda)} X_k dP + \int_{(X_n \geq -\lambda)} X_k dP - EX_k + \varepsilon \leq 2 \int_{(|X_n|>\lambda)} |X_k| dP + \varepsilon.$$

Also,

$$P(|X_n| > \lambda) \leq \frac{1}{\lambda} E|X_n| = \frac{1}{\lambda}(2EX_n^+ - EX_n) \leq \frac{1}{\lambda}(2EX_0^+ - \lim_{n\to-\infty} EX_n).$$

Therefore, we get the uniform integrability of $\{X_n\}_{n=-\infty}^0$. □

Now, we consider the continuous time case, i.e., $I = [0, \infty)$.

Theorem 1.48. *Let $X = \{X(t)\}_{t\in I}$ be a submartingale. Then, with probability 1, $Q \cap I \ni t \mapsto X(t)$ is finite valued and possesses*

$$\lim_{Q\cap I \ni s \to t+} X(s) \quad \textit{and} \quad \lim_{Q\cap I\ni s\to t-} X(s), \qquad \forall\, t \geq 0.$$

Proof. Let $T > 0$ be given and $\{r_1, r_2, \cdots\}$ be an enumeration of the set $Q \cap [0, T]$. For every n, if $\{s_1, s_2, \cdots, s_n\}$ is the rearrangement of the set $\{r_1, r_2, \cdots, r_n\}$ according to the natural order, then

$$Y_0 = X(0),\ Y_1 = X(s_1), \cdots, Y_n = X(s_n),\ Y_{n+1} = X(T)$$

defines a submartingale. Therefore, by Theorems 1.42 and 1.44, we get

$$P(\max_{0\leq i\leq n+1} |Y_i| > \lambda) \leq \frac{1}{\lambda}[E|X(0)| + E|X(T)|],$$

and for any $a, b \in I$ with $a < b$,

$$E(U_n^Y(a,b)) \leq \frac{1}{b-a} E(Y_n - a)^+ \leq \frac{1}{b-a} E(X(T) - a)^+.$$

Since this holds for every n, we have

$$P(\sup_{t\in Q\cap[0,T]} |X(t)| > \lambda) \leq \frac{2}{\lambda}[E|X(0)| + E|X(T)|],$$

and

$$E(U_\infty^{X|_{Q\cap[0,T]}}(a,b)) \leq \frac{1}{b-a} E(X(T) - a)^+.$$

By letting λ and $a < b$ run over respectively positive integers and pairs of rational, the assertion of the theorem follows. □

Theorem 1.49. *Let* $X = \{X(t)\}_{t\in I}$ *be a submartingale. Then the following assertions hold:*

i) $\widetilde{X}(t) \stackrel{\triangle}{=} \lim_{Q\cap I\ni r\to t+} X(r)$ *exists* a.s. *and* $\widetilde{X} = \{\widetilde{X}(t)\}_{t\in I}$ *is a submartingale such that* $t \mapsto \widetilde{X}(t)$ *is right-continuous with left-hand limits* a.s.*;*

ii) $X(t) \le \widetilde{X}(t)$ a.s., $\quad \forall\, t \in I$*;*

iii) $P(X(t) = \widetilde{X}(t)) = 1$ *for every* $t \in I$ *if and only if* $EX(t)$ *is right-continuous with respect to* t.

Proof. i) The well-definedness of $\widetilde{X}(t)$ is shown in Theorem 1.48. It is easy to see that $\widetilde{X}(t)$ is $\{\mathcal{F}_t\}$-adapted.

To prove $\{\widetilde{X}(t)\}_{t\in I}$ is a submartingale, fix any $s > t$. We choose arbitrarily a sequence $\{\varepsilon_n\}$ decreasing to 0 so that $t+\varepsilon_n \in Q$ for all n. Then, by Theorem 1.47, we see that

$$\lim_{n\to\infty} |X(t+\varepsilon_n) - \widetilde{X}(t)|_{L^1_{\mathcal{F}}(\Omega)} = 0.$$

Similarly, we choose a sequence $\{\varepsilon'_n\}$ decreasing to 0 such that $s + \varepsilon'_n \in Q$ for all n. Then

$$\lim_{n\to\infty} |X(s+\varepsilon'_n) - \widetilde{X}(s)|_{L^1_{\mathcal{F}}(\Omega)} = 0.$$

Hence, for any $B \in \mathcal{F}_t$,

$$\int_B \widetilde{X}(t)dP = \lim_{n\to\infty} \int_B X(t+\varepsilon_n)dP \le \lim_{n\to\infty} \int_B X(s+\varepsilon'_n)dP = \int_B \widetilde{X}(s)dP.$$

This implies that $\{\widetilde{X}(t)\}_{t\in I}$ is a submartingale.

Now, using Theorem 1.48 again, we see that, with probability 1, for any $t_0 \in I$ and $\varepsilon > 0$, there is a $\delta > 0$ such that for any $s \in (t_0, t_0+\delta) \cap Q$,

$$|\widetilde{X}(t_0) - X(s)| < \varepsilon.$$

Therefore, for any $r \in (t_0, t_0+\delta)$,

$$|\widetilde{X}(t_0) - \widetilde{X}(r)| = \lim_{Q\ni s\to r+} |\widetilde{X}(t_0) - X(s)| \le \varepsilon.$$

Hence,

$$\lim_{r\to t_0+} \widetilde{X}(r) = \widetilde{X}(t_0).$$

This yields the right-continuity of $\widetilde{X}(t)$.

To show the existence of $\lim_{t\to t_0-} \widetilde{X}(t)$, we use Theorem 1.48 again to conclude that

$$\lim_{Q\ni t\to t_0-} \widetilde{X}(t) \text{ exists a.s.}$$

Thus, for any $\varepsilon > 0$, there is a $\delta > 0$ such that

$$|\widetilde{X}(t_1) - \widetilde{X}(t_2)| < \varepsilon, \qquad \forall\, t_1, t_2 \in (t_0-\delta, t_0) \cap Q.$$

By the right-continuity of $\widetilde{X}(t)$, the above inequality can be strengthened as

$$|\widetilde{X}(t_1) - \widetilde{X}(t_2)| < \varepsilon, \qquad \forall\, t_1, t_2 \in (t_0 - \delta, t_0),$$

which implies the existence of the desired left limit.

ii) It is easy to see that

$$\int_B X(t)dP \le \int_B \widetilde{X}(t)dP, \qquad \forall\, B \in \mathcal{F}_t$$

and hence $X(t) \le \widetilde{X}(t)$ a.s.

iii) The "if" part. By the right-continuity of $EX(t)$, we get

$$EX(t) = \lim_{Q \ni s \to t+} EX(s) = E\widetilde{X}(t).$$

Note that, $\widetilde{X}(t) \ge X(t)$. Hence $P(\widetilde{X}(t) = X(t)) = 1$.

The "only if" part. Choose any sequence $\{s_n\}$ decreasing to t. Then, by the proof of Theorem 1.48, we see that, with probability 1.

$$\lim_{s_n \to t+} X(s_n) = \lim_{Q \ni s \to t+} X(s) = \widetilde{X}(t), \qquad \forall\, t \ge 0.$$

Thus,

$$\lim_{s_n \to t+} EX(s_n) = E\widetilde{X}(t) = EX(t).$$

□

$\widetilde{X} = \{\widetilde{X}(t)\}_{t\in I}$ in Theorem 1.49 is called the right-continuous modification of X. It is easy to see that when $X(\cdot)$ is a martingale, we may assume that $X(\cdot)$ itself is right-continuous.

Theorem 1.50. *If $X = \{X(t)\}_{t\in I}$ is a right-continuous martingale so that $E|X(t)|^p < \infty$ for some $p \ge 1$, then for each $T > 0$,*

$$P(\sup_{t\in[0,T]} |X(t)| > \lambda) \le \lambda^{-p} E|X(T)|^p, \qquad p \ge 1,$$
$$E(\sup_{t\in[0,T]} |X(t)|^p) \le \Big(\frac{p}{p-1}\Big)^p E|X(T)|^p, \qquad p > 1.$$

Theorem 1.51. *Let $X = \{X(t)\}_{t\in I}$ be a right-continuous submartingale so that*

$$\sup_{t\in I} EX(t)^+ < \infty.$$

Then:

i) $\lim_{t\to\infty} X(t) = X_\infty$ a.s., $X_\infty \in L^1_{\mathcal{F}}(\Omega)$;

ii) If $\{X(t)^+\}_{t\in I}$ is uniformly integrable, then $\hat{X} = \{X(t)\}_{t\in \hat{I}}$ is a submartingale;

iii) If $\{X(t)\}_{t\in I}$ is uniformly integrable, then

$$\lim_{t\to\infty} X(t) = X_\infty \quad \text{in } L^1_{\mathcal{F}}(\Omega).$$

Theorem 1.52. *Let $X = \{X(t)\}_{t\in I}$ be a martingale. Then the following conditions are equivalent:*

i) $\lim_{t\to\infty} X(t) = X_\infty$ *in* $L^1_{\mathcal{F}}(\Omega)$*;*

ii) $X(t) = E(X_\infty \mid \mathcal{F}_t)$ a.s., $\forall\, t \in I$*;*

iii) $\{X(t)\}_{t\in I}$ *is uniformly integrable.*

Furthermore, under one of these conditions, $\hat{X} = \{X(t)\}_{t\in\hat{I}}$ *is a martingale.*

Theorem 1.53. (Doob stopping theorem) *Let $X = \{X(t)\}_{t\in I}$ be a right-continuous $\{\mathcal{F}_t\}_{t\in I}$-submartingale, σ and τ be two bounded $\{\mathcal{F}_t\}_{t\in I}$-stopping times such that*

$$P(\sigma \le \tau) = 1.$$

Then $X(\sigma)$ and $X(\tau)$ are integrable, and

$$E(X(\tau) \mid \mathcal{F}_\sigma) \ge X_\sigma, \text{ a.s.}$$

Proof. Since σ and τ are bounded, one can find a positive integer N such that $\sigma \vee \tau \le N$. Put

$$\sigma_n = \sum_{k=1}^{2^n N} \frac{k}{2^n} \chi_{(\frac{k-1}{2^n} \le \sigma < \frac{k}{2^n})}, \quad \tau_n = \sum_{k=1}^{2^n N} \frac{k}{2^n} \chi_{(\frac{k-1}{2^n} \le \tau < \frac{k}{2^n})}.$$

Then

$$\sigma_n \le \tau_n \le N, \quad \text{a.s.}$$

Therefore, by Theorem 1.41, we see that

$$E(X(\tau_n) \mid \mathcal{F}_{\sigma_n}) \ge X(\sigma_n), \text{ a.s.}$$

Hence, for any $A \in \mathcal{F}_\sigma$, we have

$$\int_A X(\tau_n) dP \ge \int_A X(\sigma_n) dP, \qquad \forall\, n.$$

Hence, by Theorem 1.47, we get

$$\lim_{n\to\infty} X(\tau_n) = X(\tau), \quad \lim_{n\to\infty} X(\sigma_n) = X(\sigma), \quad \text{in } L^1_{\mathcal{F}}(\Omega).$$

Hence,

$$\int_A X(\tau) dP \ge \int_A X(\sigma) dP,$$

which gives the desired result. □

Corollary 1.54. *Let $X = \{X(t)\}_{t\in I}$ be a right-continuous $\{\mathcal{F}_t\}_{t\in I}$-submartingale and $\{\sigma_t\}_{t\in I}$ be a family of bounded stopping times such that $P(\sigma_t \le \sigma_s) = 1$ when $t < s$. Set*

$$\overline{X}(t) = X(\sigma_t), \quad \overline{\mathcal{F}}_t = \mathcal{F}_{\sigma_t}, \quad \forall\, t \in I.$$

Then $\overline{X} = \{\overline{X}(t)\}_{t\in I}$ be a right-continuous $\{\overline{\mathcal{F}}_t\}_{t\in I}$-submartingale.

Corollary 1.55. *Let* $X = \{X(t)\}_{t\in I}$ *be a* $\{\mathcal{F}_t\}_{t\in I}$*-martingale, and* $\sigma \le \tau$ *be two stopping times. Then*

$$E(X(t \wedge \tau) - X(t \wedge \sigma) \mid \mathcal{F}_\sigma) = 0, \quad \forall\, t \ge 0.$$

Proof. By Proposition 1.34, $X(t\wedge\tau)$ is $\mathcal{F}_t$-measurable. Hence, by Corollary 1.54,

$$X(t\wedge\sigma) = E(X(t\wedge\tau) \mid \mathcal{F}_{t\wedge\sigma}) = E(E(X(t\wedge\tau) \mid \mathcal{F}_t) \mid \mathcal{F}_\sigma) = E(X(t\wedge\tau) \mid \mathcal{F}_\sigma).$$

□

2

Stochastic Integrals

2.1 Itô's integrals

Let $W = \{W(t)\}_{t \geq 0}$ be a $(\mathcal{F}_t)$-Brownian motion. Since with probability 1, the function $t \to W(t)$ is nowhere differentiable, the integral $\int_0^t f(s,\omega)dW(s,\omega)$ can not be defined piontwisely. However, one can define the integral for a large class of processes by means of the martingale property of Brownian motion. This was first done by K. Itô and is now known as Itô's integral.

We first introduce the function space consisting of all possible integrands. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$ be a fixed filtered probability space satisfying the usual condition. Let $T > 0$ and recall that $L^2_{\mathcal{F}}(0,T)$ is the set of all measurable processes $f(t)$ adapted to $\{\mathcal{F}_t\}_{t \geq 0}$, such that

$$|f|^2_{L^2_{\mathcal{F}}(0,T)} \overset{\Delta}{=} E \int_0^T |f(t)|^2 dt < \infty.$$

It is seen that $L^2_{\mathcal{F}}(0,T)$ is a Hilbert space.

As one done for the Legesgue integral, we will first define the Itô's integral for simple integrands. This leads to the definition of the following space.

Let $\mathcal{L}_0$ be the sub-collection of those processes $f = f(t) \in L^2_{\mathcal{F}}(0,T)$ of the form (called simple processes):

$$f(t,\omega) = f_0(\omega)\chi_{\{0\}}(t) + \sum_{j=0}^{n} f_j(\omega)\chi_{(t_j,t_{j+1}]}(t), \quad (t,\omega) \in [0,T] \times \Omega, \tag{2.1}$$

where $0 = t_0 < t_1 < \cdots < t_n = T$, f_j is $\mathcal{F}_{t_j}$-measurable with $\sup_{j,\,\omega} |f_j(\omega)| < \infty$.

We need the following basic result. We will omit its proof since it is quite technical.

Lemma 2.1. $\mathcal{L}_0$ *is dense in* $L^2_{\mathcal{F}}(0,T)$.

Next, we introduce the following sets that are related to the integrals we are going to define:

$$\begin{cases} \mathcal{M}^2[0,T] = \{X \in L^2_{\mathcal{F}}(0,T) \bigm| X \text{ is a right-continuous} \\ \qquad\qquad\qquad \{\mathcal{F}_t\}_{t\geq 0}\text{-martingale with } X(0)=0,\ P\text{-a.s.}\}, \\ \mathcal{M}^2_c[0,T] = \{X \in \mathcal{M}^2[0,T] \bigm| t \mapsto X(t) \text{ is continuous, } P\text{-a.s.}\}. \end{cases}$$

We identify $X, Y \in \mathcal{M}^2[0,T]$ if there exists a set $N \in \mathcal{F}$ with $P(N) = 0$, such that $X(t,\omega) = Y(t,\omega)$, for all $t \geq 0$ and $\omega \notin N$. Define

$$|X|_{\mathcal{M}^2[0,T]} = \sqrt{E|X(T)|^2}, \qquad \forall X \in \mathcal{M}^2[0,T].$$

Lemma 2.2. $(\mathcal{M}^2[0,T], |\cdot|_{\mathcal{M}^2[0,T]})$ *is a Hilbert space, and* $\mathcal{M}^2_c[0,T]$ *is a closed subspace of* $\mathcal{M}^2[0,T]$.

Proof. First we note that if $|X-Y|_{\mathcal{M}^2[0,T]} = 0$, then $X = Y$. Indeed, $|X-Y|_{\mathcal{M}^2[0,T]} = 0$ implies $X(T) = Y(T)$ a.s. Therefore, $X(t) = E(X(T)|\mathcal{F}_t) = E(Y(T)|\mathcal{F}_t) = Y(T)$ for $t \leq T$. By the right-continuity of $X(t)$ and $Y(t)$, we conclude that $X = Y$.

Next, let $\{X^n\}_{n=1}^\infty$ be a Cauchy sequence in $\mathcal{M}^2[0,T]$, i.e., $\lim\limits_{n,m\to\infty} |X^n - X^m|_{\mathcal{M}^2[0,T]} = 0$. Then, by the second assertion in Theorem 1.50, we conclude that

$$E\left(\sup_{0\leq t\leq T} |X^n(t) - X^m(t)|\right) \leq 4E|X^n(T) - X^m(T)|^2.$$

Hence, there is a $X = \{X(t)\}$ such that $E|X^n(t) - X(t)| \to 0$ as $n \to \infty$ uniformly for $t \in [0,T]$, and we see from this that $X \in \mathcal{M}^2[0,T]$ and $|X^n(t) - X(t)|_{\mathcal{M}^2[0,T]} \to 0$ as $n \to \infty$.

Finally, it is also clear from this proof that if $X^n \in \mathcal{M}^2_c[0,T]$, then $X \in \mathcal{M}^2_c[0,T]$. □

We shall now define the Itô's integral as a mapping

$$f \in L^2_{\mathcal{F}}(0,T) \mapsto I(f) \in \mathcal{M}^2_c[0,T]. \tag{2.2}$$

First, assume $f \in \mathcal{L}_0$ and taking the form of (2.1). Then we set

$$I(f)(t,\omega) = \sum_{j=0}^{n} f_j(\omega)[W(t\wedge t_{j+1},\omega) - W(t\wedge t_j,\omega)]. \tag{2.3}$$

We have the following key result.

Proposition 2.3. *Let* $f \in \mathcal{L}_0$. *Then*

(i) $I(f) \in \mathcal{M}^2_c[0,T]$.

(ii) $|I(f)|_{\mathcal{M}^2[0,T]} = |f|_{L^2_{\mathcal{F}}(0,T)}$.

(iii) $\big(I(f)(t)\big)^2 - \int_0^t f^2(s)ds$ *is a* $(\mathcal{F}_t)$*-martingale.*

Next, let $f \in L^2_{\mathcal{F}}(0,T)$. Then, by Lemma 2.1, we can find a sequence of $\{f_k\} \subset \mathcal{L}_0$ such that $|f_k - f|_{L^2_{\mathcal{F}}(0,T)} \to$ as $k \to \infty$. Since $|I(f_k) - I(f_j)|_{\mathcal{M}^2[0,T]} = |f_k - f_j|_{L^2_{\mathcal{F}}(0,T)}$, one deduces that $\{I(f_k)\}$ is a Cauchy sequence in $\mathcal{M}^2[0,T]$ and therefore, by Lemma 2.2, it converges to a unique element $X \in \mathcal{M}^2[0,T]$. Clearly, X is determined uniquely from f and is independent of the particular choice of $\{f_k\}$. This process is called the Itô's stochastic integral of $f \in L^2_{\mathcal{F}}(0,T)$ with respect to the Brownian Motion $W(\cdot)$. We shall denote it by

$$\int_0^t f(s,\omega)dW(s,\omega) \quad \text{or simply} \quad \int_0^t f(s)dW(s) \quad \text{or even} \quad \int_0^t fdW.$$

Further, for any $f \in L^2_{\mathcal{F}}(0,T)$ and any two stopping times σ and τ with $0 \le \sigma \le \tau \le T$, P-a.s., we define

$$\int_\sigma^\tau f(s)dW(s) \triangleq \int_0^\tau f(s)dW(s) - \int_0^\sigma f(s)dW(s).$$

The stochastic integral with respect to a $\{\mathcal{F}_t\}$-Brownian Motion $W(\cdot)$ has the properties:

Theorem 2.4. *Let $f, g \in L^2_{\mathcal{F}}(0,T)$, $a, b \in \mathbb{R}$, $T \ge t > s \ge 0$, and σ and τ be $\{\mathcal{F}_t\}$-stopping times such that $\tau \ge \sigma$ a.s.. Then*

(i)

$$\int_0^t (af + bg)dW = a\int_0^t fdW + b\int_0^t gdW, \text{a.s.} \tag{2.4}$$

(ii)

$$E\Big(\int_s^t fdW \mid \mathcal{F}_s\Big) = 0, \text{a.s.}, \tag{2.5}$$

and

$$E\Big(\Big|\int_s^t fdW\Big|^2 \mid \mathcal{F}_s\Big) = E\Big(\int_s^t |f(u,\omega)|^2 du \mid \mathcal{F}_s\Big), \ \text{a.s.} \tag{2.6}$$

More generally,

$$E\Big(\int_{t\wedge\sigma}^{t\wedge\tau} fdW \mid \mathcal{F}_s\Big) = 0, \text{a.s.}, \tag{2.7}$$

and

$$E\Big(\Big|\int_{t\wedge\sigma}^{t\wedge\tau} fdW\Big|^2 \mid \mathcal{F}_\sigma\Big) = E\Big(\int_{t\wedge\sigma}^{t\wedge\tau} |f(u,\omega)|^2 du \mid \mathcal{F}_\sigma\Big), \quad \text{a.s.} \tag{2.8}$$

(iii)

$$E\Big(\int_s^t fdW \int_s^t gdW \mid \mathcal{F}_s\Big) = E\Big(\int_s^t f(u,\omega)g(u,\omega)du \mid \mathcal{F}_s\Big), \quad \text{a.s.} \tag{2.9}$$

and

$$E\Big(\int_{t\wedge\sigma}^{t\wedge\tau} fdW \int_{t\wedge\sigma}^{t\wedge\tau} gdW \mid \mathcal{F}_\sigma\Big) = E\Big(\int_{t\wedge\sigma}^{t\wedge\tau} f(u,\omega)g(u,\omega)du \mid \mathcal{F}_\sigma\Big), \quad \text{a.s.} \tag{2.10}$$

(iv)

$$\int_0^{t\wedge\sigma} fdW = \int_0^t f(s)\chi_{[0,\sigma]}(s)dW(s), \quad \text{a.s.} \tag{2.11}$$

Proof. We only prove (iv). First consider the case when $f \in \mathcal{L}_0$. Assume

$$f \equiv f(t,\omega) = f_0(\omega)\chi_{\{0\}}(t) + \sum_{j=0}^{n} f_j(\omega)\chi_{(t_j,t_{j+1}]}(t).$$

Let $\{s_i^n\}_{i=0}^{k(n)}$ $(n = 1, 2, \cdots)$ be a refinement of $\{t_i\}_{i=0}^n \cup \{iT2^{-n}\}_{i=0}^{2^n}$ according to the natural order. Then f may be re-written as

$$f = f_0(\omega)\chi_{\{0\}}(t) + \sum_{j=0}^{k(n)} f_j^n(\omega)\chi_{(s_j^n,s_{j+1}^n]}(t), \qquad n = 1, 2, \cdots.$$

Define

$$\sigma^n(\omega) = \sum_{j=0}^{k(n)} s_{j+1}^n \chi_{(s_j^n,s_{j+1}^n]}(\sigma).$$

Then, for each $t \geq 0$,

$$\{\sigma^n \leq t\} = \bigcup_{s_{j+1}^n \leq t,\, 0\leq j\leq k(n)} (\sigma \leq s_{j+1}^n) \in \mathcal{F}_t.$$

Hence, σ^n is a $\{\mathcal{F}_t\}$-stopping time, and $\sigma^n \downarrow \sigma$ as $n \to \infty$. It is easy to show that

$$f\chi_{[0,\sigma^n]}(t) = f_0(\omega)\chi_{\{0\}}(t) + \sum_{j=0}^{k(n)} \Big(f_j^n(\omega)\chi_{(s_j^n,\infty)}(\sigma)\Big)\chi_{(s_j^n,s_{j+1}^n]}(t) \in \mathcal{L}_0.$$

Hence,

$$\begin{aligned}\int_0^t f\chi_{[0,\sigma^n]}(u)dW(u) &= \sum_{j=0}^{k(n)} \Big(f_j^n(\omega)\chi_{(s_j^n,\infty)}(\sigma)\Big)(W(t\wedge s_{j+1}^n) - W(t\wedge s_j^n)) \\ &= \sum_{j=0}^{k(n)} \Big(f_j^n(\omega)\chi_{(\sigma>s_j^n)}\Big)(W(t\wedge\sigma^n\wedge s_{j+1}^n) - W(t\wedge\sigma^n\wedge s_j^n)) \\ &= \sum_{j=0}^{k(n)} f_j^n(\omega)(W(t\wedge\sigma^n\wedge s_{j+1}^n) - W(t\wedge\sigma^n\wedge s_j^n)) \\ &= \int_0^{t\wedge\sigma^n} fdW.\end{aligned}$$

Thus,

$$\lim_{n\to\infty}\int_0^t f\chi_{[0,\sigma^n]}(u)dW(u)=\int_0^{t\wedge\sigma} fdW.$$

On the other hand, it is clear that, for every $t>0$,

$$E\left|\int_0^t f\chi_{[0,\sigma^n]}(u)dW(u)-\int_0^t f\chi_{[0,\sigma]}(u)dW(u)\right|^2=E\int_0^t |f|^2\chi_{[\sigma,\sigma^n]}(u)du\to 0,$$

as $n\to\infty$. Hence,

$$\int_0^t f\chi_{[0,\sigma^n]}(u)dW(u)\to\int_0^t f\chi_{[0,\sigma]}(u)dW(u)\quad\text{in }\mathcal{M}_c^1[0,T].$$

Consequently,

$$\int_0^t f\chi_{[0,\sigma]}(u)dW(u)=\int_0^{t\wedge\sigma} fdW.$$

The general case can be proved by approximating f by a sequence of $\{f_n\}\subset\mathcal{L}_0$. □

Now, let $W(t)=(W^1(t),W^2(t),\cdots,W^n(t))$ be an n-dimensional $\{\mathcal{F}_t\}$-Brownian Motion and let $f_1,f_2,\cdots,f_n\in L^2_{\mathcal{F}}(0,T)$. Then one may define the stochastic integral $\int_0^t f_k dW^k$ for each $k=1,2,\cdots,n$.

Proposition 2.5. *For $t>s\geq 0$ and $k,j=1,2,\cdots,n$,*

$$E\Big(\int_s^t f_k dW^k\int_s^t f_j dW^j\ \Big|\ \mathcal{F}_s\Big)=\delta_j^k E\Big(\int_s^t f_k f_j du\ \Big|\ \mathcal{F}_s\Big).$$

More generally, for any $m\times n$-process $f\in L^2_{\mathcal{F}}(0,T;\mathbb{R}^{n\times m})$ and n-dimensional $\{\mathcal{F}_t\}$-Brownian Motion $W(t)$, one may define the Itô's integral $\int_0^T fdW$.

We now extend the Itô integral to a bigger class of integrands than $L^2_{\mathcal{F}}(0,T)$. To this end, for each $p\geq 1$, we introduce

$$\begin{aligned}L^{p,loc}_{\mathcal{F}}(0,T)=\{X:[0,T]\times\Omega\to\mathbb{R}\ \big|\ &X(\cdot)\text{ is }\{\mathcal{F}_t\}_{t\geq 0}\text{-adapted}\\ &\text{and }\int_0^T|X(t)|^p dt<\infty,\quad P\text{-a.s.}\},\end{aligned}$$

and

$$\begin{cases}\mathcal{M}^{2,loc}[0,T]=\{X:[0,T]\times\Omega\to\mathbb{R}\ \big|\ \exists\text{ nondecreasing}\\ \qquad\qquad\text{stopping times }\sigma_j\text{ with }P(\lim_{j\to\infty}\sigma_j\geq T)=1,\\ \qquad\qquad\text{and }X(\cdot\wedge\sigma_j)\in\mathcal{M}^2[0,T],\quad\forall j=1,2,\cdots\},\\ \mathcal{M}^{2,loc}_c[0,T]=\{X\in\mathcal{M}^{2,loc}[0,T]\ \big|\ t\mapsto X(t)\text{ is continuous}, P\text{-a.s.}\}.\end{cases}$$

Any element in $\mathcal{M}^{2,loc}[0,T]$ (*resp.* $\mathcal{M}_c^{2,loc}[0,T]$) is called a local square integrable (*resp.* continuous local square integrable) $\{\mathcal{F}_t\}$-martingale.

For any $f(\cdot) \in L^{2,loc}_{\mathcal{F}}(0,T)$, define

$$\sigma_j(\omega) \stackrel{\Delta}{=} \inf \left\{ t \in [0,T] \;\Big|\; \int_0^t |f(s)|^2 ds \geq j \right\}, \quad j = 1, 2, \cdots.$$

In the above, we define $\inf \emptyset \stackrel{\Delta}{=} T$. Clearly, $\{\sigma_j\}_{j\geq 1}$ is a sequence of nondecreasing stopping times satisfying $P(\lim\limits_{j\to\infty} \sigma_j \geq T) = 1$. Set $f_j(t) \stackrel{\Delta}{=} f(t)\chi_{[0,\sigma_j]}(t)$. Then

$$\int_0^T |f_j(s)|^2 ds = \int_0^{\sigma_j} |f(s)|^2 ds \leq j,$$

which implies $f_j(\cdot) \in L^2_{\mathcal{F}}(0,T)$. By (2.11), we have

$$\begin{aligned} &\int_0^{t\wedge\sigma_i} f_j(s)dW(s) = \int_0^t f_j(s)\chi_{[0,\sigma_i]}(s)dW(s) \\ &= \int_0^t f(s)\chi_{[0,\sigma_j]}(s)\chi_{[0,\sigma_i]}(s)dW(s) = \int_0^t f_i(s)dW(s), \quad \forall i \leq j. \end{aligned}$$

Hence, the following is well-defined:

$$\int_0^t f(s)dW(s) \stackrel{\Delta}{=} \int_0^t f_j(s)dW(s), \qquad \forall\, t \in [0,\sigma_j],\ j = 1, 2, \cdots.$$

This is called the *Itô integral* of $f(\cdot) \in L^{2,loc}_{\mathcal{F}}(0,T)$. It is easy to see that $\int_0^t f(s)dW(s) \in \mathcal{M}_c^{2,loc}[0,T]$ for any $f(\cdot) \in L^{2,loc}_{\mathcal{F}}(0,T)$.

We point out that for $f(\cdot) \in L^{2,loc}_{\mathcal{F}}(0,T)$, (2.5)–(2.10) do not hold in general, but (2.4) and (2.11) remain true. We list as follows some basic properties of the Itô's integral of $L^{2,loc}_{\mathcal{F}}(0,T)$-processes.

Theorem 2.6. *Let* $f, g \in L^{2,loc}_{\mathcal{F}}(0,T)$. *Then*

(i) For any $\{\mathcal{F}_t\}$*-stopping time* σ*, it holds*

$$\int_0^{t\wedge\sigma} f dW = \int_0^t f\chi_{[0,\sigma]}(s)dW(s);$$

(ii) For any $\Omega_0 \subset \Omega$ *and* $\{\mathcal{F}_t\}$*-stopping time* σ*, if* $f = 0$ *in* $A \stackrel{\Delta}{=} \{(t,\omega) \in [0,T] \times \Omega_0 \mid t \leq \sigma(\omega)\}$*, then*

$$\int_0^t f dW = 0, \quad \text{in } A;$$

(iii) For any $\varepsilon > 0$ *and* $\delta > 0$*, it holds*

$$P\Big(\sup_{s\in[0,t]}\Big|\int_0^s fdW\Big|\geq\varepsilon\Big)\leq P\Big(\int_0^t|f(u)|^2du\geq\delta\Big)+\frac{\delta}{\varepsilon^2}.$$

Hence, if $f^n, f\in L^{2,loc}_{\mathcal{F}}(0,T)$, *and*

$$\int_0^t|f^n(u)-f(u)|^2du\to 0,\quad \textit{in } P,$$

then

$$\sup_{s\in[0,t]}\Big|\int_0^s f^n dW-\int_0^s fdW\Big|\to 0,\quad \textit{in } P;$$

(iv) For any bounded $\{\mathcal{F}_t\}$*-stopping times* τ *and* σ *with* $\tau\leq\sigma$ a.s., *and any bounded* $\mathcal{F}_\tau$*-measurable random variable* ξ_1 *and* ξ_2, *it holds*

$$\int_\tau^\sigma(\xi_1 f+\xi_2 g)dW=\xi_1\int_\tau^\sigma fdW+\xi_2\int_\tau^\sigma gdW.$$

Proof. We only prove (iii). Choose $f^\delta(s)=f(s)\chi_B(s)\in L^2_{\mathcal{F}}(0,T)$, where

$$B=\Big\{s\in[0,T]\ \Big|\ \int_0^s|f(u)|^2du\leq\delta\Big\}.$$

Then, thanks to Doob's inequality, we have

$$\begin{aligned}
&P\Big(\sup_{s\in[0,t]}\Big|\int_0^s fdW\Big|\geq\varepsilon\Big)\\
&\leq P\Big(\sup_{s\in[0,t]}\Big|\int_0^s(f-f^\delta)dW\Big|\geq\varepsilon\Big)+P\Big(\sup_{s\in[0,t]}\Big|\int_0^s f^\delta dW\Big|\geq\varepsilon\Big)\\
&\leq P\Big(\int_0^t|f(u)|^2du\geq\delta\Big)+\varepsilon^{-2}E\Big(\int_0^t f^\delta dW\Big)^2\\
&=P\Big(\int_0^t|f(u)|^2du\geq\delta\Big)+\varepsilon^{-2}E\int_0^t|f^\delta(u)|^2du\\
&\leq P\Big(\int_0^t|f(u)|^2du\geq\delta\Big)+\frac{\delta}{\varepsilon^2}.
\end{aligned}$$

□

To conclude this subsection, we present an important result called the *Burkholder-Davis-Gundy inequality.*

Theorem 2.7. *Let* $(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{t\geq0},P)$ *be given as before and* $W(t)$ *be an* m*-dimensional standard Brownian motion. Let* $\sigma\in L^{2,loc}_{\mathcal{F}}(0,T;\mathbb{R}^{n\times m})$. *Then, for any* $r>0$, *there exists a constant* $K_r>0$ *such that for any stopping time* τ,

$$\begin{aligned}
&\frac{1}{K_r}E\Big\{\int_0^\tau|\sigma(s)|^2ds\Big\}^r\leq E\Big\{\sup_{0\leq t\leq\tau}\Big|\int_0^t\sigma(s)dW(s)\Big|^{2r}\Big\}\\
&\leq K_rE\Big\{\int_0^\tau|\sigma(s)|^2ds\Big\}^r.
\end{aligned}$$

2.2 Itô's formula

In this subsection, we present a stochastic version of the *chain rule* or *change-of-variable formula*, called *Itô's formula/lemma/rule*, which plays one of the most important roles in stochastic calculus.

Definition 2.8. *For any* $b(\cdot) \in L^{1,loc}_{\mathcal{F}}(0,T)$ *and* $\sigma(\cdot) \in L^{2,loc}_{\mathcal{F}}(0,T)$, $\{\mathcal{F}_t\}$-*adapted process* $X(\cdot)$ *of the form*

$$X(t) = X(0) + \int_0^t b(s)ds + \int_0^t \sigma(s)dW(s), \qquad t \in [0,T], \tag{2.12}$$

is called an Itô process.

Theorem 2.9. (Itô's formula) *Let* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ *be a filtered probability space satisfying the usual condition,* $W(t)$ *be a* 1*-dimensional* $\{\mathcal{F}_t\}_{t\geq 0}$*-Brownian motion,* $b(\cdot) \in L^{1,loc}_{\mathcal{F}}(0,T)$, $\sigma \in L^{2,loc}_{\mathcal{F}}(0,T)$, *and* $X(\cdot)$ *be given by (2.12). Let* $F(t,x) : [0,T] \times \mathbb{R} \to \mathbb{R}$ *be* C^1 *in* t *and* C^2 *in* x *with* F_t, F_x *and* F_{xx} *being continuous such that*

$$\begin{cases} F_t(\cdot, X(\cdot)), \quad F_x(\cdot, X(\cdot))b(\cdot) \in L^{1,loc}_{\mathcal{F}}(0,T), \\ F_{xx}(\cdot, X(\cdot))(\sigma(\cdot))^2 \in L^{1,loc}_{\mathcal{F}}(0,T;\mathbb{R}), \\ F_x(\cdot, X(\cdot))\sigma(\cdot) \in L^{2,loc}_{\mathcal{F}}(0,T). \end{cases}$$

Then

$$\begin{aligned} F(t, X(t)) = F(0, X(0)) &+ \int_0^t \Big\{ F_s(s, X(s)) + F_x(s, X(s))b(s) \\ &+ \frac{1}{2} F_{xx}(s, X(s))(\sigma(s))^2 \Big\} ds \\ &+ \int_0^t F_x(s, X(s))\sigma(s)dW(s), \quad \forall\, t \in [0,T], P\text{-a.s.} \end{aligned} \tag{2.13}$$

Proof. The main idea is to use the Taylor's formula. We fix any $t \in [0,T]$.

Step 1. We claim that, it suffices to prove (2.13) under the additional assumptions that $X(s)$, $\int_0^s \sigma dW$, $\int_0^s |\sigma(u)|^2 du$ and $\int_0^s |b(u)|du$ are uniformly bounded in $[0,t] \times \Omega$.

Indeed, for the general case, we set

$$\tau^n \triangleq \inf \left\{ u \geq 0 \;\Big|\; \Big| \int_0^u \sigma dW \Big| \vee \Big| \int_0^u |\sigma(s)|^2 ds \Big| \vee \Big| \int_0^u |b(s)|ds \Big| \geq n \right\},$$

($\inf \emptyset = t$). It is easy to see that $\{\tau^n\}_{n=1}^\infty$ is a sequence of $\{\mathcal{F}_t\}$-stopping times.

Put

$$X^n(t) \triangleq X^n(0) + \int_0^t \sigma^n dW + \int_0^t b^n ds,$$

where

$$X^n(0) = X(0)\chi_{(|X(0)|\leq n)}, \quad \sigma^n(s) = \sigma(s)\chi_{[0,\tau^n]}(s), \quad b^n(s) = b(s)\chi_{[0,\tau^n]}(s).$$

Hence, $X^n(0)$, $\int_0^t \sigma^n dW$, $\int_0^t |\sigma^n(s)|^2 ds$ and $\int_0^t |b^n(s)|ds$ are uniformly bounded. If the Itô's formula (2.13) holds for X^n, then we have

$$\begin{aligned} F(t, X^n(t)) = F(0, X^n(0)) &+ \int_0^t \Big\{F_s(s, X^n(s)) + F_x(s, X^n(s))b^n(s) \\ &+\frac{1}{2}F_{xx}(s, X^n(s))(\sigma^n(s))^2\Big\}ds \\ &+ \int_0^t F_x(s, X^n(s))\sigma^n(s)dW(s), \quad P\text{-a.s.} \end{aligned} \tag{2.14}$$

Clearly,

$$F(t, X^n(t)) \to F(t, X(t)), \quad F(0, X^n(0)) \to F(0, X(0)), \quad \text{a.s.} \quad \text{as } n \to \infty.$$

Now, for any fixed $\omega \in \Omega$, $F_x(s, X(s))$ is continuous with respect to $s \in [0,t]$ and therefore it is bounded by a constant C. Hence

$$|F_x(s, X^n(s))\sigma^n(s) - F_x(s, X(s))\sigma(s)|^2 \leq C|\sigma(s)|^2 \in L^{1,loc}_{\mathcal{F}}(0,T).$$

Hence,

$$\int_0^t |F_x(s, X^n(s))\sigma^n(s) - F_x(s, X(s))\sigma(s)|^2 ds \to 0, \text{ a.s. as } n \to \infty.$$

Thus, by Theorem 2.6 (iii), we conclude that

$$\int_0^t F_x(s, X^n(s))\sigma^n(s)dW(s) \to \int_0^t F_x(s, X(s))\sigma(s)dW(s), \text{ a.s. as } n \to \infty.$$

Similarly,

$$\int_0^t |F_x(s, X^n(s))b^n(s) - F_x(s, X(s))b(s)|ds, \text{ a.s. as } n \to \infty.$$

Also, using the boundedness for fixed ω, we get

$$\begin{aligned} &\int_0^t \Big\{F_s(s, X^n(s)) + \frac{1}{2}F_{xx}(s, X^n(s))(\sigma^n(s))^2\Big\}ds \\ &\quad \to \int_0^t \Big\{F_s(s, X(s)) + \frac{1}{2}F_{xx}(s, X(s))(\sigma(s))^2\Big\}ds, \text{ a.s. as } n \to \infty. \end{aligned}$$

Now, letting $n \to \infty$ in (2.14), one gets (2.13).

Step 2. Now, let us show (2.13) under the additional conditions that $X(s)$, $\int_0^s \sigma dW$, $\int_0^s |\sigma(u)|^2 du$ and $\int_0^s |b(u)|du$ are uniformly bounded in $[0,t] \times \Omega$ by some constant K. Denote the upper bound of $F_t(t,x)$, $F_x(t,x)$ and $F_{xx}(t,x)$ over $A \triangleq [0,t]\times[-K,K]$ by C.

Using the Taylor's expansion, we see that there ia a function $\varepsilon(s)$, decreasing to 0 as $s\to 0$, such that

$$\begin{aligned}&|F(s,x_2)-F(s,x_1)-F_x(s,x_1)(x_2-x_1)-\tfrac{1}{2}F_{xx}(x_1,s)(x_2-x_1)^2|\\&\leq \varepsilon(|x_2-x_1|)(x_2-x_1)^2,\quad \forall\,(s,x_1),(s,x_2)\in A,\end{aligned}$$

and

$$\begin{aligned}&|F(s_2,x)-F(s_1,x)-F_s(s_1,x)(s_2-s_1)|\\&\leq \varepsilon(|s_2-s_1|)(s_2-s_1),\quad \forall\,(s_1,x),(s_2,x)\in A.\end{aligned}$$

Step 3. For any $h>0$ and $k=0,1,2\cdots$, put

$$\begin{cases}\tau_0=0,\\ \hat{\tau}_{k+1}=\inf\left\{u\in[\tau_k,t]\;\Big|\;\Big|\int_{\tau_k}^u\sigma dW\Big|\vee\int_\tau^u|b(s)|ds\vee\int_\tau^u|\sigma(s)|^2ds\geq h\right\},\\ \tau_{k+1}=\hat{\tau}_{k+1}\wedge(\tau_k+h)\wedge t.\end{cases}$$

(Convention: $\inf\emptyset=t$). One can show that $\{\tau_k\}$ is a sequence of $\{\mathcal{F}_t\}$-stopping times, and

$$\tau_k\leq\tau_{k+1}\leq\tau_k+h,\quad \Big|\int_{\tau_k}^{\tau_{k+1}}\sigma dW\Big|\leq h,\quad \int_{\tau_k}^{\tau_{k+1}}|b(s)|ds\leq h.$$

It is easy to see that

$$\begin{aligned}I&\triangleq F(t,X(t))-F(0,X(0))\\&=\sum_k[F(\tau_{k+1},X(\tau_{k+1}))-F(\tau_k,X(\tau_{k+1}))]\\&\quad+\sum_k[F(\tau_k,X(\tau_{k+1}))-F(\tau_k,X(\tau_k))].\end{aligned}$$

Put

$$\begin{aligned}I_h=\sum_k\Big[&F_s(\tau_k,X(\tau_{k+1}))\Delta\tau_k+F_x(\tau_k,X(\tau_k))\Delta X(\tau_k)\\&+\frac{1}{2}F_{xx}(\tau_k,X(\tau_k))(\Delta X(\tau_k))^2\Big].\end{aligned}$$

Then

$$\begin{aligned}|I-I_h|&\leq\sum_k\Big[\varepsilon(h)|\Delta\tau_k|+\varepsilon(2h)(\Delta X(\tau_k))^2\Big]\\&\leq\varepsilon(2h)\Big\{t+2\sum_k\Big[\Big(\int_{\tau_k}^{\tau_{k+1}}\sigma dW\Big)^2+\Big|\int_{\tau_k}^{\tau_{k+1}}b(u)du\Big|^2\Big]\Big\}.\end{aligned}$$

Note that

$$\sum_k \Big(\int_{\tau_k}^{\tau_{k+1}} \sigma dW \Big)^2 = \sum_k \int_{\tau_k}^{\tau_{k+1}} |\sigma(s)|^2 ds = E \int_0^t |\sigma(s)|^2 ds.$$

Hence,

$$\varepsilon(2h) \sum_k \Big(\int_{\tau_k}^{\tau_{k+1}} \sigma dW \Big)^2 \to 0 \quad \text{in } P \text{ as } h \to .$$

On the other hand,

$$\sum_k \Big| \int_{\tau_k}^{\tau_{k+1}} b(u) du \Big|^2 \le h \sum_k \int_{\tau_k}^{\tau_{k+1}} |b(u)| du \le h \int_0^t |b(u)| du.$$

Hence,

$$|I_h - I| \to 0 \quad \text{in } P \text{ as } h \to .$$

In what follows, we will analyze the limit of I_h in probability as $h \to 0$.

Step 4. Clearly, the first term in I_h tends t0 $\int_0^t F_s(s, X(s)) ds$ a.s. as $h \to 0$. Recalling that

$$\Delta X(\tau_k) = \int_{\tau_k}^{\tau_{k+1}} \sigma dW + \int_{\tau_k}^{\tau_{k+1}} b(u) du.$$

Hence, the second term in I_h can be split into two terms. The last term tends to $\int_0^t F_x(s, X(s)) b(s) ds$ a.s. as $h \to 0$; while the previous term reads

$$\sum_k F_x(\tau_k, X(\tau_k)) \int_{\tau_k}^{\tau_{k+1}} \sigma dW = \sum_k \int_{\tau_k}^{\tau_{k+1}} F_x(\tau_k, X(\tau_x)) \sigma(s) dW(s)$$
$$= \int_0^t \sum_k F_x(\tau_k, X(\tau_x)) \chi_{(\tau_k, \tau_{k+1}]}(s) \sigma(s) dW(s).$$

However, it follows from the dominated convergence theorem that

$$E \int_0^t \Big| \sum_k F_x(\tau_k, X(\tau_k)) \chi_{(\tau_k, \tau_{k+1}]}(s) - F_x(s, X(s)) \Big|^2 |\sigma(s)|^2 ds \to 0$$

as $h \to 0$. Hence,

$$\sum_k F_x(\tau_k, X(\tau_k)) \int_{\tau_k}^{\tau_{k+1}} \sigma dW \to \int_0^t F_x(s, X(s)) dW(s)$$

in $\mathcal{M}_c^2[0,t]$ as $h \to 0$.

Similarly, by

$$\Big| \Delta X(\tau_k) \Big|^2 = \Big| \int_{\tau_k}^{\tau_{k+1}} \sigma dW \Big|^2 + 2 \int_{\tau_k}^{\tau_{k+1}} \sigma dW \int_{\tau_k}^{\tau_{k+1}} b(u) du + \Big| \int_{\tau_k}^{\tau_{k+1}} b(u) du \Big|^2,$$

we see that the third term in I_h can be split into three terms. The last two terms tends to 0 as $h \to 0$; while the first one reads

$$\begin{aligned}&\frac{1}{2}\sum_k F_{xx}(\tau_k, X(\tau_k))\Big|\int_{\tau_k}^{\tau_{k+1}}\sigma dW\Big|^2\\&=\frac{1}{2}\sum_k F_{xx}(\tau_k, X(\tau_k))\int_{\tau_k}^{\tau_{k+1}}|\sigma(s)|^2 ds\\&\quad+\frac{1}{2}\sum_k F_{xx}(\tau_k, X(\tau_k))\Big[\Big|\int_{\tau_k}^{\tau_{k+1}}\sigma dW\Big|^2-\int_{\tau_k}^{\tau_{k+1}}|\sigma(s)|^2 ds\Big]\\&\equiv I_1+I_2.\end{aligned}$$

Obviously, I_1 tends to $\frac{1}{2}\int_0^t F_{xx}(\tau_k, X(\tau_k))|\sigma(s)|^2 ds$ a.s. as $h \to 0$. However,

$$\begin{aligned}E(I_2)^2&\le\frac{C^2}{4}\sum_k\Big[\Big|\int_{\tau_k}^{\tau_{k+1}}\sigma dW\Big|^2-\int_{\tau_k}^{\tau_{k+1}}|\sigma(s)|^2 ds\Big]^2\\&\le\frac{C^2}{2}\sum_k\Big[\Big|\int_{\tau_k}^{\tau_{k+1}}\sigma dW\Big|^4+\Big(\int_{\tau_k}^{\tau_{k+1}}|\sigma(s)|^2 ds\Big)^2\Big]\\&\le\frac{C^2 h}{2}\Big(hE\Big|\int_0^t\sigma dW\Big|^2+E\int_0^t|\sigma(s)|^2 ds\Big)\to 0,\qquad\text{as } h\to 0.\end{aligned}$$

Hence, $\lim\limits_{h\to 0} I_2 = 0$ in probability.

Combining the above analysis, we conclude that

$$\begin{aligned}I_h \to &\int_0^t\Big\{F_s(s, X^n(s))+F_x(s, X^n(s))b^n(s)+\frac{1}{2}F_{xx}(s, X^n(s))(\sigma^n(s))^2\Big\}ds\\&+\int_0^t F_x(s, X^n(s))\sigma^n(s)dW(s)\end{aligned}$$

in probability as $h \to 0$, which gives the desired result. □

Finally, let us make an observation. Take $\sigma(\cdot) \in L^2_{\mathcal{F}}(0,T)$ and consider

$$X(t)=\int_0^t\sigma(s)dW(s),\quad t\in[0,T].$$

Then $X(\cdot) \in \mathcal{M}^2_c[0,T]$. By Itô's formula above, we have

$$X(t)^2=\int_0^t\sigma(s)^2 ds+2\int_0^t X(s)\sigma(s)dW(s),\quad t\in[0,T].$$

Note that $X(\cdot)\sigma(\cdot)$ is in $L^{2,loc}_{\mathcal{F}}(0,T)$, but not necessarily in $L^2_{\mathcal{F}}(0,T)$. Hence, the introduction of Itô's integral for integrands in $L^{2,loc}_{\mathcal{F}}(0,T)$ is not just some routine generalization. It is really necessary even for as simple calculus as the above.

3

Stochastic Ordinal Differential Equations

3.1 Stochastic differential equations

We denote $W^m \triangleq C([0,\infty);\mathbb{R}^m)$. Clearly, W^m is a Fréchet space with the usual topology. Define

$$W_t^m \triangleq \Big\{\zeta(\cdot\wedge t)\ \Big|\ \zeta(\cdot)\in W^m\Big\},\qquad \forall\, t\in[0,\infty).$$

Obviously, W_t^m is a closed subspace of W^m. Further, W_t^m is a Banach space with the inherent topology.

Denote by $\mathcal{B}(W^m)$ and $\mathcal{B}(W_t^m)$ respectively the Borel σ-fields of W^m and W_t^m. It is easy ro see that $\mathcal{B}(W_t^m)$ is not a σ-field on W^m since $W^m \notin \mathcal{B}(W_t^m)$. Define

$$\mathcal{B}_t(W^m) \triangleq \sigma(W^m,\mathcal{B}(W_t^m)),\quad \mathcal{B}_{t+}(W^m) \triangleq \bigcap_{s>t}\mathcal{B}_s(W^m),\qquad \forall\, t\in[0,\infty).$$

Obviously, both the following are filtered measurable spaces:

$$(W^m,\mathcal{B}(W^m),\{\mathcal{B}_t(W^m)\}_{t\ge 0}),\quad (W^m,\mathcal{B}(W^m),\{\mathcal{B}_{t+}(W^m)\}_{t\ge 0}).$$

For any topology space U, we denote by $\mathcal{A}^m(U)$ the set of the all $\{\mathcal{B}_{t+}(W^m)\}_{t\ge 0}$-progressively measurable processes $\eta:\ [0,\infty)\times W^m\to U$.

We need the following technical result:

Lemma 3.1. *Let $b\in\mathcal{A}^m(\mathbb{R}^n)$ and $(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{t\ge 0},P)$ be given, and X be a continuous $\mathbb{R}^n$-valued $\{\mathcal{F}_t\}_{t\ge 0}$-adapted process. Then the process $t\mapsto b(t,X)$ is $\{\mathcal{F}_t\}_{t\ge 0}$ adapted.*

Now, for any given $b\in\mathcal{A}^n(\mathbb{R}^n)$ and $\sigma\in\mathcal{A}^n(\mathbb{R}^{n\times m})$, we consider the following equation:

$$\begin{cases} dX(t) = b(t,X)dt + \sigma(t,X)dW(t), \quad t\ge 0,\\ X(0)=\xi. \end{cases} \tag{3.1}$$

Here, X is the unknown, ξ, the initial datum, is a $\mathcal{F}_0$-measurable $\mathbb{R}^n$-valued function. Such an equation is called a stochastic differential equation.

To begin with, as for ordinal differential equations, one needs to define the solution to (3.1). The following definition is the most natural one.

Definition 3.2. *An $\{\mathcal{F}_t\}_{t\geq 0}$-adapted continuous process $\{X(t)\}_{t\geq 0}$ is called a solution to (3.1) if*

(i)

$$X(0) = \xi, \qquad P-\text{a.s.}, \tag{3.2}$$

(ii) for each $t \geq 0$,

$$\int_0^t [|b(s,X)| + |\sigma(s,X)|^2] ds < \infty, \quad P-\text{a.s.}, \tag{3.3}$$

(iii) for each $t \geq 0$,

$$X(t) = X(0) + \int_0^t b(s,X)ds + \int_0^t \sigma(s,X)dW(s), \quad P-\text{a.s.} \tag{3.4}$$

If for any other solution Y to (3.1), it holds

$$P(X(t) = Y(t),\ t \geq 0) = 1,$$

then we say that the solution of (3.1) is unique.

We mention that there are some other definition for solution to (3.1).

In (3.4), the first integral on the right is the usual Lebesgue integral and the second one is the Itô's integral. Clearly, if (3.3) holds, then these two integrals are well-defined. In the sequel, we refer to $\int_0^t b(s,X)ds$ and $\int_0^t \sigma(s,X)dW(s)$ as the drift and diffusion terms, respectively.

We now show the existence and uniqueness of solution to (3.1). For this, we need the following assumptions.

(H) $b \in \mathcal{A}^n(\mathbb{R}^n)$, $\sigma \in \mathcal{A}^n(\mathbb{R}^{n\times m})$, and there is a constant $L > 0$ such that for all $t \in [0,\infty)$, $x(\cdot), y(\cdot) \in W_t^m$, it holds

$$\begin{cases} |b(t,x(\cdot)) - b(t,y(\cdot))| \leq L|x(\cdot) - y(\cdot)|_{W_t^m}, \\ |\sigma(t,x(\cdot)) - \sigma(t,y(\cdot))| \leq L|x(\cdot) - y(\cdot)|_{W_t^m}, \\ b(\cdot,0) \in L_{\mathcal{F}}^{1,loc}(0,\infty), \quad \sigma(\cdot,0) \in L_{\mathcal{F}}^{2,loc}(0,\infty). \end{cases} \tag{3.5}$$

Now, we may state the fundamental existence and uniqueness result for equation (3.1).

Theorem 3.3. *Let assumption (H) hold. Then, for any $\xi \in L^{\ell}_{\mathcal{F}_0}(\Omega;\mathbb{R}^n)$ with some $\ell \geq 1$, equation (3.1) admits a unique solution X such that for any $T > 0$, there is a constant $C = C(T,\ell)$ so that*

$$E \sup_{0\leq s\leq T} |X(s)|^{\ell} \leq C(1 + E|\xi|^{\ell}), \tag{3.6}$$

and

$$E|X(t)-X(s)|^{\ell} \le C(1+E|\xi|^{\ell})|t-s|^{\ell/2}, \quad \forall\, t,s \in [0,T]. \tag{3.7}$$

Moreover, if $\hat{\xi} \in L^{\ell}_{\mathcal{F}_0}(\Omega;\mathbb{R}^n)$ *is another initial datum and* $\widetilde{X}(t)$ *is the corresponding solution of (3.1), then*

$$E \sup_{0\le s\le T} |X(s)-\widetilde{X}(s)| \le CE|\xi-\hat{\xi}|^{\ell}. \tag{3.8}$$

Proof. For any given $s>0$, put

$$V_s = \Big\{ x(\cdot): \ [0,s]\times\Omega \mapsto \mathbb{R}^n \ \Big| \ x(\cdot) \text{ is a } \{\mathcal{F}_t\}_{t\ge 0}\text{-adapted continuous process with } E\sup_{0\le t\le T}|x(t)|^{\ell} \le \infty \Big\}.$$

Clearly, V_s is a Banach space with the following norm:

$$|x(\cdot)|_{V_s} \triangleq \Big\{ E\sup_{0\le t\le s}|x(t)|^{\ell}\Big\}^{1/\ell}, \quad x(\cdot)\in V_s.$$

Now, for any fixed $T>0$ and any $T_0\in(0,T)$. For any $x(\cdot)\in V_s$, we define

$$F(x(\cdot))(t) = \xi + \int_0^t b(u,x)du + \int_0^t \sigma(u,x)dW(u), \quad t\in[0,T_0]. \tag{3.9}$$

Then, by (3.5) and Hölder's inequality, we get

$$\begin{aligned}
&E\sup_{0\le t\le T_0}\Big|\int_0^t b(s,x)ds\Big|^{\ell} \\
&\le C\left[\sup_{0\le t\le T_0}\Big|\int_0^t b(s,0)ds\Big|^{\ell} + E\sup_{0\le t\le T_0}\Big|\int_0^t |x(s)|ds\Big|^{\ell}\right] \\
&\le C\left[\Big(\int_0^{T_0}|b(s,0)|ds\Big)^{\ell} + E\Big(\int_0^{T_0}\sup_{0\le u\le s}|x(u)|ds\Big)^{\ell}\right] \\
&\le C\left[\Big(\int_0^{T_0}|b(s,0)|ds\Big)^{\ell} + T_0^{\ell-1}\int_0^{T_0}E\sup_{0\le u\le s}|x(u)|^{\ell}ds\right].
\end{aligned} \tag{3.10}$$

Similarly, by (3.5) and Burkholder-Davis-Gundy inequality, we have

$$
\begin{aligned}
&E \sup_{0\le t\le T_0}\Big|\int_0^t \sigma(s,x)dW(s)\Big|^\ell \\
&\le C\left[E\sup_{0\le t\le T_0}\Big|\int_0^t \big(\sigma(s,x)-\sigma(s,0)\big)dW\Big|^\ell + E\sup_{0\le t\le T_0}\Big|\int_0^t \sigma(s,0)dW\Big|^\ell\right] \\
&\le C\left[E\Big(\int_0^{T_0}\big|\sigma(s,x)-\sigma(s,0)\big|^2 ds\Big)^{\ell/2} + E\Big(\int_0^{T_0}|\sigma(s,0)|^2 ds\Big)^{\ell/2}\right] \\
&\le C\left[E\Big(\int_0^{T_0}\sup_{0\le u\le s}|x(u)|^2 ds\Big)^{\ell/2} + \Big(\int_0^{T_0}|\sigma(s,0)|^2 ds\Big)^{\ell/2}\right] \\
&\le C\left[T_0^{\ell/2-1}\int_0^{T_0} E\sup_{0\le u\le s}|x(u)|^\ell ds + \Big(\int_0^{T_0}|\sigma(s,0)|^2 ds\Big)^{\ell/2}\right].
\end{aligned}
\tag{3.11}
$$

Hence, by combining (3.9)–(3.11), we end up with

$$
\begin{aligned}
&E\sup_{0\le t\le T_0}|F(x(\cdot)(t)|^\ell \\
&\le C\Bigg[T_0^{\ell/2-1}\int_0^{T_0}|x(\cdot)|_{V_s}^\ell ds \\
&\quad +|\xi|^\ell + E\Big(\int_0^{T_0}|b(s,0)|ds\Big)^\ell + \Big(\int_0^{T_0}|\sigma(s,0)|^2 ds\Big)^{\ell/2}\Bigg],
\end{aligned}
\tag{3.12}
$$

where $C = C(T,\ell) > 0$ is a constant, independent of T_0, ξ, $b(s,0)$, $\sigma(s,0)$ or $x(\cdot)$.

Now, by (3.12), it is easy to see that $F(x(\cdot)) \in V_{T_0}$, and

$$
|F(x(\cdot)) - F(y(\cdot))|_{V_{T_0}} \le C T_0^{1/2}|x(\cdot)-y(\cdot)|_{V_{T_0}}, \quad \forall\, x(\cdot), y(\cdot) \in V_{T_0}. \tag{3.13}
$$

We choose $T_0 \in (0,T)$ such that $CT_0^{1/2} < 1$. Then it is easy to see that the map F is from V_{T_0} to itself and contractive. Thus, there is a unique fixed point, which gives a solution $X(\cdot)$ to (3.1) on $[0,T_0]$. Repeating this procedure, we get a solution on $[0,T]$. Since T is arbitrary, we obtain the solution on $[0,\infty)$.

However, by (3.12), it is easy to see that solution $X(\cdot)$ to (3.1) satisfies

$$
|X|_{V_t}^\ell \le C\Big(1 + E|\xi|^\ell + t^{\ell/2-1}\int_0^t |X|_{V_s}^\ell ds\Big), \quad \forall\, t \ge 0.
$$

By this, and using Gronwall's inequality, we get (3.6).

The proof of the rest assertions follows from a similar argument. □

3.2 Martingale representation theorems

In view of (2.5), the Itô integral

$$M(t) \stackrel{\Delta}{=} \int_0^t f(s)dW(s) \tag{3.14}$$

is a martingale (or local martingale). It is natural to ask if a martingale (or local martingale) can be represented as an Itô integral of the form (3.14). Results on such a problem are called *martingale representation theorems*. They play very important roles in stochastic calculus itself as well as in stochastic control theory.

Theorem 3.4. *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ be a filtered probability space satisfying the usual condition. Assume that $\{\mathcal{F}_t\}_{t\geq 0}$ is the natural filtration generated by an m-dimensional standard Brownian motion $W(t)$. Let $X \in \mathcal{M}^2[0,T]$ (resp. $\mathcal{M}^{2,loc}[0,T]$). Then there exists a unique $\varphi(\cdot) \in L^2_{\mathcal{F}}(0,T;\mathbb{R}^m)$ (resp. $L^{2,loc}_{\mathcal{F}}(0,T;\mathbb{R}^m)$), such that*

$$X(t) = \int_0^t \langle\, \varphi(s), dW(s)\,\rangle, \qquad \forall t \in [0,T],\ P\text{-a.s.}$$

We emphasize that Theorem 3.4 works only for the case when $\{\mathcal{F}_t\}_{t\geq 0}$ is generated by the (given) Brownian motion. On the other hand, it implies that if $\{\mathcal{F}_t\}_{t\geq 0}$ *is* generated by a standard Brownian motion, then any square integrable $\{\mathcal{F}_t\}_{t\geq 0}$-martingale must be continuous, i.e., $\mathcal{M}^2[0,T] = \mathcal{M}^2_c[0,T]$.

Let us now look at an interesting consequence of Theorem 3.4. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ and $W(\cdot)$ be the same as in the above. Let $\xi \in L^2_{\mathcal{F}_T}(\Omega)$. Then $E(\xi \mid \mathcal{F}_t)$ is an $\{\mathcal{F}_t\}_{t\geq 0}$-martingale. Thus, by Theorem 3.4, there exists a $z(\cdot) \in L^2_{\mathcal{F}}(0,T;\mathbb{R}^m)$, such that

$$E(\xi \mid \mathcal{F}_t) = E\xi + \int_0^t \langle\, z(s), dW(s)\,\rangle, \qquad t \in [0,T].$$

In particular,

$$\xi = E\xi + \int_0^T \langle\, z(s), dW(s)\,\rangle\,.$$

This shows that

$$L^2_{\mathcal{F}_T}(\Omega) = \mathbb{R} + \Big\{ \int_0^T \langle\, z(s), dW(s)\,\rangle \;\Big|\; z(\cdot) \in L^2_{\mathcal{F}}(0,T;\mathbb{R}^m) \Big\}.$$

3.3 Backward stochastic differential equations

It is well-known that, for an ordinary differential equations, the terminal value problem on $[0,T]$ is equivalent to an initial value problem on $[0,T]$ under the time transformation $t \mapsto T - t$. However, things are fundamentally different for stochastic differential equations. One of the main differences between a stochastic different equation and the ordinary differential equation is that one can not reserve the "time" when we are looking for a solution that is adapted to the given filtration. Practically, this means that, one only knows about what has happen in the past, but can not foretell what is going to happen in the future. Mathematically, this means that we would like to keep the context within the framework of the Itô type stochastic calculus. As a result, one can not simply reserve the time to get a solution for a terminal value problem of stochastic case.

In what follows, the terminal value problem of stochastic differential equations is said to be backward stochastic differential equation.

Let us analyze how to correctly formulate backward stochastic differential equation.

Consider the following simplest terminal value problem of stochastic differential equation:

$$\begin{cases} dY(t) = 0, & t \in [0,T], \\ Y(T) = \xi, \end{cases} \tag{3.15}$$

where $T > 0$ and $\xi \in L^2_{\mathcal{F}_T}(\Omega)$. We want to find a $\{\mathcal{F}_t\}_{t\geq 0}$ -adapted solution $Y(\cdot)$ to (3.15). However, this is impossible since the only solution of (3.15) is

$$Y(\cdot) \equiv \xi, \qquad t \in [0,T], \tag{3.16}$$

which is not necessary $\{\mathcal{F}_t\}_{t\geq 0}$ -adapted (unless ξ is $\mathcal{F}_0$-measurable). Namely, equation (3.15) is not well-formulated if one expects to find any $\{\mathcal{F}_t\}_{t\geq 0}$-adapted solution to it.

In what follows, we will see that the modified system of (3.15):

$$\begin{cases} dY(t) = Z(t)dW(t), & t \in [0,T], \\ Y(T) = \xi \end{cases} \tag{3.17}$$

is an appropriate reformation of (3.15). Comparing (3.17) with (3.15), we see that the term $Z(t)dW(t)$ has been added. Process $Z(\cdot)$ is not *a priori* known but a part of the solution! As a matter of fact, the presence of the term $Z(t)dW(t)$ "corrects" the "non-adaptiveness" of the original $Y(\cdot)$ in (3.15).

Generally, we consider the following backward stochastic differential equation in a fixed duration $[0,T]$:

$$\begin{cases} dY(t) = h(t, Y(t), Z(t))dt + Z(t)dW(t), & t \in [0,T], \\ Y(T) = \xi, \end{cases} \tag{3.18}$$

where $h:\ [0,T]\times\mathbb{R}^k\times\mathbb{R}^{k\times m}\times\Omega \to \mathbb{R}^k$ and $\xi \in L^2_{\mathcal{F}_t}(\Omega;\mathbb{R}^k)$. The main goal of this section is to find a pair of $\{\mathcal{F}_t\}_{t\geq 0}$ -adapted process $Y:\ [0,T]\times\Omega \to \mathbb{R}^k$ and $Z:\ (0,T)\times\Omega \to \mathbb{R}^{k\times m}$ satisfying (3.18).

Put

$$\mathcal{V} \stackrel{\triangle}{=} L^2_{\mathcal{F}}(\Omega; C([0,T];\mathbb{R}^k)) \times L^2_{\mathcal{F}}(0,T;\mathbb{R}^{k\times m}). \tag{3.19}$$

Clearly, $\mathcal{V}$ is a Banach space with the norm

$$|(Y(\cdot),Z(\cdot))|_{\mathcal{V}} \triangleq \left[E\Big(\sup_{0\le t\le T}|Y(t)|^2+\int_0^T|Z(t)|^2dt\Big)\right]^{1/2}. \tag{3.20}$$

Definition 3.5. *A pari of process $(Y(\cdot),Z(\cdot))\in\mathcal{V}$ is called an adapted solution of (3.18) if*

$$Y(t)=\xi-\int_t^T h(s,Y(s),Z(s))ds-\int_t^T ZdW, \quad \forall\, t\in[0,T], \text{a.s.} \tag{3.21}$$

Equation (3.18) is said to have a unique adapted solution if for any two adapted solutions $(Y(\cdot),Z(\cdot))$ and $(\widetilde{Y}(\cdot),\widetilde{Z}(\cdot))$, it must hold

$$P(Y(t)=\widetilde{Y}(t),\ \forall\, t\in[0,T])=1, \qquad \textit{and } P(Z(t)=\widetilde{Z}(t))=1,\ \text{a.e. } t\in[0,T].$$

The main result of this section is the following:

Theorem 3.6. *Suppose for any $(y,z)\in\mathbb{R}^k\times\mathbb{R}^{k\times m}$, $h(t,y,z)$ is $\{\mathcal{F}_t\}_{t\ge0}$ -adapted with $h(\cdot,0,0)\in L^2_{\mathcal{F}}(0,T;\mathbb{R}^k)$. Moreover, there is a $L>0$ such that*

$$\begin{aligned}|h(t,y,z)-h(y,\hat{y},\hat{z})|\le L(|y-\hat{y}|+|z-\hat{z}|),&\\ \forall\, t\in[0,T],\ y,\hat{y}\in\mathbb{R}^k,\ z,\hat{z}\in\mathbb{R}^{k\times m},\text{a.s.}&\end{aligned} \tag{3.22}$$

Then, for any given $\xi\in L^2_{\mathcal{F}_T}(\Omega;\mathbb{R}^k)$, equation (3.18) admits a unique adapted solution $(Y(\cdot),Z(\cdot))\in\mathcal{V}$.

Proof. The proof is divided into several steps.

Step 1. First, we show that, for any $h(\cdot)\in L^2_{\mathcal{F}}(0,T;\mathbb{R}^k)$ and $\xi\in L^2_{\mathcal{F}_T}(\Omega;\mathbb{R}^k)$, there is an adapted solution $(Y(\cdot),Z(\cdot))\in\mathcal{V}$ such that

$$Y(t)=\xi-\int_t^T h(s)ds-\int_t^T ZdW, \quad \forall\, t\in[0,T], \text{a.s.} \tag{3.23}$$

Indeed, put

$$M(t)=E(\xi-\int_0^T h(s)ds\mid\mathcal{F}_t), \quad Y(t)=E(\xi-\int_t^T h(s)ds\mid\mathcal{F}_t).$$

Then, $M(0)=Y(0)$. In view of the Martingale Representation Theorem, one obtains a $Z(\cdot)\in L^2_{\mathcal{F}}(0,T;\mathbb{R}^n)$ such that

$$M(t)=M(0)+\int_0^t ZdW.$$

Hence,

$$\xi-\int_0^T h(s)ds=M(T)=M(0)+\int_0^T ZdW=Y(0)+\int_0^T ZdW.$$

This means

$$\begin{aligned}Y(t) &= E(\xi - \int_0^T h(s)ds + \int_0^t h(s)ds \mid \mathcal{F}_t) = M(t) + \int_0^t h(s)ds \\ &= Y(0) + \int_0^t ZdW + \int_0^t h(s)ds = \xi - \int_t^T h(s)ds - \int_t^T ZdW.\end{aligned}$$

Step 2. For any fixed $(y(\cdot), z(\cdot)) \in \mathcal{V}$, it is easy to see that

$$h(\cdot) \stackrel{\triangle}{=} h(\cdot, y(\cdot), z(\cdot)) \in L^2_{\mathcal{F}}(0, T; \mathbb{R}^k).$$

Consider the following backward stochastic differential equation

$$\begin{cases} dY(t) = h(t, y(t), z(t))dt + Z(t)dW(t), \qquad t \in [0, T), \\ Y(T) = \xi. \end{cases} \tag{3.24}$$

By the result in Step 1, equation (3.24) admits a unique adapted solution $(Y(\cdot), Z(\cdot)) \in \mathcal{V}$. This defines a map $G: \ \mathcal{V} \to \mathcal{V}$ by $(y, z) \mapsto (Y, Z)$.

Step 3. Let us show that G admits a unique fixed point, which is the unique adapted solution of (3.18). For this purpose, we need to introduce an equivalent norm in $\mathcal{V}$:

$$|(\phi, \psi)|_\beta \stackrel{\triangle}{=} \left[E\Big(\sup_{0 \le t \le T} |e^{\beta t} Y(t)|^2 + \int_0^T |e^{\beta t} Z(t)|^2 dt \Big) \right]^{1/2},$$

where $\beta \in \mathbb{R}$ is any given parameter.

It suffices to show that there is a $\beta > 0$ such that

$$|G(y, z) - G(\tilde{y}, \tilde{z})|_\beta \le \frac{1}{2} |(y, z) - (\tilde{y}, \tilde{z})|_\beta, \quad \forall\, (y, z), (\tilde{y}, \tilde{z}) \in \mathcal{V}. \tag{3.25}$$

To show (3.25), put

$$\begin{cases} (\hat{Y}(\cdot), \hat{Z}(\cdot)) = G(y, z) - G(\tilde{y}, \tilde{z}), \quad \hat{y}(\cdot) = y(\cdot) - \tilde{y}(\cdot), \quad \hat{z}(\cdot) = z(\cdot) - \tilde{z}(\cdot), \\ \hat{h}(\cdot) = h(\cdot, y(\cdot), z(cd)) - h(\cdot, \tilde{y}(\cdot), \tilde{z}(cd)). \end{cases}$$

Obviously, $(\hat{Y}(\cdot), \hat{Z}(\cdot))$ satisfies

$$\begin{cases} d\hat{Y}(t) = \hat{h}(t)dt + \hat{Z}(t)dW(t), \qquad t \in [0, T), \\ \hat{Y}(T) = 0. \end{cases} \tag{3.26}$$

Applying Itô's formula to $|e^{\beta t} \hat{Y}(t)|^2$, and noting (3.22) and (3.26), we get

$$
\begin{aligned}
&|e^{\beta t}\hat{Y}(t)|^2 + \int_t^T |e^{\beta s}\hat{Z}(s)|^2 ds \\
&= -2\int_t^T [\beta|e^{\beta s}\hat{Y}(s)|^2 + \langle\, e^{\beta s}\hat{Y}(s), e^{\beta s}\hat{h}(s)\,\rangle]ds \\
&\quad -2\int_t^T e^{2\beta s}\,\langle\, \hat{Y}(s), \hat{Z}(s)dW(s)\,\rangle \\
&\le \int_t^T [-2\beta|\hat{Y}(s)|^2 + 2L|\hat{Y}(s)|(|\hat{y}(s)| + |\hat{z}(s)|)]e^{2\beta s}ds \qquad (3.27)\\
&\quad -2\int_t^T e^{2\beta s}\,\langle\, \hat{Y}(s), \hat{Z}(s)dW(s)\,\rangle \\
&\le \int_t^T [\Big(-2\beta + \frac{2L}{\lambda}\Big)|\hat{Y}(s)|^2 + \lambda(|\hat{y}(s)|^2 + |\hat{z}(s)|^2)]e^{2\beta s}ds \\
&\quad -2\int_t^T e^{2\beta s}\,\langle\, \hat{Y}(s), \hat{Z}(s)dW(s)\,\rangle, \quad \forall\, \lambda > 0.
\end{aligned}
$$

Hence, by taking $\lambda = L^2\beta^{-1}$, we get

$$
\begin{aligned}
&|e^{\beta t}\hat{Y}(t)|^2 + \int_t^T |e^{\beta s}\hat{Z}(s)|^2 ds \\
&\le \lambda(1+T)\left[E\Big(\sup_{0\le t\le T}|e^{\beta t}\hat{y}(t)|^2 + \int_0^T |e^{\beta t}\hat{z}(t)|^2 dt\Big)\right] \qquad (3.28)\\
&\quad -2\int_t^T e^{2\beta s}\,\langle\, \hat{Y}(s), \hat{Z}(s)dW(s)\,\rangle\,.
\end{aligned}
$$

By taking expectation in both sides of (3.28), one obtains

$$
E\left[|e^{\beta t}\hat{Y}(t)|^2 + \int_t^T |e^{\beta s}\hat{Z}(s)|^2 ds\right] \le \lambda(1+T)|(\hat{y}, \hat{z})|_\beta^2. \qquad (3.29)
$$

On the other hand, by (3.29) and Burkholder-Davis-Gundy inequality, we get

$$
\begin{aligned}
&E\left(\sup_{0\le t\le T}\left|\int_t^T e^{2\beta s}\,\langle\,\hat{Y}(s),\hat{Z}(s)dW(s)\,\rangle\right|\right)\\
&\le 2E\left(\sup_{0\le t\le T}\left|\int_0^t e^{2\beta s}\,\langle\,\hat{Y}(s),\hat{Z}(s)dW(s)\,\rangle\right|\right)\\
&\le CE\left(\int_0^T e^{4\beta s}|\hat{Y}(s)|^2|Z(s)|^2ds\right)^{1/2} \qquad (3.30)\\
&\le CE\left[\left(\sup_{0\le t\le T}|e^{\beta s}\hat{Y}(s)|^2\right)^{1/2}\left(\int_0^T |e^{\beta s}\hat{Z}(s)|^2ds\right)^{1/2}\right]\\
&\le \frac{1}{4}E\left(\sup_{0\le t\le T}|e^{\beta t}\hat{Y}(t)|^2\right)+C\lambda|(\hat{y},\hat{z})|_\beta^2.
\end{aligned}
$$

Combining (3.28) and (3.30), we arrive at

$$
\begin{aligned}
&E\left(\sup_{0\le t\le T}|e^{\beta t}\hat{Y}(t)|^2\right)\\
&\le \lambda(T+1)|(\hat{y},\hat{z})|_\beta^2+E\left(\sup_{0\le t\le T}\left|\int_t^T e^{2\beta s}\,\langle\,\hat{Y}(s),\hat{Z}(s)dW(s)\,\rangle\right|\right) \qquad (3.31)\\
&\le \frac{1}{2}E\left(\sup_{0\le t\le T}|e^{\beta t}\hat{Y}(t)|^2\right)+C\lambda|(\hat{y},\hat{z})|_\beta^2.
\end{aligned}
$$

Finally, it follows from (3.29) and (3.31) that (Recall that $\lambda = L^2/\beta$)

$$
|(\hat{Y},\hat{Z})|_\beta^2 \le \frac{C}{\beta}|(\hat{y},\hat{z})|_\beta^2, \qquad (3.32)
$$

where $C > 0$ is a constant, independent of β. By taking $\beta > 0$ large enough, we get the desired contractivity of the map G from $\mathcal{V}$ into itself. □

4

Stochastic Evolution Equations

4.1 Some preliminaries

Let and H be a separable Hilbert space. For any $a, b \in H$, we denote by $a \otimes b$ the linear operator defined by $(a \otimes b)h = a(b,h)_H,\ h \in H$. Assume $\{h_k\}$ to be a complete orthonormal basis in H. For any $S \in \mathcal{L}(H)$, we define the trace of S as

$$\mathrm{Tr}\, S \stackrel{\triangle}{=} \sum_{k=1}^{\infty} (Sh_k, h_k)_H.$$

It can be shown that this number is independent of the choice of the complete orthonormal basis.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : (\Omega, \mathcal{F}) \to (H, \mathcal{B}(H))$ be a H-valued random variable. The random variable X is said to be *Bochner integrable* (or simply *integrable*) if

$$\int_{\Omega} |X(\omega)|_H dP < \infty.$$

As before, in this case we say that X has a mean, and denoted by

$$EX = \int_{\Omega} X dP.$$

We also call EX the (mathematical) expectation of X.

One can define Banach spaces $L^p_{\mathcal{F}}(\Omega; H) \stackrel{\triangle}{=} L^p(\Omega, \mathcal{F}, P; H)$ for each $p \geq 1$. In particular, $L^2_{\mathcal{F}}(\Omega; H)$ is a Hilbert space.

If $X, Y \in L^2_{\mathcal{F}}(\Omega; H)$, we define the *covariance operator* of X and Y by

$$\mathrm{Cov}\,(X, Y) = E\big((X - EX) \otimes (Y - EY)\big).$$

In particular, $\mathrm{Var}\, X \stackrel{\triangle}{=} \mathrm{Cov}\,(X, X)$ is called the *variance operator* of X. It is easy to see that

$$\mathrm{Tr}\,\mathrm{Var}\, X = E|X - EX|^2.$$

Proposition 4.1. *Assume that X is a Bochner integrable H-valued random variable defined on $(\Omega, \mathcal{F}, P)$ and let $\mathcal{J}$ be a sub-σ-field of $\mathcal{F}$. Then there is a unique integrable H-valued random variable Z, measurable with respect to $\mathcal{J}$ such that*

$$\int_A Z dP = \int_A X dP, \qquad \forall\, A \in \mathcal{J}.$$

Z will be denoted by $E(X \mid \mathcal{J})$ and called the conditional expectation of X given $\mathcal{J}$. Moreover

$$|E(X \mid \mathcal{J})|_H \leq E(|X|_H \mid \mathcal{J}).$$

One can defined the general H-valued stochastic process $\{X(t)\}_{t \in I}$, and recall the definition of $L^2_{\mathcal{F}}(0, T; H)$, etc.

If $E|X(t)|_H < \infty$ for each $t \in I$, then this process is called *integrable*. An integrable and adapted H-valued process X is said to be a *martingale* if

$$E(X(t) \mid \mathcal{F}_s) = X(s), \qquad \text{a.s.}$$

for arbitrary $t, s \in I$ with $t \geq s$.

Proposition 4.2. *Let $\{X(t)\}_{t \in I}$ be a martingale. Then the following statements hold.*

i) $\{|X(t)|_H\}_{t \in I}$ is a submartingale.

ii) For any increasing convex function g from $\mathbb{R}^+$ to itself, if $E(g(|X(t)|_H)) < \infty$ for each $t \in I$, then $g(|X(t)|_H)$ is a submartingale.

Fix any $T > 0$ and we denote by $\mathcal{M}^2_T(H)$ the space of all H-valued continuous, square integrable martingales. We need the following result.

Proposition 4.3. *The space $\mathcal{M}^2_T(H)$ equipped with the norm*

$$|M(\cdot)|_{\mathcal{M}^2_T(H)} = \sqrt{E|M(T)|^2}, \qquad \forall\, M(\cdot) \in \mathcal{M}^2_T(H),$$

is a Hilbert space.

Now, let U be another separable Hilbert space, with complete orthonormal basis $\{u_k\}$. A linear bounded operator $G: U \to H$ is said to be *Hilbert-Schmidt* if

$$|G|_2 \triangleq \sum_{k=1}^{\infty} |Gu_k|^2 < \infty. \tag{4.1}$$

It can be shown that this number is independent of the choice of $\{u_k\}$. Moreover, the set $L_2(U; H)$ of all Hilbert-Schmidt operators from U to H, equipped with the norm (4.1), is a separate Hilbert space.

4.2 Hilbert space valued Brownian Motions and stochastic integrals

We consider two separable Hilbert spaces H and U, and a symmetric nonnegative operator $Q \in \mathcal{L}(U)$. We consider first the case when $\operatorname{Tr} Q < \infty$ (which implies Q is compact). Then there is a complete orthonormal basis $\{u_k\}$ in U, and a bounded sequence of nonnegative real numbers λ_k such that

$$Qu_k = \lambda_k u_k, \qquad k = 1, 2, \cdots.$$

Definition 4.4. *A U-valued continuous stochastic process $\{W(t)\}_{t \geq 0}$ is called a Q-Brownian Motion if*
i) $W(0) = 0$;
ii) for each $0 \leq s < t < \infty$, $W(t) - W(s)$ is independent of $\mathcal{F}_s$, and

$$W(t) - W(s) \sim \mathcal{N}(0, (t-s)Q).$$

Proposition 4.5. *Assume that W is a Q-Brownian Motion, with $\operatorname{Tr} Q < \infty$. Then for any t, W has the expansion*

$$W(t) = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \beta_j(t) u_j, \tag{4.2}$$

where

$$\beta_j = \frac{1}{\sqrt{\lambda_j}} (W(t), u_k)_U, \qquad j = 1, 2, \cdots, \tag{4.3}$$

are mutually independent real valued Brownian Motions, and the series in (4.2) is convergent in $L^2_{\mathcal{F}}(\Omega; U)$. Moreover, this series is convergent in $\mathcal{M}^2_T(U)$ for each $T > 0$.

Proof. For any $t > s > 0$, by (4.3) one has

$$E(\beta_i(t)\beta_j(s)) = \tfrac{1}{\sqrt{\lambda_i \lambda_j}} E\big((W(t), u_i)_U (W(s), u_j)_U\big)$$
$$= \tfrac{1}{\sqrt{\lambda_i \lambda_j}} \Big[E\big((W(t) - W(s), u_i)_U (W(s), u_j)_U\big) + E\big((W(s), u_i)_U (W(s), u_j)_U\big) \Big]$$
$$= \tfrac{1}{\sqrt{\lambda_i \lambda_j}} s (Qu_i, u_j)_U = \sqrt{\tfrac{\lambda_i}{\lambda_j}} s \delta_{ij}.$$

Hence the independence of β_i, $i \in \mathbb{N}$, follows. To show (4.2) it is enough to notice that, for $m \geq n \geq 1$,

$$E \Big| \sum_{j=n}^{n} \sqrt{\lambda_j} \beta_j(t) u_j \Big|_U^2 = t \sum_{j=n}^{m} \lambda_j,$$

and recall that $\sum_{j=1}^{\infty} \lambda_j < \infty$. Note that this also implies the last assertion.

□

It is useful to introduce a subspace $U_0 = \mathcal{R}(Q^{1/2})$ of U, which is a Hilbert space with the norm $|\cdot|_0 \triangleq |Q^{-1/2}\cdot|_U$, or more explicitly

$$|u|_0^2 = \sum_{k=1}^{\infty} \frac{1}{\lambda_k}(u, u_k)_U^2.$$

Denote by $L_2^0 \equiv L_2(U_0; H)$ the space of all Hilbert-Schmidt operators from U_0 into H. L_2^0 is a separate Hilbert space, equipped with the norm

$$|\Psi|_{L_2^0}^2 \equiv \mathrm{Tr}\,(\Psi Q \Psi^*).$$

Let $T > 0$ and recall that $L^2_{\mathcal{F}}(0,T;L_2^0)$ is the Hilbert space consisting in all measurable L_2^0-valued processes $\Phi(t)$ adapted to $\{\mathcal{F}_t\}_{t\geq 0}$, such that

$$|\Phi|^2_{L^2_{\mathcal{F}}(0,T;L_2^0)} \triangleq E\int_0^T |\Phi(t)|^2_{L_2^0} dt < \infty.$$

For each $N \in \mathbb{N}$, we define

$$W_N(t) = \sum_{j=1}^{N} \sqrt{\lambda_j}\beta_j(t)u_j. \tag{4.4}$$

Now, for any $\Phi \in L^2_{\mathcal{F}}(0,T;L_2^0)$, we define

$$\begin{aligned}\int_0^t \Phi(s)dW_N(s) &= \sum_{j=1}^{N} \sqrt{\lambda_j}\int_0^t \Phi(s)u_j d\beta_j(t) \\ &= \sum_{k=1}^{\infty}\sum_{j=1}^{N} \sqrt{\lambda_j}h_k \int_0^t (\Phi(s)u_j, h_k)_H d\beta_j(t), \text{ in } \mathcal{M}_T^2(H).\end{aligned} \tag{4.5}$$

It can be shown that $\{\int_0^t \Phi(s)dW_N(s)\}_{N=1}^{\infty}$ is a Cauchy sequence in $\mathcal{M}_T^2(H)$. Hence, we may define the integral of Φ with respect to $dW(t)$ as follows:

$$\int_0^t \Phi(s)dW(s) = \lim_{N\to\infty}\int_0^t \Phi(s)dW_N(s), \quad \text{in } \mathcal{M}_T^2(H). \tag{4.6}$$

Similar to the scalar case, one can define the stochastic integral of $\Phi \in \mathcal{L}^{2,loc}_{\mathcal{F}}(0,T;L_2^0)$ with respect to $dW(t)$.

4.3 Stochastic integrals for cylindrical Brownian Motions

We now extend the definition of the stochastic integral to the case of general bounded, self-adjoint and nonnegative operators Q on U. For simplicity, we assume that $Qx \neq 0$ for $x \neq 0$. Let $U_0 = \mathcal{R}(Q^{1/2})$ with the norm $|\cdot|_0 \triangleq |Q^{-1/2}\cdot|_U$, and let U_1 be any fixed Hilbert space such that U is embedded continuously into U_1 and the embedding of U_0 into U_1 is Hilbert-Schmidt.

Proposition 4.6. *Let $\{g_j\}$ be a complete orthonormal basis in U_0 and $\{\beta_j\}$ be a family of independent real valued standard Brownian Motion. Then*

$$W(t) = \sum_{j=1}^{\infty} \beta_j(t) g_j, \quad t \geq 0, \tag{4.7}$$

defines a Q_1-Brownian Motion on U_1 with $\operatorname{Tr} Q_1 < \infty$. Moreover, $\mathcal{R}(Q_1^{1/2}) = U_0$ and

$$|u|_0 = |Q_1^{-1/2} u|_{U_1}. \tag{4.8}$$

Proof. Since the embedding $J: U_0 \to U_1$ is Hilbert-Schmidt, one has

$$\sum_{j=1}^{\infty} |g_j|_{U_1}^2 = \sum_{j=1}^{\infty} |J g_j|_{U_1}^2 < \infty.$$

Hence, the first assertion follows from

$$E\Big(\Big|\sum_{j=n}^{m} g_j \beta_j(t)\Big|_{U_1}^2\Big) = \sum_{j=n}^{m} |g_j|_{U_1}^2.$$

□

We will call the process $W(t)$ a *cylindrical* Brownian Motion. It is not uniquely determined but the class of integrands $L_{\mathcal{F}}^2(0,T;L_2^0)$ and $L_{\mathcal{F}}^{2,loc}(0,T;L_2^0)$ are independent of the space U_1.

Now, the method of last section, we can define the desired stochastic integrals.

4.4 Properties of the stochastic integral

Theorem 4.7. *Assume that $\Phi \in L_{\mathcal{F}}^2(0,T;L_2^0)$, then the stochastic $\int_0^t \Phi dW \in \mathcal{M}_T^2(H)$, and*

i) $|\int_0^{\cdot} \Phi dW|_{\mathcal{M}_T^2(H)} = |\Phi(\cdot)|_{L_{\mathcal{F}}^2(0,T;L_2^0)}$;

ii) for any $\Phi_1, \Phi_2 \in L_{\mathcal{F}}^2(0,T;L_2^0)$ and $t, s \in [0,T]$,

$$E\Big(\int_0^t \Phi_1 dW, \int_0^s \Phi_2 dW\Big)_H = E\int_0^{t\wedge s} \operatorname{Tr}\,[(\Phi_2(r)Q^{1/2})(\Phi_1(r)Q^{1/2})^*]dr.$$

Now, assume $\Phi \in L_{\mathcal{F}}^{2,loc}(0,T;L_2^0)$, $\phi \in L_{\mathcal{F}}^{1,loc}(0,T;L_2^0)$ and $X(0)$ is a $\mathcal{F}_0$-measurable H-valued random variable. Then the following process

$$X(t) = X(0) + \int_0^t \phi(s)ds + \int_0^t \Phi(s)dW(s), \quad t \in [0,T],$$

is well-defined. Assume a function $F: [0,T] \times H \to \mathbb{R}$ and its partial derivatives F_t, F_x and F_{xx} are uniformly continuous on bounded subsets of $[0,T]\times H$.

Theorem 4.8. *Under the above conditions, P-a.s., for all $t \in [0,T]$,*

$$F(t,X(t)) = F(0,X(0)) + \int_0^t \langle F_x(s,X(s)), \Phi(s)dW(s) \rangle + \int_0^t \Big\{ F_t(s,X(s)) + \langle F_x(s,X(s)), \phi(s) \rangle + \frac{1}{2}\mathrm{Tr}\left[F_{xx}(s,X(s))(\Phi(s)Q^{1/2})(\Phi(s)Q^{1/2})^*\right]\Big\} ds.$$

4.5 The stochastic Fubini theorem

The following stochastic version of the Fubini theorem will be frequently used.

Theorem 4.9. *Let $(G,\mathcal{G},\mu)$ be a measure space and $\Phi: \ (\Omega \times (0,T) \times G, \mathcal{F} \times \mathcal{B}(0,T) \times \mathcal{G}) \to (L_2^0, \mathcal{B}(L_2^0))$ be a measurable mapping. If*

$$\int_G |\Phi(\cdot,\cdot,x)|_{L^2_{\mathcal{F}}(0,T;L_2^0)} \mu(dx) < \infty,$$

then $P-$a.s.,

$$\int_G \Big[\int_0^T \Phi(t,x)dW(t)\Big]\mu(dx) = \int_0^T \Big[\int_G \Phi(t,x)\mu(dx)\Big] dW(t).$$

4.6 Forward stochastic evolution equations

Let H and U be two separable Hilbert spaces and Q a self-adjoint nonnegative operator on U. Let $\{W(t)\}_{t\geq 0}$ be a Q-Brownian Motion on $U_1 \supset U$ and $U_0 = \mathcal{R}(Q^{1/2})$.

In this section, we shall consider the following forward stochastic evolution equation:

$$\begin{cases} dX(t) = [AX(t) + f(t)]dt + BX(t)dW(t), & t \in (0,T], \\ X(0) = \xi, \end{cases} \tag{4.9}$$

where $A: \ D(A) \subset H \to H$ is the generator of a C_0-semigroup $S(\cdot)$, ξ is a $\mathcal{F}_0$-measurable H-valued random variable, $f \in L^{1,loc}_{\mathcal{F}}(0,T;H)$ and $B: \ D(B) \subset H \to L_2^0 = L_2(U_0;H)$ is a linear operator.

Let $\{g_j\}$ be a complete orthonormal basis in U_0. Since for any $x \in D(B)$, $B(x) \in L_2(U_0;H)$, we deduce that

$$\sum_{j=1}^{\infty} |B(x)g_j|^2 < \infty.$$

The operators

$$B_j x = B(x)g_j, \qquad j = 1, 2, \cdots.$$

are linear and

$$B(x)u = \sum_{j=1}^{\infty} B_j x(u, g_j)_{U_0}, \qquad x \in D(B),\ u \in U_0. \tag{4.10}$$

Consequently, if

$$W(t) = \sum_{j=1}^{\infty} \beta_j(t) g_j,$$

then equation (4.9) can be equivalently written as

$$\begin{cases} dX(t) = [AX(t) + f(t)]dt + \sum_{j=1}^{\infty} B_j X(t) d\beta_j(t), \quad t \in (0, T], \\ X(0) = \xi. \end{cases} \tag{4.11}$$

We define a *strong solution* of (4.9) as a H-valued adapted process $X(t)$, which takes values in $D(A) \cap D(B)$, $P-$ a.s., such that

$$P\Big(\int_0^T [|X(t)|_H + |AX(t)|_H]dt < \infty\Big) = 1, \quad P\Big(\int_0^T |BX(t)|^2_{L_2^0} ds < \infty\Big) = 1,$$

and, for any $t \in [0, T]$, and $P-$ a.s.,

$$X(t) = \xi + \int_0^t [AX(s) + f(s)]ds + \int_0^t BX(s)dW(s). \tag{4.12}$$

A H-valued adapted process $\{X(t)\}_{t\in[0,T]}$ is said to be a *weak solution* to (4.9) if it takes values in $D(B)$, $P-$ a.s.,

$$P\Big(\int_0^T |X(s)|_H ds < \infty\Big) = 1, \quad P\Big(\int_0^T |BX(t)|^2_{L_2^0} ds < \infty\Big) = 1, \tag{4.13}$$

and, for any $t \in [0, T]$ and $\zeta \in D(A^*)$,

$$\begin{aligned} (X(t), \zeta)_H = (\xi, \zeta)_H &+ \int_0^t [(X(s), A^*\zeta)_H + (f(s), \zeta)_H]ds \\ &+ \int_0^t (BX(s)dW(s), \zeta)_H, \quad P-\text{a.s.} \end{aligned} \tag{4.14}$$

Finally, A H-valued adapted process $\{X(t)\}_{t\in[0,T]}$ is said to be a *mild solution* to (4.9) if it takes values in $D(B)$, (4.13) holds $P-$ a.s., and for any $t \in [0, T]$,

$$X(t) = S(t)\xi + \int_0^t S(t-s)f(s)ds + \int_0^t S(t-s)BX(s)dW(s). \tag{4.15}$$

The relationship of the above defined solutions are stated as follows.

Theorem 4.10. *Assume that $A: D(A) \subset H \to H$ is the generator of a C_0-semigroup $S(\cdot)$ in H. Then a strong solution of (4.9) is always its weak solution, and a weak solution of (4.9) is its mild solution. Conversely, if X is a mild solution of (4.9) and $BX(\cdot) \in L^2_{\mathcal{F}}(0,T;L^0_2)$, then X is also its weak solution.*

In order to establish the well-posedness of (4.9), we need to analyze the *stochastic convolution*:

$$W_A^{\Phi}(t) = \int_0^t S(t-s)\Phi(s)dW(s), \qquad t \in [0,T],\ \Phi \in L^2_{\mathcal{F}}(0,T;L^0_2). \quad (4.16)$$

Theorem 4.11. *Assume A generates a C_0-semigroup in H and $\Phi \in L^2_{\mathcal{F}}(0,T;L^0_2)$. Then the process $W_A^{\Phi}(t)$ has a continuous and $\{\mathcal{F}_t\}$-adapted modification and there is a constant $C = C(T) > 0$ such that*

$$E\Big(\sup_{s\in[0,t]} |W_A^{\Phi}(t)|^2\Big) \le C|\Phi|^2_{L^2_{\mathcal{F}}(0,T;L^0_2)}, \quad t \in [0,T]. \quad (4.17)$$

In what follows, we will assume B to be bounded.

Theorem 4.12. *Assume A generates a C_0-semigroup in H, $B \in \mathcal{L}(H,L^0_2)$ and $\xi \in L^2_{\mathcal{F}_0}(\Omega;H)$. Then equation (4.9) admits a unique mild solution $X \in L^2_{\mathcal{F}}(\Omega;C([0,T];H))$, identicial with a weak solution.*

Example 4.13. (Stochastic wave equation) Let D be a bounded domain of $\mathbb{R}^n$ with C^2 boundary Γ. Consider the wave equation with Dirichlet boundary conditions:

$$\begin{cases} dy_t = \Delta y dt + a\cdot\nabla y d\beta(t), & (t,x,\omega) \in (0,T)\times D\times\Omega, \\ y = 0, & (t,x,\omega) \in (0,T)\times\Gamma\times\Omega, \\ y(0) = y_0, \quad y_t(0) = y_1, & x \in D, \end{cases} \quad (4.18)$$

where $a \in C^1(\overline{D})$ and $\beta(\cdot)$ is a 1-d standard Brownian Motion.

Now, for any $(y_0,y_1) \in H^1_0(D)\times L^2(D)$, by Theorem 4.11, system (4.18) admits a unique solution $(y,y_t) \in L^2_{\mathcal{F}}(\Omega;C([0,T];H^1_0(D)\times L^2(D)))$.

4.7 Backward stochastic evolution equations

We now consider following backward stochastic evolution equations:

$$\begin{cases} dX(t) = \Big(AX(t) + \sum_{i=1}^m C_i\phi^i\Big)dt + \sum_{i=1}^m \phi^i(t)d\beta_i(t), \\ X(T) = \xi, \end{cases} \quad (4.19)$$

where $\beta = (\beta^1, \cdots, \beta^m)$ are standard Brownian motions defined on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, the filtration $\{\mathcal{F}_t\}$ is generated by β, and A, $C^1, \cdots, C^m$ are linear operators defined on a Hilbert space H.

One can define the strong solution, weak solution and mild solution to (4.19).

We have the following result.

Theorem 4.14. *Assume A generates a C_0-semigroup in H, $C^i \in \mathcal{L}(H)$ for $i = 1, 2, \cdots, m$ and $\xi \in L^2_{\mathcal{F}_T}(\Omega; H)$. Then equation (4.19) admits a unique mild solution $(X, \phi^1, \cdots, \phi^m) \in L^2_{\mathcal{F}}(\Omega; C([0,T]; H)) \times (L^2_{\mathcal{F}}(0,T;H))^m$, identical with a weak solution.*

References

1. Halmos, P. R., *Measure Theory*, D. Van Nostrand Company, Inc., New York, 1950.
2. Karatzas, I., Shreve, S. E., *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, 1988.