Stochastic Calculus

Xi Geng

Contents

1 Review of probability theory

In this section, we review several aspects of probability theory that are important for our study. Most proofs are contained in standard textbooks and hence will be omitted.

Recall that a *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ which consists of a non-empty set Ω, a σ-algebra $\mathcal{F}$ over Ω and a probability measure on $\mathcal{F}$. A *random variable* over $(\Omega, \mathcal{F}, \mathbb{P})$ is a real-valued $\mathcal{F}$-measurable function. For $1 \leqslant p < \infty$, $L^p(\Omega, \mathcal{F}, \mathbb{P})$ denotes the Banach space of (equivalence classes of) random variables X satisfying $\mathbb{E}[|X|^p] < \infty$.

The following are a few conventions that we will be using in the course.

- A $\mathbb{P}$-*null set* is a subset of some $\mathcal{F}$-measurable set with zero probability.
- A property is said to *hold almost surely (a.s.)* or *with probability one* if it holds outside an $\mathcal{F}$-measurable set with zero probability, or equivalently, the set on which it does not hold is a $\mathbb{P}$-null set.

1.1 Conditional expectations

A fundamental concept in the study of martingale theory and stochastic calculus is the conditional expectation.

Definition 1.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Given an integrable random variable $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, the *conditional expectation* of X given $\mathcal{G}$ is the unique $\mathcal{G}$-measurable and integrable random variable Y such that

$$\int_A Y d\mathbb{P} = \int_A X d\mathbb{P}, \ \forall A \in \mathcal{G}. \tag{1.1}$$

It is denoted by $\mathbb{E}[X|\mathcal{G}]$.

The existence of the conditional expectation is a standard application of the Radon-Nikodym theorem, and the uniqueness follows from an easy measure theoretic argument.

Here we recall a geometric construction of the conditional expectation. We start with the Hilbert space $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Since $\mathcal{G} \subseteq \mathcal{F}$, the Hilbert space $L^2(\Omega, \mathcal{G}, \mathbb{P})$ can be regarded as a closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Given $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, let Y be the orthogonal projection of X onto $L^2(\Omega, \mathcal{G}, \mathbb{P})$. Then Y satisfies the characterizing property (1.1) of the conditional expectation. If X is a non-negative integrable random variable, we consider $X_n = X \wedge n \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and let Y_n be the orthogonal projection of X_n onto $L^2(\Omega, \mathcal{G}, \mathbb{P})$. It follows that Y_n is non-negative and increasing. Its pointwise limit, denoted by Y, is a non-negative, $\mathcal{G}$-measurable and integrable random variable which satisfies (1.1). The general case is treated by writing $X = X^+ - X^-$ and using linearity. We left it as an exercise to provide the details of the construction.

The conditional expectation satisfies the following basic properties.

(1) $X \mapsto \mathbb{E}[X|\mathcal{G}]$ is linear.

(2) If $X \leqslant Y$, then $\mathbb{E}[X|\mathcal{G}] \leqslant \mathbb{E}[Y|\mathcal{G}]$. In particular, $|\mathbb{E}[X|\mathcal{G}]| \leqslant \mathbb{E}[|X||\mathcal{G}]$.

(3) If X and ZX are both integrable, and $Z \in \mathcal{G}$, then $\mathbb{E}[ZX|\mathcal{G}] = Z\mathbb{E}[X|\mathcal{G}]$.

(4) If $\mathcal{G}_1 \subset \mathcal{G}_2$ are sub-σ-algebras of $\mathcal{F}$, then $\mathbb{E}[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1] = \mathbb{E}[X|\mathcal{G}_1]$.

(5) If X and $\mathcal{G}$ are independent, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$.

In addition, we have the following Jensen's inequality: if φ is a convex function on $\mathbb{R}$, and both X and $\varphi(X)$ are integrable, then

$$\varphi(\mathbb{E}[X|\mathcal{G}]) \leqslant \mathbb{E}[\varphi(X)|\mathcal{G}]. \tag{1.2}$$

Applying this to the function $\varphi(x) = |x|^p$ for $p \geqslant 1$, we see immediately that the conditional expectation is a contraction operator on $L^p(\Omega, \mathcal{F}, \mathbb{P})$.

The convergence theorems (the monotone convergence theorem, Fatou's lemma, and the dominated convergence theorem) also hold for the conditional expectation, stated in an obvious way.

For every measurable subset $A \in \mathcal{F}$, $\mathbb{P}(A|\mathcal{G})$ is the conditional probability of A given $\mathcal{G}$. However, $\mathbb{P}(A|\mathcal{G})$ is defined up to a null set which depends on A, and in general there does not exist a universal null set outside which the conditional probability $A \mapsto \mathbb{P}(A|\mathcal{G})$ can be regarded as a probability measure. The resolution of this issue leads to the notion of regular conditional expectations, which plays an important role in the study of Markov processes and stochastic differential equations.

Definition 1.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. A system $\{p(\omega, A)\}_{\omega\in\Omega, A\in\mathcal{F}}$ is called a *regular conditional probability* given $\mathcal{G}$ if it satisfies the following conditions:

(1) for every $\omega \in \Omega$, $A \mapsto p(\omega, A)$ is a probability measure on $(\Omega, \mathcal{F})$;

(2) for every $A \in \mathcal{F}$, $\omega \mapsto p(\omega, A)$ is $\mathcal{G}$-measurable;

(3) for every $A \in \mathcal{F}$ and $B \in \mathcal{G}$,

$$\mathbb{P}(A \bigcap B) = \int_B p(\omega, A)\mathbb{P}(d\omega).$$

The third condition tells us that for every $A \in \mathcal{F}$, $p(\cdot, A)$ is a version of $\mathbb{P}(A|\mathcal{G})$. It follows that for every integrable random variable X, $\omega \mapsto \int X(\omega')p(\omega, d\omega')$ is an almost surely well-defined and it is a version of $\mathbb{E}[X|\mathcal{G}]$.

In many situations, we are interested in the conditional distribution of a random variable taking values in another measurable space. Suppose that $\{p(\omega, A)\}_{\omega\in\Omega, A\in\mathcal{F}}$ is a regular conditional probability on $(\Omega, \mathcal{F}, \mathbb{P})$ given $\mathcal{G}$. Let X be a measurable map from $(\Omega, \mathcal{F})$ to some measurable space $(E, \mathcal{E})$. We can define

$$Q(\omega, \Gamma) = p(\omega, X^{-1}\Gamma), \ \omega \in \Omega, \Gamma \in \mathcal{E}.$$

Then the system $\{Q(\omega, \Gamma)\}_{\omega\in\Omega, \Gamma\in\mathcal{E}}$ satisfies:

(1)' for every $\omega \in \Omega$, $\Gamma \mapsto Q(\omega, \Gamma)$ is a probability measure on $(E, \mathcal{E})$;

(2)' for every $\Gamma \in \mathcal{E}$, $\omega \mapsto Q(\omega, \Gamma)$ is $\mathcal{G}$-measurable;

(3)' for every $\Gamma \in \mathcal{E}$ and $B \in \mathcal{G}$,

$$\mathbb{P}(\{X \in \Gamma\} \bigcap B) = \int_B Q(\omega, \Gamma)\mathbb{P}(d\omega).$$

In particular, we can see that $Q(\cdot,\Gamma)$ is a version of $\mathbb{P}(X\in\Gamma|\mathcal{G})$ for every $\Gamma\in\mathcal{E}$. The system $\{Q(\omega,\Gamma)\}_{\omega\in\Omega,\Gamma\in\mathcal{E}}$ is called a *regular conditional distribution* of X given $\mathcal{G}$.

It is a deep result in measure theory that if E is a complete and separable metric space, and $\mathcal{E}$ is the σ-algebra generated by open sets in E, then a regular conditional distribution of X given $\mathcal{G}$ exists. In particular, if $(\Omega,\mathcal{F})$ is a complete and separable metric space, by considering the identity map we know that a regular conditional probability given $\mathcal{G}$ exists. In this course we will mainly be interested in complete and separable metric spaces.

Sometimes we also consider conditional expectations given some random variable X. Let X be as before, and let $\mathbb{P}^X$ be the law of X on $(E,\mathcal{E})$. Similar to Definition 1.2, a system $\{p(x,A)\}_{x\in E,A\in\mathcal{F}}$ is called a *regular conditional probability* given X if it satisfies:

(1)" for every $x\in E$, $A\mapsto p(x,A)$ is a probability measure on $(\Omega,\mathcal{F})$;

(2)" for every $A\in\mathcal{F}$, $x\mapsto p(x,A)$ is $\mathcal{E}$-measurable;

(3)" for every $A\in\mathcal{F}$ and $\Gamma\in\mathcal{E}$,

$$\mathbb{P}(A\bigcap\{X\in\Gamma\})=\int_\Gamma p(x,A)\mathbb{P}^X(dx).$$

In particular, $p(\cdot,A)$ gives a version of $\mathbb{P}(A|X=\cdot)$. If $(\Omega,\mathcal{F})$ is a complete and separable metric space, then a regular conditional probability given X exists.

1.2 Uniform integrability

Now we review an important concept which is closely related to the study of L^1-convergence.

Definition 1.3. A family $\{X_t:\ t\in T\}$ of integrable random variables over a probability space $(\Omega,\mathcal{F},\mathbb{P})$ is called *uniformly integrable* if

$$\lim_{\lambda\to\infty}\sup_{t\in T}\int_{\{|X_t|>\lambda\}}|X_t|d\mathbb{P}=0.$$

Uniform integrability can be characterized by the following two properties.

Theorem 1.1. *Let $\{X_t:\ t\in T\}$ be a family of integrable random variables. Then $\{X_t:\ t\in T\}$ is uniformly integrable if and only if*

(1) (uniform boundedness in L^1) there exists $M>0$, such that

$$\int_\Omega |X_t|d\mathbb{P}\leqslant M,\ \forall t\in T;$$

(2) (uniform equicontinuity) for every $\varepsilon>0$, there exists $\delta>0$, such that for all $A\in\mathcal{F}$ with $\mathbb{P}(A)<\delta$ and $t\in T$,

$$\int_A |X_t|d\mathbb{P}<\varepsilon.$$

The two characterizing properties in Theorem 1.1 might remind us the Arzelà–Ascoli theorem (in functional analysis) for continuous functions (c.f. Theorem 1.9). Therefore, it is not unreasonable to expect that uniform integrability is equivalent to some kind of relative compactness in $L^1(\Omega, \mathcal{F}, \mathbb{P})$. This is an important result due to Dunford and Pettis.

Definition 1.4. A sequence $\{X_n\}$ of integrable random variables is said to *converge weakly in* L^1 to an integrable random variable X if for every bounded random variable Y, we have

$$\lim_{n\to\infty} \mathbb{E}[X_n Y] = \mathbb{E}[XY].$$

Theorem 1.2. *A family* $\{X_t : t \in T\}$ *of integrable random variables is uniformly integrable if and only if every sequence in* $\{X_t : t \in T\}$ *contains subsequence which converges weakly in* L^1.

Perhaps the most important property of uniform integrability for our study lies in its connection with L^1-convergence.

Theorem 1.3. *Let* $\{X_n\}$ *be a sequence of integrable random variables and let* X *be another random variable. Then the following two statements are equivalent:*

(1) X_n *converges to* X *in* L^1, *in the sense that*

$$\lim_{n\to\infty} \int |X_n - X| d\mathbb{P} = 0;$$

(2) X_n *converges to* X *in probability, in the sense that*

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$

for every $\varepsilon > 0$, *and* $\{X_n\}$ *is uniformly integrable.*

1.3 The Borel-Cantelli lemma

Now we review a simple technique which has huge applications in probability theory and stochastic processes.

Theorem 1.4. *Let* $\{A_n\}$ *be a sequence of events over some probability space* $(\Omega, \mathcal{F}, \mathbb{P})$.

(1) If $\sum_n \mathbb{P}(A_n) < \infty$, *then*

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0.$$

(2) Suppose further that $\{A_n\}$ *are independent. If* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, *then*

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1.$$

1.4 The law of large numbers and the central limit theorem

The study of limiting behaviors for random sequences is an important topic in probability theory. Here we review two classical limit theorems for sequences of independent random variables: the law of large numbers and the central limit theorem. Heuristically, given a sequence of independent random variables satisfying certain moment conditions, the (strong) law of large numbers describes the property that the sample average will eventually stabilize at the expected value, while the central limit theorem quantifies the asymptotic distribution of the stochastic fluctuation of the sample average around the expected value. Here we do not pursue the most general cases and we only state the results in a special setting which are already important on its own and relevant for our study.

Definition 1.5. Let X_n, X be random variables with distribution function $F_n(x), F(x)$ respectively. X_n is said to *converge in distribution* to X if

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for every x at which $F(x)$ is continuous.

Note that convergence in distribution is a property that only refers to distribution functions rather than underlying random variables.

Theorem 1.5. *Let $\{X_n\}$ be a sequence of independent and identically distributed random variables with $\mu = \mathbb{E}[X_1]$ and $\sigma^2 = \mathrm{Var}[X_1] < \infty$. Let $s_n = (X_1 + \cdots + X_n)/n$ be the sample average. Then with probability one,*

$$\lim_{n\to\infty} s_n = \mu.$$

Moreover, the normalized sequence $\sqrt{n}(s_n - \mu)/\sigma$ converges in distribution to the standard normal distribution $\mathcal{N}(0,1)$.

1.5 Weak convergence of probability measures

Finally, we discuss an important notion of convergence for probability measures: weak convergence. This is particularly useful in the infinite dimensional setting, for instance in studying the distributions of stochastic processes, which are probability measures on the space of paths.

Let (S, ρ) be a metric space. The *Borel σ-algebra $\mathcal{B}(S)$* over S is the σ-algebra generated by open sets in S. We use $C_b(S)$ to denote the space of bounded continuous functions on S.

Definition 1.6. Let $\mathbb{P}_n, \mathbb{P}$ be probability measures on $(S, \mathcal{B}(S))$. $\mathbb{P}_n$ is said to *converge weakly* to $\mathbb{P}$ if

$$\lim_{n\to\infty} \int_S f(x)\mathbb{P}_n(dx) = \int_S f(x)\mathbb{P}(dx), \ \forall f \in C_b(S).$$

Before the general discussion of weak convergence, let us say a bit more in the case when $S = \mathbb{R}^1$.

Definition 1.7. Let $\mathbb{P}$ be a probability measure on $(\mathbb{R}^1, \mathcal{B}(\mathbb{R}^1))$. The *characteristic function* of $\mathbb{P}$ is the complex-valued function given by

$$f(t) = \int_{\mathbb{R}^1} \mathrm{e}^{itx}\mathbb{P}(dx), \ t \in \mathbb{R}^1.$$

There are nice regularity properties for characteristic functions. For instance, it is uniformly continuous on $\mathbb{R}^1$ and uniformly bounded by 1. The uniqueness theorem for characteristic functions asserts that two probability measures on $(\mathbb{R}^1, \mathcal{B}(\mathbb{R}^1))$ are identical if and only if they have the same characteristic functions. Moreover, there is a one-to-one correspondence between probability measures on $(\mathbb{R}^1, \mathcal{B}(\mathbb{R}^1))$ and distribution functions (i.e. right continuous and increasing functions $F(x)$ with $F(-\infty) = 0$ and $F(\infty) = 1$) through the Lebesgue-Stieltjes construction.

The characteristic function is also a useful concept in studying weak convergence properties. The following result characterizes weak convergence for probability measures on $(\mathbb{R}^1, \mathcal{B}(\mathbb{R}^1))$.

Theorem 1.6. *Let $\mathbb{P}_n, \mathbb{P}$ be probability measures on $(\mathbb{R}^1, \mathcal{B}(\mathbb{R}^1))$ with distribution functions $F_n(x), F(x)$ and characteristic functions $f_n(t), f(t)$ respectively. Then the following statements are equivalent:*

(1) $\mathbb{P}_n$ converges weakly to $\mathbb{P}$;

(2) F_n converges in distribution to F;

(3) f_n converges to f pointwisely on $\mathbb{R}^1$;

Remark 1.1. When we study the distribution of a non-negative random variable T (for instance a random time), for technical convenience we usually consider the Laplace transform $\lambda > 0 \mapsto \mathbb{E}\left[\mathrm{e}^{-\lambda T}\right]$ instead of the characteristic function, which also characterizes the distribution of T.

Remark 1.2. The notion of characteristic functions extends to the multidimensional case. The previous results about the connections between characteristic functions and probability measures still hold, except for the fact that the notion of distribution functions is no longer natural–they are not in one-to-one correspondence with probability measures.

Now we come back to the general situation. The notion of characteristic functions is not well-defined on general metric spaces. However, we still have following general characterization of weak convergence. Although the proof is standard, we provide it here to help the reader get comfortable with the notions.

Theorem 1.7. *Let (S, ρ) be a metric space and let $\mathbb{P}_n, \mathbb{P}$ be probability measures on $(S, \mathcal{B}(S))$. Then the following results are equivalent:*

(1) $\mathbb{P}_n$ converges weakly to $\mathbb{P}$;

(2) for every $f \in C_b(S)$ which is uniformly continuous,

$$\lim_{n\to\infty} \int_S f(x)\mathbb{P}_n(dx) = \int_S f(x)\mathbb{P}(dx);$$

(3) for every closed subset $F \subseteq S$,

$$\limsup_{n\to\infty} \mathbb{P}_n(F) \leqslant \mathbb{P}(F);$$

(4) for every open subset $G \subseteq S$,

$$\liminf_{n\to\infty} \mathbb{P}_n(G) \geqslant \mathbb{P}(G);$$

(5) for every $A \in \mathcal{B}(S)$ satisfying $\mathbb{P}(\partial A) = 0$ where $\partial A \triangleq \overline{A}\backslash \mathring{A}$ is the boundary of A,

$$\lim_{n\to\infty} \mathbb{P}_n(A) = \mathbb{P}(A).$$

Proof. (1) $\Longrightarrow$ (2) is obvious.

(2) $\Longrightarrow$ (3). Let F be a closed subset of S. For $k \geqslant 1$, define

$$f_k(x) = \left(\frac{1}{1+\rho(x,F)}\right)^k, \ x \in S,$$

where $\rho(x, F)$ is the distance between x and F. It is easy to see that f_k is bounded and uniformly continuous. In particular,

$$\mathbf{1}_F(x) \leqslant f_k(x) \leqslant 1,$$

and $f_k \downarrow \mathbf{1}_F$ as $k \to \infty$, where $\mathbf{1}_F$ denotes the indicator function of F. Therefore, from (2) we have

$$\begin{aligned} \limsup_{n\to\infty} \mathbb{P}_n(F) &\leqslant \lim_{n\to\infty} \int_S f_k(x)\mathbb{P}_n(dx) \\ &= \int_S f_k(x)\mathbb{P}(dx) \end{aligned}$$

for every $k \geqslant 1$. From the dominated convergence theorem, by letting $k \to \infty$ we conclude that

$$\limsup_{n\to\infty} \mathbb{P}_n(F) \leqslant \mathbb{P}(F).$$

(3)$\Longleftrightarrow$(4) is obvious.

(3)+(4) $\Longrightarrow$ (5). Let $A \in \mathcal{B}(S)$ be such that $\mathbb{P}(\partial A) = 0$. It follows that

$$\mathbb{P}(\mathring{A}) = \mathbb{P}(A) = \mathbb{P}(\overline{A}).$$

From (3) and (4), we see that

$$\begin{aligned} \limsup_{n\to\infty} \mathbb{P}_n(A) &\leqslant \limsup_{n\to\infty} \mathbb{P}_n(\overline{A}) \\ &\leqslant \mathbb{P}(\overline{A}) = \mathbb{P}(A) = \mathbb{P}(\mathring{A}) \\ &\leqslant \liminf_{n\to\infty} \mathbb{P}_n(\mathring{A}) \\ &\leqslant \liminf_{n\to\infty} \mathbb{P}_n(A). \end{aligned}$$

Therefore, $\lim_{n\to\infty}\mathbb{P}_n(A)=\mathbb{P}(A)$.

(5) $\implies$ (1). Let f be a bounded continuous function on S. By translation and rescaling we may assume that $0<f<1$. Since $\mathbb{P}$ is a probability measure, we know that for each $n\geqslant 1$, the set $\{a\in\mathbb{R}^1:\ \mathbb{P}(f=a)\geqslant 1/n\}$ is finite. Therefore, the set

$$\{a\in\mathbb{R}^1:\ \mathbb{P}(f=a)>0\}$$

is at most countable. Given $k\geqslant 1$, for each $1\leqslant i\leqslant k$, we choose some $a_i\in((i-1)/k,i/k)$ such that $\mathbb{P}(f=a_i)=0$. Set $a_0=0$, $a_{k+1}=1$, and define $B_i=\{a_{i-1}\leqslant f<a_i\}$ for $1\leqslant i\leqslant k+1$. Note that $|a_i-a_{i-1}|<2/k$, and B_i are disjoint whose union is S. Moreover, from the continuity of f it is easy to see that

$$\overline{B_i}\subseteq\{a_{i-1}\leqslant f\leqslant a_i\},\ \{a_{i-1}<f<a_i\}\subseteq\mathring{B}_i.$$

Therefore, $\partial B_i\subseteq\{f=a_{i-1}\}\cup\{f=a_i\}$ and $\mathbb{P}(\partial B_i)=0$. It follows that

$$\begin{aligned}\left|\int_S f(x)\mathbb{P}_n(dx)-\int_S f(x)\mathbb{P}(dx)\right| &\leqslant \sum_{i=1}^{k+1}\left|\int_{B_i} f(x)\mathbb{P}_n(dx)-\int_{B_i} f(x)\mathbb{P}(dx)\right|\\ &\leqslant \frac{4}{k}+\sum_{i=1}^{k+1}a_{i-1}\left|\mathbb{P}_n(B_i)-\mathbb{P}(B_i)\right|.\end{aligned}$$

By letting $n\to\infty$, from (5) we conclude that

$$\limsup_{n\to\infty}\left|\int_S f(x)\mathbb{P}_n(dx)-\int_S f(x)\mathbb{P}(dx)\right|\leqslant\frac{4}{k}.$$

Now the result follows as k is arbitrary. □

Now we introduce an important characterization of relative compactness for a family of probability measures with respect to the topology of weak convergence. This is known as Prokhorov's theorem. The usefulness of relative compactness in proving weak convergence is demonstrated in Problem Sheet 2, Problem 1.

Definition 1.8. A family $\mathcal{P}$ of probability measures on a metric space $(S,\mathcal{B}(S),\rho)$ is said to be *tight* if for every $\varepsilon>0$, there exists a compact subset $K\subseteq S$, such that

$$\mathbb{P}(K)>1-\varepsilon,\ \forall\mathbb{P}\in\mathcal{P}.$$

Prokhorov's theorem relates tightness and relative compactness with respect to the topology of weak convergence.

Theorem 1.8. *Let $\mathcal{P}$ be a family of probability measures on a separable metric space $(S,\mathcal{B}(S),\rho)$.*

(1) If $\mathcal{P}$ is tight, then it is relatively compact, in the sense that every subsequence of $\mathcal{P}$ further contains a weakly convergent subsequence.

(2) Suppose in addition that (S,ρ) is complete. If $\mathcal{P}$ is relatively compact, then it is also tight.

Remark 1.3. In the language of general topology, we do not distinguish the meanings between relative compactness and sequential compactness because it is known that the topology of weak convergence is metrizable (i.e. there exists a metric d on the space of probability measures on $(S, \mathcal{B}(S))$, such that $\mathbb{P}_n$ converges weakly to $\mathbb{P}$ if and only if $d(\mathbb{P}_n, \mathbb{P}) \to 0$).

Now we study an example which plays a fundamental role in our study.

Let W^d be the space of continuous paths $w: [0,\infty) \to \mathbb{R}^d$. We define a metric ρ on W^d by

$$\rho(w, w') = \sum_{n=1}^{\infty} \frac{1 \wedge \max_{t\in[0,n]} |w_t - w'_t|}{2^n}, \quad w, w' \in W^d. \tag{1.3}$$

Therefore, ρ characterizes uniform convergence on compact intervals. It is a good exercise to show that (W^d, ρ) is a complete and separable metric space, and the Borel σ-algebra over W^d coincides with the σ-algebra generated by *cylinder sets* of the form

$$\{w \in W^d : (w_{t_1}, \cdots, w_{t_n}) \in \Gamma\}$$

for $n \in \mathbb{N}$, $0 \leqslant t_1 < \cdots < t_n$ and $\Gamma \in \mathcal{B}(\mathbb{R}^{d\times n})$.

W^d is usually known as the *(continuous) path space* over $\mathbb{R}^d$. It is important as every continuous stochastic process can be realized on W^d. Moreover, when equipped with the canonical Wiener measure (the distribution of Brownian motion), W^d carries nice analytic structure on which the Malliavin calculus (a theory of stochastic calculus of variations in infinite dimensions which constitutes a substantial part of modern stochastic analysis) is built.

We finish by proving an important criteria for tightness of probability measures on W^d. This is a simple probabilistic analogue of the Arzelà–Ascoli theorem, which is recaptured in the following. We use $\Delta(\delta, n; w)$ to denote the modulus of continuity of $w \in W^d$ over $[0, n]$, i.e.

$$\Delta(\delta, n; w) = \sup_{\substack{s,t\in[0,n] \\ |s-t|<\delta}} |w_s - w_t|, \quad \delta > 0, n \in \mathbb{N}, w \in W^d.$$

Theorem 1.9. *A subset $\Lambda \subseteq (W^d, \rho)$ is relatively compact (i.e. $\overline{\Lambda}$ is compact) if and only if the following two conditions hold:*

(1) uniform boundedness:

$$\sup\{|w_0| : w \in \Lambda\} < \infty;$$

(2) uniform equicontinuity: for every $n \in \mathbb{N}$,

$$\lim_{\delta\downarrow 0} \sup_{w\in\Lambda} \Delta(\delta, n; w) = 0.$$

Now we have the following result.

Theorem 1.10. *Let $\mathcal{P}$ be a family of probability measures on $(W^d, \mathcal{B}(W^d))$. Suppose that the following two conditions hold:*

(1)

$$\lim_{a\to\infty}\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{P}(|w_0|>a)=0;$$

(2) for every $\varepsilon>0$ and $n\in\mathbb{N}$,

$$\lim_{\delta\downarrow 0}\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{P}(\Delta(\delta,n;w)>\varepsilon)=0.$$

Then $\mathcal{P}$ is tight.

Proof. Fix $\varepsilon>0$. Condition (1) implies that there exists $a_\varepsilon>0$ such that

$$\mathbb{P}(|w_0|>a_\varepsilon)<\frac{\varepsilon}{2},\ \forall\mathbb{P}\in\mathcal{P}.$$

In addition, Condition (2) implies that there exists a sequence $\delta_{\varepsilon,n}\downarrow 0$ (as $n\to\infty$) such that

$$\mathbb{P}\left(\Delta(\delta_{\varepsilon,n},n;w)>\frac{1}{n}\right)<\varepsilon\cdot 2^{-(n+1)},\ \forall\mathbb{P}\in\mathcal{P}\text{ and }n\in\mathbb{N}.$$

Let

$$\Lambda_\varepsilon=\{|w_0|\leqslant a_\varepsilon\}\bigcap\bigcap_{n=1}^{\infty}\left\{\Delta(\delta_{\varepsilon,n},n;w)\leqslant\frac{1}{n}\right\}\subseteq W^d.$$

Then

$$\mathbb{P}(\Lambda_\varepsilon^c)<\frac{\varepsilon}{2}+\sum_{n=1}^{\infty}\varepsilon\cdot 2^{-(n+1)}=\varepsilon,\ \forall\mathbb{P}\in\mathcal{P}.$$

Moreover, it is straight forward that Λ_ε satisfies the two conditions in Arzelà–Ascoli's theorem. Therefore, Λ_ε is a relatively compact subset of W^d, and

$$\mathbb{P}\left(\overline{\Lambda_\varepsilon}\right)\geqslant\mathbb{P}(\Lambda_\varepsilon)>1-\varepsilon,\ \forall\mathbb{P}\in\mathcal{P}.$$

In other words, we conclude that $\mathcal{P}$ is tight. □

2 Generalities on continuous time stochastic processes

In this section, we study the basic notions of stochastic processes. The core concepts are filtrations and stopping times. These notions enable us to keep track of information evolving in time in a mathematical way. This is an important feature of stochastic calculus which is quite different from ordinary calculus.

2.1 Basic definitions

A stochastic process models the evolution of a random system. In this course, we will be studying the differential calculus with respect to certain important (continuous) stochastic processes.

Definition 2.1. A *(d-dimensional) stochastic process* on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a collection $\{X_t\}$ of $\mathbb{R}^d$-valued random variables indexed by $[0, \infty)$.

Because of the index set being $[0, \infty)$, t is usually interpreted as the time parameter. From the definition, we know that a stochastic process is a map

$$\begin{array}{rcl} X: \ [0,\infty) \times \Omega & \to & \mathbb{R}^d, \\ (t, \omega) & \mapsto & X_t(\omega), \end{array}$$

such that for every fixed t, as a function in $\omega \in \Omega$ it is $\mathcal{F}$-measurable. There is yet another way of looking at a stochastic process which is more important and fundamental: for every $\omega \in \Omega$, it gives a path in $\mathbb{R}^d$. More precisely, let $(\mathbb{R}^d)^{[0,\infty)}$ be the space of functions $w: \ [0, \infty) \to \mathbb{R}^d$, with Borel σ-algebra $\mathcal{B}\left((\mathbb{R}^d)^{[0,\infty)}\right)$ defined by the σ-algebra generated by cylinder sets of the form

$$\left\{ w \in (\mathbb{R}^d)^{[0,\infty)} : \ (w_{t_1}, \cdots, w_{t_n}) \in \Gamma \right\}$$

for $n \in \mathbb{N}$, $0 \leqslant t_1 < \cdots < t_n$ and $\Gamma \in \mathcal{B}(\mathbb{R}^{d \times n})$. Then the definition of a stochastic process is equivalent to a measurable map

$$X: \ (\Omega, \mathcal{F}) \to \left((\mathbb{R}^d)^{[0,\infty)}, \mathcal{B}\left((\mathbb{R}^d)^{[0,\infty)} \right) \right).$$

For every $\omega \in \Omega$, the path $X(\omega)$ is called a *sample path* of the stochastic process.

Remark 2.1. The path space $(\mathbb{R}^d)^{[0,\infty)}$ is different from the space W^d we introduced in the last section as we do not impose any regularity conditions on sample paths here. In fact it can be shown that W^d is not even a $\mathcal{B}\left((\mathbb{R}^d)^{[0,\infty)}\right)$-measurable subset of $(\mathbb{R}^d)^{[0,\infty)}$. However, if we assume that every sample path of X is continuous, then X descends to a measurable map from $(\Omega, \mathcal{F})$ to $(W^d, \mathcal{B}(W^d))$.

For technical reasons, in particular for the purpose of integration, we often require joint measurability properties on a stochastic process.

Definition 2.2. A stochastic process X is called *measurable* if it is jointly measurable in (t, ω), i.e. if the map

$$\begin{array}{rcl} X: \ [0,\infty)\times\Omega & \to & \mathbb{R}^d, \\ (t,\omega) & \mapsto & X_t(\omega), \end{array}$$

is $\mathcal{B}([0,\infty))\otimes\mathcal{F}$-measurable.

Nice consequences of measurability are: every sample path is $\mathcal{B}([0,\infty))$-measurable and Fubini's theorem is applicable to X when $([0,\infty),\mathcal{B}([0,\infty)))$ is equipped with a measure.

Another very important reason of introducing measurability is, when evaluated at a random time we always obtain a random variable. To be more precise, if X is a measurable process and τ is a finite random time (i.e. $\tau: \ \Omega\to[0,\infty)$ is $\mathcal{F}$-measurable), then $\omega\mapsto X_{\tau(\omega)}(\omega)$ is an $\mathcal{F}$-measurable random variable. This can be seen easily from the following composition of maps:

$$\begin{array}{rcccl} X_\tau: \ (\Omega,\mathcal{F}) & \to & ([0,\infty)\times\Omega,\mathcal{B}([0,\infty))\otimes\mathcal{F}) & \to & (\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d)), \\ \omega & \mapsto & (\tau(\omega),\omega) & \mapsto & X_{\tau(\omega)}(\omega). \end{array}$$

Stopping a process at a random time is a very useful notion in our study.

Sometimes we need to compare different stochastic processes in certain probabilistic sense.

Definition 2.3. Let X_t, Y_t be two stochastic processes defined on some probability space $(\Omega,\mathcal{F},\mathbb{P})$. We say:

(1) X_t and Y_t are *indistinguishable* if $X(\omega)=Y(\omega)$ a.s.;

(2) Y_t is a *modification* of X_t if for every $t\geqslant 0$, $\mathbb{P}(X_t=Y_t)=1$;

(3) X_t and Y_t have *the same finite dimensional distributions* if

$$\mathbb{P}((X_{t_1},\cdots,X_{t_n})\in\Gamma)=\mathbb{P}((Y_{t_1},\cdots,Y_{t_n})\in\Gamma)$$

for any $n\in\mathbb{N}, 0\leqslant t_1<\cdots<t_n$ and $\Gamma\in\mathbb{R}^{d\times n}$.

Apparently (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3), but none of the reverse directions is true. If X_t and Y_t have right continuous sample paths, then $(1)\Longleftrightarrow(2)$. Moreover, to make sense of (3), X_t and Y_t do not have to be defined on the same probability space.

In many situations, we are interested in infinite dimensional probabilistic properties rather than finite dimensional distributions.

Definition 2.4. The *distribution* of a stochastic process X_t is the probability measure $\mathbb{P}^X=\mathbb{P}\circ X^{-1}$ on $\left((\mathbb{R}^d)^{[0,\infty)},\mathcal{B}\left((\mathbb{R}^d)^{[0,\infty)}\right)\right)$ induced by X.

As in Remark 2.1, if X has continuous sample paths, X also induces a probability measure μ^X on $(W^d,\mathcal{B}(W^d))$. When concerning finite dimensional distribution properties, we do not have to distinguish between $\mathbb{P}^X$ and μ^X. However, it is much more convenient to use μ^X than $\mathbb{P}^X$ for studying infinite dimensional distribution properties, as $\mathcal{B}((\mathbb{R}^d)^{[0,\infty)})$ is too small to contain adequate interesting events, for instance an event like $\{w: \sup_{0\leqslant t\leqslant 1}|w_t|\leqslant 1\}$. The view of realizing a continuous stochastic process on $(W^d,\mathcal{B}(W^d),\mu^X)$ is rather important in stochastic analysis.

2.2 Construction of stochastic processes: Kolmogorov's extension theorem

The first question in the study of stochastic processes is their existence. In particular, is it possible to construct a stochastic process in a canonical way from the knowledge of its finite dimensional distributions? The answer is the content of Kolmogorov's extension theorem.

We first recapture the notion of finite dimensional distributions in a more general context.

Let X_t be a stochastic process taking values in some metric space S. We use $\mathcal{T}$ to denote the set of finite sequences $\mathfrak{t} = (t_1, \cdots, t_n)$ of distinct times on $[0, \infty)$ (they need not be ordered in an increasing manner). For each $\mathfrak{t} = (t_1, \cdots, t_n) \in \mathcal{T}$, we can define a probability measure $Q_\mathfrak{t}$ on $(S^n, \mathcal{B}(S^n))$ by

$$Q_\mathfrak{t}(\Gamma) = \mathbb{P}\left((X_{t_1}, \cdots, X_{t_n}) \in \Gamma\right), \ \Gamma \in \mathcal{B}(S^n).$$

The family $\{Q_\mathfrak{t} : \ t \in \mathcal{T}\}$ of probability measures defines the finite dimensional distributions of $\{X_t\}$. It is straight forward to see that it satisfies the following two consistency properties:

(1) let $\mathfrak{t} = (t_1, \cdots, t_n)$ and $A_1, \cdots, A_n \in \mathcal{B}(S)$, then for every permutation σ of order n,

$$Q_\mathfrak{t}(A_1 \times \cdots \times A_n) = Q_{\sigma(\mathfrak{t})}\left(A_{\sigma(1)} \times \cdots \times A_{\sigma(n)}\right),$$

where $\sigma(\mathfrak{t}) = \left(t_{\sigma(1)}, \cdots, t_{\sigma(n)}\right)$;

(2) let $\mathfrak{t} = (t_1, \cdots, t_n)$ and $\mathfrak{t}' = (t_1, \cdots, t_n, t_{n+1})$, then for every $A \in \mathcal{B}(S^n)$,

$$Q_{\mathfrak{t}'}(A \times S) = Q_\mathfrak{t}(A).$$

Definition 2.5. A family $\{Q_\mathfrak{t} : \ \mathfrak{t} \in \mathcal{T}\}$ of finite dimensional distributions is said to be *consistent* if it satisfies the previous two properties.

We are mainly interested in the reverse direction: is it possible to construct a stochastic process in a canonical way whose finite dimensional distributions coincide with a given consistent family of probability measures? The answer is yes, and the construction is made through a classical measure theoretic argument.

Recall that $S^{[0,\infty)}$ is the space of functions $w : \ [0, \infty) \to S$ and $\mathcal{B}\left(S^{[0,\infty)}\right)$ is the σ-algebra generated by cylinder sets. Then we have the following result.

Theorem 2.1. *Let S be a complete and separable metric space. Suppose that $\{Q_\mathfrak{t} : \ \mathfrak{t} \in \mathcal{T}\}$ is a consistent family of finite dimensional distributions. Then there exists a unique probability measure $\mathbb{P}$ on $\left(S^{[0,\infty)}, \mathcal{B}\left(S^{[0,\infty)}\right)\right)$, such that*

$$\mathbb{P}((w_{t_1}, \cdots, w_{t_n}) \in \Gamma) = Q_\mathfrak{t}(\Gamma)$$

for every $\mathfrak{t} = (t_1, \cdots, t_n) \in \mathcal{T}$ and $\Gamma \in \mathcal{B}(S^n)$.

We prove Theorem 2.1 by using Carathéodory's extension theorem in measure theory, and we proceed in several steps.

(1) Let $\mathcal{C}$ be the family of subsets of $S^{[0,\infty)}$ of the form $\{(w_{t_1},\cdots,w_{t_n})\in\Gamma\}$, where $\mathfrak{t}=(t_1,\cdots,t_n)\in\mathcal{T}$ and $\Gamma\in\mathcal{B}(S^n)$. It is straight forward to see that $\mathcal{C}$ is an algebra (i.e. $\emptyset, S^{[0,\infty)}\in\mathcal{C}$ and it is closed under taking complement or finite intersection) and $\mathcal{B}\left(S^{[0,\infty)}\right)=\sigma(\mathcal{C})$. It suffices to construct the probability measure on $\mathcal{C}$, as Carathéodory's extension theorem will then allow us to extend it to $\mathcal{B}\left(S^{[0,\infty)}\right)$.

(2) For $\Lambda\in\mathcal{C}$ of the form $\{(w_{t_1},\cdots,w_{t_n})\in\Gamma\}$, we define

$$\mathbb{P}(\Lambda)=Q_{\mathfrak{t}}(\Gamma),$$

where $\mathfrak{t}=(t_1,\cdots,t_n)$. From the consistency properties of $\{Q_{\mathfrak{t}}\}$, it is not hard to see that $\mathbb{P}$ is well-defined on $\mathcal{C}$ and it is finitely additive.

(3) Here comes the key step: $\mathbb{P}$ is countably additive on $\mathcal{C}$. It is a general result in measure theory that this is equivalent to showing that

$$\mathcal{C}\ni\Lambda_n\downarrow\emptyset\implies\mathbb{P}(\Lambda_n)\downarrow 0$$

as $n\to\infty$.

Now let $\Lambda_n\in\mathcal{C}$ be such a sequence and suppose on the contrary that

$$\lim_{n\to\infty}\mathbb{P}(\Lambda_n)=\varepsilon>0.$$

We are going to modify the sequence $\{\Lambda_n\}$ to another decreasing sequence $\{D_n\}$ which has a more convenient form

$$D_n=\{(w_{t_1},\cdots,w_{t_n})\in\Gamma_n\}$$

where $(t_1,\cdots,t_n,t_{n+1})$ is an extension of $(t_1,\cdots,t_n)$, while it still satisfies $D_n\downarrow\emptyset$ and $\lim_{n\to\infty}\mathbb{P}(D_n)=\varepsilon$. This is done by the following procedure.

First of all, by inserting marginals of the form $\{w_t\in S\}$ (of course that means doing nothing) and reordering, we may assume that Λ_n has the form

$$\Lambda_n=\{(w_{t_1},\cdots,w_{t_{m_n}})\in\Gamma_{m_n}\},$$

where $\Gamma_{m_n}\in\mathcal{B}(S^{m_n})$ and $m_n<m_{n+1}$ for every n. Since $\Lambda_{n+1}\subseteq\Lambda_n$, we know that $\Gamma_{m_{n+1}}\subseteq\Gamma_{m_n}\times S^{m_{n+1}-m_n}$.

Now we set

$$\begin{aligned}
D_1 &= \{w_{t_1}\in S\},\\
&\cdots\\
D_{m_1-1} &= \{(w_{t_1},\cdots,w_{t_{m_1-1}})\in S^{m_1-1}\},\\
D_{m_1} &= \Lambda_1,\\
D_{m_1+1} &= \{(w_{t_1},\cdots,w_{t_{m_1}},w_{t_{m_1+1}})\in\Gamma_{m_1}\times S\},\\
&\cdots\\
D_{m_2-1} &= \{(w_{t_1},\cdots,w_{t_{m_1}},w_{t_{m_1+1}},\cdots,w_{t_{m_2-1}})\in\Gamma_{m_1}\times S^{m_2-m_1-1}\},\\
D_{m_2} &= \Lambda_2,\\
&\cdots.
\end{aligned}$$

Apparently, $\{D_n\}$ is just constructed by copying each Λ_n several times consecutively in the original sequence. Therefore, it satisfies the properties $D_n \downarrow \emptyset$ and $\lim_{n\to\infty} \mathbb{P}(D_n) = \varepsilon$.

Now we are going to construct an element $(x_1, x_2, \cdots) \in S \times S \times \cdots$ such that $(x_1, \cdots, x_n) \in \Gamma_n$ for every n. It follows that the set

$$\Lambda = \{w \in S^{[0,\infty)} : \ w(t_i) = x_i \text{ for all } i\}$$

is a non-empty subset of D_n for every n, which leads to a contradiction. The construction of this element is made through a compactness argument, which relies crucially on the following general fact from measure theory (c.f. [7]).

Proposition 2.1. *Let X be a complete and separable metric space. Then every finite measure μ on $(X, \mathcal{B}(X))$ is (strongly) inner regular, in the sense that*

$$\mu(A) = \sup\{\mu(K) : \ K \subseteq A, \ K \text{ is compact}\}$$

for every $A \in \mathcal{B}(X)$.

According to Proposition 2.1, for every $n \geqslant 1$, there exists a compact subset K_n of Γ_n, such that

$$Q_{\mathfrak{t}^{(n)}}(\Gamma_n \backslash K_n) < \frac{\varepsilon}{2^n},$$

where $\mathfrak{t}^{(n)} = (t_1, \cdots, t_n)$. If we set

$$E_n = \{(w_{t_1}, \cdots, w_{t_n}) \in K_n\},$$

then we have $E_n \subseteq D_n$ and

$$\mathbb{P}(D_n \backslash E_n) = Q_{\mathfrak{t}^{(n)}}(\Gamma_n \backslash K_n) < \frac{\varepsilon}{2^n}.$$

Now define

$$\widetilde{E}_n = \bigcap_{k=1}^{n} E_k$$

and

$$\widetilde{K}_n = (K_1 \times S^{n-1}) \bigcap (K_2 \times S^{n-1}) \bigcap \cdots \bigcap (K_{n-1} \times S) \bigcap K_n.$$

Then we have

$$\widetilde{E}_n = \left\{(w_{t_1}, \cdots, w_{t_n}) \in \widetilde{K}_n\right\}.$$

On the other hand,

$$\begin{aligned} Q_{\mathfrak{t}^{(n)}}(\widetilde{K}_n) &= \mathbb{P}(\widetilde{E}_n) = \mathbb{P}(D_n) - \mathbb{P}(D_n \backslash \widetilde{E}_n) \\ &\geqslant \mathbb{P}(D_n) - \mathbb{P}\left(\bigcup_{k=1}^{n} (D_n \backslash E_k)\right) \\ &\geqslant \mathbb{P}(D_n) - \sum_{k=1}^{n} \mathbb{P}(D_k \backslash E_k) \\ &\geqslant \varepsilon - \sum_{k=1}^{n} \frac{\varepsilon}{2^k} > 0. \end{aligned}$$

Therefore, $\widetilde{K}_n \neq \emptyset$ and we may choose $\left(x_1^{(n)}, \cdots, x_n^{(n)}\right) \in \widetilde{K}_n$ for every $n \geqslant 1$.

From the construction of $\widetilde{K}_n$, we know that $\left\{x_1^{(n)}\right\}_{n \geqslant 1} \subseteq K_1$. By compactness, it contains a subsequence $x_1^{(m_1(n))} \to x_1 \in K_1$. Moreover, as $\left\{\left(x_1^{(m_1(n))}, x_2^{(m_1(n))}\right)\right\}_{n \geqslant 2} \subseteq K_2$, it further contains a subsequence $\left(x_1^{(m_2(n))}, x_2^{(m_2(n))}\right) \to (x_1, x_2) \in K_2$. Continuing the procedure, the desired element $(x_1, x_2, \cdots)$ is then constructed by induction.

(4) Finally, the uniqueness of $\mathbb{P}$ is a straight forward consequence of the uniqueness of Carathéodory's extension since $\mathcal{C}$ is a π-system and $\mathbb{P}$ is determined on $\mathcal{C}$ by the finite dimensional distributions.

Now the proof of Theorem 2.1 is complete.

Remark 2.2. Kolmogorov's extension theorem holds in a more general setting where the state space $(S, \mathcal{B}(S))$ can be an arbitrary measurable space without any topological or analytic structure. However, the given consistent family of finite dimensional distributions should satisfy some kind of generalized inner regularity property which roughly means that they can be well approximated by some sort of abstract "compact" sets. In any case the nature of Proposition 2.1 plays a crucial role.

2.3 Kolmogorov's continuity theorem

In the last subsection, a stochastic process is constructed on path space from its finite dimensional distributions. From this construction we have not yet seen any regularity properties of sample paths. It is natural to ask whether we could "detect" any sample path properties from the finite dimensional distributions. Kolmogorov's continuity theorem provides an answer to this question.

Theorem 2.2. *Let $\{X_t : t \in [0,T]\}$ be a stochastic process taking values in a complete metric space (S, d). Suppose that there exist constants $\alpha, \beta, C > 0$, such that*

$$\mathbb{E}[d(X_s, X_t)^\alpha] \leqslant C|t-s|^{1+\beta}, \ \forall s, t \in [0,T]. \tag{2.1}$$

Then there exists a continuous modification $\left\{\widetilde{X}_t : t \in [0,T]\right\}$ of X, such that for every $\gamma \in (0, \beta/\alpha)$, $\widetilde{X}$ has γ-Hölder continuous sample paths almost surely, i.e.

$$\mathbb{P}\left(\sup_{\substack{s,t \in [0,T] \\ s \neq t}} \frac{d\left(\widetilde{X}_s, \widetilde{X}_t\right)}{|t-s|^\gamma} < \infty\right) = 1.$$

To prove Theorem 2.2, without loss of generality we may assume that $T = 1$. The main idea of obtaining a continuous modification of X is to show that when restricted to some dense subset of $[0,1]$, with probability one X is uniformly continuous. This is based on the following simple fact.

Lemma 2.1. *Let D be a dense subset of $[0,1]$. Suppose that $f:\ D\rightarrow S$ is a uniformly continuous function taking values in a complete metric space (S,d). Then f extends to a continuous function on $[0,1]$ uniquely.*

Proof. Given $t\in[0,1]$, let $t_n\in D$ be such that $t_n\rightarrow t$. The uniform continuity of f implies that the sequence $\{f(t_n)\}_{n\geqslant 1}$ is a Cauchy sequence in S. Since S is complete, the limit $\lim_{n\rightarrow\infty}f(t_n)$ exists. We define $f(t)$ to be this limit. Apparently $f(t)$ is independent of the choice of t_n, and the resulting function $f:\ [0,1]\rightarrow S$ is indeed uniformly continuous. Uniqueness is obvious. □

For technical convenience, we are going to work with the dense subset D of dyadic points in $[0,1]$. To be precise, let $D=\cup_{n=0}^{\infty}D_n$, where $D_n=\{k/2^n: k=0,1,\cdots,2^n\}$. The following lemma is elementary.

Lemma 2.2. *Let $t\in D$. Then t has a unique expression $t=\sum_{i=0}^{\infty}a_i(t)2^{-i}$, where $a_i(t)$ is 0 or 1, and $a_i(t)=1$ for at most finitely many i . Moreover, for $n\geqslant 0$, let $t_n=\sum_{i=0}^{n}a_i(t)2^{-i}$. Then t_n is the largest point in D_n which does not exceed t.*

Now we prove Theorem 2.2.

Proof of Theorem 2.2. Let $\gamma\in(0,\beta/\alpha)$. For $n\geqslant 0$ and $1\leqslant k\leqslant 2^n$, Kolmogorov's criteria (2.1) implies that

$$\begin{aligned}\mathbb{P}\left(d\left(X_{\frac{k-1}{2^n}},X_{\frac{k}{2^n}}\right)>2^{-\gamma n}\right) &\leqslant 2^{\alpha\gamma n}\mathbb{E}\left[d\left(X_{\frac{k-1}{2^n}},X_{\frac{k}{2^n}}\right)^{\alpha}\right]\\ &\leqslant 2^{-n(1+\beta-\alpha\gamma)}.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{P}\left(\max_{1\leqslant k\leqslant 2^n}d\left(X_{\frac{k-1}{2^n}},X_{\frac{k}{2^n}}\right)>2^{-\gamma n}\right) &= \mathbb{P}\left(\bigcup_{k=1}^{2^n}\left\{d\left(X_{\frac{k-1}{2^n}},X_{\frac{k}{2^n}}\right)>2^{-\gamma n}\right\}\right)\\ &\leqslant \sum_{k=1}^{2^n}\mathbb{P}\left(d\left(X_{\frac{k-1}{2^n}},X_{\frac{k}{2^n}}\right)>2^{-\gamma n}\right)\\ &\leqslant 2^{-n(\beta-\alpha\gamma)}.\end{aligned}$$

Since $\beta-\alpha\gamma>0$, it follows from the Borel-Cantelli lemma (c.f. Theorem 1.4, (1)) that

$$\mathbb{P}\left(\max_{1\leqslant k\leqslant 2^{-n}}d\left(X_{\frac{k-1}{2^n}},X_{\frac{k}{2^n}}\right)>2^{-\gamma n}\text{ infinitely often}\right)=0.$$

In other words, there exists some measurable set Ω^* such that $\mathbb{P}(\Omega^*)=1$ and for every $\omega\in\Omega^*$,

$$d\left(X_{\frac{k-1}{2^n}}(\omega),X_{\frac{k}{2^n}}(\omega)\right)\leqslant 2^{-\gamma n},\ \forall k=1,\cdots,2^n\text{ and }n\geqslant n^*(\omega),$$

where $n^*(\omega)$ is some positive integer depending on ω.

Now fix $\omega \in \Omega^*$. Suppose that $s, t \in D$ satisfy $0 < |t-s| < 2^{-n^*(\omega)}$. Then there exists a unique $m \geqslant n^*(\omega)$, such that $2^{-(m+1)} \leqslant |t-s| < 2^{-m}$. Write $t = \sum_{i=0}^{\infty} a_i(t) 2^{-i}$ according to Lemma 2.2, and let $t_n = \sum_{i=0}^{n} a_i(t) 2^{-i}$ for $n \geqslant 0$. Define s_n in a similar way from s. Apparently $s_m = t_m$ or $|t_m - s_m| = 2^{-m}$. It follows that when evaluated at ω,

$$
\begin{aligned}
d(X_s, X_t) &\leqslant \sum_{i=m}^{\infty} d(X_{s_{i+1}}, X_{s_i}) + d(X_{s_m}, X_{t_m}) + \sum_{i=m}^{\infty} d(X_{t_i}, X_{t_{i+1}}) \\
&\leqslant 2 \sum_{i=m}^{\infty} 2^{-\gamma(i+1)} + 2^{-\gamma m} \\
&= \left(1 + \frac{2}{2^\gamma - 1}\right) 2^{-\gamma m} \\
&\leqslant 2^\gamma \left(1 + \frac{2}{2^\gamma - 1}\right) |t-s|^\gamma. \qquad (2.2)
\end{aligned}
$$

In particular, this shows that for every $\omega \in \Omega^*$, $X(\omega)$ is uniformly continuous when restricted on D.

We define $\widetilde{X}$ in the following way: if $\omega \notin \Omega^*$, define $\widetilde{X}(\omega) \equiv c$ for some fxied $c \in S$, and if $\omega \in \Omega^*$, define $\widetilde{X}(\omega)$ to be the unique extension of $X(\omega)$ to $[0,1]$ according to Lemma 2.1. Then $\widetilde{X}$ has continuous sample paths and (2.2) still holds for $\widetilde{X}(\omega)$ when $\omega \in \Omega^*$ and $|t-s| < 2^{-n^*(\omega)}$. Moreover, since $X_{t_n} \to \widetilde{X}_t$ a.s. and $X_{t_n} \to X_t$ in probability as $t_n \to t$, we conclude that $\widetilde{X}_t = X_t$ a.s. The process $\widetilde{X}$ is the desired one.

□

Remark 2.3. If the process X_t is defined on $[0, \infty)$ and Kolmogorov's criteria (2.1) holds on every finite interval $[0, T]$ with constant C possibly depending on T, then from the previous proof it is not hard to see that there is a continuous modification $\widetilde{X}$ of X on $[0, \infty)$, such that for every $\gamma \in (0, \beta/\alpha)$, with probability one, $\widetilde{X}$ is γ-Hölder continuous on every finite interval $[0, T]$.

2.4 Filtrations and stopping times

In the study of stochastic processes, it is rather important to keep track of information growth in the evolution of time. This leads to the very useful concept of a filtration.

Definition 2.6. A *filtration* over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a increasing sequence $\{\mathcal{F}_t : t \geqslant 0\}$ of sub-σ-algebras of $\mathcal{F}$, i.e. $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for $0 \leqslant s < t$. We call $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ a *filtered probability space.*

We can talk about additional measurability properties of a stochastic process when a filtration is presented.

Definition 2.7. Let $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ be a filtered probability space. A stochastic process X is called $\{\mathcal{F}_t\}$*-adapted* if X_t is $\mathcal{F}_t$-measurable for every $t \geqslant 0$. It is called $\{\mathcal{F}_t\}$-*progressively measurable* if for every $t \geqslant 0$, the map

$$\begin{aligned} X^{(t)}: \ [0,t] \times \Omega \ &\rightarrow \ \mathbb{R}^d, \\ (s, \omega) \ &\mapsto \ X_s(\omega), \end{aligned}$$

is $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$-measurable.

Intuitively, for an adapted process X, when the information of $\mathcal{F}_t$ is presented to an observer, the path $s \in [0,t] \mapsto X_s \in \mathbb{R}^d$ is then known to her.

It is apparent that if X is progressively measurable, then it is measurable and adapted. However, the converse is in general not true. It is true if the sample paths of X are right (or left) continuous.

Proposition 2.2. *Let X be an $\{\mathcal{F}_t\}$-adapted stochastic process. Suppose that every sample path of X is right continuous. Then X is $\{\mathcal{F}_t\}$-progressively measurable.*

Proof. We approximate X by step processes. Let $t \geqslant 0$. For $n \geqslant 1$, define

$$X_s^{(n)}(\omega) = \sum_{k=1}^{2^n} X_{\frac{k}{2^n}t}(\omega) \mathbf{1}_{\{s \in [\frac{k-1}{2^n}, \frac{k}{2^n})\}} + X_t(\omega) \mathbf{1}_{\{s=t\}}, \ (s, \omega) \in [0,t] \times \Omega.$$

Since X is adapted, it is obvious that $X^{(n)}$ is $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$-measurable. Moreover, by right continuity of X, we know that $X_s^{(n)}(\omega) \rightarrow X_s(\omega)$ for every $(s, \omega) \in [0,t] \otimes \Omega$. Therefore, X is progressively measurable. □

Example 2.1. Let X_t be a stochastic process on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We can define the *natural filtration* of X_t to be

$$\mathcal{F}_t^X = \sigma(X_s : \ 0 \leqslant s \leqslant t), \ t \geqslant 0.$$

Apparently, X_t is $\{\mathcal{F}_t^X\}$-adapted. According to Proposition 2.2, if X_t has right continuous sample paths, then it is $\{\mathcal{F}_t^X\}$-progressively measurable.

Another very important concept for our study is a stopping time. Intuitively, a stopping time usually models the first time that some phenomenon occurs, for instance the first time that the temperature of the classroom reaches 25 degree. A characterizing property for such time τ is: if we keep observing up to time t, we could decide whether τ is observed or not (i.e. whether the event $\{\tau \leqslant t\}$ happens), and if τ is not observed before time t, we have no idea when exactly in the future the temperature will reach 25 degree. This motivates the following definition.

Definition 2.8. Let $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ be a filtered probability space. A random time $\tau : \ \Omega \rightarrow [0, \infty]$ is called an $\{\mathcal{F}_t\}$*-stopping time* if $\{\tau \leqslant t\} \in \mathcal{F}_t$ for every $t \geqslant 0$.

Apparently, every constant time is an $\{\mathcal{F}_t\}$-stopping time. Moreover, we can easily construct new stopping times from given ones.

Proposition 2.3. *Suppose that σ, τ, τ_n are $\{\mathcal{F}_t\}$-stopping times. Then*

$$\sigma + \tau,\ \sigma \wedge \tau,\ \sigma \vee \tau,\ \sup_n \tau_n$$

are all $\{\mathcal{F}_t\}$-stopping times, where "$\wedge$" ("$\vee$", respectively) means taking minimum (taking maximum, respectively).

Proof. Consider the following decomposition:

$$\begin{aligned}\{\sigma + \tau > t\} &= \{\sigma = 0, \tau > t\} \bigcup \{0 < \sigma < t, \sigma + \tau > t\} \\ &\bigcup \{\sigma \geqslant t, \tau > 0\} \bigcup \{\sigma > t, \tau = 0\}.\end{aligned}$$

The first and fourth events are obviously in $\mathcal{F}_t$. The third event is in $\mathcal{F}_t$ because

$$\{\sigma < t\} = \bigcup_n \left\{\sigma \leqslant t - \frac{1}{n}\right\} \in \mathcal{F}_t.$$

For the second event, if $\omega \in \{0 < \sigma < t, \sigma + \tau > t\}$, then

$$\tau(\omega) > t - \sigma(\omega) > 0.$$

Keeping in mind that $\sigma(\omega) > 0$, we can certainly choose $r \in (0, t) \cap \mathbb{Q}$, such that

$$\tau(\omega) > r > t - \sigma(\omega).$$

Therefore, we see that

$$\{0 < \sigma < t, \sigma + \tau > t\} = \bigcup_{r \in (0,t) \cap \mathbb{Q}} \{\tau > r, t - r < \sigma < t\} \in \mathcal{F}_t.$$

For the other cases, we simply observe that

$$\begin{aligned}\{\sigma \wedge \tau > t\} &= \{\sigma > t, \tau > t\} \in \mathcal{F}_t, \\ \{\sigma \vee \tau \leqslant t\} &= \{\sigma \leqslant t, \tau \leqslant t\} \in \mathcal{F}_t, \\ \left\{\sup_n \tau_n \leqslant t\right\} &= \bigcap_n \{\tau_n \leqslant t\} \in \mathcal{F}_t.\end{aligned}$$

□

Remark 2.4. In general, $\inf_n \tau_n$, and therefore $\limsup_{n\to\infty} \tau_n$, $\liminf_{n\to\infty} \tau_n$ may fail to be $\{\mathcal{F}_t\}$-stopping time even though each τ_n is. However, it is a good exercise to show that they are $\{\mathcal{F}_{t+}\}$-stopping times, where $\{\mathcal{F}_{t+}\}$ is the filtration defined by

$$\mathcal{F}_{t+} = \bigcap_{u > t} \mathcal{F}_u \supseteq \mathcal{F}_t,\ t \geqslant 0. \tag{2.3}$$

We could also talk about the accumulated information up to a stopping time τ. Intuitively, the occurrence of an event A can be determined by such information if the following condition holds. Suppose that the accumulated information up to time t is presented. If we observe that $\tau \leqslant t$, we should be able to decide whether A happens or not because the information up to time τ is then known. However, if we observe that $\tau > t$, since in this case we cannot decide the exact value of τ, part of the information up to τ is missing and the occurrence of A should be undecidable. This motivates the following definition.

Definition 2.9. Let τ be an $\{\mathcal{F}_t\}$-stopping time. The *pre-τ σ-algebra* $\mathcal{F}_\tau$ is defined by

$$\mathcal{F}_\tau = \left\{ A \in \mathcal{F}_\infty : \ A \bigcap \{\tau \leqslant t\} \in \mathcal{F}_t, \quad \forall t \geqslant 0 \right\},$$

where $\mathcal{F}_\infty \triangleq \sigma\left(\cup_{t\geqslant 0}\mathcal{F}_t\right)$.

It follows from the definition that $\mathcal{F}_\tau$ is a sub-σ-algebra of $\mathcal{F}$, and τ is $\mathcal{F}_\tau$-measurable. Moreover, if $\tau \equiv t$, then $\mathcal{F}_\tau = \mathcal{F}_t$. And we have the following basic properties.

Proposition 2.4. *Suppose that σ, τ are two $\{\mathcal{F}_t\}$-stopping times.*
(1) Let $A \in \mathcal{F}_\sigma$, then $A \cap \{\sigma \leqslant \tau\} \in \mathcal{F}_\tau$. In particular, if $\sigma \leqslant \tau$, then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.
(2) $\mathcal{F}_{\sigma\wedge\tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau$, and the events

$$\{\sigma < \tau\}, \ \{\sigma > \tau\}, \ \{\sigma \leqslant \tau\}, \ \{\sigma \geqslant \tau\}, \ \{\sigma = \tau\}$$

are all $\mathcal{F}_\sigma \cap \mathcal{F}_\tau$-measurable.

Proof. (1) We have

$$\begin{aligned} A\bigcap\{\sigma \leqslant \tau\}\bigcap\{\tau \leqslant t\} &= A\bigcap\{\sigma \leqslant \tau\}\bigcap\{\tau \leqslant t\}\bigcap\{\sigma \leqslant t\} \\ &= \left(A\bigcap\{\sigma \leqslant t\}\right)\bigcap\{\tau \leqslant t\}\bigcap\{\sigma\wedge t \leqslant \tau\wedge t\}. \end{aligned}$$

From definition it is obvious that $\sigma \wedge t$ and $\tau \wedge t$ are $\mathcal{F}_t$-measurable. Therefore, by assumption we know that the above event is $\mathcal{F}_t$-measurable.

(2) Since $\sigma \wedge \tau$ is an $\{\mathcal{F}_t\}$-stopping time, from the first part we know that $\mathcal{F}_{\sigma\wedge\tau} \subseteq \mathcal{F}_\sigma \cap \mathcal{F}_\tau$. Now suppose $A \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau$, then

$$\begin{aligned} A\bigcap\{\sigma\wedge\tau \leqslant t\} &= A\bigcap\left(\{\sigma \leqslant t\}\bigcup\{\tau \leqslant t\}\right) \\ &= \left(A\bigcap\{\sigma \leqslant t\}\right)\bigcup\left(A\bigcap\{\tau \leqslant t\}\right) \in \mathcal{F}_t. \end{aligned}$$

Therefore, $A \in \mathcal{F}_{\sigma\wedge\tau}$.

Finally, by taking $A = \Omega$ in the first part, we know that $\{\sigma > \tau\} = \{\sigma \leqslant \tau\}^c \in \mathcal{F}_\tau$. It follows that

$$\{\sigma < \tau\} = \{\sigma\wedge\tau < \tau\} \in \mathcal{F}_{\sigma\wedge\tau} = \mathcal{F}_\sigma \bigcap \mathcal{F}_\tau.$$

The other cases follow by symmetry and complementation. $\square$

In the study of martingales and strong Markov processes, it is important to consider conditional expectations given $\mathcal{F}_\tau$. We give two basic properties here.

Proposition 2.5. *Suppose that σ, τ are two $\{\mathcal{F}_t\}$-stopping times and X is an integrable random variable. Then we have:*
(1) $\mathbb{E}\left[\mathbf{1}_{\{\sigma\leqslant\tau\}}X|\mathcal{F}_\sigma\right] = \mathbb{E}\left[\mathbf{1}_{\{\sigma\leqslant\tau\}}X|\mathcal{F}_{\sigma\wedge\tau}\right]$.
(2) $\mathbb{E}\left[\mathbb{E}[X|\mathcal{F}_\sigma]|\mathcal{F}_\tau\right] = \mathbb{E}[X|\mathcal{F}_{\sigma\wedge\tau}]$.

Proof. (1) According to the second part of Proposition 2.4, $\{\sigma \leqslant \tau\} \in \mathcal{F}_{\sigma\wedge\tau} \subseteq \mathcal{F}_\sigma$. Therefore, it suffices to show that $\mathbf{1}_{\{\sigma\leqslant\tau\}}\mathbb{E}[X|\mathcal{F}_\sigma]$ is $\mathcal{F}_{\sigma\wedge\tau}$-measurable. Apparently it is $\mathcal{F}_\sigma$-measurable. But the $\mathcal{F}_\tau$-measurability is a direct consequence of the first part of Proposition 2.4 (standard approximation allows us to replace A by a general $\mathcal{F}_\sigma$-measurable function in that proposition).

(2) First observe that the same argument allows us to conclude that

$$A \in \mathcal{F}_\sigma \implies A\bigcap\{\sigma < \tau\} \in \mathcal{F}_\tau, \tag{2.4}$$

and

$$\mathbb{E}\left[\mathbf{1}_{\{\sigma<\tau\}}X|\mathcal{F}_\sigma\right] = \mathbb{E}\left[\mathbf{1}_{\{\sigma<\tau\}}X|\mathcal{F}_{\sigma\wedge\tau}\right].$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}[X|\mathcal{F}_\sigma]|\mathcal{F}_\tau\right] &= \mathbb{E}\left[\mathbf{1}_{\{\sigma<\tau\}}\mathbb{E}[X|\mathcal{F}_\sigma]|\mathcal{F}_\tau\right] + \mathbb{E}\left[\mathbf{1}_{\{\sigma\geqslant\tau\}}\mathbb{E}[X|\mathcal{F}_\sigma]|\mathcal{F}_\tau\right] \\
&= \mathbb{E}\left[\mathbf{1}_{\{\sigma<\tau\}}\mathbb{E}[X|\mathcal{F}_{\sigma\wedge\tau}]|\mathcal{F}_\tau\right] + \mathbb{E}\left[\mathbf{1}_{\{\sigma\geqslant\tau\}}\mathbb{E}[X|\mathcal{F}_\sigma]|\mathcal{F}_{\sigma\wedge\tau}\right] \\
&= \mathbb{E}\left[\mathbf{1}_{\{\sigma<\tau\}}X|\mathcal{F}_{\sigma\wedge\tau}\right] + \mathbb{E}\left[\mathbf{1}_{\{\sigma\geqslant\tau\}}X|\mathcal{F}_{\sigma\wedge\tau}\right] \\
&= \mathbb{E}[X|\mathcal{F}_{\sigma\wedge\tau}].
\end{aligned}$$

□

Now we consider measurability properties for a stochastic process stopped at some stopping time.

Proposition 2.6. *Let $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ be a filtered probability space. Suppose that X_t is an $\{\mathcal{F}_t\}$-progressively measurable stochastic process and τ is an $\{\mathcal{F}_t\}$-stopping time. Then the stopped process $t \mapsto X_{\tau\wedge t}$ is also $\{\mathcal{F}_t\}$-progressively measurable. In particular, the stopped random variable $X_\tau\mathbf{1}_{\{\tau<\infty\}}$ is $\mathcal{F}_\tau$-measurable.*

Proof. Restricted on $[0,t]\times\Omega$, the stopped process is given by the following composition of maps:

$$\begin{array}{ccccc}
[0,t]\times\Omega & \to & [0,t]\times\Omega & \to & \mathbb{R}^d, \\
(s,\omega) & \mapsto & (\tau(\omega)\wedge s, \omega) & \mapsto & X_{\tau(\omega)\wedge s}(\omega).
\end{array}$$

By assumption, we know that the second map is $\mathcal{B}([0,t])\otimes\mathcal{F}_t$-measurable. The $\mathcal{B}([0,t])\otimes\mathcal{F}_t$-measurability of the first map can be easily seen from the following fact:

$$\{(s,\omega):\ (\tau(\omega)\wedge s,\omega) \in [0,c]\times A\} = ([0,c]\times A)\bigcup\Big((c,t]\times\Big(\{\tau\leqslant c\}\bigcap A\Big)\Big)$$

for every $c \in [0,t]$ and $A \in \mathcal{F}_t$.

The $\mathcal{F}_\tau$-measurability of $X_\tau \mathbf{1}_{\{\tau<\infty\}}$ follows from the $\{\mathcal{F}_t\}$-adaptedness of the stopped process $X_{\tau\wedge t}$ (because it is $\{\mathcal{F}_t\}$-progressively measurable) and the simple fact that $X_\tau \mathbf{1}_{\{\tau\leqslant t\}} = X_{\tau\wedge t}\mathbf{1}_{\{\tau\leqslant t\}}$. □

To conclude this section, we discuss a fundamental class of stopping times: hitting times for stochastic processes.

Definition 2.10. The *hitting time* of $\Gamma \subseteq \mathbb{R}^d$ by a stochastic process X is defined to be

$$H_\Gamma(\omega) = \inf\{t \geqslant 0 : \ X_t(\omega) \in \Gamma\},$$

where $\inf \emptyset \triangleq \infty$.

The following result tells us that under some conditions, a hitting time is a stopping time.

Proposition 2.7. *Let $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ be a filtered probability space. Suppose that X_t is an $\{\mathcal{F}_t\}$-adapted stochastic process such that every sample path of X_t is continuous. Then for every closed set F, H_F is an $\{\mathcal{F}_t\}$-stopping time.*

Proof. For given $t \geqslant 0$ and $\omega \in \Omega$, by continuity we know that the function

$$\varphi(s) = \mathrm{dist}(X_s(\omega), F), \ \ s \in [0,t],$$

is continuous. The result then follows from the following observation:

$$\{H_F > t\} = \bigcup_{n=1}^{\infty} \bigcap_{r\in[0,t]\cap\mathbb{Q}} \left\{\mathrm{dist}(X_r, F) > \frac{1}{n}\right\} \in \mathcal{F}_t.$$

□

On the other hand, the hitting time of an open set is in general not a stopping time even the process have continuous sample paths. The reason is intuitively simple. Suppose that a sample path of the process first hits the boundary of an open set G from the outside at time t. It is not possible to determine whether $H_G \leqslant t$ or not without looking slightly ahead into the future.

However, we do have the following result. The proof is left as an exercise.

Proposition 2.8. *Let $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ be a filtered probability space, and let X be an $\{\mathcal{F}_t\}$-adapted stochastic process such that every sample path of X is right continuous. Then for every open set G, H_G is an $\{\mathcal{F}_{t+}\}$-stopping time, where $\{\mathcal{F}_{t+}\}$ is the filtration defined by (2.3) in Remark 2.4.*

Until now, to some extend we have already seen the inconvenience caused by the difference between the filtrations $\{\mathcal{F}_t\}$ and $\{\mathcal{F}_{t+}\}$ (c.f. Remark 2.4 and Proposition 2.8). In particular, in order to include a richer class of stopping times, it is usually convenient to assumption that $\mathcal{F}_t = \mathcal{F}_{t+}$ for every $t \geqslant 0$, i.e. the filtration $\{\mathcal{F}_t\}$-is right continuous. This seemingly unnatural assumption is indeed very mild: the natural filtration of a strong Markov process, when augmented by null sets, is always right continuous (c.f. [5]).

Another mild and reasonable assumption on the filtered probability space is to make sure that most probabilistic properties, in particular for those related to adaptedness and stopping times, are preserved by another stochastic process which is indistinguishable from the original one. Mathematically speaking, this is the assumption that $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets (recall from our convention that N is a $\mathbb{P}$-null set if there exists $E \in \mathcal{F}$, such that $N \subseteq E$ and $\mathbb{P}(E) = 0$). In particular, this implies that $\mathcal{F}$ and every $\mathcal{F}_t$ are $\mathbb{P}$-complete.

Definition 2.11. A filtration is said to satisfy the *usual conditions* if it is right continuous and $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets.

Given an arbitrary filtered probability space $(\Omega, \mathcal{G}, \mathbb{P}; \{\mathcal{G}_t\})$, it can always be augmented to satisfy the usual conditions. Indeed, let $\mathcal{F}$ be the $\mathbb{P}$-completion of $\mathcal{G}$ and let $\mathcal{N}$ be the collection of $\mathbb{P}$-null sets. For every $t \geqslant 0$, we define

$$\mathcal{F}_t = \bigcap_{s>t} \sigma(\mathcal{G}_s, \mathcal{N}) = \sigma(\mathcal{G}_{t+}, \mathcal{N}).$$

Then $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ is the smallest filtered probability space containing $(\Omega, \mathcal{G}, \mathbb{P}; \{\mathcal{G}_t\})$ which satisfies the usual conditions. We call it the *usual augmentation* of $(\Omega, \mathcal{G}, \mathbb{P}; \{\mathcal{G}_t\})$. It is a good exercise to provide the details of the proof.

In Proposition 2.7, if we drop the assumption that X_t has continuous sample paths, the situation becomes very subtle. It can still be proved in a tricky set theoretic way that, under the usual conditions on $\{\mathcal{F}_t\}$, H_F is an $\{\mathcal{F}_t\}$-stopping time provided F is a compact set and every sample path of X_t is right continuous with left limits (c.f. [9]). However, the case when F is a general Borel set is even much more difficult. The result is stated as follows. The proof relies on the machinery of Choquet's capacitability theory (c.f. [2]). The usual conditions again play a crucial role in the theorem.

Theorem 2.3. *Let $(\Omega, \mathcal{G}, \mathbb{P}; \{\mathcal{G}_t\})$ be a filtered probability space. Suppose that X_t is a $\{\mathcal{G}_t\}$-progressively measurable stochastic process. Then for every $\Gamma \in \mathcal{B}(\mathbb{R}^d)$, H_Γ is an $\{\mathcal{F}_t\}$-stopping time, where $\{\mathcal{F}_t\}$ is the usual augmentation of $\{\mathcal{G}_t\}$.*

It is true that many interesting and important probabilistic properties will be preserved if we work with the usual augmentation of the original filtered probability space. Moreover, the usual conditions have more implications than just enriching the class of stopping times, for example in questions related to sample path regularity properties (c.f. Theorem 3.10, 3.11).

However, to remain fairly careful, we will not always assume that we are working under the usual conditions. We will state clearly whenever they are assumed.

3 Continuous time martingales

This section is devoted to the study of continuous time martingales. The theory of martingales and martingale methods lies in the heart of stochastic analysis. As we will see, we are adopting a very martingale flavored approach to develop the whole theory of stochastic calculus. The main results in this section are mainly due to Doob.

3.1 Basic properties and the martingale transform: discrete stochastic integration

We start with the discrete time situation. As we will see, under certain reasonable regularity conditions on sample paths, parallel results in the continuous time situation can be derived easily from the discrete case. Therefore, in most of the topics we consider in this section, we do not really see a substantial difference between the two situations. However, in Section 5, we will appreciate many deep and elegant properties of continuous time martingales which do not have their discrete time counterparts.

Let $T = \{0, 1, 2, \cdots\}$ or $[0, \infty)$.

Definition 3.1. Let $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t : t \in T\})$ be a filtered probability space. A real-valued stochastic process $\{X_t, \mathcal{F}_t : t \in T\}$ is called a *martingale* (respectively, *submartingale*, *supermartingale*) if:

(1) X_t is $\{\mathcal{F}_t\}$-adapted;
(2) X_t is integrable for every $t \in T$;
(3) for every $s < t \in T$,

$$\mathbb{E}[X_t|\mathcal{F}_s] = X_s, \quad (\text{respectively, }" \geqslant "," \leqslant ").$$

Example 3.1. Let $T = [0, \infty)$. Consider the stochastic process $\left\{\widetilde{X}_t : t \in T\right\}$ constructed in Problem Sheet 2, Problem 2 in dimension $d = 1$ (the 1-dimensional pre-Brownian motion). Let $\left\{\mathcal{F}_t^{\widetilde{X}}\right\}$ be the natural filtration of $\widetilde{X}_t$. Then for $0 \leqslant s < t$, $\widetilde{X}_t - \widetilde{X}_s$ is independent of $\mathcal{F}_s^{\widetilde{X}}$ with zero mean. Therefore, $\left\{\widetilde{X}_t, \mathcal{F}_t^{\widetilde{X}}\right\}$ is a martingale.

A very useful way of constructing a new submartingale from the old is the following.

Proposition 3.1. *Let $\{X_t, \mathcal{F}_t : t \in T\}$ be a martingale (respectively, a submartingale). Suppose that $\varphi : \mathbb{R}^1 \to \mathbb{R}^1$ is a convex function (respectively, a convex and increasing function). If $\varphi(X_t)$ is integrable for every $t \in T$, then $\{\varphi(X_t), \mathcal{F}_t : t \in T\}$ is submartingale.*

Proof. The adaptedness and integrability conditions are satisfied. To see the submartingale property, we apply Jensen's inequality (c.f. inequality (1.2)) to obtain that

$$\mathbb{E}[\varphi(X_t)|\mathcal{F}_s] \geqslant \varphi(\mathbb{E}[X_t|\mathcal{F}_s]) \geqslant \varphi(X_s)$$

for $s < t \in T$. □

Example 3.2. If $\{X_t, \mathcal{F}_t\}$ is a martingale, then $X_t^+ \triangleq \max(X_t, 0)$ and $|X_t|^p$ $(p \geqslant 1)$ are $\{\mathcal{F}_t\}$-submartingales, provided X_t is in L^p for every t.

Now we consider the situation when $T = \{0, 1, 2, \cdots\}$.

In the discrete time setting, stopping times and pre-stopping time σ-algebras are defined analogously to the continuous time case in an obvious way.

We are going to construct a class of martingales which plays a central role in this section, in particular in the study of martingale convergence and the optional sampling theorem.

Definition 3.2. Let $\{\mathcal{F}_n : n \geqslant 0\}$ be a filtration. A real-valued random sequence $\{C_n : n \geqslant 1\}$ is said to be $\{\mathcal{F}_n\}$-*predictable* if C_n is $\mathcal{F}_{n-1}$-measurable for every $n \geqslant 1$.

Let $\{X_n : n \geqslant 0\}$ and $\{C_n : n \geqslant 1\}$ be two sequences. We define another sequence $\{Y_n : n \geqslant 0\}$ by $Y_0 = 0$ and

$$Y_n = \sum_{k=1}^{n} C_k (X_k - X_{k-1}).$$

Definition 3.3. The sequence $\{Y_n : n \geqslant 0\}$ is called the *martingale transform* of X_n by C_n. It is denoted by $(C \bullet X)_n$.

Comparing with Section 5, the martingale transform can be regarded as a discrete version of stochastic integration.

The following result verifies the name. Its proof is straight forward.

Theorem 3.1. *Let $\{X_n, \mathcal{F}_n : n \geqslant 0\}$ be a martingale (respectively, submartingale, supermartingale) and let $\{C_n : n \geqslant 1\}$ be an $\{\mathcal{F}_n\}$-predictable random sequence which is bounded (respectively, bounded and non-negative). Then the martingale transform $\{(C \bullet X)_n, \mathcal{F}_n : n \geqslant 0\}$ of X_n by C_n is a martingale (respectively, submartingale, supermartingale).*

Remark 3.1. The boundedness of C_n is not important–we only need to guarantee the integrability of Y_n.

It is very helpful to have the following intuition of the martingale transform in mind. Suppose that you are playing games over the time horizon $\{1, 2, \cdots\}$. C_n is interpreted as your stake on game n. Predictability means that you are making your decision on the stake amount C_n based on the history $\mathcal{F}_{n-1}$. $X_n - X_{n-1}$ represents your winning at game n per unit stake. Therefore, Y_n is your total winning up to time n.

3.2 The martingale convergence theorems

The (sub or super)martingale property exhibits a trend on average in the long run. It is therefore not unreasonable to expect that a (sub or super)martingale can converge (almost surely) if its mean sequence is well controlled.

We first explain a general way of proving the almost sure convergence of a random sequence.

Let $\{X_n : n \geqslant 0\}$ be a random sequence. Then $X_n(\omega)$ is convergent if and only if $\liminf_{n\to\infty} X_n(\omega) = \limsup_{n\to\infty} X_n(\omega)$. Therefore,

$$\begin{aligned}\{X_n \text{ does not converge}\} &\subseteq \left\{\liminf_{n\to\infty} X_n < \limsup_{n\to\infty} X_n\right\} \\ &\subseteq \bigcup_{\substack{a<b \\ a,b\in\mathbb{Q}}} \left\{\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n\right\}.\end{aligned}$$

Therefore, in order to prove that X_n converges a.s., it suffices to show that

$$\mathbb{P}\left(\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n\right) = 0 \tag{3.1}$$

for every given $a < b$. But the event in the bracket implies that a subsequence of X_n lies below a while another subsequence of X_n lies above b. This further implies that the total number of upcrossings by the sequence X_n from below a to above b must be infinite.

Therefore, the convergence of X_n is closely related to controlling the upcrossing number of an interval $[a, b]$.

Now we define this number mathematically.

Consider the following two sequences of random times:

$$\begin{aligned}
&\sigma_0 = 0, \\
&\sigma_1 = \inf\{n \geqslant 0 : X_n < a\}, \qquad \tau_1 = \inf\{n > \sigma_1 : X_n > b\}, \\
&\sigma_2 = \inf\{n > \tau_1 : X_n < a\}, \qquad \tau_2 = \inf\{n > \sigma_2 : X_n > b\}, \\
&\cdots \\
&\sigma_k = \inf\{n > \tau_{k-1} : X_n < a\}, \qquad \tau_k = \inf\{n > \sigma_k : X_n > b\}, \\
&\cdots .
\end{aligned}$$

Definition 3.4. For $N \geqslant 0$, the *upcrossing number* $U_N(X; [a, b])$ of the interval $[a, b]$ by the sequence X_n up to time N is define to be random number

$$U_N(X; [a, b]) = \sum_{k=1}^{\infty} \mathbf{1}_{\{\tau_k \leqslant N\}}.$$

Note that $U_N(X; [a, b]) \leqslant N/2$. Moreover, if $\{\mathcal{F}_n : n \geqslant 0\}$ is a filtration and X_n is $\{\mathcal{F}_n\}$-adapted, then σ_k, τ_k are $\{\mathcal{F}_n\}$-stopping times. In particular, in this case $U_N(X; [a, b])$ is $\mathcal{F}_N$-measurable.

The main result of controlling $U_N(X; [a, b])$ in our context is the following. Here we are in particular working with supermartingales. The technique of dealing with the submartingale case is actually quite different. However, as they both lead to the same general convergence theorems, we will omit the discussion of the submartingale case and focus on supermartingales.

Proposition 3.2 (Doob's upcrossing inequality). *Let $\{X_n, \mathcal{F}_n : n \geqslant 0\}$ be a supermartingale. Then the upcrossing number $U_N(X;[a,b])$ of $[a,b]$ by X_n up to time N satisfies the following inequality:*

$$\mathbb{E}[U_N(X;[a,b])] \leqslant \frac{\mathbb{E}[(X_N-a)^-]}{b-a}, \tag{3.2}$$

where $x^- \triangleq \max(-x,0)$.

The proof of this inequality can be fairly easy as long as we can find a good way of looking at this upcrossing number.

Suppose in the aforementioned gambling model that $X_n - X_{n-1}$ represents the winning at game n per unit stake. Now consider the following gambling strategy: repeat the following two steps until forever:

(1) Wait until X_n gets below a;

(2) Play unit stakes onwards until X_n gets above b and stop playing.

Mathematically, this is to say that we define

$$\begin{aligned} C_1 &= \mathbf{1}_{\{X_0<a\}}, \\ C_n &= \mathbf{1}_{\{C_{n-1}=0\}}\mathbf{1}_{\{X_{n-1}<a\}} + \mathbf{1}_{\{C_{n-1}=1\}}\mathbf{1}_{\{X_{n-1}\leqslant b\}}, \quad n \geqslant 2. \end{aligned}$$

Let $\{Y_n\}$ be the martingale transform of X_n by C_n. Then Y_N represents the total winning up to time N. Note that Y_N comes from two parts: the playing intervals corresponding to complete upcrossings, and the last playing interval corresponding to the last incomplete upcrossing (possibly non-existing). The total winning Y_N from the first part is obviously bounded below by $(b-a)U_N(X;[a,b])$, and the total winning in the last playing interval (possibly non-existing) is bounded below $-(X_N-a)^-$. Therefore, we have

$$Y_N \geqslant (b-a)U_N(X;[a,b]) - (X_N-a)^-.$$

On the other hand, by definition it is apparent that $\{C_n\}$ is a bounded and non-negative $\{\mathcal{F}_n\}$-predictable sequence. According to Theorem 3.1, $\{Y_n, \mathcal{F}_n\}$ is a supermartingale. Therefore, $\mathbb{E}[Y_N] \leqslant \mathbb{E}[Y_0] = 0$, which implies (3.2).

Now since $U_N(X;[a,b])$ is increasing in N, we may define

$$U_\infty(X;[a,b]) = \lim_{N\to\infty} U_N(X;[a,b]),$$

which is the upcrossing number for the whole time horizon.

From Doob's upcrossing inequality, we can immediately see that if the supermartingale $\{X_n, \mathcal{F}_n\}$ is bounded in L^1, i.e. $\sup_{n\geqslant 0}\mathbb{E}[|X_n|] < \infty$, then

$$\mathbb{E}[U_\infty(X;[a,b])] = \lim_{N\to\infty}\mathbb{E}[U_N(X;[a,b])] \leqslant \frac{\sup_{n\geqslant 0}\mathbb{E}[|X_n|]+|a|}{b-a} < \infty.$$

In particular, $U_\infty(X;[a,b]) < \infty$ a.s.

But from the discussion at the beginning of this subsection, we know that

$$\left\{\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n\right\} \subseteq \{U_\infty(X;[a,b]) = \infty\}.$$

Therefore, (3.1) holds and we conclude that X_n converges to some X_∞ a.s. Moreover, Fatou's lemma shows that

$$\mathbb{E}[|X_\infty|] \leqslant \liminf_{n\to\infty} \mathbb{E}[|X_n|] \leqslant \sup_{n\geqslant 0} \mathbb{E}[|X_n|] < \infty.$$

In other words, we have proved the following convergence result.

Theorem 3.2 (Doob's supermartingale convergence theorem). *Let $\{X_n, \mathcal{F}_n : n \geqslant 0\}$ be a supermartingale which is bounded in L^1. Then X_n converges almost surely to an integrable random variable X_∞.*

Remark 3.2. In the theorem, we can define $X_\infty = \limsup_{n\to\infty} X_n$, so that X_∞ is $\mathcal{F}_\infty$-measurable, where $\mathcal{F}_\infty \triangleq \sigma\left(\cup_{n=0}^\infty \mathcal{F}_n\right)$.

Now we consider the question about when the convergence holds in L^1. This is closely related to uniform integrability.

Theorem 3.3. *Let $\{X_n, \mathcal{F}_n : n \geqslant 0\}$ be a supermartingale which is bounded in L^1, so that X_n converges almost surely to some $X_\infty \in L^1$. Then the following statements are equivalent:*

(1) $\{X_n\}$ is uniformly integrable;

(2) X_n converges to X_∞ in L^1.

In this case, we have

(3) $\mathbb{E}[X_\infty|\mathcal{F}_n] \leqslant X_n$ a.s.

In addition, if $\{X_n, \mathcal{F}_n\}$ is a martingale, then (1) or (2) is also equivalent to (3) with "$\leqslant$" replaced by "$=$".

Proof. Since almost sure convergence implies convergence in probability, the equivalence of (1) and (2) is a direct consequence of Theorem 1.3. To see (3), it suffices to show that

$$\int_A X_\infty d\mathbb{P} \leqslant \int_A X_n d\mathbb{P}, \quad \forall A \in \mathcal{F}_n. \tag{3.3}$$

But from the supermartingale property, we know that

$$\int_A X_m d\mathbb{P} \leqslant \int_A X_n d\mathbb{P}, \quad \forall m \geqslant n,\ A \in \mathcal{F}_n.$$

Therefore, (3.3) follows from letting $m \to \infty$.

The last part of the theorem in the martingale case is seen from Problem Sheet 1, Problem 2, (1). ☐

Corollary 3.1 (Lévy's forward theorem). *Let Y be an integrable random variable and let $\{\mathcal{F}_n : n \geqslant 0\}$ be a filtration. Then $X_n = \mathbb{E}[Y|\mathcal{F}_n]$ is a uniformly integrable martingale such that X_n converges to $\mathbb{E}[Y|\mathcal{F}_\infty]$ almost surely and in L^1.*

Proof. The martingale property follows from

$$\mathbb{E}[X_m|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[Y|\mathcal{F}_m]|\mathcal{F}_n] = \mathbb{E}[Y|\mathcal{F}_n] = X_n, \quad \forall m > n.$$

Uniform integrability follows from Problem Sheet 1, Problem 2, (1). In particular, from Theorem 1.1 we know that $\{X_n\}$ is bounded in L^1. According to Theorem 3.3, X_n converges to some X_∞ almost surely and in L^1.

Now it suffices to show that $X_\infty = \mathbb{E}[Y|\mathcal{F}_\infty]$ a.s. Since $\cup_{n=0}^\infty \mathcal{F}_n$ is a π-system, we only need to verify

$$\int_A X_\infty d\mathbb{P} = \int_A Y d\mathbb{P}, \quad \forall A \in \mathcal{F}_n, \ n \geqslant 0.$$

This follows from letting $m \to \infty$ in the identity:

$$\int_A X_m d\mathbb{P} = \int_A Y d\mathbb{P}, \quad \forall m \geqslant n, \ A \in \mathcal{F}_n.$$

□

It is sometimes very useful to consider martingales running backward in time, or equivalently, to work with negative time parameter, in particular in the study of continuous time martingales. As we shall see, due to the natural ordering of negative integers, convergence properties for backward martingales are simpler and stronger than the forward case.

Let $T = \{-1, -2, \cdots\}$. By using the natural ordering on T, we define the notions of (sub or super)martingales in the same way as the non-negative time parameter case. Now observe that we have a decreasing filtration

$$\mathcal{G}_{-\infty} \triangleq \bigcap_{n=1}^{\infty} \mathcal{G}_{-n} \subseteq \cdots \subseteq \mathcal{G}_{-(m+1)} \subseteq \mathcal{G}_{-m} \cdots \subseteq \mathcal{G}_{-1}$$

as $n \to -\infty$.

The following convergence theorem plays a crucial role in the passage from discrete to continuous time.

Theorem 3.4 (The Lévy-Doob backward theorem). *Let $\{X_n, \mathcal{G}_n : \ n \leqslant -1\}$ be a supermartingale. Suppose that $\sup_{n \leqslant -1} \mathbb{E}[X_n] < \infty$. Then X_n is uniformly integrable, and the limit*

$$X_{-\infty} \triangleq \lim_{n \to -\infty} X_n$$

exists almost surely and in L^1. Moreover, for $n \leqslant -1$, we have

$$\mathbb{E}[X_n|\mathcal{G}_{-\infty}] \leqslant X_{-\infty}, \quad \text{a.s.},$$

with equality if $\{X_n, \mathcal{G}_n\}$ is a martingale.

Proof. To see that X_n converges almost surely, we use the same technique as in the proof of Doob's supermartingale convergence theorem. The main difference here is that the right hand side of Doob's upcrossing inequality is now in terms of X_{-1} since we are working with negative times. Therefore, the limit $X_{-\infty} \triangleq \lim_{n\to-\infty} X_n$ exists almost surely (possibly infinite) without any additional assumptions. It is apparent that $X_{-\infty}$ can be defined to be $\mathcal{G}_{-\infty}$-measurable (c.f. Remark 3.2).

Now we show uniform integrability.

Let $\lambda > 0$ and $n \leqslant k \leqslant -1$. According to the supermartingale property, we have

$$\begin{aligned}
\mathbb{E}\left[|X_n|\mathbf{1}_{\{|X_n|>\lambda\}}\right] &= \mathbb{E}\left[X_n\mathbf{1}_{\{X_n>\lambda\}}\right] - \mathbb{E}\left[X_n\mathbf{1}_{\{X_n<-\lambda\}}\right] \\
&= \mathbb{E}[X_n] - \mathbb{E}\left[X_n\mathbf{1}_{\{X_n\leqslant\lambda\}}\right] - \mathbb{E}\left[X_n\mathbf{1}_{\{X_n<-\lambda\}}\right] \\
&\leqslant \mathbb{E}[X_n] - \mathbb{E}\left[X_k\mathbf{1}_{\{X_n\leqslant\lambda\}}\right] - \mathbb{E}\left[X_k\mathbf{1}_{\{X_n<-\lambda\}}\right] \\
&= \mathbb{E}[X_n] - \mathbb{E}[X_k] + \mathbb{E}\left[X_k\mathbf{1}_{\{X_n>\lambda\}}\right] - \mathbb{E}\left[X_k\mathbf{1}_{\{X_n<-\lambda\}}\right] \\
&\leqslant \mathbb{E}[X_n] - \mathbb{E}[X_k] + \mathbb{E}\left[|X_k|\mathbf{1}_{\{|X_n|>\lambda\}}\right].
\end{aligned}$$

Given $\varepsilon > 0$, by the assumption $\sup_{n\leqslant-1}\mathbb{E}[X_n] < \infty$, there exists $k \leqslant -1$, such that

$$0 \leqslant \mathbb{E}[X_n] - \mathbb{E}[X_k] \leqslant \frac{\varepsilon}{2}, \quad \forall n \leqslant k.$$

Moreover, for this particular k, by integrability there exists $\delta > 0$, such that

$$A \in \mathcal{F}, \mathbb{P}(A) < \delta \implies \mathbb{E}[|X_k|\mathbf{1}_A] < \frac{\varepsilon}{2}.$$

On the other hand, since $\{X_n, \mathcal{G}_n\}$ is a supermartingale, we know that $\{X_n^-, \mathcal{G}_n\}$ is a submartingale. Therefore,

$$\mathbb{E}[|X_n|] = \mathbb{E}[X_n] + 2\mathbb{E}[X_n^-] \leqslant \mathbb{E}[X_n] + 2\mathbb{E}[X_{-1}^-], \quad \forall n \leqslant -1.$$

This implies that $M \triangleq \sup_{n\leqslant-1}\mathbb{E}[|X_n|] < \infty$ (which by Fatou's lemma already implies that X_∞ is a.s. finite), and

$$\mathbb{P}(|X_n| > \lambda) \leqslant \frac{\mathbb{E}[|X_n|]}{\lambda} \leqslant \frac{M}{\lambda}, \quad \forall n \leqslant -1, \ \lambda > 0.$$

Now we choose $\Lambda > 0$ such that if $\lambda > \Lambda$, then

$$\mathbb{P}(|X_n| > \lambda) < \delta, \quad \forall n \leqslant k,$$

and

$$\mathbb{E}\left[|X_n|\mathbf{1}_{\{|X_n|>\lambda\}}\right] < \varepsilon, \quad \forall k < n \leqslant -1.$$

The uniform integrability then follows.

With uniform integrability, it follows immediately from Theorem 1.3 that $X_n \to X_\infty$ almost surely and in L^1 as $n \to -\infty$.

Finally, the last part of the theorem follows from

$$\int_A X_n d\mathbb{P} \leqslant \int_A X_m d\mathbb{P}, \quad \forall A \in \mathcal{G}_{-\infty}, \ m \leqslant n \leqslant -1,$$

and letting $m \to -\infty$. □

Remark 3.3. We have seen that the fundamental condition that guarantees convergence is the boundedness in L^1. In particular, all the convergence results we discussed before hold for submartingales as well, since $-X_n$ is a supermartingale if X_n is a submartingale, and applying a minus sign does not affect the L^1-boundedness.

3.3 Doob's optional sampling theorems

Given a martingale, under certain conditions it is reasonable to expect that the martingale property is preserved when sampling along stopping times. The study of this problem leads to the so-called Doob's optional sampling theorems.

Again we will mainly work with (super)martingales, and all the results apply to submartingales by applying a minus sign.

Let $\{X_n, \mathcal{F}_n : n \geqslant 0\}$ be a (super)martingale, and let τ be an $\{\mathcal{F}_n\}$-stopping time. We first consider the stopped process $X_n^\tau \triangleq X_{\tau\wedge n}$.

As in the proof of Doob's upcrossing inequality, we can interpret X_n^τ by a gambling model. The model in this case is very easy: we keep playing unit stakes from the beginning and quit immediately after τ. Mathematically, set

$$C_n^\tau = \mathbf{1}_{\{n\leqslant\tau\}}, \quad n \geqslant 1.$$

Then $(C^\tau \bullet X)_n = X_{\tau\wedge n} - X_0$ (recall that $(C^\tau \bullet X)_n$ represents the total winning up to time n). Apparently, the sequence C_n^τ is bounded, non-negative and $\{\mathcal{F}_n\}$-predictable. According to Theorem 3.1, we have proved the following result.

Theorem 3.5. *The stopped process X_n^τ is an $\{\mathcal{F}_n\}$-(super)martingale.*

Now we consider the situation when we also stop our filtration at some stopping time. We first consider the case in which the stopping times are bounded.

Theorem 3.6. *Let $\{X_n, \mathcal{F}_n : n \geqslant 0\}$ be a (super)martingale. Suppose that σ, τ are two bounded $\{\mathcal{F}_n\}$-stopping times such that $\sigma \leqslant \tau$. Then $\{X_\sigma, \mathcal{F}_\sigma; X_\tau, \mathcal{F}_\tau\}$ is a two-step (super)martingale.*

In particular, if τ is an $\{\mathcal{F}_n\}$-stopping time, then $\{X_{\tau\wedge n}, \mathcal{F}_{\tau\wedge n} : n \geqslant 0\}$ is a (super)martingale.

Proof. We only consider the supermartingale case. Assume that $\sigma \leqslant \tau \leqslant N$ for some constant $N \geqslant 0$. Adaptedness and integrability of $\{X_\sigma, \mathcal{F}_\sigma; X_\tau, \mathcal{F}_\tau\}$ are obvious. To see the supermartingale property, let $F \in \mathcal{F}_\sigma$. Consider the sequence

$$C_n = \mathbf{1}_F \mathbf{1}_{\{\sigma<n\leqslant\tau\}}, \quad n \geqslant 1.$$

Then $(C \bullet X)_N = (X_\tau - X_\sigma)\mathbf{1}_F$. On the other hand, C_n is $\{\mathcal{F}_n\}$-predictable because

$$F \bigcap \{\sigma < n \leqslant \tau\} = F \bigcap \{\sigma \leqslant n-1\} \bigcap (\tau \leqslant n-1)^c \in \mathcal{F}_{n-1}, \quad \forall n \geqslant 1.$$

It is also bounded and non-negative. Therefore, according to Theorem 3.1, $\{(C \bullet X)_n, \mathcal{F}_n : n \geqslant 0\}$ is a supermartingale. In particular,

$$\mathbb{E}[(C \bullet X)_N] = \mathbb{E}[(X_\tau - X_\sigma)\mathbf{1}_F] \leqslant 0,$$

which is the desired supermartingale property. □

The case when σ, τ are unbounded is more involved.

In general, since a stopping time τ can be infinite, the definition of X_τ involves its value at ∞. Therefore, a natural assumption on our (super)martingale is to included a "last" element X_∞.

Definition 3.5. A (super)martingale *with a last element* is a (super)martingale $\{X_t, \mathcal{F}_t : t \in T\}$ over the index set $T = \{0, 1, 2, \cdots\} \cup \{\infty\}$.

According to Lévy's forward theorem (c.f. Theorem 3.1), a martingale $\{X_n, \mathcal{F}_n : 0 \leqslant n \leqslant \infty\}$ with a last element is uniformly integrable and $X_n \to X_\infty$ almost surely and in L^1 as $n \to \infty$.

The general optional sample theorem for martingales is easy.

Theorem 3.7. *Let $\{X_n, \mathcal{F}_n : \ 0 \leqslant n \leqslant \infty\}$ be a martingale with a last element. Suppose that σ, τ are two $\{\mathcal{F}_n\}$-stopping times such that $\sigma \leqslant \tau$. Then $\{X_\sigma, \mathcal{F}_\sigma; X_\tau, \mathcal{F}_\tau\}$ is a two-step martingale.*

Proof. Adaptedness is easy. Integrability follows from

$$\begin{aligned}
\mathbb{E}[|X_\tau|] &= \sum_{n=0}^{\infty} \mathbb{E}\left[|X_n| \mathbf{1}_{\{\tau=n\}}\right] + \mathbb{E}\left[|X_\infty| \mathbf{1}_{\{\tau=\infty\}}\right] \\
&\leqslant \sum_{n=0}^{\infty} \mathbb{E}\left[|X_\infty| \mathbf{1}_{\{\tau=n\}}\right] + \mathbb{E}\left[|X_\infty| \mathbf{1}_{\{\tau=\infty\}}\right] \\
&= \mathbb{E}[|X_\infty|],
\end{aligned}$$

where we have used the fact that $\{|X_n|, \mathcal{F}_n : \ 0 \leqslant n \leqslant \infty\}$ is a submartingale with a last element.

Now we show that

$$\mathbb{E}[X_\infty | \mathcal{F}_\tau] = X_\tau \ \ \text{a.s.}$$

The martingale property will follow from further conditioning on $\mathcal{F}_\sigma$. Let $A \in \mathcal{F}_\tau$. For every $n \geqslant 0$, we have

$$\mathbb{E}\left[X_\tau \mathbf{1}_{A \cap \{\tau \leqslant n\}}\right] = \sum_{k=0}^{n} \mathbb{E}\left[X_k \mathbf{1}_{A \cap \{\tau = k\}}\right] = \mathbb{E}\left[X_\infty \mathbf{1}_{A \cap \{\tau \leqslant n\}}\right].$$

Since X_τ and X_∞ are both integrable, by the dominated convergence theorem, we have

$$\mathbb{E}\left[X_\tau \mathbf{1}_{A \cap \{\tau < \infty\}}\right] = \mathbb{E}\left[X_\infty \mathbf{1}_{A \cap \{\tau < \infty\}}\right].$$

But the identity over $\{\tau = \infty\}$ is obvious. Therefore, we have

$$\mathbb{E}[X_\tau \mathbf{1}_A] = \mathbb{E}[X_\infty \mathbf{1}_A].$$

□

To study the case for supermartingales, we need the following lemma.

Lemma 3.1. *Every supermartingale with a last element can be written as the sum of a martingale with a last element and a non-negative supermartingale with zero last element.*

Proof. Let $\{X_n, \mathcal{F}_n : 0 \leqslant n \leqslant \infty\}$ be a supermartingale with a last element. Define

$$Y_n = \mathbb{E}[X_\infty | \mathcal{F}_n], \ Z_n = X_n - Y_n, \quad 0 \leqslant n \leqslant \infty.$$

Then $X_n = Y_n + Z_n$ is the desired decomposition. □

Now we are able to prove the general optional sampling theorem for supermartingales.

Theorem 3.8. *Let $\{X_n, \mathcal{F}_n : 0 \leqslant n \leqslant \infty\}$ be a supermartingale with a last element. Suppose that σ, τ are two $\{\mathcal{F}_n\}$-stopping times such that $\sigma \leqslant \tau$. Then $\{X_\sigma, \mathcal{F}_\sigma; X_\tau, \mathcal{F}_\tau\}$ is a two-step supermartingale.*

Proof. According to Theorem 3.7 and Lemma 3.1, it suffices to consider the case when $\{X_n, \mathcal{F}_n : 0 \leqslant n \leqslant \infty\}$ is a non-negative supermartingale with a last element $X_\infty = 0$.

Adaptedness is easy. To see integrability, since

$$X_\sigma = X_\sigma \mathbf{1}_{\{\sigma < \infty\}} + X_\infty \mathbf{1}_{\{\sigma = \infty\}}$$

and X_∞ is integrable, we only need to show that $X_\sigma \mathbf{1}_{\{\sigma < \infty\}}$ is integrable. But

$$X_\sigma \mathbf{1}_{\{\sigma < \infty\}} = \lim_{n \to \infty} X_{\sigma \wedge n} \mathbf{1}_{\{\sigma < \infty\}},$$

and according to Theorem 3.6, we know that

$$\mathbb{E}\left[X_{\sigma \wedge n} \mathbf{1}_{\{\sigma < \infty\}}\right] \leqslant \mathbb{E}[X_{\sigma \wedge n}] \leqslant \mathbb{E}[X_0].$$

The integrability of $X_\sigma \mathbf{1}_{\{\sigma < \infty\}}$ then follows from Fatou's lemma.

Now we show the supermartingale property. Let $A \in \mathcal{F}_\sigma$. According to Proposition 2.4, for every $n \geqslant 0$, $A \cap \{\sigma \leqslant n\} \in \mathcal{F}_\sigma \cap \mathcal{F}_n = \mathcal{F}_{\sigma \wedge n}$. From Theorem 3.6, we know that $\{X_{\sigma \wedge n}, \mathcal{F}_{\sigma \wedge n}; X_{\tau \wedge n}, \mathcal{F}_{\tau \wedge n}\}$ is a two-step supermartingale. Therefore,

$$\int_{A \cap \{\sigma \leqslant n\}} X_{\tau \wedge n} d\mathbb{P} \leqslant \int_{A \cap \{\sigma \leqslant n\}} X_{\sigma \wedge n} d\mathbb{P} = \int_{A \cap \{\sigma \leqslant n\}} X_\sigma d\mathbb{P}.$$

Moreover, we know that

$$\int_{A \cap \{\sigma \leqslant n\}} X_{\tau \wedge n} d\mathbb{P} \geqslant \int_{A \cap \{\sigma \leqslant n\} \cap \{\tau < \infty\}} X_{\tau \wedge n} d\mathbb{P}$$

as X_n is non-negative, and we also have

$$X_\tau \mathbf{1}_{A \cap \{\tau < \infty\}} = \lim_{n \to \infty} X_{\tau \wedge n} \mathbf{1}_{A \cap \{\sigma \leqslant n\} \cap \{\tau < \infty\}}.$$

Fatou's lemma then implies that

$$\int_{A \cap \{\tau < \infty\}} X_\tau d\mathbb{P} \leqslant \lim_{n \to \infty} \int_{A \cap \{\sigma \leqslant n\}} X_\sigma d\mathbb{P} \leqslant \int_A X_\sigma d\mathbb{P}.$$

But the left hand side of the above inequality is equal to $\int_A X_\tau d\mathbb{P}$ since $X_\infty = 0$. This yields the desired supermartingale property. □

Corollary 3.2. *Let $\{X_n, \mathcal{F}_n : 0 \leqslant n \leqslant \infty\}$ be a (super)martingale with a last element. Suppose that $\{\tau_m : m \geqslant 1\}$ is a increasing sequence of $\{\mathcal{F}_n\}$-stopping times. Define $\widetilde{X}_m = X_{\tau_m}$ and $\widetilde{\mathcal{F}}_m = \mathcal{F}_{\tau_m}$. Then $\{\widetilde{X}_m, \widetilde{\mathcal{F}}_m : m \geqslant 1\}$ is a (super)martingale.*

3.4 Doob's martingale inequalities

By using Doob's optional sampling theorem for bounded stopping times, we are going to derive several fundamental inequalities in martingale theory which are important in the analytic aspect of stochastic calculus.

Here we will work with submartingales instead.

The central inequality is known as Doob's maximal inequality, which is the first part of the following result. As a submartingale exhibits an increasing trend on average, it is not surprising that its running maximum can be controlled by the terminal value in some sense.

Theorem 3.9. *Let $\{X_n, \mathcal{F}_n : n \geqslant 0\}$ be a submartingale. Then for every $N \geqslant 0$ and $\lambda > 0$, we have the following inequalities:*

(1) $\lambda \mathbb{P}\left(\sup_{0\leqslant n\leqslant N} X_n \geqslant \lambda\right) \leqslant \mathbb{E}[X_N^+]$;

(2) $\lambda \mathbb{P}\left(\inf_{0\leqslant n\leqslant N} X_n \leqslant -\lambda\right) \leqslant \mathbb{E}[X_N^+] - \mathbb{E}[X_0]$.

Proof. (1) Let $\sigma = \inf\{n \leqslant N : X_n \geqslant \lambda\}$ and we define $\sigma = N$ if no such $n \leqslant N$ exists. Clearly σ is an $\{\mathcal{F}_n\}$-stopping time bounded by N. According to Theorem 3.6, we have

$$\begin{aligned} \mathbb{E}[X_N] &\geqslant \mathbb{E}[X_\sigma] = \mathbb{E}\left[X_\sigma \mathbf{1}_{\{\sup_{0\leqslant n\leqslant N} X_n \geqslant \lambda\}}\right] + \mathbb{E}\left[X_N \mathbf{1}_{\{\sup_{0\leqslant n\leqslant N} X_n < \lambda\}}\right] \\ &\geqslant \lambda \mathbb{P}\left(\left\{\sup_{0\leqslant n\leqslant N} X_n \geqslant \lambda\right\}\right) + \mathbb{E}\left[X_N \mathbf{1}_{\{\sup_{0\leqslant n\leqslant N} X_n < \lambda\}}\right]. \end{aligned}$$

Therefore,

$$\lambda \mathbb{P}\left(\left\{\sup_{0\leqslant n\leqslant N} X_n \geqslant \lambda\right\}\right) \leqslant \mathbb{E}\left[X_N \mathbf{1}_{\{\sup_{0\leqslant n\leqslant N} X_n \geqslant \lambda\}}\right] \leqslant \mathbb{E}[X_N^+].$$

The desired inequality then follows.

(2) Let $\tau = \inf\{n \leqslant N : X_n \leqslant -\lambda\}$ and we define $\tau = N$ if no such $n \leqslant N$ exists. Then τ is an $\{\mathcal{F}_n\}$-stopping time bounded by N. The desired inequality follows in a similar manner by considering $\mathbb{E}[X_0] \leqslant \mathbb{E}[X_\tau]$. □

An important corollary of Doob's maximal inequality is Doob's L^p-inequality for the maximal functional. We first need the following lemma.

Lemma 3.2. *Suppose that X, Y are two non-negative random variables such that*

$$\mathbb{P}(X \geqslant \lambda) \leqslant \frac{\mathbb{E}[Y \mathbf{1}_{\{X\geqslant\lambda\}}]}{\lambda}, \quad \forall \lambda > 0. \tag{3.4}$$

Then for any $p > 1$, we have

$$\|X\|_p \leqslant q\|Y\|_p, \tag{3.5}$$

where $q \triangleq p/(p-1)$ (so that $1/p + 1/q = 1$).

Proof. Suppose $\|Y\|_p < \infty$ (otherwise the result is trivial). Write

$$\mathbb{E}[X^p] = \mathbb{E}\left[\int_0^X p\lambda^{p-1}d\lambda\right] = \mathbb{E}\left[\int_0^\infty p\lambda^{p-1}\mathbf{1}_{\{X\geqslant\lambda\}}d\lambda\right].$$

By Fubini's theorem, we have

$$\begin{aligned}
\mathbb{E}\left[\int_0^\infty p\lambda^{p-1}\mathbf{1}_{\{X\geqslant\lambda\}}d\lambda\right] &= \int_0^\infty p\lambda^{p-1}\mathbb{P}(X\geqslant\lambda)d\lambda \\
&\leqslant \int_0^\infty p\lambda^{p-2}\mathbb{E}\left[Y\mathbf{1}_{\{X\geqslant\lambda\}}\right]d\lambda \\
&= \mathbb{E}\left[Y\int_0^X p\lambda^{p-2}d\lambda\right] \\
&= \frac{p}{p-1}\mathbb{E}[YX^{p-1}].
\end{aligned}$$

First we assume that $\|X\|_p < \infty$. It follows from Hölder's inequality that

$$\mathbb{E}[YX^{p-1}] \leqslant \|Y\|_p\|X^{p-1}\|_q = \|Y\|_p\|X\|_p^{p-1}.$$

Therefore, (3.5) follows.

For the general case, let $X^N = X \wedge N$ ($N \geqslant 1$). By considering $\lambda > N$ and $\lambda \leqslant N$, it is easy to see that the condition (3.4) also holds for X^N and Y. The desired inequality (3.5) follows by first considering X^N and then applying the monotone convergence theorem. □

Corollary 3.3 (Doob's L^p-inequality). *Let $\{X_n, \mathcal{F}_n : n \geqslant 0\}$ be a non-negative submartingale. Suppose that $p > 1$ and $X_n \in L^p$ for all n. Then for every $N \geqslant 0$, $X_N^* \triangleq \sup_{0\leqslant n\leqslant N} X_n \in L^p$, and we have the following inequality:*

$$\|X_N^*\|_p \leqslant q\|X_N\|_p,$$

where $q \triangleq p/(p-1)$.

Proof. The result follows from the first part of Theorem 3.9 and Lemma 3.2. □

3.5 The continuous time analogue

The key to the passage from discrete to continuous time is an additional assumption on the right continuity for sample paths. With this right continuity assumption, the generalizations of all the results in Section 3.1, 3.2 and 3.3 to the continuous time setting are almost straight forward. It will be seen in Theorem 3.10 that this is not at all a luxurious assumption.

Suppose that $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t : t \geqslant 0\})$ is a filtered probability space, and all stochastic processes are defined on $[0,\infty)$ (except in the backward case in which the parameter set is $(-\infty, -1]$). In the following discussion, we always assume that *the underlying (sub or super)martingales have right continuous sample paths.*

1. The martingale convergences theorems: Theorem 3.2, Theorem 3.3, Corollary 3.1 and Theorem 3.4 (just for backward martingales) also hold in the continuous time setting.

Proof. Indeed, the only place which needs care is the definition of upcrossing numbers. Suppose that X_t is an $\{\mathcal{F}_t\}$-adapted stochastic process. Let $a < b$ be two real numbers. For a finite subset $F \subseteq [0, \infty)$, we define $U_F(X; [a, b])$ to be the upcrossing number of $[a, b]$ by the process $\{X_t : t \in F\}$, defined in the same way as in the discrete time case. For a general subset $I \subseteq [0, \infty)$, set

$$U_I(X; [a, b]) = \sup\{U_F(X; [a, b]) : F \subseteq I, \ F \text{ is finite}\}.$$

If X_t has right continuous sample paths, we may approximate $U_{[0,n]}(X; [a, b])$ by rational time indices to conclude that $U_{[0,n]}(X; [a, b])$ and $U_{[0,\infty)}(X; [a, b])$ are measurable. The remaining details in proving the continuous time analogue of these convergence results are then obvious. □

2. Doob's optional sampling theorems: Theorem 3.6 and Theorem 3.8 also hold in the continuous time setting.

Proof. We only prove the analogue of Theorem 3.8. The case for bounded stopping times is treated in a similar way.

Suppose that $\{X_t, \mathcal{F}_t : 0 \leqslant t \leqslant \infty\}$ is a supermartingale with a last element. Let $\sigma \leqslant \tau$ be two $\{\mathcal{F}_t\}$-stopping times. For $n \geqslant 1$, define

$$\sigma_n = \sum_{k=1}^{\infty} \frac{k}{2^n} \mathbf{1}_{\{\frac{k-1}{2^n} \leqslant \sigma < \frac{k}{2^n}\}} + \sigma \cdot \mathbf{1}_{\{\sigma=\infty\}}$$

and define τ_n in the same way. It is apparent that $\sigma_n \leqslant \tau_n$, and $\sigma_n \downarrow \sigma$, $\tau_n \downarrow \tau$. Moreover, given $t \geqslant 0$, let k be the unique integer such that $t \in [(k-1)/2^n, k/2^n)$. From $\{\sigma_n \leqslant t\} = \{\sigma < (k-1)/2^n\}$, we can see that σ_n is an $\{\mathcal{F}_t\}$-stopping time and the same is true for τ_n.

Since σ_n, τ_n take discrete values $\{k/2^n : k \geqslant 1\} \cup \{\infty\}$, we can apply Theorem 3.8 to conclude that

$$\mathbb{E}[X_{\tau_n} | \mathcal{F}_{\sigma_n}] \leqslant X_{\sigma_n}, \ \forall n \geqslant 1.$$

In other words,

$$\int_A X_{\tau_n} d\mathbb{P} \leqslant \int_A X_{\sigma_n} d\mathbb{P}, \quad \forall A \in \mathcal{F}_{\sigma_n}, \ n \geqslant 1. \tag{3.6}$$

By Proposition 2.4, we know that $\mathcal{F}_\sigma \subseteq \mathcal{F}_{\sigma_n}$ and hence (3.6) is true for every $A \in \mathcal{F}_\sigma$.

Now a key observation is that $\{X_{\sigma_n}, \mathcal{F}_{\sigma_n} : n \geqslant 1\}$ is a backward supermartingale with $\sup_{n \geqslant 1} \mathbb{E}[X_{\sigma_n}] < \infty$. This follows from the fact that $\sigma_{n+1} \leqslant \sigma_n$ and $\mathbb{E}[X_{\sigma_n}] \leqslant \mathbb{E}[X_0]$ for all n. The same is true for $\{X_{\tau_n}, \mathcal{F}_{\tau_n} : n \geqslant 1\}$. Under the assumption that X_t has right continuous sample paths (so that $X_{\sigma_n} \to X_\sigma$, $X_{\tau_n} \to X_\tau$ as $n \to \infty$), Theorem 3.4 enables us to conclude that $X_\sigma, X_\tau \in L^1$ and to take limit on both sides of (3.6) for every $A \in \mathcal{F}_\sigma$. Therefore, $\{X_\sigma, \mathcal{F}_\sigma; X_\tau, \mathcal{F}_\tau\}$ is a two-step supermartingale. □

Remark 3.4. Although here the backward supermartingale is indexed by positive integers, the reader should easily find it equivalent with having negative time parameter as in Theorem 3.4 by setting $m = -n$ for $n \geqslant 1$.

3. **Doob's martingale inequalities:** Theorem 3.9 and Corollary 3.3 also hold in the continuous time setting.

Proof. The right continuity of sample paths implies that

$$\sup_{t\in[0,N]} X_t = \sup_{t\in[0,N]\cap\mathbb{Q}} X_t,$$

and the same is true for the infimum. Now the results follow easily. □

Finally, we demonstrate that we can basically only work with right continuous (sub or super)martingales without much loss of generality. We can also see how the usual conditions for filtration (c.f. Definition 2.11) come in naturally.

Definition 3.6. A function $x : [0,\infty) \to \mathbb{R}$ is called *càdlàg* if it is right continuous with left limits everywhere.

Definition 3.7. A function $x : \mathbb{Q}^+ \to \mathbb{R}$ is called *regularizable* if

$$\lim_{q\downarrow t} x_q \text{ exists finitely for every } t \geqslant 0$$

and

$$\lim_{q\uparrow t} x_q \text{ exists finitely for every } t > 0.$$

The following classical fact about real functions is important.

Lemma 3.3. *Let $x : \mathbb{Q}^+ \to \mathbb{R}$ be a real function. Suppose that for every $N \in \mathbb{N}$, $a, b \in \mathbb{Q}$ with $a < b$, we have*

$$\sup_{q\in\mathbb{Q}^+\cap[0,N]} |x_q| < \infty \text{ and } U_{\mathbb{Q}^+\cap[0,N]}(x;[a,b]) < \infty,$$

where $U_{[0,N]}(x;[a,b])$ is the upcrossing number of $[a,b]$ by $x|_{\mathbb{Q}^+\cap[0,N]}$. Then the function x is regularizable. Moreover, the regularization $\widetilde{x}$ of x, defined by

$$\widetilde{x}_t = \lim_{q\downarrow t} x_q, \quad t \geqslant 0,$$

is a càdlàg function on $[0,\infty)$.

Proof. Suppose on the contrary that at some $t \geqslant 0$ we have

$$\liminf_{q\downarrow t} x_q < \limsup_{q\downarrow t} x_q.$$

Then there exist $a < b \in \mathbb{Q}$, such that $U_{[0,N]}(x;[a,b]) = \infty$ for every $N > t$, which is a contradiction. Therefore, $\lim_{q\downarrow t} x_q$ exists in $[-\infty,\infty]$. The boundedness assumption guarantees the finiteness of this limit. The case of $q \uparrow t$ is treated in the same way. Therefore, x is regularizable.

Now suppose $t_n \downarrow t \geqslant 0$. We choose $q_n \in (t_{n+1}, t_n)$ be such that $|x_{q_n} - \widetilde{x}_{t_{n+1}}| \leqslant 1/n$. It follows that $q_n \downarrow t$ and hence $\widetilde{x}_{t_n} \to \widetilde{x}_t$. Similarly, we can show that $\lim_{s\uparrow t} \widetilde{x}_s = \lim_{q\uparrow t} x_q$ for every $t > 0$. Therefore, $\widetilde{x}$ is a càdlàg function. □

Recall that the usual augmentation of a filtered probability space $(\Omega, \mathcal{G}, \mathbb{P}; \{\mathcal{G}_t\})$ is given by $\mathcal{F}_t = \sigma(\mathcal{G}_{t+}, \mathcal{N})$, where $\mathcal{N}$ is the collection of $\mathbb{P}$-null sets. Now we have the following regularization theorem due to Doob.

Theorem 3.10. *Let $\{X_t, \mathcal{G}_t\}$ be a supermartingale defined over a filtered probability space $(\Omega, \mathcal{G}, \mathbb{P}; \{\mathcal{G}_t\})$. Then:*

(1) almost every sample path of X_t, when restricted to $\mathbb{Q}^+$, is regularizable;

(2) the regularization $\widetilde{X}_t$ of X_t, defined as in Lemma 3.3, is a supermartingale with respect to the usual augmentation $\{\mathcal{F}_t\}$ of $(\Omega, \mathcal{G}, \mathbb{P}; \{\mathcal{G}_t\})$;

(3) $\widetilde{X}_t$ is a modification of X_t if and only if X_t is right continuous in L^1, i.e.

$$\lim_{t\downarrow s} \mathbb{E}[|X_t - X_s|] = 0, \quad \forall s \geqslant 0.$$

Proof. (1) According to Lemma 3.3, it suffices to show that, for any given $N \in \mathbb{N}$, $a, b \in \mathbb{Q}$ with $a < b$, with probability one we have

$$\sup_{q\in\mathbb{Q}^+\cap[0,N]} |X_q| < \infty, \ U_{\mathbb{Q}^+\cap[0,N]}(X; [a,b]) < \infty.$$

Indeed, let Q_n be an increasing sequence of finite subsets of $\mathbb{Q}^+ \cap [0, N]$ containing $\{0, N\}$, such that $\cup_{n=1}^{\infty} Q_n = \mathbb{Q}^+ \cap [0, N]$. According to Theorem 3.9, we have

$$\mathbb{P}\left(\sup_{q\in\mathbb{Q}^+\cap[0,N]} |X_q| > \lambda\right) = \lim_{n\to\infty} \mathbb{P}\left(\sup_{q\in Q_n} |X_q| > \lambda\right) \leqslant \frac{2}{\lambda}\mathbb{E}[X_N^+] + \frac{1}{\lambda}\mathbb{E}[X_0^-],$$

and according to Doob's upcrossing inequality (3.2), we have

$$\mathbb{E}[U_{\mathbb{Q}^+\cap[0,N]}(X; [a,b])] = \lim_{n\to\infty} \mathbb{E}[U_{Q_n}(X; [a,b])] \leqslant \frac{\mathbb{E}[|X_N|] + |a|}{b-a}.$$

The result then follows.

(2) Define $\widetilde{X}_t$ as in Lemma 3.3 on the set where X_t is regularizable and set $\widetilde{X} \equiv 0$ otherwise. From the construction it is apparent that $\widetilde{X}_t$ is $\{\mathcal{F}_t\}$-adapted. Now suppose $s < t$, and let $p_n < q_n \in \mathbb{Q}^+$ be such that $p_n \downarrow s, q_n \downarrow t$. It follows that $\{X_{p_n}, \mathcal{G}_{p_n}\}$ is a backward supermartingale with $\sup_n \mathbb{E}[X_{p_n}] < \infty$, and the same is true for $\{X_{q_n}, \mathcal{G}_{q_n}\}$. Therefore,

$$X_{p_n} \to \widetilde{X}_s, \ X_{q_n} \to \widetilde{X}_t,$$

almost surely and in L^1 as $n \to \infty$. In particular, $\widetilde{X}_s$ and $\widetilde{X}_t$ are integrable. The supermartingale property follows by taking limit in the following inequality:

$$\int_A X_{q_n} d\mathbb{P} \leqslant \int_A X_{p_n} d\mathbb{P}, \quad \forall A \in \mathcal{G}_{s+}.$$

(3) Necessity. We only need a weaker assumption that X has a right continuous modification $\widehat{X}$. In this case, given $t_0 \geqslant 0$, let t_n $(n \geqslant 1)$ be an arbitrary sequence such that $t_n \downarrow t_0$. Then with probability one, $\widehat{X}_{t_n} = X_{t_n}$ for all $n \geqslant 0$. Since $\widehat{X}_t$ has right continuous sample paths, we obtain that $X_{t_n} \to X_{t_0}$ almost surely. On the other hand, the same backward supermartingale argument as in the second part implies that $X_{t_n} \to X_{t_0}$ in L^1. Therefore, X_t is right continuous in L^1.

Sufficiency. Given $t \geqslant 0$, let $q_n \in \mathbb{Q}^+$ be such that $q_n \downarrow t$. Then $X_{q_n} \to \widetilde{X}_t$ in L^1. Since X_{q_n} also converges to X_t in L^1 by assumption, we conclude that $\widetilde{X}_t = X_t$. □

Remark 3.5. By adjusting the proof slightly, one can show that if $\{X_t, \mathcal{G}_t\}$ is a supermartingale with right continuous sample paths, then almost every sample path of X_t also has left limits everywhere.

If we start with a filtration satisfying the usual conditions, we have the following very nice and stronger result.

Theorem 3.11. *Let $\{X_t, \mathcal{F}_t\}$ be a supermartingale defined over a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ which satisfies the usual conditions. Then X has a càdlàg modification $\widetilde{X}$ if and only if the function $t \mapsto \mathbb{E}[X_t]$ is right continuous. Moreover, in this case $\left\{\widetilde{X}_t, \mathcal{F}_t\right\}$ is also a supermartingale.*

Proof. Necessity follows from the same argument as in the proof of Theorem 3.10, (3). For the sufficiency, let $\widetilde{X}$ be the regularization of X given by Theorem 3.10. Then for $t \geqslant 0$, $q_n \in \mathbb{Q}^+$ with $q_n > t$, we have

$$\int_A X_{q_n} d\mathbb{P} \leqslant \int_A X_t d\mathbb{P}, \quad \forall A \in \mathcal{F}_t.$$

By letting $q_n \downarrow t$, we conclude that $\mathbb{E}\left[\widetilde{X}_t | \mathcal{F}_t\right] \leqslant X_t$ a.s. But $\widetilde{X}_t$ is $\{\mathcal{F}_t\}$-adapted since $\{\mathcal{F}_t\}$ satisfies the usual conditions. Therefore, $\widetilde{X}_t \leqslant X_t$ a.s. But from the right continuity of $t \mapsto \mathbb{E}[X_t]$, we see that $\mathbb{E}\left[\widetilde{X}_t\right] = \mathbb{E}[X_t]$. Therefore, $\widetilde{X}_t = X_t$ a.s. Finally, it is trivial that every modification of X is also an $\{\mathcal{F}_t\}$-supermartingale. □

It follows from Theorem 3.11 that every martingale has a càdlàg modification, provided that the underlying filtration satisfies the usual conditions.

3.6 The Doob-Meyer decomposition

Now we discuss a result which is fundamental in the study of stochastic integration and lies in the heart of continuous time martingale theory. This will also be the first time that the continuous time situation becomes substantially harder than the discrete time setting.

Roughly speaking, the intuition behind the whole discussion can be summarized as: the tendency of increase for a submartingale can be extracted in a pathwise way, and what remains is a martingale part.

As before, we first consider the easy bit: the discrete time situation. This is known as Doob's decomposition.

Definition 3.8. An *increasing sequence* $\{A_n : n \geqslant 0\}$ over a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_n\})$ is an $\{\mathcal{F}_n\}$-adapated sequence such that with probability one we have $0 = A_0(\omega) \leqslant A_1(\omega) \leqslant A_2(\omega) \leqslant \cdots$, and $\mathbb{E}[A_n] < \infty$ for all n.

Theorem 3.12 (Doob's decomposition). *Let $\{X_n, \mathcal{F}_n : n \geqslant 0\}$ be a submartingale defined over a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_n\})$. Then X has a decomposition*

$$X_n = M_n + A_n, \quad n \geqslant 0,$$

where $\{M_n, \mathcal{F}_n\}$ is a martingale, $\{A_n, \mathcal{F}_n\}$ is an increasing sequence which is $\{\mathcal{F}_n\}$-predictable. Moreover, such a decomposition is unique with probability one.

Proof. We first show uniqueness. Suppose that X_n has such a decomposition. Then

$$X_n - X_{n-1} = M_n - M_{n-1} + A_n - A_{n-1}.$$

Since M_n is an $\{\mathcal{F}_n\}$-martingale and A_n is $\{\mathcal{F}_n\}$-predictable, we have

$$\mathbb{E}[X_n - X_{n-1} | \mathcal{F}_{n-1}] = A_n - A_{n-1}.$$

Therefore,

$$A_n = \sum_{k=1}^{n} \mathbb{E}[X_k - X_{k-1} | \mathcal{F}_{k-1}]. \tag{3.7}$$

Existence follows from defining A_n by (3.7) and $M_n \triangleq X_n - A_n$. □

Remark 3.6. Predictability is an important condition for the uniqueness of Doob's decomposition. Indeed, if $\{M_n, \mathcal{F}_n\}$ is a square integrable martingale with $M_0 = 0$, then

$$M_n^2 = V_n + B_n$$

where $B_n = \sum_{k=1}^{n} (M_k - M_{k-1})^2$ is another decomposition of the submartingale M_n^2 into a martingale part V_n and an increasing sequence B_n. However, B_n here is not $\{\mathcal{F}_n\}$-predictable.

To understand the continuous time analogue of Theorem 3.12, we first need to recapture the predictability property in a way which extends to the continuous time case naturally.

Definition 3.9. An increasing sequence A_n defined over some filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_n\})$ is called *natural* if

$$\sum_{k=1}^{n} \mathbb{E}[m_k(A_k - A_{k-1})] = \sum_{k=1}^{n} \mathbb{E}[m_{k-1}(A_k - A_{k-1})], \quad \forall n \geqslant 1, \tag{3.8}$$

for every bounded martingale $\{m_n, \mathcal{F}_n\}$.

Note that from the martingale property, the left hand side of (3.8) is equal to $\mathbb{E}[m_nA_n]$. Moreover, simple calculation yields

$$\sum_{k=1}^{n} m_{k-1}(A_k - A_{k-1}) = m_nA_n - \sum_{k=1}^{n} A_k(m_k - m_{k-1}) = m_nA_n - (A \bullet m)_n,$$

where $(A \bullet m)_n$ is the martingale transform of m_n by A_n. Therefore, A_n is natural if and only if

$$\mathbb{E}[(A \bullet m)_n] = 0, \quad \forall n \geqslant 1,$$

for every bounded martingale $\{m_n, \mathcal{F}_n\}$.

Now we have the following simple fact.

Proposition 3.3. *Suppose that $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets. Then an increasing sequence A_n is $\{\mathcal{F}_n\}$-predictable if and only if it is natural.*

Proof. Suppose that A_n is $\{\mathcal{F}_n\}$-predictable. Given bounded martingale $\{m_n, \mathcal{F}_n\}$, according to Theorem 3.1, we know that $\{(A \bullet m)_n, \mathcal{F}_n\}$ is a martingale null at 0. Therefore, A_n is natural.

Conversely, suppose that A_n is natural and hence we know that

$$\mathbb{E}[A_n(m_n - m_{n-1})] = 0, \quad \forall n \geqslant 1,$$

for every bounded martingale $\{m_n, \mathcal{F}_n\}$. It follows that for every $n \geqslant 1$,

$$\begin{aligned}
&\mathbb{E}[m_n(A_n - \mathbb{E}[A_n|\mathcal{F}_{n-1}])] \\
&= \mathbb{E}[m_nA_n] - \mathbb{E}[m_n\mathbb{E}[A_n|\mathcal{F}_{n-1}]] \\
&= \mathbb{E}[m_nA_n] - \mathbb{E}[A_nm_{n-1}] \quad \text{(by Problem Sheet 1, Problem 1, (1))} \\
&= 0. \qquad (3.9)
\end{aligned}$$

Now for fixed $n \geqslant 1$, define $Z = \operatorname{sgn}(A_n - \mathbb{E}[A_n|\mathcal{F}_{n-1}])$, and set $m_k = \mathbb{E}[Z|\mathcal{F}_k]$ if $k \leqslant n$ and $m_k = Z$ if $k > n$. It follows from $\{m_n, \mathcal{F}_n\}$ is a bounded martingale, and from (3.9) we know that $\mathbb{E}[|A_n - \mathbb{E}[A_n|\mathcal{F}_{n-1}]|] = 0$. Therefore, $A_n = \mathbb{E}[A_n|\mathcal{F}_{n-1}]$ almost surely, which implies that A_n is $\{\mathcal{F}_n\}$-predictable. □

Now we discuss the continuous time situation. In the rest of this subsection, we will always work over a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t : t \geqslant 0\})$ which satisfies the usual conditions.

To study the corresponding decomposition, we first need the analogue of increasing sequences.

Definition 3.10. A *increasing process* $\{A_t : t \geqslant 0\}$ is an $\{\mathcal{F}_t\}$-adapted process A_t such that with probability one we have $A_0(\omega) = 0$ and $t \mapsto A_t(\omega)$ is increasing and right continuous, and $\mathbb{E}[A_t] < \infty$ for all $t \geqslant 0$.

Definition 3.11. An increasing process A_t is called *natural* if for every bounded and càdlàg martingale $\{m_t, \mathcal{F}_t\}$, we have

$$\mathbb{E}\left[\int_0^t m_s dA_s\right] = \mathbb{E}\left[\int_0^t m_{s-} dA_s\right], \quad \forall t \geqslant 0, \tag{3.10}$$

where the integrals inside the expectations are understood in the Lebesgue-Stieltjes sense.

Note that every continuous, increasing process is natural, since a càdlàg function can have at most countably many jumps. Moreover, the left hand side of (3.10) is equal to $\mathbb{E}[m_t A_t]$. Indeed, let $\mathcal{P}: \ 0 = t_0 < t_1 < \cdots < t_n = t$ be an arbitrary finite partition of $[0, t]$. Then as in the discrete time case, we see that

$$\sum_{k=1}^n \mathbb{E}[m_{t_k}(A_{t_k} - A_{t_{k-1}})] = \mathbb{E}[m_t A_t]. \tag{3.11}$$

Since m_t is right continuous, by the dominated convergence theorem, the left hand side of (3.11) converges to $\mathbb{E}\left[\int_0^t m_s dA_s\right]$ as $\mathrm{mesh}(\mathcal{P}) \to 0$.

Remark 3.7. In the continuous time setting, there is also a notion of predictability which is crucial for the study of stochastic calculus for processes with jumps. This is technically much more complicated than the discrete time case. Under this notion of predictability, it can be shown that an increasing process is natural if and only if it is predictable (c.f. [1]).

Unlike discrete time submartingales, *not* every càdlàg submartingale has a Doob-type decomposition (see Problem Sheet 5, Problem 2 for a counterexample). We first examine what condition should the submartingale satisfy if such a decomposition exists.

Suppose that $\{X_t, \mathcal{F}_t\}$ is a càdlàg submartingale with a decomposition

$$X_t = M_t + A_t,$$

where $\{M_t, \mathcal{F}_t\}$ is a càdlàg martinagle and A_t is an increasing process. Given $T > 0$, let $\mathcal{S}_T$ be the set of $\{\mathcal{F}_t\}$-stopping times τ satisfying $\tau \leqslant T$ a.s. According to the optional sampling theorem, we have $M_\tau = \mathbb{E}[M_T|\mathcal{F}_\tau]$ for every $\tau \in \mathcal{S}_T$. It follows from Problem Sheet 1, Problem 2, (1) that $\{M_\tau : \ \tau \in \mathcal{S}_T\}$ is uniformly integrable. Moreover, since A_T is integrable and $A_\tau \leqslant A_T$ a.s. for every $\tau \in \mathcal{S}_T$, we conclude that $\{X_\tau : \ \tau \in \mathcal{S}_T\}$ is uniformly integrable.

Definition 3.12. A càdlàg submartingale $\{X_t, \mathcal{F}_t\}$ is said to be of *class (DL)* if for every $T > 0$, the family $\{X_\tau : \ \tau \in \mathcal{S}_T\}$ is uniformly integrable.

Now we prove the converse. This is the famous Doob-Meyer decomposition theorem.

Theorem 3.13. *Let $\{X_t, \mathcal{F}_t\}$ be a càdlàg submartingale of class (DL). Then X_t can be written as the sum of a càdlàg martingale and an increasing process which is natural. Moreover, such decomposition is unique with probability one.*

Proof. The main idea is to apply discrete approximation by using Doob's decomposition for discrete time submartingales.

(1) We first prove uniqueness.

Suppose that X_t has two such decompositions:

$$X_t = M_t + A_t = M'_t + A'_t.$$

Then $\Delta_t \triangleq A'_t - A_t = M_t - M'_t$ is an $\{\mathcal{F}_t\}$-martingale. Therefore, fix $t > 0$, for any bounded and càdlàg martingale $\{m_s, \mathcal{F}_s\}$, we have

$$\mathbb{E}\left[\int_0^t m_{s-} d\Delta_s\right] = \lim_{\mathrm{mesh}(\mathcal{P})\to 0} \sum_{k=1}^n \mathbb{E}\left[m_{t_{k-1}}\left(\Delta_{t_k} - \Delta_{t_{k-1}}\right)\right] = 0,$$

where $\mathcal{P}: \ 0 = t_0 < t_1 < \cdots < t_n = t$ is a finite partition of $[0, t]$. Since A and A' are both natural, it follows that $\mathbb{E}[m_t\Delta_t] = 0$. For an arbitrary bounded $\mathcal{F}_t$-measurable random variable ξ, let m_s be a càdlàg version of $\mathbb{E}[\xi|\mathcal{F}_s]$. It follows that $\mathbb{E}[\xi\Delta_t] = 0$. By taking $\xi = \mathbf{1}_{\{A_t < A'_t\}}$, we conclude that almost surely $\xi\Delta_t = 0$ and hence $A_t \geqslant A'_t$. Similarly, $A_t \leqslant A'_t$ almost surely. Therefore, $A_t = A'_t$ almost surely. The uniqueness follows from right continuity of sample paths.

(2) To prove existence, it suffices to prove it on every finite interval $[0, T]$, as the uniqueness will then enable us to extend the construction to the whole interval $[0, \infty)$.

For $n \geqslant 1$, let $D_n: \ t_k^n = kT/2^n \ (0 \leqslant k \leqslant 2^n)$ be the n-th dyadic partition of $[0, T]$. Let

$$X_t = M_t^{(n)} + A_t^{(n)}, \quad t \in D_n,$$

be the Doob decomposition for the discrete time submartingale $X|_{D_n}$ given by Theorem 3.12. Since $\left\{M_t^{(n)}, \mathcal{F}_t: \ t \in D_n\right\}$ is a martingale, we have

$$A_t^{(n)} = X_t - \mathbb{E}[X_T|\mathcal{F}_t] + \mathbb{E}\left[A_T^{(n)}|\mathcal{F}_t\right], \quad t \in D_n. \tag{3.12}$$

(3) The key step is to prove that $\left\{A_T^{(n)}: \ n \geqslant 1\right\}$ is uniformly integrable, which then enables us to get a weak limit A_T along a subsequence according to the Dunford-Pettis theorem (c.f. Theorem 1.2). In view of (3.12), the desired increasing process A_t can then be constructed in terms of A_T and X_t easily.

Given $\lambda > 0$, define

$$\tau_\lambda^{(n)} \triangleq \inf\left\{t_k^n \in D_n: \ A_{t_{k+1}^n}^{(n)} > \lambda\right\},$$

where $\inf\emptyset \triangleq T$. Since $\left\{A_t^{(n)}: \ t \in D_n\right\}$ is $\{\mathcal{F}_t: \ t \in D_n\}$-predictable, we see that $\tau_\lambda^{(n)} \in \mathcal{S}_T$. Moreover, $\left\{A_T^{(n)} > \lambda\right\} = \left\{\tau_\lambda^{(n)} < T\right\} \in \mathcal{F}_{\tau_\lambda^{(n)}}$. By applying the optional sampling theorem to the identity (3.12), we obtain that

$$\begin{aligned}
&\mathbb{E}\left[A_T^{(n)}\mathbf{1}_{\left\{A_T^{(n)}>\lambda\right\}}\right] \\
&= \mathbb{E}\left[X_T\mathbf{1}_{\left\{\tau_\lambda^{(n)}<T\right\}}\right] - \mathbb{E}\left[X_{\tau_\lambda^{(n)}}\mathbf{1}_{\left\{\tau_\lambda^{(n)}<T\right\}}\right] + \mathbb{E}\left[A_{\tau_\lambda^{(n)}}^{(n)}\mathbf{1}_{\left\{\tau_\lambda^{(n)}<T\right\}}\right] \\
&\leqslant \mathbb{E}\left[X_T\mathbf{1}_{\left\{\tau_\lambda^{(n)}<T\right\}}\right] - \mathbb{E}\left[X_{\tau_\lambda^{(n)}}\mathbf{1}_{\left\{\tau_\lambda^{(n)}<T\right\}}\right] + \lambda\mathbb{P}\left(\tau_\lambda^{(n)}<T\right). \qquad (3.13)
\end{aligned}$$

On the other hand, since $\left\{\tau_\lambda^{(n)}<T\right\} \subseteq \left\{\tau_{\lambda/2}^{(n)}<T\right\}$, we have

$$\mathbb{E}\left[\left(A_T^{(n)} - A_{\tau_{\lambda/2}^{(n)}}^{(n)}\right)\mathbf{1}_{\left\{\tau_{\lambda/2}^{(n)}<T\right\}}\right] \geqslant \mathbb{E}\left[\left(A_T^{(n)} - A_{\tau_{\lambda/2}^{(n)}}^{(n)}\right)\mathbf{1}_{\left\{\tau_\lambda^{(n)}<T\right\}}\right] \geqslant \frac{\lambda}{2}\mathbb{P}\left(\tau_\lambda^{(n)}<T\right). \qquad (3.14)$$

Again from the optional sampling theorem, we know that the left hand side of (3.14) is equal to $\mathbb{E}\left[\left(X_T - X_{\tau_{\lambda/2}^{(n)}}\right)\mathbf{1}_{\left\{\tau_{\lambda/2}^{(n)}<T\right\}}\right]$. Therefore, from (3.13) we arrive at

$$\begin{aligned}
\mathbb{E}\left[A_T^{(n)}\mathbf{1}_{\left\{A_T^{(n)}>\lambda\right\}}\right] \leqslant\ & \mathbb{E}\left[X_T\mathbf{1}_{\left\{\tau_\lambda^{(n)}<T\right\}}\right] - \mathbb{E}\left[X_{\tau_\lambda^{(n)}}\mathbf{1}_{\left\{\tau_\lambda^{(n)}<T\right\}}\right] \\
&+2\mathbb{E}\left[X_T\mathbf{1}_{\left\{\tau_{\lambda/2}^{(n)}<T\right\}}\right] - 2\mathbb{E}\left[X_{\tau_{\lambda/2}^{(n)}}\mathbf{1}_{\left\{\tau_{\lambda/2}^{(n)}<T\right\}}\right]. \qquad (3.15)
\end{aligned}$$

Now observe that

$$\mathbb{P}\left(\tau_\lambda^{(n)}<T\right) = \mathbb{P}\left(A_T^{(n)}>\lambda\right) \leqslant \frac{1}{\lambda}\mathbb{E}\left[A_T^{(n)}\right] = \frac{1}{\lambda}\mathbb{E}[X_T - X_0] \to 0$$

uniformly in n as $\lambda \to \infty$. Since X_t is of class (DL), we conclude that the right hand side of (3.15) converges to 0 uniformly in n as $\lambda \to \infty$. In particular, this implies that $\left\{A_T^{(n)} : n \geqslant 1\right\}$ is uniformly integrable.

(4) According to the Dunford-Pettis Theorem (c.f. Theorem 1.2), there exist a subsequence $A_T^{(n_j)}$ and some $A_T \in L^1(\Omega, \mathcal{F}_T, \mathbb{P})$, such that $A_T^{(n_j)}$ converges to A_T weakly in $L^1(\Omega, \mathcal{F}_T, \mathbb{P})$. In view of (3.12), now we define (taking a right continuous version)

$$A_t = X_t - \mathbb{E}[X_T|\mathcal{F}_t] + \mathbb{E}[A_T|\mathcal{F}_t], \quad t \in [0, T]. \qquad (3.16)$$

It remains to show that A_t is a natural increasing process.

First of all, given any bounded $\mathcal{F}_T$-measurable random variable ξ, from Problem Sheet 1, Problem 1, (1), (i), we know that

$$\mathbb{E}[\xi\mathbb{E}[A_T|\mathcal{F}_t]] = \mathbb{E}[A_T\mathbb{E}[\xi|\mathcal{F}_t]],$$

and the same is true when A_T is replaced by $A_T^{(n)}$. In view of (3.12) and (3.16), we see that $A_t^{(n_j)}$ converges to A_t weakly in $L^1(\Omega, \mathcal{F}_T, \mathbb{P})$ for any given $t \in D \triangleq \cup_{n=1}^\infty D_n$.

To see the increasingness of A, let $s < t \in D$ and $\xi = \mathbf{1}_{\{A_s > A_t\}}$. It follows that $\xi(A_s - A_t) \geqslant 0$. But

$$\mathbb{E}[\xi(A_s - A_t)] = \lim_{j\to\infty} \mathbb{E}\left[\xi\left(A_s^{(n_j)} - A_t^{(n_j)}\right)\right] \leqslant 0.$$

Therefore, $A_s \leqslant A_t$ almost surely. The increasingness of A_t then follows from right continuity.

To see the naturality of A, first note that $\left\{A_t^{(n)} : t \in D_n\right\}$ is $\{\mathcal{F}_t : t \in D_n\}$-predictable, and $A_t^{(n)}, A_t$ differ from X_t by martingales. Therefore, for any given bounded and càdlàg martingale $\{m_t, \mathcal{F}_t : t \in [0, T]\}$, we have

$$\begin{aligned}
\mathbb{E}\left[m_T A_T^{(n)}\right] &= \sum_{k=1}^{2^n} \mathbb{E}\left[m_T\left(A_{t_k^n}^{(n)} - A_{t_{k-1}^n}^{(n)}\right)\right] = \sum_{k=1}^{2^n} \mathbb{E}\left[m_{t_{k-1}^n}\left(A_{t_k^n}^{(n)} - A_{t_{k-1}^n}^{(n)}\right)\right] \\
&= \sum_{k=1}^{2^n} \mathbb{E}\left[m_{t_{k-1}^n}\left(X_{t_k^n} - X_{t_{k-1}^n}\right)\right] = \sum_{k=1}^{2^n} \mathbb{E}\left[m_{t_{k-1}^n}\left(A_{t_k^n} - A_{t_{k-1}^n}\right)\right].
\end{aligned}$$

By taking limit along the subsequence n_j, we conclude that

$$\mathbb{E}[m_T A_T] = \mathbb{E}\left[\int_0^T m_{s-} dA_s\right]. \tag{3.17}$$

Now given $t \in [0, T]$, by applying (3.17) to the bounded and càdlàg martingale $m_s^t \triangleq m_{t\wedge s}$ $(s \in [0, T])$, we obtain that

$$\mathbb{E}[m_t A_t] = \mathbb{E}\left[\int_0^t m_{s-} dA_s\right].$$

Therefore, the naturality of A follows.

Now the proof of Theorem 3.13 is complete. □

It is usually important to understand the relationship between regularity properties of a submartingale and of its Doob-Meyer decomposition.

Definition 3.13. A càdlàg submartingale $\{X_t, \mathcal{F}_t\}$ is called *regular* if for every $T > 0$ and $\tau_n, \tau \in \mathcal{S}_T$ with $\tau_n \uparrow \tau$, we have

$$\lim_{n\to\infty} \mathbb{E}[X_{\tau_n}] = \mathbb{E}[X_\tau].$$

Note that from the optional sampling theorem, every càdlàg martingale is regular.

Lemma 3.4. *Let $\{X_t, \mathcal{F}_t\}$ be a regular càdlàg submartingale which is of class (DL), and let A_t be the natural increasing process in the Doob-Meyer decomposition of X_t. Then for any $\tau_n, \tau \in \mathcal{S}_T$ with $\tau_n \uparrow \tau$, $A_{\tau_n} \uparrow A_\tau$ almost surely.*

Proof. $0 \leqslant A_{\tau_n} \leqslant A_\tau \leqslant A_T$ implies that $\{A_{\tau_n} : n \geqslant 1\}$ is uniformly integrable. Let $B = \lim_{n\to\infty} A_{\tau_n}$. Then $\mathbb{E}[B] = \lim_{n\to\infty} \mathbb{E}[A_{\tau_n}]$. On the other hand, it is apparent that A_t is also regular and $B \leqslant A_\tau$. Therefore, $\mathbb{E}[B] = \mathbb{E}[A_\tau]$, which implies that $B = A_\tau$ almost surely. □

Now we have the following general result.

Theorem 3.14. *Let $\{X_t, \mathcal{F}_t\}$ be a càdlàg submartingale which is of class (DL), and let $X_t = M_t + A_t$ be its Doob-Meyer decomposition. Then X_t is regular if and only if A_t is continuous.*

Proof. Sufficiency is obvious. Now we show necessity.

Since A_t is increasing and right continuous, we use the following global way to establish the continuity of A_t over an arbitrary finite interval $[0, T]$: it suffices to show that, for every $\lambda > 0$,

$$\mathbb{E}\left[\int_0^T A_t \wedge \lambda dA_t\right] = \mathbb{E}\left[\int_0^T A_{t-} \wedge \lambda dA_t\right]. \tag{3.18}$$

Indeed, since $A_t \wedge \lambda$ is also increasing and right continuous, it has at most countably many discontinuities. Therefore,

$$\int_0^T A_t\wedge\lambda dA_t - \int_0^T A_{t-}\wedge\lambda dA_t = \sum_{t\leqslant T}(A_t\wedge\lambda - A_{t-}\wedge\lambda)(A_t - A_{t-}) \geqslant \sum_{t\leqslant T}(A_t\wedge\lambda - A_{t-}\wedge\lambda)^2,$$

where the summation is over all discontinuities of $A_t \wedge \lambda$ on $[0, T]$. Therefore, (3.18) implies that with probability one, $A_t \wedge \lambda$ does not have jumps on $[0, T]$. Since λ is arbitrary, we conclude that A_t is continuous almost surely.

Now we establish (3.18). This looks very similar to the naturality property of A_t, except for the fact that the integrand $A_t \wedge \lambda$ is not a martingale. To get around this issue, we use piecewise martingales to approximate $A_t \wedge \lambda$ in a reasonable sense.

As in the proof of Theorem 3.13, let D_n be the n-th dyadic partition of $[0, T]$. Define (taking a right continuous version)

$$A_t^{(n)} = \mathbb{E}\left[A_{t_k^n} \wedge \lambda | \mathcal{F}_t\right], \quad t \in (t_{k-1}^n, t_k^n],\ 1 \leqslant k \leqslant 2^n.$$

By applying the naturality property of A_t on each $(t_{k-1}^n, t_k^n]$, we obtain that

$$\mathbb{E}\left[\int_0^T A_t^{(n)} dA_t\right] = \mathbb{E}\left[\int_0^T A_{t-}^{(n)} dA_t\right].$$

Now we prove that $A_t^{(n)}$ converges uniformly in $t \in [0, T]$ to $A_t \wedge \lambda$ in probability. This will imply that along a subsequence $A_t^{(n_j)}$ converges uniformly in $t \in [0, T]$ to $A_t \wedge \lambda$ almost surely, which concludes (3.18) by the dominated convergence theorem.

Note that with probability one, $A_t^{(n)} \geqslant A_t \wedge \lambda$ and $A_t^{(n)}$ is decreasing in n for every $t \in [0, T]$. Given $\varepsilon > 0$, let $\sigma_\varepsilon^{(n)} = \inf\left\{t \in [0, T] : A_t^{(n)} - A_t \wedge \lambda > \varepsilon\right\}$ $(\inf \emptyset \triangleq T)$. Then

$\sigma_\varepsilon^{(n)}$ is an increasing sequence of $\{\mathcal{F}_t\}$-stopping times in $\mathcal{S}_T$. Let $\sigma_\varepsilon = \lim_{n\to\infty}\sigma_\varepsilon^{(n)}$. Now define another $\tau_\varepsilon^{(n)} \in \mathcal{S}_T$ by $\tau_\varepsilon^{(n)} = t_k^n$ if $\sigma_\varepsilon^{(n)} \in (t_{k-1}^n, t_k^n]$. It is apparent that $\sigma_\varepsilon^{(n)} \leqslant \tau_\varepsilon^{(n)}$ and $\tau_\varepsilon^{(n)} \uparrow \sigma_\varepsilon$ as well.

For fixed $1 \leqslant k \leqslant 2^n$, by applying the optional sampling theorem to the martingale $\widetilde{A}_t^{(n)} = \mathbb{E}\left[A_{t_k^n} \wedge \lambda | \mathcal{F}_t\right]$ $(t \in [0,T])$, we obtain that

$$\begin{aligned}
\mathbb{E}\left[A^{(n)}_{\sigma_\varepsilon^{(n)}}\mathbf{1}_{\left\{t_{k-1}^n<\sigma_\varepsilon^{(n)}\leqslant t_k^n\right\}}\right] &= \mathbb{E}\left[\widetilde{A}^{(n)}_{\sigma_\varepsilon^{(n)}}\mathbf{1}_{\left\{t_{k-1}^n<\sigma_\varepsilon^{(n)}\leqslant t_k^n\right\}}\right] \\
&= \mathbb{E}\left[A_{t_k^n}\wedge\lambda\mathbf{1}_{\left\{t_{k-1}^n<\sigma_\varepsilon^{(n)}\leqslant t_k^n\right\}}\right] \\
&= \mathbb{E}\left[A_{\tau_\varepsilon^{(n)}}\wedge\lambda\mathbf{1}_{\left\{t_{k-1}^n<\sigma_\varepsilon^{(n)}\leqslant t_k^n\right\}}\right].
\end{aligned}$$

By summing over k, we arrive at $\mathbb{E}\left[A^{(n)}_{\sigma_\varepsilon^{(n)}}\right] = \mathbb{E}\left[A_{\tau_\varepsilon^{(n)}}\wedge\lambda\right]$. Therefore,

$$\mathbb{E}\left[A_{\tau_\varepsilon^{(n)}}\wedge\lambda - A_{\sigma_\varepsilon^{(n)}}\wedge\lambda\right] = \mathbb{E}\left[A^{(n)}_{\sigma_\varepsilon^{(n)}} - A_{\sigma_\varepsilon^{(n)}}\wedge\lambda\right] \geqslant \varepsilon\mathbb{P}\left(\sigma_\varepsilon^{(n)} < T\right).$$

On the other hand, according to Lemma 3.4, we know that

$$\lim_{n\to\infty} A_{\sigma_\varepsilon^{(n)}}\wedge\lambda = \lim_{n\to\infty} A_{\tau_\varepsilon^{(n)}}\wedge\lambda = A_{\sigma_\varepsilon}\wedge\lambda, \quad \text{a.s.}$$

Therefore, by the monotone convergence theorem, we conclude that

$$\lim_{n\to\infty}\mathbb{P}\left(\sigma_\varepsilon^{(n)} < T\right) = 0.$$

But $\left\{\sigma_\varepsilon^{(n)} < T\right\} = \left\{\sup_{t\in[0,T]}(A_t^{(n)} - A_t\wedge\lambda) > \varepsilon\right\}$. In other words, $A_t^{(n)}$ converges uniformly in $t \in [0,T]$ to $A_t\wedge\lambda$ in probability.

Now the proof of Theorem 3.14 is complete. □

4 Brownian motion

In this section, we study a fundamental example of stochastic processes: the Brownian motion.

In 1905, based on principles of statistical physics, Albert Einstein discovered the mechanism governing the random movement of particles suspended in a fluid, a phenomenon first observed by the botanist Robert Brown. In physics, such random motion is known as the Brownian motion. However, it was Louis Bachelier in 1900 who first used the distribution of Brownian motion to model Paris stock market and evaluate stock options. The precise mathematical model of Brownian motion was established by Nobert Wiener in 1923.

Brownian motion is the most important object in stochastic analysis, since it lies in the intersection of all fundamental stochastic processes: it is a Gaussian process, a martingale, a (strong) Markov process and a diffusion. Moreover, being an elegant mathematical object on its own, it also connects stochastic analysis with other parts of mathematics, e.g. partial differential equations, harmonic analysis, differential geometry etc. as well as applied areas such as physics and mathematical finance.

From this section we will start appreciating the great power of martingale methods developed in the last section.

4.1 Basic properties

Definition 4.1. A (d-dimensional) stochastic process $\{B_t : t \geqslant 0\}$ is called a *(d-dimensional) Brownian motion* if:

(1) $B_0 = 0$ almost surely;

(2) for every $0 \leqslant s < t$, $B_t - B_s$ is normally distributed with mean zero and covariance matrix $(t-s)\mathrm{I}_d$, where I_d is the $d \times d$ identity matrix;

(3) for every $0 < t_1 < \cdots < t_n$, the random variables $B_{t_1}, B_{t_2} - B_{t_1}, \cdots, B_{t_n} - B_{t_{n-1}}$ are independent;

(4) with probability one, $t \mapsto B_t(\omega)$ is contiuous.

Direct computation shows that a Brownian motion is a d-dimensional Gaussian process with i.i.d. components, each having covariance function $\rho(s,t) = s \wedge t$ $(s, t \geqslant 0)$.

As usual, it is also important to keep track of information when a filtration is presented.

Definition 4.2. Let $\{\mathcal{F}_t : t \geqslant 0\}$ be a filtration. A stochastic process $\{B_t : t \geqslant 0\}$ is called an *$\{\mathcal{F}_t\}$-Brownian motion* if it is a Brownian motion such that it is $\{\mathcal{F}_t\}$-adapted and $B_t - B_s$ is independent of $\mathcal{F}_s$ for every $s < t$.

Apparently, every Brownian motion is a Brownian motion with respect to its natural filtration. Moreover, every $\{\mathcal{F}_t\}$-Brownian motion is an $\{\mathcal{F}_t\}$-martingale.

The existence of a Brownian motion on some probability space is proved in Problem Sheet 2, Problem 2 by using Kolmogorov's extension and continuity theorems. From a mathematical point of view, it is also important and convenient to realize a Brownian motion on the continuous path space $(W^d, \mathcal{B}(W^d), \rho)$ (c.f. Section 1.5). Suppose that B_t is a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$ and every sample path of B_t is continuous. It is straight forward to see that the map $B: \omega \mapsto (B_t(\omega))_{t \geqslant 0}$ is $\mathcal{F}/\mathcal{B}(W^d)$ measurable. Let $\mu_d \triangleq \mathbb{P} \circ B^{-1}$.

Definition 4.3. μ_d is called the *(d-dimensional) Wiener measure (or the law of Brownian motion).*

From the definition of Brownian motion and the uniqueness of Carathéodory's extension, we can see that μ^d is the unique probability measure on $(W^d, \mathcal{B}(W^d))$ under which the coordinate process $X_t(w) \triangleq w_t$ is a Brownian motion.

The following invariance properties of Brownian motion are obvious.

Proposition 4.1. *Let B_t be a Brownian motion. Then we have:*

(1) translation invariance: for every $s \geqslant 0$, $\{B_{t+s} - B_s : \ t \geqslant 0\}$ is a Brownian motion;

(2) reflection symmetry: $-B_t$ is a Brownian motion;

(3) scaling invariance: for every $\lambda > 0$, $\{\lambda^{-1} B_{\lambda^2 t} : \ t \geqslant 0\}$ is a Brownian motion.

4.2 The strong Markov property and the reflection principle

Now we demonstrate a very important property of Brownian motion: the strong Markov property.

Heuristically, the Markov property means that knowing the present state, history on the past does not provide any new information on predicting the distribution of future states. "Strong" means that the meaning of "present" can be randomized by a stopping time.

Let

$$p_t(x,y) = \frac{1}{(2\pi t)^{\frac{d}{2}}} \mathrm{e}^{-\frac{|x-y|^2}{2}}, \quad t > 0, \ x, y \in \mathbb{R}^d. \tag{4.1}$$

It is easy to see that

$$\frac{\partial}{\partial t} p_t(x,y) = \frac{1}{2} \Delta_x p_t(x,y), \quad t > 0.$$

Let $B_b(\mathbb{R}^d)$ be the space of bounded measurable functions on $\mathbb{R}^d$, equipped with the supremum norm. Define a family $\{P_t : \ t \geqslant 0\}$ of continuous linear operators on $B_b(\mathbb{R}^d)$ by

$$P_t f(x) = \begin{cases} \int_{\mathbb{R}^d} p_t(x,y) f(y) dy, & t > 0; \\ f(x), & t = 0. \end{cases}$$

Routine calculation shows the following semigroup property:

$$P_{t+s} = P_t \circ P_s = P_s \circ P_t, \, \forall s, t \geqslant 0.$$

This is known as the *Chapman-Kolmogorov equation*, which is a basic feature of Markov processes.

Now let B_t be an $\{\mathcal{F}_t\}$-Brownian motion with respect to some filtration $\{\mathcal{F}_t\}$. The Markov property of B_t can be stated as follows.

Theorem 4.1. *For every $0 \leqslant s < t$ and $f \in B_b(\mathbb{R}^d)$, we have*

$$\mathbb{E}[f(B_{t+s})|\mathcal{F}_s] = \mathbb{E}[f(B_{t+s})|B_s] \ \text{ a.s.}$$

Proof. Note that

$$\mathbb{E}[f(B_{t+s} - B_s + x)] = P_t f(x), \quad \forall x \in \mathbb{R}^d.$$

Since $B_{t+s} - B_s$ is independent of $\mathcal{F}_s$ and B_s is $\mathcal{F}_s$-measurable, from Problem Sheet 1, Problem 1, (1), (ii), we conclude that

$$\mathbb{E}[f(B_{t+s})|\mathcal{F}_s] = \mathbb{E}[f(B_{t+s} - B_s + B_s)|\mathcal{F}_s] = P_t f(B_s) \ \text{ a.s.}$$

The result then follows by conditioning on B_s. □

The kernel $p_t(x,y)$ is called the *Brownian transition density* (or the *heat kernel*). Heuristically, it gives the probability density of finding a Brownian particle at position y after time t whose initial position is x. Respectively, the semigroup $\{P_t\}$ is called the *Brownian transition semigroup* (or the *heat semigroup*). The relationship between the Brownian motion B_t and the Laplace operator Δ lies in the fact that Δ is the *infinitesimal generator* of B_t, in the sense that

$$\frac{1}{2}\Delta f = \lim_{t \to 0}(P_t f - f)/t \ \text{ in } B_b(\mathbb{R}^d),$$

at least for $f \in C_c^2(\mathbb{R}^d)$, the space of twice continuously differentiable functions with compact support. We will come back to this point when we study stochastic differential equations and diffusion processes.

The strong Markov property of Brownian motion takes essentially the same form as Theorem 4.1, but with s replaced by a stopping time. Indeed, we will establish a finer result. The proof exploits the optional sampling theorem of martingales.

Theorem 4.2. *Suppose that $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets. Let B_t be an $\{\mathcal{F}_t\}$-Brownian motion and let τ be an $\{\mathcal{F}_t\}$-stopping time which is finite almost surely. Then the process $B^{(\tau)} = \{B_{\tau+t} - B_\tau : \ t \geqslant 0\}$ is a Brownian motion which is independent of $\mathcal{F}_\tau$. In particular, for every $t \geqslant 0$ and $f \in B_b(\mathbb{R}^d)$, we have*

$$\mathbb{E}[f(B_{\tau+t})|\mathcal{F}_\tau] = \mathbb{E}[f(B_{\tau+t})|B_\tau] \ \text{ a.s.} \tag{4.2}$$

Proof. From classical probability theory, it is sufficient to show that:

$$\mathbb{E}\left[\xi \mathrm{e}^{i\sum_{k=1}^n \left\langle \theta_k, B_{t_k}^{(\tau)} - B_{t_{k-1}}^{(\tau)} \right\rangle}\right] = \mathbb{E}[\xi] \cdot \mathrm{e}^{-\frac{1}{2}\sum_{k=1}^n |\theta_k|^2 (t_k - t_{k-1})} \tag{4.3}$$

for every ξ bounded $\mathcal{F}_\tau$-measurable, $0 = t_0 < t_1 < \cdots < t_n$ and $\theta_1, \cdots, \theta_n \in \mathbb{R}^d$.

In general, given $\theta \in \mathbb{R}^d$, define

$$M_t^{(\theta)} = \mathrm{e}^{i\langle \theta, B_t \rangle + \frac{1}{2}|\theta|^2 t}, \quad t \geqslant 0.$$

It is easily seen that $\left\{M_t^{(\theta)}, \mathcal{F}_t\right\}$ is a continuous martingale. Therefore, given an almost surely finite $\{\mathcal{F}_t\}$-stopping time σ and $t \geqslant 0$, according to the optional sampling theorem, we have

$$\mathbb{E}\left[\mathrm{e}^{i\langle\theta, B_{\sigma\wedge N+t}\rangle+\frac{1}{2}|\theta|^2(\sigma\wedge N+t)}|\mathcal{F}_{\sigma\wedge N}\right]=\mathrm{e}^{i\langle\theta, B_{\sigma\wedge N}\rangle+\frac{1}{2}|\theta|^2\sigma\wedge N},\quad \forall N\in\mathbb{N}.$$

Equivalently,

$$\mathbb{E}\left[\mathrm{e}^{i\langle\theta, B_{\sigma\wedge N+t}-B_{\sigma\wedge N}\rangle}|\mathcal{F}_{\sigma\wedge N}\right]=\mathrm{e}^{-\frac{1}{2}|\theta|^2 t},\quad \forall N\in\mathbb{N}. \tag{4.4}$$

Note that $\mathcal{F}_\sigma = \cup_{N\in\mathbb{N}}\mathcal{F}_{\sigma\wedge N}$ ($A \in \mathcal{F}_\sigma \implies A\cap\{\sigma \leqslant N\} \in \mathcal{F}_{\sigma\wedge N}$ for all N, so $A\cap\{\sigma<\infty\}\in\cup_{N\in\mathbb{N}}\mathcal{F}_{\sigma\wedge N}$. But $A\cap\{\sigma=\infty\}\in\mathcal{F}_0$ by assumption.) By using the definition of conditional expectation and the dominated convergence theorem, we may take limit $N\to\infty$ in (4.4) to conclude that

$$\mathbb{E}\left[\mathrm{e}^{i\langle\theta, B_{\sigma+t}-B_\sigma\rangle}|\mathcal{F}_\sigma\right]=\mathrm{e}^{-\frac{1}{2}|\theta|^2 t}. \tag{4.5}$$

Now (4.3) follows from taking conditional expectations and applying (4.5) recursively, starting from $\sigma=\tau+t_{n-1}$, $t=t_n-t_{n-1}$ and $\theta=\theta_n$.

In the same way as before, the strong Markov property (4.2) follows from the fact that

$$\mathbb{E}[f(B_{\tau+t})|\mathcal{F}_\tau]=P_t f(B_\tau)\ \text{ a.s.}$$

□

In the one dimensional case, a very nice application of the strong Markov property is the so-called *reflection principle*, which yields immediately the explicit distributions of passage times and maximal functionals.

Let B_t be a one dimensional Brownian motion, and let $\{\mathcal{F}_t^B\}$ be the augmented natural filtration of B_t (i.e. $\mathcal{F}_t^B=\sigma(\mathcal{G}_t^B,\mathcal{N})$ where $\{\mathcal{G}_t^B\}$ is the natural filtration of B_t and $\mathcal{N}$ is the collection of $\mathbb{P}$-null sets).

For $x>0$, set

$$\tau_x=\inf\{t\geqslant 0:\ B_t=x\}.$$

Then τ_x is an $\{\mathcal{F}_t^B\}$-stopping time. To see it is finite almost surely, we need the following simple fact.

Proposition 4.2. *We have:*

$$\mathbb{P}\left(\sup_{t\geqslant 0}B_t=\infty\right)=\mathbb{P}\left(\inf_{t\geqslant 0}B_t=-\infty\right)=1.$$

In particular, $\tau_x<\infty$ almost surely.

Proof. Let $M=\sup_{t\geqslant 0}B_t$. According to the scaling invariance of Brownian motion (c.f. Proposition 4.1, (2)), for every $\lambda>0$, we have

$$\{\lambda^{-1}B_{\lambda^2 t}:\ t\geqslant 0\}\stackrel{\text{law}}{=}\{B_t:\ t\geqslant 0\}.$$

This certainly implies that $\lambda^{-1}M \overset{\text{law}}{=} M$. Now we have:

$$\begin{aligned}\mathbb{P}(M>\lambda)=\mathbb{P}(M>1) &\overset{\lambda\to\infty}{\Longrightarrow} \mathbb{P}(M=\infty)=\mathbb{P}(M>1),\\ \mathbb{P}(M\leqslant\lambda)=\mathbb{P}(M\leqslant 1) &\overset{\lambda\to 0}{\Longrightarrow} \mathbb{P}(M\leqslant 0)=\mathbb{P}(M\leqslant 1).\end{aligned}$$

Since $M \geqslant 0$ almost surely, we conclude that M is supported on $\{0, \infty\}$.

On the other hand, observe that

$$\begin{aligned}\mathbb{P}(M=0) &\leqslant \mathbb{P}(B_1\leqslant 0,\ B_u\leqslant 0\ \forall u\geqslant 1)\\ &\leqslant \mathbb{P}\left(B_1\leqslant 0,\ \sup_{t\geqslant 0}(B_{1+t}-B_1)=0\right)\\ &= \mathbb{P}(B_1\leqslant 0)\cdot\mathbb{P}(M=0)\\ &= \frac{1}{2}\mathbb{P}(M=0),\end{aligned}$$

where the second inequality follows from the fact that $\sup_{t\geqslant 0}(B_{1+t}-B_1)$ is either 0 or ∞ since $\{B_{1+t}-B_1 :\ t\geqslant 0\}$ is again a Brownian motion. Therefore, $\mathbb{P}(M=0)=0$ and $M=\infty$ almost surely.

The infimum case follows from the reflection symmetry of Brownian motion. □

The reflection principle asserts that the law of Brownian motion is invariant under reflection with respect to the position x after time τ_x. Here is the mathematical statement.

Proposition 4.3. *Define*

$$\widetilde{B}_t = \begin{cases} B_t, & t<\tau_x;\\ 2x-B_t, & t\geqslant \tau_x.\end{cases} \tag{4.6}$$

Then $\widetilde{B}_t$ is also a Brownian motion.

Proof. According to Theorem 4.2, the process $B_t^{(\tau_x)} \triangleq B_{\tau_x+t}-B_{\tau_x} = B_{\tau_x+t}-x$ is a Brownian motion which is independent of $\mathcal{F}_{\tau_x}$. Therefore, $-B_t^{(\tau_x)}$ is also a Brownian motion being independent of $\mathcal{F}_{\tau_x}$. Let $Y_t \triangleq B_{\tau_x\wedge t}$ be the Brownian motion stopped at τ_x. Note that the map $\omega\mapsto Y_\cdot(\omega)$ is $\mathcal{F}_{\tau_x}/\mathcal{B}(W_0^1)$-measurable, where W_0^1 is the space of continuous path $w\in W^1$ with $w_0=0$. It follows that as random variables taking values in the space $W_0^1\times[0,\infty)\times W_0^1$, $\left(Y,\tau_x,B^{(\tau_x)}\right)$ has the same distribution as $\left(Y,\tau_x,-B^{(\tau_x)}\right)$.

Now define a map $\varphi:\ W_0^1\times[0,\infty)\times W_0^1\to W_0^1$ by $\varphi(x,t,y)(s)=x_s+y_{s-t}\mathbf{1}_{\{s>t\}}$ $(s\geqslant 0)$. Then $\varphi\left(Y,\tau_x,B^{(\tau_x)}\right)=B$ and $\varphi\left(Y,\tau_x,-B^{(\tau_x)}\right)=\widetilde{B}$. Therefore, $\widetilde{B}$ is also a Brownian motion. □

4.3 The Skorokhod embedding theorem and the Donsker invariance principle

In this subsection, we apply the strong Markov property to study the fundamental connections between Brownian motion and random walk in dimension one. On the one hand, it is not hard to imagine that the Brownian motion can be regarded as the continuum limit of scaled random walks. The precise form of this result is known as the *Donsker invariance principle*. On the other hand, it is highly non-trivial that every random walk can be embedded into a Brownian motion evaluated along a sequence of stopping times. This embedding result is an easy consequece of the well known *Skorokhod embedding theorem*.

We first establish the Skorokhod embedding theorem. As we will see, based on this theorem, the Donsker invariance principle is a consequence of the continuity of Brownian motion.

Suppose that B_t is a one dimensional Brownian motion with $\{\mathcal{F}_t^B\}$ being its augmented natural filtration.

The Skorokhod embedding theorem can be stated as follows.

Theorem 4.3. *Let X be a real-valued random variable such that $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] < \infty$. Then there exists an integrable $\{\mathcal{F}_t^B\}$-stopping time τ, such that $B_\tau \stackrel{\text{law}}{=} X$ and $\mathbb{E}[\tau] = \mathbb{E}[X^2]$.*

The proof of this theorem is highly non-trivial but the starting point is simple.

Consider the simplest case where X takes two values $a < 0 < b$. The condition $\mathbb{E}[X] = 0$ implies that

$$\mathbb{P}(X = a) = \frac{b}{b-a}, \quad \mathbb{P}(X = b) = \frac{-a}{b-a}, \quad \mathbb{E}[X^2] = -ab.$$

On the other hand, define

$$\tau_{a,b} = \inf\{t \geqslant 0 : \ B_t \notin (a, b)\}.$$

From Proposition 4.2, we know that $\tau_{a,b}$ is an almost surely finte $\{\mathcal{F}_t^B\}$-stopping time.

Proposition 4.4. *We have:*

$$\mathbb{P}(B_{\tau_{a,b}} = a) = \frac{b}{b-a}, \ \mathbb{P}(B_{\tau_{a,b}} = b) = \frac{-a}{b-a}, \ \mathbb{E}[\tau_{a,b}] = -ab.$$

In particular, $\tau_{a,b}$ gives a solution to the Skorokhod embedding problem for the distribution of X.

Proof. By applying the optional sampling theorem to the $\{\mathcal{F}_t^B\}$-martingales B_t and $B_t^2 - t$, we have

$$\mathbb{E}[B_{\tau_{a,b} \wedge n}] = 0, \ \mathbb{E}[B^2_{\tau_{a,b} \wedge n}] = \mathbb{E}[\tau_{a,b} \wedge n].$$

Since $|B_{\tau_{a,b} \wedge n}| \leqslant \max(|a|, |b|)$ for every n, by the dominated convergence theorem, we conclude that

$$\mathbb{E}[B_{\tau_{a,b}}] = 0, \ \mathbb{E}[B^2_{\tau_{a,b}}] = \mathbb{E}[\tau_{a,b}]. \tag{4.7}$$

As $B_{\tau_{a,b}}$ takes the two values a and b, the first identity of (4.7) shows that $B_{\tau_{a,b}} \stackrel{\text{law}}{=} X$. The second identity of (4.7) then shows that $\tau_{a,b}$ is integrable and $\mathbb{E}[\tau_{a,b}] = \mathbb{E}[X^2] = -ab$. □

Remark 4.1. In general, it is a good exercise to show that: for every integrable $\{\mathcal{F}_t^B\}$-stopping time τ, B_τ is square integrable, and

$$\mathbb{E}[B_\tau]=0,\ \mathbb{E}[B_\tau^2]=\mathbb{E}[\tau].$$

This is called *Wald's identities*. Therefore, $\mathbb{E}[X]=0$ and $\mathbb{E}[X^2]=0$ are necessary conditions for the existence of Skorokhod's embedding.

The general solution of the Skorokhod embedding is motivated from the simple two-value case. The idea is the following. We approximate the general random variable X by a binary splitting martingale sequence X_n, so that the desired stopping time τ can be constructed as the limit of a sequence τ_n of stopping times each of which corresponding to a two-value case but starting from the previous one. The strong Markov property will play an important role in the construction.

Definition 4.4. A sequence $\{X_n:\ n\geqslant 1\}$ of random variables is called *binary splitting* if for each $n\geqslant 1$, there exists some Borel measurable function $f_n:\ \mathbb{R}^{n-1}\times\{\pm 1\}\to\mathbb{R}$ and a $\{\pm 1\}$-valued random variable D_n, such that

$$X_n=f_n(X_1,\cdots,X_{n-1},D_n)\ \text{ a.s.}$$

It is called a *binary splitting martingale* if it is also martingale with respect to its natural filtration.

Intuitively, if $\{X_n\}$ is binary splitting sequence, the conditional distribution of X_n given $(X_1,\cdots,X_{n-1})$ is supported on at most two values.

We first establish the approximation result.

Proposition 4.5. *Let X be a square integrable random variable. Then there exists a binary splitting martingale $\{X_n:\ n\geqslant 1\}$ which is square integrable, such that $X_n\to X$ almost surely and in L^2 as $n\to\infty$.*

Proof. Define

$$D_1=\begin{cases}1, & X\geqslant\mathbb{E}[X];\\ -1, & \text{otherwise},\end{cases}$$

$\mathcal{F}_1=\sigma(D_1)$, and $X_1=\mathbb{E}[X|\mathcal{F}_1]$. Inductively, for $n\geqslant 2$, define

$$D_n=\begin{cases}1, & X\geqslant X_{n-1};\\ -1, & \text{otherwise},\end{cases}$$

$\mathcal{F}_n=\sigma(D_1,\cdots,D_n)$, and $X_n=\mathbb{E}[X|\mathcal{F}_n]$. It follows that $X_n=g_n(D_1,\cdots,D_n)$ almost surely for some measurable function g_n on $\{\pm 1\}^n$.

Now the key observation is that for each $n\geqslant 1$, D_n is a function of $X_1,\cdots,X_n$. When $n=1$, this is apparent since $X_1=a\mathbf{1}_{\{D_1=1\}}+b\mathbf{1}_{\{D_1=-1\}}$ for some constants

a, b, so that we can obtain $\mathbf{1}_{\{D_1=1\}}$ (and hence D_1) from X_1. Suppose that this fact is true for $k \leqslant n-1$. By the definition of X_n, we can write

$$X_n = \sum_{i_1,\cdots,i_n=\pm 1} c_{i_1,\cdots,i_n} \mathbf{1}_{\{D_1=i_1,\cdots,D_n=i_n\}} = \xi_1 \mathbf{1}_{\{D_n=1\}} + \xi_2 \mathbf{1}_{\{D_n=-1\}} \text{ a.s.},$$

where ξ_1, ξ_2 are functions of $\mathbf{1}_{\{D_1=1\}}, \cdots, \mathbf{1}_{\{D_{n-1}=1\}}$. By the induction hypothesis, ξ_1, ξ_2 are functions of $X_1, \cdots, X_{n-1}$.Therefore, $\mathbf{1}_{\{D_n=1\}}$ (and hence D_n) can be obtained from $X_1, \cdots, X_n$. Therefore, X_n has the form $X_n = f_n(X_1, \cdots, X_{n-1}, D_n)$ almost surely, which shows that X_n is a binary splitting sequence. The fact that it is a martingale with respect to its natural filtration follows from $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$ (up to $\mathbb{P}$-null sets).

Since $X \in L^2$, from Jensen's inequality we know that X_n is bounded in L^2. By Problem Sheet 3, Problem 3, we conclude that $X_n \to X_\infty$ almost surely and in L^2 for some X_∞. It remains to show that $X = X_\infty$ almost surely.

First of all, we have

$$\lim_{n\to\infty} D_n(X - X_n) = |X - X_\infty| \text{ a.s.} \tag{4.8}$$

Indeed, if $\omega \in \{X = X_\infty\}$, (4.8) is trivial. If $\omega \in \{X > X_\infty\}$, then $X(\omega) > X_n(\omega)$ when n is large. By the definition of D_n, it follows that $D_n(\omega) = 1$ when n is large. Therefore, (4.8) holds at ω. The case when $\omega \in \{X < X_\infty\}$ is similar. Now observe that

$$\mathbb{E}[D_n(X - X_n)] = \mathbb{E}[D_n \mathbb{E}[(X - X_n)|\mathcal{F}_n]] = 0, \quad \forall n \geqslant 1.$$

Since $D_n(X - X_n)$ is bounded in L^2 (and hence uniformly integrable), we conclude that

$$\mathbb{E}[|X - X_\infty|] = \lim_{n\to\infty} \mathbb{E}[D_n(X - X_n)] = 0,$$

which implies that $X = X_\infty$ almost surely. □

Now we are able to prove the Skorokhod embedding theorem.

Proof of Theorem 4.3. Let X_n be the binary splitting martingale given by Proposition 4.5, so that X_n has the form $X_n = f_n(X_1, \cdots, X_n, D_n)$ for every n. By the martingale property, we know that

$$X_{n-1} = \mathbb{E}[X_n | X_1, \cdots, X_{n-1}].$$

Moreover, X_n takes values in $\{f_n(X_1, \cdots, X_{n-1}, 1), f_n(X_1, \cdots, X_{n-1}, -1)\}$ when conditioned on $(X_1, \cdots, X_{n-1})$. This implies that the conditional distribution of X_n given $(X_1, \cdots, X_{n-1})$ is a two-point distribution with mean X_{n-1}.

Now define $\tau_0 = 0$, and for $n \geqslant 1$, define

$$\tau_n = \inf\left\{t \geqslant \tau_{n-1} : \ B_t \notin \left(f_n(B_{\tau_1}, \cdots, B_{\tau_{n-1}}, -1), f_n(B_{\tau_1}, \cdots, B_{\tau_{n-1}}, +1)\right)\right\}.$$

By the strong Markov property, for each $n \geqslant 1$, $\{B_{\tau_{n-1}+t} - B_{\tau_{n-1}} : \ t \geqslant 0\}$ is a Brownian motion independent of $\mathcal{F}^B_{\tau_{n-1}}$. According to Proposition 4.4, the conditional distribution of B_{τ_n} given $(B_{\tau_1}, \cdots, B_{\tau_{n-1}})$ is a two-point distribution with mean $B_{\tau_{n-1}}$.

Therefore, $(X_1,\cdots,X_n)\overset{\text{law}}{=}(B_{\tau_1},\cdots,B_{\tau_n})$ for every n. Moreover,

$$\begin{aligned}\mathbb{E}[\tau_n-\tau_{n-1}] &= \mathbb{E}[\mathbb{E}[\tau_n-\tau_{n-1}|\mathcal{F}^B_{\tau_{n-1}}]]=\mathbb{E}[\mathbb{E}[(B_{\tau_n}-B_{\tau_{n-1}})^2|\mathcal{F}_{\tau_{n-1}}]]\\ &= \mathbb{E}[(B_{\tau_n}-B_{\tau_{n-1}})^2]=\mathbb{E}[(X_n-X_{n-1})^2].\end{aligned}$$

Since $X_n\in L^2$, the martingale property gives

$$\mathbb{E}[\tau_n]=\sum_{k=1}^n\mathbb{E}[(X_k-X_{k-1})^2]=\mathbb{E}[X_n^2].$$

Finally, as τ_n is increasing, we set $\tau=\lim_{n\to\infty}\tau_n$. Since $X_n\to X$ almost surely and in L^2, we conclude that $\mathbb{E}[\tau]=\mathbb{E}[X^2]<\infty$. This particularly implies that $\tau<\infty$ almost surely, and thus $B_{\tau_n}\to B_\tau$ almost surely which yields that $B_\tau\overset{\text{law}}{=}X$.

□

By applying the strong Markov property, we easily obtain the following important fact: in the distributional sense, a random walk can be embedding into a Brownian motion evaluated along a sequence of stopping times.

Proposition 4.6. *Let F be a distribution function on $\mathbb{R}^1$ with mean zero and finite variance σ^2. Suppose that $\{S_n:\ n\geqslant 1\}$ is a random walk with step distribution F (i.e. $S_n=X_1+\cdots+X_n$ where $\{X_n\}$ is an i.i.d. sequence with distribution F).Then there exists a sequence $\{\tau_n:\ n\geqslant 1\}$ of integrable $\{\mathcal{F}^B_t\}$-stopping times, such that $\{\tau_n-\tau_{n-1}\}$ are i.i.d. with mean σ^2, and $B_{\tau_n}\overset{\text{law}}{=}S_n$ for every n.*

Proof. By the Skorokhod embedding theorem, there exists an integrable $\{\mathcal{F}^B_t\}$-stopping time τ_1, such that $B_{\tau_1}\overset{\text{law}}{=}F$ and $\mathbb{E}[\tau_1]=\sigma^2$. Applying the Skorokhod embedding theorem again to the Brownian motion $B^{(\tau_1)}_t=B_{\tau_1+t}-B_{\tau_1}$ with its augmented natural filtration $\left\{\mathcal{F}^{B^{(\tau_1)}}_t\right\}$, we get an integrable $\left\{\mathcal{F}^{B^{(\tau_1)}}_t\right\}$-stopping time τ_2', such that $B^{(\tau_1)}_{\tau_2'}\overset{\text{law}}{=}F$ and $\mathbb{E}[\tau_2']=\sigma^2$. Define $\tau_2=\tau_1+\tau_2'$. According to Theorem 4.2, we know that $B_{\tau_2}\overset{\text{law}}{=}S_2$. Moreover, according to Problem Sheet 2, Problem 4, (2), we know that $\{\mathcal{F}^B_t\}$ is right continuous. Therefore, the fact that τ_2 is an $\{\mathcal{F}^B_t\}$-stopping time follows from Problem Sheet 2, Problem 3, (2), (ii).

Now the result follows by induction. □

Finally, we establish the Donsker invariance principle, which asserts that the Brownian motion is the weak scaling limit of a random walk.

Let S_n be a random walk with step distribution F, where F has mean zero and finite variance σ^2. From the identity $\mathbb{E}[(B_t-B_s)^2]=t-s$, it is not hard to write down the right scaling of S_n: define

$$S^{(n)}_t=\frac{n^{\frac{1}{2}}}{\sigma}\left(\left(\frac{k}{n}-t\right)S_{k-1}+\left(t-\frac{k-1}{n}\right)S_k\right),\quad t\in\left[\frac{k-1}{n},\frac{k}{n}\right],\ k\geqslant 1,\qquad (4.9)$$

with $S_0 = 0$. In other words, $S_t^{(n)}$ is a piecewise linear continuous process taking value $S_k/(\sigma\sqrt{n})$ at each vertex point k/n ($k \geqslant 0$). Let $\mathbb{P}^{(n)}$ be the distribution of the process $S_t^{(n)}$ on W^1. The Donsker invariance principle can be stated as follows.

Theorem 4.4. *Let μ_1 be the one dimensional Wiener measure. Then $\mathbb{P}^{(n)}$ converges weakly to μ_1 as $n \to \infty$.*

Proof. Without loss of generality, we may assume that $\sigma = 1$. Since we are only concerned with distributions, we may futher assume that the random walk $S_n = B_{\tau_n}$, where $\{\tau_n\}$ is the sequence of $\{\mathcal{F}_t^B\}$-stopping times given in Proposition 4.6. Construct $S^{(n)}$ by (4.9) based on this random walk and define $B_t^{(n)} = n^{-1/2}B_{nt}$, which is again a Brownian motion. Note that S_n and B_t are defined on some given probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

It is sufficient to show that: for every fixed $T > 0$,

$$\sup_{0\leqslant t\leqslant T}\left|S_t^{(n)} - B_t^{(n)}\right| \to 0 \text{ in prob.} \tag{4.10}$$

as $n \to \infty$. Indeed, suppose that (4.10) holds. From the definition of the metric ρ on W^1 (c.f. (1.3)), it is then easy to see that $\rho\left(S^{(n)}, B^{(n)}\right) \to 0$ in probability. Now let F be an arbitrary closed subset of W^1. Then for every $\varepsilon > 0$,

$$\begin{aligned} \mathbb{P}^{(n)}(F) &= \mathbb{P}(S^{(n)} \in F) \\ &\leqslant \mathbb{P}(\rho(S^{(n)}, B^{(n)}) > \varepsilon) + \mathbb{P}(\rho(B^{(n)}, F) \leqslant \varepsilon) \\ &= \mathbb{P}(\rho(S^{(n)}, B^{(n)}) > \varepsilon) + \mu_1(\rho(w, F) \leqslant \varepsilon). \end{aligned}$$

By letting $n \to \infty$, we conclude that

$$\limsup_{n\to\infty} \mathbb{P}^{(n)}(F) \leqslant \mu^1(\rho(w, F) \leqslant \varepsilon).$$

Since ε is arbitrary, the result of the theorem follows from Theorem 1.7, (3).

Now we prove (4.10). Without loss of generality, we assume $T = 1$.

If $\omega \in \left\{\sup_{0\leqslant t\leqslant 1}\left|S_t^{(n)} - B_t^{(n)}\right| > \varepsilon\right\}$, then $\left|S_t^{(n)}(\omega) - B_t^{(n)}(\omega)\right| > \varepsilon$ for some t and k with $t \in [(k-1)/n, k/n]$. Since $S_k(\omega) = B_{\tau_k(\omega)}(\omega)$, from the definition of $S_t^{(n)}$ and the intermediate value theorem, there exists some $v \in [\tau_{k-1}(\omega), \tau_k(\omega)]$, such that $S_t^{(n)}(\omega) = B_v(\omega)/\sqrt{n}$. Write $v = nu$, we then have $u \in [\tau_{k-1}(\omega)/n, \tau_k(\omega)/n]$ and $S_t^{(n)}(\omega) = B_u^{(n)}(\omega)$. Therefore, $\left|B_u^{(n)}(\omega) - B_t^{(n)}(\omega)\right| > \varepsilon$. From the continuity of Brownian motion, it is now clear that the key step is to demonstrate τ_k/n and k/n are very close to each other (so are u and t) in a suitable sense.

Given $\varepsilon > 0$, by the continuity of Brownian motion, there exists $0 < \delta < 1$, such that

$$\mathbb{P}\left(\bigcup_{\substack{s,t\in[0,2]\\|t-s|\leqslant\delta}} \{|B_t - B_s| > \varepsilon\}\right) < \varepsilon/2.$$

On the other hand, from Proposition 4.6 we know that $\{\tau_n - \tau_{n-1}\}$ are i.i.d. with unit mean and $\tau_n = \sum_{k=1}^n (\tau_k - \tau_{k-1})$. By the strong law of large numbers, $\tau_n/n \to 1$ almost surely. In general, it is an elementary fact that

$$a_n \geqslant 0,\ \frac{a_n}{n} \to 0 \implies \frac{1}{n} \sup_{1\leqslant k\leqslant n} a_k \to 0.$$

Taking $a_n = |\tau_n - n|$ in our case, we conclude that $\sup_{1\leqslant k\leqslant n} |\tau_k - k|/n \to 0$ almost surely. In particular, the convergence holds in probability. Therefore, there exists $N \geqslant 1$, such that for any $n > N$, we have

$$\mathbb{P}\left(\frac{1}{n} \sup_{1\leqslant k\leqslant n} |\tau_k - k| > \frac{\delta}{5}\right) < \frac{\varepsilon}{2}.$$

In addition, if $n > 5/\delta$, by the previous discussion, we see that

$$\left\{\sup_{0\leqslant t\leqslant 1} \left|S_t^{(n)} - B_t^{(n)}\right| > \varepsilon,\ \frac{1}{n} \sup_{1\leqslant k\leqslant n} |\tau_k - k| \leqslant \frac{\delta}{5}\right\} \subseteq \left\{\bigcup_{\substack{s,t\in[0,2]\\ |t-s|\leqslant\delta}} \{|B_t^{(n)} - B_s^{(n)}| > \varepsilon\right\}.$$

Therefore, we conclude that for any $n > \max(N, 3/\delta)$,

$$\mathbb{P}\left(\sup_{0\leqslant t\leqslant 1} \left|S_t^{(n)} - B_t^{(n)}\right| > \varepsilon\right) < \varepsilon,$$

which gives the desired convergence in probability. □

By the Donsker invariance principle, we can easily obtain the central limit theorem in the i.i.d. case without using any characteristic function method!

Corollary 4.1. *Let $\{X_n\}$ be a sequence of i.i.d. random variables with $\sigma^2 \triangleq \mathbb{E}[X_1^2] < \infty$, and define $S_n = X_1 + \cdots + X_n$. Then for every $x \in \mathbb{R}^1$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - \mathbb{E}[S_n]}{\sigma\sqrt{n}} \leqslant x\right) = \Phi(x),$$

where $\Phi(x) \triangleq (2\pi)^{-1/2} \int_{-\infty}^x \mathrm{e}^{-u^2/2} du$ is the standard normal distribution function.

Proof. Without loss of generality, we may assume that $\mathbb{E}[X_1] = 0$. Let $\pi^1 :\ W^1 \to \mathbb{R}^1$ be the projection defined by $\pi_1(w) = w_1$. It follows that $\pi_1\left(S^{(n)}\right) = S_n/(\sigma\sqrt{n})$. By the Donsker invariance principle, for every bounded continuous function f on $\mathbb{R}^1$, we have:

$$\lim_{n\to\infty} \mathbb{E}\left[f\left(\frac{S_n}{\sigma\sqrt{n}}\right)\right] = \lim_{n\to\infty} \mathbb{E}\left[f \circ \pi_1\left(S^{(n)}\right)\right] = \mathbb{E}[f \circ \pi_1(B)] = \mathbb{E}[f(B_1)].$$

But B_1 is a standard normal random variable. Now the result follows from Theorem 1.6. □

4.4 Passage time distributions

In this subsection, we apply martingale methods, the strong Markov property and the reflection principle to perform a series of explicit computations related to passage times.

Given $c \in \mathbb{R}^1$, define a process $X_t = B_t + ct$, where B_t is a one dimensional Brownian motion. X_t is called the *Brownian motion with drift* c.

Consider $M_t \triangleq \exp(\theta X_t - \lambda t)$, where $\theta \in \mathbb{R}^1$ and $\lambda > 0$ are two parameters. It is straight forward to see that M_t is a martingale (with respect to the aumented natural filtration of B_t) if and only if

$$\frac{1}{2}\theta^2 - (\lambda - c\theta) = 0,$$

i.e. $\theta = \alpha \triangleq -c - \sqrt{c^2 + 2\lambda} < 0$ or $\theta = \beta \triangleq -c + \sqrt{c^2 + 2\lambda} > 0$. Now we use M_t to compute the Laplace transform of passage times for X_t.

We first consider the passage time of a single barrier.

For $x > 0$, define

$$\tau_x = \inf\{t \geqslant 0 : \ X_t = x\}.$$

Proposition 4.7. *The Laplace transform of τ_x is given by:*

$$\mathbb{E}[\mathrm{e}^{-\lambda \tau_x}] = \mathrm{e}^{-x(\sqrt{c^2+2\lambda} - c)}, \quad \lambda > 0. \tag{4.11}$$

In particular, we have

$$\mathbb{P}(\tau_x < \infty) = \begin{cases} 1, & c \geqslant 0; \\ \mathrm{e}^{2cx}, & c < 0. \end{cases} \tag{4.12}$$

Proof. By applying the optional sampling theorem to the martingale M_t with $\theta = \beta$, we know that

$$\mathbb{E}\left[\mathrm{e}^{\beta X_{\tau_x \wedge n} - \lambda \tau_x \wedge n}\right] = 1, \quad \forall n \geqslant 1.$$

But

$$\mathrm{e}^{\beta x - \lambda \tau_x} \mathbf{1}_{\{\tau_x < \infty\}} \overset{n \to \infty}{\longleftarrow} \mathrm{e}^{\beta X_{\tau_x \wedge n} - \lambda \tau_x \wedge n} \leqslant \mathrm{e}^{\beta x}.$$

Therefore,

$$\mathbb{E}\left[\mathrm{e}^{\beta x - \lambda \tau_x} \mathbf{1}_{\{\tau_x < \infty\}}\right] = 1,$$

which yields (4.11). (4.12) follows from letting $\lambda \downarrow 0$ in (4.11). $\square$

Now we consider the passage time of a double barrier.

For $a < 0 < b$, define τ_a, τ_b as before and $\tau_{a,b} \triangleq \tau_a \wedge \tau_b$.

Proposition 4.8. *The Laplace transform of $\tau_{a,b}$ is given by:*

$$\mathbb{E}\left[\mathrm{e}^{-\lambda \tau_{a,b}}\right] = \frac{\mathrm{e}^{\beta b} - \mathrm{e}^{\alpha b} + \mathrm{e}^{\alpha a} - \mathrm{e}^{\beta a}}{\mathrm{e}^{\beta b + \alpha a} - \mathrm{e}^{\beta a + \alpha b}}, \quad \lambda > 0. \tag{4.13}$$

Proof. Similarly with the proof of Proposition 4.7, by applying the optional sampling theorem to the martingale M_t with $\theta = \alpha$ and $\theta = \beta$ respectively, we conclude that

$$\mathbb{E}\left[\mathrm{e}^{\alpha X_{\tau_{a,b}} - \lambda \tau_{a,b}}\right] = 1, \ \mathbb{E}\left[\mathrm{e}^{\beta X_{\tau_{a,b}} - \lambda \tau_{a,b}}\right] = 1. \tag{4.14}$$

The first identity of (4.14) gives

$$\mathbb{E}\left[\mathrm{e}^{\alpha a - \lambda \tau_{a,b}} \mathbf{1}_{\{\tau_a < \tau_b\}}\right] + \mathbb{E}\left[\mathrm{e}^{\alpha b - \lambda \tau_{a,b}} \mathbf{1}_{\{\tau_a > \tau_b\}}\right] = 1,$$

and the second identity (4.14) gives

$$\mathbb{E}\left[\mathrm{e}^{\beta a - \lambda \tau_{a,b}} \mathbf{1}_{\{\tau_a < \tau_b\}}\right] + \mathbb{E}\left[\mathrm{e}^{\beta b - \lambda \tau_{a,b}} \mathbf{1}_{\{\tau_a > \tau_b\}}\right] = 1.$$

By solving these two equations for $\mathbb{E}\left[\mathrm{e}^{-\lambda \tau_{a,b}} \mathbf{1}_{\{\tau_a < \tau_b\}}\right]$ and $\mathbb{E}\left[\mathrm{e}^{-\lambda \tau_{a,b}} \mathbf{1}_{\{\tau_a > \tau_b\}}\right]$, we obtain (4.13) which is the sum of these two terms. □

By using the martingale M_t, we have seen how convenient it is in computing the Laplace transform of passage times. The Laplace transform can be used to compute moments very easily via differentiation. However, in many situations we are more interested in the entire distribution than just in the Laplace transform, and inverting the Laplace transform is often rather difficult.

In what follows, we are going to use the strong Markov property and the reflection principle to directly compute distributions related to passage times. The results are more general and powerful. For simplicity, we only consider the Brownian motion case without drift. The case with drift is treated in Problem Sheet 4, Problem 6 by a very inspiring and far-reaching method: change of measure.

Again we first consider the single barrier case.

For $t > 0$, let $S_t = \max_{0 \leqslant s \leqslant t} B_s$ be the running maximum of Brownian motion up to time t. We start by establishing a general formula for the joint distribution of (S_t, B_t). The distribution of passage times then follows easily.

Proposition 4.9. *For any $x, y \geqslant 0$, we have*

$$\mathbb{P}(S_t \geqslant x, \ B_t \leqslant x - y) = \mathbb{P}(B_t \geqslant x + y) = \frac{1}{\sqrt{2\pi}} \int_{\frac{x+y}{\sqrt{t}}}^{\infty} \mathrm{e}^{-\frac{u^2}{2}} du. \tag{4.15}$$

In particular, the joint density of (S_t, B_t) is given by

$$\mathbb{P}(S_t \in dx, \ B_t \in dy) = \frac{2(2x - y)}{\sqrt{2\pi t^3}} \mathrm{e}^{-\frac{(2x-y)^2}{2t}} dx dy, \quad x \geqslant 0, \ x \geqslant y, \tag{4.16}$$

and the density of τ_x ($x > 0$) is given by

$$\mathbb{P}(\tau_x \in dt) = \frac{x}{\sqrt{2\pi t^3}} \mathrm{e}^{-\frac{x^2}{2t}} dt, \quad t > 0. \tag{4.17}$$

Proof. Let $\widetilde{B}_t$ be the reflection of B_t at x defined by (4.6), and define $\widetilde{S}_t$ accordingly. From the reflection principle (c.f. Proposition 4.3), we know that $\widetilde{B}_t$ is also a Brownian motion. Together with the simple observation that $\{S_t \geqslant x\} = \left\{\widetilde{S}_t \geqslant x\right\}$, we arrive at

$$\begin{aligned}\mathbb{P}(S_t \geqslant x,\ B_t \leqslant x-y) &= \mathbb{P}\left(\widetilde{S}_t \geqslant x,\ \widetilde{B}_t \leqslant x-y\right) = \mathbb{P}\left(S_t \geqslant x,\ \widetilde{B}_t \leqslant x-y\right) \\ &= \mathbb{P}(S_t \geqslant x,\ B_t \geqslant x+y) = \mathbb{P}(B_t \geqslant x+y).\end{aligned}$$

Therefore, (4.15) follows. Now (4.16) follows by differentiation, and (4.17) follows from the fact that

$$\begin{aligned}\mathbb{P}(\tau_x \leqslant t) &= \mathbb{P}(S_t \geqslant x) = \mathbb{P}(S_t \geqslant x,\ B_t \leqslant x) + \mathbb{P}(S_t \geqslant x,\ B_t > x) \\ &= \mathbb{P}(B_t \geqslant x) + \mathbb{P}(B_t > x) = 2\mathbb{P}(B_t \geqslant x). \end{aligned} \tag{4.18}$$

□

Remark 4.2. From the formula (4.15), it is not hard to show that for each $t \geqslant 0$, $S_t - B_t \stackrel{\text{law}}{=} |B_t|$, and $2S_t - B_t \stackrel{\text{law}}{=} |B_t^{(3)}|$, where $B_t^{(3)}$ is the standard 3-dimensional Brownian motion. What is much more remarkable is the fact that $S - B \stackrel{\text{law}}{=} |B|$ and $2S - B \stackrel{\text{law}}{=} |\widetilde{B}|$ as stochastic processes. This result is closely related to the study of local times and excursion theory (c.f. Theorem 5.23 which we will not prove in this course).

The formula (4.15) also gives the marginal distribution of the absorbed Brownian motion. Given $x > 0$, let B_t^x be a one dimensional Brownian motion starting at position x. Define τ_0 to be the passage time of position 0 for B_t^x.

Corollary 4.2. *For $t > 0$, we have:*

$$\mathbb{P}(B_t^x \in dy,\ \tau_0 > t) = p_t(x,y) - p_t(x,-y), \quad y > 0,$$

where $p_t(x,y)$ is the Brownian transition density defined by (4.1) for $d = 1$.

Proof. According to the formula (4.15), we have

$$\mathbb{P}(B_t^x \geqslant y,\ \tau_0 \leqslant t) = \mathbb{P}(B_t^x \leqslant -y).$$

Therefore,

$$\mathbb{P}(B_t^x \geqslant y,\ \tau_0 > t) = \mathbb{P}(B_t^x \geqslant y) - \mathbb{P}(B_t^x \leqslant -y).$$

Now the result follows from differentiation. □

Finally, we consider the double barrier case. This is much more involved than the single barrier case.

Again let B_t^x be a one dimensional Brownian motion starting at x, where $0 < x < a$. Define $\tau_{0,a} = \tau_0 \wedge \tau_a$ to be the first exit time of the interval $(0,a)$ by B_t^x.

Proposition 4.10. *For $t > 0$, we have:*

$$\mathbb{P}(B_t^x \in dy,\ \tau_{0,a} > t) = \sum_{n=-\infty}^{\infty} (p_t(x, y+2na) - p_t(x, -y-2na))\, dy, \quad 0 < y < a. \tag{4.19}$$

In particular,

$$\begin{aligned}\mathbb{P}(\tau_{0,a} \in dt) &= \frac{1}{\sqrt{2\pi t^3}} \sum_{n=-\infty}^{\infty} \Bigg((2na+x)\mathrm{e}^{-\frac{(2na+x)^2}{2t}} \\ &\qquad + (2na+a-x)\mathrm{e}^{-\frac{(2na+a-x)^2}{2t}} \Bigg)\, dt, \quad t > 0. \end{aligned} \tag{4.20}$$

Proof. Define $\sigma_0 = 0$, $\theta_0 = \tau_0$, and for $n \geqslant 1$, define $\sigma_n = \inf\{t \geqslant \theta_{n-1} :\ B_t = a\}$, $\theta_n = \inf\{t \geqslant \sigma_n :\ B_t = 0\}$. By using the reflection principle (indeed a slightly more general version for the stopping time θ_n, but the proof is the same), we can see that for every $y > 0$,

$$\mathbb{P}(B_t^x \geqslant y,\ \theta_n \leqslant t) = \mathbb{P}(B_t^x \leqslant -y,\ \theta_n \leqslant t).$$

But by the definition of σ_n and θ_n, we know that $\{B_t^x \leqslant -y,\ \theta_n \leqslant t\} = \{B_t^x \leqslant -y,\ \sigma_n \leqslant t\}$. Therefore, we have

$$\mathbb{P}(B_t^x \geqslant y,\ \theta_n \leqslant t) = \mathbb{P}(B_t^x \leqslant -y,\ \sigma_n \leqslant t). \tag{4.21}$$

Similarly, for every $y < a$, we have

$$\mathbb{P}(B_t^x \leqslant y,\ \sigma_n \leqslant t) = \mathbb{P}(B_t^x \geqslant 2a - y,\ \sigma_n \leqslant t) = \mathbb{P}(B_t^x \geqslant 2a - y,\ \theta_{n-1} \leqslant t). \tag{4.22}$$

Now (4.21) and (4.22) can be used recursively in pair to obtain that for every $0 < y < a$,

$$\begin{aligned}\mathbb{P}(B_t^x \geqslant y,\ \theta_n \leqslant t) &= \mathbb{P}(B_t^x \leqslant -y - 2na), \\ \mathbb{P}(B_t^x \leqslant y,\ \sigma_n \leqslant t) &= \mathbb{P}(B_t^x \leqslant y - 2na).\end{aligned}$$

By differentiation, we arrive at

$$\begin{aligned}\mathbb{P}(B_t^x \in dy,\ \theta_n \leqslant t) &= p_t(x, -y - 2na)dy, \\ \mathbb{P}(B_t^x \in dy,\ \sigma_n \leqslant t) &= p_t(x, y - 2na)dy,\end{aligned} \tag{4.23}$$

for $0 < y < a$.

Symmetrically, we define $\pi_0 = 0$, $\rho_0 = \tau_a$ and for $n \geqslant 1$, define $\pi_n = \inf\{t \geqslant \rho_{n-1} :\ B_t = 0\}$, $\rho_n = \inf\{t \geqslant \pi_n :\ B_t = a\}$. By the same argument, we conclude that for every $0 < y < a$,

$$\begin{aligned}\mathbb{P}(B_t^x \in dy,\ \rho_n \leqslant t) &= p_t(x, -y + 2(n+1)a)dy, \\ \mathbb{P}(B_t^x \in dy,\ \pi_n \leqslant t) &= p_t(x, y + 2na)dy.\end{aligned} \tag{4.24}$$

Now the key observation is that $\theta_{n-1} \vee \rho_{n-1} = \sigma_n \wedge \pi_n$ and $\sigma_n \vee \pi_n = \theta_n \wedge \rho_n$ for every $n \geqslant 1$, which can be seen easily by considering the cases $\tau_0 < \tau_a$ and $\tau_0 > \tau_a$. Therefore, we have

$$
\begin{aligned}
& \mathbb{P}(B_t^x \in dy,\ \theta_0 \wedge \rho_0 \leqslant t) \\
= \ & \mathbb{P}(B_t^x \in dy,\ \theta_0 \leqslant t) + \mathbb{P}(B_t^x \in dy,\ \rho_0 \leqslant t) - \mathbb{P}(B_t^x \in dy,\ \theta_0 \vee \rho_0 \leqslant t) \\
= \ & \mathbb{P}(B_t^x \in dy,\ \theta_0 \leqslant t) + \mathbb{P}(B_t^x \in dy,\ \rho_0 \leqslant t) - \mathbb{P}(B_t^x \in dy,\ \sigma_1 \wedge \pi_1 \leqslant t) \\
= \ & \mathbb{P}(B_t^x \in dy,\ \theta_0 \leqslant t) + \mathbb{P}(B_t^x \in dy,\ \rho_0 \leqslant t) - \mathbb{P}(B_t^x \in dy,\ \sigma_1 \leqslant t) \\
& -\mathbb{P}(B_t^x \in dy,\ \pi_1 \leqslant t) + \mathbb{P}(B_t^x \in dy,\ \theta_1 \wedge \rho_1 \leqslant t).
\end{aligned}
$$

By induction, we arrive at

$$
\begin{aligned}
\mathbb{P}(B_t^x \in dy,\ \theta_0 \wedge \rho_0 \leqslant t) \ = \ & \sum_{k=1}^{n} (\mathbb{P}(B_t^x \in dy,\ \theta_{k-1} \leqslant t) + \mathbb{P}(B_t^x \in dy,\ \rho_{k-1} \leqslant t) \\
& -\mathbb{P}(B_t^x \in dy,\ \sigma_k \leqslant t) - \mathbb{P}(B_t^x \in dy,\ \pi_k \leqslant t)) \\
& +\mathbb{P}(B_t^x \in dy,\ \theta_n \wedge \rho_n \leqslant t). \qquad (4.25)
\end{aligned}
$$

Finally, to see that the last term vanishes as $n \to \infty$, first note that according to the strong Markov property, both $\theta_n - \sigma_n$ and $\sigma_n - \theta_{n-1}$ have the same distribution as the passage time of a for a Brownian motion starting at the origin. In particular, according to (4.11), the Laplace transforms of $\theta_n - \sigma_n$ and $\sigma_n - \theta_{n-1}$ are both given by $\mathrm{e}^{-a\sqrt{2\lambda}}$. Moreover,

$$\theta_n = \theta_0 + (\sigma_1 - \theta_0) + (\theta_1 - \sigma_1) + \cdots + (\sigma_n - \theta_{n-1}) + (\theta_n - \sigma_n)$$

is a sum of indepentent random variables. Therefore, the Laplace transform of θ_n is given by $\mathrm{e}^{-x\sqrt{2\lambda}} \cdot \left(\mathrm{e}^{-a\sqrt{2\lambda}}\right)^{2n} = \mathrm{e}^{-(x+2na)\sqrt{2\lambda}}$. In particular, θ_n has the same distribution as the passage time of $x + 2na$ for a Brownian motion starting at the origin. From (4.18) we conclude that

$$\mathbb{P}(\theta_n \leqslant t) = 2\mathbb{P}(B_t \geqslant x + 2na) \to 0$$

as $n \to \infty$. Similarly, $\lim_{n\to\infty} \mathbb{P}(\rho_n \leqslant t) = 0$. Therefore,

$$\lim_{n\to\infty} \mathbb{P}(B_t^x \in dy,\ \theta_n \wedge \rho_n \leqslant t) = 0.$$

Now (4.19) follows from substituting (4.23), (4.24) in (4.25), letting $n \to \infty$ and rearranging terms.

(4.20) follows from integrating over $0 < y < a$ in (4.19) and differentiating with respect to t. □

The careful reader might use (4.11) and (4.17) to conclude that

$$
\begin{aligned}
\mathcal{L}\left[\frac{1}{\sqrt{2\pi t^3}} \sum_{n=-\infty}^{\infty} (2na + x)\mathrm{e}^{-\frac{(2na+x)^2}{2t}} dt \ (t > 0)\right] \ &= \ \mathbb{E}\left[\mathrm{e}^{-\tau_0} \mathbf{1}_{\{\tau_0 < \tau_a\}}\right], \\
\mathcal{L}\left[\frac{1}{\sqrt{2\pi t^3}} \sum_{n=-\infty}^{\infty} (2na + a - x)\mathrm{e}^{-\frac{(2na+a-x)^2}{2t}} dt \ (t > 0)\right] \ &= \ \mathbb{E}\left[\mathrm{e}^{-\tau_a} \mathbf{1}_{\{\tau_a < \tau_0\}}\right],
\end{aligned}
$$

where $\mathcal{L}$ denotes the Laplace transform operator, and the expectations on the right hand side are indeed computed in the proof of Proposition 4.8 by using martingale methods. Therefore, we obtain the following corollary.

Corollary 4.3. *We have:*

$$\begin{aligned}\mathbb{P}(\tau_0 \in dt,\ \tau_0 < \tau_a) &= \frac{1}{\sqrt{2\pi t^3}} \sum_{n=-\infty}^{\infty} (2na + x)\mathrm{e}^{-\frac{(2na+x)^2}{2t}} dt, \\ \mathbb{P}(\tau_a \in dt,\ \tau_a < \tau_0) &= \frac{1}{\sqrt{2\pi t^3}} \sum_{n=-\infty}^{\infty} (2na + a - x)\mathrm{e}^{-\frac{(2na+a-x)^2}{2t}} dt, \quad t > 0.\end{aligned}$$

Remark 4.3. From the results obtained so far, we are indeed able to derive the distributions of $B^x_{t\wedge\tau_0}$ (the single barrier case) and $B^x_{t\wedge\tau_{0,a}}$ (the double barrier case) respectively for given $t > 0$.

4.5 Sample path properties: an overview

So far we have been dealing with distributional properties of Brownian motion. However, the study of sample path properties of Brownian motion is also a huge and important topic, in which we may find a variety of interesting and striking results. As we are mainly interested in the probabilistic side in this course, we will only give an overview on the basic results along this direction. We do give a detailed proof of the fact that almost every Brownian sample path has infinite p-variation ($1 \leqslant p < 2$) on every finite interval. This reveals the fundamental obstacle to expecting a classical deterministic theory of differential calculus for Brownian motion.

We assume that B_t is a one dimensional Brownian motion.

1. Oscillations

We know from Problem Sheet 2, Problem 2, (2) that with probability one, the Brownian motion is γ-Hölder continuous on every finite interval for $0 < \gamma < 1/2$, and it fails to be so if $\gamma = 1/2$. It is very natural to ask what is the precise rate of oscillation for Brownian motion.

At every given point $t \geqslant 0$, the exact rate of oscillation is given by Khinchin in his celebrated law of the iterated logarithm.

Theorem 4.5. *Let $\varphi(h) = \sqrt{2h \log\log 1/h}$ ($h > 0$). Then for every given $t \geqslant 0$, we have:*

$$\mathbb{P}\left(\limsup_{h\downarrow 0} \frac{B_{t+h} - B_t}{\varphi(h)} = 1\right) = 1.$$

It is far from being true that Khinchin's law of the iterated logarithm holds uniformly in t with probability one. Indeed, it is Lévy who discovered the exact modulus of continuity for Brownian motion.

Theorem 4.6. *Let $\psi(h)=\sqrt{2h\log 1/h}$ $(h>0)$. Then for every $T>0$, we have:*

$$\mathbb{P}\left(\limsup_{h\downarrow 0}\sup_{\substack{0\leqslant s<t\leqslant T\\ t-s\leqslant h}}\frac{|B_t-B_s|}{\psi(h)}=1\right)=1.$$

The curious reader might wonder how big the gap is between Khinchin's law of the iterated logarithm and Lévy's modulus of continuity theorem. Indeed, the set of times at which Khinchin's law of the iterated logarithm fails is much larger than we can imagine: with probability one, the random set

$$\left\{t\in[0,1]:\ \limsup_{h\downarrow 0}\frac{B_{t+h}-B_t}{\psi(h)}=1\right\}$$

is uncountable and dense in $[0,1]$, and random the set

$$\left\{t\in[0,1]:\ \limsup_{h\downarrow 0}\frac{B_{t+h}-B_t}{\varphi(h)}=\infty\right\}$$

has Hausdorff dimension one (c.f. [6]).

2. Irregularity

If this is the first time that we encounter Brownian motion, it is really hard to believe how irregular a Brownian sample path can be.

Theorem 4.7. *With probability one, the following properties hold:*

(1) $t\mapsto B_t(\omega)$ is nowhere differentiable;

(2) the set of local maximum points for $t\mapsto B_t(\omega)$ is countable and dense in $[0,\infty)$, and every local maximum is a strict local maximum;

(3) $t\mapsto B_t(\omega)$ has no points of increase (t is a point of increase of a path x if there exists $\delta>0$, such that $x_s\leqslant x_t\leqslant x_u$ for all $s\in((t-\delta)^+,t)$ and $u\in(t,t+\delta)$);

(4) for given $x\in\mathbb{R}^1$, the level set $\{t\geqslant 0:\ B_t(\omega)=x\}$ is closed,unbounded, with zero Lebesgue measure, and does not contain isolated points.

3. The p-variation of Brownian motion

Let $x:\ [0,\infty)\to(E,\rho)$ be a continuous path in some metric space (E,ρ). Recall that for $p\geqslant 1$, the *p-variation* of x over $[s,t]$ is define to be

$$\|x\|_{p-\mathrm{var};[s,t]}=\sup_{\mathcal{P}}\left(\sum_k\rho(x_{t_{k-1}},x_{t_k})^p\right)^{\frac{1}{p}},$$

where the supremum runs over all finite partitions $\mathcal{P}$ of $[s,t]$.

We first show that B_t has finite 2-variation (or finite quadratic variation) on any finite interval in certain probabilistic sense.

Proposition 4.11. *Given $t>0$, let $\mathcal{P}_n:\ 0=t_0^n<t_1^n<\cdots<t_{m_n}^n=t$ be a sequence of finite partitions of $[0,t]$ such that $\mathrm{mesh}(\mathcal{P}_n)\to 0$ as $n\to\infty$. Then*

$$\lim_{n\to\infty}\sum_{k=1}^{m_n}(B_{t_k^n}-B_{t_{k-1}^n})^2=t\ \text{ in } L^2.$$

If we further assume that $\sum_{n=1}^{\infty}\mathrm{mesh}(\mathcal{P}_n)<\infty$, then the convergence holds almost surely.

Proof. Since B has independent increments, we have

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{k=1}^{m_n}\left(B_{t_k^n}-B_{t_{k-1}^n}\right)^2-t\right)^2\right] &= \mathbb{E}\left[\left(\sum_{k=1}^{m_n}\left(\left(B_{t_k^n}-B_{t_{k-1}^n}\right)^2-(t_k^n-t_{k-1}^n)\right)\right)^2\right]\\ &= \sum_{k=1}^{m_n}\mathbb{E}\left[\left(\left(B_{t_k^n}-B_{t_{k-1}^n}\right)^2-(t_k^n-t_{k-1}^n)\right)^2\right]\\ &= 2\sum_{k=1}^{m_n}(t_k^n-t_{k-1}^n)^2\\ &\leqslant 2t\cdot\mathrm{mesh}(\mathcal{P}_n), \qquad (4.26)\end{aligned}$$

where we have also used the fact that $B_v-B_u\sim\mathcal{N}(0,v-u)$ for $u<v$ and $\mathbb{E}[Y^4]=3\left(\mathbb{E}[Y^2]\right)^2$ for a centered Gaussian random variable Y. The L^2-convergence then follows immediately from (4.26). If we further assume that $\sum_{n=1}^{\infty}\mathrm{mesh}(\mathcal{P}_n)<\infty$, then by the Chebyshev inequality and (4.26), we conclude that

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\sum_{k=1}^{m_n}\left(B_{t_k^n}-B_{t_{k-1}^n}\right)^2-t\right|>\varepsilon\right)<\infty$$

for every $\varepsilon>0$. The almost sure convergence then follows from the Borel-Cantelli lemma. □

Proposition 4.11 enables us to prove the following sample path property, which puts a serious negative effect to the theory.

Corollary 4.4. *For every $1\leqslant p<2$, with probability one, B_t has infinite p-variation on every finite interval $[s,t]$.*

Proof. Given any finite partition $\mathcal{P}:\ s=t_0<t_1<\cdots<t_n=t$ of $[s,t]$, we know that

$$\begin{aligned}\sum_k\left|B_{t_k}-B_{t_{k-1}}\right|^2 &\leqslant \left(\max_k\left|B_{t_k}-B_{t_{k-1}}\right|^{2-p}\right)\cdot\sum_k\left|B_{t_k}-B_{t_{k-1}}\right|^p\\ &\leqslant \left(\max_k\left|B_{t_k}-B_{t_{k-1}}\right|^{2-p}\right)\|B\|_{p-\mathrm{var};[s,t]}^p. \qquad (4.27)\end{aligned}$$

If we take a sequence of finite partitions $\mathcal{P}_n$ of $[s,t]$ such that $\sum_{n=1}^{\infty} \mathrm{mesh}(\mathcal{P}_n) < \infty$, by Proposition 4.11 and the continuity of Brownian motion, we know that with probability one, the left hand side of (4.27) converges to $t-s>0$ and the first term on the right hand side of (4.27) converges to zero. Therefore, $\|B\|_{p-\mathrm{var};[s,t]} = \infty$ almost surely. To see that the statement is uniform with respect to all $[s,t]$, we only need to run over all possible $s,t \in \mathbb{Q}$. □

Remark 4.4. From the local Hölder continuity of Brownian sample paths, it is easy to see that for every $p>2$, with probability one, B_t has finite p-variation on every finite interval. However, on the borderline $p=2$, the fact that $\|B\|_{2-\mathrm{var};[s,t]} = \infty$ almost surely is much harder to establish (c.f. [3]).

The result of Corollary 4.4 destroys any hope of establishing a pathwise theory of integration and differential equations for Brownian motion in the classical sense of Lebesgue-Stieltjes ($p=1$) or Young ($1<p<2$). It is indeed the fact that B_t and $B_t^2 - t$ are both martingales leads us to the realm of Itô's calculus, an elegant L^2*-theory* of stochastic integration and differential equations, which has profound impacts on pure and applied mathematics.

5 Stochastic integration

In this section, we develop Itô's theory of stochastic integration. As we have seen in the last section, the sample path properties of Brownina motion force us to deviate from the classical approach, and we should look for a more probabilistic counterpart of calculus in the context of Brownian motion, or more generally, of continuous semimartingales. A price to pay is that differentiation is no longer meaningful, and everything is understood in an integral sense. The core result in the theory is the renowed Itô's formula–a general change of variable formula for continuous semimartingales which is fumdamentally differential from the classical one. We will see a long series of exciting and important applications of Itô's formula in the rest of our study.

Through out the rest of this section, unless otherwise stated, we always assume that $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ is a filtered probability space which satisfies the usual conditions. All stochastic processes are defined on this setting.

5.1 L^2-bounded martingales and the bracket process

Taking a functional analytic viewpoint, the key ingredient to establishing Itô's integration is the use of a Hilbert structure and isometry. Hence we start with the study of L^2-bounded martingales.

Definition 5.1. A càdlàg martingale $\{M_t, \mathcal{F}_t\}$ is called an L^2*-bounded martingale* if

$$\sup_{t\geqslant 0} \mathbb{E}[M_t^2] < \infty.$$

The space of L^2-bounded martingales is denoted by $\mathbb{H}^2$. We use H^2 (H_0^2, respectively) to denote the subspace of L^2-bounded continuous martingales (vanishing at $t = 0$, respectively).

It is immediate that an L^2-bounded martingale $\{M_t, \mathcal{F}_t\}$ is uniformly integrable. Therefore, M_t converges to some $M_\infty \in \mathcal{F}_\infty$ almost surely and in L^1, and we have

$$M_t = \mathbb{E}[M_\infty | \mathcal{F}_t]. \tag{5.1}$$

Moreover, from Problem Sheet 3, Problem 3, we know that the convergence holds in L^2 as well. Therefore, the relation (5.1) sets up a one-to-one correspondence between $\mathbb{H}^2$ (modulo indistinguishability) and $L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$.

Proposition 5.1. *The space $\mathbb{H}^2$ (modulo indistinguishability) is a Hilbert space when equipped with the inner product*

$$\langle M, N \rangle_{\mathbb{H}^2} \triangleq \mathbb{E}[M_\infty N_\infty], \quad M, N \in \mathbb{H}^2.$$

The space H_0^2 is a closed subspace of $\mathbb{H}^2$.

Proof. The first claim follows simply because $L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$ is a Hilbert space. To prove the second claim, let $M^{(n)}$ be a sequence of L^2-bounded continuous martingales converging to some $M \in \mathbb{H}^2$ under the $\mathbb{H}^2$-metric. An application of Doob's L^p-inequality with $p = 2$ (c.f Corollary 3.3 and Section 3.5) shows that

$$\mathbb{E}\left[\left(\sup_{t \geqslant 0}\left|M_t^{(n)} - M_t\right|\right)^2\right] \leqslant 4\left\|M_\infty^{(n)} - M_\infty\right\|_{L^2}^2.$$

In particular, along a subsequence $M^{(n_k)}$, we know that with probability one, $M_t^{(n_k)}$ converges to M_t uniformly in t as $k \to \infty$. This shows that M_t is continuous. □

In this course, we will only focus on the continuous situation.

Lemma 5.1. *Let $\{M_t, \mathcal{F}_t\}$ be a continuous martingale such that with probability one, the sample paths of M_t has bounded variation on every finite interval. Then $M_t \equiv M_0$ for all t.*

Proof. By considering $M_t - M_0$ we may assume that $M_0 = 0$. Let $V_t = \|M\|_{1-\mathrm{var};[0,t]}$ be the one variation process of M_t. We first consider the case when V_t is uniformly bounded by some constant $C > 0$. In this case, for a given finite partition $\mathcal{P}$ of $[0, t]$, we have

$$\begin{aligned} \mathbb{E}[M_t^2] &= \sum_{i \in \mathcal{P}} \mathbb{E}[M_{t_i}^2 - M_{t_{i-1}}^2] = \sum_{i \in \mathcal{P}} \mathbb{E}[(M_{t_i} - M_{t_{i-1}})^2] \\ &\leqslant \mathbb{E}\left[V_t \cdot \max_{i \in \mathcal{P}} |M_{t_i} - M_{t_{i-1}}|\right] \leqslant C\mathbb{E}\left[\max_{i \in \mathcal{P}} |M_{t_i} - M_{t_{i-1}}|\right]. \end{aligned} \quad (5.2)$$

Since M is continuous, from the dominated convergence theorem we know that the right hand side of (5.2) converges to zero as $\mathrm{mesh}(\mathcal{P}) \to 0$. Therefore, $M_t = 0$.

In the general case, let $\tau_n = \inf\{t \geqslant 0 : \ V_t \geqslant n\}$. Then τ_n is an $\{\mathcal{F}_t\}$-stopping time with $\tau_n \uparrow \infty$ almost surely. From Problem Sheet 3, Problem 1, (1), we know that the stopped process $M_t^{\tau_n} \triangleq M_{\tau_n \wedge t}$ is an $\{\mathcal{F}_t\}$-martingale whose one variation process is bounded by n. Therefore, $M_t^{\tau_n} = 0$. By letting $n \to \infty$, we conclude that $M_t = 0$. □

The following result plays a fundamental role in establishing an L^2-theory of stochastic integration.

Theorem 5.1. *Let $M \in H_0^2$. Then there exists a unique (up to indistinguishability) continuous, $\{\mathcal{F}_t\}$-adapted process $\langle M \rangle_t$ which vanishes at $t = 0$ and has bounded variation on every finite interval, such that $M_t^2 - \langle M \rangle_t$ is an $\{\mathcal{F}_t\}$-martingale.*

Proof. We first prove uniqueness. Suppose A_t and A'_t are two such processes. Then $\{A'_t - A_t, \mathcal{F}_t\}$ is a continuous martingale with bounded variation on every finite interval. According to Lemma 5.1, we conclude that $A = A'$.

Since $\{M_t^2, \mathcal{F}_t\}$ is a non-negative and continuous submartingale, from Problem Sheet 3, Problem 7, (1), we know that M_t^2 is of class (DL) and regular (c.f. Definition 3.12 and Definition 3.13). The existence of $\langle M \rangle_t$ then follows immediately from the Doob-Meyer decomposition theorem (c.f Theorem 3.13) and Theorem 3.14. □

It is immediate to see that $\langle M\rangle_t$ is an increasing process in the sense of Definition 3.10 and $\|M\|_{\mathbb{H}^2} = \sqrt{\mathbb{E}[\langle M\rangle_\infty]} < \infty$ for $M \in H_0^2$.

Definition 5.2. The process $\langle M\rangle_t$ defined in Theorem 5.1 is called the *quadratic variation process* of M_t.

In general, the class H_0^2 is too restrictive to serve our study in many interesting situations. It is unnatural to impose a priori integrability conditions on the process we are considering. To extend our study, it is important to have some kind of localization method. We have already seen this in the proof of Lemma 5.1.

Definition 5.3. A continuous, $\{\mathcal{F}_t\}$-adapted process M_t is called a *continuous local martingale* if there exists a sequence τ_n of $\{\mathcal{F}_t\}$-stopping times such that $\tau_n \uparrow \infty$ almost surely, and the stopped process $(M - M_0)_t^{\tau_n} \triangleq M_{\tau_n \wedge t} - M_0$ is an $\{\mathcal{F}_t\}$-martingale for every n. We use $\mathcal{M}^{\mathrm{loc}}$ ($\mathcal{M}_0^{\mathrm{loc}}$, respectively) to denote the space of continuous local martingales (vanishing at $t = 0$, respectively).

Remark 5.1. If M_t is a continuous, $\{\mathcal{F}_t\}$-adapted process vanishing at $t = 0$, we can define a sequence of finite $\{\mathcal{F}_t\}$-stopping times by $\sigma_n = \inf\{t \geqslant 0 : \ |M_t| \geqslant n\} \wedge n$. It is obvious that $\sigma_n \uparrow \infty$ almost surely. If $M \in \mathcal{M}_0^{\mathrm{loc}}$ with a localization sequence τ_n, then $M_t^{\sigma_n \wedge \tau_n}$ is a bounded $\{\mathcal{F}_t\}$-martingale for each n. Therefore, for $M \in \mathcal{M}_0^{\mathrm{loc}}$, whenever convenient, it is not harmful to assume that the stopped martingale $M_t^{\tau_n}$ in Definition 5.3 is bounded for each n.

From the definition, it is easy to see that $\mathcal{M}^{\mathrm{loc}}$ is a vector space. Moreover, if $\{M_t, \mathcal{F}_t\}$ is a continuous local martingale and τ is an $\{\mathcal{F}_t\}$-stopping time, then the stopped process M_t^τ is also a continuous local martingale.

Every continuous martingale is a continuous local martingale (simply take $\tau_n = n$). However, we must point out that a continuous local martingale can fail to be a martingale, even if we impose strong integrability conditions (for instance, exponential integrability or uniform integrability). We will encounter important examples of continuous local martingales which are not martingales in the study of stochastic differential equations.

The following result gives us a simple idea about the relationships between local martingales and martingales. The proof is easy and hence omitted.

Proposition 5.2. *A non-negative, integrable, continuous local martingale is a supermartingale. A continuous local martingale is a martingale if and only if it is of class (DL).*

By certain localization argument, we can also define the quadratic variation of a local martinagle $M \in \mathcal{M}_0^{\mathrm{loc}}$.

Theorem 5.2. *Let $M \in \mathcal{M}_0^{\mathrm{loc}}$. Then there exists a unique (up to indistinguishability) continuous, $\{\mathcal{F}_t\}$-adapted process $\langle M\rangle_t$ which vanishes at $t = 0$ and has bounded variation on every finite interval, such that $M^2 - \langle M\rangle \in \mathcal{M}_0^{\mathrm{loc}}$. The sample paths of the process $\langle M\rangle_t$ are indeed increasing.*

Proof. We first prove existence.

According to Remark 5.1, we may assume that there exists a sequence τ_n of finite $\{\mathcal{F}_t\}$-stopping times such that $\tau_n \uparrow \infty$ almost surely and $M_t^{\tau_n}$ is a bounded $\{\mathcal{F}_t\}$-martingale vanishing at $t = 0$ for each n. According to Theorem 5.1, we can define the quadratic variation process $\langle M^{\tau_n}\rangle_t$ for $M_t^{\tau_n}$ such that $(M_t^{\tau_n})^2 - \langle M^{\tau_n}\rangle_t$ is an $\{\mathcal{F}_t\}$-martingale.

Now we know that $M^2_{\tau_{n+1}\wedge\tau_n\wedge t} - \langle M^{\tau_{n+1}}\rangle_{\tau_n\wedge t} = M^2_{\tau_n\wedge t} - \langle M^{\tau_{n+1}}\rangle_{\tau_n\wedge t}$ and $M^2_{\tau_n\wedge t} - \langle M^{\tau_n}\rangle_t$ are both $\{\mathcal{F}_t\}$-martingales. By Lemma 5.1, with probability one, we have

$$\langle M^{\tau_{n+1}}\rangle_{\tau_n\wedge t} = \langle M^{\tau_n}\rangle_t, \quad \forall t \geqslant 0.$$

In other words, $\langle M^{\tau_{n+1}}\rangle_t = \langle M^{\tau_n}\rangle_t$ on $[0, \tau_n]$. This enables us to define a continuous, $\{\mathcal{F}_t\}$-adapted process $\langle M\rangle_t \triangleq \lim_{n\to\infty}\langle M^{\tau_n}\rangle_t$ which vanishes at $t = 0$ and obviously has increasing sample paths. Moreover, since $\langle M\rangle_{\tau_n\wedge t} = \langle M^{\tau_n}\rangle_t$, we conclude that $M^2_{\tau_n\wedge t} - \langle M\rangle_{\tau_n\wedge t}$ is an $\{\mathcal{F}_t\}$-martingale. Therefore, $M^2 - \langle M\rangle \in \mathcal{M}_0^{\mathrm{loc}}$.

The uniqueness of $\langle M\rangle_t$ follows from the fact that Lemma 5.1 holds for continuous local martingales as well, which can be easily shown by a similar localization argument. □

For $M \in \mathcal{M}_0^{\mathrm{loc}}$, the process $\langle M\rangle_t$ is also called the *quadratic variation process* of M_t.

In the intrinsic characterization of stochastic integrals as we will see later on, it is important to consider more generally the "bracket" of two local martingales.

Let $M, N \in \mathcal{M}_0^{\mathrm{loc}}$. Define

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

Since $\mathcal{M}_0^{\mathrm{loc}}$ is a vector space, we can see that $\langle M, N\rangle_t$ is the unique (up to indistinguishability) continuous, $\{\mathcal{F}_t\}$-adapted process which vanishes at $t = 0$ and has bounded variation on every finite interval, such that $M \cdot N - \langle M, N\rangle \in \mathcal{M}_0^{\mathrm{loc}}$.

Definition 5.4. For $M, N \in \mathcal{M}_0^{\mathrm{loc}}$, the process $\langle M, N\rangle_t$ is called the *bracket process* of M and N.

The bracket process is compatible with localization.

Proposition 5.3. *Let $M, N \in \mathcal{M}_0^{\mathrm{loc}}$ and let τ be an $\{\mathcal{F}_t\}$-stopping time. Then*

$$\langle M^\tau, N^\tau\rangle = \langle M^\tau, N\rangle = \langle M, N\rangle^\tau.$$

Proof. The fact that $\langle M^\tau, N^\tau\rangle = \langle M, N\rangle^\tau$ follows from the stability of $\mathcal{M}_0^{\mathrm{loc}}$ under stopping and the uniqueness property of the bracket process. To see the other identity, it suffices to show that $M^\tau(N - N^\tau) \in \mathcal{M}_0^{\mathrm{loc}}$. By localization along a suitable sequence of $\{\mathcal{F}_t\}$-stopping times, we may assume that M, N are both bounded $\{\mathcal{F}_t\}$-martingales. In this case, for $s < t$, we have

$$\begin{aligned}\mathbb{E}[M_{\tau\wedge t}(N_t - N_{\tau\wedge t})|\mathcal{F}_s] &= \mathbb{E}[M_{\tau\wedge t}\mathbf{1}_{\{\tau\leqslant s\}}(N_t - N_{\tau\wedge t})|\mathcal{F}_s] \\ &\quad +\mathbb{E}[M_{\tau\wedge t}\mathbf{1}_{\{\tau> s\}}(N_t - N_{\tau\wedge t})|\mathcal{F}_s].\end{aligned}$$

The first term equals

$$\begin{aligned}\mathbb{E}[M_{\tau\wedge s}\mathbf{1}_{\{\tau\leqslant s\}}(N_t-N_{\tau\wedge t})|\mathcal{F}_s] &= M_{\tau\wedge s}\mathbf{1}_{\{\tau\leqslant s\}}\mathbb{E}[N_t-N_{\tau\wedge t}|\mathcal{F}_s]\\ &= M_{\tau\wedge s}\mathbf{1}_{\{\tau\leqslant s\}}(N_s-N_{\tau\wedge s})\\ &= M_{\tau\wedge s}(N_s-N_{\tau\wedge s}).\end{aligned}$$

The second term equals

$$\begin{aligned}\mathbb{E}[M_{\tau\wedge t}\mathbf{1}_{\{\tau>s\}}(N_t-N_{\tau\wedge t})|\mathcal{F}_{\tau\wedge s}] &= \mathbb{E}\left[\mathbb{E}[M_{\tau\wedge t}\mathbf{1}_{\{\tau>s\}}(N_t-N_{\tau\wedge t})|\mathcal{F}_{\tau\wedge t}]|\mathcal{F}_{\tau\wedge s}\right]\\ &= \mathbb{E}\left[M_{\tau\wedge t}\mathbf{1}_{\{\tau>s\}}\mathbb{E}[N_t-N_{\tau\wedge t}|\mathcal{F}_{\tau\wedge t}]|\mathcal{F}_{\tau\wedge s}\right]\\ &= 0.\end{aligned}$$

Therefore, $M_t^\tau(N_t-N_t^\tau)$ is an $\{\mathcal{F}_t\}$-martingale. □

The bracket process behaves pretty much like an inner product. Indeed, we have the following simple but useful properties.

Proposition 5.4. *Let $M,M_1,M_2\in\mathcal{M}_0^{\mathrm{loc}}$, and let $\alpha,\beta\in\mathbb{R}^1$. Then with probability one, we have:*
(1) $\langle\alpha M_1+\beta M_2,M\rangle=\alpha\langle M_1,M\rangle+\beta\langle M_2,M\rangle$;
(2) $\langle M_1,M_2\rangle=\langle M_2,M_1\rangle$;
(3) $\langle M,M\rangle=\langle M\rangle\geqslant 0$, and $\langle M\rangle=0$ if and only if $M=0$.

Proof. We only prove the last part of (3). All the rest assertions are straight forward applications of the uniqueness property of the bracket process. Suppose that $\langle M\rangle=0$. It follows that $M^2\in\mathcal{M}_0^{\mathrm{loc}}$. Let τ_n be a sequence of $\{\mathcal{F}_t\}$-stopping times such that $\tau_n\uparrow\infty$ almost surely and $(M^2)_t^{\tau_n}$ is a bounded $\{\mathcal{F}_t\}$-martingale. Then we have $\mathbb{E}[M_{\tau_n\wedge t}^2]=\mathbb{E}[M_0^2]=0$ for any given $t\geqslant 0$, which implies that $M_{\tau_n\wedge t}=0$. By letting $n\to\infty$, we conclude that $M_t=0$. □

In exactly the same way as for inner products, Proposition 5.4 enables us to prove the following Cauchy-Schwarz inequality.

Proposition 5.5. *Let $M,N\in\mathcal{M}_0^{\mathrm{loc}}$. Then $|\langle M,N\rangle|\leqslant\langle M\rangle^{1/2}\cdot\langle N\rangle^{1/2}$ almost surely. More generally, with probability one, we have:*

$$|\langle M,N\rangle_t-\langle M,N\rangle_s|\leqslant(\langle M\rangle_t-\langle M\rangle_s)^{\frac{1}{2}}\cdot(\langle N\rangle_t-\langle N\rangle_s)^{\frac{1}{2}},\quad\forall 0\leqslant s<t. \tag{5.3}$$

What is really useful for us is the following extension of inequality (5.3).

Proposition 5.6 (The Kunita-Watanabe inequality)**.** *Let $M,N\in\mathcal{M}_0^{\mathrm{loc}}$, and let X_t,Y_t be two stochastic processes which have measurable sample paths almost surely. Then with probability one, we have:*

$$\int_0^t|X_s|\cdot|Y_s|d\|\langle M,N\rangle\|_s\leqslant\left(\int_0^t X_s^2d\langle M\rangle_s\right)^{\frac{1}{2}}\cdot\left(\int_0^t Y_s^2d\langle N\rangle_s\right)^{\frac{1}{2}},\quad\forall t\geqslant 0, \tag{5.4}$$

where $\|\langle M,N\rangle\|_t$ denotes the total variation process of $\langle M,N\rangle_t$.

Proof. We may assume that the right hand side of (5.4) is always finite, otherwise there is nothing to prove.

Define

$$\varphi_t = \frac{1}{2}(\langle M\rangle_t + \langle N\rangle_t),\ t \geqslant 0.$$

From (5.3), we know that with probability one, the measures $d\|\langle M,N\rangle\|_t$, $d\langle M\rangle_t$ and $d\langle N\rangle_t$ are all absolutely continuous with respect to $d\varphi_t$. Therefore, we may write

$$\begin{aligned}\langle M,N\rangle_t(\omega) &= \int_0^t f_1(u,\omega)d\varphi_u(\omega),\\ \langle M\rangle_t(\omega) &= \int_0^t f_2(u,\omega)d\varphi_u(\omega),\\ \langle N\rangle_t(\omega) &= \int_0^t f_3(u,\omega)d\varphi_u(\omega),\end{aligned}$$

for some measurable functions $f_i(t,\omega)$ $(i=1,2,3)$.

Therefore, according to Proposition 5.4, for each pair (α,β) of rational numbers, there exists $\Omega_{\alpha,\beta}\in\mathcal{F}$ with $\mathbb{P}(\Omega_{\alpha,\beta})=1$, such that for every $\omega\in\Omega_{\alpha,\beta}$, we have:

$$\begin{aligned}0 &\leqslant \langle\alpha M+\beta N\rangle_t - \langle\alpha M+\beta N\rangle_s\\ &= \int_s^t \left(\alpha^2 f_2(u,\omega)+2\alpha\beta f_1(u,\omega)+\beta^2 f_3(u,\omega)\right)d\varphi_u(\omega),\quad \forall 0\leqslant s<t.\end{aligned}$$

This implies that there exists some $T_{\alpha,\beta}(\omega)\in\mathcal{B}([0,\infty))$ depending on ω and (α,β), such that $\int_{T_{\alpha,\beta}(\omega)} d\varphi_u(\omega)=0$ and

$$\alpha^2 f_2(t,\omega)+2\alpha\beta f_1(t,\omega)+\beta^2 f_3(t,\omega)\geqslant 0 \tag{5.5}$$

is true for all $t\notin T_{\alpha,\beta}(\omega)$.

Now take $\widetilde{\Omega}=\cap_{(\alpha,\beta)\in\mathbb{Q}^2}\Omega_{\alpha,\beta}$ and $\widetilde{T}(\omega)=\cup_{(\alpha,\beta)\in\mathbb{Q}^2}T_{\alpha,\beta}(\omega)$ for every $\omega\in\widetilde{\Omega}$. It follows that (5.5) is true for $\omega\in\widetilde{\Omega}$, $t\notin\widetilde{T}(\omega)$ and $(\alpha,\beta)\in\mathbb{Q}^2$ (thus for all $(\alpha,\beta)\in\mathbb{R}^2$). Fix such ω and t, replace α by $\alpha|X_t(\omega)|$ and β by $|Y_t(\omega)|\cdot\mathrm{sgn}(f_1(t,\omega))$ respectively, we obtain that

$$\alpha^2|X_t(\omega)|^2 f_2(t,\omega)+2\alpha|X_t(\omega)|\cdot|Y_t(\omega)|\cdot|f_1(t,\omega)|+|Y_t(\omega)|^2 f_3(t,\omega)\geqslant 0$$

for every $\omega\in\widetilde{\Omega}$, $t\in\widetilde{T}$ and $\alpha\in\mathbb{R}^1$.

Inequality (5.4) then follows from integrating against $d\varphi_t(\omega)$ and optimizing α. □

Now we illustrate the reason why $\langle M\rangle_t$ is called the quadratic variation process of M_t.

Proposition 5.7. *Let $M\in\mathcal{M}_0^{\mathrm{loc}}$. Given $t\geqslant 0$, let $\mathcal{P}_n$ be a sequence of finite partitions over $[0,t]$ such that $\mathrm{mesh}(\mathcal{P}_n)\to 0$. Then*

$$\sum_{t_i\in\mathcal{P}_n}(M_{t_i}-M_{t_{i-1}})^2\to\langle M\rangle_t \text{ in probability}$$

as $n\to\infty$.

Proof. To simplify the notation, for $t_i \in \mathcal{P}_n$, we write $\Delta^i M \triangleq M_{t_i} - M_{t_{i-1}}$ and $\Delta^i \langle M \rangle \triangleq \langle M \rangle_{t_i} - \langle M \rangle_{t_{i-1}}$.

We first assume that M and $\langle M \rangle$ are both uniformly bounded by some constant K. In this case, M_t and $M_t^2 - \langle M \rangle_t$ are both martinagles. Now we show that

$$\sum_i (\Delta^i M)^2 \to \langle M \rangle_t$$

in L^2 as $\mathrm{mesh}(\mathcal{P}_n) \to 0$. Indeed, we have:

$$\begin{aligned} \mathbb{E}\left[\left|\sum_i (\Delta^i M)^2 - \langle M \rangle_t\right|^2\right] &= \mathbb{E}\left[\left|\sum_i \left((\Delta^i M)^2 - \Delta^i \langle M \rangle\right)\right|^2\right] \\ &= \sum_i \mathbb{E}\left[\left((\Delta^i M)^2 - \Delta^i \langle M \rangle\right)^2\right] \\ &\leqslant 2\left(\sum_i \mathbb{E}\left[(\Delta^i M)^4\right] + \sum_i \mathbb{E}\left[(\Delta^i \langle M \rangle)^2\right]\right), \end{aligned}$$

where the second equality follows from the fact that

$$\mathbb{E}[((\Delta^i M)^2 - \Delta^i \langle M \rangle)((\Delta^j M)^2 - \Delta^j \langle M \rangle)] = 0$$

for $i \neq j$, which can be easily shown by conditioning.

On the one hand, since $\langle M \rangle$ is continuous, we have

$$\sum_i (\Delta^i \langle M \rangle)^2 \leqslant \langle M \rangle_t \cdot \max_i \Delta^i \langle M \rangle \leqslant K \cdot \max_i \Delta^i \langle M \rangle \to 0$$

as $\mathrm{mesh}(\mathcal{P}_n) \to 0$. According to the dominated convergence theorem, we see that

$$\sum_i \mathbb{E}[(\Delta^i \langle M \rangle)^2 \to 0$$

as $\mathrm{mesh}(\mathcal{P}_n) \to 0$.

On the other hand,

$$\sum_i (\Delta^i M)^4 \leqslant \left(\sum_i (\Delta^i M)^2\right) \cdot \max_i (\Delta^i M)^2, \tag{5.6}$$

and thus

$$\sum_i \mathbb{E}[(\Delta^i M)^4] \leqslant \left(\mathbb{E}\left[\left(\sum_i (\Delta^i M)^2\right)^2\right]\right)^{\frac{1}{2}} \cdot \left(\mathbb{E}\left[\max_i (\Delta^i M)^4\right]\right)^{\frac{1}{2}}. \tag{5.7}$$

We first show that $\mathbb{E}[(\sum_i (\Delta^i M)^2)^2]$ is uniformly bounded. Indeed,

$$\mathbb{E}\left[\left(\sum_i (\Delta^i M)^2\right)^2\right] = \sum_i \mathbb{E}[(\Delta^i M)^4] + 2\sum_i \sum_{j>i} \mathbb{E}[(\Delta^i M)^2 (\Delta^j M)^2]. \qquad (5.8)$$

Since $\mathbb{E}[\sum_i (\Delta^i M)^2] = \mathbb{E}[M_t^2] \leqslant K^2$, from (5.6) we can easily see that $\sum_i \mathbb{E}[(\Delta^i M)^4] \leqslant 4K^4$. Moreover, by conditioning we can also see that the second term of (5.8) equals

$$2\sum_i \mathbb{E}\left[(\Delta^i M)^2 (M_t^2 - M_{t_i}^2)\right] \leqslant 2K^2 \sum_i \mathbb{E}[(\Delta^i M)^2] \leqslant 2K^4.$$

Therefore, $\mathbb{E}[(\sum_i (\Delta^i M)^2)^2] \leqslant 6K^4$. Applying the dominated convergence theorem to (5.7), we obtain that

$$\sum_i \mathbb{E}[(\Delta^i M)^4] \to 0$$

as $\mathrm{mesh}(\mathcal{P}_n) \to 0.$

Therefore, we conclude that

$$\sum_i (\Delta^i M)^2 \to \langle M \rangle_t$$

in L^2 as $\mathrm{mesh}(\mathcal{P}_n) \to 0.$

Coming back to the local martingale situation, we again apply a localization argument. Let τ_m be a sequence of $\{\mathcal{F}_t\}$-stopping times increasing to ∞ such that $M_t^{\tau_m}$ is a bounded $\{\mathcal{F}_t\}$-martingale and $\langle M^{\tau_m} \rangle_t$ is bounded. Given $\delta > 0,$ there exists $m \geqslant 1$ such that $\mathbb{P}(\tau_m \leqslant t) < \delta$. For this particular m, we have

$$\begin{aligned} & \mathbb{P}\left(\left|\sum_i (\Delta^i M)^2 - \langle M \rangle_t\right| > \varepsilon\right) \\ \leqslant \ & \mathbb{P}(\tau_m \leqslant t) + \mathbb{P}\left(\left|\sum_i (\Delta^i M)^2 - \langle M \rangle_t\right| > \varepsilon,\ \tau_m > t\right) \\ \leqslant \ & \delta + \mathbb{P}\left(\left|\sum_i (\Delta^i M^{\tau_m})^2 - \langle M^{\tau_m} \rangle_t\right| > \varepsilon\right). \end{aligned}$$

Since L^2 convergence implies convergence in probability, by applying what we just proved in the bounded case, we obtain that

$$\limsup_{n\to\infty} \mathbb{P}\left(\left|\sum_i (\Delta^i M)^2 - \langle M \rangle_t\right| > \varepsilon\right) \leqslant \delta.$$

As δ is arbitrary, we get the desired convergence in probability. □

Combining the existence of quadratic variation in the sense of Prosposition 5.7 and the global positive definiteness of the quadradic variation process in the sense of Proposition 5.4, (3), we can further show the following local positive definiteness property.

Proposition 5.8. *Let $M \in \mathcal{M}_0^{\mathrm{loc}}$. Then there exists a $\mathbb{P}$-null set N, such that for every $\omega \in N^c$, we have*

$$M_t = M_a \ \ \forall t \in [a,b] \iff \langle M \rangle_a = \langle M \rangle_b,$$

for each $a < b$.

Proof. First of all, since convergence in probability implies almost sure convergence along a subsequence, according to Proposition 5.7, we can see that for each given pair of rational numbers $p < q$, there exists a $\mathbb{P}$-null set $N_{p,q}$, such that for every $\omega \notin N_{p,q}$, we have

$$M_t(\omega) = M_p(\omega) \ \forall t \in [p,q] \implies \langle M \rangle_p(\omega) = \langle M \rangle_q(\omega).$$

Take $N_1 \triangleq \cup_{p,q \in \mathbb{Q}, p<q} N_{p,q}$. Given any $\omega \notin N_1$ and $a < b$, if $M_t(\omega) = M_a(\omega)$ on $[a,b]$, then the same holds on any subinterval $[p,q] \subseteq [a,b]$ with $p,q \in \mathbb{Q}$. Therefore, $\langle M \rangle_p(\omega) = \langle M \rangle_q(\omega)$. By the continuity of $t \mapsto \langle M \rangle_t(\omega)$, we conclude that $\langle M \rangle_a(\omega) = \langle M \rangle_b(\omega)$. This is true for arbitrary $a < b$.

To see the other direction, first assume that M_t is a bounded $\{\mathcal{F}_t\}$-martinagale. For each $q \in \mathbb{Q}$, define $\widetilde{M}_t = M_{t+q} - M_q$ and $\mathcal{G}_t = \mathcal{F}_{t+q}$. Then $\left\{\widetilde{M}_t, \mathcal{G}_t\right\}$ is a martingale with quaratic variation process $\left\langle \widetilde{M} \right\rangle_t = \langle M \rangle_{t+q} - \langle M \rangle_q$. Let $\tau_q \triangleq \inf\left\{t \geqslant 0 : \left\langle \widetilde{M} \right\rangle_t > 0\right\}$. It follows that $\left\langle \widetilde{M}^{\tau_q} \right\rangle = \left\langle \widetilde{M} \right\rangle^{\tau_q} = 0$, and thus $\widetilde{M}^{\tau_q} = 0$ by Proposition 5.4, (3). In particular, for every ω outside some $\mathbb{P}$-null set N_q', we have $M_t(\omega) = M_q(\omega)$ for every $t \in [q, q + \tau_q(\omega)]$. Let $N_2 \triangleq \cup_{p \in \mathbb{Q}^+} N_q'$. Given $\omega \notin N_2$ and $a < b$, suppose that $\langle M \rangle_a(\omega) = \langle M \rangle_b(\omega)$. Then for any $q \in (a,b)$, $\langle M \rangle_q(\omega) = \langle M \rangle_b(\omega)$. This implies that $\tau_q(\omega) \geqslant b - q$. In particular, $M_t(\omega) = M_q(\omega)$ for every $t \in [q,b]$. This is true for every $q \in (a,b)$. By the continuity of $t \mapsto M_t(\omega)$, we conclude that $M_t(\omega) = M_a(\omega)$ on $[a,b]$.

Therefore, the result of the proposition is proved for the case of bounded martingales. For a general $M \in \mathcal{M}_0^{\mathrm{loc}}$, let τ_n be a sequence of $\{\mathcal{F}_t\}$-stopping times such that $\tau_n \uparrow \infty$ almost surely, and $M_t^{\tau_n}$ is a bounded $\{\mathcal{F}_t\}$-martingale for every n. Then there exists a $\mathbb{P}$-null set N_n for each n, such that outside N_n the result holds for the martingale $M_t^{\tau_n}$. By taking $N \triangleq \cup_{n=1}^{\infty} N_n$, we know that outside N the result holds for M_t. □

5.2 Stochastic integrals

The most natural way of defining the integral $\int \Phi_t dB_t$ for a stochastic process Φ_t and a Brownian motion B_t is to consider the Riemann sum approximation $\sum_i \Phi_{u_i}(B_{t_i} - B_{t_{i-1}})$ for a given partition $\mathcal{P}$, where $u_i \in [t_{i-1}, t_i]$. However, as $\mathrm{mesh}(\mathcal{P}) \to 0$, we can not expect that the Riemnann sum would converge in a pathwise sense due to the fact that sample paths of B_t have infinite 1-variation on every finite interval. If instead we look for convergence in some probabilistic sense, we have to be careful about the choice of u_i.

Suppose that Φ_t is uniformly bounded and $\{\mathcal{F}_t\}$-adapted. If we choose $u_i = t_{i-1}$ (the left endpoint), nice things will occur: for $m > n$,

$$\begin{aligned}
& \mathbb{E}\left[\sum_{i=1}^{m}\Phi_{t_{i-1}}(B_{t_i}-B_{t_{i-1}})|\mathcal{F}_{t_n}\right] \\
= & \sum_{i=1}^{n}\Phi_{t_{i-1}}(B_{t_i}-B_{t_{i-1}})+\sum_{i=n+1}^{m}\mathbb{E}\left[\mathbb{E}[\Phi_{t_{i-1}}(B_{t_i}-B_{t_{i-1}})|\mathcal{F}_{t_{i-1}}]|\mathcal{F}_{t_n}\right] \\
= & \sum_{i=1}^{n}\Phi_{t_{i-1}}(B_{t_i}-B_{t_{i-1}}).
\end{aligned}$$

This suggests that we might look for a construction under which $\int \Phi_t dB_t$ is a martingale. Another observation is that

$$\mathbb{E}\left[\left(\sum_{i=1}^{n}\Phi_{t_{i-1}}(B_{t_i}-B_{t_{i-1}})\right)^2\right]=\mathbb{E}\left[\sum_{i=1}^{n}\Phi_{t_{i-1}}^2(t_i-t_{i-1})\right].$$

This suggests that if we define a norm on Φ by $\|\Phi\|_B = \left(\mathbb{E}\left[\int \Phi_t^2 dt\right]\right)^{1/2}$, then the integration map $\Phi \mapsto \int \Phi_t dB_t$ should be an isometry into L^2. Therefore, it sheds light on constructing stochastic integrals through a functional analytic approach (more precisely, a Hilbert space approach).

A technical point is to identify suitable functional spaces on which the integration map is to be built. To make sure $\int \Phi_t dB_t$ will again be $\{\mathcal{F}_t\}$-adapted, a natural measurability condition on Φ_t is progressive measurability (c.f. Definition 2.7).

It is remarkable that Itô already had this deep insight in his original construction of stochastic integrals before Doob's martingale theory was available. The more intrinsic approach within the martingale framework that we are going to present here is due to Kunita-Watanabe.

Suppose that $M \in H_0^2$ is an L^2-bounded continuous martingale vanishing at $t=0$ (c.f. Definition 5.1).

Define $\mathcal{L}^2(M)$ to be the space of progressively measurable processes Φ_t such that

$$\|\Phi\|_M \triangleq \left(\mathbb{E}\left[\int_0^\infty \Phi_t^2 d\langle M\rangle_t\right]\right)^{\frac{1}{2}} < \infty. \tag{5.9}$$

If we define a measure $\mathbb{P}_M$ on $([0,\infty)\times\Omega, \mathcal{B}([0,\infty))\otimes\mathcal{F})$ by

$$\mathbb{P}_M(\Lambda) \triangleq \mathbb{E}\left[\int_0^\infty \mathbf{1}_\Lambda(t,\omega)d\langle M\rangle_t(\omega)\right],\quad \Lambda\in\mathcal{B}([0,\infty))\otimes\mathcal{F},$$

then $\mathcal{L}^2(M)$ is just the space of $\mathbb{P}_M$-square integrable, progressively measurable processes. Note that $\mathbb{P}_M$ is a finite measure since $M \in H_0^2$. Define $L^2(M)$ to be the space of $\mathbb{P}_M$-equivalence classes of elements in $\mathcal{L}^2(M)$.

Remark 5.2. We will adopt the convention of not being very careful in distinguishing between a process and its equivalence class. It will be clear that if Φ, Ψ are equivalent, then $\int \Phi_t dM_t, \int \Psi_t dM_t$ are indistinguishable.

Lemma 5.2. *$(L^2(M), \|\cdot\|_M)$ is a Hilbert space.*

Proof. The only thing which is not immediately clear is the progressive measurability for a limit process. Let $\Phi^{(n)}$ be a sequence in $\mathcal{L}^2(M)$ converging to some measurable process Φ under $\|\cdot\|_M$. Along a subsequence $\Phi^{(n_k)}$ we know that the set $\{(t,\omega) : \lim_{k\to\infty} \Phi_t^{(n_k)}(\omega) \neq \Phi_t(\omega)\}$ is a $\mathbb{P}_M$-null set. In general, Φ_t might not be progressively measurable. But the process $\mathbf{1}_A$, where

$$A \triangleq \{(t,\omega) : \lim_{k\to\infty} \Phi_t^{(n_k)}(\omega) \text{ exists finitely}\},$$

is easily seen to be progressively measurale. Moreover, the process $\Psi \triangleq \limsup_{k\to\infty} \Phi^{(n_k)} \cdot \mathbf{1}_A$ is $\mathbb{P}_M$-equivalent to Φ. Since $\Phi^{(n_k)}$ is progressively measurable for each k, we conclude that Ψ is progressively measurable. □

The construction of stochastic integrals with respect to M is contained in the following result.

Theorem 5.3. *For each $\Phi \in L^2(M)$, there exists a unique $I^M(\Phi) \in H_0^2$ (up to indistinguishability), such that for any $N \in H_0^2$,*

$$\langle I^M(\Phi), N\rangle = \Phi \bullet \langle M, N\rangle, \tag{5.10}$$

where $\Phi \bullet \langle M, N\rangle$ denotes the integral process $\int_0^t \Phi_s d\langle M, N\rangle_s$, defined pathwisely. Moreover, the map $I^M : \Phi \mapsto I^M(\Phi)$ defines a linear isometry from $L^2(M)$ into H_0^2.

Proof. We first prove uniqueness. Suppose that $X, Y \in H_0^2$ both satisfy (5.10). It follows that

$$\langle Y - X, N\rangle = 0, \quad \forall N \in H_0^2.$$

In particular, by taking $N = Y - X$, we know that $\langle Y - X\rangle = 0$. Therefore, $X = Y$.

Now we show existence. Given $\Phi \in L^2(M)$, define a linear functional F^Φ on H_0^2 by

$$F^\Phi(N) \triangleq \mathbb{E}\left[\int_0^\infty \Phi_t d\langle M, N\rangle_t\right], \quad N \in H_0^2.$$

According to the Kunita-Watanabe inequality (c.f. (5.4)), we have

$$\left|\int_0^\infty \Phi_t d\langle M, N\rangle_t\right| \leqslant \left(\int_0^\infty \Phi_t^2 d\langle M\rangle_t\right)^{\frac{1}{2}} \cdot \langle N\rangle_\infty^{\frac{1}{2}}.$$

Therefore, by the Cauchy-Schwarz inequality, we have

$$\left|F^\Phi(N)\right| \leqslant \left(\mathbb{E}\left[\int_0^\infty \Phi_t^2 d\langle M\rangle_t\right]\right)^{\frac{1}{2}} \cdot \mathbb{E}[\langle N\rangle_\infty]^{\frac{1}{2}} = \|\Phi\|_M \cdot \|N\|_{\mathbb{H}^2}.$$

In particular, F^{Φ} defines a bounded linear functional on H_0^2. It follows from the Riesz representation theorem that there exists $X \in H_0^2$, such that

$$F^{\Phi}(N) = \langle X, N \rangle_{\mathbb{H}^2} = \mathbb{E}[X_\infty N_\infty], \quad \forall N \in H_0^2. \tag{5.11}$$

To establish (5.10) for X, suppose that τ is an arbitrary $\{\mathcal{F}_t\}$-stopping time. Then for any $N \in H_0^2$, we have

$$\mathbb{E}[X_\tau N_\tau] = \mathbb{E}[\mathbb{E}[X_\infty | \mathcal{F}_\tau] N_\tau] = \mathbb{E}[X_\infty N_\tau].$$

Note that $N^\tau \in H_0^2$ and $N_\tau = N_\infty^\tau$. Therefore, according to (5.11) and Proposition 5.3, we arrive at

$$\begin{aligned}
\mathbb{E}[X_\tau N_\tau] &= \mathbb{E}[X_\infty N_\infty^\tau] = F^{\Phi}(N^\tau) = \mathbb{E}\left[\int_0^\infty \Phi_t d\langle M, N^\tau \rangle_t\right] \\
&= \mathbb{E}\left[\int_0^\infty \Phi_t d\langle M, N \rangle_t^\tau\right] = \mathbb{E}\left[\int_0^\tau \Phi_t d\langle M, N \rangle_t\right].
\end{aligned}$$

By Problem Sheet 3, Problem 1, (2), we conclude that $XN - \Phi \bullet \langle M, N \rangle$ is a martingale, which implies (5.10).

It is apparent that the map $I^M: \Phi \mapsto X = I^M(\Phi)$ is linear. Moreover, from (5.11) and (5.10), we have:

$$\|X\|_{\mathbb{H}^2}^2 = \mathbb{E}[X_\infty^2] = \mathbb{E}\left[\int_0^\infty \Phi_t d\langle M, X \rangle_t\right] = \mathbb{E}\left[\int_0^\infty \Phi_t^2 d\langle M \rangle_t\right] = \|\Phi\|_M,$$

Therefore, I^M is a linear isometry from $L^2(M)$ into H_0^2. □

Definition 5.5. For $\Phi \in L^2(M)$, $I^M(\Phi)$ is called the *stochastic integral* of Φ with respect to M. As a stochastic process, $I^M(\Phi)_t$ is also denoted as $\int_0^t \Phi_s dM_s$. The map $I^M: L^2(M) \to H_0^2$ is called the *stochastic integration map*.

The reason why I^M is called the stochastic integration map is the following. Let Φ_t be a stochastic process of the form

$$\Phi = \Phi_0 \mathbf{1}_{\{0\}} + \sum_{i=1}^{\infty} \Phi_{t_{i-1}} \mathbf{1}_{(t_{i-1}, t_i]},$$

where $0 = t_0 < t_1 < \cdots < t_n < \cdots$ is a partition of $[0, \infty)$, Φ_{t_n} is $\{\mathcal{F}_{t_n}\}$-measurable for each n, and they are uniformly bounded by some constant $C > 0$. Then $\Phi \in L^2(M)$ and

$$\int_0^t \Phi_s dM_s = \sum_{i=1}^{n-1} \Phi_{t_{i-1}}(M_{t_i} - M_{t_{i-1}}) + \Phi_{t_{n-1}}(M_t - M_{t_{i-1}}), \quad t \in [t_{n-1}, t_n]. \tag{5.12}$$

The proof follows easily by computing the bracket with $N \in H_0^2$ of the right hand side of (5.12), which is left as an exercise.

Now we present some basic properties of stochastic integrals.

Proposition 5.9. *Let $M, N \in H_0^2$ and let $\Phi \in L^2(M), \Psi \in L^2(N)$ respectively. Suppose that $\sigma \leqslant \tau$ are two $\{\mathcal{F}_t\}$-stopping times. Then we have:*

(1)

$$\mathbb{E}\left[I^M(\Phi)_{t\wedge\tau} - I^M(\Psi)_{t\wedge\sigma}|\mathcal{F}_\sigma\right] = 0.$$

(2)

$$\begin{aligned}&\mathbb{E}\left[(I^M(\Phi)_{t\wedge\tau} - I^M(\Phi)_{t\wedge\sigma})(I^N(\Psi)_{t\wedge\tau} - I^N(\Psi)_{t\wedge\sigma})|\mathcal{F}_\sigma\right] \\ = \quad &\mathbb{E}\left[\int_{t\wedge\sigma}^{t\wedge\tau} \Phi_s\Psi_s d\langle M, N\rangle_s|\mathcal{F}_\sigma\right].\end{aligned}$$

Proof. The result follows easily from applying the optional sampling theorem to the underlying martinagles stopped at t. Note that

$$\left\langle I^M(\Phi), I^N(\Psi)\right\rangle = \Phi \bullet \left\langle M, I^N(\Psi)\right\rangle = (\Phi\Psi) \bullet \langle M, N\rangle.$$

□

Remark 5.3. $\sigma = s < \tau = t$ or $M = N$ are important special cases of Proposition 5.9. In particular, $\langle I^M(\Phi)\rangle = \Phi^2 \bullet \langle M\rangle$.

The next property is associativity.

Proposition 5.10. *Let $M \in H_0^2$. Suppose that $\Phi \in L^2(M)$ and $\Psi \in L^2(I^M(\Phi))$. Then $\Psi \cdot \Phi \in L^2(M)$ and $I^M(\Psi\Phi) = I^{I^M(\Phi)}(\Psi)$.*

Proof. Since $\left\langle I^M(\Phi)\right\rangle_t = \int_0^t \Phi_s^2 d\langle M\rangle_s$ and $\Psi \in L^2(I^M(\Phi))$, we see that $\Psi\cdot\Phi \in L^2(M)$. Moreover, for every $N \in H_0^2$, we have

$$\begin{aligned}\left\langle I^M(\Psi\Phi), N\right\rangle &= (\Psi\Phi) \bullet \langle M, N\rangle = \Psi \bullet (\Phi \bullet \langle M, N\rangle) \\ &= \Psi \bullet \left(\langle I^M(\Phi), N\rangle\right) = \left\langle I^{I^M(\Phi)}(\Psi), N\right\rangle.\end{aligned}$$

Therefore, $I^M(\Psi\Phi) = I^{I^M(\Phi)}(\Psi)$. □

The associativity enables us to show compatibility with stopping easily.

Proposition 5.11. *Let τ be an $\{\mathcal{F}_t\}$-stopping time. Then for $M \in H_0^2$ and $\Phi \in L^2(M)$, we have*

$$I^{M^\tau}(\Phi) = I^M(\Phi\mathbf{1}_{[0,\tau]}) = I^{M^\tau}(\Phi^\tau) = I^M(\Phi)^\tau.$$

Proof. Firstly, observe that $\mathbf{1}_{[0,\tau]} \in L^2(M)$ and $I^M(\mathbf{1}_{[0,\tau]}) = M^\tau$ (note that the process $(t,\omega) \mapsto \mathbf{1}_{[0,\tau(\omega)]}(t)$ is progressively measurable). Moreover, it is apparent that $\Phi, \Phi^\tau \in L^2(M^\tau)$. Therefore, the first equality follows from Proposition 5.10, and the other inequalities follow from taking bracket with $N \in H_0^2$. □

So far our stochastic integration does not even cover the case of Brownian motion, as the Brownian motion is not bounded in L^2. The way to enlarging our scope of stochastic integration is localization.

Definition 5.6. Let $M \in \mathcal{M}_0^{\text{loc}}$ be a continuous local martingale vanishing at $t = 0$. We use $L^2_{\text{loc}}(M)$ to denote the space of progressively measurable processes X_t, such that with probability one,

$$\int_0^t \Phi_s^2 d\langle M \rangle_s < \infty, \quad \forall t \geqslant 0.$$

We aim at defining the stochastic integral $I^M(\Phi) \in \mathcal{M}_0^{\text{loc}}$ for $\Phi \in L^2_{\text{loc}}(M)$. This is contained in the following theorem.

Theorem 5.4. *Let $M \in \mathcal{M}_0^{\text{loc}}$ and let $\Phi \in L^2_{\text{loc}}(M)$. Then there exists a unique $I^M(\Phi) \in \mathcal{M}_0^{\text{loc}}$, such that for any $N \in \mathcal{M}_0^{\text{loc}}$, we have*

$$\langle I^M(\Phi), N \rangle = \Phi \bullet \langle M, N \rangle, \tag{5.13}$$

where the integral process $\Phi \bullet \langle M, N \rangle$ is finitely almost surely according to the Kunita-Watanabe inequality (c.f. (5.4)).

Proof. Uniqueness is obvious.

Now we show existence. For $n \geqslant 1$, define

$$\tau_n = \inf \left\{ t \geqslant 0 : \ |M_t| > n \text{ or } \int_0^t \Phi_s^2 d\langle M \rangle_s > n \right\}.$$

Then τ_n is a sequence of $\{\mathcal{F}_t\}$-stopping times such that $\tau_n \uparrow \infty$ almost surely. Moreover, for each n, we have $M^{\tau_n} \in H_0^2$ and $\Phi^{\tau_n} \in L^2(M^{\tau_n})$. Therefore, $X^{(n)} \triangleq I^{M^{\tau_n}}(\Phi^{\tau_n}) \in H_0^2$ is well-defined. According to Proposition 5.11, we know that

$$\left(X^{(n+1)} \right)^{\tau_n} = I^{M^{\tau_n}}(\Phi^{\tau_n}) = X^{(n)}. \tag{5.14}$$

This implies that we can define a process X_t on $[0, \infty)$ such that $X_t = X_t^{(n)}$ on $[0, \tau_n]$. It is apparent that X_t is continuous and $\{\mathcal{F}_t\}$-adapted. From (5.14) we also know that $X_t^{(n)} = \text{const.}$ for $t > \tau_n$. Therefore, $X^{\tau_n} = X^{(n)}$. This implies that $X \in \mathcal{M}_0^{\text{loc}}$. Finally, to see (5.13), let $N \in H_0^2$ (the general case where $N \in \mathcal{M}_0^{\text{loc}}$ follows easily by further localizing N to be bounded). Then

$$\langle X, N \rangle_t^{\tau_n} = \langle X^{(n)}, N \rangle_t = \int_0^t \Phi_s^{\tau_n} d\langle M^{\tau_n}, N \rangle_s = \int_0^{\tau_n \wedge t} \Phi_s d\langle M, N \rangle_s$$

for every n. (5.13) follows from letting $n \to \infty$. □

Remark 5.4. For $M \in \mathcal{M}_0^{\text{loc}}$, we can define the space $L^2(M) \subseteq L^2_{\text{loc}}(M)$ in the same way as (5.9). Exactly the same proof of Theorem 5.3 allows us to conclude that for each $\Phi \in L^2(M)$, there exists a unique $X \in H_0^2$ satisfying the characterizing property (5.10). The map $\Phi \mapsto X$ is a linear isometry from $L^2(M)$ into H_0^2. This part has nothing to do with the martingale property of M. Of course X coincides with $I^M(\Phi)$ which is defined in Theorem 5.4 in the sense of local martingales.

Although the stochastic integral $I^M(\Phi)$ is constructed from a global point of view, we also have the following local property.

Proposition 5.12. *Let $M \in \mathcal{M}_0^{\mathrm{loc}}$ and let $\Phi \in L^2_{\mathrm{loc}}(M)$. Then there exists a $\mathbb{P}$-null set N, such that for every $\omega \in N^c$,*

$$\Phi_t \equiv 0 \text{ or } M_t \equiv M_a \text{ on } [a,b] \implies I^M(\Phi)_t \equiv I^M(\Phi)_a \text{ on } [a,b]$$

for each $a < b$.

Proof. The result is a direct consequence of Proposition 5.8 and the fact that $\langle I^M(\Phi)\rangle = \Phi^2 \bullet \langle M \rangle$. □

As we will see in the next subsection, if M is a local martingale and f is a nice function, $f(M)$ is in general not a local martingale, but it is a local martingale (a stochastic integral) plus a process with bounded variation. Moreover, in the study of stochastic differential equations, we also consider systems having such a general decomposition, namely $dX_t = \mu(X_t)dt + \sigma(X_t)dB_t$. Therefore, it is necessary to further extend our scope of integration.

Definition 5.7. A continuous, $\{\mathcal{F}_t\}$-adapted process X_t is called a *continuous semimartingale* if it has the decomposition

$$X_t = X_0 + M_t + A_t, \tag{5.15}$$

where $M \in \mathcal{M}_0^{\mathrm{loc}}$ is a continuous local martingale vanishing at $t = 0$, and A_t is a continuous, $\{\mathcal{F}_t\}$-adapted process such that with probability one, $A_0(\omega) = 0$ and $t \mapsto A_t(\omega)$ has bounded variation on every finite interval.

Given two continuous semimartingales $X_t = X_0 + M_t + A_t$ and $Y_t = Y_0 + N_t + B_t$, the *bracket process* of X and Y is $\langle X, Y\rangle_t \triangleq \langle M, N\rangle_t$, and the *qradratic variation process* of X is $\langle X\rangle_t \triangleq \langle M \rangle_t$.

It is apparent that the decomposition (5.15) for a continuous semimartingale is unique. Moreover, the quadratic variation process also satisfies Proposition 5.7.

When we talk about stochastic integrals with respect to continuous semimartingales, it is convenient to have a universal class of integrands which is independent of the underlying semimartingales.

Definition 5.8. A progressively measurable process Φ_t is called *locally bounded* if there exists a sequence τ_n of $\{\mathcal{F}_t\}$-stopping times increasing to infinity and positive constants C_n, such that

$$|\Phi_t^{\tau_n}| \leqslant C_n, \quad \forall t \geqslant 0,$$

for every $n \geqslant 1$.

It is apparent that every continuous, $\{\mathcal{F}_t\}$-adapted process Φ_t with bounded Φ_0 is locally bounded. Indeed, we can simply define $\tau_n = \inf\{t \geqslant 0 : \ |\Phi_t| > n\}$. Moreover, if Φ_t is locally bounded, then for every $M \in \mathcal{M}_0^{\mathrm{loc}}$, we have $\Phi \in L^2_{\mathrm{loc}}(M)$.

Definition 5.9. Let $X_t = X_0 + M_t + A_t$ be a continuous semimartingale and let Φ_t be a locally bounded process. The *stochastic integral* of Φ_t with respect to X_t is defined to be the continuous semimartingale

$$I^X(\Phi)_t = I^M(\Phi)_t + I^A(\Phi)_t, \quad t \geqslant 0,$$

where the second term $I^A(\Phi)_t \triangleq \int_0^t \Phi_s dA_s$ is understood in the Lebesgue sense. The stochastic integral $I^X(\Phi)_t$ is also denoted as $\int_0^t \Phi_s dX_s$.

When X_t is a Brownian motion, the stochastic integral is usually known as *Itô's integral*.

It is important to point out that $I^M(\Phi)$ can fail to be a martingale if $M \in \mathcal{M}_0^{\mathrm{loc}}$, so the integrability properties in Proposition 5.9 may not hold in general. However, we still have the following properties. The proof is similar to the non-local case and is hence omitted.

Proposition 5.13. *(1)* $I^X(\Phi)$ *is linear in* X *and in* Φ.
(2) $I^X(\Psi\Phi) = I^{I^X(\Phi)}(\Psi)$ *for any locally bounded* Φ, Ψ.
(3) $I^{X^\tau}(\Phi) = I^X(\Phi \mathbf{1}_{[0,\tau]}) = I^{X^\tau}(\Phi^\tau) = I^X(\Phi)^\tau$ *for any* $\{\mathcal{F}_t\}$*-stopping time* τ.

Remark 5.5. In the definition of stochastic integrals and in Proposition 5.13, assuming local boundedness is just for technical convenience. Everything works well as long as we assume that all the Itô integals and Lebesgue integrals are well defined in their right sense respectively.

To conclude this subsetion, we prove a very useful tool which acts as the stochastic counterpart of the dominated convergence theorem.

Proposition 5.14. *Let* X_t *be a continuous semimartingale. Suppose that* Φ_t^n *is a sequence of locally bounded processes converges to zero pointwisely, and there exists a locally bounded process* Φ *such that* $|\Phi^n| \leqslant \Phi$. *Then* $I^X(\Phi^n)_t$ *converges to zero in probability uniformly on every finite interval, i.e. for every* $T > 0$,

$$\sup_{t \in [0,T]} \left| \int_0^t \Phi_s^n dX_s \right| \to 0 \text{ in probability}$$

as $n \to \infty$.

Proof. We only consider the situation where $X \in \mathcal{M}_0^{\mathrm{loc}}$ as the other case is easier. Let τ_m be a sequence of $\{\mathcal{F}_t\}$-stopping times increasing to infinity, such that for each m, $X^{\tau_m} \in H_0^2$ and $\Phi^{\tau_m}, X^{\tau_m}$, $\langle X \rangle^{\tau_m}$ are all bounded. Given $\varepsilon, \delta > 0$, choose m such that

$\mathbb{P}(\tau_m \leqslant T) < \delta$. It follows that

$$\begin{aligned}
& \mathbb{P}\left(\sup_{t\in[0,T]}\left|\int_0^t \Phi_s^n dX_s\right| > \varepsilon\right) \\
\leqslant\ & \mathbb{P}(\tau_m \leqslant t) + \mathbb{P}\left(\sup_{t\in[0,T]}\left|\int_0^t \Phi_s^n dX_s\right| > \varepsilon,\ \tau_m > T\right) \\
=\ & \mathbb{P}(\tau_m \leqslant t) + \mathbb{P}\left(\sup_{t\in[0,T]}\left|\int_0^{\tau_m\wedge t} \Phi_s^n dX_s\right| > \varepsilon,\ \tau_m > T\right) \\
=\ & \mathbb{P}(\tau_m \leqslant t) + \mathbb{P}\left(\sup_{t\in[0,T]}\left|\int_0^t (\Phi^n)_s^{\tau_m}\, dX_s^{\tau_m}\right| > \varepsilon,\ \tau_m > T\right) \\
\leqslant\ & \delta + \mathbb{P}\left(\sup_{t\in[0,T]}\left|\int_0^t (\Phi^n)_s^{\tau_m}\, dX_s^{\tau_m}\right| > \varepsilon\right).
\end{aligned}$$

Since $(\Phi^n)^{\tau_m} \in L^2(X^{\tau_m})$ for each n, we know that $I^{X^{\tau_m}}((\Phi^n)^{\tau_m}) \in H_0^2$. According to Doob's L^p-inequality, we have

$$\begin{aligned}
\mathbb{P}\left(\sup_{t\in[0,T]}\left|\int_0^t (\Phi^n)_s^{\tau_m}\, dX_s^{\tau_m}\right| > \varepsilon\right) &\leqslant \frac{1}{\varepsilon^2}\mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t (\Phi^n)_s^{\tau_m}\, dX_s^{\tau_m}\right|^2\right] \\
&\leqslant \frac{4}{\varepsilon^2}\mathbb{E}\left[\left(\int_0^T (\Phi^n)_s^{\tau_m}\, dX_s^{\tau_m}\right)^2\right] \\
&= \frac{4}{\varepsilon^2}\mathbb{E}\left[\int_0^T |(\Phi^n)_s^{\tau_m}|^2\, d\langle X^{\tau_m}\rangle_s\right],
\end{aligned}$$

which converges to zero as $n \to \infty$ by the dominated convergence theorem. Therefore,

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{t\in[0,T]}\left|\int_0^t \Phi_s^n dX_s\right| > \varepsilon\right) \leqslant \delta,$$

which implies the desired convergence as δ is arbitrary. □

A direct consequence of Proposition 5.14 is the following intuitive interpretation of stochastic integrals.

Corollary 5.1. *Let Φ_t be a left continuous and locally bounded process, and let X_t be a continuous semimartingale. For given $t > 0$, let $\mathcal{P}_n$ be a sequence of finite partitions over $[0, t]$ such that $\mathrm{mesh}(\mathcal{P}_n) \to 0$. Then*

$$\lim_{n\to\infty} \sum_{t_i\in\mathcal{P}_n} \Phi_{t_{i-1}}(X_{t_i} - X_{t_{i-1}}) = \int_0^t \Phi_s dX_s \ \text{ in probability.} \tag{5.16}$$

Proof. We only consider the case where $X \in \mathcal{M}_0^{\text{loc}}$. Suppose that $X \in H_0^2$ and Φ is bounded. Define

$$\Phi_s^n = \Phi_0 \mathbf{1}_{\{0\}}(s) + \sum_{t_i \in \mathcal{P}_n} \Phi_{t_{i-1}} \mathbf{1}_{(t_{i-1}, t_i]}(s) + \Phi_t \mathbf{1}_{(t,\infty)}(s).$$

Then Φ^n is uniformly bounded and $\Phi^n \to \Phi$ pointwisely on $[0,t] \times \Omega$. Note that (c.f. (5.12))

$$\int_0^t \Phi_s^n dX_s = \sum_{t_i \in \mathcal{P}_n} \Phi_{t_{i-1}} (X_{t_i} - X_{t_{i-1}}).$$

According to Proposition 5.14, we know that

$$\sum_{t_i \in \mathcal{P}_n} \Phi_{t_{i-1}} (X_{t_i} - X_{t_{i-1}}) \to \int_0^t \Phi_s dX_s$$

in probability as $n \to \infty$. The general case follows from the same localization argument as in the proof of Proposition 5.14. ☐

Remark 5.6. Taking left endpints in the Riemann sum approximation is an important feature of stochastic integrals. Indeed, (5.16) does not hold any more if we are not taking left endpoints.

5.3 Itô's formula

In classical analysis, if x_t is a smooth path, we have the differentiation rule $df(x_t) = f'(x_t)dx_t$, or equivalently, $f(x_t) - f(x_0) = \int_0^t f'(x_s)dx_s$. In the probabilistic setting, a natural question is: what happens if we replace x_t by a Brownian motion B_t? The answer is surprisingly different from the classical situation: we have $f(B_t) - f(B_0) = \int_0^t f'(B_s)dB_s + \frac{1}{2}\int_0^t f''(B_s)ds$. This is the renowned Itô's formula. We can see why it takes this form in the following naive way. Take the Taylor approximation up to degree 2 (it is reasonable to expect that all higher degrees are negligible): $df(B_t) = f'(B_t)dB_t + (1/2)f''(B_t)(dB_t)^2$. Here comes the key point: we have $(dB_t)^2 = dt \neq 0$. This is not entirely obvious at the moment, and it is crucially related to the martingale nature of B_t and the existence of its quadratic variation process. Therefore, Itô's formula follows naively.

Now we develop the mathematics.

We first consider the case when $f(x) = x^2$. This is also known as the *integration by parts* formula.

Proposition 5.15. *Suppose that X_t, Y_t are two continuous semimartingales. Then*

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

In particular,

$$X_t^2 = X_0^2 + 2\int_0^t X_s dX_s + \langle X \rangle_t. \tag{5.17}$$

Proof. It suffices to prove (5.17). The general case follows immediately from considering $(X_t+Y_t)^2$, $(X_t-Y_t)^2$ and linearity. Indeed, for any given finite partition $\mathcal{P}$ of $[0,t]$, we have

$$X_t^2 - X_0^2 = 2\sum_{t_i\in\mathcal{P}} X_{t_{i-1}}(X_{t_i}-X_{t_{i-1}}) + \sum_{t_i\in\mathcal{P}}(X_{t_i}-X_{t_{i-1}})^2.$$

According to Corollary 5.1 and the semimartingale version of Proposition 5.7, the result follows from taking limit in probability as $\mathrm{mesh}(\mathcal{P})\to 0$. □

If we take $X = M \in \mathcal{M}_0^{\mathrm{loc}}$, (5.17) tells us that

$$M_t^2 - \langle M\rangle_t = 2\int_0^t M_s dM_s.$$

We have already seen in Subsection 5.1 that $M^2 - \langle M\rangle \in \mathcal{M}_0^{\mathrm{loc}}$. Therefore, (5.17) gives us an explicit formula for this local martingale.

The general Itô's formula is stated as follows.

Theorem 5.5. *Let $X_t = (X_t^1,\cdots,X_t^d)$ be a vector of d continuous semimartingales. Suppose that $F\in C^2(\mathbb{R}^d)$ (continuously differentiable up to degree 2). Then $F(X_t)$ is a continuous semimartingale given by*

$$F(X_t) = F(X_0) + \sum_{i=1}^d \int_0^t \frac{\partial F}{\partial x^i}(X_s)dX_s^i + \frac{1}{2}\sum_{i,j=1}^d \int_0^t \frac{\partial^2 F}{\partial x^i\partial x^j}(X_s)d\langle X^i, X^j\rangle_s. \quad (5.18)$$

Proof. Suppose that $F\in C^2(\mathbb{R}^d)$ satisfies Itô's formula (5.18). Let $G(x) = x^i F(x)$ for some $1\leqslant i\leqslant d$. According to the integration by parts formula (c.f. Proposition 5.15), it is easy to see that $G(X_t)$ also satisfies Itô's formula. Therefore, Itô's formula holds for all polynomials.

For a general $F\in C^2(\mathbb{R}^d)$, we first assume that $|X_t|\leqslant K$ uniformly for some $K>0$. Let $G\in C^2(\mathbb{R}^d)$ be such that $G=F$ for $|x|\leqslant K$ and $G=0$ for $|x|>2K$. We only need to verify Itô's formula for G in this case. From classical analysis, we know that there exists a sequence p_n of polynomials on $\mathbb{R}^d$, such that for $|\alpha|\leqslant 2$,

$$\sup_{|x|\leqslant 2K} |D^\alpha p_n(x) - D^\alpha G(x)| \to 0$$

as $n\to\infty$, where "D^α" means the α-th derivative. Since Itô's formula holds for each p_n, according to the stochastic dominated convergence theorem (c.f. Proposition 5.14), we conclude that Itô's formula holds for G as well.

For a general X_t, we need to apply a localization argument. For each $n\geqslant 1$, define

$$\tau_n = \begin{cases} 0, & |X_0|\geqslant n; \\ \inf\{t\geqslant 0:\ |X_t|>n\}, & |X_0|<n, \end{cases}$$

and set

$$X_t^{(n)} = X_0\mathbf{1}_{\{|X_0|<n\}} + M_t^{\tau_n} + A_t^{\tau_n},$$

where M_t, A_t are the (vector-valued) martingale and bounded variation parts of X_t respectively. Then $X_t^{(n)}$ is a uniformly bounded continuous semimartingale. By the previous discussion, Itô's formula holds for $X_t^{(n)}$. On the other hand, since the stopped process $X^{\tau_n} = X^{(n)}$ on $\{\tau_n > 0\}$, by Proposition 5.12 we conclude that Itô's formula holds for X^{τ_n} on $\{\tau_n > 0\}$. Since $\cup_{n=1}^{\infty}\{\tau_n > 0\} = \Omega$, by letting $n \to \infty$, we conclude that Itô's formula holds for X_t and F. □

Remark 5.7. The same result holds if $F \in C^2(U)$ for some open subset $U \subseteq \mathbb{R}^d$ and with probability one, the process X_t takes values in U. The proof is identical but we need to use compact subsets to approximate U and localize on each of these compact subsets.

Formally, we usually write Itô's formula in the following differential form although it should always be understood in the integral sense:

$$dF(X_t) = \sum_{i=1}^{d} \frac{\partial F}{\partial x^i}(X_t)dX_t^i + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2 F}{\partial x^i \partial x^j}(X_t)d\langle X^i, X^j\rangle_t.$$

Now we present an important class of examples for Itô's formula.

Proposition 5.16. *Suppose that $f(x,y) \in C^2(\mathbb{R} \times \mathbb{R})$ is a complex-valued function which satisfies*

$$\frac{\partial f}{\partial y} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2} = 0.$$

Then for every $M \in \mathcal{M}_0^{\mathrm{loc}}$, $f(M_t, \langle M\rangle_t)$ is a continuous local martingale.

Proof. This is a straight forward application of Itô's formua to the vector semimartingale $(M_t, \langle M\rangle_t)$ and the function f. □

A particular f satisfying Proposition 5.16 is the exponential function

$$f^\lambda(x,y) = \mathrm{e}^{\lambda x - \frac{1}{2}\lambda^2 y}$$

for given $\lambda \in \mathbb{C}$. The resulting continuous local martingale

$$\mathcal{E}^\lambda(M)_t \triangleq \mathrm{e}^{\lambda M_t - \frac{1}{2}\lambda^2\langle M\rangle_t} = 1 + \lambda\int_0^t \mathcal{E}^\lambda(M)_s dM_s$$

is known as the *exponential martingale*. This (local) martingale is very important in the study of change of measure. In the case when $M_t = B_t$ (Brownian motion), from the distribution of B_t we can see directly that $\mathcal{E}^\lambda(B)_t$ is a martingale, a fact which was already used to prove the strong Markov property of Brownian motion and to compute passage time distributions.

5.4 The Burkholder-Davis-Gundy Inequalities

It is absolutely not unreasonable to say that Itô's formula is the most fundamental result in stochastic calculus. Starting from here we will begin a long journey of applying Itô's formula to a very rich class of interesting topics.

To appreciate the profoundness of Itô's formula, in this subsection we are going to (solely) use it in a pretty non-trivial way to obtain a fundamental type of martingale inequalities. These inequalities were first proved by Burkholder, Davis and Gundy and we usually refer them as the BDG inequalities. They play a fundamental role in the connection with harmonic analysis.

Let M_t be a continuous and square integrable martingale. Define $M_t^* \triangleq \sup_{0\leqslant s\leqslant t}|M_s|$ to be the running maximum process. According to Doob's L^p-inequality (in the case when $p=2$) and the definition of quadratic variation, it is seen that

$$\mathbb{E}[\langle M\rangle_t] = \mathbb{E}[M_t^2] \leqslant \mathbb{E}[(M_t^*)^2] \leqslant 4\mathbb{E}[M_t^2] = 4\mathbb{E}[\langle M\rangle_t] \tag{5.19}$$

for every $t\geqslant 0$. In other words, the running maximum and the quadratic variation control each other in some universal way which is independent of the underlying martingale. In a more functional analytic language, it suggests that the norm

$$\|M\|' \triangleq \sqrt{\mathbb{E}[(M_\infty^*)^2]}$$

on H_0^2 is equivalent to the original norm $\|\cdot\|_{\mathbb{H}^2}$, where $M_\infty^* \triangleq \sup_{0\leqslant t<\infty}|M_t|$. However, this simple fact relies on the very special L^2-structure, in which we have $\mathbb{E}[M_t^2] = \mathbb{E}[\langle M\rangle_t]$ making the story a lot easier.

The BDG inequalities investigates the L^p-situation for all $0<p<\infty$. Here is the main theorem.

Theorem 5.6. *For each $0<p<\infty$, there exist universal constants $C_{1,p}, C_{2,p}>0$, such that for every continuous local martingale $M\in\mathcal{M}_0^{\mathrm{loc}}$, we have*

$$C_{1,p}\mathbb{E}[\langle M\rangle_t^p] \leqslant \mathbb{E}[(M_t^*)^{2p}] \leqslant C_{2,p}\mathbb{E}[\langle M\rangle_t^p], \quad \forall t\geqslant 0, \tag{5.20}$$

where $M_t^* \triangleq \sup_{0\leqslant s\leqslant t}|M_s|$.

Proof. We prove the theorem by several steps. To simplify our notation, we will always use C_p to denote a universal constant which depends only on p although it may be different from line to line.

(1) By localization, we may assume that M and $\langle M\rangle$ are both uniformly bounded. Indeed, if we are able to prove the theorem for this case, since the constants $C_{1,p}, C_{2,p}$ will be universal, it is not hard to see that the general case follows by removing the localization.

(2) The case $p=1$. This is done in view of (5.19). In this case, we have $C_{1,1}=1$ and $C_{2,1}=4$.

(3) The case $p>1$.

We first prove the right hand side of (5.20). Let $f(x) = x^{2p}$. Then $f \in C^2(\mathbb{R}^1)$, and

$$f'(x) = 2px^{2p-1},\ f''(x) = 2p(2p-1)x^{2(p-1)}.$$

According to Itô's formula, we have

$$M_t^{2p} = 2p\int_0^t M_s^{2p-1}dM_s + p(2p-1)\int_0^t M_s^{2(p-1)}d\langle M\rangle_s. \tag{5.21}$$

Since M and $\langle M\rangle$ are bounded, we can see that the local martingale part in (5.21) is indeed a martingale. Therefore,

$$\mathbb{E}[M_t^{2p}] = p(2p-1)\mathbb{E}\left[\int_0^t M_s^{2(p-1)}d\langle M\rangle_s\right] \leqslant p(2p-1)\mathbb{E}\left[(M_t^*)^{2(p-1)}\langle M\rangle_t\right].$$

On the one hand, Doob's L^p-inequality gives that

$$\mathbb{E}[(M_t^*)^{2p}] \leqslant C_p\mathbb{E}[M_t^{2p}],$$

while on the other hand, Hölder's inequality gives that

$$\mathbb{E}\left[(M_t^*)^{2(p-1)}\langle M\rangle_t\right] \leqslant \|\langle M\rangle_t\|_p \left\|(M_t^*)^{2(p-1)}\right\|_q,$$

where $q \triangleq p/(p-1)$ is the Hölder conjugate of p. By rearranging the resulting inequality, we arrive at

$$\mathbb{E}\left[(M_t^*)^{2p}\right] \leqslant C_p\mathbb{E}[\langle M\rangle_t^p].$$

To see the left hand side of (5.20), let $A_t \triangleq \langle M\rangle_t$. To estimate $A_t^p = \int_0^t pA_s^{p-1}dA_s$, the key is to regard it as the quadratic variation process of some martingale and use Itô's formula to estimate this martingale. More precisely, define $N_t \triangleq \int_0^t A_s^{\frac{p-1}{2}}dM_s$. Then $\langle N\rangle_t = A_t^p/p$. On the other hand, since the process $A_t^{\frac{p-1}{2}}$ is bounded and increasing, Itô's formula yields that

$$M_tA_t^{\frac{p-1}{2}} = N_t + \int_0^t M_s dA_s^{\frac{p-1}{2}}.$$

Therefore, $|N_t| \leqslant 2M_t^*A_t^{\frac{p-1}{2}}$ and thus

$$\frac{1}{p}\mathbb{E}[A_t^p] = \mathbb{E}[\langle N\rangle_t] = \mathbb{E}[|N_t|^2] \leqslant 4\mathbb{E}[(M_t^*)^2A_t^{p-1}].$$

By applying Hölder's inequality on the right hand side and rearranging the resulting inequality, we arrive at

$$\mathbb{E}[A_t^p] \leqslant C_p\mathbb{E}[(M_t^*)^{2p}].$$

(4) The case $0 < p < 1$. We still use A_t to denote $\langle M\rangle_t$.

We first prove the right hand side of (5.20). Again we define $N_t \triangleq \int_0^t A_s^{\frac{p-1}{2}} dM_s$ so that $\langle N\rangle_t = A_t^p/p$. According to the associativity of stochastic integrals (c.f. Proposition 5.10), we see that $M_t = \int_0^t A_s^{\frac{1-p}{2}} dN_s$. Since the process $A_t^{\frac{1-p}{2}}$ is bounded and increasing, by Itô's formula, we have

$$N_t A_t^{\frac{1-p}{2}} = M_t + \int_0^t N_s dA_s^{\frac{1-p}{2}}.$$

Therefore, $|M_t| \leqslant 2N_t^* A_t^{\frac{1-p}{2}}$. Since this is true for all $t \geqslant 0$, we see that $M_t^* \leqslant 2N_t^* A_t^{\frac{1-p}{2}}$ and thus

$$\begin{aligned}
\mathbb{E}[(M_t^*)^{2p}] &\leqslant 2^{2p}\mathbb{E}\left[(N_t^*)^{2p} A_t^{p(1-p)}\right] \\
&\leqslant 2^{2p}\mathbb{E}[(N_t^*)^2]^p \cdot \mathbb{E}[A_t^p]^{1-p} \\
&\leqslant 2^{2p}4^p\mathbb{E}[N_t^2]^p \cdot \mathbb{E}[A_t^p]^{1-p} \\
&= 2^{2p}4^p\mathbb{E}[\langle N\rangle_t]^p \cdot \mathbb{E}[A_t^p]^{1-p} \\
&= \frac{16^p}{p^p}\mathbb{E}[A_t^p].
\end{aligned}$$

Finally, we prove the left hand side of (5.20). Given $\alpha > 0$, consider

$$A_t^p = A_t^p(\alpha + M_t^*)^{-2p(1-p)}(\alpha + M_t^*)^{2p(1-p)}$$

and Hölder's inequality gives

$$\mathbb{E}[A_t^p] \leqslant \left(\mathbb{E}[A_t(\alpha + M_t^*)^{-2(1-p)}]\right)^p \left(\mathbb{E}[(\alpha + M_t^*)^{2p}]\right)^{1-p}. \qquad (5.22)$$

Here the reason of introducing α is to avoid the singularity in the term $(\alpha + M_t^*)^{-2(1-p)}$. Now since

$$A_t(\alpha + M_t^*)^{-2(1-p)} \leqslant \int_0^t (\alpha + M_s^*)^{-2(1-p)} dA_s, \qquad (5.23)$$

we introduce the martingale $N_t = \int_0^t (\alpha + M_s^*)^{-(1-p)} dM_s$ so that its quadratic variation process coincides with the right hand side of (5.23). Since the process $(\alpha + M_t^*)^{-(1-p)}$ is bounded and has bounded variation, from Itô's formula, we know that

$$\begin{aligned}
(\alpha + M_t^*)^{-(1-p)} M_t &= N_t + \int_0^t M_s d(\alpha + M_s^*)^{p-1} \\
&= N_t + (p-1)\int_0^t M_s(\alpha + M_s^*)^{p-2} dM_s^*.
\end{aligned}$$

Therefore,

$$\begin{aligned}
|N_t| &\leqslant (\alpha + M_t^*)^{p-1} M_t^* + (1-p)\int_0^t M_s^*(\alpha + M_s^*)^{p-2} dM_s^* \\
&\leqslant (M_t^*)^p + \frac{1-p}{p}(M_t^*)^p \\
&= \frac{1}{p}(M_t^*)^p.
\end{aligned}$$

It follows that

$$\mathbb{E}[\langle N\rangle_t] = \mathbb{E}[N_t^2] \leqslant \frac{1}{p^2}\mathbb{E}[(M_t^*)^{2p}].$$

Combining this with inequalities (5.22) and (5.23), we arrive at

$$\mathbb{E}[A_t^p] \leqslant \frac{1}{p^{2p}}\left(\mathbb{E}[(M_t^*)^{2p}]\right)^p \cdot \left(\mathbb{E}[(\alpha+M_t^*)^{2p}]\right)^{1-p}.$$

Since this is true for all $\alpha > 0$, the result follows by letting $\alpha \downarrow 0$. □

Remark 5.8. From the proof, we can actually see that the constants $C_{1,p}$ and $C_{2,p}$ can be written down explicitly, although there is no need to do so.

5.5 Lévy's characterization of Brownian motion

It is a rather deep and remarkable fact that most Markov processes can be characterized by certain martingale properties. This is the renowned martingale problem of Stroock and Varadhan, which we will touch at an introductory level when we study stochastic differential equations. Here we investigate the special case of Brownian motion, which is the content of Lévy's characterization theorem. This result, along with the series of martingale representation theorems that we shall prove in the sequel, reveals the intimacy between continuous martingales and Brownian motion. Probably this explains why martingale methods are so powerful and why the Brownian motion is so fundamental in the theory of Itô's calculus.

Suppose that $B_t = (B_t^1, \cdots, B_t^d)$ is a d-dimensional $\{\mathcal{F}_t\}$-Brownian motion. Apparently, we have $\langle B^i\rangle_t = t$ for each $1 \leqslant i \leqslant d$. Moreover, for $i \neq j$, from the simple observation that $\frac{\sqrt{2}}{2}(B_t^i \pm B_t^j)$ are both $\{\mathcal{F}_t\}$-Brownian motions, we conclude that $\left\langle \frac{\sqrt{2}}{2}(B^i \pm B^j)\right\rangle_t = t$. Therefore, $\langle B^i, B^j\rangle_t = 0$. In other words, we know that $\langle B^i, B^j\rangle_t = \delta_{ij}t$ for all $1 \leqslant i,j \leqslant d$. Lévy's characterization theorem tells us that this property characterizes the Brownian motion.

Theorem 5.7. *Let $M_t = (M_t^1, \cdots, M_t^d)$ be a vector of continuous $\{\mathcal{F}_t\}$-local martingales vanishing at $t = 0$. Suppose that*

$$\langle M^i, M^j\rangle_t = \delta_{ij}t, \quad t \geqslant 0.$$

Then M_t is an $\{\mathcal{F}_t\}$-Brownian motion.

Proof. The key is to use the following neat characterization of an $\{\mathcal{F}_t\}$-Brownian motion in terms of characteristic functions (see also the proof of Theorem 4.2): it suffices to show that

$$\mathbb{E}\left[\mathrm{e}^{i\langle\theta, M_t - M_s\rangle}|\mathcal{F}_s\right] = \mathrm{e}^{-\frac{1}{2}|\theta|^2(t-s)},\ \forall \theta \in \mathbb{R}^d \text{ and } s < t. \tag{5.24}$$

Let $f = (f_1, \cdots, f_d) \in L^2([0,\infty); \mathbb{R}^d)$ and define the (complex-valued) exponential martingale

$$\mathcal{E}^{if}(M)_t \triangleq \exp\left(i\sum_{j=1}^d \int_0^t f_j(s)dM_s^j + \frac{1}{2}\sum_{j=1}^d \int_0^t f_j^2(s)ds\right), \quad t \geqslant 0.$$

By applying Itô's formula to the vector semimartingale

$$\left(\sum_{j=1}^{d}\int_0^t f_j(s)dM_s^j, \sum_{j=1}^{d}\int_0^t f_j^2(s)ds\right)$$

in $\mathbb{R}^2$ and the function $f(x,y) = \mathrm{e}^{ix+y/2}$, it can be easily seen that $\mathcal{E}^{if}(M)_t$ is a continuous local martingale (starting at 1). Since it is uniformly bounded, we know that it is indeed a martingale (c.f. Proposition 5.2).

Now let $\theta \in \mathbb{R}^d$. For $T > 0$, consider $f \triangleq \theta \mathbf{1}_{[0,T]} \in L^2([0,\infty);\mathbb{R}^d)$. In this case, we conclude that

$$\mathcal{E}^{if}(M)_t = \mathrm{e}^{i\langle\theta, M_{T\wedge t}\rangle + \frac{1}{2}|\theta|^2 T\wedge t}, \quad t \geqslant 0,$$

is a martingale. This is true for every $T > 0$. Therefore, if we consider $s < t < T$, then for any $A \in \mathcal{F}_s$, we have

$$\begin{aligned}\mathbb{E}\left[\mathrm{e}^{i\langle\theta, M_t - M_s\rangle}\mathbf{1}_A\right] &= \mathbb{E}\left[\mathrm{e}^{-i\langle\theta, M_s\rangle}\mathbf{1}_A\mathbb{E}\left[\mathrm{e}^{i\langle\theta, M_t\rangle}|\mathcal{F}_s\right]\right] \\ &= \mathbb{E}\left[\mathbf{1}_A\mathrm{e}^{-\frac{1}{2}|\theta|^2(t-s)}\right] \\ &= \mathbb{P}(A)\mathrm{e}^{-\frac{1}{2}|\theta|^2(t-s)},\end{aligned}$$

which implies (5.24). □

5.6 Continuous local martingales as time-changed Brownian motions

Sometimes it can be very useful if we change the speed of a process. In particular, if we change the speed of a continuous martingale in a proper way, we can get a Brownian motion! Because the Brownian motion is so simple and explicit, this technique could have lots of nice applications.

We should not be too surprised about this fact. Heuristically, let M_t be a continuous martingale with quadratic variation process $\langle M\rangle_t$. Since $\langle M\rangle_t$ is increasing, we can define an "inverse" process τ_t of $\langle M\rangle_t$. If we run M at speed τ, i.e. considering the process $\widehat{M_t} \triangleq M_{\tau_t}$, the optional sampling theorem will imply that $\widehat{M_t}$ is a martingale with respect to the filtration $\{\mathcal{F}_{\tau_t}\}$. Therefore, it is not unreasonable to expect that $\langle\widehat{M}\rangle = \langle M\rangle_{\tau_t} = t$ as τ_t is the "inverse" of $\langle M\rangle_t$. Lévy's characterization theorem then implies that $\widehat{M_t}$ is a Brownian motion.

Now we put this philosophy in a rigorous mathematical form, which is however technically quite involved.

We start with the discussion of a general time-change. This part is completely deterministic. Let $a: [0,\infty) \to [0,\infty)$ be a continuous and increasing function which vanishes at $t = 0$. Define

$$c_t \triangleq \inf\{s \geqslant 0 : a_s > t\}, \quad t \geqslant 0.$$

Definition 5.10. The function c_t is called the *time-change* associated with a_t.

The following properties are elementary and should be clear when a picture is drawn. They provide a good intuition about what the time-change looks like. The proof is easy and we leave it to the reader as an exercise. Denote $a_\infty \triangleq \lim_{t\to\infty} a_t$.

Proposition 5.17. *The time-change c_t of a_t satisfies the following properties.*

(1) c_t is strictly increasing and right continuous for $t < a_\infty$, and $c_t = \infty$ if $t \geqslant a_\infty$. If $a_\infty = \infty$, then $c_\infty \triangleq \lim_{t\to\infty} c_t = \infty$.

(2) For every s, t, $c_t < s \iff a_s > t$.

(3) Let $t = a_s$. Then $c_{t-} \leqslant s \leqslant c_t$. Moreover, for every t, $a \equiv \text{constant}$ on $[c_{t-}, c_t]$. This implies that the size of every jump for c_t corresponds to an interval of constancy for a_t and vice versa.

(4) For every $t \leqslant a_\infty$, $a_{c_t} = t$, and for every $s \leqslant \infty$, $c_{a_s} \geqslant s$. If s is an increasing point of a (i.e. $a(s') > a(s)$ for all $s' > s$), then $c_{a_s} = s$.

The time-change can give us a useful change of variable formula for integration. But we need to be a bit careful as it should involve some continuity property with respect to the time-change.

Definition 5.11. A continuous function $x: [0, \infty) \to \mathbb{R}^1$ is called *c-continuous* if x is constant on $[c_{t-}, c_t]$ for each t, where $c_{0-} \triangleq 0$.

Under c-continuity, we can prove the following change of variable formula.

Proposition 5.18. *Let x be a c-continuous function which has bounded variation on each finite interval. Then for any measurable function $y: [0, \infty) \to \mathbb{R}^1$, we have*

$$\int_{[c_{t_1}, c_{t_2}]} y_u dx_u = \int_{[t_1, t_2]} y_{c_v} dx_{c_v},$$

provided that $t_1 < t_2 < a_\infty$ and the integrals make sense.

Proof. First of all, observe that $a|_{[c_{t_1}, c_{t_2}]}: [c_{t_1}, c_{t_2}] \to [t_1, t_2]$ is well-defined and surjective. For simplicity we still denote it by a. Respectively, dx denotes the Lebesgue-Stieltjes measure induced by x on $[c_{t_1}, c_{t_2}]$ and μ denotes the push-forward of dx by a on $[t_1, t_2]$. It follows that for any $[v_1, v_2] \subseteq [t_1, t_2]$,

$$[c_{v_1}, c_{v_2}] \subseteq \{u \in [c_{t_1}, c_{t_2}]: v_1 \leqslant a_u \leqslant v_2\} \subseteq [c_{v_1 -}, c_{v_2}].$$

Since x is c-continuous, we conclude that

$$\mu([v_1, v_2]) = dx(\{u \in [c_{t_1}, c_{t_2}]: v_1 \leqslant a_u \leqslant v_2\}) = x_{c_{v_2}} - x_{c_{v_1}}.$$

In particular, μ coincides with the Lebesgue-Stieltjes measure induced by the function x_{c_v} on $[t_1, t_2]$.

According to the classical change of variable formula in measure theory, we have

$$\int_{[t_1, t_2]} y_{c_v} dx_{c_v} = \int_{[t_1, t_2]} y_{c_v} d\mu = \int_{[c_{t_1}, c_{t_2}]} y_{c_{a_u}} dx_u,$$

whenever the integrals make sense. But we know that $c_{a_u} = u$ for every increasing point u of a, and apparently there are at most countably many intervals of constancy for a. Therefore, from the c-continuity of x, we conclude that

$$\int_{[c_{t_1}, c_{t_2}]} y_{c_{a_u}} dx_u = \int_{[c_{t_1}, c_{t_2}]} y_u dx_u,$$

which then completes the proof. □

Now we put everything in a probabilistic context.

Recall that $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ is a filtered probability space which satisfies the usual conditions. Let A_t be an $\{\mathcal{F}_t\}$-adapted process such that with probability one, every sample path $t \mapsto A_t(\omega)$ is continuous and increasing which vanishes at $t = 0$. Define the process

$$C_t \triangleq \inf\{s \geqslant 0 : \ A_s > t\}, \quad t \geqslant 0,$$

to be the *time-change* associated with A_t.

Since the filtration $\{\mathcal{F}_s\}$ is right continuous, according to Proposition 5.17, (2), for every $t \geqslant 0$, C_t is an $\{\mathcal{F}_s\}$-stopping time. Therefore, we may define a new filtration $\widehat{\mathcal{F}}_t \triangleq \mathcal{F}_{C_t}$ associated with the time-change. Since C_t is right continuous, from Problem Sheet 2, Problem 4, (2), (i), we know that $\{\widehat{\mathcal{F}}_t\}$ also satisfies the usual conditions. In addition, for every $s \geqslant 0$, A_s is an $\{\widehat{\mathcal{F}}_t\}$-stopping time.

Now assume further that $A_\infty = \infty$ almost surely, so that with probability one, $C_t < \infty$ for all t.

Definition 5.12. Let $\{X_t\}$ be an $\{\mathcal{F}_t\}$-progressively measurable stochastic process. $\widehat{X}_t \triangleq X_{C_t}$ is called the *time-changed process* of X_t by C_t.

We are mainly interested in how a continuous local martingale behaves under a time-change. To emphasize the dependence on the filtration, we use $\mathcal{M}_0^{\mathrm{loc}}(\{\mathcal{F}_t\})$ to denote that space of continuous $\{\mathcal{F}_t\}$-local martingales vanishing at $t = 0$.

It is quite a subtle point that a time-changed continuous martingale can fail to be a local martingale even it is continuous. Here is a counterexample.

Example 5.1. Let B_t be a Brownian motion with augmented natural filtration $\{\mathcal{F}_t^B\}$. Define $A_t \triangleq \max_{0 \leqslant s \leqslant t} B_s$. Then A_t is a continuous and increasing process vanishing at $t = 0$, and $A_\infty = \infty$ almost surely. Let C_t be the time-change associated with A_t. Then we have $B_{C_t} = t$. Indeed, apparently we have $B_{C_t} \leqslant A_{C_t} = t$. If $B_{C_t} < A_{C_t}$, by the continuity of Brownian motion, we know that $A_u = A_{C_t} = t$ on $[C_t, C_t + \delta]$ for some small $\delta > 0$. It follows from the definition of C_t that $C_t \geqslant C_t + \delta$, which is a contradiction. Therefore, $B_{C_t} = t$. In particular, B_{C_t} cannot be a continuous local martingale regardless of what filtration we take since it has bounded variation on finite intervals.

To expect that a time-changed continuous local martingale is again a continuous local martingale, we need the C-continuity. A stochastic process X_t is called *C-continuous* if with probability one, X is constant on $[C_{t-}, C_t]$ for each t, where $C_{0-} \triangleq 0$ (c.f. Definition 5.11).

Proposition 5.19. *Let $A_\infty = \infty$ almost surely, and let C_t be the time-change associated with A_t. Suppose that $M \in \mathcal{M}_0^{\mathrm{loc}}(\{\mathcal{F}_t\})$ is C-continuous.*

(1) Let $\widehat{M}_t$ be the time-changed process of M_t by C_t. Then $\widehat{M} \in \mathcal{M}_0^{\mathrm{loc}}(\{\widehat{\mathcal{F}}_t\})$ and $\langle \widehat{M} \rangle = \widehat{\langle M \rangle}$.

(2) Let $\Phi \in L^2_{\mathrm{loc}}(M)$ with respect to $\{\mathcal{F}_t\}$. Then $\widehat{\Phi} \in L^2_{\mathrm{loc}}(\widehat{M})$ with respect to $\{\widehat{\mathcal{F}}_t\}$ and $I^{\widehat{M}}(\widehat{\Phi}) = \widehat{I^M(\Phi)}$.

Proof. (1) Since M_t is C-continuous, we know that $\widehat{M}_t$ is continuous and $\widehat{M}_0 = 0$. Moreover, it is obvious that $\widehat{M}_t$ is $\{\widehat{\mathcal{F}}_t\}$-adapted.

Now let τ be a finite $\{\mathcal{F}_t\}$-stopping time such that the stopped process M_t^τ is a bounded $\{\mathcal{F}_t\}$-martingale. Define $\widehat{\tau} = A_\tau$. From Proposition 5.17, (2), we see that $\{\widehat{\tau} > t\} = \{C_t < \tau\}$. Therefore, $\widehat{\tau}$ is a finite $\{\widehat{\mathcal{F}}_t\}$-stopping time. In addition, by definition we have

$$\widehat{M_t^{\widehat{\tau}}} = \widehat{M}_{\widehat{\tau} \wedge t} = M_{C_{\widehat{\tau} \wedge t}}.$$

If $\widehat{\tau} > t$, then $C_t < \tau$ and $\widehat{M_t^{\widehat{\tau}}} = M_{C_t} = M_{\tau \wedge C_t}$. If $\widehat{\tau} \leqslant t$, then $\widehat{M_t^{\widehat{\tau}}} = M_{C_{\widehat{\tau}}}$. But from Proposition 5.17, (3), we know that $C_{\widehat{\tau}-} \leqslant \tau \leqslant C_{\widehat{\tau}}$. By the C-continuity of M, M is constant on $[\tau, C_{\widehat{\tau}}]$. Therefore, $\widehat{M_t^{\widehat{\tau}}} = M_\tau = M_{\tau \wedge C_t}$ since $C_t \geqslant \tau$ in this case. In other words, we conclude that $\widehat{M_t^{\widehat{\tau}}} = M_{C_t}^\tau$ for all t. In particular, $\widehat{M_t^{\widehat{\tau}}}$ is a bounded process. Applying the optional sampling theorem to the bounded $\{\mathcal{F}_t\}$-martingale M_t^τ which thus has a last element (c.f. Corollary 3.2), we conclude that $\widehat{M_t^{\widehat{\tau}}}$ is an $\{\widehat{\mathcal{F}}_t\}$-martingale. If we let $\tau = \tau_n \uparrow \infty$, then $\widehat{\tau} = A_{\tau_n} \uparrow A_\infty = \infty$. Therefore, $\widehat{M} \in \mathcal{M}_0^{\mathrm{loc}}(\{\widehat{\mathcal{F}}_t\})$.

Finally, since M_t is C-continuous, according to Proposition 5.8, we see that $\langle M \rangle_t$ is also C-continuous. Therefore, $M^2 - \langle M \rangle \in \mathcal{M}_0^{\mathrm{loc}}(\{\mathcal{F}_t\})$ is C-continuous. From what was just proved, we know that $\widehat{M}^2 - \widehat{\langle M \rangle} \in \mathcal{M}_0^{\mathrm{loc}}(\{\widehat{\mathcal{F}}_t\})$. Therefore, $\langle \widehat{M} \rangle = \widehat{\langle M \rangle}$.

(2) First of all, it is apparent that $\int_0^t \widehat{\Phi}_s^2 d\langle \widehat{M} \rangle_s = \int_{C_0}^{C_t} \Phi_s^2 d\langle M \rangle_s < \infty$ almost surely for every t.

To prove the last claim, we first need to observe a slightly more general fact than what was proved in (1): if $M, N \in \mathcal{M}_0^{\mathrm{loc}}(\{\mathcal{F}_t\})$ are C-continuous, then $\langle M, N \rangle_t$ is C-continuous, and $\langle \widehat{M}, \widehat{N} \rangle = \widehat{\langle M, N \rangle}$. Indeed, the fact that $\langle M, N \rangle_t$ is C-continuous follows from the identity

$$\langle M, N \rangle_t = \frac{1}{4}(\langle M + N \rangle_t - \langle M - N \rangle_t).$$

Therefore, $MN - \langle M, N \rangle \in \mathcal{M}_0^{\mathrm{loc}}$ is C-continuous, which proves the claim according to (1).

Coming back to the proposition, in order to show that $I^{\widehat{M}}(\widehat{\Phi}) = \widehat{I^M(\Phi)}$, it suffices to show that $\left\langle I^{\widehat{M}}(\widehat{\Phi}) - \widehat{I^M(\Phi)} \right\rangle = 0$. On the one hand, we have

$$\left\langle I^{\widehat{M}}(\widehat{\Phi}) \right\rangle_t = \int_0^t \widehat{\Phi}_s^2 d\langle \widehat{M} \rangle_s = \int_0^{C_t} \Phi_s^2 d\langle M \rangle_s = \left\langle \widehat{I^M(\Phi)} \right\rangle_t.$$

On the other hand,

$$\begin{aligned}\left\langle I^{\widehat{M}}(\widehat{\Phi}), \widehat{I^M(\Phi)}\right\rangle_t &= \widehat{\Phi} \bullet \left\langle \widehat{M}, \widehat{I^M(\Phi)}\right\rangle = \widehat{\Phi} \bullet \langle M, \widehat{I^M(\Phi)}\rangle \\ &= \widehat{\Phi} \bullet \left(\widehat{\Phi \bullet \langle M\rangle}\right) = \left\langle I^{\widehat{M}}(\widehat{\Phi})\right\rangle.\end{aligned}$$

Therefore, the result follows. □

Now we are able to prove the main result of this subsection. This is known as the Dambis-Dubins-Schwarz theorem.

Theorem 5.8. *Let $M \in \mathcal{M}_0^{\mathrm{loc}}(\{\mathcal{F}_t\})$ be such that $\langle M\rangle_\infty = \infty$ almost surely. Define C_t to be the time-change associated with $\langle M\rangle_t$. Then $B_t \triangleq M_{C_t}$ is an $\{\mathcal{F}_{C_t}\}$-Brownian motion and $M_t = B_{\langle M\rangle_t}$.*

Proof. From Proposition 5.17, (3), we know that $\langle M\rangle_t$, and hence M_t, is C-continuous. Therefore, by Proposition 5.19, (1), $B \in \mathcal{M}_0^{\mathrm{loc}}(\{\mathcal{F}_{C_t}\})$ and $\langle B\rangle_t = \widehat{\langle M\rangle}_t = \langle M\rangle_{C_t} = t$. According to Lévy's characterization theorem (c.f. Theorem 5.7), we conclude that B_t is an $\{\mathcal{F}_{C_t}\}$-Brownian motion. Finally, for each $t \geqslant 0$, from Proposition 5.17, (3) again, we know that $\langle M\rangle$, and hence M, is constant on $[C_{\langle M\rangle_t -}, C_{\langle M\rangle_t}]$, as well as $t \in [C_{\langle M\rangle_t -}, C_{\langle M\rangle_t}]$. Therefore, $M_t = M_{C_{\langle M\rangle_t}} = B_{\langle M\rangle_t}$. □

The condition $\langle M\rangle_\infty = \infty$ almost surely in the Dambis-Dubins-Schwarz theorem ensures that the underlying probability space is rich enough to support a Brownian motion. To generalize the theorem to the case when $\langle M\rangle_\infty < \infty$ with positive probability, we need to enlarge the underlying probability space.

Definition 5.13. An *enlargement* of a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ is another filtered probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}; \{\widetilde{\mathcal{F}}_t\})$ together with a projection $\pi : \widetilde{\Omega} \to \Omega$, such that $\widetilde{\mathbb{P}} \circ \pi^{-1} = \mathbb{P}$, $\pi^{-1}(\mathcal{F}) \subseteq \widetilde{\mathcal{F}}$ and $\pi^{-1}(\mathcal{F}_t) \subseteq \pi^{-1}(\widetilde{\mathcal{F}}_t)$ for all t.

If $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}; \{\widetilde{\mathcal{F}}_t\})$ is an enlargement of $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$, then associated with any given stochastic process X_t on Ω, we can define a process $\widetilde{X}_t$ on $\widetilde{\Omega}$ by

$$\widetilde{X}_t(\widetilde{\omega}) \triangleq X_t(\pi(\widetilde{\omega})), \quad \widetilde{\omega} \in \widetilde{\Omega}, \tag{5.25}$$

canonically. Apparently, the law of $\widetilde{X}$ is the same as the law of X by the definition of enlargement. For simplicity, we may use the same notation X to denote $\widetilde{X}$.

Now we have the following extension of the Dambis-Dubins-Schwarz theorem. Recall from Problem Sheet 5, Problem 1, (3) that if $M \in \mathcal{M}_0^{\mathrm{loc}}$, then with probability one, $M_\infty = \lim_{t\to\infty} M_t$ exists finitely on the event $\{\langle M\rangle_\infty < \infty\}$.

Theorem 5.9. *Let $M \in \mathcal{M}_0^{\mathrm{loc}}(\{\mathcal{F}_t\})$. Define C_t to be the time-change associated with $\langle M\rangle_t$. Then there exists an enlargement $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}; \{\widetilde{\mathcal{F}}_t\})$ of $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_{C_t}\})$ and a Brownian motion $\widetilde{\beta}$ on $\widetilde{\Omega}$ which is independent of M, such that the process*

$$B_t \triangleq \begin{cases} M_{C_t}, & t < \langle M\rangle_\infty; \\ M_\infty + & \widetilde{\beta}_t - \widetilde{\beta}_{t\wedge\langle M\rangle_\infty}, t \geqslant \langle M\rangle_\infty, \end{cases}$$

is an $\{\widetilde{\mathcal{F}}_t\}$-Brownian motion. Moreover, $M_t = B_{\langle M\rangle_t}$.

Proof. Let $(\Omega', \mathcal{F}', \mathbb{P}'; \{\mathcal{F}'_t\})$ be a filtered probability space on which an $\{\mathcal{F}'_t\}$-Brownian motion β_t is defined. Let $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}; \{\widetilde{\mathcal{F}}_t\})$ be the usual augmentation of $(\Omega \times \Omega', \mathcal{F} \times \mathcal{F}', \mathbb{P}\times\mathbb{P}'; \{\mathcal{F}_{C_t}\times\mathcal{F}'_t\})$. Apparently, $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}; \{\widetilde{\mathcal{F}}_t\})$ is an enlargement of $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_{C_t}\})$ with projection $\pi((\omega, \omega')) = \omega$. Define $\widetilde{\beta}_t((\omega, \omega')) \triangleq \beta_t(\omega')$. Then $\widetilde{\beta}_t$ is a Brownian motion on $\widetilde{\Omega}$. It is apparent that $\widetilde{\beta}$ and M are independent.

An important general fact for this enlargement is that for every $X \in \mathcal{M}_0^{\mathrm{loc}}(\{\mathcal{F}_{C_t}\})$ on Ω, we have $\widetilde{X} \in \mathcal{M}_0^{\mathrm{loc}}(\{\widetilde{\mathcal{F}}_t\})$ and $\langle \widetilde{X} \rangle = \widetilde{\langle X \rangle}$ almost surely on $\widetilde{\Omega}$, where $\widetilde{X}$ ($\widetilde{\langle X \rangle}$, respectively) is the process on $\widetilde{\Omega}$ defined by pulling back X ($\langle X \rangle$, respectively) via the projection π (c.f. (5.25)). Similarly, for every $\{\mathcal{F}_{C_t}\}$-stopping time τ on Ω, $\widetilde{\tau}$ is an $\{\widetilde{\mathcal{F}}_t\}$-stopping time on $\widetilde{\Omega}$.

Now we rewrite the definition of B_t in the following form:

$$B_t = M_{C_t} + \int_0^t \mathbf{1}_{(\langle M \rangle_\infty, \infty)}(s) d\widetilde{\beta}_s.$$

On the one hand, by adapting the argument in the proof of Proposition 5.19, (1), we can see that $M_{C_\cdot} \in \mathcal{M}_0^{\mathrm{loc}}(\{\widetilde{\mathcal{F}}_t\})$ with quadratic variation process $t \wedge \langle M \rangle_\infty$. On the other hand, $\int_0^\cdot \mathbf{1}_{(\langle M \rangle_\infty, \infty)}(s) d\widetilde{\beta}_s \in \mathcal{M}_0^{\mathrm{loc}}(\{\widetilde{\mathcal{F}}_t\})$ with quadratic variation process $t - t \wedge \langle M \rangle_\infty$. Therefore, $B \in \mathcal{M}_0^{\mathrm{loc}}(\{\widetilde{\mathcal{F}}_t\})$ and

$$\langle B \rangle_t = t + 2 \int_0^t \mathbf{1}_{(\langle M \rangle_\infty, \infty)}(s) d\langle M_{C_\cdot}, \widetilde{\beta} \rangle_s.$$

Finally, by the independence of of $\pi^{-1}(\mathcal{F}_\infty)$ and $(\pi')^{-1}(\mathcal{F}'_\infty)$, it is not hard to see that $M_{C_\cdot}\widetilde{\beta} \in \mathcal{M}_0^{\mathrm{loc}}(\{\widetilde{\mathcal{F}}_t\})$. Therefore, $\langle M_{C_\cdot}, \widetilde{\beta} \rangle = 0$, which implies that $\langle B \rangle_t = t$. According to Lévy's characterization theorem, B_t is an $\{\widetilde{\mathcal{F}}_t\}$-Brownian motion. The fact that $M_t = B_{\langle M \rangle_t}$ follows from the same reason as in the proof of Theorem 5.8. □

A natural question is whether the Dambis-Dubins-Schwarz theorem can be extended to multidimensions. This is the content of Knight's theorem.

Theorem 5.10. *Let $M_t = (M_t^1, \cdots, M_t^d)$ be d continuous $\{\mathcal{F}_t\}$-local martingales vanishing at $t = 0$, such that $\langle M^j, M^k \rangle_t = 0$ for $j \neq k$. Then there exists an enlargement $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ of $(\Omega, \mathcal{F}, \mathbb{P})$ and a d-dimensional Brownian motion β on $\widetilde{\Omega}$ which is independent of M, such that the process $B_t = (B_t^1, \cdots, B_t^d)$ defined by*

$$B_t^j = \begin{cases} M^j_{C^j_t}, & t < \langle M^j \rangle_\infty; \\ M^j_\infty + \beta^j_t - \beta^j_{t \wedge \langle M^j \rangle_\infty}, & t \geqslant \langle M^j \rangle_\infty, \end{cases}$$

is a d-dimensional Brownian motion.

Proof. From the proof of Theorem 5.9, we can see that on some enlargement $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ of $(\Omega, \mathcal{F}, \mathbb{P})$, every B_t^j is a one dimensional Brownian motion. It remains to show that

$B^1, \cdots, B^d$ are independent. To this end, we again use the method of characteristic functions. Let f_j $(1 \leqslant j \leqslant d)$ be a real step function of the form

$$f_j(t) = \sum_{k=1}^{m} \lambda_j^k \mathbf{1}_{(t_{k-1}, t_k]}(t).$$

We only need to show that $\mathbb{E}[L] = 1$, where

$$L \triangleq \exp\left(i \sum_{j=1}^{d} \int_0^\infty f_j(s) dB_s^j + \frac{1}{2} \sum_{j=1}^{d} \int_0^\infty f_j^2(s) ds \right).$$

The independence then follows immediately since the equation $\mathbb{E}[L] = 1$ (for arbitrary λ_j^k and t_k) gives the right characteristic functions for the finite dimensional distributions of B_t.

We use A_t^j to denote $\langle M^j \rangle_t$. Since $s < A_u^j \leqslant t \iff C_s^j < u \leqslant C_t^j$, we know that

$$M_{C_t^j}^j - M_{C_s^j}^j = \int_0^\infty \mathbf{1}_{(C_s^j, C_t^j]}(u) dM_u^j = \int_0^\infty \mathbf{1}_{(s,t]}(A_u^j) dM_u^j.$$

Therefore, by the definition of B^j and f_j, we have

$$\begin{aligned} \int_0^\infty f_j(t) dB_t^j &= \int_0^{A_\infty^j} f_j(t) dB_t^j + \int_{A_\infty^j}^\infty f_j(t) d\beta_t^j \\ &= \int_0^\infty f_j(A_t^j) dM_t^j + \int_{A_\infty^j}^\infty f_j(t) d\beta_t^j. \end{aligned} \tag{5.26}$$

On the other hand, a simple change of variables also shows that

$$\int_0^\infty f_j^2(t) dt = \int_0^\infty f_j^2(A_t^j) dA_t^j + \int_{A_\infty^j}^\infty f_j^2(t) dt. \tag{5.27}$$

Now define

$$I_t \triangleq \exp\left(i \sum_{j=1}^{d} \int_0^t f_j(A_s^j) dM_s^j + \frac{1}{2} \sum_{j=1}^{d} \int_0^t f_j^2(A_s^j) dA_s^j \right), \quad t \geqslant 0.$$

From Itô's formula and the assumption that $\langle M^j, M^k \rangle_t = 0$ for $j \neq k$, I_t is a bounded $\{\mathcal{F}_t\}$-martingale. Therefore, $\mathbb{E}[I_\infty] = \mathbb{E}[I_0] = 1$. Moreover, define

$$J \triangleq \exp\left(i \sum_{j=1}^{d} \int_{A_\infty^j}^\infty f_j(t) d\beta_t^j + \frac{1}{2} \sum_{j=1}^{d} \int_{A_\infty^j}^\infty f_j^2(t) dt \right).$$

From (5.26) and (5.27) we know that $L = I_\infty J$. But $\mathbb{E}\left[J | \mathcal{F}^M\right] = 1$ where $\mathcal{F}^M$ is the σ-algebra generated by M, since the conditional distribution of $\sum_{j=1}^d \int_{A_\infty^j}^\infty f_j(t) d\beta_t^j$ given M is Gaussian with mean 0 and variance $\sum_{j=1}^d \int_{A_\infty^j}^\infty f_j^2(t) dt$. Therefore,

$$\mathbb{E}[L] = \mathbb{E}[I_\infty J] = \mathbb{E}\left[\mathbb{E}\left[I_\infty J | \mathcal{F}^M\right]\right] = \mathbb{E}\left[I_\infty \mathbb{E}\left[J | \mathcal{F}^M\right]\right] = \mathbb{E}[I_\infty] = 1,$$

which completes the proof. □

Remark 5.9. Although Knight's theorem is a generalization of the Dambis-Dubins-Schwarz theorem to higher dimensions, it is somehow less precise because there is no counterpart of a filtration with respect to which the time-changed process B_t is a Brownian motion.

5.7 Continuous local martingales as Itô's integrals

Now we take up the question about when a continuous local martingale M_t can be represented as an Itô's integral $\int_0^t \Phi_s dB_s$ where B_t is a Brownian motion. Formally speaking, the main results can be summarized as two parts:

(1) If a Brownian motion is given, then *every* continuous local martingale with respect to the *Brownian filtration* has such a representation.

(2) Given general continuous local martingale M, if $d\langle M\rangle_t$ is *absolutely continuous* with respect to dt, then M has such a representation for *some* Brownian motion defined possibly on an enlarged probability space.

Now we develop the first part, which is indeed much more surprising than the second one.

Suppose that B_t is a one dimensional Brownian motion and $\{\mathcal{F}_t^B\}$ is its augmented natural filtration.

Let $\mathcal{T}$ be the space of real step functions on $[0,\infty)$ of the form

$$f(t) = \sum_{k=1}^m \lambda_k \mathbf{1}_{(t_{k-1},t_k]}(t), \quad t \geqslant 0,$$

For an $f \in \mathcal{T}$, define

$$\mathcal{E}_t^f \triangleq \exp\left(\int_0^t f(s)dB_s - \frac{1}{2}\int_0^t f^2(s)ds\right), \quad t \geqslant 0,$$

to be the associated exponential martingale. It is apparent that $\mathcal{E}_t^f$ is uniformly bounded in L^2.

The following lemma reveals why the Brownian filtration is crucial.

Lemma 5.3. *The set* $\{\mathcal{E}_\infty^f : f \in \mathcal{T}\}$ *is total in* $L^2(\Omega, \mathcal{F}_\infty^B, \mathbb{P})$.

Proof. Let $Y \in L^2(\Omega, \mathcal{F}_\infty^B, \mathbb{P})$ be such that $\mathbb{E}[Y\mathcal{E}_\infty^f] = 0$ for all $f \in \mathcal{T}$. We want to show that $Y = 0$. Define a finite signed measure μ on $(\Omega, \mathcal{F}_\infty^B)$ by

$$\mu(A) \triangleq \int_A Y d\mathbb{P}, \quad A \in \mathcal{F}_\infty^B.$$

It is equivalent to showing that $\mu = 0$. Since $\mathcal{F}_\infty^B$ is generated by the Brownian motion, it then suffices to prove that the induced finite signed measure ν on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ given by

$$\nu(\Gamma) = \int_\Omega Y \mathbf{1}_{\{(B_{t_1},\cdots,B_{t_n})\in\Gamma\}} d\mathbb{P}, \quad \Gamma \in \mathcal{B}(\mathbb{R}^n),$$

is identically zero, for every choice of $n \geqslant 1$ and $0 \leqslant t_1 < t_2 < \cdots < t_n < \infty$. But this is equivalent to showing that the Fourier transform of ν is zero.

By definition, the Fourier transform of ν is given by

$$\varphi(\lambda) \triangleq \int_{\mathbb{R}^n} \mathrm{e}^{i(\lambda_1 x_1 + \cdots + \lambda_n x_n)} \nu(dx), \quad \lambda = (\lambda_1, \cdots, \lambda_n) \in \mathbb{R}^n.$$

Moreover, by the definition of ν and a standard approximation argument by simple functions, it is easy to see that

$$\int_{\mathbb{R}^n} g(x_1, \cdots, x_n) \nu(dx) = \mathbb{E}[Y g(B_{t_1}, \cdots, B_{t_n})]$$

for any bounded Borel measurable function g on $\mathbb{R}^n$. In particular,

$$\varphi(\lambda) = \mathbb{E}\left[Y \mathrm{e}^{i(\lambda_1 B_{t_1} + \cdots + \lambda_n B_{t_n})}\right].$$

To see why $\varphi(\lambda)$ is identically zero, we define a complex-valued function Φ on $\mathbb{C}^n$ by

$$\Phi(z) \triangleq \mathbb{E}\left[Y \mathrm{e}^{z_1 B_{t_1} + \cdots + z_n B_{t_n}}\right], \quad z = (z_1, \cdots, z_n) \in \mathbb{C}^n.$$

It is apparent that $\Phi(z)$ is analytic on $\mathbb{C}^n$. Moreover, when $z \in \mathbb{R}^n$, by assumption we have

$$\Phi(z) = \mathrm{e}^{\frac{1}{2}\int_0^\infty f^2(s)ds} \cdot \mathbb{E}\left[Y \mathcal{E}_\infty^f\right] = 0,$$

where

$$f(t) \triangleq \sum_{k=1}^n z_k' \mathbf{1}_{(t_{k-1}, t_k]}(t) \in \mathcal{T}$$

with $z_k' \triangleq z_k + \cdots + z_n$. According to the identity theorem in complex analysis, we conclude that Φ is identically zero on $\mathbb{C}^n$. Therefore, by taking $z = i\lambda$, we know that $\varphi(\lambda) = 0$. $\square$

Now we are able to prove the following representation theorem.

Theorem 5.11. *Let $\xi \in L^2(\Omega, \mathcal{F}_\infty^B, \mathbb{P})$. Then there exists a unique element $\Phi \in L^2(B)$, such that*

$$\xi = \mathbb{E}[\xi] + \int_0^\infty \Phi_s dB_s. \tag{5.28}$$

Proof. Suppose that Φ and Φ' both satisfy (5.28). Then $\int_0^\infty (\Phi_s - \Phi_s')dB_s = 0$, which implies that

$$\mathbb{E}\left[\int_0^\infty (\Phi_s - \Phi_s')^2 ds\right] = 0,$$

as $\int_0^\cdot (\Phi_s - \Phi_s')dB_s \in H_0^2$. Therefore, $\Phi = \Phi'$ in $L^2(B)$ and the uniqueness holds.

To see the existence, we first show that the space $\mathcal{H}$ of elements $\xi \in L^2(\Omega, \mathcal{F}_\infty^B, \mathbb{P})$ which has a representation (5.28) is a closed subspace of $L^2(\Omega, \mathcal{F}_\infty^B, \mathbb{P})$. Indeed, let

$$\xi_n = \mathbb{E}[\xi_n] + \int_0^\infty \Phi_s^{(n)} dB_s$$

be a sequence converging to some $\xi \in L^2(\Omega, \mathcal{F}_\infty^B, \mathbb{P})$. It follows that $\mathbb{E}[\xi_n] \to \mathbb{E}[\xi]$. Moreover, from

$$\left\| \int_0^\infty \Phi_s^{(m)} dB_s - \int_0^\infty \Phi_s^{(n)} dB_s \right\|_{L^2}^2 = \mathbb{E}\left[\int_0^\infty (\Phi_s^{(m)} - \Phi_s^{(n)})^2 ds \right],$$

we know that $\Phi^{(n)}$ is a Cauchy sequence in $L^2(B)$. According to Lemma 5.2, $\Phi^{(n)} \to \Phi \in L^2(B)$. Therefore, $\int_0^\cdot \Phi_s^{(n)} dB_s \to \int_0^\cdot \Phi_s dB_s$ in H_0^2, which implies that

$$\xi = \mathbb{E}[\xi] + \int_0^\infty \Phi_s dB_s.$$

Therefore, $\mathcal{H}$ is a closed subspace of $L^2(\Omega, \mathcal{F}_\infty^B, \mathbb{P})$.

Now the existence follows from the simple fact that $\mathcal{H}$ contains elements of the form $\mathcal{E}_\infty^f$ for $f \in \mathcal{T}$ and Lemma 5.3, since

$$\mathcal{E}_\infty^f = 1 + \int_0^\infty f(s) \mathcal{E}_s^f dB_s,$$

where $\mathcal{E}_t^f$ is the exponential martingale defined by

$$\mathcal{E}_t^f \triangleq \exp\left(\int_0^t f(s) dB_s - \frac{1}{2} \int_0^t f^2(s) ds \right) = 1 + \int_0^t f(s) \mathcal{E}_s^f dB_s$$

according to Itô's formula, and apparently $f \cdot \mathcal{E}^f \in L^2(B)$. □

Remark 5.10. From the proof of Theorem 5.11, we can see that the uniqueness of Φ is equivalent to saying that if $\Phi, \Phi' \in L^2(B)$ both satisfy (5.28), then with probability one, we have

$$\int_0^\infty (\Phi_s - \Phi_s')^2 ds = 0.$$

On the other hand, if we remove the restriction that $\Phi \in L^2(B)$, then uniqueness fails in the class $L^2_{\text{loc}}(B)$ provided $\int_0^\infty \Phi_s dB_s = \lim_{t \to \infty} \int_0^t \Phi_s dB_s$ exists finitely (c.f. Problem Sheet 5, Problem 5).

Remark 5.11. The reader should find it easy to obtain a local version of Theorem 5.11, i.e. the representation of $\xi \in L^2(\Omega, \mathcal{F}_T^B, \mathbb{P})$ as an Itô's integral over $[0, T]$ for given $T > 0$.

Theorem 5.11 enables us to prove the following representation for continuous local martingales with respect to the Brownian filtration. This is the main result of the first part.

Theorem 5.12. *Let M_t be a continuous $\{\mathcal{F}_t^B\}$-local martingale. Then M_t has the representation*

$$M_t = M_0 + \int_0^t \Phi_s dB_s, \tag{5.29}$$

for some $\Phi \in L^2_{\text{loc}}(B)$. Such representation is unique in the following sense: if Φ' is another process in $L^2_{\text{loc}}(B)$ which also satisfies (5.29), then $\Phi_\cdot(\cdot) = \Phi'_\cdot(\cdot)$ $\mathbb{P} \times dt$-almost everywhere, or equivalently, with probability one, $\Phi_\cdot(\omega) = \Phi'_\cdot(\omega)$ dt-almost everywhere.

Proof. We may assume that $M_0 = 0$ so that $M \in \mathcal{M}_0^{\text{loc}}(\{\mathcal{F}_t^B\})$.

If $M \in H_0^2$, according to Theorem 5.11, we know that

$$M_\infty = \int_0^\infty \Phi_s dB_s$$

for some $\Phi \in L^2(B)$. Therefore,

$$M_t = \mathbb{E}\left[M_\infty | \mathcal{F}_t^B\right] = \mathbb{E}\left[\int_0^\infty \Phi_s dB_s | \mathcal{F}_t^B\right] = \int_0^t \Phi_s dB_s,$$

which proves the representation for M.

In general, suppose τ_n is a sequence of finite $\{\mathcal{F}_t^B\}$-stopping times increasing to infinity such that $M^{\tau_n} \in H_0^2$ for each n. Write $M_t^{\tau_n} = \int_0^t \Phi_s^{(n)} dB_s$ for $\Phi^{(n)} \in L^2(B)$. According to Proposition 5.11, we have

$$\int_0^t \Phi_s^{(n)} dB_s = M_t^{\tau_n} = (M^{\tau_{n+1}})_t^{\tau_n} = \int_0^t \Phi_s^{(n+1)} \mathbf{1}_{[0,\tau_n]}(s) dB_s.$$

Therefore, with probability one, $\Phi_\cdot^{(n)}(\omega) = \Phi_\cdot^{(n+1)}(\omega)\mathbf{1}_{[0,\tau_n(\omega)]}(\cdot)$ dt-almost everywhere. This implies that with probability one, $\Phi_\cdot^{(n+1)}(\omega) = \Phi_\cdot^{(n)}(\omega)$ on $[0, \tau_n(\omega)]$, which enables us to patch all those $\Phi^{(n)}$'s to define a single process Φ. More precisely, let

$$\Phi \triangleq \left(\limsup_{n\to\infty} \Phi^{(n)}\right) \cdot \mathbf{1}_{\left\{\limsup_{n\to\infty} \Phi^{(n)} \text{ is finite}\right\}}.$$

Apparently, Φ is progressively measurable, and $\Phi \in L^2_{\text{loc}}(B)$. To see that $M_t = \int_0^t \Phi_s dB_s$, let $N \in \mathcal{M}_0^{\text{loc}}(\{\mathcal{F}_t^B\})$. Then

$$\langle M, N\rangle_{\tau_n \wedge t} = \langle M^{\tau_n}, N\rangle_t = \int_0^t \Phi_s^{(n)} d\langle B, N\rangle_s$$

for each n. But from the Kunita-Watanabi inequality (c.f. (5.4)) and the fact that with probability one, $\Phi_\cdot^{(n)}(\omega) = \Phi_\cdot(\omega)$ dt-almost everywhere on $[0, \tau_n(\omega)]$, we know that $\int_0^t \Phi_s^{(n)} d\langle B, N\rangle_s = \int_0^t \Phi_s d\langle B, N\rangle_s$ whenever $t \leqslant \tau_n$. Therefore, by letting $n \to \infty$, we conclude that

$$\langle M, N\rangle_t = \int_0^t \Phi_s d\langle B, N\rangle_s.$$

Since this is true for arbitrary $N \in \mathcal{M}_0^{\text{loc}}(\{\mathcal{F}_t^B\})$, we obtain the desired representation.

Finally, the uniqueness follows from the fact that if $\Phi \in L^2_{\text{loc}}(B)$ satisfies $\int_0^t \Phi_s dB_s = 0$ for every t, then with probability one,

$$\int_0^t \Phi_s^2 ds = 0, \quad \forall t \geqslant 0.$$

□

Remark 5.12. Since $B_0 = 0$ and $\{\mathcal{F}_t^B\}$ is the augmented natural filtration of B_t, every $\mathcal{F}_0$-measurable random variable is therefore a constant. In particular, M_0 is a constant for a continuous $\{\mathcal{F}_t^B\}$-local martingale.

The same argument extends to the multidimensional case without any difficulty. We only state the main result and leave the details to the reader.

Theorem 5.13. *Let B_t be a d-dimensional Brownian motion and let $\{\mathcal{F}_t^B\}$ be its augmented natural filtration. Then for any continuous $\{\mathcal{F}_t^B\}$-local martingale M_t, there exists $\Phi^j \in L^2_{\text{loc}}(B^j)$, such that*

$$M_t = M_0 + \sum_{j=1}^d \int_0^t \Phi_s^j dB_s^j.$$

These Φ_j's are unique in the sense that if Ψ^j's satisfy the same property, then with probability one,

$$(\Phi_\cdot^1(\omega), \cdots, \Phi_\cdot^d(\omega)) = (\Psi_\cdot^1(\omega), \cdots, \Psi_\cdot^d(\omega)), \quad dt - \text{a.e.}$$

An analogous result of Theorem 5.11 also holds in the multidimensional case, and we will not state it here.

Now we turn to the second part: what if the underlying filtration is not the Brownian filtration?

Suppose that $M_t = \int_0^t \Phi_s dB_s$ for some Brownian motion. Then $\langle M \rangle_t = \int_0^t \Phi_s^2 ds$. Therefore, a necessary condition for M having the representation as a stochastic integral is that $d\langle M \rangle_t$ is absolutely continuous with respect to the Lebesgue measure. Moreover, if we know the Radon-Nikodym derivative $d\langle M \rangle_t / dt = \gamma_t > 0$, then $B_t \triangleq \int_0^t \gamma_s^{-1/2} dM_s$ will be a Brownian motion by Lévy's characterization theorem, and from the associativity of stochastic integrals, of course we have $M_t = \int_0^t \gamma_s^{1/2} dB_t$. If γ_t is simply non-negative, in order to support a Brownian motion we need to enlarge the underlying probability space as we have seen in the last subsection.

To be precise, we are going to prove the following main result in the multidimensional setting. Let $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ be a filtered probability space which satisfies the usual conditions.

We are going to use matrix notation exclusively and apply results from standard linear algebra. To treat things in an elegant way, we first fix some notation. For a real

$m \times n$ matrix A, A^* is denoted as the transpose of A, and we define the norm of A to be $\|A\| \triangleq \max_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n} |A^i_j|$. The space of real $m \times n$ matrices is denoted by $\mathrm{Mat}(m,n)$. If $M_t = (M^1_t, \cdots, M^d_t)$ is a vector of continuous $\{\mathcal{F}_t\}$-local martingales, we use $\langle M \rangle_t$ to denote the matrix $\left(\langle M^i, M^j \rangle_t\right)_{1 \leqslant i,j \leqslant d}$. It is apparent that this matrix is symmetric and non-negative definite for each t. If Ψ_t is a matrix-valued process, we use $\Psi \bullet M$ to denote the vector-valued stochastic integral $\int \Psi \cdot dM$ as long as the matrix multiplication and the stochastic integral make sense in a componentwise manner. Apparently, $\langle \Psi \bullet M \rangle = \Psi \cdot \langle M \rangle \cdot \Psi^*$.

Recall that every real $d \times d$ matrix A has a singular value decomposition as $A = U \Lambda V^*$, where U, V are orthogonal matrices, Λ is a diagonal matrix with non-negative entries on the diagonal. Moreover, the nonzero elements on the diagonal of Λ are the square roots of nonzero eigenvalues of AA^* counted with multiplicity.

Theorem 5.14. *Let $M_t = (M^1_t, \cdots, M^d_t)$ be a vector of d continuous $\{\mathcal{F}_t\}$-local martingales. Suppose that there exist matrix-valued progressively measurable processes γ_t and Φ_t taking values in $\mathrm{Mat}(d,d)$ and $\mathrm{Mat}(d,r)$ respectively, such that:*

(1) with probability one, $\int_0^t \|\Phi_s\|^2 ds < \infty$ for every $t \geqslant 0$;

(2) $\langle M \rangle_t = \int_0^t \gamma_s ds$ for every $t \geqslant 0$;

(3) $\gamma_t = \Phi_t \cdot \Phi^*_t$ for every $t \geqslant 0$.

Then on an enlargement $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}; \{\widetilde{\mathcal{F}}_t\})$ of $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$, there exists an r-dimensional $\{\widetilde{\mathcal{F}}_t\}$-Brownian motion, such that

$$M_t = M_0 + \int_0^t \Phi_s \cdot dB_s.$$

Proof. We may assume that $M_0 = 0$. By adding $M^i = 0$ or $\Phi^i_j = 0$ when necessary, it suffices to prove the theorem in the case $d = r$.

First of all, let $\Phi = \beta \rho \sigma^*$ be a singular value decomposition of Φ. Note that β, ρ, σ are matrix-valued processes, where β, σ are orthogonal and ρ is diagonal . It follows that

$$\gamma = \Phi \Phi^* = \beta \rho \sigma^* \sigma \rho \beta^* = \beta \rho^2 \beta^*.$$

This also gives the diagonalization of γ. Let $\alpha \triangleq \beta \rho$ and $\lambda \triangleq \theta \beta^*$, where θ is the diagonal matrix formed by replacing each nonzero element on the diagonal of ρ by its reciprocal. It is important to note that all these matrix-valued process constructed here are progressively measurable, as they are constructed from a pointwise manner.

Now define $\zeta_t \triangleq \mathrm{rank}(\gamma_t)$, and let P_{ζ_t} to be the matrix-valued process given by $(P_{\zeta_t})^i_j = 1$ if $i = j \leqslant \zeta_t$ and $(P_{\zeta_t})^i_j = 0$ otherwise. Define the stochastic integral process $N \triangleq \lambda \bullet M$. It follows that

$$d\langle N \rangle_t = \lambda \cdot d\langle M \rangle \cdot \lambda^* = \lambda \gamma \lambda^* dt = \theta \beta^* \beta \rho^2 \beta^* \beta \theta dt = P_\zeta dt.$$

Next define $X \triangleq \alpha \bullet N$. Then we have

$$d\langle X \rangle_t = \alpha P_\zeta \alpha^* dt = \beta \rho P_\zeta \rho \beta^* dt = \beta \rho^2 \beta^* dt = \gamma dt = d\langle M \rangle_t,$$

and

$$
\begin{aligned}
d\langle X, M\rangle_t &= d\langle M, X\rangle^* = \alpha\lambda d\langle M\rangle_t = \alpha\lambda\gamma dt \\
&= \beta\rho\theta\beta^*\beta\rho^2\beta^* dt = \gamma dt.
\end{aligned}
$$

Therefore, $\langle X - M\rangle = 0$, which implies that $X = M$.

To finish the proof, let $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}; \{\widetilde{\mathcal{F}}_t\})$ be an enlargement of $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ which supports a d-dimensional Brownian motion W_t independent of M. The construction of $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}; \{\widetilde{\mathcal{F}}_t\})$ is similar to the one in the proof of Theorem 5.9. Define

$$
\overline{W} \triangleq N + (\mathrm{Id} - P_\zeta) \bullet W.
$$

Then

$$
d\langle \overline{W}\rangle_t = d\langle N\rangle_t + (\mathrm{Id} - P_\zeta)dt = dt,
$$

where we have used the fact that $\langle N, W\rangle = 0$ due to independence (the same reason as in the last part of the proof of Theorem 5.9). Therefore, $\overline{W}_t$ is an $\{\widetilde{\mathcal{F}}_t\}$-Brownian motion according to Lévy's characterization theorem. Moreover, by the definition of α, we know that

$$
\alpha \bullet \overline{W} = \alpha \bullet N + (\alpha(\mathrm{Id} - P_\zeta)) \bullet W = X = M.
$$

Since $\alpha = \Phi\sigma$, we conclude that $M = \Phi \bullet (\sigma \bullet \overline{W})$. But $B \triangleq \sigma \bullet \overline{W}$ is also an $\{\widetilde{\mathcal{F}}_t\}$-Brownian motion according to Lévy's characterization theorem as σ takes values in the space of orthogonal matrices. Therefore, we arrive at the representation

$$
M_t = \int_0^t \Phi_s \cdot dB_s.
$$

□

Remark 5.13. The underlying idea of proving Theorem 5.14 is very simple. The complexity arises from the possibility that γ is degenerate. If we further assume that γ_t is positive definite everywhere, then $B \triangleq \Phi^{-1} \bullet M$ will be an $\{\mathcal{F}_t\}$-Brownian motion, and $M = \Phi \bullet B$. In particular, in this case we do not need to enlarge the underlying probability space.

5.8 The Cameron-Martin-Girsanov transformation

In Section 5.6, we have seen the notion of a random time-change. Now we study another important technique: change of measure. This technique is very useful in the study of stochastic differential equations.

It is well known that the Lebesgue measure on $\mathbb{R}^d$ is translation invariant, in the sense that given any $h \in \mathbb{R}^d$, the measure induced by the translation map $x \mapsto x + h$ is again the Lebesgue measure. The Lebesgue measure is essentially a finite dimensional object: there is no counterpart of Lebesgue measure in infinite dimensions in any obvious way.

However, a Gaussian measure is quite different: for instance, a natural infinite dimensional counterpart of a finite dimensional Gaussian measure is just the law of Brownian motion defined on the continuous path space. A natural question therefore arises: what is the invariance property for a Gaussian measure with respect to translation?

We first illustrate the motivation by doing a series of formal calculations.

Let us first consider the finite dimensional situation. Let

$$\mu(dx) = \frac{1}{(2\pi)^{d/2}} \mathrm{e}^{-\frac{|x|^2}{2}} dx$$

be the standard Gaussian measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, so that the coordinate random variables $\xi^i(x) \triangleq x^i$ define a standard Gaussian vector $\xi = (\xi^1, \cdots, \xi^d) \sim \mathcal{N}(0, \mathrm{Id})$ under the probability measure μ. Now fix $h \in \mathbb{R}^d$. Consider the translation map $T_h : \mathbb{R}^d \to \mathbb{R}^d$ defined by $T_h(x) \triangleq x + h$, and let $\mu^h \triangleq \mu \circ (T^h)^{-1}$ be the push-forward of μ by T_h. From the simple relation that for any nice test function $f : \mathbb{R}^d \to \mathbb{R}^1$,

$$\begin{aligned} \int_{\mathbb{R}^d} f(y)\mu^h(dy) &= \int_{\mathbb{R}^d} f(x+h)\mu(dx) \\ &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} f(x+h) \mathrm{e}^{-\frac{|x|^2}{2}} dx \\ &= \int_{\mathbb{R}^d} f(y) \mathrm{e}^{\langle h, y\rangle - \frac{1}{2}|h|^2} \mu(dy), \end{aligned}$$

we see that μ^h is absolutely continuous with respect to μ, and the Radon-Nikodym derivative is given by

$$\frac{d\mu^h}{d\mu} = \mathrm{e}^{\langle h, x\rangle - \frac{1}{2}|h|^2}. \tag{5.30}$$

This property is usually known as the *quasi-invariance* of Gaussian measures.

Another way of looking that this fact is the following: if we define μ^h by the formula (5.30), then $\eta \triangleq \xi - h$ is a standard Gaussian vector under μ^h, since its distribution, which is the push-forward of μ^h by the map $T_{-h} : x \mapsto x - h$, is just μ.

Now we look for the infinite dimensional counterpart of this simple observation. For simplicity, let W_0 be the space of continuous paths $w : [0,1] \to \mathbb{R}^1$ vanishing at $t = 0$, and let μ be the law of a one dimensional Brownian motion over $[0,1]$, which is a probability measure on $(W_0, \mathcal{B}(W_0))$. Define $B_t(w) \triangleq w_t$, so that B_t is a Brownian motion under μ. Now fix $h \in W_0$, which is in this case a continuous path. We assume that h has "nice" regularity properties and let us do not bother with what they are at the moment. Again consider the translation map $T_h : W_0 \to W_0$ defined by $T_h(w) = w + h$, and let μ^h be the push-forward of μ by T_h.

To understand the relationship between μ^h and μ, we need some kind of finite dimensional approximations. For each $n \geqslant 1$, consider the partition $\mathcal{P}_n : 0 = t_0 < t_1 < \cdots < t_n = 1$ of $[0,1]$ into n sub-intervals with equal length $1/n$. Given $w \in W_0$, let $w^{(n)} \in W_0$ be the piecewise linear interpolation of w over $\mathcal{P}_n$. More precisely, $w^{(n)}_{t_i} = w_{t_i}$ for $t_i \in \mathcal{P}_n$ and $w^{(n)}$ is linear on each sub-interval associated with $\mathcal{P}_n$. Given a nice test function $f: W \to \mathbb{R}^1$, we define an approximation $f^{(n)}$ of f by $f^{(n)}(w) \triangleq f(w^{(n)})$. A crucial observation is that $f^{(n)}$ depends only on the values $\{w_{t_1}, \cdots, w_{t_n}\}$. Therefore, $f^{(n)}$ is a finite dimensional function, in the sense that there exists $H: \mathbb{R}^n \to \mathbb{R}^1$, such that $f^{(n)}(w) = H(w_{t_1}, \cdots, w_{t_n})$ for all $w \in W_0$.

Now we do a similar calculation as in the finite dimensional case:

$$\begin{aligned} & \int_{W_0} f^{(n)}(w)\mu^h(dw) \\ = & \int_{W_0} f^{(n)}(w+h)\mu(dw) \\ = & \int_{W_0} H(w_{t_1}+h_{t_1}, \cdots, w_{t_n}+h_{t_n})\mu(dw) \\ = & C\int_{\mathbb{R}^n} H(x_1+h_{t_1}, \cdots, x_n+h_{t_n}) \exp\left(-\frac{1}{2}\sum_{i=1}^n \frac{|x_i - x_{i-1}|^2}{t_i - t_{i-1}}\right) dx \\ = & C\int_{\mathbb{R}^n} H(y_1, \cdots, y_n) \exp\left(\sum_{i=1}^n \frac{h_{t_i} - h_{t_{i-1}}}{t_i - t_{i-1}} \cdot (y_i - y_{i-1}) - \frac{1}{2}\sum_{i=1}^n \frac{(h_{t_i} - h_{t_{i-1}})^2}{t_i - t_{i-1}} \right. \\ & \left. -\frac{1}{2}\sum_{i=1}^n \frac{(y_i - y_{i-1})^2}{t_i - t_{i-1}}\right) dy \\ = & \int_{W_0} f^{(n)}(w) \exp\left(\sum_{i=1}^n \frac{h_{t_i} - h_{t_{i-1}}}{t_i - t_{i-1}} \cdot (w_{t_i} - w_{t_{i-1}}) - \frac{1}{2}\sum_{i=1}^n \frac{(h_{t_i} - h_{t_{i-1}})^2}{t_i - t_{i-1}}\right) \mu(dw), \end{aligned}$$

where $C \triangleq (2\pi)^{-n/2}(t_1(t_2-t_1)\cdots(t_n-t_{n-1}))^{-1/2}$. Here comes the crucial observation. If we let $n \to \infty$, it is natural to expect that $f^{(n)}(w) \to f(w)$, and also

$$\begin{aligned} \sum_{i=1}^n \frac{h_{t_i} - h_{t_{i-1}}}{t_i - t_{i-1}} \cdot (w_{t_i} - w_{t_{i-1}}) & \to \int_0^1 h_t' dB_t, \\ \sum_{i=1}^n \frac{(h_{t_i} - h_{t_{i-1}})^2}{t_i - t_{i-1}} = \sum_{i=1}^n \frac{(h_{t_i} - h_{t_{i-1}})^2}{(t_i - t_{i-1})^2} \cdot (t_i - t_{i-1}) & \to \int_0^1 (h_t')^2 dt, \end{aligned} \quad (5.31)$$

where the first limit is Itô's integral! Therefore, formally we arrive at

$$\int_{W_0} f(w)\mu^h(dw) = \int_{W_0} f(w) \exp\left(\int_0^1 h_t' dB_t - \frac{1}{2}\int_0^1 (h_t')^2 dt\right) \mu(dw),$$

which suggests that μ^h is absolutely continuous with respect to μ, and the Radon-Nikodym derivative is given by

$$\frac{d\mu^h}{d\mu} = \exp\left(\int_0^1 h_t' dB_t - \frac{1}{2}\int_0^1 (h_t')^2 dt\right). \quad (5.32)$$

Another way of looking at this fact is the following: if we define μ^h by the formula (5.32), then under the new measure μ^h, $\widetilde{B}_t \triangleq B_t - h_t = B_t - \int_0^t h_s' ds$ is a Brownian motion, since its distribution, which is the push-forward of μ^h by the map $T_{-h}: \ w \mapsto w - h$, is just μ.

The above argument outlines the philosophy of Cameron-Martin's original work. The main technical difficulty lies in verifying the convergence in (5.31) for the right class of h. Here the right regularity assumption on h is the following: h needs to be absolutely continuous and $\int_0^1 (h_t')^2 dt < \infty$. Cameron-Martin's result can be stated as follows. We refer the reader to [10] for a modern proof.

Theorem 5.15. *Let $\mathcal{H}$ be the space of absolutely continuous paths $h \in W_0$ with $\int_0^1 (h_t')^2 dt < \infty$. Then for any $h \in \mathcal{H}$, μ^h is absolutely continuous with respect to μ with Radon-Nikodym derivative given by (5.32), and $w_t - \int_0^t h_s' ds$ is a Brownian motion under μ^h. In addition, for any $h \notin \mathcal{H}$, μ^h and μ are singular to each other.*

Remark 5.14. We can see from Cameron-Martin's theorem that the infinite dimensional situation is very different from the finite dimensional one: the quasi-invariance property is true and only true along directions in $\mathcal{H}$. This space $\mathcal{H}$, which is known as the *Cameron-Martin subspace*, plays a fundamental role in the stochastic analysis on the space $(W_0, \mathcal{B}(W_0), \mu)$.

After Cameron-Martin's important work, Girsanov pushed this idea further into a more general situation. It is Girsanov's work that we will explore in details with the help of martingale methods.

Let $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ be a filtered probability space which satisfies the usual conditions, and let $B_t = (B_t^1, \cdots, B_t^d)$ be a d-dimensional $\{\mathcal{F}_t\}$-Brownian motion. Suppose that $X_t = (X_t^1, \cdots, X_t^d)$ is a stochastic process with $X^i \in L^2_{\text{loc}}(B^i)$ for each i.

Motivated from the previous discussion on Cameron-Martin's work, we define the exponential martingale

$$\mathcal{E}_t^X \triangleq \exp\left(\sum_{i=1}^d \int_0^t X_s^i dB_s^i - \frac{1}{2}\int_0^t |X_s|^2 ds\right), \quad t \geqslant 0. \tag{5.33}$$

According to Itô's formula, we have

$$\mathcal{E}_t^X = 1 + \sum_{i=1}^d \int_0^t \mathcal{E}_s^X X_s^i dB_s^i.$$

Therefore, $\mathcal{E}_t^X$ is a continuous local martingale. Now take a localization sequence $\tau_n \uparrow \infty$ of stopping times such that $\mathcal{E}_{\tau_n \wedge t}^X$ is martingale for each n, i.e.

$$\mathbb{E}[\mathcal{E}_{\tau_n \wedge t}^X | \mathcal{F}_s] = \mathcal{E}_{\tau_n \wedge s}^X.$$

Fatou's lemma then allows us to conclude that $\mathcal{E}_t^X$ is a supermartingale and $\mathbb{E}[\mathcal{E}_t^X] \leqslant \mathbb{E}[\mathcal{E}_0^X] = 1$ for all $t \geqslant 0$. In general, $\mathcal{E}_t^X$ can fail to be a martingale. However, we have the following simple fact.

Proposition 5.20. *$\mathcal{E}_t^X$ is a martingale if and only if $\mathbb{E}[\mathcal{E}_t^X] = 1$ for all $t \geqslant 0$.*

Proof. Since $\mathcal{E}_t^X$ is a supermartingale, given $s < t$, we have

$$\int_A \mathcal{E}_t^X d\mathbb{P} \leqslant \int_A \mathcal{E}_s^X d\mathbb{P}, \quad \forall A \in \mathcal{F}_s. \tag{5.34}$$

If $\mathcal{E}_t^X$ has constant expectation, then

$$\int_{A^c} \mathcal{E}_t^X d\mathbb{P} \geqslant \int_{A^c} \mathcal{E}_s^X d\mathbb{P}, \quad \forall A \in \mathcal{F}_s. \tag{5.35}$$

But (5.34) and (5.35) are true for all $A \in \mathcal{F}_s$. It follows that

$$\int_A \mathcal{E}_t^X d\mathbb{P} = \int_A \mathcal{E}_s^X d\mathbb{P}, \quad \forall A \in \mathcal{F}_s,$$

which implies the martingale property. □

Remark 5.15. In Cameron-Martin's work, given $h \in \mathcal{H}$, since $\int_0^t h'_s dB_s$ is Gaussian distributed with mean 0 and variance $\int_0^t (h'_s)^2 ds$ (c.f. Problem Sheet 5, Problem 1, (2)), we know from Proposition 5.20 that the exponential martingale

$$\mathcal{E}_t^h \triangleq \exp\left(\int_0^t (h'_s) dB_s - \frac{1}{2}\int_0^t (h'_s)^2 ds\right)$$

is indeed a martingale.

Now we make the following assumption exclusively and explore its consequences. At the end of this subsection, we will establish a useful condition which verifies the assumption.

Assumption 5.1. *$\{\mathcal{E}_t^X, \mathcal{F}_t\}$ is a martingale.*

As in Cameron-Martin's formula (5.32), for each given $T > 0$, we define

$$\widetilde{\mathbb{P}}_T(A) \triangleq \mathbb{E}[\mathbf{1}_A \mathcal{E}_T^X], \quad A \in \mathcal{F}_T.$$

According to Assumption 5.1, $\widetilde{\mathbb{P}}_T$ is a probability measure on $(\Omega, \mathcal{F}_T)$ which is obviously equivalent to $\mathbb{P}$.

The following lemma tells us how to compute conditional expectations under $\widetilde{\mathbb{P}}_T$.

Lemma 5.4. *Let $0 \leqslant s \leqslant t \leqslant T$. Suppose that Y is an $\{\mathcal{F}_t\}$-measurable random variable which is integrable with respect to $\widetilde{\mathbb{P}}_T$. Then we have:*

$$\widetilde{\mathbb{E}}_T[Y|\mathcal{F}_s] = \frac{1}{\mathcal{E}_s^X}\mathbb{E}[Y\mathcal{E}_t^X|\mathcal{F}_s], \quad \mathbb{P} \text{ and } \widetilde{\mathbb{P}}_T - \text{a.s.},$$

where $\widetilde{\mathbb{E}}_T$ is the expectation under $\widetilde{\mathbb{P}}_T$.

Proof. For any $A \in \mathcal{F}_s$, by the martingale property of $\mathcal{E}^X$ under $\mathbb{P}$, we have

$$\begin{aligned}
\widetilde{\mathbb{E}}_T[Y\mathbf{1}_A] &= \mathbb{E}\left[Y\mathbf{1}_A\mathcal{E}_T^X\right] = \mathbb{E}\left[Y\mathbf{1}_A\mathcal{E}_t^X\right] \\
&= \mathbb{E}\left[\mathbf{1}_A\mathbb{E}[Y\mathcal{E}_t^X|\mathcal{F}_s]\right] = \mathbb{E}\left[\frac{\mathcal{E}_T^X}{\mathcal{E}_s^X}\mathbf{1}_A\mathbb{E}[Y\mathcal{E}_t^X|\mathcal{F}_s]\right] \\
&= \widetilde{\mathbb{E}}_T\left[\frac{1}{\mathcal{E}_s^X}\mathbf{1}_A\mathbb{E}[Y\mathcal{E}_t^X|\mathcal{F}_s]\right].
\end{aligned}$$

Therefore, the result follows. □

With the help of Lemma 5.4, we are able to understand the relationship between continuous local martingales under $\mathbb{P}$ and $\widetilde{\mathbb{P}}_T$. Given $T > 0$, we use the notation $\mathcal{M}_{0;T}^{\text{loc}}$ (respectively, $\widetilde{\mathcal{M}}_{0;T}^{\text{loc}}$) to denote the space of continuous local martingales $\{M_t, \mathcal{F}_t : 0 \leqslant t \leqslant T\}$ on $(\Omega, \mathcal{F}_T, \mathbb{P})$ (respectively, on $(\Omega, \mathcal{F}_T, \widetilde{\mathbb{P}}_T)$) which vanishes at $t = 0$. The meaning of a local martingale defined on a finite interval $[0, T]$ should be clear to the reader.

Theorem 5.16. *For $T > 0$, the transformation map*

$$\begin{aligned}
G_T : \mathcal{M}_{0;T}^{\text{loc}} &\to \widetilde{\mathcal{M}}_{0;T}^{\text{loc}}, \\
M_t &\mapsto \widetilde{M}_t \triangleq M_t - \sum_{i=1}^d \int_0^t X_s^i d\langle M, B^i\rangle_s,
\end{aligned}$$

is a linear isomorphism and respects the bracket, i.e. $\langle \widetilde{M}, \widetilde{N}\rangle = \langle M, N\rangle$ for all $M, N \in \mathcal{M}_{0;T}^{\text{loc}}$, where the bracket processes are computed under the appropriate probability measures.

Proof. We first show that $\widetilde{M} = G_T(M) \in \widetilde{\mathcal{M}}_{0;T}^{\text{loc}}$ for $M \in \mathcal{M}_{0;T}^{\text{loc}}$. By localization, we may assume that all involved local martingales and bounded variation processes are uniformly bounded. By the definition of $\widetilde{M}$ and the integration by parts formula (c.f. Proposition5.15), we have

$$\widetilde{M}_t\mathcal{E}_t^X = \int_0^t \mathcal{E}_s^X dM_s + \sum_{i=1}^d \int_0^t \widetilde{M}_s\mathcal{E}_s^X X_s^i dB_s^i,$$

which shows that $\widetilde{M}_t\mathcal{E}_t^X$ is a martingale under $\mathbb{P}$. Therefore, by Lemma 5.4,

$$\widetilde{\mathbb{E}}_T[\widetilde{M}_t|\mathcal{F}_s] = \frac{1}{\mathcal{E}_s^X}\mathbb{E}[\widetilde{M}_t\mathcal{E}_t^X|\mathcal{F}_s] = \widetilde{M}_s,$$

showing that $\widetilde{M}_t$ is a martingale under $\widetilde{\mathbb{P}}_T$. This proves that G_T maps $\mathcal{M}_{0;T}^{\text{loc}}$ to $\mathcal{M}_{0;T}^{\text{loc}}$. It is apparent that G_T is linear.

Now we show that G_T respects the bracket. Indeed, again localizing in the bounded setting, exactly the same but longer calculation based on the integration by parts formula shows that $(\widetilde{M}_t\widetilde{N}_t - \langle M, N\rangle_t)\mathcal{E}_t^X$ is a linear combination of stochastic integrals, which proves that it is a martingale under $\mathbb{P}$. Therefore, Lemma 5.4 again shows that $\widetilde{M}_t\widetilde{N}_t - \langle M, N\rangle_t$ is a martingale under $\widetilde{\mathbb{P}}_T$. This proves that $\langle\widetilde{M}, \widetilde{N}\rangle = \langle M, N\rangle$.

In particular, G_T is injective since

$$\widetilde{M} = 0 \implies \langle\widetilde{M}\rangle = \langle M\rangle = 0 \implies M = 0.$$

Finally, we show that G_T is surjective. Let $\widetilde{M} \in \widetilde{\mathcal{M}}_{0;T}^{\mathrm{loc}}$. If $\widetilde{M}$ is bounded, by Lemma 5.4, we know that

$$\mathbb{E}[\widetilde{M}_t\mathcal{E}_t^X|\mathcal{F}_s] = \mathcal{E}_s^X\widetilde{\mathbb{E}}_T[\widetilde{M}_t|\mathcal{F}_s] = \widetilde{M}_s\mathcal{E}_s^X.$$

Therefore, $\widetilde{M}_t\mathcal{E}_t^X$ is a martingale under $\mathbb{P}$. Since $\widetilde{M}_t = (\widetilde{M}_t\mathcal{E}_t^X)/\mathcal{E}_t^X$, after removing the localization, Itô's formula shows that $\widetilde{M}_t$ is a continuous semimartingale under $\mathbb{P}$. Therefore, we may assume that under $\mathbb{P}$, $\widetilde{M}_t = M_t + A_t$ for some $M \in \mathcal{M}_{0;T}^{\mathrm{loc}}$ and some bounded variation process A. Now define $\overline{M} \triangleq G_T(M) \in \widetilde{\mathcal{M}}_{0;T}^{\mathrm{loc}}$. It follows that

$$\widetilde{M}_t - \overline{M}_t = A_t + \sum_{i=1}^{d}\int_0^t X_s^i d\langle M, B^i\rangle_s.$$

This shows that $\widetilde{M} - \overline{M}$ is a bounded variation process. But $\widetilde{M} - \overline{M} \in \widetilde{\mathcal{M}}_{0;T}^{\mathrm{loc}}$. Therefore, $\widetilde{M} = \overline{M} = G_T(M)$, which shows that G_T is also surjective. □

From the characterization of stochastic integrals, a direct corollary of Theorem 5.16 is that the transformation map G_T respects stochastic integration.

Corollary 5.2. *Let $M \in \mathcal{M}_{0;T}^{\mathrm{loc}}$, and let Φ_t be a progressively measurable process on $[0, T]$ such that $\mathbb{P}\left(\int_0^T \Phi_s^2 d\langle M\rangle_s < \infty\right) = 1$. Then $\widetilde{I^M(\Phi)} = I^{\widetilde{M}}(\Phi)$.*

Proof. Since $\langle\widetilde{M}\rangle = \langle M\rangle$, we have $\widetilde{\mathbb{P}}_T\left(\int_0^T \Phi_s^2 d\langle\widetilde{M}\rangle_s < \infty\right) = 1$. The last claim follows from the fact that

$$\begin{aligned}\left\langle\widetilde{I^M(\Phi)}, \widetilde{N}\right\rangle &= \langle I^M(\Phi), N\rangle = \Phi \bullet \langle M, N\rangle \\ &= \Phi \bullet \langle\widetilde{M}, \widetilde{N}\rangle = \left\langle I^{\widetilde{M}}(\Phi), \widetilde{N}\right\rangle, \quad \forall \widetilde{N} \in \mathcal{M}_{0;T}^{\mathrm{loc}}.\end{aligned}$$

□

Another direct consequence of Theorem 5.16 is the following result. This is the original *Girsanov's theorem*.

Theorem 5.17. *Define the process $\widetilde{B}_t = (\widetilde{B}_t^1, \cdots, \widetilde{B}_t^d)$ by*

$$\widetilde{B}_t^i \triangleq B_t^i - \int_0^t X_s^i ds, \quad t \geqslant 0,\ 1 \leqslant i \leqslant d. \tag{5.36}$$

Then for each $T > 0$, the process $\{\widetilde{B}_t, \mathcal{F}_t :\ 0 \leqslant t \leqslant T\}$ is a d-dimensional Brownian motion on $(\Omega, \mathcal{F}_T, \widetilde{\mathbb{P}}_T)$.

Proof. From Theorem 5.16, we know that $\widetilde{B}^i = G_T(B^i) \in \widetilde{\mathcal{M}}_{0;T}^{\mathrm{loc}}$ for each $1 \leqslant i \leqslant d$. Moreover, we have

$$\langle \widetilde{B}^i, \widetilde{B}^j \rangle_t = \langle B^i, B^j \rangle_t = \delta_{ij} dt, \quad t \in [0, T].$$

Therefore, according to Lévy's characterization theorem, we conclude that $\widetilde{B}_t$ is an $\{\mathcal{F}_t\}$-Brownian motion on $[0, T]$ under $\widetilde{\mathbb{P}}_T$. □

The careful reader might ask if there exists a single probability measure $\widetilde{\mathbb{P}}$ on $(\Omega, \mathcal{F}_\infty)$, such that $\widetilde{\mathbb{P}} = \widetilde{\mathbb{P}}_T$ on $\mathcal{F}_T$ for every $T \geqslant 0$. This is not true in general. Indeed, if such $\widetilde{\mathbb{P}}$ exists, then $\widetilde{\mathbb{P}}$ is absolutely continuous with respect to $\mathbb{P}$ on $\mathcal{F}_\infty$ ($A \in \mathcal{F}_\infty$, $\mathbb{P}(A) = 0$ implies $A \in \mathcal{F}_0$, and hence $\widetilde{\mathbb{P}}(A) = \widetilde{\mathbb{P}}_0(A) = 0$). In this case, if we let $\xi \triangleq d\widetilde{\mathbb{P}}/d\mathbb{P}$ on $(\Omega, \mathcal{F}_\infty)$, then it is not hard to see that $\mathcal{E}_t^X = \mathbb{E}[\xi|\mathcal{F}_t]$ so that $\mathcal{E}_t^X$ is uniformly integrable. Certainly this is too strong to assume in general (for instance, the martingale $\mathrm{e}^{B_t - \frac{1}{2}t^2}$ is not uniformly integrable). Conversely, if $\mathcal{E}_t^X$ is uniformly integrable, then $\mathcal{E}_t^X = \mathbb{E}[\xi|\mathcal{F}_t]$ for $\xi \triangleq \lim_{t\to\infty} \mathcal{E}_t^X \in \mathcal{F}_\infty$. If we define $\widetilde{\mathbb{P}}(A) = \int_A \xi d\mathbb{P}$ for $A \in \mathcal{F}_\infty$, then $\widetilde{\mathbb{P}} = \widetilde{\mathbb{P}}_T$ on $\mathcal{F}_T$ for every $T \geqslant 0$. Therefore, we see that an extension $\widetilde{\mathbb{P}}$ of $\{\widetilde{\mathbb{P}}_T : \ T \geqslant 0\}$ exists on $(\Omega, \mathcal{F}_\infty)$ if and only if $\mathcal{E}_t^X$ is uniformly integrable, in which case $\widetilde{\mathbb{P}}$ is absolutely continuous with respect to $\mathbb{P}$ on $\mathcal{F}_\infty$. This is crucially related to the fact that we assume $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets, which is part of the usual conditions. Also note that in this case, the process $\widetilde{B}_t$ defined by (5.36) is an $\{\mathcal{F}_t\}$-Brownian motion on $[0, \infty)$ under $\widetilde{\mathbb{P}}$.

However, if we only consider the natural filtration and do not take its usual augmentation, then we do have such an extension $\widetilde{\mathbb{P}}$ even without the uniform integrability of $\mathcal{E}_t^X$, and the process $\widetilde{B}_t$ is a Brownian motion on $[0, \infty)$ under $\widetilde{\mathbb{P}}$.

To be more precise, let us consider the continuous path space $(W^d, \mathcal{B}(W^d), \mu)$, where μ is the d-dimensional Wiener measure. Let $B_t(w) \triangleq w_t$ be the coordinate process and let $\{\mathcal{G}_t^B\}$ be the natural filtration of B_t. It follows that $\{B_t, \mathcal{G}_t^B\}$ is a Brownian motion under μ. Now consider a $\{\mathcal{G}_t^B\}$-progressively measurable process X_t which satisfies $\int_0^t X_s^2(\omega) ds < \infty$ for every (t, ω). Define the exponential martingale $\mathcal{E}_t^X$ by (5.33) and assume that it is a martingale (technically speaking, in order to make sense of the stochastic integrals involved, we need to define $\mathcal{E}_t^X$ with respect to the augmented natural filtration $\{\mathcal{F}_t^B\}$, and assume that $\mathcal{E}_t^X$ is a martingale under this filtration). Then

$$\widetilde{\mathbb{P}}_T(A) \triangleq \int_A \mathcal{E}_T^X d\mu, \quad A \in \mathcal{G}_T^B,$$

defines a compatible family of probability measures. Therefore, they extend to a probability measure on the π-system $\cup_{T \geqslant 0} \mathcal{G}_T^B$. By verifying the conditions in Carathéodory's extension theorem, we get a single probability measure $\widetilde{\mathbb{P}}$ on $\mathcal{G}_\infty^B = \mathcal{B}(W^d)$ which extends those $\widetilde{\mathbb{P}}_T$'s. Apparently, $\widetilde{B}_t \triangleq B_t - \int_0^t X_s ds$ is $\{\mathcal{G}_t^B\}$-adapted and it is indeed a $\{\mathcal{G}_t^B\}$-Brownian motion on $[0, \infty)$ under $\widetilde{\mathbb{P}}$.

In general, althought $\widetilde{\mathbb{P}}$ is absolutely continuous with respect to μ when restricted on each $\mathcal{G}_T^B$ (because $\widetilde{\mathbb{P}}_T$ is by definition), it can fail to be so on $\mathcal{G}_\infty^B$. A simple example is the following: consider $X \equiv c \neq 0$. Then under the new probability measure $\widetilde{\mathbb{P}}$, $\widetilde{B}_t = B_t - ct$

is a Brownian motion and therefore $B_t = \widetilde{B}_t + ct$ is a Brownian motion with drift c. Let

$$\Lambda \triangleq \{w \in W^d : \lim_{t\to\infty} w_t/t = c\} \in \mathcal{G}^B_\infty.$$

Then $\widetilde{\mathbb{P}}(\Lambda) = 1$ but $\mu(\Lambda) = 0$.

We leave the reader to think about these details.

To finish this part, we give a useful condition, known as *Novikov's condition*, under which Assumption 5.1 holds.

Theorem 5.18. *Let $M \in \mathcal{M}_0^{\mathrm{loc}}$. Suppose that*

$$\mathbb{E}\left[\mathrm{e}^{\frac{1}{2}\langle M\rangle_t}\right] < \infty, \quad \forall t \geqslant 0.$$

Then

$$\mathcal{E}^M_t \triangleq \mathrm{e}^{M_t - \frac{1}{2}\langle M\rangle_t}, \quad t \geqslant 0,$$

is an $\{\mathcal{F}_t\}$-martingale.

The idea of the proof is not hard: we try to use the Dambis-Dubins-Schwarz theorem, which tells us that $M_t = B_{\langle M\rangle_t}$ for a Brownian motion possibly defined on some enlarged probability space. Since $\mathrm{e}^{B_s - \frac{1}{2}s}$ is obviously a martingale and $\langle M\rangle_t$ is a stopping time with respect to the relevant filtration, by applying the optional sampling theorem formally, it is entirely reasonable to expect that

$$\mathbb{E}\left[\mathrm{e}^{M_t - \frac{1}{2}\langle M\rangle_t}\right] = \mathbb{E}\left[\mathrm{e}^{B_{\langle M\rangle_t} - \frac{1}{2}\langle M\rangle_t}\right] = 1.$$

Therefore, the result follows according to Proposition 5.20. To make this idea work, we need to overcome the issue of integrability by a technical trick.

Proof of Theorem 5.18. According to the generalized Dambis-Dubins-Schwarz theorem (c.f. Theorem 5.9), there exists an $\{\widetilde{\mathcal{F}}_t\}$-Brownian motion B_t, possibly defined on some enlarged space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}; \{\widetilde{\mathcal{F}}_t\})$, such that $M_t = B_{\langle M\rangle_t}$. Moreover, $\langle M\rangle_t$ is an $\{\widetilde{\mathcal{F}}_s\}$-stopping time for every $t \geqslant 0$.

For each $b < 0$, define

$$\tau_b \triangleq \inf\{s \geqslant 0 : B_s - s = b\}.$$

Note that in Problem Sheet 4, Problem 6, (2), we have computed the marginal distribution of the running maximum process for the Brownian motion with drift. From that formula it is not hard to see that the density of τ_b is given by

$$\mathbb{P}(\tau_b \in ds) = \frac{|b|}{\sqrt{2\pi s^3}} \mathrm{e}^{-\frac{(b+s)^2}{2s}} ds, \quad t > 0.$$

In particular,

$$\mathbb{E}\left[\mathrm{e}^{\frac{1}{2}\tau_b}\right] = \int_0^\infty \mathrm{e}^{\frac{1}{2}s} \cdot \frac{|b|}{\sqrt{2\pi s^3}} \mathrm{e}^{-\frac{(b+s)^2}{2s}} ds = \mathrm{e}^{-b}, \tag{5.37}$$

where we applied the change of variables $u = |b|/\sqrt{s}$.

Apparently, $Z_s \triangleq \mathrm{e}^{B_s - \frac{1}{2}s}$ is an $\{\widetilde{\mathcal{F}}_s\}$-martingale. Therefore, $Z_s^{\tau_b}$ is also an $\{\widetilde{\mathcal{F}}_s\}$-martingale. Moreover, since $\tau_b < \infty$ almost surely,

$$Z_\infty^{\tau_b} = \mathrm{e}^{B_{\tau_b} - \frac{1}{2}\tau_b} = \mathrm{e}^{\frac{1}{2}\tau_b + b}.$$

Now on the one hand, Fatou's lemma tells us that $\{Z_s^{\tau_b}, \widetilde{\mathcal{F}}_s : 0 \leqslant s \leqslant \infty\}$ is a supermartingale with a last element. On the other hand, (5.37) tells us that $\mathbb{E}[Z_\infty^{\tau_b}] = \mathbb{E}[Z_s^{\tau_b}] = 1$ for all s. Therefore, similar to the proof of Proposition 5.20, we know that $\{Z_s^{\tau_b}, \widetilde{\mathcal{F}}_s : 0 \leqslant s \leqslant \infty\}$ is a martingale with a last element. This allows us to use the optional sampling theorem to conclude that

$$\begin{aligned} \mathbb{E}\left[Z_{\langle M\rangle_t}^{\tau_b}\right] &= \mathbb{E}\left[\mathrm{e}^{B_{\tau_b \wedge \langle M\rangle_t} - \frac{1}{2}\tau_b \wedge \langle M\rangle_t}\right] = 1 \\ &= \mathbb{E}\left[\mathbf{1}_{\{\langle M\rangle_t \geqslant \tau_b\}} \mathrm{e}^{\frac{1}{2}\tau_b + b}\right] + \mathbb{E}\left[\mathbf{1}_{\{\langle M\rangle_t < \tau_b\}} \mathrm{e}^{M_t - \frac{1}{2}\langle M\rangle_t}\right]. \end{aligned}$$

As $b \to -\infty$, the first term goes to zero by the dominated convergence theorem, since the integrand is controlled by $\mathrm{e}^b \mathrm{e}^{\frac{1}{2}\langle M\rangle_t}$ (note that $\mathbb{E}[\mathrm{e}^{\frac{1}{2}\langle M\rangle_t}] < \infty$ according to the assumption) and $\tau_b \to \infty$. Therefore,

$$\mathbb{E}\left[\mathrm{e}^{M_t - \frac{1}{2}\langle M\rangle_t}\right] = 1.$$

As this is true for all t, according to Proposition 5.20, we conclude that $\mathcal{E}_t^M$ is an $\{\mathcal{F}_t\}$-martingale.

□

Combing back to the setting of the Cameron-Martin-Girsanov theorem, we have the following direct corollary.

Corollary 5.3. *Suppose that $X^i \in L^2_{\mathrm{loc}}(B^i)$ for $i = 1, \cdots, d$. Suppose that*

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^t |X_s|^2 ds\right)\right] < \infty, \quad \forall t \geqslant 0.$$

Then the exponential martingale $\mathcal{E}_t^X$ defined by (5.33) is indeed a martingale.

5.9 Local times for continuous semimartingales

From Itô's formula, we know that the space of continuous semimartingales is stable under composition by C^2-functions. Now a natural question is: what happens if the function fails to be in C^2?

Let us consider the simplest case: $f(x) = |x|$. Then $f'(x) = \mathrm{sgn}(x)$ and $f''(x) = 2\delta_0(x)$, where δ_0 is the Dirac δ-function at 0. Applying Itô's formula for the one dimensional Brownian motion and f in a formal way, we have

$$|B_t| = \int_0^t \mathrm{sgn}(B_s) dB_s + \int_0^t \delta_0(B_s) ds.$$

Heuristically, $\int_0^t \delta_0(B_s)ds$ measures the "amount of time" before t that the Brownian motion is at the zero level. Of course this is not $m(\{s \in [0,t] : B_s = 0\})$ (m is the Lebesgue measure), because level sets of Brownian motion are Lebesgue null sets with probability one. More precise, the term $\int_0^t \delta_0(B_s)ds$ should be understood as

$$\int_0^t \delta_0(B_s)ds = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} m(\{s \in [0,t] : |B_s| < \varepsilon\}), \tag{5.38}$$

so it measures some kind of occupation density at the zero level.

This motivates the definition of a local time, and with which we can extend Itô's formula to functions with singularities. The theory of local times for Brownian motion is a very rich subject, and it leads to a large amount of deep distributional properties related to the Brownian motion. Here we only introduce the basic theory for local times of general continuous semimartingales, and we will not touch those deep computational aspects.

We start with the following result. Let $X_t = X_0 + M_t + A_t$ be a continuous semimartingale.

Theorem 5.19. *Let f be a convex function on $\mathbb{R}^1$. Then there exists a unique $\{\mathcal{F}_t\}$-adapted process A_t^f with continuous, increasing sample paths vanishing at $t = 0$, such that*

$$f(X_t) = f(X_0) + \int_0^t f'_-(X_s)dX_s + A_t^f,$$

where f'_- is the left derivative of f. In particular, $f(X_t)$ is a continuous semimartingale.

Proof. Let $\rho \in C^\infty(\mathbb{R}^1)$ be a non-negative function with compact support on $(-\infty, 0]$ and $\int_{\mathbb{R}^1} \rho(y)dy = 1$. We can think of ρ as a mollifier. For each $n \geqslant 1$, define $\rho_n(y) = n\rho(ny)$ and $f_n(x) \triangleq \int_{\mathbb{R}^1} f(x+y)\rho_n(y)dy$. Then $f_n \in C^\infty(\mathbb{R}^1)$ and $f_n(x) \to f(x)$ for every $x \in \mathbb{R}^1$. Moreover, since f is convex, it is locally Lipschitz. Therefore, f' exists almost everywhere and f'_- is locally bounded. By the dominated convergence theorem, we have

$$f'_n(x) = \int_{\mathbb{R}^1} f'_-(x+y)\rho_n(y)dy = \int_{\mathbb{R}^1} f'_-\left(x + \frac{z}{n}\right)\rho(z)dz.$$

But we know that for a convex function f, f'_- is left continuous. As ρ is supported on $(-\infty, 0]$, we conclude that $f'_n(x) \to f'_-(x)$ for every $x \in \mathbb{R}^1$.

Now we define

$$\tau_m \triangleq \begin{cases} 0, & |X_0| \geqslant m; \\ \inf\{t \geqslant 0 : |X_t| > m\}, & |X_0| < m, \end{cases}$$

and $X_t^{(m)} \triangleq X_0 \mathbf{1}_{\{|X_0| < m\}} + M_t^{\tau_m} + A_t^{\tau_m}$ in the same way as in the proof of Itô's formula. Then each $X^{(m)}$ is a bounded continuous semimartingale. By applying Itô's formula to $X^{(m)}$ and the function f_n, we have

$$f_n\left(X_t^{(m)}\right) = f_n\left(X_0^{(m)}\right) + \int_0^t f'_n\left(X_s^{(m)}\right)dX_s^{(m)} + A_t^{n,m},$$

where $A_t^{n,m} \triangleq \frac{1}{2}\int_0^t f_n''\left(X_s^{(m)}\right) d\left\langle X^{(m)}\right\rangle_s$. If we let $n \to \infty$, according to the stochastic and ordinary dominated convergence theorems (c.f. Proposition 5.14), we conclude that

$$f\left(X_t^{(m)}\right) = f\left(X_0^{(m)}\right) + \int_0^t f_-'\left(X_s^{(m)}\right) dX_s^{(m)} + A_t^m,$$

where $A_t^m \triangleq \lim_{n\to\infty} A_t^{n,m}$ which has to exist. In addition, as $\{\tau_m > 0\} \uparrow \Omega$ and $X^{(m)} = X^{\tau_m}$ on $\{\tau_m > 0\}$, by letting $m \to \infty$, we arrive that

$$f(X_t) = f(X_0) + \int_0^t f_-'(X_s)dX_s + A_t,$$

where $A_t \triangleq \lim_{m\to\infty} A_t^m$ which also has to exist. Since f_n is convex, we know that $f_n'' \geqslant 0$. Therefore, $A_t^{n,m}$ is increasing in t for every n, m. This implies that A_t is increasing in t. Therefore, we can simply define

$$A_t^f \triangleq f(X_t) - f(X_0) - \int_0^t f_-'(X_s)dX_s, \tag{5.39}$$

which is continuous and has to be a modification of A_t. A_t^f will be the desired process, and uniqueness is obvious as it has to be given by the formula (5.39). □

The reader might think that Theorem 5.19 is very general and the increasing process A_t^f can depend on f in some complicated way. In fact, this is not true. The process A_t^f can be written down in a very explicit way in terms of the local time of X which we are going to define now.

We define $\mathrm{sgn}(x) = 1$ if $x > 0$ and and $\mathrm{sgn}(x) = -1$ if $x \leqslant 0$. If $f(x) = |x|$, then $f_-'(x) = \mathrm{sgn}(x)$.

Theorem 5.20 (Tanaka's formula). *For any real number $a \in \mathbb{R}^1$, there exists a unique $\{\mathcal{F}_t\}$-adapted process L_t^a with continuous, increasing sample paths vanishing at $t = 0$, such that*

$$\begin{aligned}
|X_t - a| &= |X_0 - a| + \int_0^t \mathrm{sgn}(X_s - a)dX_s + L_t^a, \\
(X_t - a)^+ &= (X_0 - a)^+ + \int_0^t \mathbf{1}_{\{X_s > a\}}dX_s + \frac{1}{2}L_t^a, \\
(X_t - a)^- &= (X_0 - a)^- - \int_0^t \mathbf{1}_{\{X_s \leqslant a\}}dX_s + \frac{1}{2}L_t^a.
\end{aligned}$$

In particular, $|X_t - a|$, $(X_t - a)^\pm$ are all continuous semimartingales.

Proof. We apply Theorem 5.19 for the function $f(x) = |x - a|$ and define $L_t^a \triangleq A_t^f$ in the theorem. Then the first identity holds. Let B_t and C_t be the increasing processes arising from Theorem 5.19 applied to the functions $(x-a)^{\pm}$ respectively, i.e.

$$\begin{aligned} (X_t-a)^+ &= (X_0-a)^+ + \int_0^t \mathbf{1}_{\{X_s>a\}}dX_s + B_t, \\ (X_t-a)^- &= (X_0-a)^- - \int_0^t \mathbf{1}_{\{X_s\leqslant a\}}dX_s + C_t. \end{aligned}$$

Adding the two identities gives $B_t+C_t=L_t^a$, while subtracting them gives $B_t-C_t=0$ as

$$X_t = X_0 + \int_0^t dX_s.$$

Therefore, $B_t=C_t=L_t^a/2$. □

Definition 5.14. The process $\{L_t^a : t\geqslant 0\}$ is called the *local time at* a of the continuous semimartingale X.

Example 5.2. Let B_t be a one dimensional Brownian motion. Then the first identity in Tanaka's formula gives the Doob-Meyer decomposition for the submartingale $|B_t-a|$, where the corresponding increasing process is the local time at a, and the martingale part is $|B_0-a|+\int_0^t \operatorname{sgn}(B_s-a)dB_s$, which interestingly, is a Brownian motion starting at $|B_0-a|$ according to Lévy's characterization theorem.

Since L_t^a is increasing in t, it defines a (random) measure dL^a on $[0,\infty)$. The first property of L_t^a is that the random measure dL^a is almost surely carried by the set $\Lambda_a \triangleq \{t\geqslant 0 : X_t=a\}$.

Proposition 5.21. *With probability one,* $dL^a(\Lambda_a^c)=0$.

Proof. By applying Itô's formula to the continuous semimartingale $|X_t-a|$ given by the first identity of Tanaka's formula and the function $f(x)=x^2$, we have

$$\begin{aligned} (X_t-a)^2 &= (X_0-a)^2 + 2\int_0^t |X_s-a|\cdot\operatorname{sgn}(X_s-a)dX_s \\ &\quad +2\int_0^t |X_s-a|dL_s^a + \langle X\rangle_t \\ &= (X_0-a)^2 + 2\int_0^t (X_s-a)dX_s + 2\int_0^t |X_s-a|dL_s^a + \langle X\rangle_t. \end{aligned}$$

On the other hand, Itô's formula applied to X_t-a and the same function $f(x)=x^2$ gives that

$$(X_t-a)^2 = (X_0-a)^2 + 2\int_0^t (X_s-a)dX_s + \langle X\rangle_t.$$

Therefore, $\int_0^t |X_s-a|dL_s^a=0$ for all $t\geqslant 0$. This implies that $dL^a(\Lambda_a^c)=0$. □

So far the local time process is defined for each given $a \in \mathbb{R}^1$. In order to obtain more interesting results from the analysis of local times, we should first look for better versions of L_t^a as a process in the pair (a,t). At the very least, we should expect a jointly measurable version of L_t^a. This is the content of the next result.

Proposition 5.22. *There exists a $\mathcal{B}(\mathbb{R}^1)\otimes\mathcal{B}([0,\infty))\otimes\mathcal{F}$-measurable $\widetilde{L}: (a,t,\omega)\mapsto \widetilde{L}_t^a(\omega)$, such that for every $a\in\mathbb{R}^1$, the processes $\{\widetilde{L}_t^a: t\geqslant 0\}$ and $\{L_t^a: t\geqslant 0\}$ are indistinguishable.*

Proof. This is a direct consequence of the stochastic Fubini's theorem (c.f. Problem Sheet 5, Problem 3). □

With this jointly measurable version of local time process (which is still denoted as L_t^a), we are able to prove the following so-called *Itô-Tanaka's formula*. This result gives an explicit formula for the process A_t^f arising from Theorem 5.19 in terms of the local time L_t^a.

Theorem 5.21. *Let f be a convex function on $\mathbb{R}^1$ and let X_t be a continuous semimartingale. Then*

$$f(X_t)=f(X_0)+\int_0^t f'_-(X_s)dX_s+\frac{1}{2}\int_{\mathbb{R}^1}L_t^a\mu(da),$$

where μ is the second derivative measure of f on $(\mathbb{R}^1,\mathcal{B}(\mathbb{R}^1))$ induced by $\mu([a,b))\triangleq f'_-(b)-f'_-(a)$ for $a<b$. In particular, $f(X_t)$ is a continuous semimartingale.

Proof. The main idea is to represent a convex function in some more explicit way. This part involves some notions from generalized functions.

First assume that μ is compactly supported. Define the convex function

$$g(x)\triangleq\frac{1}{2}\int_{\mathbb{R}^1}|x-a|\mu(da),\quad x\in\mathbb{R}^1. \tag{5.40}$$

We claim that $f(x)-g(x)=\alpha x+\beta$ for some $\alpha,\beta\in\mathbb{R}^1$. To this end, it suffices to show that $\mu=g''$ in the sense of distributions. Let $\varphi\in C_c^\infty(\mathbb{R}^1)$ be a smooth function with compact support. Then

$$\begin{aligned}
T_{g''}(\varphi) &= -\int_{\mathbb{R}^1}g'\varphi'dx=\int_{\mathbb{R}^1}g\varphi''dx\\
&= \frac{1}{2}\int_{\mathbb{R}^1}\left(\int_{\mathbb{R}^1}|x-a|\mu(da)\right)\varphi''(x)dx\\
&= \frac{1}{2}\int_{\mathbb{R}^1}\mu(da)\int_{\mathbb{R}^1}2\delta_a(x)\varphi(x)dx\\
&= \int_{\mathbb{R}^1}\varphi(a)\mu(da),
\end{aligned}$$

where we have used the fact that $|x-a|'' = 2\delta_a(x)$ in the sense of distributions. Therefore, the claim holds. Since the theorem is apparently true for any affine function $\alpha x+\beta$ (in which case $\mu = 0$), it remains to show that it is true for g given by (5.40).

Integrating the first identity of Tanaka's formula with respect to μ and applying the stochastic Fubini's theorem (c.f. Problem Sheet 5, Problem 3), we have

$$\begin{aligned} g(X_t) &= g(X_0) + \int_0^t \left(\frac{1}{2}\int_{\mathbb{R}^1} \mathrm{sgn}(X_s - a)\mu(da)\right) dX_s + \frac{1}{2}\int_{\mathbb{R}^1} L_t^a \mu(da) \\ &= g(X_0) + \int_0^t g'_-(X_s) dX_s + \frac{1}{2}\int_{\mathbb{R}^1} L_t^a \mu(da). \end{aligned}$$

Therefore, the theorem holds for g.

In general, if μ is not compactly supported, we define f_n to be a convex function such that $f_n = f$ on $[-n, n]$ and its second derivative measure μ_n is compactly supported on $[-n, n]$. By stopping along a sequence τ_n of stopping times, we then localize X_t inside $[-n, n]$ in the same way as in the proofs of Itô's formula and Theorem 5.19. It follows that the theorem holds for f on each $[0, \tau_n]$ provided $\{\tau_n > 0\}$, and therefore holds globally by letting $n \to \infty$. □

Itô-Tanaka's formula immediately gives the following so-called *occupation times formula*.

Corollary 5.4. *There exists a $\mathbb{P}$-null set outside which we have*

$$\int_0^t \Phi(X_s) d\langle X\rangle_s = \int_{\mathbb{R}^1} \Phi(x) L_t^x dx \tag{5.41}$$

for all $t \geqslant 0$ and all non-negative Borel measurable functions Φ.

Proof. Let $\Phi \in C_c(\mathbb{R}^1)$ be a non-negative continuous function with compact support. Let $f \in C^2(\mathbb{R}^1)$ be a convex function whose second derivative is Φ. By comparing Itô's formula and Itô-Tanaka's formula for f, we conclude that outside a $\mathbb{P}$-null set N_Φ,

$$\int_0^t \Phi(X_s) d\langle X\rangle_s = \int_{\mathbb{R}^1} \Phi(x) L_t^x dx, \quad \forall t \geqslant 0. \tag{5.42}$$

To obtain a single $\mathbb{P}$-null set independent of Φ, let $\mathcal{H} = \{\Phi_{q_1,q_2,q_3,q_4} : q_1 < q_2 < q_3 < q_4 \in \mathbb{Q}\}$ be the countable family of functions defined by

$$\Phi_{q_1,q_2,q_3,q_4}(x) \triangleq \begin{cases} 0, & x \leqslant q_1 \text{ or } x \geqslant q_4; \\ \frac{x-q_1}{q_2-q_1}, & q_1 < x < q_2; \\ 1, & q_2 \leqslant x \leqslant q_3; \\ \frac{q_4-x}{q_4-q_3}, & q_3 < x < q_4. \end{cases}$$

Let $N \triangleq \cup_{\Phi\in\mathcal{H}} N_\Phi$. Then N is a $\mathbb{P}$-null set outside which (5.42) holds for all $\Phi \in \mathcal{H}$. From a standard approximation argument, this is sufficient to conclude that (5.42) holds for all non-negative Borel measurable functions. □

It is tempting to choose $\Phi_n \to \delta_a$ so that we obtain $L_t^a = \int_0^t \delta_a(X_s) d\langle X\rangle_s$, at least in the sense of (5.38), which verifies the intuitive meaning of local time (in the Brownian motion case) that we explained at the beginning. To do so, we need an even better version of L_t^a.

Theorem 5.22. *Suppose that $X_t = X_0 + M_t + A_t$ is a continuous semimartingale. Then there exists a modification $\{\widetilde{L}_t^a : a \in \mathbb{R}^1, t \geqslant 0\}$ of the process $\{L_t^a : a \in \mathbb{R}^1, t \geqslant 0\}$, such that with probability one, the map $(a,t) \mapsto \widetilde{L}_t^a(\omega)$ is continuous in t and càdlàg in a. Moreover, for each $a \in \mathbb{R}^1$,*

$$\widetilde{L}_t^a - \widetilde{L}_t^{a-} = 2\int_0^t \mathbf{1}_{\{X_s=a\}} dA_s = 2\int_0^t \mathbf{1}_{\{X_s=a\}} dX_s. \tag{5.43}$$

In particular, if X_t is a continuous local martingale, then there is a bicontinuous modification of the process $\{L_t^a : a \in \mathbb{R}^1, t \geqslant 0\}$.

Proof. We start with the jointly measurable modification L_t^a given by Proposition 5.22, which allows us to integrate with respect to a. From the second identity of Tanaka's formula, we have

$$\frac{1}{2}L_t^a = (X_t - a)^+ - (X_0 - a)^+ - \int_0^t \mathbf{1}_{\{X_s>a\}} dM_s - \int_0^t \mathbf{1}_{\{X_s>a\}} dA_s. \tag{5.44}$$

We first show that the family $\widehat{M}_t^a \triangleq \int_0^t \mathbf{1}_{\{X_s>a\}} dM_s$ of continuous local martingales possesses a bicontinuous modification in the pair (a,t). To this end, given $T > 0$, let W_T be the space of continuous paths on $[0,T]$, equipped with the uniform topology. It suffices to show that, when restricted on $t \in [0,T]$, the W_T-valued stochastic process $\{\widehat{M}^a : a \in \mathbb{R}^1\}$ possesses a continuous modification in a.

Indeed, for given $a < b$ and $k \geqslant 1$, the BDG inequalities (c.f. (5.20)) implies that

$$\mathbb{E}\left[\sup_{0\leqslant t\leqslant T}\left|\widehat{M}_t^a - \widehat{M}_t^b\right|^{2k}\right] \leqslant C_k \mathbb{E}\left[\left(\int_0^T \mathbf{1}_{\{a<X_s\leqslant b\}} d\langle M\rangle_s\right)^k\right]. \tag{5.45}$$

By applying the occupation times formula (c.f. (5.41)) to the function $\Phi = \mathbf{1}_{(a,b]}$, the right hand side of (5.45) is equal to

$$\begin{aligned}
C_k \mathbb{E}\left[\left(\int_a^b L_T^x dx\right)^k\right] &\leqslant C_k (b-a)^k \mathbb{E}\left[\left(\frac{1}{b-a}\int_a^b L_\infty^x dx\right)^k\right] \\
&\leqslant C_k (b-a)^k \mathbb{E}\left[\frac{1}{b-a}\int_a^b (L_\infty^x)^k dx\right] \\
&\leqslant C_k (b-a)^k \sup_{x\in\mathbb{R}^1} \mathbb{E}[(L_\infty^x)^k].
\end{aligned}$$

Now from (5.44), we can see that

$$L_\infty^x \leqslant 2\left(\sup_{t\geqslant 0}|X_t - X_0| + \sup_{t\geqslant 0}\left|\int_0^t \mathbf{1}_{\{X_s>x\}} dM_s\right| + \int_0^\infty d\|A\|_s\right),$$

where $\|A\|_t$ is the total variation process of A_t. The BDG inequalities again implies that

$$\mathbb{E}[(L^x_\infty)^k] \leqslant C'_k \mathbb{E}\left[\sup_{t\geqslant 0}|X_t - X_0|^k + \langle M\rangle_\infty^{k/2} + \left(\int_0^\infty d\|A\|_s\right)^k\right].$$

Observe that the right hand side is independent of x. If it is finite, then our claim follows from Kolmogorov's continuity theorem with state space W_T which is a complete metric space. In general, we define

$$\tau_n \triangleq \inf\left\{t \geqslant 0: \ |X_t - X_0|^k + \langle M\rangle_t^{k/2} + \left(\int_0^t d\|A\|_s\right)^k > n\right\}.$$

Then the previous argument applied to the stopped process X^{τ_n} (note that in this case the corresponding local time L^x_∞ will be the local time of the stopped process) implies that for each n, the family $(\widehat{M^a})^{\tau_n}$ possesses a bicontinuous modification in $(a,t) \in \mathbb{R}^1 \times [0,T]$. We denote such modification as $\widetilde{M}^{a,n}_t$. Note that when a, n are fixed, the relevant processes are always continuous in t. Therefore, for each given $n \geqslant 1$ and $a, t \geqslant 0$,

$$\widetilde{M}^{a,n+1}_{\tau_n \wedge t} = \widetilde{M}^{a,n}_t \quad \text{a.s.} \tag{5.46}$$

From the bicontinuity property, outside a single null set (5.46) holds for all n, a, t. In particular, we are able to define a single process $\widetilde{M}^a_t$ on $\mathbb{R}^1 \times [0,T]$ such that $\widetilde{M}^a_t = \widetilde{M}^{a,n}_t$ on $[0, \tau_n \wedge T]$. Of course $\widetilde{M}^a_t$ is bicontinuous in (a,t) with probability one and it is a modification of $\widehat{M}^a_t$.

Now consider the family of pathwise integral processes $\widehat{A}^a_t \triangleq \int_0^t \mathbf{1}_{\{X_s > a\}} dA_s$. Apparently,

$$\begin{aligned}
\widehat{A}^{a-}_t &= \lim_{\varepsilon \downarrow 0} \int_0^t \mathbf{1}_{\{X_s > a-\varepsilon\}} dA_s = \int_0^t \mathbf{1}_{\{X_s \geqslant a\}} dA_s, \qquad (5.47)\\
\widehat{A}^{a+}_t &= \lim_{\varepsilon \downarrow 0} \int_0^t \mathbf{1}_{\{X_s > a+\varepsilon\}} dA_s = \int_0^t \mathbf{1}_{\{X_s > a\}} dA_s = \widehat{A}^a_t.
\end{aligned}$$

Since $\widehat{A}^a_t$ is already continuous in t and càdlàg in a pathwisely, there is no way to improve the continuity of $\widehat{A}^a_t$ by taking a modification.

Therefore, there exists a modification $\widetilde{L}^a_t$ of L^a_t which is continuous in t and càdlàg in a with probability one. If X_t is a continuous local martingale, then $A = 0$ and we obtain a bicontinuous modification.

It remains to show (5.43). The first part is clear from (5.47). To see the second part, it suffices to show that

$$\int_0^t \mathbf{1}_{\{X_s = a\}} dM_s = 0, \quad \forall t \geqslant 0,$$

for each given a. But from the occupation times formula applied to the function $\Phi = \mathbf{1}_{\{a\}}$, we know that

$$\int_0^t \mathbf{1}_{\{X_s=a\}} d\langle M\rangle_s = \int_0^t \mathbf{1}_{\{X_s=a\}} d\langle X\rangle_s = \int_{\mathbb{R}^1} \mathbf{1}_{\{a\}}(x) \widetilde{L}^x_t dx = 0, \quad \forall t \geqslant 0.$$

Therefore the result follows. □

Remark 5.16. It is important to point out that in the presence of A_t, we cannot expect a modification which is bicontinuous in (a,t) in general (a good example is illustrated in Problem Sheet 5, Problem 7). However, such possible discontinuity is *not* a pure effect of the presence of A_t. Indeed, if $M=0$, by the occupation times formula we have $\int_{\mathbb{R}^1}\Phi(x)L_t^x dx=0$ for all non-negative Borel measurable Φ. In particular, $L_t^x=0$ for almost every $x\in\mathbb{R}^1$. Since L_t^x is càdlàg in x, we conclude that $L_t^x=0$ for all (x,t). Therefore, the possible discontinuity of L_t^a in a is a consequence of the interaction between the martingale part and the bounded variation part of X.

Remark 5.17. From the proof of Theorem 5.22, if X is a continuous local martingale, we have indeed shown that there exists a modification $\widetilde{L}_t^a$ of $\{L_t^a:\ a\in\mathbb{R}^1, t\geqslant 0\}$, such that with probability one, $a\mapsto\widetilde{L}_t^a$ is locally γ-Hölder continuous uniformly on every finite t-interval for every $\gamma\in(0,1/2)$:

$$\mathbb{P}\left(\sup_{t\in[0,T]}\sup_{0<|a-b|<C}\frac{\left|\widetilde{L}_t^a-\widetilde{L}_t^b\right|}{|a-b|^\gamma}<\infty\right)=1$$

for every $T,C>0$ and $\gamma\in(0,1/2)$.

Now we use the version L_t^a of local time that we obtain in Theorem 5.22. Then we have the following result which verifies (5.38) at the beginning.

Corollary 5.5. *With probability one, we have*

$$L_t^a=\lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}\int_0^t\mathbf{1}_{[a,a+\varepsilon)}(X_s)d\langle X\rangle_s\ \ \forall a\in\mathbb{R}^1, t\geqslant 0. \tag{5.48}$$

If X_t is a continuous local martingale, we also have

$$L_t^a=\lim_{\varepsilon\downarrow 0}\frac{1}{2\varepsilon}\int_0^t\mathbf{1}_{(a-\varepsilon,a+\varepsilon)}(X_s)d\langle X\rangle_s,\ \ \forall a\in\mathbb{R}^1, t\geqslant 0. \tag{5.49}$$

In particular, in the Brownian motion case, (5.38) holds with the left hand side being the local time at 0 of the Brownian motion.

Proof. From the occupation times formula, we know that with probability one,

$$\frac{1}{\varepsilon}\int_0^t\mathbf{1}_{[a,a+\varepsilon)}(X_s)d\langle X\rangle_s=\frac{1}{\varepsilon}\int_a^{a+\varepsilon}L_t^x dx,\ \ \forall a,t,\varepsilon.$$

But L_t^a is right continuous in a, so we have (5.48) by letting $\varepsilon\to 0$. If X_t is a continuous local martingale, then L_t^a is continuous in a, in which case (5.49) follows from the same reasoning but with $\Phi=\mathbf{1}_{(a-\varepsilon,a+\varepsilon)}$ when applying the occupation times formula. □

Although we are not going to touch any distributional properties related to local times, we finish this section by stating an elegant result for the Brownian local time along this direction. This result is due to Lévy. We refer the reader to [8] for the proof.

Theorem 5.23. *Let B_t be a one dimensional Brownian motion, and let L_t be the local time at 0 of B. Then the two-dimensional processes $\{(S_t - B_t, S_t) : \ t \geqslant 0\}$ and $\{(|B_t|, L_t) : \ t \geqslant 0\}$ have the same distribution, where $S_t \triangleq \max_{0 \leqslant s \leqslant t} B_s$ is the running maximum of B.*

6 Stochastic differential equations

Consider a second order differential operator $\mathcal{A}$ over $\mathbb{R}^n$ of the form

$$\mathcal{A} = \frac{1}{2}\sum_{i,j=1}^{n} a^{ij}(x)\frac{\partial^2}{\partial x^i \partial x^j} + \sum_{i=1}^{n} b^i(x)\frac{\partial}{\partial x^i}.$$

There are two fundamental questions one could ask in general:

(1) How can we construct a Markov process (or more precisely, a Markov family $((\Omega, \mathcal{F}, \mathbb{P}), \{X_t^x, \mathcal{F}_t : x \in \mathbb{R}^n, t \geqslant 0\}))$ with $\mathcal{A}$ being its infinitesimal generator, in the sense that

$$\lim_{t\to 0} \frac{1}{t}\left(\mathbb{E}[f(X_t^x)] - f(x)\right) = (\mathcal{A}f)(x), \quad \forall x \in \mathbb{R}^n,$$

for all $f \in C_b^2(\mathbb{R}^n)$?

(2) How can we construct the fundamental solution to the parabolic PDE $\frac{\partial u}{\partial t} - \mathcal{A}u = 0$?

The first question is purely probabilistic and the second one is purely analytic. However, to some extent, these two questions are indeed equivalent. If a Markov family solves Question (1) with a nice transition probability density function $p(t, x, y) \triangleq \mathbb{P}(X_t^x \in dy)/dy$, then $p(t, x, y)$ solves Question (2). Conversely, if $p(t, x, y)$ is a solution to Question (2), then a standard Kolmogorov's extension argument allows us to construct a Markov family on path space which solves Question (1).

It was Lévy who suggested a purely probabilistic approach to study these questions, and Itô carried out this program in a series of far-reaching works. The philosophy of this approach can be summarized as follows. Let $a = \sigma\sigma^*$ for some matrix σ. Suppose that there exists a stochastic process X_t which solves the following *stochastic differential equation* (in matrix notation):

$$\begin{cases} dX_t^x = \sigma(X_t^x)dB_t + b(X_t^x)dt, & t \geqslant 0, \\ X_0 = x, \end{cases}$$

which is of course understood in Itô's integral sense. Then the family $\{X_t^x\}$ solves Question (1), or equivalently, the probability density function $p(t, x, y) \triangleq \mathbb{P}(X_t^x \in dy)/dy$ solves Question (2) provided that it exists and is reasonably regular. The existence and regularity of the density $p(t, x, y)$ is a rich subject under the framework of Malliavin's calculus, in which the theory is well developed in the case when $\mathcal{A}$ is a hypoelliptic operator. The reader may consult [10] for a nice introduction to this theory.

This general discussion provides us with a natural motivation to study the theory of stochastic differential equations in depth. This is the main focus of the present section.

6.1 Itô's theory of stochastic differential equations

We start with Itô's classical approach.

Recall that $(W^n, \mathcal{B}(W^n))$ is the space of continuous paths in $\mathbb{R}^n$, equipped with a metric ρ defined by (1.3) which characterizes uniform convergence on compact intervals. We use $\{\mathcal{B}_t(W^n)\}$ to denote the natural filtration of the coordinate process on W^n.

In its full generality, we are interested in a stochastic differential equation (we simply call it an SDE hereafter) of the form

$$dX_t = \alpha(t, X)dB_t + \beta(t, X)dt. \tag{6.1}$$

Here B_t is a d-dimensional Brownian motion, X_t is an n-dimensional continuous stochastic process, α, β are maps defined on $[0, \infty) \times W^n$ taking values in $\mathrm{Mat}(n, d)$ (the space of real $n \times d$ matrices) and in $\mathbb{R}^n$ respectively. Note that α, β here can depend on the whole trajectory of X, and we write X to emphasize that it is a random variable taking values in W^n.

From now on, when we are concerned with an SDE of the form (6.1), we always make the following measurability assumption on the coefficients α and β.

Assumption 6.1. *Regarded as stochastic processes defined on* $(W^n, \mathcal{B}(W^n))$, α *and* β *are* $\{\mathcal{B}_t(W^n)\}$*-progressively measurable.*

Remark 6.1. Recall from the solution of Problem Sheet 2, Problem 6 that $A \in \mathcal{B}_t(W^n)$ if and only if for any two $w, w' \in W^n$, if $w \in A$, $w = w'$ on $[0, t]$, then $w' \in \mathcal{B}_t(W^n)$. Given $t \geqslant 0$, consider

$$A \triangleq \{w \in W^n : \ \alpha(t, w) = \alpha(t, w^t)\} \in \mathcal{B}_t(W^n),$$

where $w^t \triangleq (w_{t \wedge s})_{s \geqslant 0}$ is the path obtained by stopping w at t. For every $w \in W^n$, since $w = w^t$ on $[0, t]$ and $w^t \in A$, we conclude that $w \in A$. Therefore, $\alpha(t, w) = \alpha(t, w^t)$ for every $(t, w) \in [0, \infty) \times W^n$. Similar result holds for β.

Remark 6.2. If X_t is a progressively measurable process defined on some filtered probability space, then the process $\alpha(t, X)$ is progressively measurable. Similar result is true for $\beta(t, X)$.

Now we can talk about the meaning of solutions to (6.1). Unlike ordinary differential equations, the meaning of an SDE is not just about a solution process itself; it should also involve the underlying filtered probability space together with a Brownian motion. This leads to two notions of solutions: strong and weak solutions. Heuristically, being solutions in the strong sense means that we are solving the SDE on a *given* filtered probability space with a *given* Brownian motion on it, while being solutions in the weak sense means that the SDE is solvable on *some* filtered probability space with *some* Brownian motion on it. In the strong setting, we are particularly interested in how a solution can be constructed from the given initial data and the given Brownian motion. In the weak setting, we are mainly interested in distributional properties of the solution process and do not care what the underlying space and Brownian motion are (they can be arbitrary as long as the equation is verified). From the next subsection, we will discuss the strong and weak notions of solutions in detail .

As an introduction to the theory, we start with Itô's classical approach which falls in the context of strong solutions. Therefore, we assume that $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ is a given filtered probability space which satisfies the usual conditions, and B_t is an $\{\mathcal{F}_t\}$-Brownian motion. Suppose that the coefficients α, β satisfy Assumption 6.1.

Itô's theory, which is essentially an L^2-theory, asserts that the SDE (6.1) is uniquely solvable for any given initial data $\xi \in L^2(\Omega, \mathcal{F}_0, \mathbb{P})$, provided that the coefficients satisfy the Lipschitz condition and have linear growth.

Theorem 6.1. *Suppose that the coefficients α, β satisfy the following two conditions: there exists a constant $K > 0$, such that*

(1) (Lipschitz condition) for any $w, w' \in W^n$ and $t \geqslant 0$,

$$\|\alpha(t,w) - \alpha(t,w')\| + \|\beta(t,w) - \beta(t,w')\| \leqslant K(w - w')_t^*; \tag{6.2}$$

(2) (linear growth condition) for any $w \in W^n$ and $t \geqslant 0$,

$$\|\alpha(t,w)\| + \|\beta(t,w)\| \leqslant K(1 + w_t^*), \tag{6.3}$$

where $w_t^* \triangleq \sup_{s \leqslant t} |w_s|$ is the running maximum of w. Then for any initial data $\xi \in L^2(\Omega, \mathcal{F}_0, \mathbb{P})$, there exists a unique continuous, $\{\mathcal{F}_t\}$-adapted process X_t in $\mathbb{R}^n$, such that

$$X_t^i = \xi + \sum_{k=1}^{d} \int_0^t \alpha_k^i(s, X) dB_s^k + \int_0^t \beta^i(s, X) ds, \quad t \geqslant 0, \ 1 \leqslant i \leqslant n. \tag{6.4}$$

In addition, for each $T > 0$, there exists some constant $C_{T,K}$ depending only on T, K and dimensions, such that

$$\mathbb{E}[(X_T^*)^2] \leqslant C_{T,K}(1 + \mathbb{E}[|\xi|^2]), \quad t \geqslant 0. \tag{6.5}$$

In particular, the martingale part of X_t is a square integrable $\{\mathcal{F}_t\}$-martingale.

The key ingredient in proving the theorem is the following estimate. Although here we only need the case when $p = 2$, the estimate for arbitrary p is quite useful for many purposes.

Lemma 6.1. *Let $X_t = (X_t^1, \cdots, X_t^n)$ be a vector of continuous semimartingales of the form*

$$X_t = \xi + \int_0^t \alpha_s dB_s + \int_0^t \beta_s ds, \tag{6.6}$$

provided the integrals are well defined in the appropriate sense, where (6.6) is written in the matrix form. Then for each $T > 0$ and $p \geqslant 2$, there exists some constant $C_{T,p}$ depending only on T, p and dimensions, such that

$$\mathbb{E}[(X_t^*)^p] \leqslant C_{T,p} \left(\mathbb{E}[|\xi|^p] + \mathbb{E}\left[\int_0^t (\|\alpha_s\|^p + \|\beta_s\|^p) \, ds \right] \right), \quad 0 \leqslant t \leqslant T,$$

where $X_t^* \triangleq \sup_{s \leqslant t} |X_s|$.

Proof. Note that

$$(X_t^*)^p \leqslant C_p\left(|\xi|^p + \sup_{0\leqslant s\leqslant t}\left|\int_0^s \alpha_u dB_u\right|^p + \left(\int_0^t \|\beta_u\| du\right)^p\right).$$

The result then follows easily from the BDG inequalities (c.f. (5.20)) and Hölder's inequality. □

Coming back to Theorem 6.1, we first prove uniqueness. It then allows us to patch solutions defined on finite intervals to obtain a global solution defined on $[0,\infty)$.

Similar to ordinary differential equations, uniqueness is usually obtained by applying the following *Gronwall's inequality*.

Lemma 6.2. *Let $g: [0,T] \to [0,\infty)$ be a non-negative, continuous function defined on $[0,T]$. Suppose that*

$$g(t) \leqslant c(t) + k\int_0^t g(s)ds, \quad 0 \leqslant t \leqslant T, \tag{6.7}$$

for some $k \geqslant 0$ and some integrable $c: [0,T] \to \mathbb{R}^1$. Then

$$g(t) \leqslant c(t) + k\int_0^t c(s)\mathrm{e}^{k(t-s)}ds, \quad 0 \leqslant t \leqslant T.$$

Proof. From (6.7), we have

$$\begin{aligned} g(t) &\leqslant c(t) + k\int_0^t \left(c(s) + k\int_0^s g(u)du\right)ds \\ &= c(t) + k\int_0^t c(s)ds + k^2\int_{0<u<s<t} g(u)duds \\ &= c(t) + k\int_0^t c(s)ds + k^2\int_0^t g(u)(t-u)du. \end{aligned}$$

By applying (6.7) inductively, for every $m \geqslant 1$, we have

$$g(t) \leqslant c(t) + \sum_{l=1}^m k^l \int_0^t \frac{c(s)(t-s)^{l-1}}{(l-1)!}ds + k^{m+1}\int_0^t \frac{(t-s)^m}{m!}g(s)ds. \tag{6.8}$$

Since g is continuous on $[0,T]$, we know that it is bounded on $[0,T]$. Therefore, the last term of (6.8) tends to zero as $n \to \infty$. As c is integrable on $[0,T]$, it follows from the dominated convergence theorem that

$$g(t) \leqslant c(t) + k\int_0^t c(s)\mathrm{e}^{k(t-s)}ds.$$

□

Now suppose that X, Y are two solutions to the SDE (6.1) (i.e. satisfying Theorem 6.1), so in matrix form we have

$$\begin{aligned} X_t &= \xi + \int_0^t \alpha(s, X) dB_s + \int_0^t \beta(s, X) ds, \\ Y_t &= \xi + \int_0^t \alpha(s, Y) dB_s + \int_0^t \beta(s, Y) ds. \end{aligned}$$

Define $\tau_m \triangleq \inf\{t \geqslant 0 : \ |X_t - Y_t| \geqslant m\}$. Then we have

$$\begin{aligned} (X-Y)_t^{\tau_m} &= \int_0^t (\alpha(s, X) - \alpha(s, Y)) \mathbf{1}_{[0,\tau_m]} dB_s \\ &\quad + \int_0^t (\beta(s, X) - \beta(s, Y)) \mathbf{1}_{[0,\tau_m]} ds. \end{aligned}$$

By applying Lemma 6.1 in the case when $p = 2$ and the Lipschitz condition (6.2), we conclude that for every given $T > 0$,

$$\mathbb{E}\left[\left((X-Y)^*_{t\wedge\tau_m}\right)^2\right] \leqslant C_{T,K} \int_0^t \mathbb{E}\left[\left((X-Y)^*_{s\wedge\tau_m}\right)^2\right] ds, \quad \forall 0 \leqslant t \leqslant T.$$

Now define

$$f(t) = \mathbb{E}\left[\left((X-Y)^*_{t\wedge\tau_m}\right)^2\right], \quad t \in [0, T].$$

From the dominated convergence theorem, we easily see that f is non-negative and continuous on $[0, T]$. Therefore, according to Gronwall's inequality (c.f. Lemma 6.2), $f = 0$. As $\tau_m \uparrow \infty$, we conclude that $X = Y$ on $[0, T]$, which implies that $X = Y$ as T is arbitrary.

Now we consider existence. From the uniqueness part, it suffices to show existence on every finite interval $[0, T]$. Indeed, if $X^{(T)}$ satisfies (6.4) on $[0, T]$, then the uniqueness argument will imply that $X^{(T+1)} = X^{(T)}$ on $[0, T]$, which allows us to define a single process X such that $X = X^{(T)}$ on $[0, T]$. In view of Remark 6.1, we see that

$$\alpha(s, X) = \alpha(s, X^s) = \alpha\left(s, \left(X^{(T)}\right)^s\right) = \alpha\left(s, X^{(T)}\right)$$

for every $s \leqslant T$, and similar result is true for β. Therefore, X is a global solution to the SDE (6.1) in the sense of Theorem 6.1.

For fixed $T > 0$, define L_T^2 to be the space of all continuous, $\{\mathcal{F}_t\}$-adapted processes X_t on $[0, T]$ such that $\mathbb{E}[(X_T^*)^2] < \infty$ (technically we define $X_t \triangleq X_T$ for $t > T$ so that X_t is well defined on $[0, \infty)$). Then L_T^2 is a Banach space. Indeed, if $X^{(m)}$ is a Cauchy sequence in L_T^2, then along a subsequence m_k we have

$$\mathbb{E}\left[\sup_{0\leqslant t\leqslant T}\left|X_t^{(m_{k+1})} - X_t^{(m_k)}\right|^2\right] < \frac{1}{2^k}, \quad \forall k \geqslant 1.$$

From Chebyshev's inequality and the first Borel-Cantelli's lemma, we know that with probability one, $X^{(m_k)}$ is a Cauchy sequence in the space of continuous paths on $[0, T]$

under uniform topology. Therefore, with probability one, $X_t^{(m_k)}$ converges to some continuous X_t uniformly on $[0,T]$. It is apparent that X_t is $\{\mathcal{F}_t\}$-adapted and $\mathbb{E}[(X_T^*)^2] < \infty$. Moreover, from Fatou's lemma, we have

$$\lim_{m\to\infty}\mathbb{E}\left[\sup_{0\leqslant t\leqslant T}\left|X_t^{(m)}-X_t\right|^2\right]\leqslant\lim_{m\to\infty}\liminf_{k\to\infty}\mathbb{E}\left[\sup_{0\leqslant t\leqslant T}\left|X_t^{(m)}-X_t^{(m_k)}\right|^2\right]=0.$$

Therefore, L_T^2 is a Banach space. For $t\leqslant T$, we denote $\|X\|_t\triangleq\sqrt{\mathbb{E}[(X_t^*)^2]}$.

The proof of existence on $[0,T]$ is a standard Picard's iteration argument. Therefore we consider the map $\mathcal{R}:\ L_T^2\to L_T^2$ defined by

$$(\mathcal{R}X)_t\triangleq\xi+\int_0^t\alpha(s,X)dB_s+\int_0^t\beta(s,X)ds,\quad t\in[0,T].$$

Similar to the uniqueness argument, Lemma 6.1 and the Lipschitz condition show that

$$\|\mathcal{R}X-\mathcal{R}Y\|_t^2\leqslant C_{T,K}\int_0^t\|X-Y\|_s^2ds,\quad\forall 0\leqslant t\leqslant T. \tag{6.9}$$

Now define $X^{(0)}\triangleq\xi$, and for each $m\geqslant 1$, define $X^{(m)}\triangleq\mathcal{R}X^{(m-1)}$. By the linear growth condition (6.3), It is apparent that

$$\|X^{(1)}-X^{(0)}\|_T^2\leqslant C_{T,K}(1+\mathbb{E}[|\xi|^2]).$$

In addition, from (6.9), we have

$$\begin{aligned}\|X^{(m+1)}-X^{(m)}\|_T^2 &\leqslant C_{T,K}\int_0^T\|X^{(m)}-X^{(m-1)}\|_s^2ds\\ &\cdots\\ &\leqslant C_{T,K}^m\int_{0<s_1<\cdots<s_m<T}\|X^{(1)}-X^{(0)}\|_{s_1}^2ds_1\cdots ds_m\\ &\leqslant\frac{C_{T,K}^{m+1}T^m}{m!}(1+\mathbb{E}[|\xi|^2]).\end{aligned} \tag{6.10}$$

Since the right hand side of (6.10) is summable, we conclude that $X^{(m)}$ is a Cauchy sequence in L_T^2. Suppose that $X=\lim_{n\to\infty}X^{(m)}$ in L_T^2. It follows that

$$\|X-\mathcal{R}X\|_T\leqslant\|X-X^{(m)}\|_T+\|X^{(m)}-\mathcal{R}X^{(m)}\|_T+\|\mathcal{R}X^{(m)}-\mathcal{R}X\|_T.$$

Combining with (6.9), (6.10) and the fact that $X^{(m+1)}=\mathcal{R}X^{(m)}$, we conclude that $X=\mathcal{R}X$, which shows that X is a solution to the SDE (6.1) on $[0,T]$.

Finally, since $X\in L_T^2$, (6.5) follows immediately from Lemma 6.1 and Gronwall's inequality.

Now the proof of Theorem 6.1 is complete.

Let us take a second thought on the proof of Theorem 6.1. On the one hand, to expect (pathwise) uniqueness, we can see that some kind of Lipschitz condition is necessary. Because of localization, this part does not really rely on the integrability of solution. On the other hand, in the existence part, it is not so clear whether the Lipschitz condition is playing a crucial role as the fixed point argument is certainly not the only way to obtain existence. Moreover, in the previous argument we can see that the square integrability of ξ does play an important role for the existence. It is not so clear from the argument whether existence still holds if ξ is simply an $\mathcal{F}_0$-measurable random variable.

Therefore, to some extent, it is more fundamental to separate the study of existence and uniqueness in different contexts, and to understand how they are combined to give a single well-posed theory of SDE. This leads us to the realm of Yamada-Watanabe's theory.

6.2 Different notions of solutions and the Yamada-Watanabe theorem

In this subsection, we study different notions of existence and uniqueness for an SDE, which are all very natural and important on their own. Then we present the fundamental theorem of Yamada and Watanabe, which outlines the structure of the theory of SDEs.

We first make the following convention.

Definition 6.1. By a set-up $((\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\}), \xi, B_t)$, we mean that

(1) $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ is a filtered probability space which satisfies the usual conditions;

(2) ξ is an $\mathcal{F}_0$-measurable random variable;

(3) B_t is a d-dimensional $\{\mathcal{F}_t\}$-Brownian motion.

Now let $\alpha : [0, \infty) \times W^n \to \mathrm{Mat}(n, d)$ and $\beta : [0, \infty) \times W^n \to \mathbb{R}^n$ be two maps satisfying Assumption 6.1. We are interested in an SDE of the general form (6.1). For simplicity, we always use matrix notation in writing our equations.

Motivated from Itô's classical result, it is natural to introduce the following definition in the strong sense.

Definition 6.2. We say that the SDE (6.1) is *(pathwise) exact* if on any given set-up $((\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\}), \xi, B_t)$, there exists exactly one (up to indistinguishability) continuous, $\{\mathcal{F}_t\}$-adapted n-dimensional process X_t, such that with probability one,

$$\int_0^t \left(\|\alpha(s, X)\|^2 + \|\beta(s, X)\|\right) ds < \infty, \quad \forall t \geqslant 0, \tag{6.11}$$

and

$$X_t = \xi + \int_0^t \alpha(s, X) dB_s + \int_0^t \beta(s, X) ds, \quad t \geqslant 0. \tag{6.12}$$

As we mentioned at the end of last subsection, it is even not clear if exactness is true in Itô's setting (i.e. under the conditions in Theorem 6.1) although we do have uniqueness. Therefore, it is a fundamental problem to understand how one can prove exactness in general. Before doing so, we need to introduce different notions of existence and uniqueness, which are all important and natural on their own.

Definition 6.3. Let μ be a probability measure on $\mathbb{R}^n$. We say that the SDE (6.1) *has a weak solution with initial distribution* μ if there exists a set-up $((\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\}), \xi, B_t)$ together with a continuous, $\{\mathcal{F}_t\}$-adapted n-dimensional process X_t, such that

(1) ξ has distribution μ;

(2) X_t satisfies (6.11) and (6.12).

If for every probability measure μ on $\mathbb{R}^n$, the SDE (6.1) has a weak solution with initial distribution μ, we say that it *has a weak solution*.

From Definition 6.3, a weak solution is the existence of a set-up on which the SDE is satisfied in Itô's integral sense. A particular feature of a weak solution is that we have large flexibility on choosing a set-up; it could be any set-up as long as conditions (1) and (2) are verified on it. Therefore, in some sense a weak solution only reflects its distributional properties.

Corresponding to weak solutions, we have the notion of uniqueness in law.

Definition 6.4. We say that the solution to the SDE (6.1) is *unique in law* if whenever X_t and X'_t are two weak solutions (possibly defined on two different probability set-ups) with the same initial distributions, they have the same law on W^n.

In contrast to the weak formulation, we have another (strong) notion of uniqueness.

Definition 6.5. We say that *pathwise uniqueness* holds for the SDE (6.1) if the following statement is true. Given any set-up $((\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\}), \xi, B_t)$, if X_t and X'_t are two continuous, $\{\mathcal{F}_t\}$-adapted n-dimensional process satisfying (6.11) and (6.12), then $\mathbb{P}(X_t = X'_t \ \forall t \geqslant 0) = 1$.

It is part of the Yamada-Watanabe theorem that pathwise uniqueness implies uniqueness in law (c.f. Theorem 6.2 below). However, the converse is not true and the following is a famous counterexample due to Tanaka.

Example 6.1. Consider the one dimensional SDE

$$dX_t = \sigma(X_t)dB_t \tag{6.13}$$

where $\sigma(x) = -1$ if $x \leqslant 0$ and $\sigma(x) = 1$ if $x > 0$.

Suppose that X_t is a weak solution with initial distribution μ on some given set-up $((\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\}), \xi, B_t)$, so that we have

$$X_t = \xi + \int_0^t \sigma(X_s)dB_s.$$

Since $\int_0^t \sigma(X_s)dB_s$ is an $\{\mathcal{F}_t\}$-Brownian motion according to Lévy's characterization theorem, we see immediately that the distribution of X is uniquely determined by μ and the law of Brownian motion. Therefore, uniqueness in law holds. Moreover, for given initial distribution μ, let $((\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\}), \xi, B_t)$ be an arbitrary set-up in which ξ has distribution μ. Define $X_t \triangleq \xi + B_t$ and let $\widetilde{B}_t \triangleq \int_0^t \sigma(X_s)dB_s$. Lévy's characterization theorem again tells us that $\widetilde{B}_t$ is an $\{\mathcal{F}_t\}$-Brownian motion, and the associativity of stochastic integrals implies that

$$X_t = \xi + \int_0^t \sigma(X_s)d\widetilde{B}_s.$$

Therefore, the SDE (6.13) has a weak solution.

However, pathwise uniqueness does not hold. Indeed, suppose that $X_t = \int_0^t \sigma(X_s)dB_s$ on some set-up (so $X_0 = 0$ in this case). According to the occupation time formula (c.f. (5.41)), we know that $\int_0^t \mathbf{1}_{\{X_s=0\}}ds = 0$, which implies that $\int_0^t \mathbf{1}_{\{X_s=0\}}dB_s = 0$. Therefore, $(-X_t) = \int_0^t \sigma(-X_s)dB_s$. This shows that pathwise uniqueness fails as $X \neq -X$.

Now we can state the renowned Yamada-Watanabe theorem which has far-reaching consequences. The proof is beyond the scope of the course and hence omitted. The interested reader may consult N. Ikeda and S. Watanabe, *Stochastic differential equations and diffusion processes,* 1989 for basically the original proof.

Theorem 6.2. *The SDE (6.1) is exact if and only if it has a weak solution and pathwise uniqueness holds. In addition, pathwise uniqueness implies uniqueness in law.*

If we have an exact SDE, it is natural to expect that there is some *universal* way to produce the unique solution (as the output) whenever an initial data and a Brownian motion are given (as the input), regardless of the set-up we are working on. In other words, it is natural to look for a single function $F: \mathbb{R}^n \times W^d \to W^n$, such that on any given set-up $((\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\}), \xi, B_t)$, $X \triangleq F(\xi, B)$ produces the unique solution. This is indeed the original spirit of Yamada and Watanabe.

Definition 6.6. A function $F: \mathbb{R}^n \times W^d \to W^n$ is called $\widehat{\mathcal{E}}(\mathbb{R}^n \times W^d)$*-measurable* if for any probability measure μ on $\mathbb{R}^n$, there exists a function $F_\mu: \mathbb{R}^n \times W^d \to W^n$ which is $\overline{\mathcal{B}(\mathbb{R}^n \times W^d)}^{\mu\times\mathbb{P}^W}/\mathcal{B}(W^n)$-measurable, where $\mathbb{P}^W$ is the distribution of Brownian motion and $\overline{\mathcal{B}(\mathbb{R}^n \times W^d)}^{\mu\times\mathbb{P}^W}$ is the $\mu \times \mathbb{P}^W$-completion of $\mathcal{B}(\mathbb{R}^n \times W^d)$, such that for μ-almost all $x \in \mathbb{R}^n$, we have

$$F(x, w) = F_\mu(x, w) \text{ for } \mathbb{P}^W - \text{almost all } w \in W^d.$$

If ξ is an $\mathbb{R}^n$-valued random variable with distribution μ and B_t is a Brownian motion, we set $F(\xi, B) \triangleq F_\mu(\xi, B)$.

Definition 6.7. We say that the SDE (6.1) *has a unique strong solution* if there exists an $\widehat{\mathcal{E}}(\mathbb{R}^n \times W^d)$-measurable function $F: \mathbb{R}^n \times W^d \to W^n$, such that:

(1) for every fixed $x \in \mathbb{R}^n$, $w \mapsto F(x, w)$ is $\overline{\mathcal{B}_t(W^d)}^{\mathbb{P}^W}/\mathcal{B}_t(W^n)$-measurable for each $t \geqslant 0$;

(2) given any set-up $((\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\}), \xi, B_t)$, $X \triangleq F(\xi, B)$ is a continuous, $\{\mathcal{F}_t\}$-adapted process which satisfies (6.11) and (6.12);

(3) for any continuous, $\{\mathcal{F}_t\}$-adapted process X_t satisfying (6.11) and (6.12) on a given set-up $((\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\}), \xi, B_t)$, we have $X = F(\xi, B)$ almost surely.

The following elegant result puts the philosophy of "constructing the unique solution out of initial data and Brownian motion in a universal way" on firm mathematical basis. This is essentially another form of the Yamada-Watanabe theorem.

Theorem 6.3. *The SDE (6.1) is exact if and only if it has a unique strong solution.*

6.3 Existence of weak solutions

The Yamada-Watanabe theorem tells us that the structure of exactness is very simple: we only need to study weak existence and pathwise uniqueness independently, and they combine to give exactness.

We first study weak existence in this subsection. The main result is that (surprisingly) continuity of coefficients is sufficient to guarantee weak existence (up to an intrinsic explosion time), and it has nothing to do with any Lipschitz property (compare Theorem 6.1 in Itô's theory).

In general, the weak existence has an elegant martingale characterization, which is known as *Stroock and Varadhan's martingale problem*. Let α, β be the coefficients of the SDE (6.1) which satisfy Assumption 6.1. We define the *generator* $\mathcal{A}$ of the SDE in the following way: for $f \in C_b^2(\mathbb{R}^n)$ (the space of twice continuously differentiable with bounded derivatives of up to second order), define $\mathcal{A}f$ to be the function on $[0,\infty)\times W^n$ given by

$$\begin{aligned}(\mathcal{A}f)(t,w) \triangleq{}& \frac{1}{2}\sum_{i,j=1}^{n} a^{ij}(t,w)\frac{\partial^2 f}{\partial x^i \partial x^j}(w_t)\\ &+\sum_{i=1}^{n}\beta^i(t,w)\frac{\partial f}{\partial x^i}(w_t), \quad (t,w)\in[0,\infty)\times W^n, \end{aligned} \tag{6.14}$$

where a is the $n\times n$ matrix defined by $a \triangleq \alpha\alpha^*$.

Suppose that X_t satisfies (6.11) and (6.12) on a given set-up $((\Omega,\mathcal{F},\mathbb{P};\{\mathcal{F}_t\}),\xi,B_t)$. For any $f\in C_b^2(\mathbb{R}^n)$, according to Itô's formula, we have

$$\begin{aligned} f(X_t) ={}& f(\xi)+\sum_{i=1}^{n}\sum_{k=1}^{d}\int_0^t \frac{\partial f}{\partial x^i}(X_s)\alpha_k^i(s,X)dB_s^k+\sum_{i=1}^{n}\int_0^t\frac{\partial f}{\partial x^i}(X_s)\beta^i(s,X)ds\\ &+\frac{1}{2}\sum_{i,j=1}^{n}\sum_{k=1}^{d}\int_0^t\frac{\partial^2 f}{\partial x^i\partial x^j}(X_s)\alpha_k^i(s,X)\alpha_k^j(s,X)ds. \end{aligned}$$

Therefore,

$$f(X_\cdot)-f(\xi)-\int_0^\cdot(\mathcal{A}f)(s,X)ds\in\mathcal{M}_0^{\mathrm{loc}} \tag{6.15}$$

on $(\Omega,\mathcal{F},\mathbb{P};\{\mathcal{F}_t\})$.

Conversely, suppose that X_t is a continuous, $\{\mathcal{F}_t\}$-adapted process defined on a given filtered probability space $(\Omega,\mathcal{F},\mathbb{P};\{\mathcal{F}_t\})$ satisfying the usual conditions, such that (6.11) and (6.15) hold for every $f \in C_b^2(\mathbb{R}^n)$ (of course with $\xi = X_0$). Let

$$M_t \triangleq X_t - X_0 - \int_0^t \beta(s, X)ds, \quad t \geqslant 0.$$

For each $R > 0$, define $f_R^i \in C_b^2(\mathbb{R}^n)$ to be such that $f_R^i(x) = x^i$ if $|x| \leqslant R$. Let

$$\sigma_R \triangleq \inf\left\{t \geqslant 0 : \ |X_t| > R \text{ or } \left|\int_0^t \beta^i(s, X)ds\right| > R\right\}.$$

From (6.15), we know that

$$\left(M^i\right)^{\sigma_R} = f_R^i(X_{\sigma_R \wedge \cdot}) - f_R^i(X_0) - \int_0^{\sigma_R \wedge \cdot} \beta^i(s, X)ds \in \mathcal{M}_0^{\text{loc}}.$$

But $(M^i)^{\sigma_R}$ is uniformly bounded, so $(M^i)^{\sigma_R}$ is indeed a martingale. Since $\sigma_R \uparrow \infty$ as $R \to \infty$, we conclude that $M^i \in \mathcal{M}_0^{\text{loc}}$.

Similarly, by considering $f_R^{ij} \in C_b^2(\mathbb{R}^n)$ with $f_R^{ij}(x) = x^i x^j$ when $|x| \leqslant R$, and by defining σ_R in a similar way but for f_R^{ij}, we know that

$$N_\cdot^{ij} \triangleq X_\cdot^i X_\cdot^j - X_0^i X_0^j - \int_0^\cdot a^{ij}(s, X)ds - \int_0^\cdot (X_s^i \beta^j(s, X) + X_s^j \beta^i(s, X))ds \in \mathcal{M}_0^{\text{loc}}. \tag{6.16}$$

On the other hand, from the integration by parts formula, we have

$$\begin{aligned} X_t^i X_t^j &= X_0^i X_0^j + \int_0^t X_s^i dX_s^j + \int_0^t X_s^j dX_s^i + \langle X^i, X^j \rangle_t \\ &= X_0^i X_0^j + \int_0^t X_s^i dM_s^j + \int_0^t X_s^j dM_s^i \\ &\quad + \int_0^t X_s^i \beta^j(s, X)ds + \int_0^t X_s^j \beta^i(s, X)ds + \langle M^i, M^j \rangle. \end{aligned} \tag{6.17}$$

By comparing (6.16) and (6.17), we conclude that

$$\langle M^i, M^j \rangle_t = \int_0^t a^{ij}(s, X)ds.$$

According to the martingale representation theorem for general filtrations (c.f. Theorem 5.14), possibly on an enlargement of $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$, we have

$$M_t = \int_0^t \alpha(s, X)dB_s$$

for some Brownian motion B_t. Therefore,

$$X_t = X_0 + \int_0^t \alpha(s, X)dB_s + \int_0^t \beta(s, X)ds.$$

To summarize, we have proved the following result.

Theorem 6.4. *Let μ be a probability measure on $\mathbb{R}^n$. Then the SDE (6.1) has a weak solution with initial distribution μ if and only if there exists a continuous, $\{\mathcal{F}_t\}$-adapted process n-dimensional process X_t defined on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ which satisfies the usual conditions, such that X_0 has distribution μ, and (6.11) and (6.15) hold for every $f \in C_b^2(\mathbb{R}^n)$.*

There is yet a more intrinsic way to formulate the martingale characterization described in Theorem 6.4. Recall that $(W^n, \mathcal{B}(W^n))$ is the continuous path space over $\mathbb{R}^n$, and $\{\mathcal{B}_t(W^n)\}$ is the natural filtration of the coordinate process.

Theorem 6.5. *Let μ be a probability measure on $\mathbb{R}^n$. Then the SDE (6.1) has a weak solution with initial distribution μ if and only if there exists a probability measure $\mathbb{P}^\mu$ on $(W^n, \mathcal{B}(W^n))$, such that:*

(1) $\mathbb{P}^\mu(w_0 \in \Gamma) = \mu(\Gamma)$ for every $\Gamma \in \mathcal{B}(\mathbb{R}^n)$;

(2) $\mathbb{P}^\mu$-almost surely, we have

$$\int_0^t \left(\|\alpha(s,w)\|^2 + \|\beta(s,w)\|\right) ds < \infty, \quad \forall t \geqslant 0;$$

(3) for every $f \in C_b^2(\mathbb{R}^n)$, under $\mathbb{P}^\mu$ we have

$$f(w_\cdot) - f(w_0) - \int_0^\cdot (\mathcal{A}f)(s,w)ds \in \mathcal{M}_0^{\mathrm{loc}}(\mathcal{H}_t(W^n)),$$

where $\{\mathcal{H}_t(W^n)\}$ is the usual augmentation of $\{\mathcal{B}_t(W^n)\}$ under $\mathbb{P}^\mu$.

Proof. Sufficiency is already proved before.

Now we consider necessity. Suppose that X_t is a continuous, $\{\mathcal{F}_t\}$-adapted process on some $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ satisfying the usual conditions, such that X_0 has distribution μ, and (6.11) and (6.15) hold for every $f \in C_b^2(\mathbb{R}^n)$. Consider the distribution $\mathbb{P}^X$ of X on $(W^n, \mathcal{B}(W^n))$. Apparently, (1) and (2) are satisfied for $\mathbb{P}^X$. To see (3), given $f \in C_b^2(\mathbb{R}^n)$, define

$$\sigma_R \triangleq \inf\left\{t \geqslant 0 : \left|\int_0^t (\mathcal{A}f)(s,X)ds\right| > R\right\}$$

on Ω, and consider the stopped process

$$Y_t \triangleq f(X_{\sigma_R \wedge t}) - f(X_0) - \int_0^{\sigma_R \wedge t} (\mathcal{A}f)(s,X)ds.$$

Y_t is indeed an $\{\mathcal{F}_t\}$-martingale since it is uniformly bounded. Correspondingly, define

$$\tau_R \triangleq \inf\left\{t \geqslant 0 : \left|\int_0^t (\mathcal{A}f)(s,w)ds\right| > R\right\}$$

on W^n, and consider the stopped process

$$z_t \triangleq f(w_{\tau_R\wedge t}) - f(w_0) - \int_0^{\tau_R\wedge t}(\mathcal{A}f)(s,w)ds$$

which is also uniformly bounded. Now a crucial observation is that σ_R and Y are determined by X pathwisely. Therefore, for $\Lambda \in \mathcal{H}_s(W^n)$, since $\{\mathcal{F}_t\}$ satisfies the usual conditions, we have $X^{-1}\Lambda \in \mathcal{F}_s$, and

$$\int_\Lambda z_t d\mathbb{P}^X = \int_{X^{-1}\Lambda} Y_t d\mathbb{P} = \int_{X^{-1}\Lambda} Y_s d\mathbb{P} = \int_\Lambda z_s d\mathbb{P}^X.$$

This shows that z_t is an $\{\mathcal{H}_t(W^n)\}$-martingale under $\mathbb{P}^X$. As $\tau_R \uparrow \infty$, we conclude that (3) holds. □

Remark 6.3. If we assume that the coefficients α, β are bounded, then all the relevant local martingale properties in the previous discussion become true martingale properties, and $\int_0^t(\mathcal{A}f)(s,w)ds$ is finite for every $(t,w) \in [0,\infty)\times W^n$. In this case, we do not need to pass to the usual augmentation, and (3) is equivalent to the following statement: for every $f \in C_b^2(\mathbb{R}^n)$, the process

$$m_t^f \triangleq f(w_t) - f(w_0) - \int_0^t (\mathcal{A}f)(s,w)ds$$

is a $\{\mathcal{B}_t(W^n)\}$-martingale under $\mathbb{P}^\mu$. Indeed, the only missing gap is perhaps the fact that m_t^f is a $\{\mathcal{B}_t(W^n)\}$-martingale if and only if it is an $\{\mathcal{H}_t(W^n)\}$-martingale, which can be shown easily by using the discrete backward martingale convergence theorem.

In view of Theorem 6.5, there are lots of advantages working on the path space. For instance, it has a nice metric structure, and we can apply the powerful tools of weak convergence and regular conditional expectations. The search of a probability measure $\mathbb{P}^\mu$ on $(W^n, \mathcal{B}(W^n))$ satisfying (1), (2), (3) in Theorem 6.5 is known as the martingale problem.

As a byproduct, we have indeed proved the following nice result.

Corollary 6.1. *(1) A continuous, $\{\mathcal{F}_t\}$-adapted process X_t on $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ satisfying the usual conditions in an $\{\mathcal{F}_t\}$-Brownian motion if and only if $X_0 = 0$ almost surely and*

$$f(X_t) - f(X_0) - \frac{1}{2}\int_0^t (\Delta f)(X_s)ds$$

is an $\{\mathcal{F}_t\}$-martingale for every $f \in C_b^2(\mathbb{R}^n)$.

(2) A probability measure $\mathbb{P}$ on $(W^n, \mathcal{B}(W^n))$ is the n-dimensional Wiener measure if and only if $\mathbb{P}(w_0 = 0) = 1$ and

$$f(w_t) - f(w_0) - \frac{1}{2}\int_0^t (\Delta f)(w_s)ds$$

is an $\{\mathcal{B}_t(W^n)\}$-martingale under $\mathbb{P}$ for every $f \in C_b^2(\mathbb{R}^n)$.

By the same reasoning, it is easy to show that uniqueness in law holds if and only if for every probability measure μ on $\mathbb{R}^n$, there exists at most one probability measure on $(W^n, \mathcal{B}(W^n))$ which satisfies (1), (2), (3) in Theorem 6.5. Remark 6.3 also applies for the uniqueness.

We will appreciate the power of the martingale characterization of weak existence in the following general result.

Theorem 6.6. *Suppose that α, β satisfy Assumption 6.1, and they are bounded and continuous. Then for any probability measure μ on $\mathbb{R}^n$ with compact support, the SDE (6.1) has a weak solution with initial distribution μ.*

Proof. For $m \geqslant 1$, define $\alpha_m : [0,\infty) \times W^n \to \mathrm{Mat}(n,d)$ by $\alpha_m(t,w) \triangleq \alpha(\phi_m(t), w)$, where $\phi_m(t)$ is the unique dyadic partition point $k/2^m$ such that $k/2^m \leqslant t < (k+1)/2^m$. Define β_m similarly. Now we construct a weak solution to the SDE with coefficients α_m, β_m with initial distribution μ explicitly.

Let $((\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\}), \xi, B_t)$ be a set-up in which ξ has distribution μ. Define a process $X_t^{(m)}$ inductively in the following way. Set $X_0^{(m)} \triangleq \xi$. If $X_t^{(m)}$ is defined for $t \leqslant k/2^m$, then for $t \in [k/2^m, (k+1)/2^m]$, define

$$X_t^{(m)} \triangleq X_{k/2^m}^{(m)} + \alpha\left(\frac{k}{2^m}, X^{(m,k)}\right)(B_t - B_{k/2^m}) + \beta\left(\frac{k}{2^m}, X^{(m,k)}\right)\left(t - \frac{k}{2^m}\right),$$

where $X^{(m,k)}$ is the stopped process defined by

$$X_t^{(m,k)} \triangleq \begin{cases} X_t^{(m)}, & t \leqslant k/2^m; \\ X_{k/2^m}^{(m)}, & t > k/2^m. \end{cases}$$

It follows from Remark 6.1 that

$$\alpha\left(\frac{k}{2^m}, X^{(m,k)}\right) = \alpha_m\left(t, X^{(m)}\right), \quad \beta\left(\frac{k}{2^m}, X^{(m,k)}\right) = \beta_m\left(t, X^{(m)}\right),$$

provided that $t \in [k/2^m, (k+1)/2^m]$. In particular, we conclude that

$$X_t^{(m)} = \xi + \int_0^t \alpha_m\left(s, X^{(m)}\right) dB_s + \int_0^t \beta_m\left(s, X^{(m)}\right) ds, \quad t \geqslant 0.$$

In other words, $X_t^{(m)}$ is a weak solution to the SDE with coefficients α_m, β_m with initial distribution μ. Now define $\mathbb{P}^{(m)}$ to be the distribution of $X^{(m)}$ on $(W^n, \mathcal{B}(W^n))$. According to Theorem 6.5 and Remark 6.3, we know that for given $f \in C_b^2(\mathbb{R}^n)$, the process

$$f(w_t) - f(w_0) - \int_0^t (\mathcal{A}_m f)(u, w) ds$$

is a $\{\mathcal{B}_t(W^n)\}$-martingale under $\mathbb{P}^{(m)}$, where the differential operator $\mathcal{A}_m$ is defined by (6.14) in terms of the coefficients α_m, β_m.

In addition, given constants $\gamma, p \geqslant 1$, we have

$$\sup_{m\geqslant 1}\mathbb{E}\left[\left|X_0^{(m)}\right|^\gamma\right]=\mathbb{E}[|\xi|^\gamma]\leqslant C_\gamma,$$

and by the BDG inequalities, we have

$$\sup_{m\geqslant 1}\mathbb{E}\left[\left|X_t^{(m)}-X_s^{(m)}\right|^{2p}\right]\leqslant C_{T,p}|t-s|^p,\quad \forall s,t\in[0,T]. \tag{6.18}$$

According to Problem Sheet 3, Problem 3, (3) (we take $p=2$ in (6.18)), we conclude that $\{\mathbb{P}^{(m)}\}$ is tight. Without loss of generality, we may assume that $\mathbb{P}^{(m)}$ converges weakly to some probability measure $\mathbb{P}$.

In view of Theorem 6.5 and Remark 6.3 again, it suffices to show that $\mathbb{P}(w_0\in\Gamma)=\mu(\Gamma)$ for $\Gamma\in\mathcal{B}(\mathbb{R}^n)$ (which is trivial), and for every $f\in C_b^2(\mathbb{R}^n)$, the process

$$f(w_t)-f(w_0)-\int_0^t(\mathcal{A}f)(u,w)du$$

is a $\{\mathcal{B}_t(W^n)\}$-martingale under $\mathbb{P}$.

Indeed, let $s<t$, and $\Phi(w)=\varphi(w_{s_1},\cdots,w_{s_k})$ for some $s_1<\cdots<s_k\leqslant s$ and $\varphi\in C_b(\mathbb{R}^{n\times k})$. Then from the $\mathbb{P}^{(m)}$-martingale property, we know that

$$\int_{W^n}\Phi(w)\cdot\left(f(w_t)-f(w_s)-\int_s^t(\mathcal{A}_mf)(u,w)du\right)d\mathbb{P}^{(m)}=0$$

for every m. To simplify our notation, set

$$\zeta_{s,t}^m(w)\triangleq f(w_t)-f(w_s)-\int_s^t(\mathcal{A}_mf)(u,w)du$$

and

$$\zeta_{s,t}(w)\triangleq f(w_t)-f(w_s)-\int_s^t(\mathcal{A}f)(u,w)du$$

respectively. From the tightness of $\{\mathbb{P}^{(m)}\}$, given $\varepsilon>0$, there exists a compact set $K\subseteq W^n$, such that

$$\mathbb{P}^{(m)}(K^c)<\varepsilon,\quad\forall m\geqslant 1.$$

By the definition of α_m,β_m and the uniform continuity of α,β on $[s,t]\times K$, when m is large, we have

$$\sup_{w\in K}|\zeta_{s,t}^m(w)-\zeta_{s,t}(w)|<\varepsilon.$$

Therefore, when m is large,

$$\left|\int_{W^n}\Phi(w)\zeta_{s,t}^m(w)d\mathbb{P}^{(m)}-\int_{W^n}\Phi(w)\zeta_{s,t}(w)d\mathbb{P}^{(m)}\right|\leqslant C\left(\mathbb{P}^{(m)}(K^c)+\varepsilon\right)<2C\varepsilon.$$

On the other hand, by weak convergence we know that

$$\lim_{m\to\infty}\int_{W^n}\Phi(w)\zeta_{s,t}(w)d\mathbb{P}^{(m)}=\int_{W^n}\Phi(w)\zeta_{s,t}(w)d\mathbb{P}.$$

Since ε is arbitrary, we conclude that

$$\int_{W^n}\Phi(w)\zeta_{s,t}(w)d\mathbb{P}=0,$$

which implies the desired $\mathbb{P}$-martingale property. □

Remark 6.4. The assumption that μ is compactly supported in Theorem 6.6 is just for technical convenience. Indeed, we have shown that for every $x\in\mathbb{R}^n$, there exists $\mathbb{P}^x$ which solves the martingale problem with initial distribution δ_x. For a general probability measure μ on $\mathbb{R}^n$, we can simply define

$$\mathbb{P}^{\mu}(\Lambda)\triangleq\int_{\mathbb{R}^n}\mathbb{P}^x(\Lambda)\mu(dx),\quad \Lambda\in\mathcal{B}(W^n). \tag{6.19}$$

This $\mathbb{P}^{\mu}$ will then solve the martingale problem with initial distribution μ. Of course here a rather subtle point is whether $x\mapsto\mathbb{P}^x$ is measurable before making sense of (6.19). This is not true in general, but we can always *select* $\mathbb{P}^x$ so that the map $x\mapsto\mathbb{P}^x$ is measurable. This selection theorem is rather deep and technical, and we will not get into the details.

From now on, we will restrict ourselves to a special but very important type of SDEs:

$$dX_t=\sigma(X_t)dB_t+b(X_t)dt, \tag{6.20}$$

where $\sigma:\ \mathbb{R}^n\to\mathrm{Mat}(n,d)$ and $b:\ \mathbb{R}^n\to\mathbb{R}^n$. This type of SDEs is usually known as *time homogeneous Markovian type*, and it is closely related to the study of diffusion processes. In particular, we are going to develop a relatively complete solution theory along the line of Yamada and Watanabe's philosophy.

Of course the general weak existence theorem (c.f. Theorem 6.6) that we just proved covers this special case (with $\alpha(t,w)=\sigma(w_t)$, $\beta(t,w)=b(w_t)$). However, in general it is not natural to assume uniform boundedness on the coefficients. And it is not so clear how a localization argument should yield weak existence without the boundedness assumption. Indeed, if we do not assume boundedness, the solution can possibly explode in finite amount of time. Therefore, it is a good idea to start with this general situation independently, and then to explore under what conditions on the coefficients will a solution be globally defined in time without explosion. Uniform boundedness will be too strong to assume and not so satisfactory.

To include the possibility of explosion, we take $\varDelta$ to be some given point outside $\mathbb{R}^n$ which captures the explosion. Define $\widehat{\mathbb{R}}^n\triangleq\mathbb{R}^n\cup\{\varDelta\}$. Topologically, $\widehat{\mathbb{R}}^n$ is the one-point compactification of $\mathbb{R}^n$. In particular, $\widehat{\mathbb{R}}^n$ is homeomorphic to the n-sphere S^n. In this sphere model, $\varDelta$ corresponds to the north pole N, and $\mathbb{R}^n$ is homeomorphic to $S^n\backslash\{N\}$.

We consider the following continuous path space over $\widehat{\mathbb{R}}^n$. Let $\widehat{W}^n$ be the space of continuous paths $w: [0,\infty) \to \widehat{\mathbb{R}}^n$, such that if $w_t = \Delta$, then $w_{t'} = \Delta$ for all $t' \geqslant t$. The Borel σ-algebra $\mathcal{B}(\widehat{W}^n)$ is the σ-algebra generated by cylinder sets in the usual way. For each $w \in \widehat{W}^n$, we can define an intrinsic quantity

$$e(w) \triangleq \inf\{t \geqslant 0 : \ w_t = \Delta\}.$$

$e(w)$ is called the *explosion time* of w.

Unless otherwise stated, we assume exclusively that the coefficients σ, b are continuous. Note that σ, b are defined on $\mathbb{R}^n$ instead of on $\widehat{\mathbb{R}}^n$.

Definition 6.8. Let μ be a probability measure on $\mathbb{R}^n$. We say that the SDE (6.20) *has a weak solution with initial distribution* μ if there exists a set-up $((\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\}), \xi, B_t)$ together with a continuous, $\{\mathcal{F}_t\}$-adapted process X_t in $\widehat{\mathbb{R}}^n$, such that:

(1) ξ has distribution μ;

(2) if $e(\omega) \triangleq e(X(\omega))$ is the explosion time of $X(\omega) \in \widehat{W}^n$, then we have

$$X_t = \xi + \int_0^t \sigma(X_s)dB_s + \int_0^t b(X_s)ds, \quad t \in [0, e). \tag{6.21}$$

Remark 6.5. The stochastic integral in (6.21) is defined in the following way. Let $\sigma_m \triangleq \inf\{t \geqslant 0 : \ |X_t| \geqslant m\}$. From the continuity of σ, we know that for each fixed $m \geqslant 1$, the process $\sigma(X_t)\mathbf{1}_{[0,\sigma_m]}(t)$ is uniformly bounded, and hence the stochastic integral

$$I_t^{(m)} \triangleq \int_0^t \sigma(X_s)\mathbf{1}_{[0,\sigma_m]}(s)dB_s$$

is well-defined on $[0,\infty)$. Moreover, by stopping we see that $I^{(m+1)} = I^{(m)}$ on $[0, \sigma_m]$. Therefore, since $\sigma_m \uparrow e$, we can define a single process I_t on $[0, e)$, such that $I_t = I_t^{(m)}$ if $t < \sigma_m \leqslant e$. This I_t is our stochastic integral in (6.21).

Similarly, the notions of exactness, uniqueness in law, pathwise uniqueness, and unique strong solution carry through without much difficulty. In particular, the Yamada-Watanabe theorem remains true in this setting. Moreover, the martingale characterization is also valid as long as we localize in the same way as Remark 6.5 when we describe the martingale property.

To establish a solution theory for the SDE (6.20) in the spirit of Yamada and Watanabe, we first take up the question about weak existence. We have the following main result.

Theorem 6.7. *Suppose that the coefficients σ, b are continuous. Then for any probability measure μ on $\mathbb{R}^n$ with compact support, the SDE (6.20) has a weak solution with initial distribution μ.*

The most convenient way to prove this result is to use the martingale characterization in the sense of Theorem 6.4. To this end, we need to show that, on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ satisfying the usual conditions, there exists a continuous, $\{\mathcal{F}_t\}$-adapted process X_t in $\widehat{\mathbb{R}}^n$, such that:

(1) $\mathbb{P}(X_0 \in \Gamma) = \mu(\Gamma)$ for $\Gamma \in \mathcal{B}(\mathbb{R}^n)$;

(2) for every $f \in C_b^2(\mathbb{R}^n)$ and $m \geqslant 1$, the process

$$f(X_{\sigma_m \wedge t}) - f(X_0) - \int_0^{\sigma_m \wedge t} (\mathcal{A}f)(X_s)ds$$

is an $\{\mathcal{F}_t\}$-martingale, where the differential operator $\mathcal{A}$ is defined by (6.14) in terms of the coefficients σ, b, and

$$\sigma_m \triangleq \inf\{t \geqslant 0 : \ |X_t| \geqslant m\}. \tag{6.22}$$

The idea of proving Theorem 6.7 is to obtain X as the time-change of some $\widetilde{X}$ which is a weak solution to some SDE with bounded coefficients.

For this purpose, we choose a function $\rho(x)$ on $\mathbb{R}^n$ such that $0 < \rho(x) \leqslant 1$ for every $x \in \mathbb{R}^n$, and $\rho(x)a(x), \rho(x)b(x)$ are both bounded, where $a(x) \triangleq \sigma(x)\sigma(x)^*$. It is not hard to see that such ρ exists. Consider the differential operator defined by $(\widetilde{\mathcal{A}}f)(x) \triangleq \rho(x)(\mathcal{A}f)(x)$ for $f \in C_b^2(\mathbb{R}^n)$. According to Theorem 6.6 and the martingale characterization, there exists a continuous, $\{\widetilde{\mathcal{F}}_t\}$-adapted process $\widetilde{X}_t$ in $\mathbb{R}^n$ defined on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\widetilde{\mathcal{F}}_t\})$ which satisfies the usual conditions, such that $\widetilde{X}_0$ has distribution μ, and for every $f \in C_b^2(\mathbb{R}^n)$,

$$f(\widetilde{X}_t) - f(\widetilde{X}_0) - \int_0^t (\widetilde{\mathcal{A}}f)(\widetilde{X}_s)ds$$

is an $\{\widetilde{\mathcal{F}}_t\}$-martingale.

Consider the strictly increasing process

$$A_t \triangleq \int_0^t \rho(\widetilde{X}_s)ds$$

and define

$$e \triangleq \int_0^\infty \rho(\widetilde{X}_s)ds.$$

Let C_t be the time-change associated with A_t, so that $C_t < \infty$ if and only if $t < e$. Define $\mathcal{F}_t \triangleq \widetilde{\mathcal{F}}_{C_t}$, and

$$X_t \triangleq \begin{cases} \widetilde{X}_{C_t}, & t < e; \\ \Delta, & t \geqslant e. \end{cases}$$

The key ingredient of the proof is to show that e is the explosion time of X. We assume this is true for the moment and postpone its proof for a little while.

Now we show that X_t satisfies the desired martingale characterization on $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ for the differential operator $\mathcal{A}$. Recall that σ_m is defined by (6.22). Set $\widetilde{\sigma}_m \triangleq \inf\{t \geqslant$

$0: |\widetilde{X}_t| \geqslant m\}$. It follows that $C_{\sigma_m} = \widetilde{\sigma}_m$, and thus $C_{\sigma_m \wedge t} = \widetilde{\sigma}_m \wedge C_t$ for all $t \geqslant 0$. Now we know that

$$f(\widetilde{X}_{\widetilde{\sigma}_m \wedge t}) - f(\widetilde{X}_0) - \int_0^{\widetilde{\sigma}_m \wedge t} (\rho\mathcal{A}f)(\widetilde{X}_s)ds$$

is a bounded $\{\widetilde{\mathcal{F}}_t\}$-martingale. According to the optional sampling theorem,

$$\begin{aligned} & f(\widetilde{X}_{\widetilde{\sigma}_m \wedge C_t}) - f(\widetilde{X}_0) - \int_0^{\widetilde{\sigma}_m \wedge C_t} (\rho\mathcal{A}f)(\widetilde{X}_s)ds \\ = & f(\widetilde{X}_{C_{\sigma_m \wedge t}}) - f(\widetilde{X}_0) - \int_0^{C_{\sigma_m \wedge t}} (\rho\mathcal{A}f)(\widetilde{X}_s)ds \\ = & f(X_{\sigma_m \wedge t}) - f(X_0) - \int_0^{C_{\sigma_m \wedge t}} (\rho\mathcal{A}f)(\widetilde{X}_s)ds \end{aligned}$$

is an $\{\mathcal{F}_t\}$-martingale. But a change of variables $s = C_u$ ($u = A_s$) together with the definition of A_s yields immediately that

$$\int_0^{C_{\sigma_m \wedge t}} (\rho\mathcal{A}f)(\widetilde{X}_s)ds = \int_0^{\sigma_m \wedge t} (\mathcal{A}f)(X_u)du.$$

Therefore, we have the desired martingale characterization property for X_t on $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$.

Now it remains to prove the following key lemma.

Lemma 6.3. *With probability one, if* $e(\omega) < \infty$, *then* $\lim_{t\uparrow e} X_t = \Delta$.

Proof. It is equivalent to showing that, with probability one, if $\int_0^\infty \rho(\widetilde{X}_s)ds < \infty$, then $\lim_{t\to\infty} \widetilde{X}_t = \Delta$ in $\widehat{\mathbb{R}}^n$. This is not surprising to expect. Indeed, since $\rho > 0$, we know that the minimum of ρ on any compact subset of $\mathbb{R}^n$ is strictly positive. Therefore, if $\widetilde{X}_t$ spends too much time being trapped inside a compact set, then the integral $\int_0^\infty \rho(\widetilde{X}_s)ds$ will have a high chance of being infinity.

Now fix $r < R$ such that $|\widetilde{X}_0| < r$ almost surely. The key point is to demonstrate that, with probability one, if $\int_0^\infty \rho(\widetilde{X}_s)ds < \infty$, then after exiting the R-ball and coming back into the r-ball for at most finitely many times, $\widetilde{X}_t$ will stay outside the r-ball forever. As r can be arbitrarily large, this shows that $\widetilde{X}_t$ has to explode to infinity as $t \to \infty$.

To be precise, define

$$\begin{aligned} \widetilde{\sigma}_1 &\triangleq 0, & \widetilde{\tau}_1 &\triangleq \inf\{t \geqslant \widetilde{\sigma}_1 : |\widetilde{X}_t| \geqslant R\}, \\ \widetilde{\sigma}_2 &\triangleq \inf\{t \geqslant \widetilde{\tau}_1 : |\widetilde{X}_t| \leqslant r\}, & \widetilde{\tau}_2 &\triangleq \inf\{t \geqslant \widetilde{\sigma}_2 : |\widetilde{X}_t| \geqslant R\}, \\ &\cdots & &\cdots. \end{aligned}$$

We want to show that, with probability one,

$$\left\{ \int_0^\infty \rho(\widetilde{X}_s)ds < \infty \right\} \subseteq \{\exists m \geqslant 1, \text{ s.t. } \widetilde{\tau}_m < \infty \text{ and } \widetilde{\sigma}_{m+1} = \infty\}. \qquad (6.23)$$

This is equivalent to showing that, with probability one,

$$\bigcap_{m=1}^{\infty}\{\widetilde{\tau}_m=\infty \text{ or } \widetilde{\sigma}_{m+1}<\infty\}\subseteq\left\{\int_0^\infty \rho(\widetilde{X}_s)ds=\infty\right\}. \tag{6.24}$$

Observe that the left hand side of (6.24) is equal to the event that

$$\{\exists m\geqslant 1, \text{ s.t. } \widetilde{\sigma}_m<\infty \text{ and } \widetilde{\tau}_m=\infty\}\bigcup\{\widetilde{\sigma}_m<\infty\ \forall m\geqslant 1\}.$$

Therefore, we need to show that with probability one, this event triggers $\int_0^\infty \rho(\widetilde{X}_s)ds=\infty$.

Case one. Suppose that there exists $m\geqslant 1$, such that $\widetilde{\sigma}_m<\infty$ but $\widetilde{\tau}_m=\infty$. Then $|\widetilde{X}_t|\leqslant R$ for all $t\geqslant\widetilde{\sigma}_m$. Since $\min_{|x|\leqslant R}\rho(x)>0$, we have

$$\int_0^\infty \rho(\widetilde{X}_s)ds\geqslant\int_{\widetilde{\sigma}_m}^\infty \rho(\widetilde{X}_s)ds\geqslant\left(\min_{|x|\leqslant R}\rho(x)\right)\cdot\int_{\widetilde{\sigma}_m}^\infty ds=\infty.$$

Case two. Suppose that $\widetilde{\sigma}_m<\infty$ for every $m\geqslant 1$. In this case, we only need to show that

$$\sum_{m=1}^\infty(\widetilde{\tau}_m-\widetilde{\sigma}_m)=\infty. \tag{6.25}$$

Indeed, if this is true, then

$$\int_0^\infty \rho(\widetilde{X}_s)ds\geqslant\sum_{m=1}^\infty\int_{\widetilde{\sigma}_m}^{\widetilde{\tau}_m}\rho(\widetilde{X}_s)ds\geqslant\left(\min_{|x|\leqslant R}\rho(x)\right)\sum_{m=1}^\infty(\widetilde{\tau}_m-\widetilde{\sigma}_m)=\infty.$$

Observe that (6.25) is equivalent to showing that with probability one,

$$\left(\prod_{m=1}^\infty \mathbf{1}_{\{\widetilde{\sigma}_m<\infty\}}\right)\mathrm{e}^{-\sum_{m=1}^\infty(\widetilde{\tau}_m-\widetilde{\sigma}_m)}=\prod_{m=1}^\infty\left(\mathbf{1}_{\{\widetilde{\sigma}_m<\infty\}}\mathrm{e}^{-(\widetilde{\tau}_m-\widetilde{\sigma}_m)}\right)=0,$$

which is also equivalent to showing that

$$\mathbb{E}\left[\prod_{m=1}^\infty\left(\mathbf{1}_{\{\widetilde{\sigma}_m<\infty\}}\mathrm{e}^{-(\widetilde{\tau}_m-\widetilde{\sigma}_m)}\right)\right]=0. \tag{6.26}$$

We write

$$\begin{aligned}&\mathbb{E}\left[\prod_{k=1}^{m+1}\left(\mathbf{1}_{\{\widetilde{\sigma}_k<\infty\}}\mathrm{e}^{-(\widetilde{\tau}_k-\widetilde{\sigma}_k)}\right)|\widetilde{\mathcal{F}}_{\widetilde{\sigma}_{m+1}}\right]\\&=\prod_{k=1}^{m}\left(\mathbf{1}_{\{\widetilde{\sigma}_k<\infty\}}\mathrm{e}^{-(\widetilde{\tau}_k-\widetilde{\sigma}_k)}\right)\cdot\mathbf{1}_{\{\widetilde{\sigma}_{m+1}<\infty\}}\mathbb{E}\left[\mathrm{e}^{-(\widetilde{\tau}_{m+1}-\widetilde{\sigma}_{m+1})}|\widetilde{\mathcal{F}}_{\widetilde{\sigma}_{m+1}}\right].\end{aligned} \tag{6.27}$$

Now we estimate the quantity $\mathbf{1}_{\{\widetilde{\sigma}_{m+1}<\infty\}}\mathbb{E}\left[\mathrm{e}^{-(\widetilde{\tau}_{m+1}-\widetilde{\sigma}_{m+1})}|\widetilde{\mathcal{F}}_{\widetilde{\sigma}_{m+1}}\right]$. In particular, we are going to show that this quantity is bounded by some deterministic constant $\gamma<1$, which depends only on R, r and $\widetilde{\mathcal{A}}$. If this is true, then by taking expectation on both sides of (6.27), we get (6.26) immediately.

Let

$$M^i_\cdot \triangleq \widetilde{X}^i_\cdot - \widetilde{X}^i_0 - \int_0^\cdot \rho(\widetilde{X}_s)b^i(\widetilde{X}_s)ds \in \mathcal{M}_0^{\mathrm{loc}}(\{\widetilde{\mathcal{F}}_t\})$$

for each i (the reader may recall from the proof of Theorem 6.4). According to Itô's formula, we may write

$$|\widetilde{X}_t|^2 = |\widetilde{X}_0|^2 + N_t + A_t$$

in semimartingale form, where

$$\begin{aligned} N_t &= 2\sum_{i=1}^n \int_0^t \widetilde{X}^i_s dM^i_s, \\ A_t &= \sum_{i=1}^n \left(2\int_0^t \widetilde{X}^i_s \rho(\widetilde{X}_s)b^i(\widetilde{X}_s)ds + \langle M^i\rangle_t\right). \end{aligned}$$

Since $\rho\cdot a$ and $\rho\cdot b$ are both bounded, by the definition of $\widetilde{\sigma}_{m+1}$, it is not hard to see that there exists a constant $C>0$ depending only on $\widetilde{\mathcal{A}}$, such that on $\{\widetilde{\sigma}_{m+1}<\infty\}$, for every $0\leqslant t\leqslant \widetilde{\tau}_{m+1}-\widetilde{\sigma}_{m+1}$, we have

$$\begin{aligned} \langle N\rangle_{\widetilde{\sigma}_{m+1}+t} - \langle N\rangle_{\widetilde{\sigma}_{m+1}} &\leqslant CR^2 t, \\ \left|A_{\widetilde{\sigma}_{m+1}+t} - A_{\widetilde{\sigma}_{m+1}}\right| &\leqslant C(2R+1)t. \end{aligned}$$

If $\widetilde{\tau}_{m+1}<\infty$, then we know that

$$|\widetilde{X}_{\widetilde{\tau}_{m+1}}| = R,\ |\widetilde{X}_{\widetilde{\sigma}_{m+1}}| = r.$$

Therefore,

$$|N_{\widetilde{\tau}_{m+1}} - N_{\widetilde{\sigma}_{m+1}}| \geqslant \frac{R^2-r^2}{2} \ \text{ or } \ |A_{\widetilde{\tau}_{m+1}} - A_{\widetilde{\sigma}_{m+1}}| \geqslant \frac{R^2-r^2}{2}.$$

In particular, if we define

$$\theta \triangleq \inf\left\{t\geqslant 0:\ |N_{\widetilde{\sigma}_{m+1}+t} - N_{\widetilde{\sigma}_{m+1}}| \geqslant \frac{R^2-r^2}{2}\right\},$$

then

$$\widetilde{\tau}_{m+1} - \widetilde{\sigma}_{m+1} \geqslant \theta \wedge \frac{R^2-r^2}{2C(2R+1)}.$$

In addition, according to the generalized Dambis-Dubins-Schwarz theorem (c.f. Theorem 5.9), there exists an $\{\widehat{\mathcal{F}}_t\}$-Brownian motion defined possibly on some enlargement of the underlying filtered probability space, such that $N_t = B_{\langle N\rangle_t}$, where $\widehat{\mathcal{F}}_t \triangleq \widetilde{\mathcal{F}}_{D_t}$ and D_t is the time-change associated with $\langle N\rangle_t$. Let

$$\eta \triangleq \inf\left\{u \geqslant 0: \ \left|B_{\langle N\rangle_{\widetilde{\sigma}_{m+1}}+u} - B_{\langle N\rangle_{\widetilde{\sigma}_{m+1}}}\right| \geqslant \frac{R^2-r^2}{2}\right\}.$$

It follows that

$$\eta \leqslant \langle N\rangle_{\widetilde{\sigma}_{m+1}+\theta} - \langle N\rangle_{\widetilde{\sigma}_{m+1}} \leqslant CR^2\theta.$$

Therefore, we obtain that on $\{\widetilde{\sigma}_{m+1} < \infty\}$,

$$\widetilde{\tau}_{m+1} - \widetilde{\sigma}_{m+1} \geqslant \frac{\eta}{CR^2} \wedge \frac{R^2-r^2}{2C(2R+1)}.$$

It follows that

$$\mathbf{1}_{\{\widetilde{\sigma}_{m+1}<\infty\}}\mathbb{E}\left[\mathrm{e}^{-(\widetilde{\tau}_{m+1}-\widetilde{\sigma}_{m+1})}|\widetilde{\mathcal{F}}_{\widetilde{\sigma}_{m+1}}\right] \leqslant \mathbf{1}_{\{\widetilde{\sigma}_{m+1}<\infty\}}\mathbb{E}\left[\mathrm{e}^{-\frac{\eta}{CR^2}\wedge\frac{R^2-r^2}{2C(2R+1)}}\Big|\widetilde{\mathcal{F}}_{\widetilde{\sigma}_{m+1}}\right].$$

In addition, observe that $\widetilde{\mathcal{F}}_{\widetilde{\sigma}_{m+1}} \subseteq \widehat{\mathcal{F}}_{\langle N\rangle_{\widetilde{\sigma}_{m+1}}}$. Indeed, if $A \in \widetilde{\mathcal{F}}_{\widetilde{\sigma}_{m+1}}$, according to Proposition 2.4, we know that

$$A \cap \{\langle N\rangle_{\widetilde{\sigma}_{m+1}} \leqslant t\} = A \cap \{\widetilde{\sigma}_{m+1} \leqslant D_t\} \in \widetilde{\mathcal{F}}_{D_t} = \widehat{\mathcal{F}}_t,$$

therefore $A \in \widehat{\mathcal{F}}_{\langle N\rangle_{\widetilde{\sigma}_{m+1}}}$. It follows that

$$\begin{aligned} & \mathbf{1}_{\{\widetilde{\sigma}_{m+1}<\infty\}}\mathbb{E}\left[\mathrm{e}^{-\frac{\eta}{CR^2}\wedge\frac{R^2-r^2}{2C(2R+1)}}\Big|\widetilde{\mathcal{F}}_{\widetilde{\sigma}_{m+1}}\right] \\ = \ & \mathbf{1}_{\{\widetilde{\sigma}_{m+1}<\infty\}}\mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_{\{\langle N\rangle_{\widetilde{\sigma}_{m+1}}<\infty\}}\mathrm{e}^{-\frac{\eta}{CR^2}\wedge\frac{R^2-r^2}{2C(2R+1)}}\Big|\widehat{\mathcal{F}}_{\langle N\rangle_{\widetilde{\sigma}_{m+1}}}\right]\Big|\widetilde{\mathcal{F}}_{\widetilde{\sigma}_{m+1}}\right]. \end{aligned}$$

Now a crucial observation is that on $\{\langle N\rangle_{\widetilde{\sigma}_{m+1}} < \infty\}$, $B_{\langle N\rangle_{\widetilde{\sigma}_{m+1}+u}} - B_{\langle N\rangle_{\widetilde{\sigma}_{m+1}}}$ is a Brownian motion independent of $\widehat{\mathcal{F}}_{\langle N\rangle_{\widetilde{\sigma}_{m+1}}}$ by the strong Markov property. Therefore,

$$\mathbb{E}\left[\mathrm{e}^{-\frac{\eta}{CR^2}\wedge\frac{R^2-r^2}{2C(2R+1)}}\Big|\widehat{\mathcal{F}}_{\langle N\rangle_{\widetilde{\sigma}_{m+1}}}\right] = \mathbb{E}[\mathrm{e}^{-\frac{\tau}{CR^2}\wedge\frac{R^2-r^2}{2C(2R+1)}}] =: \gamma, \quad \text{on } \{\langle N\rangle_{\widetilde{\sigma}_{m+1}} < \infty\},$$

where τ is the hitting time of the level set $(R^2 - r^2)/2$ by a one dimensional Brownian motion. Apparently $\gamma < 1$, otherwise $\tau = 0$ which is absurd. Therefore, we arrive at

$$\mathbf{1}_{\{\widetilde{\sigma}_{m+1}<\infty\}}\mathbb{E}\left[\mathrm{e}^{-(\widetilde{\tau}_{m+1}-\widetilde{\sigma}_{m+1})}|\widetilde{\mathcal{F}}_{\widetilde{\sigma}_{m+1}}\right] \leqslant \gamma,$$

which concludes (6.26).

To summarize, we have show that with probability one, (6.23) holds. Since r, R are arbitrary, we conclude that with probability one, on $\left\{\int_0^\infty \rho(\widetilde{X}_s)ds < \infty\right\}$, $\widetilde{X}_t \to \Delta$ as $t \to \infty$. □

The next question is about non-explosion criteria. The following result shows that explosion will not happen if the coefficients have linear growth. This is compatible with Theorem 6.1 in Itô's classical theory.

Theorem 6.8. *Suppose that the coefficients σ, b are continuous, and satisfy the following linear growth condition: there exists some $K > 0$, such that*

$$\|\sigma(x)\| + \|b(x)\| \leqslant K(1 + |x|), \quad \forall x \in \mathbb{R}^n. \tag{6.28}$$

Then for any weak solution X_t to the SDE (6.20) with $\mathbb{E}[|X_0|^2] < \infty$, we have $\mathbb{E}[|X_t|^2] < \infty$ for all $t \geqslant 0$. In particular, $e = \infty$ almost surely.

Proof. Suppose that X_t is a weak solution on some set-up with $\mathbb{E}[|X_0|^2] < \infty$. Let $\sigma_m \triangleq \inf\{t \geqslant 0 : \ |X_t| \geqslant m\}$. Then for any $f \in C_b^2(\mathbb{R}^n)$, the process

$$f(X_{\sigma_m \wedge t}) - f(X_0) - \int_0^{\sigma_m \wedge t} (\mathcal{A}f)(X_s)ds$$

is a bounded martingale. In particular, if we choose $f \in C_b^2(\mathbb{R}^n)$ to be such that $f(x) = |x|^2$ when $|x| \leqslant m$, then by the martingale property and the condition (6.28), we have

$$\begin{aligned}
\mathbb{E}\left[|X_{\sigma_m \wedge t}|^2\right] &\leqslant \mathbb{E}[|X_0|^2] + \sum_{i=1}^n \mathbb{E}\left[\int_0^{\sigma_m \wedge t} \left(a^{ii}(X_s) + 2b^i(X_s)X_s^i\right) ds\right] \\
&\leqslant \mathbb{E}[|X_0|^2] + C_K \mathbb{E}\left[\int_0^{\sigma_m \wedge t} (1 + |X_s|^2) ds\right] \\
&\leqslant \mathbb{E}[|X_0|^2] + C_K \int_0^t \left(1 + \mathbb{E}[|X_{\sigma_m \wedge s}|^2]\right) ds.
\end{aligned}$$

Gronwall's inequality then implies that

$$\mathbb{E}[|X_{\sigma_m \wedge t}|^2] \leqslant \left(1 + \mathbb{E}[|X_0|^2]\right) \mathrm{e}^{C_K t} - 1.$$

By letting $m \to \infty$, we conclude that $e > t$ almost surely and $\mathbb{E}[|X_t|^2] < \infty$. Since t is arbitrary, we know that $e = \infty$ almost surely. □

Remark 6.6. By the same reason as in Remark 6.4, we can remove the compactness assumption on μ in Theorem 6.7. In addition, by Theorem 6.8, we know that every weak solution with initial distribution $\mu = \delta_x$ does not explode. By using the martingale formulation on the continuous path space $\widehat{W}^n$, we have

$$\mathbb{P}^x \left(\lim_{m \to \infty} \sigma_m = \infty \right) = 1, \quad \forall x \in \mathbb{R}^n,$$

where $\mathbb{P}^x$ is a solution to the martingale problem with initial distribution δ_x, and $\sigma_m \triangleq \inf\{w \in \widehat{W}^n : \ |w_t| \geqslant m\}$. For an arbitrary probability measure μ on $\mathbb{R}^n$, we define $\mathbb{P}^\mu$

by (6.19) as in Remark 6.4 (after suitable measurable selection on the family $\{\mathbb{P}^x\}$). It follows that

$$\mathbb{P}^\mu\left(\lim_{m\to\infty}\sigma_m=\infty\right)=1.$$

Therefore, $\mathbb{P}^\mu$ is a non-exploding solution to the martingale problem with initial distribution μ. In particular, a non-exploding weak solution with initial distribution μ exists.

6.4 Pathwise uniqueness results

Now we study pathwise uniqueness for the SDE (6.20). It is a standard result that (local) Lipschitz condition implies pathwise uniqueness.

Theorem 6.9. *Suppose that the coefficients σ, b are locally Lipschitz, i.e. for every $N \geqslant 1$, there exists $K_N > 0$, such that*

$$\|\sigma(x)-\sigma(y)\|+\|b(x)-b(y)\|\leqslant K_N|x-y|,\quad \forall x,y\in B_N, \tag{6.29}$$

where B_N is the Euclidean ball of radius N. Then pathwise uniqueness holds for the SDE (6.20).

Proof. Suppose that X_t, X_t' are two solutions to the SDE (6.20) on a given set-up $((\Omega,\mathcal{F},\mathbb{P};\{\mathcal{F}_t\}),\xi,B_t)$ with the same initial condition ξ. Define

$$\sigma_N\triangleq\inf\{t\geqslant0:\ |X_t|\geqslant N\},\quad \sigma_N'\triangleq\inf\{t\geqslant0:\ |X_t'|\geqslant N\},$$

respectively. Then we have

$$\begin{aligned}X_{\sigma_N\wedge\sigma_N'\wedge t}-X'_{\sigma_N\wedge\sigma_N'\wedge t} &= \int_0^{\sigma_N\wedge\sigma_N'\wedge t}\left(\sigma(X_s)-\sigma(X_s')\right)dB_s\\ &\quad+\int_0^{\sigma_N\wedge\sigma_N\wedge t}\left(b(X_s)-b(X_s')\right)ds. \end{aligned}\tag{6.30}$$

Therefore, given $T>0$, for every $t\in[0,T]$, we have

$$\begin{aligned}&\mathbb{E}\left[\left|X_{\sigma_N\wedge\sigma_N'\wedge t}-X'_{\sigma_N\wedge\sigma_N'\wedge t}\right|^2\right]\\ \leqslant\;& 2\mathbb{E}\left[\left|\int_0^{\sigma_N\wedge\sigma_N'\wedge t}\left(\sigma(X_s)-\sigma(X_s')\right)dB_s\right|^2+\left|\int_0^{\sigma_N\wedge\sigma_N\wedge t}\left(b(X_s)-b(X_s')\right)ds\right|^2\right]\\ \leqslant\;& 2\mathbb{E}\left[\int_0^{\sigma_N\wedge\sigma_N'\wedge t}\|\sigma(X_s)-\sigma(X_s')\|^2ds\right]+2T\mathbb{E}\left[\int_0^{\sigma_N\wedge\sigma_N'\wedge t}\|b(X_s)-b(X_s')\|^2ds\right]\\ \leqslant\;& 2\mathbb{E}\left[\int_0^t\|\sigma(X_{\sigma_N\wedge\sigma_N'\wedge s})-\sigma(X'_{\sigma_N\wedge\sigma_N'\wedge s})\|^2ds\right]\\ &+2T\mathbb{E}\left[\int_0^t\|b(X_{\sigma_N\wedge\sigma_N'\wedge s})-b(X'_{\sigma_N\wedge\sigma_N'\wedge s})\|^2ds\right]\\ \leqslant\;& 2K_N(1+T)\int_0^t\mathbb{E}\left[|X_{\sigma_N\wedge\sigma_N'\wedge s}-X'_{\sigma_N\wedge\sigma_N'\wedge s}|^2\right]ds.\end{aligned}$$

Since the function $t \in [0,T] \mapsto \mathbb{E}\left[|X_{\sigma_N\wedge\sigma'_N\wedge t} - X'_{\sigma_N\wedge\sigma'_N\wedge t}|^2\right]$ is non-negative and continuous, according to Gronwall's inequality, we conclude that

$$\mathbb{E}\left[|X_{\sigma_N\wedge\sigma'_N\wedge t} - X'_{\sigma_N\wedge\sigma'_N\wedge t}|^2\right] = 0$$

for every $t \in [0,T]$. As T is arbitrary, it follows that with probability one,

$$X_{\sigma_N\wedge\sigma'_N\wedge t} = X_{\sigma_N\wedge\sigma'_N\wedge t}, \quad \forall t \geqslant 0.$$

This implies that $X_t = X'_t$ on $[0, \sigma_N \wedge \sigma'_N)$. By the definition of σ_N and σ'_N, we must have $\sigma_N = \sigma'_N$. In particular, by letting $N \to \infty$, we conclude that with probability one, $e(X) = e(X')$ and $X_t = Y_t$ on $[0, e(X))$, where $e(X)$ and $e(Y)$ are the explosion times of X and Y respectively. □

The local Lipschitz condition can be weakened in the one dimensional case. In particular, pathwise uniqueness holds if σ is $1/2$-Hölder continuous and b is locally Lipschitz continuous.

Theorem 6.10. *Suppose that $n = 1$ and the coefficients σ, b are continuous. Assume further that the following two conditions hold:*

(1) there exists a strictly increasing function ρ on $[0,\infty)$ such that $\rho(0) = 0$, $\int_{0+}\rho^{-2}(u)du = \infty$, and $\|\sigma(x) - \sigma(y)\| \leqslant \rho(|x-y|)$ for all $x, y \in \mathbb{R}^1$;

(2) b is locally Lipschitz in the sense of (6.29).

Then pathwise uniqueness holds for the SDE (6.20).

Proof. According to Condition (1), we can find a sequence $0 < \cdots < a_n < a_{n-1} < \cdots < a_2 < a_1 < 1$ such that

$$\int_{a_1}^{1}\rho^{-2}(u)du = 1, \ \int_{a_2}^{a_1}\rho^{-2}(u)du = 2, \cdots, \ \int_{a_n}^{a_{n-1}}\rho^{-2}(u)du = n, \cdots.$$

Apparently $a_n \downarrow 0$ as $n \to \infty$. For each n, choose a continuous function ψ_n supported on $[a_n, a_{n-1}]$, such that

$$0 \leqslant \psi_n(u) \leqslant \frac{2\rho^{-2}(u)}{n}, \quad \forall u \geqslant 0,$$

and

$$\int_{a_n}^{a_{n-1}}\psi_n(u)du = 1.$$

Define

$$\varphi_n(x) \triangleq \int_0^{|x|}dy\int_0^y\psi_n(u)du.$$

It is not hard to see that $\varphi_n \in C^2(\mathbb{R}^1)$, $\varphi_n(x) \uparrow |x|$, $|\varphi'_n(x)| \leqslant 1$, and $\varphi''_n(x) = \psi_n(|x|)$.

Now suppose that X_t, X'_t are two solutions to the SDE (6.20) on a given set-up $((\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\}), \xi, B_t)$ with the same initial condition ξ. Define σ_N, σ'_N in the same way as in the proof of Theorem 6.9. For the simplicity of notation, we set $\tau \triangleq \sigma_N \wedge \sigma'_N$, $Y_t \triangleq X_{\sigma_N \wedge \sigma'_N \wedge t}$ and $Y'_t \triangleq X'_{\sigma_N \wedge \sigma'_N \wedge t}$. By rewriting the equation (6.30), we obtain that

$$Y_t - Y'_t = \int_0^t \left(\sigma(Y_s) - \sigma(Y'_s)\right) \mathbf{1}_{[0,\tau]}(s) dB_s + \int_0^t \left(b(Y_s) - b(Y'_s)\right) \mathbf{1}_{[0,\tau]}(s) ds.$$

According to Itô's formula,

$$\begin{aligned}
\varphi_n(Y_t - Y'_t) &= \int_0^t \varphi'_n(Y_s - Y'_s) \left(\sigma(Y_s) - \sigma(Y'_s)\right) \mathbf{1}_{[0,\tau]}(s) dB_s \\
&\quad + \int_0^t \varphi'_n(Y_s - Y'_s) \left(b(Y_s) - b(Y'_s)\right) \mathbf{1}_{[0,\tau]}(s) ds \\
&\quad + \frac{1}{2} \int_0^t \varphi''_n(Y_s - Y'_s) \|\sigma(Y_s) - \sigma(Y'_s)\|^2 \mathbf{1}_{[0,\tau]}(s) ds.
\end{aligned}$$

Since φ'_n, σ are bounded, we know that the first term is a martingale. Therefore,

$$\mathbb{E}\left[\varphi_n(Y_t - Y'_t)\right] = I_n^1 + I_n^2,$$

where

$$\begin{aligned}
I_n^1 &\triangleq \mathbb{E}\left[\int_0^t \varphi'_n(Y_s - Y'_s) \left(b(Y_s) - b(Y'_s)\right) \mathbf{1}_{[0,\tau]}(s) ds\right], \\
I_n^2 &\triangleq \frac{1}{2}\mathbb{E}\left[\int_0^t \varphi''_n(Y_s - Y'_s) \left(\sigma(Y_s) - \sigma(Y'_s)\right)^2 \mathbf{1}_{[0,\tau]}(s) ds\right].
\end{aligned}$$

On the one hand, according to Condition (2), we have

$$I_n^1 \leqslant K_N \int_0^t \mathbb{E}[|Y_s - Y'_s|] ds.$$

On the other hand, since $0 \leqslant \varphi''_n(x) = \psi_n(|x|) \leqslant 2\rho^{-2}(|x|)/n$, according to Condition (1), we have

$$I_n^2 \leqslant \frac{1}{2}\mathbb{E}\left[\int_0^t \frac{2\rho^{-2}(|Y_s - Y'_s|)}{n} \cdot \rho^2(|Y_s - Y'_s|) ds\right] = \frac{t}{n}.$$

Since $\varphi_n(x) \uparrow |x|$, by the monotone convergence theorem, we arrive at

$$\mathbb{E}[|Y_t - Y'_t|] \leqslant K_N \int_0^t \left(\mathbb{E}[|Y_s - Y'_s|]\right) ds.$$

Gronwall's inequality then implies that $Y_t = Y'_t$ for all $t \geqslant 0$. Since N is arbitrary, the same reason as in the proof of Theorem 6.9 shows that $e(X) = e(X')$ and $X_t = X'_t$ on $[0, e(X))$. □

The integrability condition on σ in Theorem 6.10 is essentially the best we can have. The following example shows what can go wrong if the integrability condition is not satisfied. This also gives an example in which uniqueness in law does not hold.

Example 6.2. Consider the case $n = d = 1$ with $b = 0$. Suppose that σ is a function such that $\sigma(0) = 0$, $\int_{-1}^{1} \sigma^{-2}(x)dx < \infty$ and $|\sigma(x)| \geqslant 1$ for $|x| \geqslant 1$ (for instance, $\sigma(x) = |x|^{\alpha}$ with $0 < \alpha < 1/2$). Let W_t be a one dimensional Brownian motion, and let L_t^x be its local time process. For each $\lambda > 0$, define

$$A_t^{\lambda} \triangleq \int_{\mathbb{R}^1} \sigma^{-2}(x) L_t^x dx + \lambda L_t^0 = \int_0^t \sigma^{-2}(W_s)ds + \lambda L_t^0.$$

According to the assumptions on σ, apparently A_t^{λ} is well defined and strictly increasing. Moreover, since $L_t^0 \stackrel{\text{law}}{=} S_t$ (c.f. Theorem 5.23), we know that $A_{\infty}^{\lambda} = \infty$. Let C_t^{λ} be the time-change associated with A_t^{λ}, and define $X_t^{\lambda} \triangleq W_{C_t^{\lambda}}$. It follows from Proposition 5.19 that X_t^{λ} is a local martingale with respect to the time-changed filtration, and

$$\langle X^{\lambda} \rangle_t = \langle W \rangle_{C_t^{\lambda}} = C_t^{\lambda}.$$

On the other hand, we know that

$$dA_t^{\lambda} = \sigma^{-2}(W_t)dt + \lambda L_t^0.$$

Now a crucial observation is that $\sigma^2(W_t)dL_t^0 \equiv 0$ because $\sigma(0) = 0$ and $dL^0(\{t \geqslant 0 : W_t \neq 0\}) = 0$. Therefore, $dt = \sigma^2(W_t)dA_t^{\lambda}$ and

$$C_t^{\lambda} = \int_0^{C_t^{\lambda}} ds = \int_0^{C_t^{\lambda}} \sigma^2(W_s)dA_s^{\lambda} = \int_0^t \sigma^2(X_u^{\lambda})du.$$

According to the martingale representation theorem (c.f. Theorem 5.14), we conclude that

$$X_t^{\lambda} = \int_0^t \sigma(X_u^{\lambda})dB_u$$

for some Brownian motion B_t. Therefore, we have a family of weak solutions X_t^{λ} with the same initial distribution $X_0^{\lambda} = 0$, which are possible defined on different set-ups because the filtrations, which depend on λ, can be different. Apparently, the distribution of X^{λ} varies for different λ (if $\lambda_1 < \lambda_2$, then $C_t^{\lambda_2} < C_t^{\lambda_1}$ for all t, so that as λ increases, X^{λ} is defined by running the Brownian motion in strictly slower speed which certainly results in a different distribution). Therefore, uniqueness in law for the corresponding SDE does not hold, and of course pathwise uniqueness fails as well.

Remark 6.7. The weak existence theorem and uniqueness theorems that we just proved for the time homogeneous SDE (6.20) extend to the time inhomogeneous case

$$dX_t = \sigma(t, X_t)dB_t + b(t, X_t)dt$$

without any difficulty. Indeed, by adding an additional equation $dX_t^0 = dt$, the time inhomogeneous equation reduces to the time homogeneous case. In particular, Theorem 6.7 and Theorem 6.9 hold for this case. Moreover, it can be easily seen that the proof of Theorem 6.10 works in exactly the same way without any difficulty in the time inhomogeneous case, although we cannot simply apply this reduction argument as the nature of this theorem is one dimensional. In the time inhomogeneous case, the corresponding assumptions on the coefficients in the uniqueness theorems should be made uniform with respect to the time variable.

Remark 6.8. In the context of Theorem 6.9 or Theorem 6.10, we also know that weak solution always exists. Therefore, according to the Yamada-Watanabe theorem, the SDE (6.20) is exact and has a unique strong solution. In addition, if we assume that the coefficients satisfy the linear growth condition, then for every probability measure μ on $\mathbb{R}^n$, a non-exploding weak solution with initial distribution μ exists (c.f. Theorem 6.8 and Remark 6.6). But we also know that pathwise uniqueness implies uniqueness in law, which is part of the Yamada-Watanabe theorem (c.f. Theorem 6.2). Therefore, every weak solution must not explode (note that the explosion time is intrinsically determined by the process, so if two processes have the same distribution, their explosion times will also have the same distribution).

6.5 A comparison theorem for one dimensional SDEs

Now let us use the same technique as in the proof of Theorem 6.10 to establish a useful comparison result in dimension one.

Let $\sigma(t,x), b_i(t,x)$ $(i=1,2)$ be real-valued continuous on $[0,\infty)\times\mathbb{R}^1$. We consider the SDEs

$$dX_t = \sigma(t,X_t)dB_t + b_i(t,X_t)dt \tag{6.31}$$

for $i=1,2$.

We assume the same condition on σ as in Theorem 6.10, i.e. there exists a strictly increasing function ρ on $[0,\infty)$ such that $\rho(0)=0$, $\int_{0+}\rho^{-2}(u)du=\infty$, and $\|\sigma(t,x)-\sigma(t,y)\|\leqslant\rho(|x-y|)$ for all $t\geqslant 0$ and $x,y\in\mathbb{R}^1$.

Theorem 6.11. *Suppose that*

(1) $b_1(t,x)\leqslant b_2(t,x)$ for all $t\geqslant 0$ and $x\in\mathbb{R}^1$;

(2) at least one of $b^i(t,x)$ is locally Lipschitz, i.e. for some $i=1,2$, for each $N\geqslant 1$, there exists $K_N>0$, such that

$$|b^i(t,x)-b^i(t,y)|\leqslant K_N|x-y|$$

for all $t\geqslant 0$ and $x,y\in\mathbb{R}^1$ with $|x|,|y|\leqslant N$.
Let X_t^i $(i=1,2)$ be a solution to the SDE (6.31) on the same given filtered probability space up to the intrinsic explosion time, i.e. X_t^i satisfies

$$X_t^i = X_0^i + \int_0^t \sigma(s,X_s^i)dB_s + \int_0^t b_i(s,X_s^i)ds,\quad 0\leqslant t<e(X^i),$$

for $i = 1, 2$. Suppose further that $X_0^1 \leqslant X_0^2 < \infty$ almost surely. Then with probability one, we have

$$X_t^1 \leqslant X_t^2 \ \ \forall t < e(X^1) \wedge e(X^2).$$

Proof. Suppose that $b_1(t,x)$ is locally Lipschitz.

Define ψ_n in the same way as in the proof of Theorem 6.10, but we set

$$\phi_n(x) \triangleq \begin{cases} 0, & x \leqslant 0; \\ \int_0^x dy \int_0^y \psi_n(u)du, & x > 0. \end{cases}$$

Then $\phi_n \in C^2(\mathbb{R}^1)$, $\phi_n(x) \uparrow x^+$, $0 \leqslant \phi_n'(x) \leqslant 1$, and $\phi_n''(x) = \psi_n(x)\mathbf{1}_{\{x>0\}}$.

We also localize X_t^1, X_t^2 in the same way as in the proof of Theorem 6.10, so we define $Y_t^i \triangleq X_{\tau\wedge t}^i$ $(i = 1, 2)$ where $\tau \triangleq \sigma_N^1 \wedge \sigma_N^2$.

By applying Itô's formula, we have

$$\phi_n(Y_t^1 - Y_t^2) = I_n^1 + I_n^2 + I_n^3, \tag{6.32}$$

where

$$\begin{aligned} I_n^1 &\triangleq \int_0^t \phi_n'(Y_s^1 - Y_s^2)(\sigma(s,Y_s^1) - \sigma(s,Y_s^2))\mathbf{1}_{[0,\tau]}(s)dB_s, \\ I_n^2 &\triangleq \int_0^t \phi_n'(Y_s^1 - Y_s^2)(b_1(s,Y_s^1) - b_2(s,Y_s^2))\mathbf{1}_{[0,\tau]}(s)ds, \\ I_n^3 &\triangleq \frac{1}{2}\int_0^t \phi_n''(Y_s^1 - Y_s^2)\left(\sigma(s,Y_s^1) - \sigma(s,Y_s^2)\right)^2 \mathbf{1}_{[0,\tau]}(s)ds. \end{aligned}$$

From the boundedness assumption, we know that $\mathbb{E}[I_n^1] = 0$. Moreover,

$$\begin{aligned} \mathbb{E}[I_n^3] &= \frac{1}{2}\mathbb{E}\left[\int_0^t \psi_n(Y_s^1 - Y_s^2)\mathbf{1}_{\{Y_s^1 > Y_s^2\}}\left(\sigma(s,Y_s^1) - \sigma(s,Y_s^2)\right)^2 \mathbf{1}_{[0,\tau]}(s)ds\right] \\ &\leqslant \frac{1}{2}\mathbb{E}\left[\int_0^t \frac{2\rho^{-2}(|Y_s^1 - Y_s^2|)}{n} \cdot \rho^2(|Y_s^1 - Y_s^2|)ds\right] \\ &\leqslant \frac{t}{n}. \end{aligned}$$

And also we have

$$\begin{aligned}
I_n^2 &= \int_0^t \phi_n'(Y_s^1 - Y_s^2)(b_1(s, Y_s^1) - b_2(s, Y_s^2))\mathbf{1}_{[0,\tau]}(s)ds \\
&= \int_0^t \phi_n'(Y_s^1 - Y_s^2)(b_1(s, Y_s^1) - b_1(s, Y_s^2))\mathbf{1}_{[0,\tau]}(s)ds \\
&\quad + \int_0^t \phi_n'(Y_s^1 - Y_s^2)(b_1(s, Y_s^2) - b_2(s, Y_s^2))\mathbf{1}_{[0,\tau]}(s)ds \\
&\leqslant \int_0^t \phi_n'(Y_s^1 - Y_s^2)(b_1(s, Y_s^1) - b_1(s, Y_s^2))\mathbf{1}_{[0,\tau]}(s)ds \\
&\leqslant K\int_0^t \mathbf{1}_{\{Y_s^1 > Y_s^2\}}|Y_s^1 - Y_s^2|ds \\
&= K\int_0^t (Y_s^1 - Y_s^2)^+ ds.
\end{aligned}$$

Therefore, by taking expectation on (6.32) and letting $n \to \infty$, we arrive at

$$\mathbb{E}[(Y_t^1 - Y_t^2)^+] \leqslant K\int_0^t \mathbb{E}[(Y_s^1 - Y_s^2)^+]ds.$$

According to Gronwall's inequality, we conclude that

$$\mathbb{E}[(Y_t^1 - Y_t^2)^+] = 0, \quad \forall t \geqslant 0,$$

which implies that with probability one,

$$Y_t^1 \leqslant Y_t^2 \quad \forall t \geqslant 0. \tag{6.33}$$

Now the result follows from letting $N \to \infty$.

In the case when $b_2(t, x)$ is locally Lipschitz, the same argument gives the desired result. □

Remark 6.9. If we make a more restrictive assumption that $b_1(t, x) < b_2(t, x)$ for all $t \geqslant 0$ and $x \in \mathbb{R}^1$, then we do not need to assume that at least one of b^i is locally Lipschitz. Indeed, after suitable localization, we may assume that $b_i(t, x)$ is uniformly bounded on $[0, T] \times \mathbb{R}^1$ for each fixed $T > 0$. In this case, it is possible to choose some $b(t, x)$ defined on $[0, T] \times \mathbb{R}^1$ which is Lipschitz continuous. If we consider the unique solution to the SDE

$$\begin{cases} dX_t = \sigma(t, X_t)dB_t + b(t, X_t)dt, \quad 0 \leqslant t \leqslant T; \\ X_0 = X_0^2, \end{cases}$$

then the argument in the proof of Theorem 6.11 shows that with probability one

$$Y_t^1 \leqslant X_t \leqslant Y_t^2, \quad \forall t \in [0, T].$$

As T is arbitrary, we conclude that (6.33) holds, which implies the desired result by letting $N \to \infty$.

6.6 Two useful techniques: transformation of drift and time-change

In this subsection, we introduce two important probabilistic techniques of solving SDEs in the weak sense. These techniques usually apply to SDEs with discontinuous coefficients, hence they are not covered by the existence and uniqueness theorems that we have proven so far.

The first technique is transformation of drift, which is an application of the Cameron-Martin-Girsanov transformation. For this part, we are interested transforming an SDE

$$dX_t = \alpha(t,X)dB_t + \beta(t,X)dt, \quad 0 \leqslant t \leqslant T, \tag{6.34}$$

to another SDE

$$dX_t = \alpha(t,X)dB_t + \beta'(t,X)dt, \quad 0 \leqslant t \leqslant T, \tag{6.35}$$

which has the same diffusion coefficient α but a different drift coefficient β'. In practice, we usually want $\beta' = 0$. In view of the Cameron-Martin-Girsanov transformation, we will always fix $T > 0$ and consider SDEs defined on the finite interval $[0,T]$.

In this part, we will always make the following assumption.

Assumption 6.2. *There exists some* $\gamma: \ [0,T] \times W^n \to \mathbb{R}^d$, *such that* γ *is bounded,* $\{\mathcal{B}_t(W^n)\}$*-progressively measurable, and*

$$\beta' = \beta + \alpha\gamma.$$

This is the case if α is invertible with bounded inverse, and β, β' are bounded.

Suppose that X_t is a solution to the SDE (6.34) on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t: \ 0 \leqslant t \leqslant T\})$ with an $\{\mathcal{F}_t\}$-Brownian motion B_t. Define

$$\mathcal{E}_t^\gamma \triangleq \exp\left(\int_0^t \gamma^*(s,X)dB_s - \frac{1}{2}\int_0^t \|\gamma(s,X)\|^2 ds\right).$$

Since γ is bounded, according to Novikov's condition (c.f. Theorem 5.18), we know that $\{\mathcal{E}_t^\gamma, \mathcal{F}_t: \ 0 \leqslant t \leqslant T\}$ is a martingale. Define a probability measure $\widetilde{\mathbb{P}}$ on $\mathcal{F}_T$ by

$$\widetilde{\mathbb{P}}(A) \triangleq \mathbb{E}[\mathbf{1}_A \mathcal{E}_T^\gamma], \quad A \in \mathcal{F}_T.$$

It follows from Girsanov's theorem (c.f. Theorem 5.17) that

$$\widetilde{B}_t \triangleq B_t - \int_0^t \gamma(s,X)ds, \quad 0 \leqslant t \leqslant T,$$

is an $\{\mathcal{F}_t\}$-Brownian motion under the probability measure $\widetilde{\mathbb{P}}$. Therefore, under $\widetilde{\mathbb{P}}$, X_t satisfies

$$\begin{aligned} X_t &= X_0 + \int_0^t \alpha(s,X)d\widetilde{B}_s + \int_0^t (\beta(s,X) + \alpha(s,X)\gamma(s,X))\, ds \\ &= X_0 + \int_0^t \alpha(s,X)d\widetilde{B}_s + \int_0^t \beta'(s,X)ds. \end{aligned}$$

In other words, X_t solves the SDE (6.35) with the new Brownian motion $\widetilde{B}_t$ under $\widetilde{\mathbb{P}}$.

On the other hand, suppose that X_t solves the SDE (6.35) with Brownian motion B_t. By defining

$$\mathcal{E}_t^{-\gamma} \triangleq \exp\left(-\int_0^t \gamma^*(s,X)dB_s - \frac{1}{2}\int_0^t \|\gamma(s,X)\|^2 ds\right),$$

the same argument shows that X_t solves the SDE (6.34) with Brownian motion

$$\widetilde{B}_t \triangleq B_t + \int_0^t \gamma(s,X)ds, \quad 0 \leqslant t \leqslant T, \tag{6.36}$$

under the probability measure

$$\widetilde{\mathbb{P}}(A) \triangleq \mathbb{E}[\mathbf{1}_A \mathcal{E}_T^{-\gamma}], \quad A \in \mathcal{F}_T. \tag{6.37}$$

Now we consider uniqueness. A crucial point is the following: we can assume without loss of generality that $\gamma = \alpha^*\eta$ for some $\{\mathcal{B}_t(W^n)\}$-progressively measurable $\eta: [0,T] \times W^n \to \mathbb{R}^n$. Indeed, for each $(t,w) \in [0,T] \times W^n$, write

$$\mathbb{R}^d = (\mathrm{Im}(\alpha^*(t,w))) \bigoplus \left(\mathrm{Im}(\alpha^*(t,w))^\perp\right),$$

where we regard $\alpha^*(t,w)$ as a linear map from $\mathbb{R}^n$ to $\mathbb{R}^d$. Under this decomposition, we can write

$$\gamma(t,w) = \gamma_1(t,w) + \gamma_2(t,w)$$

for some $\gamma_1(t,w) \in \mathrm{Im}(\alpha^*(t,w))$. Since γ_1 is defined pointwisely and $\|\gamma_1\| \leqslant \|\gamma\|$, we know that γ_1 is bounded, $\{\mathcal{B}_t(W^n)\}$-progressively measurable. Moreover, since

$$\langle \alpha(t,w)\gamma_2(t,w), y\rangle = \langle \gamma_2(t,w), \alpha^*(t,w)y\rangle = 0, \quad \forall y \in \mathbb{R}^n,$$

we have $\alpha(t,w)\gamma_2(t,w) = 0$, and hence $\alpha(t,w)\gamma(t,w) = \alpha(t,w)\gamma_1(t,w)$. As $\gamma_1(t,w) \in \mathrm{Im}(\alpha^*(t,w))$, there is a canonical way of choosing η such that $\gamma_1 = \alpha^*\eta$ and η is $\{\mathcal{B}_t(W^n)\}$-progressively measurable. For instance, we can define $\eta(t,w)$ to be the unique element in the affine space $\{\eta \in \mathbb{R}^n : \alpha^*(t,w)\eta(t,w) = \gamma_1(t,w)\}$ which minimizes its Euclidean norm. In the following, we will assume that $\gamma = \alpha^*\eta$.

Suppose that uniqueness in law holds for the SDE (6.34). Let X_t be a solution to the SDE (6.35) with Brownian motion B_t. It follows from the previous discussion that X_t solves the SDE (6.34) with a new Brownian motion $\widetilde{B}_t$ defined by (6.36) under the new probability measure $\widetilde{\mathbb{P}}$ defined by (6.37). However, we know that for every $A \in \mathcal{F}_T$,

$$\begin{aligned}
& \mathbb{P}(A) \\
= \; & \widetilde{\mathbb{E}}\left[\mathbf{1}_A \exp\left(\int_0^T \gamma^*(s,X)dB_s + \frac{1}{2}\int_0^T \|\gamma(s,X)\|^2 ds\right)\right] \\
= \; & \widetilde{\mathbb{E}}\left[\mathbf{1}_A \exp\left(\int_0^T \eta^*\alpha dB_s + \frac{1}{2}\int_0^T \|\gamma\|^2 ds\right)\right] \\
= \; & \widetilde{\mathbb{E}}\left[\mathbf{1}_A \exp\left(\int_0^T \eta^*\alpha d\widetilde{B}_s - \frac{1}{2}\int_0^T \|\gamma\|^2 ds\right)\right] \\
= \; & \widetilde{\mathbb{E}}\left[\mathbf{1}_A \exp\left(\int_0^T \eta^*(s,X)dX_s - \int_0^T \left(\eta^*(s,X)\beta(s,X) + \frac{1}{2}\|\gamma(s,X)\|^2\right) ds\right)\right].
\end{aligned}$$

In particular, for given $k \geqslant 1$, $0 \leqslant t_1 < \cdots < t_k \leqslant T$ and $f \in C_b(\mathbb{R}^{n\times k})$, we have

$$\begin{aligned}
&\mathbb{E}[f(X_{t_1}, \cdots, X_{t_n})] \\
= &\widetilde{\mathbb{E}}\left[f(X_{t_1}, \cdots, X_{t_k})\right. \\
&\left.\exp\left(\int_0^T \eta^*(s, X)dX_s - \int_0^T \left(\eta^*(s, X)\beta(s, X) + \frac{1}{2}\|\gamma(s, X)\|^2\right) ds\right)\right].
\end{aligned}$$

But the integrand of the expectation on the right hand side is a functional of $\{X_t : 0 \leqslant t \leqslant T\}$. So its distribution is uniquely determined by the distribution of X. Since uniqueness in law holds for the SDE (6.34) and X_t solves this SDE under $\widetilde{\mathbb{P}}$, we conclude that the distribution of X under $\mathbb{P}$ is uniquely determined by the distribution of X under $\widetilde{\mathbb{P}}$. In particular, uniqueness in law holds for the SDE (6.35). Conversely, similar argument shows that uniqueness in law for the SDE (6.35) implies uniqueness in law for the SDE (6.34).

To summarize, we have obtained the following result.

Theorem 6.12. *Under Assumption 6.2, the existence of weak solutions and uniqueness in law are equivalent for the two SDEs (6.34) and (6.35).*

The following is an important example which is not covered by our existence and uniqueness theorems so far, but it is within the scope of Theorem 6.12.

Example 6.3. Let $\beta : [0, T] \times W^d \to \mathbb{R}^d$ be bounded, $\{\mathcal{B}_t(W^d)\}$-progressively measurable (here $n = d$). Then weak existence and uniqueness in law hold for the SDE

$$dX_t = dB_t + \beta(t, X)dt, \quad 0 \leqslant t \leqslant T. \tag{6.38}$$

Indeed, we can just take $\gamma \triangleq \beta$ in the previous discussion to see that weak existence and uniqueness in law are both equivalent for the SDE

$$dX_t = dB_t, \quad 0 \leqslant t \leqslant T,$$

and the SDE (6.38). In particular, the law of the weak solution with initial distribution δ_x is determined by

$$\begin{aligned}
&\mathbb{E}\left[f(X_{t_1}, \cdots, X_{t_k})\right] \\
= \quad &\mathbb{E}^x\left[f(w_{t_1}, \cdots, w_{t_k})\exp\left(\int_0^T \beta^*(t, w)dw_t - \frac{1}{2}\int_0^T \|\beta(t, w)\|^2 ds\right)\right]
\end{aligned}$$

for $k \geqslant 1$, $0 \leqslant t_1 < \cdots < t_k \leqslant T$, and $f \in C_b(\mathbb{R}^{d\times k})$, where $\mathbb{P}^x$ is the law of the Brownian motion starting at x, and w_t is the coordinate process on path space.

It should be pointed out that existence and uniqueness only hold in the weak sense. In general, the SDE (6.38) can fail to be exact.

Now we study another useful technique: time-change. This technique applies to the one dimensional SDE of the following form:

$$dX_t = \alpha(t, X)dB_t, \tag{6.39}$$

where we assume that there exists constants $C_1, C_2 > 0$, such that

$$C_1 \leqslant \alpha(t, w) \leqslant C_2, \quad \forall (t, w) \in [0, \infty) \times W^1.$$

We have already seen the notion of a time-change in Section 5.6. Here we consider a more restrictive class of time-change. Denote $\mathcal{I}$ as the space of continuous functions $a: [0, \infty) \to [0, \infty)$ such that $a_0 = 0$, $t \mapsto a_t$ is strictly increasing and $\lim_{t\to\infty} a_t = \infty$. An adapted process A_t on some filtered probability space is called a *process of time-change* if for almost all ω, $A(\omega) \in \mathcal{I}$. As in Section 5.6, we use C_t to denote the time-change associated with A_t. In this case, C_t is really the inverse of A_t. Given a process X_t, we use $T^A X \triangleq X_{C_t}$ to denote the time-changed process of X_t by C_t (c.f. Definition 5.12).

The following result gives a method of solving the SDE (6.39) by using a time-change technique.

Theorem 6.13. *(1) Let b_t be a one dimensional Brownian motion defined on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\widetilde{\mathcal{F}}_t\})$ which satisfies the usual conditions. Let X_0 be an $\mathcal{F}_0$-measurable random variable. Define $\xi_t \triangleq X_0 + b_t$. Suppose that there exists a process A_t of time-change on Ω, such that with probability one, we have*

$$A_t = \int_0^t \alpha(A_s, T^A\xi)^{-2} ds, \quad \forall t \geqslant 0. \tag{6.40}$$

If we set $X \triangleq T^A\xi = X_0 + b_{C.}$ and $\mathcal{F}_t \triangleq \widetilde{\mathcal{F}}_{C_t}$, then there exists an $\{\mathcal{F}_t\}$-Brownian motion B_t, such that X_t solves the SDE (6.39) on $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ with Brownian motion B_t.

(2) Every solution to the SDE (6.39) arises in the way described by (1).

Proof. (1) Let b_t, X_0 and A_t be given as in the assumption. According to Proposition 5.19, we know that $M \triangleq T^A b \in \mathcal{M}_0^{\mathrm{loc}}(\{\mathcal{F}_t\})$ and $\langle M \rangle_t = C_t$. In view of (6.40), we have

$$\alpha(A_t, T^A\xi)^2 dA_t = dt,$$

so that

$$t = \int_0^t \alpha(A_s, T^A\xi)^2 dA_s = \int_0^t \alpha(A_s, X)^2 dA_s,$$

and hence

$$C_t = \int_0^{C_t} \alpha(A_s, X)^2 dA_s = \int_0^t \alpha(u, X)^2 du.$$

If we define $B_t \triangleq \int_0^t \alpha(s, X)^{-1} dM_s$, by Lévy's characterization theorem we know that B_t is an $\{\mathcal{F}_t\}$-Brownian motion, and

$$X_t - X_0 = M_t = \int_0^t \alpha(s, X) dB_s.$$

In other words, X_t solves the SDE (6.39) on $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ with Brownian motion B_t.

(2) Let X_t solves the SDE (6.39) on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\})$ with Brownian motion B_t. Then $M. \triangleq X. - X_0 \in \mathcal{M}_0^{\text{loc}}(\{\mathcal{F}_t\})$ and $\langle M \rangle_t = \int_0^t \alpha(s, X)^2 ds$. Let A_t be the inverse of $\langle M \rangle_t$. Define $b \triangleq T^{\langle M \rangle} M$ and $\widetilde{\mathcal{F}}_t \triangleq \mathcal{F}_{A_t}$. It follows that b_t is an $\{\widetilde{\mathcal{F}}_t\}$-Brownian motion, and $M = T^A b$. If we define $\xi_t \triangleq X_0 + b_t$, then apparently $X = T^A \xi$. In addition, since

$$t = \int_0^t \alpha(s, X)^{-2} d\langle M \rangle_s,$$

a simple change of variables shows that

$$A_t = \int_0^t \alpha(A_s, X)^{-2} ds = \int_0^t \alpha(A_s, T^A \xi)^{-2} ds.$$

Therefore, X_t arises in the way described by (1). $\square$

An important corollary of the second part of Theorem 6.13 is that if the ordinary differential equation (6.40) is always uniquely solvable from (X_0, b) in a pathwise sense, then weak existence and uniqueness in law hold for the SDE (6.39). Indeed, in this case the solution X is some deterministic functional of X_0 and b, so that its distribution is uniquely determined by the initial distribution and the distribution of Brownian motion.

Example 6.4. Consider the time homogeneous SDE

$$dX_t = a(X_t) dB_t,$$

where $a: \mathbb{R}^1 \to \mathbb{R}^1$ is a Borel measurable function such that $C_1 \leqslant a \leqslant C_2$ for some positive constants C_1, C_2. By setting $\alpha(t, w) \triangleq a(w_t)$, the differential equation (6.40) becomes

$$A_t = \int_0^t a\left((T^A \xi)_{A_s}\right)^{-2} ds = \int_0^t a(\xi_s)^{-2} ds. \tag{6.41}$$

Apparently A_t is uniquely determined by X_0 and b, and it is simply given by the formula (6.41). Therefore, $X \triangleq T^A \xi$ defines a weak solution and uniqueness in law holds.

Example 6.5. Consider the time inhomogeneous SDE

$$dX_t = a(t, X_t) dB_t,$$

where $a: [0, \infty) \times \mathbb{R}^1 \to \mathbb{R}^1$ is a Borel measurable function such that $C_1 \leqslant a \leqslant C_2$ for some positive constants C_1, C_2. By setting $\alpha(t, w) \triangleq a(t, w_t)$, the differential equation (6.40) becomes

$$A_t = \int_0^t a(A_s, \xi_s)^{-2} ds,$$

or

$$\begin{cases} \frac{dA_t}{dt} = \frac{1}{a(A_t, \xi_t)^2}, \\ A_0 = 0. \end{cases}$$

This equation has a unique solution along each fixed sample path of ξ, for instance if $a(t,x)$ is Lipschitz continuous in t. In this case, $X \triangleq T^A\xi$ defines a weak solution and uniqueness in law holds.

Example 6.6. Let $f(x)$ be a locally bounded, Borel measurable function on $\mathbb{R}^1$, and let $a(x)$ be a Borel measurable function on $\mathbb{R}^1$ such that $C_1 \leqslant a \leqslant C_2$ for some positive constants C_1, C_2. For a fixed $y \in \mathbb{R}^1$, define

$$\alpha(t,w) \triangleq a\left(y + \int_0^t f(w_s)ds\right), \quad (t,w) \in [0,\infty) \times W^1.$$

Consider the SDE

$$dX_t = \alpha(t,X)dB_t = a\left(y + \int_0^t f(X_s)ds\right)dB_t.$$

The differential equation (6.40) now becomes

$$\begin{aligned} A_t &= \int_0^t a\left(y + \int_0^{A_s} f(\xi_{C_u})du\right)^{-2} ds \\ &= \int_0^t a\left(y + \int_0^s f(\xi_u)\dot{A}_u du\right)^{-2} ds. \end{aligned}$$

Define

$$Z_t \triangleq \int_0^t f(\xi_u)\dot{A}_u du.$$

It follows that

$$\frac{dZ_t}{dt} = f(\xi_t)\dot{A}_t = \frac{f(\xi_t)}{a\left(y+Z_t\right)^2},$$

and hence

$$\int_0^t a(y+Z_s)^2 dZ_s = \int_0^t f(\xi_s)ds.$$

If we set

$$\Phi(x) \triangleq \int_0^x a(y+z)^2 dz, \quad x \in \mathbb{R}^1,$$

then $\Phi(x)$ is continuous, strictly increasing and

$$\Phi(Z_t) = \int_0^t f(\xi_s)ds.$$

Therefore,

$$Z_t = \Phi^{-1}\left(\int_0^t f(\xi_s)ds\right),$$

and A_t is uniquely solved as

$$A_t = \int_0^t a\left(y + \Phi^{-1}\left(\int_0^s f(\xi_u)du\right)\right)^{-2} ds.$$

In this case, $X \triangleq T^A \xi$ defines a weak solution and uniqueness in law holds.

A particular example is that $f(x) = x$. In this case, the SDE is equivalent to the following equation of motion with random acceleration:

$$\begin{cases} dY_t = X_t dt, \\ dX_t = a(Y_t) dB_t, \\ Y_0 = y. \end{cases}$$

6.7 Examples: linear SDEs and Bessel processes

In this subsection, we discuss several useful examples of SDEs.

The first type of examples that we are going to study are linear SDEs. This is the very nice case where we can obtain explicit formulae.

Consider the SDE

$$dX_t = (A(t)X_t + a(t))dt + \sigma(t)dB_t, \tag{6.42}$$

where $A(t), a(t)$ and $\sigma(t)$ are bounded, deterministic functions taking values in $\mathrm{Mat}(n, n)$, $\mathrm{Mat}(n, 1)$ and $\mathrm{Mat}(n, d)$ respectively.

This SDE can be solved explicitly in the following way. First of all, from classical ODE theory, we know that the matrix equation

$$\begin{cases} \frac{d\Phi(t)}{dt} = A(t)\Phi(t), \quad t \geqslant 0; \\ \Phi(0) = \mathrm{Id}_n, \end{cases}$$

has a unique solution $\Phi(t)$ which is absolutely continuous. Moreover, it is not hard to see that $\Phi(t)$ is non-singular for every $t \geqslant 0$. Indeed, suppose on the contrary that for some $t_0 \geqslant 0$ and some non-zero $\lambda \in \mathbb{R}^n$, we have $\Phi(t_0)\lambda = 0$. Since the function $x(t) \triangleq \Phi(t)\lambda$ solves the ODE

$$\frac{dx(t)}{dt} = A(t)x(t) \tag{6.43}$$

with condition $x(t_0) = 0$, from the uniqueness of (6.43), we know that $x(t) = 0$ for all t. But this contradicts the fact that $x(0) = \Phi(0)\lambda = \lambda \neq 0$. Therefore, $\Phi(t)$ is non-singular for every $t \geqslant 0$. In the case when $A(t) = A$, $\Phi(t)$ is explicitly given by

$$\Phi(t) = \mathrm{e}^{tA} \triangleq \sum_{k=0}^{\infty} \frac{A^k t^k}{k!},$$

and $\Phi^{-1}(t) = \mathrm{e}^{-tA}$.

$\Phi(t)$ is called the *fundamental solution* to the ODE (6.43). The reason for this name is very simple: the solution to the inhomogeneous ODE

$$\frac{dx(t)}{dt} = (A(t)x(t) + a(t)) \tag{6.44}$$

and even to the SDE (6.42) can be expressed in terms of $\Phi(t)$ and the coefficients. Indeed, it is classical that the solution to the ODE (6.44) is given by

$$x(t) = \Phi(t)\left(x(0) + \int_0^t \Phi^{-1}(s)a(s)ds\right).$$

Moreover, by using Itô's formula, it is not hard to see that

$$X_t \triangleq \Phi(t)\left(X_0 + \int_0^t \Phi^{-1}(s)a(s)ds + \int_0^t \Phi^{-1}(s)\sigma(s)dB_s\right) \tag{6.45}$$

is a solution to the SDE (6.42). This is the unique solution because pathwise uniqueness holds as a consequence of Lipschitz condition.

Since the integrands in the formula (6.45) are deterministic functions, we know that X_t is a Gaussian process provided X_0 is a Gaussian random variable. In this case, the mean function $m(t) \triangleq \mathbb{E}[X_t]$ is given by

$$m(t) = \Phi(t)\left(m(0) + \int_0^t \Phi^{-1}(s)a(s)ds\right)$$

and the covariance function $\rho(s,t) \triangleq \mathbb{E}[(X_s - m(s)) \cdot (X_t - m(t))^*]$ is given by

$$\rho(s,t) = \Phi(s)\left(\rho(0,0) + \int_0^{s\wedge t} \Phi^{-1}(u)\sigma(u)\sigma^*(u)\left(\Phi^{-1}(u)\right)^* du\right)\Phi^*(t), \quad s,t \geqslant 0.$$

A important example is the case when $n = d = 1$, $A(t) = -\gamma$ ($\gamma > 0$), $a(t) = 0$, $\sigma(t) = \sigma$, so the SDE takes the form

$$dX_t = -\gamma X_t dt + \sigma dB_t.$$

This is known as the *Langevin equation* and the solution is called the *Ornstein-Uhlenbeck* process. In this case, the solution is given by

$$X_t = X_0 e^{-\gamma t} + \sigma \int_0^t e^{-\gamma(t-s)} dB_s.$$

If X_0 is a Gaussian random variable with mean zero and variance η^2, then X_t is a centered Gaussian process with covariance function

$$\begin{aligned} \rho(s,t) &= e^{-\gamma(s+t)}\left(\eta^2 + \sigma^2 \int_0^s e^{2\gamma u} du\right) \\ &= \left(\eta^2 - \frac{\sigma^2}{2\gamma}\right) e^{-\gamma(s+t)} + \frac{\sigma^2}{2\gamma} e^{-\gamma(t-s)} \end{aligned}$$

provided $s < t$. In particular, if $\eta^2 = \sigma^2/(2\gamma)$, then X_t is also stationary in the sense that the distribution of $(X_{t_1+h}, \cdots, X_{t_k+h})$ is independent of h for any given $k \geqslant 1$ and $t_1 < \cdots < t_k$.

We can generalize the SDE (6.42) to the case where the diffusion coefficient depends linearly on X_t. For simplicity, we only consider the one dimensional case $(n = d = 1)$, in which the SDE takes the form

$$dX_t = (A(t)X_t + a(t))dt + (C(t)X_t + c(t))dB_t. \tag{6.46}$$

Of course we also have pathwise uniqueness in this case.

To write down the explicit solution for this SDE, it is helpful to first understand heuristically how to obtain the formula (6.45) in the previous discussion. Indeed, in the SDE (6.42), the linear dependence on X_t appears only in the term $A(t)X_t dt$, which can be viewed as contributing to an "exponential form" of the solution. Since $\Phi(t)$ behaves like the exponential of $\int A(t)dt$, it is reasonable to expect that if we apply Itô's formula to the process $\Phi^{-1}(t)X_t$, all those terms involving X_t should get killed and we will obtain that

$$d(\Phi^{-1}(t)X_t) = \Phi^{-1}(t)a(t)dt + \Phi^{-1}(t)\sigma(t)dB_t.$$

The reader might do the computation to see that this is indeed the case.

Now to solve the SDE (6.46), note that the linear dependence on X_t appears in the terms $A(t)X_t$ and $C(t)X_t$. These two terms should contribute to the "exponential form" of X_t. Therefore, taking into account the fact that $C(t)$ appears in the diffusion coefficient, we may define the "exponential" of $(A(t), C(t))$ by

$$Z_t \triangleq \exp\left(\int_0^t A(s)ds + \int_0^t C(s)dB_s - \frac{1}{2}\int_0^t C^2(s)ds\right).$$

Then it is reasonable to expect that after applying Itô's formula to the process $Z_t^{-1}X_t$, we should arrive at an expression which does not involve X_t. This is indeed the case, and we can obtain that the solution to the SDE (6.46) is given by

$$X_t = Z_t\left(X_0 + \int_0^t Z_s^{-1}(a(s) - C(s)c(s))ds + \int_0^t Z_s^{-1}c(s)dB_s\right).$$

An important example is the case when $A(t) = \mu$, $C(t) = \sigma$, and $a(t) = c(t) = 0$. In this case, the SDE takes the form

$$dX_t = \mu X_t dt + \sigma X_t dB_t$$

and the solution is given by

$$X_t = X_0 e^{\mu t + \sigma B_t - \frac{1}{2}\sigma^2 t}.$$

X_t is known as the *geometric Brownian motion*.

Another type of examples that we are going to study are Bessel processes. These are one dimensional SDEs. They are important and useful because a lot of explicit computations are possible and many interesting SDE models can be reduced to Bessel processes.

We first take a slight detour to discuss a bit more about one dimensional SDEs.

In the one dimensional case, explosion can be described in a more precise way as there are exactly two possible ways to explode in finite time: to the left or to the right. Therefore, we may consider more generally an SDE

$$\begin{cases} dX_t = \sigma(X_t)dB_t + b(X_t)dt, & t \geqslant 0, \\ X_0 = x \in I, \end{cases} \tag{6.47}$$

defined on an open interval $I = (l, r)$ with $-\infty \leqslant l < r \leqslant \infty$, and we think of l and r as two points of explosion. In Section 6.3, we were essentially identifying l and r in the case of explosion, which we do not want to do here. The argument here fails in higher dimensions, and indeed very little is known for the geometry of multidimensional diffusions.

Suppose that σ, b are continuously differentiable on I. According to the Yamada-Watanabe theory (continuity gives weak existence and local Lipschitz condition gives pathwise uniqueness), the SDE (6.47) has a unique solution X_t^x defined up to an explosion time e. More precisely, $e = \lim_{n\to\infty} \tau_n$, where $\tau_n \triangleq \inf\{t \geqslant 0 : X_t^x \notin [a_n, b_n]\}$ and a_n, b_n are two sequences of real numbers such that $a_n \downarrow l$, $b_n \uparrow r$. If we define $\widehat{W}^I$ to be the space of continuous paths $w : [0, \infty) \to [l, r]$ such that $w_0 \in I$ and $w_t = w_{e(w)}$ for $t \geqslant e(w)$, where $e(w) \triangleq \inf\{t \geqslant 0 : w_t = l \text{ or } r\}$, then X^x is a random variable taking values in $\widehat{W}^I$, and the explosion time of X^x is $e = e(X^x)$. What is more precise than Section 6.3 is that $\lim_{t\uparrow e} X_t^x$ exists and is equal to l or r on the event that $\{e < \infty\}$, a fact which can be proved by the same argument as in the proof of Theorem 6.7 and the continuity of X^x.

From now on, we always assume that $\sigma^2 > 0$ on I.

The following quantities play a fundamental role in studying the geometry of one dimensional diffusions.

Definition 6.9. Let $c \in I$ be a fixed real number.

(1) The *scale function* is defined by

$$s(x) \triangleq \int_c^x \exp\left(-2\int_c^\xi \frac{b(\zeta)}{\sigma^2(\zeta)}d\zeta\right)d\xi, \quad x \in I.$$

(2) The *speed measure* is defined by

$$m(dx) \triangleq \frac{2dx}{s'(x)\sigma^2(x)}, \quad x \in I.$$

(3) The *Green function* is defined by

$$G_{a,b}(x, y) \triangleq \frac{(s(x \wedge y) - s(a))(s(b) - s(x \vee y))}{s(b) - s(a)}, \quad x, y \in [a, b] \subseteq I.$$

We first show that these quantities can be used to compute expected exit times.

Let $[a,b] \subseteq I$. Consider the ODE

$$\begin{cases} b(x)M'(x) + \frac{1}{2}\sigma^2(x)M''(x) = -1, & a < x < b; \\ M(a) = M(b) = 0. \end{cases} \tag{6.48}$$

It is not hard to see that a solution is given by

$$M_{a,b}(x) \triangleq \int_a^b G_{a,b}(x,y)m(dy), \quad x \in [a,b].$$

In particular, $M_{a,b}$ is non-negative.

Define $\tau_{a,b} \triangleq \inf\{t < e : X_t^x \notin [a,b]\}$. According to Itô's formula and the ODE (6.48), we see that

$$M_{a,b}(X^x_{\tau_{a,b}\wedge t}) = M_{a,b}(x) + \int_0^{\tau_{a,b}\wedge t} M'_{a,b}(X_s^x)\sigma(X_s^x)dB_s - \tau_{a,b} \wedge t.$$

Therefore,

$$\mathbb{E}[\tau_{a,b} \wedge t] = M_{a,b}(x) - \mathbb{E}[M_{a,b}(X^x_{\tau_{a,b}\wedge t})] \leqslant M_{a,b}(x) < \infty. \tag{6.49}$$

In particular, $\tau_{a,b}$ is integrable. Indeed, in view of the boundary condition for the ODE (6.48), by letting $t \to \infty$ in (6.49), we have obtained the following result.

Proposition 6.1. *The expected exit time $\mathbb{E}[\tau_{a,b}]$ equals $M_{a,b}(x)$. In particular, $\tau_{a,b} < \infty$ almost surely.*

Note that Proposition 6.1 does not imply that $e < \infty$ almost surely. In fact, we are going to study the limiting behavior of X_t^x as $t \to e$ and give a simple non-explosion criterion for X_t^x. This is the content of the following elegant result.

Theorem 6.14. *(1) Suppose that $s(l+) = -\infty$ and $s(r-) = \infty$, then*

$$\mathbb{P}(e = \infty) = \mathbb{P}\left(\limsup_{t\to\infty} X_t^x = r\right) = \mathbb{P}\left(\liminf_{t\to\infty} X_t^x = l\right) = 1.$$

(2) If $s(l+) > -\infty$ and $s(r-) = \infty$, then $\lim_{t\uparrow e} X_t^x$ exists almost surely and

$$\mathbb{P}\left(\lim_{t\uparrow e} X_t^x = l\right) = \mathbb{P}\left(\sup_{t<e} X_t^x < r\right) = 1.$$

A similar assertion holds if the roles of l and r are interchanged.

(3) If $s(l+) > -\infty$ and $s(r-) < \infty$, then

$$\mathbb{P}\left(\lim_{t\uparrow e} X_t^x = l\right) = 1 - \mathbb{P}\left(\lim_{t\uparrow e} X_t^x = r\right) = \frac{s(r-) - s(x)}{s(r-) - s(l+)}. \tag{6.50}$$

Proof. The main feature of the scale function is that $(\mathcal{A}s)(x) = 0$ for all $x \in I$, where

$$(\mathcal{A}f)(x) = \frac{1}{2}\sigma^2(x)f''(x) + b(x)f'(x)$$

is the generator of the SDE. Given $l < a < x < b < r$, we again consider the first exit time $\tau_{a,b}$. According to the martingale characterization for the weak solution, we know that $s(X^x_{\tau_{a,b}\wedge t}) - s(x)$ is a martingale. In particular, we have

$$\mathbb{E}\left[s(X^x_{\tau_{a,b}\wedge t})\right] = s(x).$$

By letting $t \to \infty$ and noting that $\tau_{a,b} < \infty$ almost sure (c.f. Proposition 6.1), we obtain that

$$s(a)\mathbb{P}\left(X^x_{\tau_{a,b}} = a\right) + s(b)\mathbb{P}\left(X^x_{\tau_{a,b}} = b\right) = s(x)$$

as well as

$$\mathbb{P}\left(X^x_{\tau_{a,b}} = a\right) + \mathbb{P}\left(X^x_{\tau_{a,b}} = b\right) = 1.$$

Therefore,

$$\mathbb{P}\left(X^x_{\tau_{a,b}} = a\right) = \frac{s(b) - s(x)}{s(b) - s(a)}, \quad \mathbb{P}\left(X^x_{\tau_{a,b}} = b\right) = \frac{s(x) - s(a)}{s(b) - s(a)}.$$

(1) Suppose that $s(l+) = -\infty$ and $s(r-) = \infty$.

In this case, $\lim_{a\downarrow l}\mathbb{P}(X^x_{\tau_{a,b}} = b) = 1$. Since $\{X^x_{\tau_{a,b}} = b\} \subseteq \{\sup_{t<e} X^x_t \geqslant b\}$ for all a, we conclude that $\mathbb{P}(\sup_{t<e} X^x_t \geqslant b) \geqslant 1$. This is true for all b, which implies that $\mathbb{P}(\sup_{t<e} X^x_t = r) = 1$. Similarly, we have $\mathbb{P}(\inf_{t<e} X^x_t = l) = 1$. This in particular implies that $\mathbb{P}(e = \infty) = 1$, for otherwise on the event that $\{e < \infty\}$, we know that $\lim_{t\uparrow e} X_t$ exists and equals l or r, which is a contradiction.

(2) Suppose that $s(l+) > -\infty$ and $s(r-) = \infty$.

In this case, by the discussion in the first case, we see that

$$\mathbb{P}\left(\inf_{t<e} X^x_t = l\right) = 1. \tag{6.51}$$

On the other hand, the process

$$Y^{a,b}_t \triangleq s(X^x_{\tau_{a,b}\wedge t}) - s(l+)$$

is a non-negative martingale. According to Fatou's lemma, by letting $a \downarrow l, b \uparrow r$, we conclude that

$$Y_t \triangleq s(X^x_{e\wedge t}) - s(l+)$$

is a non-negative supermartingale. In particular, with probability one $\lim_{t\to\infty} Y_t$ exists finitely. This implies that $\lim_{t\uparrow e} s(X^x_t)$ exists finitely. But s is continuous and strictly increasing, we therefore know that $\lim_{t\uparrow e} X^x_t$ exists almost surely. In view of (6.51), we conclude that $\mathbb{P}(\lim_{t\uparrow e} X^x_t = l) = 1$. This also implies that $\mathbb{P}(\sup_{t<e} X^x_t < r) = 1$ since $\{\lim_{t\uparrow e} X^x_t = l\} \subseteq \{\sup_{t<e} X^x_t < r\}$.

The case when $s(l+) = -\infty$ and $s(r-) < \infty$ is treated in a similar way.
(3) Suppose that $s(l+) > -\infty$ and $s(r-) < \infty$.
In this case, we have

$$\mathbb{P}\left(\sup_{t<e} X_t^x \geqslant b\right) \geqslant \frac{s(x) - s(l+)}{s(b) - s(l+)}$$

for all b. Therefore,

$$\mathbb{P}\left(\sup_{t<e} X_t^x = r\right) \geqslant \frac{s(x) - s(l+)}{s(r-) - s(l+)}. \tag{6.52}$$

Similarly,

$$\mathbb{P}\left(\inf_{t<e} X_t^x = l\right) \geqslant \frac{s(r-) - s(x)}{s(r-) - s(l+)}. \tag{6.53}$$

On the other hand, by the discussion in the second case, we see that $\lim_{t\uparrow e} X_t^x$ exists almost surely. Therefore,

$$\left\{\lim_{t\uparrow e} X_t^x = r\right\} = \left\{\sup_{t<e} X_t^x = r\right\}, \ \left\{\lim_{t\uparrow e} X_t^x = l\right\} = \left\{\inf_{t<e} X_t^x = l\right\}.$$

In particular, these two events are disjoint. Since the right hand sides of (6.52) and (6.53) add up to one, we arrive at (6.50). □

Remark 6.10. In the first case in Theorem 6.14, we see that X_t^x is recurrent, in the sense that $\mathbb{P}(\sigma_y < \infty) = 1$ for every $y \in I$, where $\sigma_y \triangleq \inf\{t > 0 : X_t = y\}$. This case gives a simple non-explosion criterion for the SDE. In the second and third cases, there exists an open subset U of (l, r), such that with positive probability X_t^x never enters U. In these cases, it is not clear whether X_t^x explodes in finite time with positive probability. The famous *Feller's test* studies explosion and non-explosion criteria in a very elegant way, in terms of a more complicated quantity than just the scale function. We are not going to discuss Feller's test here, and we refer the reader to [4] for a detailed introduction.

Now we come back to the example of Bessel processes.

Let $B_t = (B_t^1, \cdots, B_t^n)$ be an n-dimensional Brownian motion. $R_t \triangleq |B_t|$ is known as the classical n-dimensional *Bessel process*. By applying Itô's formula to the process $\rho_t \triangleq R_t^2 = \sum_{i=1}^n (B_t^i)^2$, formally we have

$$\begin{aligned} d\rho_t &= 2\sum_{i=1}^n B_t^i dB_t^i + ndt \\ &= 2\sqrt{\rho_t} \cdot \frac{\sum_{i=1}^n B_t^i dB_t^i}{\sqrt{\rho_t}} + ndt. \end{aligned}$$

But the process

$$W_t \triangleq \sum_{i=1}^n \int_0^t \frac{B_s^i dB_s^i}{\sqrt{\rho_s}}$$

is a one dimensional Brownian motion according to Lévy's characterization theorem. Therefore, ρ_t solves the SDE

$$d\rho_t = 2\sqrt{\rho_t}dW_t + ndt.$$

In terms of the SDE, we can generalize the previous notion of Bessel processes to arbitrary (real) dimensions. To be precise, consider the one dimensional SDE

$$d\rho_t = 2\sqrt{|\rho_t|}dB_t + \alpha dt, \tag{6.54}$$

where $\alpha \geqslant 0$ is a constant.

Proposition 6.2. *The SDE (6.54) is exact without explosion to infinity in finite time. Moreover, if $\rho_0 \geqslant 0$, then $\rho_t \geqslant 0$ for all t.*

Proof. Since the coefficients are continuous, weak existence holds. From the inequality $|\sqrt{x} - \sqrt{y}| \leqslant \sqrt{x-y}$, pathwise uniqueness holds. Moreover, since $\sqrt{|x|} \leqslant (1+|x|)/2$, the solution cannot explode. Therefore, the SDE is exact without explosion.

Now observe that the unique solution to the SDE

$$\begin{cases} d\rho_t = 2\sqrt{|\rho_t|}dB_t, \quad t \geqslant 0, \\ \rho_0 = 0, \end{cases}$$

is the trivial solution $\rho \equiv 0$. According to the comparison theorem (c.f. Theorem 6.11), we conclude that the solution ρ_t to the SDE (6.54) is non-negative, provided that $\rho_0 \geqslant 0$. □

Due to Proposition 6.2, when writing the SDE (6.54), we may drop the absolute value inside the square root.

Definition 6.10. Given $\alpha > 0$ and $x \geqslant 0$, the solution to the equation

$$\rho_t = x + 2\int_0^t \sqrt{\rho_s}dB_s + \alpha t$$

is called a *squared Bessel process starting at x with dimension α* and it is simply denoted as BSEQ$^\alpha$. The process $R_t \triangleq \sqrt{\rho_t}$ is called a *Bessel process starting at $\sqrt{x}$ with dimension* α and it is simply denoted as BSE$^\alpha$.

From the comparison theorem again, we easily see that if ρ_t^1, ρ_t^2 are BESQ$^{\alpha_1}$,BESQ$^{\alpha_2}$ starting at x_1, x_2 respectively such that $\alpha_1 \leqslant \alpha_2, x_1 \leqslant x_2$, then $\rho_t^1 \leqslant \rho_t^2$ for all t.

An important property for Bessel processes is the additivity property. For $\alpha \geqslant 0, x \geqslant 0$, we use $\mathbb{Q}_x^\alpha$ to denote the law of BESQ$^\alpha$ starting at x on the path space W^1.

Proposition 6.3. *Let $\alpha_1, \alpha_2 \geqslant 0$ and $x_1, x_2 \geqslant 0$. Then $\mathbb{Q}_{x_1}^{\alpha} * \mathbb{Q}_{x_2}^{\alpha_2} = \mathbb{Q}_{x_1+x_2}^{\alpha_1+\alpha_2}$, where $*$ means convolution of two measures.*

Proof. In terms of processes, it is equivalent to prove the following. Suppose that B^1, B^2 are two independent Brownian motions. Let ρ_t^i $(i = 1, 2)$ be the unique solution to the following SDE:

$$\begin{cases} d\rho_t^i = 2\sqrt{\rho_t^i} dB_t^i + \alpha_i dt, & t \geqslant 0, \\ \rho_0^i = x_i. \end{cases}$$

Then $\rho_t^3 \triangleq \rho_t^1 + \rho_t^2$ solves the SDE

$$\begin{cases} d\rho_t^3 = 2\sqrt{\rho_t^3} dB_t + (\alpha_1 + \alpha_2) dt, & t \geqslant 0, \\ \rho_0^3 = x_1 + x_2, \end{cases}$$

for some Brownian motion B_t. Indeed, this is true because

$$\begin{aligned} \rho_t^3 &= x_1 + x_2 + 2\int_0^t \left(\sqrt{\rho_s^1} dB_s^1 + \sqrt{\rho_s^2} dB_s^2\right) + (\alpha_1 + \alpha_2)t \\ &= x_1 + x_2 + 2\int_0^t \sqrt{\rho_s^3} \cdot \left(\frac{\sqrt{\rho_s^1} dB_s^1}{\sqrt{\rho_s^3}} + \frac{\sqrt{\rho_s^2} dB_s^2}{\sqrt{\rho_s^3}}\right) + (\alpha_1 + \alpha_2)t \\ &= x_1 + x_2 + 2\int_0^t \sqrt{\rho_s^3} dB_s + (\alpha_1 + \alpha_2)t, \end{aligned}$$

where

$$B_t \triangleq \int_0^t \left(\frac{\sqrt{\rho_s^1} dB_s^1}{\sqrt{\rho_s^3}} + \frac{\sqrt{\rho_s^2} dB_s^2}{\sqrt{\rho_s^3}}\right)$$

is a Brownian motion according to Lévy's characterization theorem. □

We can also derive the Laplace transform of a BSEQ$^\alpha$ explicitly. In fact, if $\alpha \in \mathbb{N}$, let ρ_t^i $(1 \leqslant i \leqslant \alpha)$ be a BESQ^1 starting at x/α (driven by independent Brownian motions). According to Proposition 6.3, we have

$$\mathbb{E}\left[e^{-\lambda(\rho_t^1 + \cdots + \rho_t^\alpha)}\right] = \left(\mathbb{E}\left[e^{-\lambda \rho_t^1}\right]\right)^\alpha = \frac{1}{(1 + 2\lambda t)^{\alpha/2}} e^{-\frac{\lambda x}{(1+2\lambda t)}}, \tag{6.55}$$

where we have used the formula for the Laplace transform of the square of a Gaussian random variable. This formula must be true as the additivity holds even when α is not an integer, as indicated by Proposition 6.3. In other words, we have the following result.

Proposition 6.4. *The formula (6.55) gives the Laplace transform for a BESQ$^\alpha$ starting at x for arbitrary $\alpha, x \geqslant 0$. In particular,*

$$\mathbb{P}(R_t \in dy) = \frac{e^{-\frac{(x^2+y^2)}{2t}}}{t(xy)^{\alpha/2-1}} y^{\alpha-1} I_{\alpha/2-1}\left(\frac{xy}{t}\right), \quad y > 0,$$

where R_t is a BES$^\alpha$ starting at x, and

$$I_\nu(x) \triangleq \left(\frac{x}{2}\right)^\nu \sum_{n=0}^\infty \frac{(x/2)^{2n}}{n!\Gamma(\nu + n + 1)}$$

is the modified Bessel function.

Proof. The function $v(t,x)$ defined by the right hand side of (6.55) is in $C_b^{1,2}([0,\infty)\times[0,\infty))$ and satisfies $\partial v/\partial t=\mathcal{A}v$, where

$$(\mathcal{A}f)(x)\triangleq 2|x|f''(x)+\alpha f'(x)$$

is the generator of the SDE (6.54). Given $t_0>0$, define

$$u(t,x)\triangleq v(t_0-t,x),\quad (t,x)\in[0,t_0]\times[0,\infty).$$

By applying Itô's formula to the process $u(t,\rho_t)$, we can see that $u(t,\rho_t)-u(0,x)$ is martingale (one may see the integrability of quadratic variation process by comparing ρ_t with a BESQ$^{\alpha'}$ where $\alpha'\geqslant\alpha$ is an integer). Therefore,

$$\mathbb{E}[u(t_0,\rho_{t_0})]=\mathbb{E}[u(0,x)],$$

which is exactly the formula (6.55) at t_0.

The second part follows from inverting the Laplace transform and the fact that $R_t=\sqrt{\rho_t}$. □

Another interesting property is the behavior at the boundary point 0.

Proposition 6.5. *Let R_t be a BES$^\alpha$ starting at $\sqrt{x}>0$. If $\alpha\geqslant 2$, then with probability one, R_t never reaches 0.*

Proof. It is equivalent to looking at $\rho_t=R_t^2$. Let $e\triangleq\inf\{t\geqslant 0:\ \rho_t=0\}$. As we are only concerned with the process before e, we can use the model (6.47) with $I=(0,\infty)$ and think of 0 and ∞ as two explosion times. In this framework, consider the scale function

$$s(x)=\int_1^x\exp\left(-\int_1^y\frac{2\alpha dz}{4z}\right)dy=\int_1^x y^{-\frac{\alpha}{2}}dy.$$

In particular, $s(0+)=-\infty$ if and only if $\alpha\geqslant 2$, and $s(\infty)=\infty$ if and only if $\alpha\leqslant 2$. Therefore, the result follows from Theorem 6.14 (Case (1) for $\alpha=2$ and Case (2) for $\alpha>2$). □

Remark 6.11. By using Feller's test of explosion, it is possible to show that if $0\leqslant\alpha<2$, then with probability one, R_t reaches 0 in finite time. Moreover, by using pathwise uniqueness and local times (we leave this part as a good exercise), one can show that the point 0 is absorbing (i.e. the process remains at 0 once it reaches it) if $\alpha=0$ and reflecting (i.e. whenever $R_{t_0}=0$, we have for any $\delta>0$, there exists $t\in(t_0,t_0+\delta)$ such that $R_t>0$) if $0<\alpha<2$.

As we mentioned before, Bessel processes are useful because they are related to many interesting SDE models. Here we present two of them: the Cox-Ingersoll-Ross processes and the Constant Elasticity of Variance processes.

(1) The Cox-Ingersoll-Ross (CIR) model

Consider the SDE

$$\begin{cases} dr_t = k(\theta - r_t)dt + \sigma\sqrt{r_t}dB_t, & t \geqslant 0, \\ r_0 = x \geqslant 0, \end{cases} \tag{6.56}$$

where $k \cdot \theta > 0$ and $\sigma \neq 0$. Apparently, the SDE (6.56) is exact without explosion. Moreover, according to the comparison theorem, we know that $r_t \geqslant 0$ for all t. The solution to this SDE is nothing but a time-change and rescaling of a BESQ.

Proposition 6.6. *The solution r_t to the SDE (6.56) is given by*

$$r_t = \mathrm{e}^{-kt}\rho\left(\frac{\sigma^2}{4k}(\mathrm{e}^{kt} - 1)\right),$$

where ρ_t is a BESQ$^\alpha$ starting at x with $\alpha \triangleq 4k\theta/\sigma^2$.

Proof. Let $Z_t \triangleq r_t \mathrm{e}^{kt}$. According to the integration by parts formula, we have

$$Z_t = x + \theta(\mathrm{e}^{kt} - 1) + \int_0^t \sigma \mathrm{e}^{\frac{ks}{2}}\sqrt{Z_s}dB_s.$$

To relate Z_t with a BESQ, let A_t be a non-negative, increasing function to be determined and let C_t be the associated time-change. It follows that

$$\begin{aligned} Z_{C_t} &= x + \theta(\mathrm{e}^{kC_t} - 1) + \int_0^{C_t} \sigma \mathrm{e}^{\frac{ks}{2}}\sqrt{Z_s}dB_s \\ &= x + \theta(\mathrm{e}^{kC_t} - 1) + 2\int_0^t \frac{\sigma \mathrm{e}^{\frac{kC_s}{2}}}{2} \cdot \sqrt{Z_{C_s}}dB_{C_s}. \end{aligned}$$

In view of the SDE (6.54) for a BESQ, we want

$$W_t \triangleq \int_0^t \frac{\sigma \mathrm{e}^{\frac{kC_s}{2}}}{2}dB_{C_s}$$

to be a Brownian motion. This is equivalent to saying that

$$d\langle W\rangle_t = \frac{\sigma^2 \mathrm{e}^{kC_t}}{4}C_t' dt = dt,$$

and hence

$$\frac{\sigma^2 \mathrm{e}^{kC_t}}{4}C_t' = 1.$$

Since we also want $C_0 = 0$, we obtain easily that

$$t = \frac{(\mathrm{e}^{kc} - 1)\sigma^2}{4k}, \quad c \geqslant 0,$$

and this is the increasing function $c \mapsto t = A_c$ that we need. C_t would just be the inverse of A_t. If we set $\rho_t \triangleq Z_{C_t}$, then

$$\rho_t = x + \frac{4k\theta}{\sigma^2}t + 2\int_0^t \sqrt{\rho_s}dW_s,$$

which shows that ρ_t is a BESQ$^\alpha$ starting at x with $\alpha \triangleq 4k\theta/\sigma^2$. The result then follows. □

From Proposition 6.6 and Theorem 6.14, we know that r_t never reaches 0 if $k\theta \geqslant \sigma^2/2$, provided $r_0 = x > 0$.

(2) The Constant Elasticity of Variance (CEV) model

Consider the SDE

$$\begin{cases} dS_t = S_t(\mu dt + \sigma S_t^\beta dB_t), & t \geqslant 0, \\ S_0 = x > 0, \end{cases} \tag{6.57}$$

where $\beta > 0$ and $\sigma \neq 0$. Apparently, the SDE (6.57) is exact.

We first look at the case when $\mu = 0$. Let $\tau_0 \triangleq \inf\{t \geqslant 0 : \ S_t = 0\}$. Define

$$\rho_t \triangleq \frac{1}{\sigma^2\beta^2}S_t^{-2\beta}, \quad t < \tau_0 \wedge e.$$

According to Itô's formula, we have

$$\rho_t = \frac{1}{\sigma^2\beta^2}x^{-2\beta} + 2\int_0^t \sqrt{\rho_s}dW_s + \left(2 + \frac{1}{\beta}\right)t, \quad t < \tau_0 \wedge e,$$

where $W_t \triangleq -B_t$ is a Brownian motion. Therefore, ρ_t is a BESQ$^\alpha$ starting at $\frac{1}{\sigma^2\beta^2}x^{-2\beta}$ with $\alpha \triangleq 2 + \beta^{-1} > 2$. In particular, we know that ρ_t does not explode in finite time and hence $\mathbb{P}(\tau_0 = \infty) = 1$. In other words, S_t never reaches 0. Moreover, according to Proposition 6.5, we know that ρ_t never reaches 0. Therefore, S_t does not explode in finite time. Since

$$S_t = \left(\frac{1}{\sigma^2\beta^2}\right)^{\frac{1}{2\beta}} \rho_t^{-\frac{1}{2\beta}} = x + \sigma\int_0^t S_u^{1+\beta}dB_u,$$

we see that S_t is a local martingale with strictly decreasing expectation. In particular, S_t is not a martingale (but of course it is a non-negative supermartingale). The distribution of S_t can of course be computed from the distribution of BESQ explicitly.

Now we consider the case when $\mu \neq 0$. Similar to the previous discussion, let $\tau_0 \triangleq \inf\{t \geqslant 0 : \ S_t = 0\}$, and define

$$r_t \triangleq \frac{1}{4\beta^2}S_t^{-2\beta}, \quad t < \tau_0 \wedge e.$$

According to Itô's formula, we obtain that

$$r_t = \frac{1}{4\beta^2}x^{-2\beta} + \int_0^t k(\theta - r_s)ds + (-\sigma)\cdot\int_0^t \sqrt{r_s}dB_s, \quad t < \tau_0 \wedge e.$$

where $k \triangleq 2\mu\beta$, $\theta \triangleq (2\beta+1)\sigma^2/(4k\beta)$. In particular, r_t satisfies the CIR model with parameters $k, \theta, -\sigma$. Therefore, r_t does not explode and hence $\mathbb{P}(\tau_0 = \infty) = 1$. Moreover, since $k\theta \geqslant \sigma^2/2$, we know from the previous discussion on the CIR model that r_t never reaches 0. In other words, S_t does not explode. Since

$$S_t = \left(\frac{1}{4\beta^2}\right)^{\frac{1}{2\beta}} r_t^{-\frac{1}{2\beta}},$$

according to Proposition 6.6, the distribution of S_t can also be computed from the distribution of BESQ explicitly.

6.8 Itô's diffusion processes and partial differential equations

To conclude this course, we study Itô's diffusion processes and explore their relationship with partial differential equations. In particular, we are going to show that solutions to a class of elliptic and parabolic equations admit stochastic representations.

Throughout this subsection, we consider the multidimensional SDE

$$\begin{cases} dX_t = \sigma(X_t)dB_t + b(X_t)dt, \quad t \geqslant 0, \\ X_0 = x, \end{cases} \tag{6.58}$$

where σ, b are Lipschitz continuous. We know from the Yamada-Watanabe theory that this SDE is exact. Let $\{X_t^x : x \in \mathbb{R}^n, t \geqslant 0\}$ be the unique solution to the SDE (6.58) on some given filtered probability space.

Definition 6.11. $\{X_t^x : x \in \mathbb{R}^n, t \geqslant 0\}$ is called a *time homogeneous Itô's diffusion process.*

A important consequence of exactness is the following strong Markov property.

Theorem 6.15. *Suppose that τ is a finite $\{\mathcal{F}_t\}$-stopping time. Then for any $x \in \mathbb{R}^n$, $\theta > 0$ and $f \in B_b(\mathbb{R}^n)$, we have*

$$\mathbb{E}[f(X_{\tau+\theta}^x)|\mathcal{F}_\tau] = \mathbb{E}[f(X_\theta^y)]|_{y=X_\tau^x}.$$

Proof. From the SDE (6.58), for $\theta \geqslant 0$, we have

$$\begin{aligned} X_{\tau+\theta}^x &= x + \int_0^{\tau+\theta} \sigma(X_s^x)dB_s + \int_0^{\tau+\theta} b(X_s^x)ds \\ &= X_\tau^x + \int_\tau^{\tau+\theta} \sigma(X_s^x)dB_s + \int_\tau^{\tau+\theta} b(X_s^x)ds \\ &= X_\tau^x + \int_0^\theta \sigma(X_{\tau+u}^x)dB_u^{(\tau)} + \int_0^\theta b(X_{\tau+u}^x)du, \end{aligned}$$

where $B^{(\tau)}_u \triangleq B_{\tau+u} - B_\tau$ is a Brownian motion. Therefore, the process $\theta \mapsto X^x_{\tau+\theta}$ solves the SDE (6.58) with initial data X^x_τ and Brownian motion $B^{(\tau)}$. According to exactness, $X^x_{\tau+\theta}$ is a deterministic functional of X^x_τ and $B^{(\tau)}$, and we may write

$$X^x_{\tau+\theta} = F(\theta, X^x_\tau, B^{(\tau)})$$

for some deterministic functional F, where $\theta \mapsto F(\theta, y, B)$ gives the unique solution to the SDE (6.58) starting at y with a given Brownian motion B. Now since X^x_τ is $\mathcal{F}_\tau$-measurable and $B^{(\tau)}$ is independent of $\mathcal{F}_\tau$, we see immediately that

$$\begin{aligned} \mathbb{E}[f(X^x_{\tau+\theta})|\mathcal{F}_\tau] &= \mathbb{E}\left[f\left(F(\theta, X^x_\tau, B^{(\tau)})\right)|\mathcal{F}_\tau\right] \\ &= \mathbb{E}\left[f\left(F(\theta, y, B)\right)\right]|_{y=X^x_\tau} \\ &= \mathbb{E}[f(X^y_\theta)]|_{y=X^x_\tau}. \end{aligned}$$

□

Now recall that the generator $\mathcal{A}$ of the SDE (6.58) is the second order differential operator

$$(\mathcal{A}f)(x) = \frac{1}{2}\sum_{i,j=1}^n a^{ij}(x)\frac{\partial^2 f}{\partial x^i \partial x^j} + \sum_{i=1}^n b^i(x)\frac{\partial f}{\partial x^i},$$

where $a \triangleq \sigma \cdot \sigma^*$.

In the theory of elliptic PDEs, we are usually interested in the boundary value problem associated with the operator $\mathcal{A}$. More precisely, suppose that D is a bounded domain in $\mathbb{R}^n$. Let $k : \overline{D} \to [0, \infty)$, $g : \overline{D} \to \mathbb{R}^1$ and $f : \partial D \to \mathbb{R}^1$ be continuous functions. We consider the following (Dirichlet) boundary value problem: find a function $u \in C(\overline{D}) \cap C^2(D)$, such that

$$\begin{cases} \mathcal{A}u - k \cdot u = -g, & x \in D, \\ u = f, & x \in \partial D. \end{cases} \tag{6.59}$$

The existence of such u is well studied in PDE theory under suitable regularity conditions on the coefficients and the geometry of ∂D. Here we are not concerned with existence, but we are interested in representing the solution in terms of Itô's diffusion process under the assumption of existence.

Theorem 6.16. *Suppose that there exists $u \in C(\overline{D}) \cap C^2(D)$ which solves the boundary value problem (6.59). Let X^x_t be the solution to the SDE (6.58). Assume that the exit time*

$$\tau_D \triangleq \inf\{t \geqslant 0 : \ X^x_t \notin D\}$$

satisfies

$$\mathbb{E}[\tau_D] < \infty \tag{6.60}$$

for every $x \in D$. Then u is given by

$$\begin{aligned} u(x) &= \mathbb{E}\left[f(X^x_{\tau_D})\exp\left(-\int_0^{\tau_D} k(X^x_s)ds\right)\right. \\ &\quad \left. + \int_0^{\tau_D} g(X^x_s)\exp\left(-\int_0^s k(X^x_\theta)d\theta\right)ds\right]. \qquad (6.61) \end{aligned}$$

In particular, the solution to the boundary value problem is unique in $C(\overline{D}) \cap C^2(D)$.

Proof. Fix $x \in D$. According to Itô's formula and the elliptic equation for u, we have

$$\begin{aligned} u(X^x_{\tau_D\wedge t}) &= u(x) + \sum_{i=1}^n \sum_{k=1}^d \int_0^{\tau_D\wedge t} \frac{\partial u}{\partial x^i}(X^x_s)\sigma^i_k(X^x_s)dB^k_s + \int_0^{\tau_D\wedge t}(\mathcal{A}u)(X^x_s)ds \\ &= u(x) + \sum_{i=1}^n \sum_{k=1}^d \int_0^{\tau_D\wedge t} \frac{\partial u}{\partial x^i}(X^x_s)\sigma^i_k(X^x_s)dB^k_s \\ &\quad + \int_0^{\tau_D\wedge t} k(X^x_s)u(X^x_s)ds - \int_0^{\tau_D\wedge t} g(X^x_s)ds. \end{aligned}$$

By further applying the integration by parts formula to the process

$$u(X^x_{\tau_D\wedge t})\cdot\exp(-\int_0^{\tau_D\wedge t} k(X^x_s)ds), \quad t \geqslant 0,$$

we conclude that the process

$$\begin{aligned} M_t &\triangleq u(X^x_{\tau_D\wedge t})\cdot\exp\left(-\int_0^{\tau_D\wedge t} k(X^x_s)ds\right) + \int_0^{\tau_D\wedge t} g(X^x_s)\exp\left(-\int_0^s k(X^x_\theta)d\theta\right)ds \\ &= u(x) + \sum_{i=1}^n \sum_{k=1}^d \int_0^{\tau_D\wedge t} \frac{\partial u}{\partial x^i}(X^x_s)\sigma^i_k(X^x_s)\exp\left(-\int_0^s k(X^x_\theta)d\theta\right)dB^k_s \end{aligned}$$

is a continuous local martingale. Moreover, from continuity we see that

$$|M_t| \leqslant C(1+\tau_D), \quad \forall t \geqslant 0.$$

According to the assumption (6.60), we conclude that M_t is of class (DL). In particular, M_t is a martingale. Indeed, the same reason shows that M_t is a uniformly integrable martingale. Therefore,

$$\mathbb{E}[M_0] = \mathbb{E}[M_\infty],$$

which yields the desired formula (6.61). □

One might wonder when the condition (6.60) holds. Here is a simple sufficient condition.

Proposition 6.7. *Suppose that for some* $1 \leqslant i \leqslant n$, *we have*

$$\inf_{x\in\overline{D}} a^{ii}(x) > 0.$$

Then (6.60) holds for every $x \in D$.

Proof. Define

$$p \triangleq \inf_{x\in\overline{D}} a^{ii}(x),\ q \triangleq \sup_{x\in\overline{D}} |b(x)|,\ r \triangleq \inf_{x\in\overline{D}} x^i.$$

Let $\lambda > 2q/p$, and define

$$h(x) \triangleq -\mathrm{e}^{\lambda x^i}, \quad x \in \mathbb{R}^n.$$

Then we have

$$-(\mathcal{A}h)(x) = \mathrm{e}^{\lambda x^i} \cdot \left(\frac{1}{2}\lambda^2 a^{ii}(x) + \lambda b^i(x)\right) \geqslant \frac{1}{2}\lambda p \mathrm{e}^{\lambda r}\left(\lambda - \frac{2q}{p}\right) =: \gamma > 0,$$

for every $x \in D$. On the other hand, for fixed $x \in D$, we know from the martingale characterization that the process $h(X^x_{\tau_D\wedge t}) - h(x) - \int_0^{\tau_D\wedge t}(\mathcal{A}h)(X^x_s)ds$ is a martingale. Therefore,

$$\begin{aligned}\mathbb{E}[h(X^x_{\tau_D\wedge t})] &= h(x) + \mathbb{E}\left[\int_0^{\tau_D\wedge t}(\mathcal{A}h)(X^x_s)ds\right] \\ &\leqslant h(x) - \gamma\mathbb{E}[\tau_D \wedge t],\end{aligned}$$

which implies that

$$\mathbb{E}[\tau_D \wedge t] \leqslant \frac{h(x) - \mathbb{E}[h(X^x_{\tau_D\wedge t})]}{\gamma} \leqslant \frac{2\sup_{x\in\overline{D}}|h(x)|}{\gamma}.$$

Therefore, by letting $t \to \infty$, we conclude that (6.60) holds. □

Next we turn to the parabolic problem. The idea of the analysis is similar to the elliptic problem.

We fix an arbitrary $T > 0$. Let $f:\ \mathbb{R}^n \to \mathbb{R}^1$, $k:\ [0,T]\times\mathbb{R}^n \to [0,\infty)$ and $g:\ [0,T]\times\mathbb{R}^n \to [0,\infty)$ be continuous functions. We consider the following backward Cauchy problem: find a function $v \in C([0,T]\times\mathbb{R}^n)\cap C^{1,2}([0,T)\times\mathbb{R}^n)$ (continuously differentiable in t and twice continuously differentiable in x), such that

$$\begin{cases}-\frac{\partial v}{\partial t} + k\cdot v = \mathcal{A}v + g, & (t,x)\in[0,T)\times\mathbb{R}^n,\\ v(T,x) = f(x), & x\in\mathbb{R}^n.\end{cases} \tag{6.62}$$

Here $\mathcal{A}$ acts on v by differentiating with respect to the spatial variable. Again we are not concerned with existence which is contained in the standard parabolic theory, and we are looking for a stochastic representation for the existing solution.

Suppose that f, g satisfy the following polynomial growth condition:

$$|f(x)| \leqslant C(1+|x|^\mu), \quad \sup_{0\leqslant t\leqslant T}|g(t,x)| \leqslant C(1+|x|^\mu), \quad \forall x\in\mathbb{R}^n, \tag{6.63}$$

for some $C, \mu > 0$. Then we have the following result which is known as the *Feynman-Kac formula*.

Theorem 6.17. *Suppose that $v \in C([0,T]\times\mathbb{R}^n)\cap C^{1,2}([0,T)\times\mathbb{R}^n)$ is a solution to the backward Cauchy problem (6.62) which satisfies the polynomial growth condition:*

$$\sup_{0\leqslant t\leqslant T}|v(t,x)|\leqslant K(1+|x|^{\lambda}),\quad \forall x\in\mathbb{R}^n,$$

for some $K,\lambda>0$. Then v has a representation

$$\begin{aligned} v(t,x) &= \mathbb{E}\bigg[f(X^x_{T-t})\exp\left(-\int_0^{T-t}k(t+s,X^x_s)ds\right) \\ &\quad +\int_0^{T-t}g(t+s,X^x_s)\exp\left(-\int_0^s k(t+\theta,X^x_\theta)d\theta\right)ds\bigg] \end{aligned}$$

In particular, the solution is unique in the space of $C([0,T]\times\mathbb{R}^n)\cap C^{1,2}([0,T)\times\mathbb{R}^n)$-functions which satisfy the polynomial growth condition.

Proof. The proof is essentially the same is the one of Theorem 6.16. Given $0\leqslant t\leqslant T$, by applying Itô's formula to the process together with the parabolic equation for v,

$$Y_s\triangleq v(t+s,X^x_s)\cdot\exp\left(-\int_0^s k(t+\theta,X^x_\theta)d\theta\right),\quad 0\leqslant s\leqslant T-t,$$

we arrive at

$$\begin{aligned} dY_s &= -g(t+s,X^x_s)\cdot\exp\left(-\int_0^s k(t+\theta,X^x_\theta)d\theta\right)ds \\ &\quad +\sum_{i=1}^n\sum_{k=1}^d\exp\left(-\int_0^s k(t+\theta,X^x_\theta)d\theta\right)\cdot\frac{\partial v}{\partial x^i}(t+s,X^x_s)\sigma^i_k(X^x_s)dB^k_s \end{aligned}$$

for $0\leqslant s\leqslant T-t$. On the other hand, since σ,b satisfy the linear growth condition which is a consequence of Lipschitz continuity, we see from the BDG inequalities and Gronwall's inequality that

$$\mathbb{E}\left[\sup_{0\leqslant s\leqslant t}|X^x_s|^p\right]<\infty,\quad\forall x\in\mathbb{R}^n,t\geqslant 0,p\geqslant 2.$$

In particular, together with the polynomial growth condition for g and v, we conclude that the local martingale

$$Y_s-v(t,x)+\int_0^s g(t+\theta,X^x_\theta)\cdot\exp\left(-\int_0^\theta k(t+r,X^x_r)dr\right)d\theta,\quad 0\leqslant s\leqslant T-t,$$

is indeed a martingale (one might show that this local martingale is of class (DL)). Therefore, we arrive at

$$v(t,x)=\mathbb{E}\left[Y_{T-t}+\int_0^{T-t}g(t+s,X^x_s)\exp\left(-\int_0^s k(t+\theta,X^x_\theta)d\theta\right)ds\right],$$

which yields the desired formula. □

We can see from the proof of Theorem 6.17 that the backward Cauchy problem is a more natural one to consider from the probabilistic point of view. Of course one can consider the following classical forward Cauchy problem, which is indeed a direct consequence of the backward case.

Corollary 6.2. *Let* $f:\ \mathbb{R}^n \to \mathbb{R}^1$, $k:\ [0,\infty)\times\mathbb{R}^n \to [0,\infty)$ *and* $g:\ [0,\infty)\times\mathbb{R}^n \to \mathbb{R}^1$ *be continuous functions such that* f,g *have local polynomial growth in the sense that for each* $T>0$, *there exists* $C,\mu>0$ *such that (6.63) holds. Suppose that* $u \in C([0,\infty)\times\mathbb{R}^n)\times C^{1,2}((0,\infty)\times\mathbb{R}^n)$ *is a solution to the Cauchy problem*

$$\begin{cases} \frac{\partial u}{\partial t} + k\cdot u = \mathcal{A}u + g, & (t,x)\in(0,\infty)\times\mathbb{R}^n, \\ u(0,x) = f(x), & x\in\mathbb{R}^n, \end{cases} \tag{6.64}$$

which has local polynomial growth in the same sense. Then u *is given by*

$$\begin{aligned} u(t,x) &= \mathbb{E}\bigg[f(X_t^x)\exp\left(-\int_0^t k(t-s,X_s^x)ds\right) \\ &\quad + \int_0^t g(t-s,X_s^x)\exp\left(-\int_0^s k(t-\theta,X_\theta^x)d\theta\right)ds\bigg]. \end{aligned} \tag{6.65}$$

In particular, the solution is unique in the space of $C([0,\infty)\times\mathbb{R}^n)\cap C^{1,2}((0,\infty)\times\mathbb{R}^n)$*-functions which have local polynomial growth.*

Proof. For fixed $T>0$, define

$$v(t,x) \triangleq u(T-t,x), \quad 0\leqslant t\leqslant T.$$

Then v solves the backward Cauchy problem. The result follows from applying Theorem 6.62 to v. □

Remark 6.12. A nice consequence of Theorem 6.16 (the elliptic problem) is a maximum principle: suppose that $g\geqslant 0, f\geqslant 0,$ then $u\geqslant 0.$ Similar result holds for the parabolic problem (backward and forward).

In the end, we give a brief answer (not entirely rigorous) to the two fundamental questions that we raised in the introduction of this section.

(1) According to the martingale characterization for the SDE (6.58) and some boundedness estimates, we know that the process

$$f(X_t^x) - f(x) - \int_0^t (\mathcal{A}f)(X_s^x)ds$$

is a martingale. Therefore, it is very natural to expect that

$$\frac{\mathbb{E}[f(X_t^x)] - f(x)}{t} = \frac{1}{t}\int_0^t (\mathcal{A}f)(X_s^x)ds \to (\mathcal{A}f)(x)$$

as $t\downarrow 0$.

(2) Let $p(t,x,y)$ be the fundamental solution (if it exists) to the parabolic equation $\frac{\partial u}{\partial t} - \mathcal{A}u = 0$. In other words, $p(t,x,y)$ satisfies

$$\begin{cases} \frac{\partial p}{\partial t} - (\mathcal{A}p) = 0, & t > 0, \\ p(0,x,y) = \delta_x(y), \end{cases}$$

where $\mathcal{A}$ acts on p by differentiating with respect to the x variable. It follows that for every $f: \mathbb{R}^n \to \mathbb{R}^1$, the function

$$u(t,x) \triangleq \int_{\mathbb{R}^n} p(t,x,y) f(y) dy$$

solves the forward Cauchy problem (6.64) ($k = 0, g = 0$) with initial condition given by f. According to Corollary 6.2, we know that

$$u(t,x) = \mathbb{E}[f(X_t^x)].$$

This implies that $p(t,x,y)$ has to be given by

$$p(t,x,y) = \mathbb{P}(X_t^x \in dy)/dy.$$

Remark 6.13. There are still technical gaps to fill in order to make the previous argument work. Even so, a rather subtle point is that it is not at all clear that if we define $u(t,x)$ by the right hand side of (6.65) (respectively, define $p(t,x,y) \triangleq \mathbb{P}(X_t^x \in dy)/dy$ if it exists), then $u(t,x)$ (respectively, $p(t,x,y)$) solves the forward Cauchy problem (respectively, defines a fundamental solution). This philosophy of proving existence was not fully explored because it turns out to be not as efficient as traditional PDE methods in general. The elegance of the stochastic representation lies in the fact that once a solution exists, it has to be in the neat probabilistic form that we have seen here, which gives us solid intuition about its structure and probabilistic ways to study its properties. On the practical side, it enables us to simulate the solution to the PDE from a probabilistic point of view (the so-called *Monte Carlo method*), which proves to be rather efficient and successful.

References

[1] C. Dellacherie, *Capacités et processus stochastiques*, 1972.

[2] C. Dellacherie and P.A. Meyer, *Probabilités et potentiel*, 1975.

[3] P.K. Friz and N.B. Victoir, *Multidimensional stochastic processes as rough paths,* Cambridge University Press, 2010.

[4] N. Ikeda and S. Watanabe, *Stochastic differential equations and diffusion processes,* 2nd edition, North-Holland, 1989.

[5] I.A. Karatzas and S.E. Shreve, *Brownian motion and stochastic calculus,* 2nd edition, Springer-Verlag,1991.

[6] S. Orey and S. J. Taylor, How often on a Brownian path does the law of iterated logarithm fail?, *Proc. London Math. Soc.* 28 (3): 174–192, 1974

[7] K.R. Parthasarathy: *Probability measures on metric spaces*, American Mathematical Society, 1967.

[8] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, 3rd edition, Springer, 2004.

[9] L.C.G. Rogers and D. Williams, *Diffusions, Markov processes and martingales*, Volume 1 and 2, 2nd edition, John Wiley & Sons, 1994.

[10] I. Shigekawa, *Stochastic analysis*, American Mathematical Society, 2004.